Introducing a Novel Method for Simulating Stochastic Movement and Occupancy in Residential Spaces Using Time-Use Survey Data

Ruiming Zhang[1,2,3,4], Tongyu Zhou[1*], Hong Ye[2,3,4,5], Jo Darkwa[6]

[1] Department of Architecture and Built Environment, University of Nottingham Ningbo China, Ningbo 315100, China

[2] Key Laboratory of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China

[3] Xiamen Key Laboratory of Urban Metabolism, Xiamen 361021, China

[4] University of Chinese Academy of Sciences, Beijing 100049, China

[5]CAS Haixi Industrial Technology Innovation Center in Beilun, Ningbo 315800, China

[6] Department of Architecture and Built Environment, University of Nottingham, University Park, Nottingham, NG7 2RD, United Kingdom

*Tongyu Zhou: tongyu.zhou@nottingham.edu.cn

Abstract: In the context of growing concerns over energy consumption and sustainability, accurate modelling of occupancy patterns within residential buildings is critical. In this study, a novel stochastic occupancy model is introduced for simulating human behaviour within residential buildings by employing Time Use Survey (TUS) data and utilising Markov chains and probabilistic sampling algorithms. The novelty of this research lies in its approach to represent the dynamic nature of occupancy across different functional spaces and age groups, a gap not yet adequately addressed in existing studies. The model's accuracy is ascertained through ten-fold cross-validation, achieving an average R^2 value of 0.91 across key functional rooms (bedroom, bathroom, kitchen, living room), indicating a high degree of precision. Applied to a case study of a two-story detached house in the UK, the model effectively reflects varied behaviour patterns and room occupancy among different age groups. For instance, the average daily appliance energy consumption for occupants aged 8-14 ranged from 0 to 3.77 kWh (median 1.71 kWh), for ages 15-64 from 0 to 4.93 kWh (median 2.61 kWh), and for over 65 from 0.87 to 5.65 kWh (median 3.60 kWh). This model, with its scalability and accuracy in capturing the inherent randomness of human behaviour, is a valuable tool for improving energy consumption simulations and contributing to sustainable residential building design and management.

Keywords: Stochastic Occupancy Model; Residential Energy Consumption; Time Use Survey; Markov Chains; Occupant Behaviour

1. Introduction

1.1 Background

Building energy consumption is a key parameter by which the performance of the indoor environment can be assessed and improved, and occupant behaviours are one of the drivers causing difference in building energy consumption among buildings with the same function and in similar climatic conditions [1]. For this reason, the United States Department of Energy (DOE) issued Standard Building Operating Conditions (SBOC) in 1979 and defined occupancy schedules for 14 building occupancy types [2, 3]. In 1989, the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) issued the first standard occupancy schedules for nine building occupancy types based on the DOE SBOC standards (Fig.1 a)), which were revised and refined in 2004 and 2013. Additionally, international standards such as prEN16798-1 and ISO 17772-1:2017 have also addressed occupant schedules for energy calculations (Fig. 1 b)) [4].

However, these occupancy schedules were based on statistical results, and the descriptions of occupancy were static and simplified [5]. They fail to capture the temporal and spatial stochastic nature of occupancy, which is critical for accurately assessing energy consumption in residential buildings. This limitation is particularly evident when considering the diversity of occupancy patterns across different regions. For instance, research conducted by Mitra et, al. showed that Canadians are on average 6.6% less actively occupied in residential buildings than British [6]. Moreover, even within a single country, occupancy patterns can vary widely. In China, for example, the time spent in living rooms ranges from 10-11 hours in Beijing to as little as 5.4 hours in Yinchuan and 4.8 hours in Chengdu [7]. These variations underscore the importance of developing occupancy models that can more accurately reflect the diverse and dynamic nature of building occupancy.

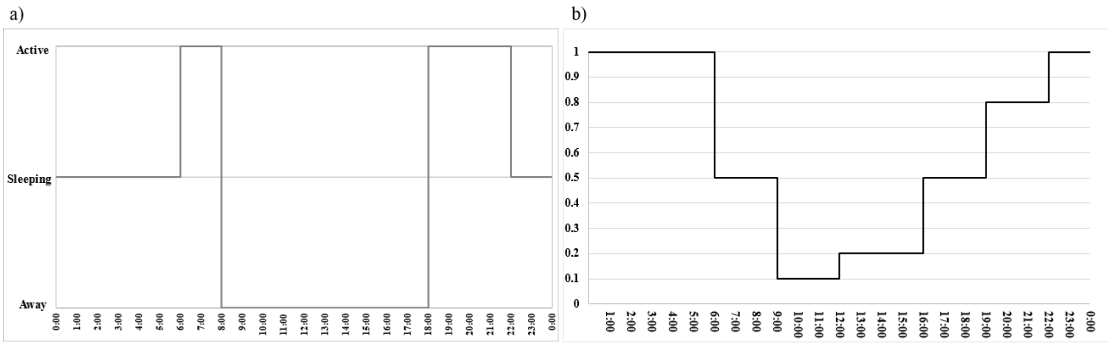

Fig. 1 a) An example of standard ASHRAE schedule for residential buildings [8] (Source: Section 13 of ASHRAE standard 90.1-1989); b) Hourly residential building pattern for energy calculation

The behaviour of occupants is one of the most important sources of uncertainty in predicting building energy use through modelling procedures [1]. Given the diversity in individual behaviours, applying such oversimplified schedules in building simulation often result in a significant discrepancy between the simulated results and actual energy consumption. For example, Duarte et, al. carried out a study in 2013 which showed a 46% difference between private office and ASHRAE reference occupancy rates [9]. Hong et, al. found that for a typical single-person office room, different working styles could result in energy consumption varying from 50% less to 90% more compared to the standard or reference working style[1].

Many studies have shown that the use of occupancy information can save around 10% - 40% of a building's energy consumption [10]. Erickson and Cerpa indicated HVAC control strategies with predictive and real-time occupancy monitoring via camera sensor networks have a potential energy saving of 20% [11]. Peng et, al. adaptation to local occupancy scenarios can save 20.3% of energy [12]. The expected energy use simulated using sensors to detect occupancy and sleep patterns in the home saves an average of 28% of energy compared to existing energy simulation methods [13].

1.2 Occupancy models

In recent years, researchers have done a number of studies to achieve accurate building occupancy estimation. Back in 2001, Macdonald and Strachan proposed using the Monte Carlo method to build a basic stochastic model as inputs to simulation tools [14]. Nowadays, some advanced models are proposed to randomly generate plausible building occupancy models. Most of the research focuses on the types of public buildings. For instance, Wang et, al. investigated the occupancy patterns in single-person offices within a large office building in San Francisco. Their study revealed that the vacancy intervals across the 35 single-person offices involved in the study were exponentially distributed. They demonstrated three typical occupancy models for these offices using non-homogeneous Poisson process simulations [15]. However, the findings, being limited to a single office building, raised questions about their universal applicability. Addressing this limitation, Page et al. introduced a more versatile approach using the Markov chain to simulate occupant presence. Their algorithm, implemented in Matlab, was adaptable to both residential and public buildings, enabling the generation of occupancy statuses (absence or presence) in various zones over different time series[16]. Further refining the focus on occupancy modelling, Wang et, al. proposed an innovative approach that included a basic movement module and an advanced event module. This model was able to simulate not only the spatial location of each occupant but also zone-level occupancy of the whole building [17]. Furthermore, a part of the research focused on the generation or prediction of typical standard occupancy models. Liang et, al. used data mining methods to learn and predict the occupancy schedule of a whole office building [18]. Happle et, al. used location-based services (LBS) data to create occupancy schedules of a retail or restaurant building use type in different cities and compared them with standard schedules [3]. These studies highlighted the potential of leveraging big data and advanced analytics in occupancy modelling.

While the majority of earlier occupancy studies concentrated on public buildings, there has been a notable shift towards residential occupancy models in recent times. This transition is characterized by the increasing use of national-level Time Use Survey data and the Markov chain as foundational methods in developing residential occupancy models. For instance, Richardson et, al. developed a stochastic occupancy model to describe the active or inactive states in the house using the first-order Markov–Chain technique [19]. Building on this, McKenna et, al. refined the approach with the first-order time-inhomogeneous Markov-chain technique. This optimization allowed for modelling four stochastic states of occupants (absent/present and active/inactive) within a household [20]. Buttitta and Finn applied the first-order Markov–Chain technique to generate high-time resolution occupancy models and used them as input parameters to calculate high-time resolution heating load in buildings[21]. In addition to traditional survey data, the use of monitoring data from household devices has emerged as a valuable source for occupancy modelling. Huchuk et, al. utilised real consumer longitudinal data from the connected thermostat devices to predict household occupancy using different methods, including Logistic regression, Markov model, Random Forest, Hidden Markov model and Recurrent neural network. Their findings indicated that the Random forest algorithm outperforms the other models [22]. Causone et al. took a different approach by collecting energy metering data and employing machine learning algorithms to infer occupant-related input data [23]. Similarly, Diao et, al. proposed to use direct energy consumption results and energy time use data to identify and classify occupant behaviour through unsupervised clustering. Their behavioural model offered more accurate and reliable predictions than the ASHRAE standard schedule [24]. Additionally, Sayed et, al. developed a simple and effective image conversion technique for predicting occupancy [25]. These diverse methodologies highlight the evolving complexity and precision in residential occupancy modelling. A detailed overview of the research methods applied to different types of buildings is shown in Table 1 below.

Table 1 An overview of the occupancy model generation methods

Forecast object	Method/ Algorithm	Data source	Description	Ref.
Commercial buildings	Non-homogeneous Poisson process model	Infrared sensor	Proposed statistical properties of single-person office occupancy	[15]
	Markov Chain Model	Movement sensor	Generated the occupancy status (absence or presence) of occupants in different time series in the zones	[16]
	Homogeneous Markov chain	Experience	Modelled the location of occupancy and the building's zone-level occupancy	[17]
	Inhomogeneous Markov chain	Real-Time Locating System (RTLS)	Provided an adaptive probabilistic occupancy prediction model capturing the actual behaviour of open office occupants and zone-level occupants with high accuracy	[26]
	Inhomogeneous Markov chain	Wireless camera	Offered two stochastic building occupancy models for multi-residential single-area and multi-area scenarios respectively	[27]

	Data mining	Sensors	Provided a building occupancy schedule available in most office buildings	[18]
	Generative Adversarial Network (GAN)	Camera	Introduced methods to build occupancy model without prior assumptions	[28]
	Feature scaled extreme learning machine (FS-ELM) algorithm	CO_2 concentration data	Developed an occupancy simulator based on a discrete-time dynamic model of real-time carbon dioxide concentration measurements	[29]
	Statistical methods	Switch lighting equipment data	Determined five typical occupancy patterns through analysis of 200 open-plan offices	[30]
	The S-curve method and the probabilistic methods	Questionnaire	Proposed prediction formulas of occurrence and frequency for activities inside and outside the office during the workday	[31]
	Adaptive neural-fuzzy inference system (ANFIS) model	Sensor monitoring	Estimated non-residential building occupancy	[32]
Retail or restaurant building	Statistical method	Location-based services (LBS) data	Created a data-driven situation-specific and representative occupancy schedule for different building use types	[3]
Airport building	The Bayesian model	Wi-Fi IPS data	Predicted high-resolution occupancy of the airport	[33]
Large exhibition hall	Recurrent neural network (RNN) model with long short-term memory units (LSTM)	Image sensors and counting devices	Predicted short/long-term real-time occupancy in exhibition events	[34]
Laboratory	The auto-regressive hidden Markov model (ARHMM)	Wireless sensor network	Estimated the number of occupants in the laboratory	[35]
Residential buildings	Probabilistic model and the Hierarchical clustering algorithm	TUS and Household Budget Survey (HBS)	Identified seven significant occupancy schedules and reconstructed individual daily and annual occupancy	[36]
	Markov Chain Monte Carlo (MCMC) technique	Time-Use Survey (TUS)	Generated the stochastic occupancy (active/inactive) in the house	[19]
	Markov Chain Monte Carlo	Time-Use Survey (TUS)	Modelled occupant's state (absent/present and	[20]

	(MCMC) technique		active/inactive) in the house	
	A new Markov model	Passive infrared sensors	Predicted short-term occupancy in the buildings	[37]
	Machine learning algorithms	Smart meters	Generated standardized occupancy profiles using the electricity records from smart meters	[23]
	Logistic regression; Markov model; Random forest; Hidden Markov model; Recurrent neural network	Connected thermostats	Generated household occupancy prediction models; Random forest algorithm outperforms other models	[22]
	Unsupervised clustering; First-order inhomogeneous Markov chain	American Time Use Survey (ATUS)	Identified ten occupant behaviour model	[24]
	First-order Markov–Chain technique	TUS data	Generated high-temporal resolution occupancy model	[21]
	Deep learning	Sensors	Developed a method to detect building occupancy	[25]
	Generic data-driven framework (including clustering and changepoint detection (CPD))	Home energy management system (HEMS)	Explored occupant patterns and presence probabilities for a set of residential buildings	[38]
	Semi-Markov chain mode	Smart thermostat data	Modelled annual occupancy schedules for urban-scale	[39]

With the increasing emphasis on the study of human behaviour in buildings, many research efforts are being made to accurately capture occupancy patterns and behaviours. We found that some studies for building occupancy schedule use occupancy sensors, cameras, the passive infrared (PI R) sensor, radio frequency identification (RFID) instruments or other devices that can be used to collect occupancy data to achieve the purpose of obtaining occupied data and for occupancy prediction [10, 40-42]. However, the data collected by this method is limited to a small sample size and is difficult to apply to the entire residential building due to the privacy issues involved for the occupants and the difficulty of installing the sensors without disturbing the occupants' activities [7].

As occupants of residential buildings often refuse direct data collection by researchers or research institutions entering their homes, several studies attempt to investigate behaviours through indirect data sources. Fortunately, several countries conduct regular national-wide time use surveys to gather information about household time use, including time spent and appliance usage at home. Since 1996, The Japan Bureau of Statistics has conducted a time use survey in every five years [43]. American Time Use Survey, UK Time Use Survey (TUS) and other time use surveys collect the amount of time people spend sleeping, working at home, preparing food and other activities. This time-use data helps to understand household activities and can be used to roughly determine the locations of individuals in different rooms within the residential buildings. Therefore, this type of data source aids in the development of more accurate occupancy schedules for building simulations.

1.3 Research aim and objectives

In summary, several probabilistic and data-driven approaches to assessing occupancy levels of buildings have been established in recent years. However, current occupancy forecasting methods have limitations. First of all, most studies focus on public buildings, while there are relatively few studies on residential buildings. For those who focus on residential buildings, few have considered the occupancy of different functional rooms in residential buildings. Given the prevailing use of static schedules in building energy modelling for occupancy and the predominant focus on public buildings like offices in existing studies of occupancy schedules, this research aims to address the need for a stochastic occupancy model in residential buildings. By utilising extensive real data from TUS, this model captures the randomness of residents' behaviours in residential buildings and dynamically quantifies the probabilities of different groups of people being present in various rooms at different times. It enables direct integration into building simulations, thereby enhancing the accuracy of simulation outcomes to closely align with real-world scenarios.

2. Methodology

The methodology employed in this study consists of three primary components, and an overview of the methodology is given in Fig. 2. The first part involves analytical processing of the TUS data. This data is carefully analysed and processed to extract relevant information about residents' activities, their durations, and the rooms they occupy within residential buildings. Comprehensive examination of the TUS data provides insights into occupant patterns and behaviours.

The second component utilises the extracted TUS data to construct probabilistic transfer matrices and generate Markov chains. These matrices capture the transition probabilities of occupants moving from one room to another within a residential building. By leveraging these transfer matrices, the stochastic nature of occupancy patterns over time can be simulated. This enables the modelling of dynamic movements of residents and their presence in different rooms at different time intervals.

The third component involves the use of probabilistic sampling models. These models enable the prediction of room occupancy within a residential sample and the generation of occupant movements between various spaces. By incorporating the probabilistic sampling method, the inherent uncertainty and randomness in occupant behaviour are accommodated.

Additionally, to validate the accuracy and reliability of our methodology, we employed a rigorous validation process using 10-fold cross-validation. The dataset was divided into ten subsets of approximately equal size. In each iteration, nine subsets were used for training the occupancy models, while the remaining subset was held out for testing. This process was repeated ten times, with each subset serving as the test set once.

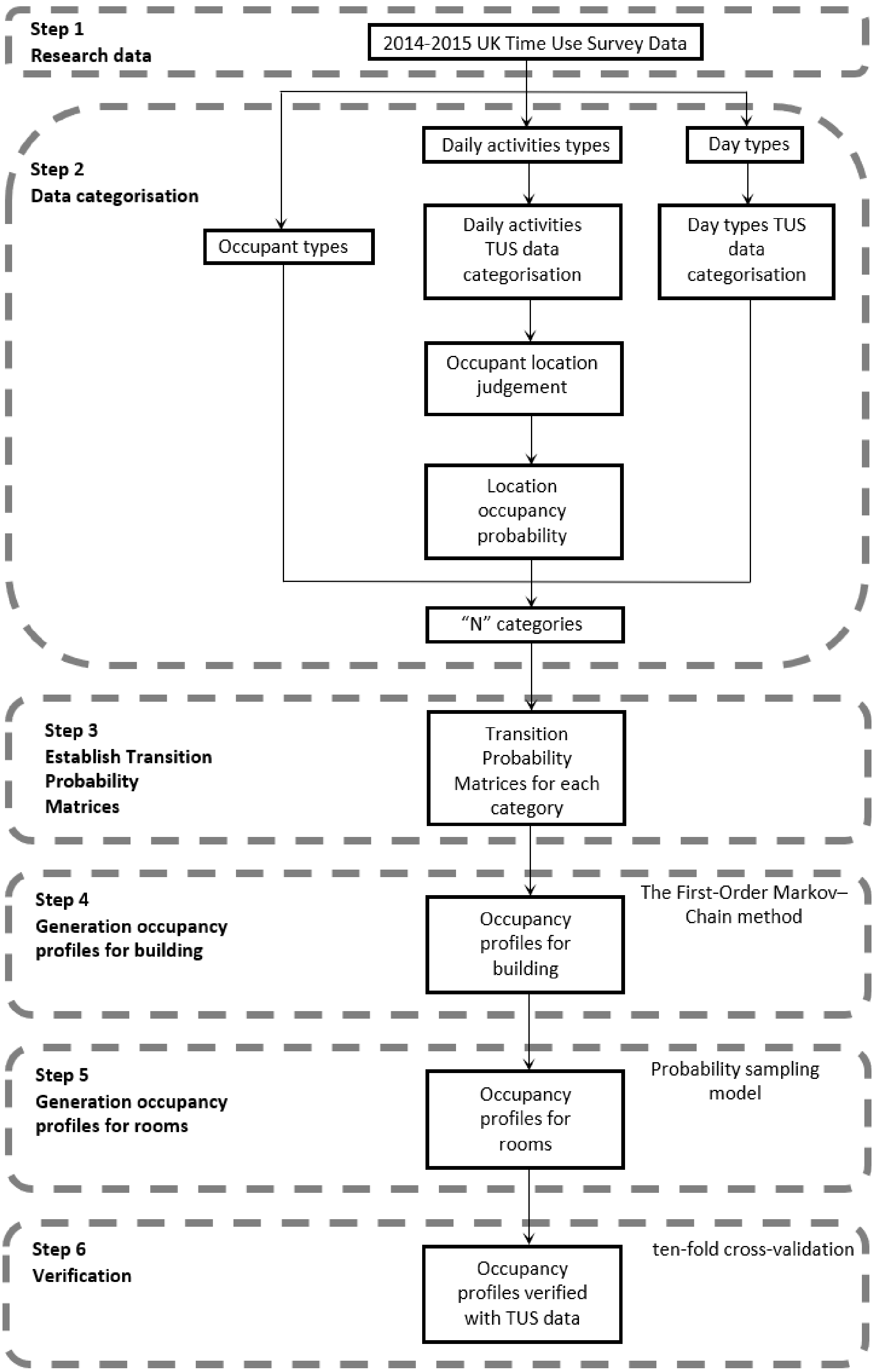

Fig. 2 Methodology flow chart

2.1 Data Description and Processing

To capture the stochastic nature of occupancy patterns in various rooms, it is essential to have a database that records the activities of each occupant with fine time granularity, such as ten-minute intervals. Additionally, a sufficient sample size is crucial to ensure the representativeness and reliability of the data. TUS data sets from various countries are ideal for this purpose as they provide detailed and comprehensive activity records. This study drew on data extracted from the UK TUS conducted in 2014–2015. The data can be downloaded from the Economic and Social Research Council (ESRC) website [44]. This large-scale household-level survey, which examined how people used their time, was conducted by the National Centre for Social Research and the Northern Ireland Statistics and Research Agency on behalf of the University of Oxford’s Centre for Time Use Research. The sample for the UK TUS comprised households from England, Scotland, Wales and Northern Ireland. A total of 4,238 family interviews were conducted with 10,208 eligible respondents. These respondents completed 16,550 records of their daily routine, of which 16,533 contained valid data on their daily behaviours. The data compiled include the participants’ basic information, their locations and their activities. Each participant aged 8 years and above was provided with two 24-hour schedules and instructed to record their activities at 10-minute intervals.

The analysis of the TUS data reveals a spectrum of twelve typical activities that characterise the day-to-day life of a residential building occupant. These include: sleeping, eating, personal care, employment-related activities, studying, household and family care activities, voluntary work and meetings, social life and entertainment, sports and outdoor activities, hobbies and computing activities, mass media activities and travelling. We assume that these activities occur in one of the functional rooms within the building, such as the kitchen (including the dining room), bathroom, bedroom, living room, or occur outside of this building. To elaborate, during a specific 10-minute interval, an individual’s change in location can be classified into one of three types: remaining static, transitioning from one room to another within the building, or moving from an outdoor location to an indoor one. For instance, when the occupant is engaged in eating or cooking, it is associated with a change in location to the kitchen from either another room within the building or from an outside location. Similarly, personal care activities correspond to the occupant's change of location from either inside or outside the building to the bathroom; sleep, employment-related and study activities correspond to the occupant's change of location from either inside or outside the building to the bedroom, and other activities correspond to the change of location from either inside or outside the building to the living room. Table 2 presents the typical activities of occupants and their corresponding functional rooms.

Table 2 Examples of typical activities and corresponding functional rooms

Activities	**Corresponding rooms**
Sleeping	Bedroom
Eating	Kitchen (including the dining room)
Personal care	Bathroom
Employment-related activities	Living room
Studying	Living room
Household and family care activities	Living room
Voluntary work and meetings	Living room
Social life and entertainment	Living room
Sports and outdoor activities	Outdoor
Hobbies and computing activities	Living room

Mass media activities	Living room
Travelling	Outdoor

Predicting energy consumption patterns in residential buildings presents complex challenge due to the different behaviours of households, which are influenced by many factors [45, 46]. Therefore, this study considers factors that influence occupancy schedules, specifically focusing on the age of occupants and differentiating between weekdays and weekends. The amount of valid data on occupants is shown in Table 3, and the TUS data is classified according to age groups: 8-14 years, 15-64 years and 65 years and above. This classification is in line with the standardised statistical breakdown of the UK's age distribution from 2011 to 2021 as summarized by O'Neill[47].

Table 3 Classification of sample respondents' basic information

Background	Groups	Description	Day Type	Frequency	Percentage (%)
Age	Group 1	8-14 years	Weekday	1016	6.15
			Weekend	559	3.38
	Group 2	15-64 years	Weekday	7514	45.45
			Weekend	4088	24.73
	Group 3	65 years and over	Weekday	2127	12.87
			Weekend	1229	7.43

2.2 The First-Order Markov–Chain Monte Carlo method

Markov chain is a statistical method which has been widely used in building occupancy modelling [17, 26, 46]. In this study, we utilised the Markov chain approach to construct a profile of the overall occupancy within a residential building. It is used to model the sequences of an occupant's movements - specifically between being within or outside the building. We adopted the concept of 'stochastic movement', indicating that the transitions of occupants between inside and outside states are random and unpredictable. This movement of occupants forms the foundation of our occupancy profile. This hypothesis allows the transformation of occupants between inside and outside a residential building to be modelled as a Markov chain process. Thus, the occupant's subsequent occupancy status of the residential building depends only on his/her current state and a certain probability which is defined based on observed patterns in the data. To elaborate, in the First-Order Markov–Chain Monte Carlo method, the presence of occupants at a given time step only depends on the presence of occupants at the previous time step, taking into account factors such as the hour of the day and the day of the week [6]. The process begins with a defined starting state. At each time step, a random number within the interval [0,1] is generated. The transition of the occupant's state is then determined by comparing this random number with the probabilities indicated in the transition probability matrix, which links a given time step to a specific class [21]. This approach allows for the generation of data that accurately simulates the unpredictable nature of occupancy movements within residential spaces.

Markov chains are stochastic processes in the state space that undergo transitions from one state to another. It is described in Eq. 1 that the state of the next stage is only related to the state of the previous stage and the probability of state change.

$$Pr(X_{n+1} = x | X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = Pr(X_{n+1} = x | X_n = x_n) \qquad \text{Eq. 1}$$

For the First-Order Markov–Chain method, the previous state and the probabilities of the state change, which are stored in a "transition probability matrix (TPM) " [19]. Transfer probabilities between states with more than one step are more easily calculated by means of transfer matrices [37]. At any time step t, the probability transition matrix is denoted as [24]:

$$TransitionprobabilityMatrix_t = \begin{bmatrix} P_t^{11} & P_t^{12} & \cdots & P_t^{1n} \\ P_t^{21} & P_t^{22} & \cdots & P_t^{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_t^{n1} & P_t^{n2} & \cdots & P_t^{nn} \end{bmatrix} (t > 1) \qquad \text{Eq. 2}$$

Where, P_t^{ij} denotes the observed probability of transition from activity i to activity j at time step t. It is the conditional probability of activity j at time step t, given that activity i is at time step t - 1. The sum of each row in the matrix is equal to 1. P_t^{ij} is calculates as

$$P_t^{ij} = \frac{O_{ij}}{\sum_{k=1}^{m} O_{mk}} \qquad \text{Eq. 3}$$

where, O_{ij} is the observed number of transitions from state i to state j, O_{mk} is the observed number of transitions from state i to state k, and m is the number of possible states.

The TUS data operates on a ten-minute interval basis. This means that a full day's active occupancy time series data for a specific household comprises 144 states. Each state signifies the likelihood of occupants being present in the house during each ten-minute segment. Consequently, 144 conversion matrices were created to represent the transition of the occupancy situation in the household from time i to the next time i+1. The dimension of the transition probability matrix is 2 × 2, as shown in Fig. 3.

$$\begin{array}{rc} & \textbf{Next state (at time t+1)} \\ \textbf{Current state (at time t)} & \begin{array}{cc} \textbf{Absence} & \textbf{Presence} \end{array} \\ \begin{array}{r} \textbf{Absence} \\ \textbf{Presence} \end{array} & \begin{bmatrix} P_{abs,abs} & P_{abs,pre} \\ P_{pre,abs} & P_{pre,pre} \end{bmatrix} \end{array}$$

Fig. 3. Transition probability matrix at time t

2.3 Probability sampling model

In the previous section, the overall occupancy of a residential house was determined by the Markov chain method. However, to track the occupancy patterns within various functional rooms in the house, a more complex approach was necessary. The probability sampling model was developed primarily on the basis of a probability distribution map of historical presence which was calculated using TUS data. In this study, we use this model to generate occupancy of different function rooms in the household. Predictions are made by inverse sampling method during periods when individuals were present in the rooms. The algorithm for using probability sampling to predict presence is shown in Fig. 4. For each time step in the day that needs to be predicted, the occupancy status is determined by comparing the presence probability at that time step in the profile with a random number drawn from a uniform distribution. If the probability of occupancy surpasses the random number, the respective time step is considered as "occupied.". This method is implemented using MATLAB.

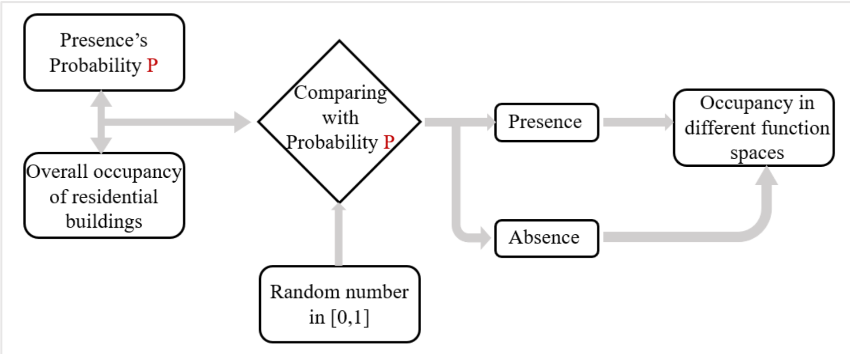

Fig. 4. Flowchart for the probability sampling model

2.4 Ten-fold cross-validation

In our study, we employ the ten-fold cross-validation method [48] to assess the performance across six groups defined by age and weekdays/weekends. Cross-validation is widely used as a statistical method to evaluate generalization performance of models. This method repeatedly divides the data into a training set and a test set for testing and training respectively. Unlike a single split of the dataset into training and test sets, which can lead to variability in model performance, cross-validation provides a more stable and thorough assessment. k-fold cross-validation is the most common cross-validation method, where k is usually 5 or 10. In k steps, a set of data is retained as a test set and the remaining data is used as a training set to train the model. The resulting k accuracy scores are averaged and the cross-validation accuracy is summarized into a performance metric for easy comparison [49].

In our case, we opt for ten-fold cross-validation. Fig.5 illustrates the procedural steps involved in the ten-fold cross-validation methodology. The dataset is divided into ten folds, with each fold containing 10% of the data as the validation set, while the remaining data serves as the training set. This approach allows a comprehensive evaluation of the performance and generalizability of the proposed methodology within each group. By evaluating the performance of the occupancy models across multiple iterations, we were able to assess their consistency and effectiveness in predicting room occupancy and occupant movements.

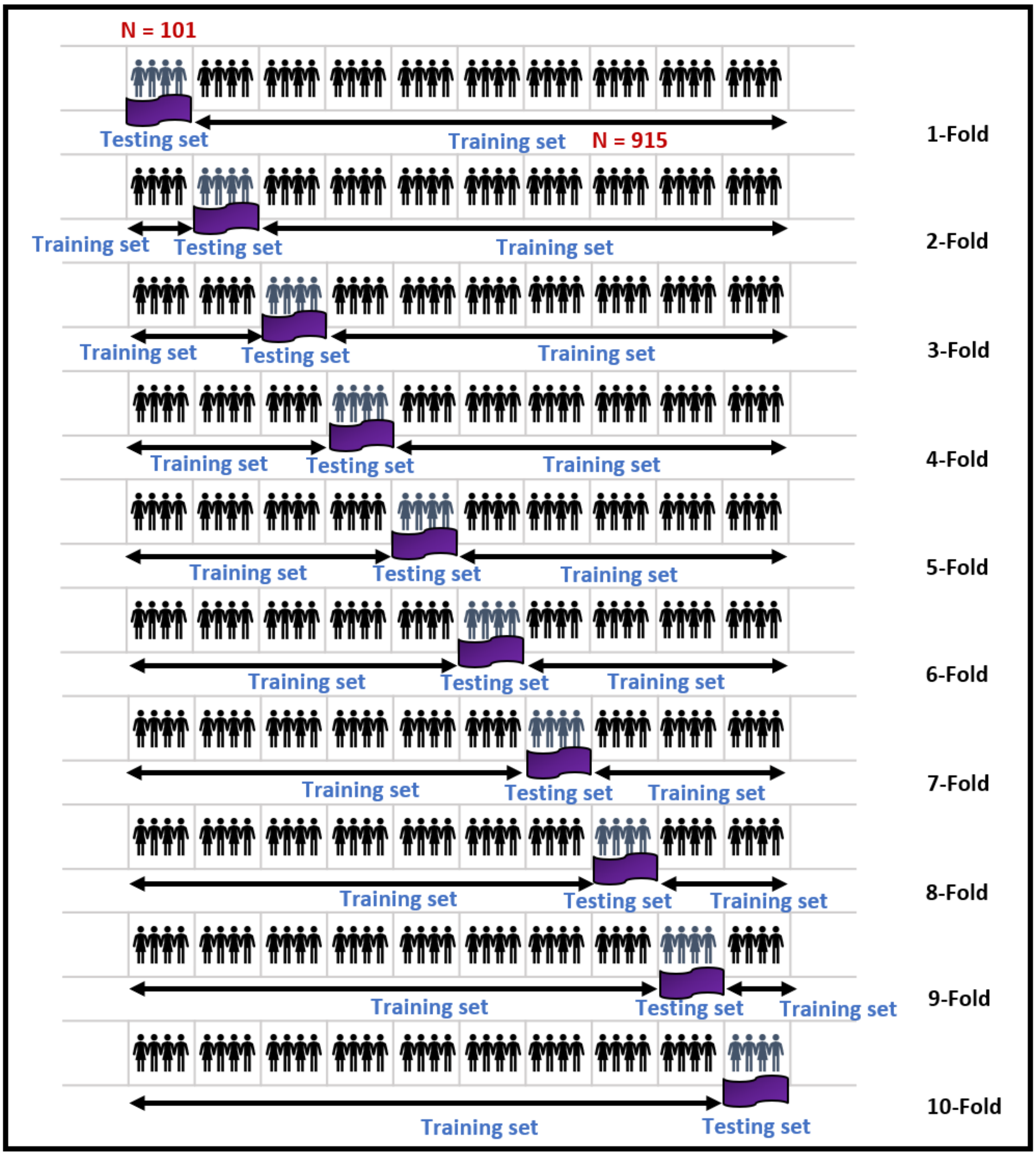

Fig. 5. The procedure of ten-fold cross-validation method for Group 1 (age 8-14 years) on weekday

Four evaluation indexes, R^2, root mean square error (RMSE), mean absolute error (MAE), and median absolute error (MedAE), are used to verify the proposed model, the definitions are described below.

The coefficient of determination (R^2) indicates how well the predicted values in a model compare to a scenario where only the mean is used. It is given by the formula for the sum of squared residuals as shown below:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(E_i - \widehat{E_i})^2}{\sum_{i=1}^{n}(E_i - \overline{E_i})^2} \qquad \text{Eq. 4}$$

$$\overline{E_i} = \frac{1}{n}\sum_{i=1}^{n} E_i \qquad \text{Eq. 5}$$

Where, E_i denotes the actual data of occupants, $\widehat{E_\iota}$ denotes the simulation results of occupants, $\overline{E_\iota}$ is the is the average of the actual data, n is the total number of those data.

RMSE is the mean of the square root of the error between the predicted value and the true value. It quantifies the typical size of the error in the predictions, expressed in absolute units [18], expressed in the following formula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(E_i - \widehat{E_\iota})^2}{n}} \qquad \text{Eq. 6}$$

For the perfect model, RMSE is equal to zero when the predicted value exactly matches the true value, the larger the error, the larger the value.

The Mean Absolute Error (MAE) is similar to the RMSE:

$$MAE = \frac{\sum_{i=1}^{n}|E_i - \widehat{E_\iota}|}{n} \qquad \text{Eq. 7}$$

For the perfect model, MAE is equal to zero when the predicted value exactly matches the true value; the larger the error, the larger the value.

The MedAE indicates whether the model has a systematic tendency to overestimate or underestimate. If the value of MedAE is 0, there is no population bias in the prediction method. The equation is as following [18]:

$$MedAE = median\left|E_i - \widehat{E_\iota}\right| \qquad \text{Eq. 8}$$

2.5 Estimating energy consumption associated with energy-related behaviours.

To contrast the standard ARSHRA occupancy schedule with the stochastic occupancy model put forth in this research, we undertook a comparative study, with a primary focus on energy-related behaviours and the resultant energy consumption inherent to each schedule. The concept of energy-related behaviours refers to those activities that involve the direct use of energy. In the context of a residential setting, these activities encompass the operation of various household appliances such as televisions, washing machines, computers, microwave ovens and the like. In essence, each of these appliances forms part of the daily energy consumption profile of a household, thereby establishing a clear link between occupancy patterns, activities, and energy usage.

The energy consumption of these appliances can be determined by [24]:

$$E_{appliance} = P_{appliance} \times Activityduration \qquad \text{Eq. 9}$$

where, $P_{appliance}$ is the equivalent power of an active application device. The power of common household appliances, as shown in Table 4 [50], were carefully selected from a comprehensive dataset provided by Generatorist. This dataset compiles power consumption data from a variety of authoritative sources, including government websites and well-known generator manufacturers such as Generac, Honda, and Yamaha, as well as major retailers like Lowe's, Home Depot, and Sears. These sources offer a mix of average and typical usage values, making the data robust and applicable to a wide range of residential buildings. In our research, we assume that when a room is occupied, the energy consumption can be estimated as the average energy usage of all appliances within that room. Specifically, in the context of a bedroom, we consider the scenario where occupants primarily use the room for sleeping, and hence, the only appliances accounted for are two electronic device chargers.

Table 4 The power consumption of household appliances[50]

KITCHEN APPLIANCES		BATHROOM APPLIANCES		LIVINGROOM APPLIANCES	
Household Appliances	**Watts**	**Household Appliances**	**Watts**	**Household Appliances**	**Watts**
Coffee Maker	1,000 W	**Bathroom Towel Heater**	60 W	**Apple TV**	3 W
Cooker Hood	20 W	**Clothes Dryer (Electric)**	5,400 W	**AV Receiver**	450 W
Dishwasher	1,500 W	**Curling Iron**	1,500 W	**Computer Monitor**	25 W
Electric Kettle	1,200 W	**Electric Shaver**	15 W	**Desktop Computer**	100 W
Electric Oven	2,150 W	**Extractor Fan**	12 W	**Guitar Amplifier**	20 W
Food Processor /Blender	400 W	**Hair Dryer**	1,250 W	**Home Internet Router**	5 W
Fryer	1,000 W	**Iron**	1,200 W	**Home Phone**	3 W
Induction Hob (Per Hob)	1,400 W	**Straightening Iron**	75 W	**Home Sound System**	95 W
Microwave	1,000 W	**Vacuum Cleaner**	200 W	**Laptop**	50 W
Percolator	800 W	**Washing Machine**	1,150 W	**Mi Box**	5 W
Pressure Cooker	700 W			**Monitor**	200 W
Refrigerator / Freezer	700 W			**Set Top Box**	27 W
Rice Cooker	200 W			**Television**	85 W
Sandwich Maker	700 W	**BEDROOM APPLIANCES**		**VCR / DVD Player**	100 W
Slow Cooker	160 W	**Charger (2)**	20 W	**Video Game System**	40 W
Steriliser	650 W				
Toaster	850 W				
Water Dispenser	100 W				
Water Filter & Cooler	70 W				
Wine Cooler (18 Bottles)	83 W				

3. Verification the stochastic occupancy model

This section discusses in detail the accuracy of this stochastic model in terms of ten-fold cross-validation. The generated occupancy data for group 1 at a house on weekdays is presented in Fig.6 as an example of validation.

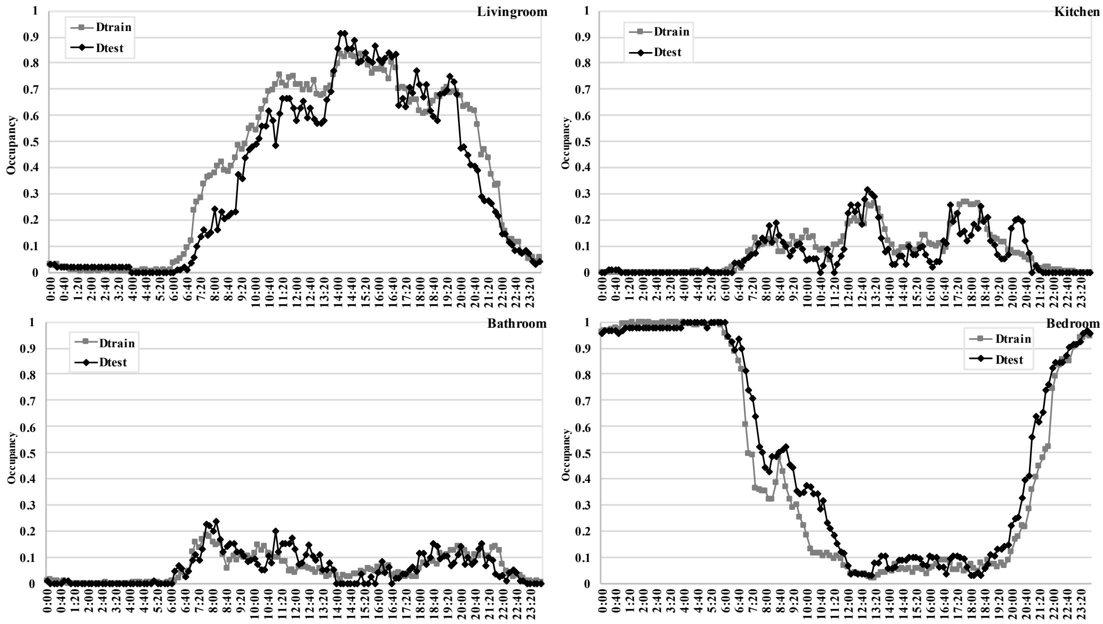

Fig. 6. Comparison of a generated stochastic occupancy and an actual occupancy.

There are a total of 1,016 sets of data for weekdays in group1, of which 915 sets are used to generate stochastic occupancy in rooms. The actual occupancy is derived from the remaining 101 sets. In Fig. 6, the lines illustrate the occupancy patterns of the different functional rooms. The black folded line represents a generated stochastic occupancy from the training set and the grey folded line denotes the actual occupancy from the test set. As Fig. 6 reveals, the training and test sets yield curves with nearly identical trends. This similarity suggests that the stochastic model generated on the training set is accurately capturing the underlying patterns and behaviours in the data.

Observing the data for different rooms, we see distinct patterns that reflect the occupancy of people aged 8-14 in real world. It is worth clarifying that the occupancy rate refers to the likelihood of an individual being in different rooms once they are already at home. In the living room, both the training and test sets show a peak in occupancy around 6:20. The highest occupancy is observed between 14:00 and 17:00, after which there is a significant drop at 20:00. This suggests that the living room is most frequently used in the mid to late afternoon. In the kitchen, both data sets indicate marked increases in occupancy during the morning, noon, and evening, respectively. This pattern likely corresponds with meal times, demonstrating the kitchen's role as a hub of activity at these key points in the day. The bathroom data presents a more random pattern, with occupancy fluctuating more unpredictably. In the bedroom, it shows a significant decline in occupancy starting around 6:00, with occupancy rates of less than 0.1 from noon to 20:00. After that, there is a sharp increase to nearly 100% occupancy and maintained between 23:00 to 6:00. This pattern aligns with typical sleeping hours, indicating that the bedroom is primarily used during the night.

In the performance evaluation of the proposed model, Table 5 shows the values of four evaluation indexes, derived from the ten-fold cross-validation of the whole dataset. The performance of the model was found to be satisfactory across all the rooms. Normally, a model with R^2 values greater or equal to 0.7 was considered good models [51]. This criterion suggests that the method proposed in this study exhibits a high degree of accuracy in modelling the stochasticity inherent in occupancy patterns. The RMSE, a measure of the model's prediction accuracy, yielded values close to zero across all folds. This suggests that the proposed model outperform a model generated solely on the mean of the TUS data. The MAE, a metric that quantifies the difference between the model's predictions and the actual data, also produced values near zero. This implies that the proposed model's error is minimal. The MedAE, another measure of prediction error, yielded values close to zero, further attesting to the model's excellent fit. In conclusion, these indexes collectively validate the accuracy of the proposed model.

Table 5 The value of four evaluation indexes in each function room

Livingroom	**R^2**	**RMSE**	**MAE**	**MedAE**
1-fold	0.98	0.07	0.05	0.03
2-fold	0.98	0.07	0.05	0.03
3-fold	0.97	0.06	0.06	0.04
4-fold	0.96	0.09	0.07	0.05
5-fold	0.97	0.06	0.06	0.04
6-fold	0.97	0.06	0.05	0.03
7-fold	0.96	0.07	0.07	0.05
8-fold	0.96	0.07	0.08	0.07
9-fold	0.96	0.07	0.06	0.04
10-fold	0.98	0.05	0.05	0.04
Average	0.97	0.07	0.06	0.04
Kitchen				
1-fold	0.75	0.06	0.04	0.01
2-fold	0.89	0.04	0.02	0.01
3-fold	0.86	0.08	0.03	0.01
4-fold	0.80	0.09	0.03	0.02
5-fold	0.84	0.08	0.03	0.01
6-fold	0.85	0.08	0.03	0.01
7-fold	0.84	0.09	0.03	0.01
8-fold	0.88	0.10	0.03	0.01
9-fold	0.85	0.09	0.03	0.02
10-fold	0.84	0.07	0.03	0.01
Average	0.84	0.08	0.03	0.01
Bathroom				
1-fold	0.80	0.03	0.02	0.01
2-fold	0.85	0.02	0.02	0.02
3-fold	0.84	0.04	0.02	0.02
4-fold	0.78	0.04	0.03	0.02
5-fold	0.83	0.04	0.02	0.02
6-fold	0.85	0.04	0.03	0.02
7-fold	0.85	0.04	0.03	0.01
8-fold	0.79	0.04	0.03	0.02
9-fold	0.77	0.05	0.03	0.02
10-fold	0.82	0.04	0.03	0.02
Average	0.82	0.04	0.03	0.02

Bedroom				
1-fold	0.99	0.06	0.04	0.03
2-fold	0.99	0.05	0.05	0.02
3-fold	0.99	0.05	0.04	0.02
4-fold	0.98	0.05	0.06	0.03
5-fold	0.99	0.04	0.05	0.03
6-fold	0.99	0.04	0.04	0.02
7-fold	0.99	0.05	0.05	0.03
8-fold	0.99	0.04	0.05	0.03
9-fold	0.99	0.05	0.04	0.02
10-fold	0.99	0.04	0.03	0.02
Average	0.99	0.05	0.05	0.03

4. Application of the method: Case Study in the UK

4.1 Case study house

To validate the proposed approach towards establishing stochastic occupant occupancy in residential buildings, we applied our method to a typical residential building, serving as our model case study. This case study aims to cover the occupants in all the groups we divided for the TUS data. It pivots around a two-storey detached house, presumed to be inhabited by a six-member family, with a room distribution that aligns with the UK Office of National Statistics data [52]. As illustrated in Fig. 7, it encompasses six rooms with varying functionalities, namely: three bedrooms, one kitchen, one bathroom and one living room.

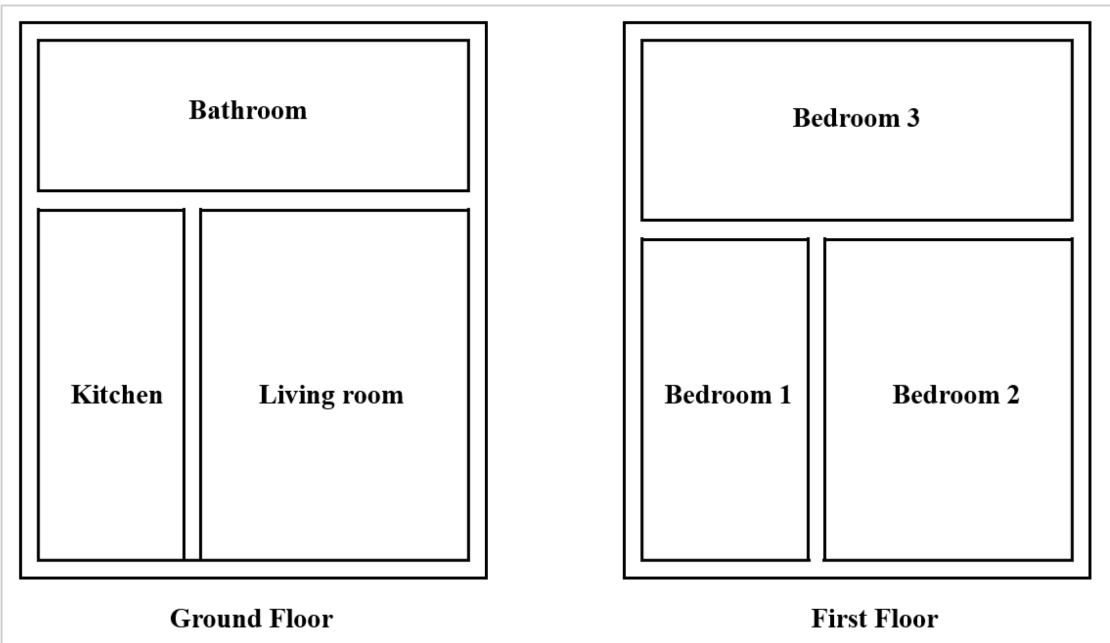

Fig. 7. Different functions spaces of the case study building

The occupants of this building are divided into three distinct age groups: two children aged between 8-14 years, two young adults aged 15-64 years, and two retirees aged over 65 years. This categorization serves to provide a more detailed understanding of occupancy patterns as influenced by age.

We further analyse the stochastic movement of these building occupants, focusing on transitions both within different rooms and between inside and outside of the house. This analysis aims to depict the model's ability to effectively represent these unpredictable movement patterns. The step-by-step application of this method is detailed in Fig. 8 below. We derived the necessary input data for this case from the TUS dataset. The calculations were performed on a desktop computer with Intel(R) Core (TM) i9-10900 CPU @ 2.80GHz, 32.0 GB of RAM, and running Windows 11 Professional. The time taken to complete a single run of the stochastic indoor occupancy pattern output was less than 1 second. This level of computational efficiency indicates that our model can be executed swiftly on standard modern computing hardware, enhancing its scalability and adaptability for various research and practical applications.

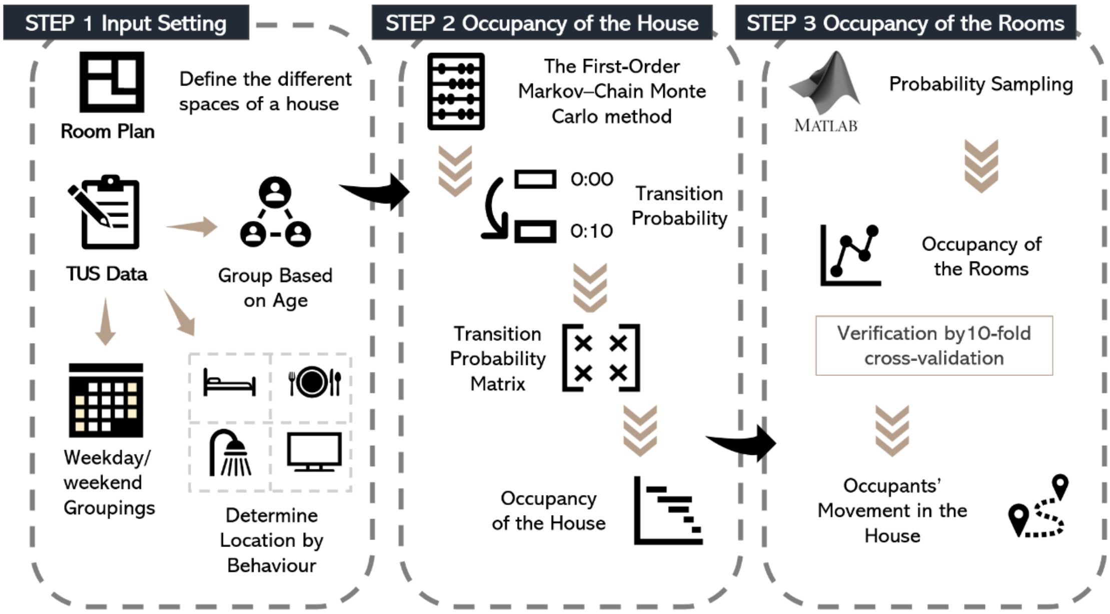

Fig. 8. The steps for application for the method

4.2 The generation of the overall occupancy in the house

To construct the transitional probability matrices of the occupancy of the six occupants, we implemented the Markov chain Monto Carlo method. Using the UK TUS data (2014-2015), in total 144 matrices were built for each group, representing the transition probabilities of occupants moving between inside and outside the house throughout the day, at 10 minutes intervals. Fig. 9 provides an example of such a matrix for occupants between 8-14 years old for the time interval from 12:00 to 12:10 noon. It reveals that, if a person (aged 8-14) was at home at 12:00, then there is a 0.977 probability that this person will still be at home at 12:10. Conversely, if the person is not at home at 12:00, there is a 0.047 probability that this person will return home at 12:10.

$$
\begin{array}{r c}
 & \textbf{Next state (at 12:10)} \\
\textbf{Current state (at 12:00)} & \begin{array}{cc} \textbf{Absence} & \textbf{Presence} \end{array} \\
\begin{array}{r} \textbf{Absence} \\ \textbf{Presence} \end{array} & \begin{bmatrix} 0.977 & 0.230 \\ 0.047 & 0.953 \end{bmatrix}
\end{array}
$$

Fig. 9. An example transition probability matrix for occupants between Ages 8 to 14 years old at 12:00 noon

Upon applying the obtained transition probability matrices to this case study house, we were able to derive the house's full-day occupancy status. We conducted three separate tests for both weekdays and weekends, with the results presented in Fig.10. As can be observed, each occupancy test is different. However, due to the use of the same transition probability matrix, they bear similarities. These consistent observations across all the random simulation iterations not only underscore the reliability of the proposed model but also attest to its ability to effectively represent stochastic occupancy behaviours.

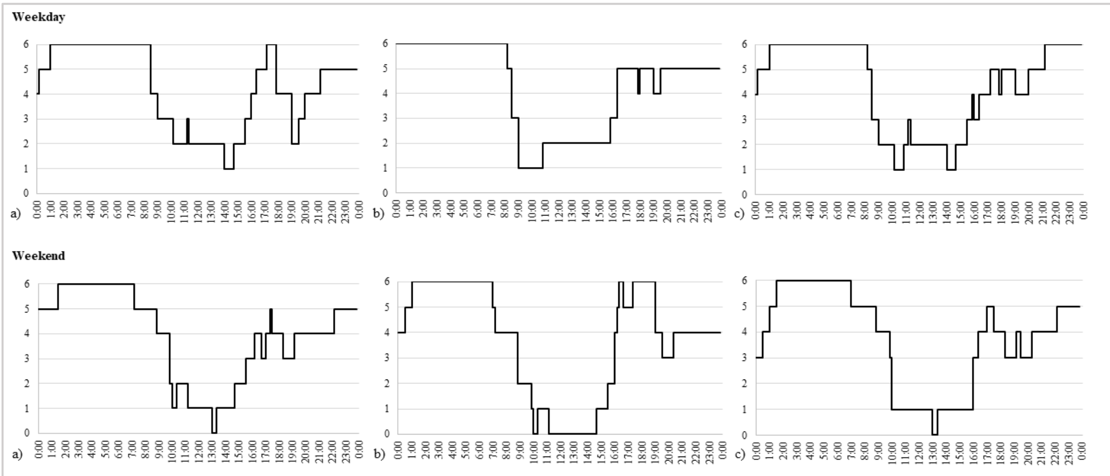

Fig. 10. Three random examples of the case study model: Occupancy results for the whole residential buildings on weekdays and weekends

4.3 The generation of the occupancy for different functional spaces

Building on the findings from Section 4.2 about the overall occupancy in the house, we employed a probabilistic sampling algorithm to determine the occupancy of each functional space. This enabled us to generate movement trajectories for occupants as shown in Fig.11. Occupant 1 and 2 represents individuals aged 8-14 years, occupant 3 and 4 represents individuals aged 15-64 years and occupant 5 and 6 represents individuals aged over 65 years.

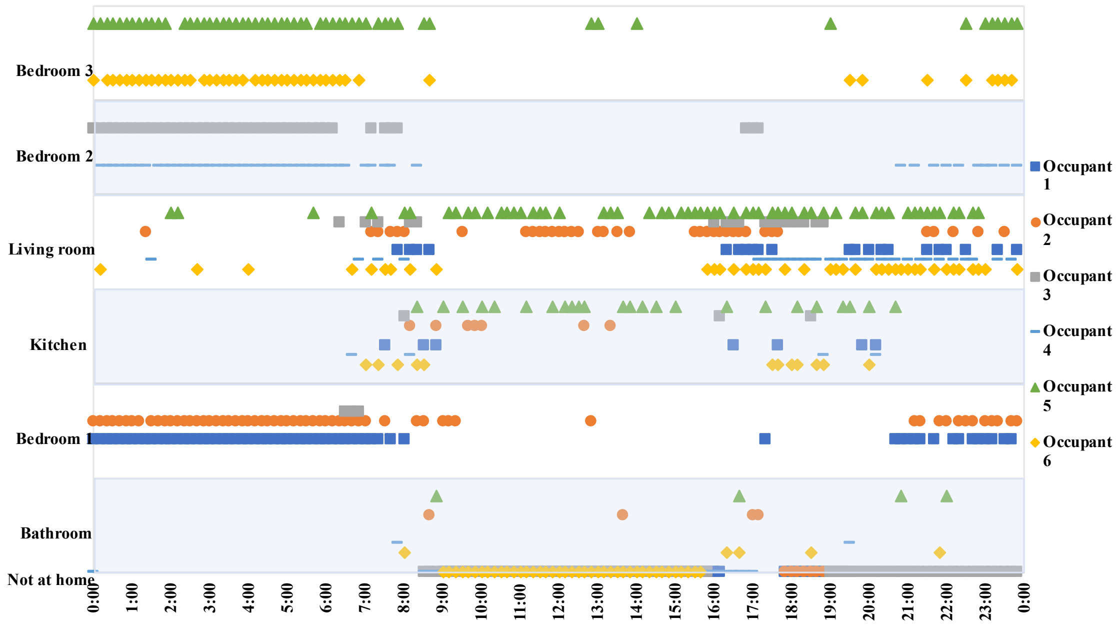

Fig. 11 Samples of occupancy in the house on a weekday

Figure 11 provides a compelling visualization of how occupancy fluctuates in a household over a typical weekday. It's crucial to clarify that the daily trajectories illustrated are not direct empirical data gathered from the UK TUS dataset. Instead, these trajectories are a result of a data generation process aimed at capturing and reflecting the inherent stochastic nature of occupant occupancy and movement. Each simulation, or 'run', exhibits its own unique pattern due to the inherent randomness of the generation process. Nevertheless, these runs all stem from the same probability basis for the sampling calculations. This ensures that the occupant behaviour, although unique in each run, exhibits overall similarity in terms of its characteristics.

A deeper look into the generated data reveals recognizable patterns. For instance, occupants generally leave their bedrooms in the morning, spending most of the daytime in the living room if they are home (particularly for those aged above 65), migrate to the kitchen around mealtimes, and return to the bedrooms in the evening.

It is important to note that the time spent in each room — the bedroom, living room, bathroom, and kitchen — varies significantly among occupants. These variations signify the stochastic simulation's effectiveness in capturing the random and unpredictable nature of occupant movement and occupancy. The data generation process thus successfully encapsulates the true complexity and dynamism inherent in human behaviour within residential environments.

5 Estimation of the appliance energy consumption

To study the efficacy of the stochastic model in calculating appliance energy consumption related to occupant behaviours, we used a typical weekday as an example. We compared the actual data derived from the TUS data, energy consumption calculated by the stochastic model, and energy consumption based on the ASHRAE standard schedule.

Due to the large size of the TUS data, it was not feasible to simulate all of it. Therefore, a random sampling method was used to select subsets of the data for the energy calculation. We randomly selected subsets of 120 data sets from the weekday dataset across three distinct groups. These subsets serve as representative samples, providing a snapshot of the larger dataset. The selection process adhered to a statistical standard of an alpha level less than 0.05, a common threshold in statistical hypothesis testing that ensures a less than 5% probability of incorrectly rejecting the null hypothesis, thereby affirming the statistical significance of our chosen sample sizes for large populations [53].

The distribution of energy consumption of a typical weekday for each group is shown in boxplots in Fig. 12. Group 1, comprising individuals aged 8-14, demonstrated an energy consumption range of 0 to 4.66 kWh in the actual model, with a median value of 2.15 kWh. The stochastic model for this group showed a similar range of 0 to 3.77 kWh, with a similar median value of 1.71 kWh. Group 2, consisting of individuals aged 15-64, exhibited an energy consumption range of 0 to 7.36 kWh in the actual model, with a median value of 2.81 kWh. The stochastic model for this group presented a range of 0 to 4.93 kWh, with a median value of 2.61 kWh. For Group 3, which includes individuals aged 65 and above, the actual model recorded an energy consumption range of 0 to 8.04 kWh, with a median of 4.09 kWh. The stochastic model for this group showed a range of 0.87 to 5.65 kWh, with a median of 3.60 kWh. The interquartile ranges, representing the spread of the middle 50% of the data, were found to be similar across all three groups. This similarity suggests comparable variability in energy consumption between the stochastic model and actual occupancy. The study also referenced a standard model, which consistently reported an energy consumption of 2.73 kWh. When compared with this standard model, the data from the actual and stochastic models either surpassed or fell below the standard model's energy consumption. These findings indicate that the standard model may not accurately represent the inherent variability in energy consumption of appliances within residential buildings, suggesting that a single, fixed value may not sufficiently capture the dynamic nature of energy consumption.

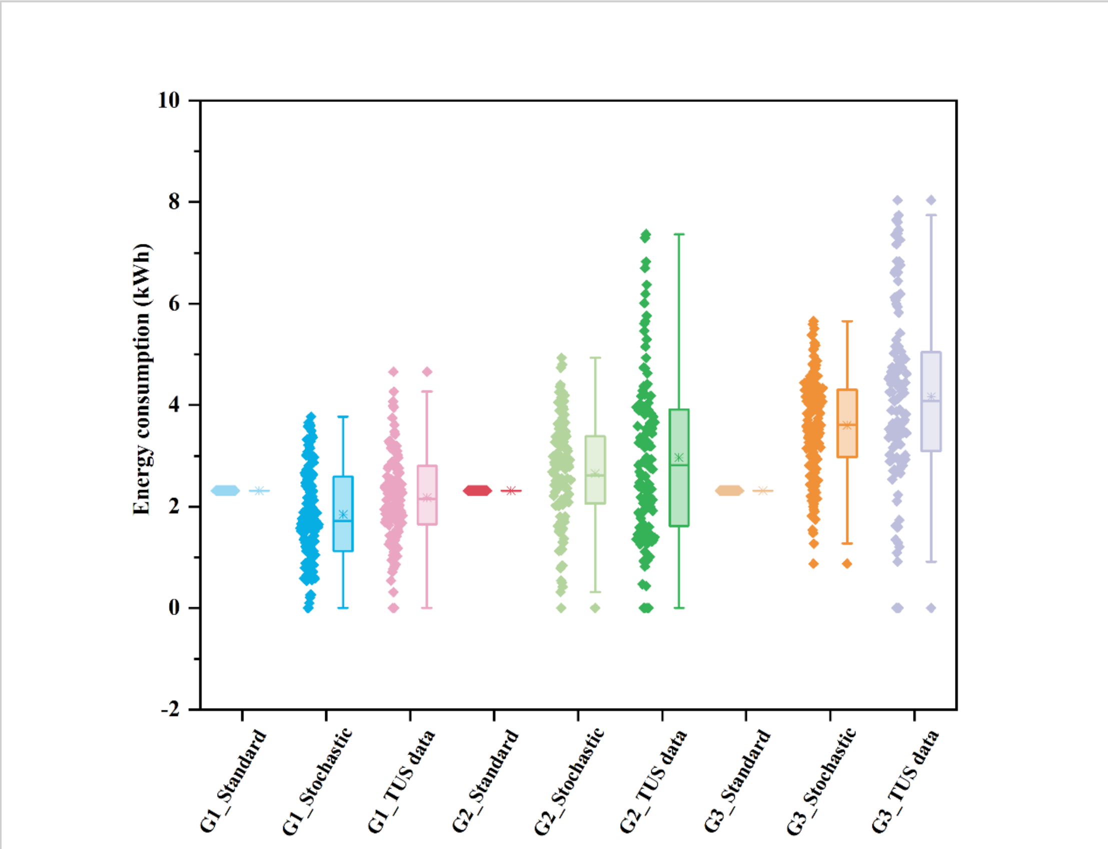

Fig. 12 Appliance energy consumption for occupants at different ages on a typical weekday

6 Discussion

In the present study, we have devised a novel method capable of reflecting the stochastic occupancy patterns in different functional spaces within residential buildings. The method was validated using an extensive dataset from the UK TUS, and the results revealed a close match between the generated activity data and the actual indoor activity statistics of the occupants. This method not only supplements the lesser-known methods of simulating occupancy in residential buildings [7, 19], but it also provides a more detailed simulation of occupancy by categorizing occupants by age. In section 5, we examined the application of the model by performing appliances energy consumption calculations using the stochastic model, the standard model, and real data. The results showed that the method was able to better reflect the stochastic nature of occupancy behaviour.

It is important to note that the energy consumption of these appliances represents only a part of the total energy consumption in residential buildings. Taking the abovementioned case study of a six-member family in the UK as an example, we can estimate their daily energy consumption based on the UK's per capita daily energy usage for heating (3.28 kWh), lighting (0.62 kWh), and hot water (0.55 kWh) [54]. Roughly, this type of household's daily energy consumption would range between 28.44 to 55.40 kWh, with a median value of approximately 42.06 kWh. It is crucial to consider that these figures can vary significantly due to external factors such as weather conditions. To more accurately simulate the entire building's energy consumption, the stochastic occupancy data generated for each room should be integrated into building simulation software, such as EnergyPlus, to calculate the energy consumption of all energy-consuming devices in residential buildings, including HVAC, lighting, domestic hot water, and appliance usage. This comprehensive approach is our next research goal, aiming to provide a more complete understanding of residential energy consumption patterns.

Furthermore, this method facilitates the achievement of more accurate predictions via a relatively simple algorithm. The proposed method employs Markov method and probabilistic sampling method to model occupancy patterns, enabling the random generation of numerous data sets that align with actual occupant activity. The true probabilities extracted using TUS are utilised to predict occupancy, which will yield commendable performance and facilitate the application of the proposed method to real building energy simulation [18].

While this study is predicated on the UK TUS data for validation and simulation purposes, the proposed method exhibits scalability. By obtaining occupancy rates for different functional rooms in residential buildings from the TUS data alone, the occupancy patterns of rooms in residential buildings can be established. Given the availability of TUS data in several, this method can be employed to simulate the occupancy of residential buildings in countries with diverse living habits, and to simulate energy consumption as well.

Furthermore, while the TUS data from 2014-2015 has provided a robust foundation for our study, we must acknowledge that lifestyles and occupancy patterns are subject to change over time. The COVID-19 pandemic, in particular, has significantly altered how residential spaces are used, with more people working and studying from home. The methodology and framework of our model are designed to be adaptable and can be updated with more recent data as it becomes available. Future research should consider updating the occupancy data to reflect these recent lifestyle changes, ensuring the continued relevance of occupancy models in a rapidly evolving world.

7 Conclusion

The study presented herein sought to address the challenge of accurately modelling occupancy patterns within residential buildings by developing a stochastic occupancy model based on TUS data. The importance of such models lies in their ability to effectively inform energy consumption simulations, which in turn aids in the design and management of energy-efficient buildings.

The proposed stochastic occupancy model was verified through an extensive ten-fold cross-validation process. The model’s performance was evident from the similarity between the occupancy trends generated by the model and the actual occupancy data. For the four functional rooms – bedroom, bathroom, kitchen, and living room – the model achieved an average R^2 value of 0.91, indicating a high degree of accuracy. Additionally, the average RMSE, MAE, and MedAE values for these rooms were 0.06, 0.04, and 0.03, respectively, further attesting to the model's precision in capturing occupancy patterns.

The model was applied to a case study of a two-story detached house in the UK. The application incorporated an examination of occupancy patterns in different functional spaces within the residential building and across different age groups. It was found that the model effectively reflects different behaviour patterns and room occupancies among occupants of different ages, as well as the resulting variations in appliance energy consumption. For occupants aged 8-14, a typical day's average appliance energy consumption ranged from 0 to 3.77 kWh, with a median of 1.71 kWh. For occupants aged 15-64, the range was 0 to 4.93 kWh, with a median of 2.61 kWh. For the elderly aged over 65, the range was 0.87 to 5.65 kWh, with a median of 3.60 kWh. These findings highlight the variability in energy consumption and underscore the importance of considering age-specific occupancy and behaviour patterns in residential energy consumption studies.

Looking ahead, the developed stochastic occupancy data can be integrated into building simulation software like EnergyPlus. This integration will enable more detailed calculations of energy consumption for all energy-consuming devices in residential buildings, including HVAC, lighting, hot water and appliance usage, thereby enhancing the accuracy and applicability of our model in real-world scenarios.

In conclusion, this research contributes a simple and stochastic model for simulating occupancy in residential buildings. The method, grounded in a combination of Markov chains and probabilistic sampling, proved to be effective in generating data that closely aligns with real-world occupancy patterns. Importantly, it is worth mentioning that the method has the potential for scalability and can be adapted to various contexts given the availability of TUS data in numerous countries at different times. Future research could explore the extension of this model to other building types and the incorporation of additional parameters such as outdoor environmental conditions or cultural differences in occupancy patterns.

Acknowledgement

This research was funded by the International Partnership Program of Chinese Academy of Sciences (grand No. 132c35kysb2020007) and Ningbo Commonweal Science and Technology Planning Project (grand No. 2021S081).

Statement: During the preparation of this work the authors used ChatGPT in order to improve readability and language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

1. Hong, T. and H.-W. Lin, *Occupant behavior: impact on energy use of private offices*. 2013, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States).
2. OVENS, K.R., *TECHNICAL SUPPORT DOCUMENT: ENERGY EFFICIENCY STANDARDS.* 1993.
3. Happle, G., J.A. Fonseca, and A. Schlueter, *Context-specific urban occupancy modeling using location-based services data.* Building and Environment, 2020. **175**: p. 106803.
4. Ahmed, K., et al., *Occupancy schedules for energy simulation in new prEN16798-1 and ISO/FDIS 17772-1 standards.* Sustainable cities and society, 2017. **35**: p. 134-144.
5. Hoes, P., et al., *User behavior in whole building simulation.* Energy and Buildings, 2009. **41**(3): p. 295-302.

6. Rouleau, J., et al., *A unified probabilistic model for predicting occupancy, domestic hot water use and electricity use in residential buildings.* Energy and Buildings, 2019. **202**: p. 109375.
7. Hu, S., et al., *Investigation and analysis of Chinese residential building occupancy with large-scale questionnaire surveys.* Energy and Buildings, 2019. **193**: p. 289-304.
8. *Energy Efficient Design of New Buildings Except Low-rise Residential Buildings*. Vol. 90. 1989: ASHRAE.
9. Duarte, C., K. Van Den Wymelenberg, and C. Rieger, *Revealing occupancy patterns in an office building through the use of occupancy sensor data.* Energy and Buildings, 2013. **67**: p. 587-595.
10. Sun, K., Q. Zhao, and J. Zou, *A review of building occupancy measurement systems.* Energy and Buildings, 2020. **216**: p. 109965.
11. Erickson, V.L. and A.E. Cerpa. *Occupancy based demand response HVAC control strategy*. in *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*. 2010.
12. Peng, Y., et al., *Occupancy learning-based demand-driven cooling control for office spaces.* Building and Environment, 2017. **122**: p. 145-160.
13. Lu, J., et al. *The smart thermostat: using occupancy sensors to save energy in homes*. in *Proceedings of the 8th ACM conference on embedded networked sensor systems*. 2010.
14. Macdonald, I. and P. Strachan, *Practical application of uncertainty analysis.* Energy and Buildings, 2001. **33**(3): p. 219-227.
15. Wang, D., C.C. Federspiel, and F. Rubinstein, *Modeling occupancy in single person offices.* Energy and Buildings, 2005. **37**(2): p. 121-126.
16. Page, J., et al., *A generalised stochastic model for the simulation of occupant presence.* Energy and Buildings, 2008. **40**(2): p. 83-98.
17. Wang, C., D. Yan, and Y. Jiang, *A novel approach for building occupancy simulation.* Building Simulation, 2011. **4**(2): p. 149-167.
18. Liang, X., T. Hong, and G.Q. Shen, *Occupancy data analytics and prediction: A case study.* Building and Environment, 2016. **102**: p. 179-192.
19. Richardson, I., M. Thomson, and D. Infield, *A high-resolution domestic building occupancy model for energy demand simulations.* Energy and Buildings, 2008. **40**(8): p. 1560-1566.
20. McKenna, E., M. Krawczynski, and M. Thomson, *Four-state domestic building occupancy model for energy demand simulations.* Energy and Buildings, 2015. **96**: p. 30-39.
21. Buttitta, G. and D.P. Finn, *A high-temporal resolution residential building occupancy model to generate high-temporal resolution heating load profiles of occupancy-integrated archetypes.* Energy and Buildings, 2020. **206**: p. 109577.
22. Huchuk, B., S. Sanner, and W. O'Brien, *Comparison of machine learning models for occupancy prediction in residential buildings using connected thermostat data.* Building and Environment, 2019. **160**: p. 106177.
23. Causone, F., et al., *A data-driven procedure to model occupancy and occupant-related electric load profiles in residential buildings for energy simulation.* Energy and

Buildings, 2019. **202**: p. 109342.

24. Diao, L., et al., *Modeling energy consumption in residential buildings: A bottom-up analysis based on occupant behavior pattern clustering and stochastic simulation.* Energy and Buildings, 2017. **147**: p. 47-66.
25. Sayed, A.N., Y. Himeur, and F. Bensaali, *From time-series to 2D images for building occupancy prediction using deep transfer learning.* Engineering Applications of Artificial Intelligence, 2023. **119**: p. 105786.
26. Salimi, S., Z. Liu, and A. Hammad, *Occupancy prediction model for open-plan offices using real-time location system and inhomogeneous Markov chain.* Building and Environment, 2019. **152**: p. 1-16.
27. Chen, Z., J. Xu, and Y.C. Soh, *Modeling regular occupancy in commercial buildings using stochastic models.* Energy and Buildings, 2015. **103**: p. 216-223.
28. Chen, Z. and C. Jiang, *Building occupancy modeling using generative adversarial network.* Energy and Buildings, 2018. **174**: p. 372-379.
29. Jiang, C., et al., *Indoor occupancy estimation from carbon dioxide concentration.* Energy and Buildings, 2016. **131**: p. 132-141.
30. Chang, W.-K. and T. Hong, *Statistical analysis and modeling of occupancy patterns in open-plan offices using measured lighting-switch data.* Building Simulation, 2013. **6**(1): p. 23-32.
31. Tabak, V. and B. de Vries, *Methods for the prediction of intermediate activities by office occupants.* Building and Environment, 2010. **45**(6): p. 1366-1372.
32. Ekwevugbe, T., N. Brown, and D. Fan. *A design model for building occupancy detection using sensor fusion*. in *2012 6th IEEE International Conference on Digital Ecosystems and Technologies (DEST)*. 2012.
33. Huang, W., et al., *Modeling and predicting the occupancy in a China hub airport terminal using Wi-Fi data.* Energy and Buildings, 2019. **203**: p. 109439.
34. Kim, S., et al., *Real-time occupancy prediction in a large exhibition hall using deep learning approach.* Energy and Buildings, 2019. **199**: p. 216-222.
35. Ai, B., Z. Fan, and R.X. Gao. *Occupancy estimation for smart buildings by an auto-regressive hidden Markov model*. in *2014 American Control Conference*. 2014.
36. Aerts, D., et al., *A method for the identification and modelling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison.* Building and Environment, 2014. **75**: p. 67-78.
37. Li, Z. and B. Dong, *A new modeling approach for short-term prediction of occupancy in residential buildings.* Building and Environment, 2017. **121**: p. 277-290.
38. Panchabikesan, K., F. Haghighat, and M.E. Mankibi, *Data driven occupancy information for energy simulation and energy use assessment in residential buildings.* Energy, 2021. **218**: p. 119539.
39. Doma, A. and M. Ouf. *Developing Urban-Scale Markovian Occupancy Profiles for Residential Buildings Using Smart Thermostat Data*. in *Proceedings of the 5th International Conference on Building Energy and Environment*. 2023. Singapore: Springer Nature Singapore.
40. Erickson, V.L., M.Á. Carreira-Perpiñán, and A.E. Cerpa, *Occupancy modeling and prediction for building energy management.* ACM Transactions on Sensor Networks

(TOSN), 2014. **10**(3): p. 1-28.

41. Dong, B., Z. Li, and G. Mcfadden, *An investigation on energy-related occupancy behavior for low-income residential buildings.* Science and Technology for the Built Environment, 2015. **21**(6): p. 892-901.
42. Rueda, L., et al., *A comprehensive review of approaches to building occupancy detection.* Building and Environment, 2020. **180**: p. 106966.
43. Yamaguchi, Y. and Y. Shimoda. *Evaluation of a behavior model of occupants in home based on japanese national time use survey*. in *Proceedings of Building Simulation*. 2015.
44. Gershuny, J. and O. Sullivan, *United kingdom time use survey, 2014-2015. centre for time use research, university of oxford.* URL: http://doi. org/10.5255/UKDA-SN-8128-1, 2017.
45. Martinaitis, V., et al., *Importance of occupancy information when simulating energy demand of energy efficient house: A case study.* Energy and Buildings, 2015. **101**: p. 64-75.
46. Flett, G. and N. Kelly, *An occupant-differentiated, higher-order Markov Chain method for prediction of domestic occupancy.* Energy and Buildings, 2016. **125**: p. 219-230.
47. O'Neill, A. *Age distribution in the United Kingdom 2011-2021*. 2023; Available from: https://www.statista.com/statistics/270370/age-distribution-in-the-united-kingdom/.
48. Vakharia, V. and R. Gujar, *Prediction of compressive strength and portland cement composition using cross-validation and feature ranking techniques.* Construction and Building Materials, 2019. **225**: p. 292-301.
49. Chidurala, V. and X. Li, *Occupancy Estimation Using Thermal Imaging Sensors and Machine Learning Algorithms.* IEEE Sensors Journal, 2021. **21**(6): p. 8627-8638.
50. Gerther, M. *POWER CONSUMPTION OF HOUSEHOLD APPLIANCES*. 2023; Available from: https://generatorist.com/power-consumption-of-household-appliances.
51. Dangeti, P., *Statistics for machine learning*. 2017: Packt Publishing Ltd.
52. Statistics, O.o.N. *Housing, England and Wales: Census 2021*. 2021; Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/housing/bulletins/housingenglandandwales/census2021#glossary.
53. Kotrlik, J. and C. Higgins, *Organizational research: Determining appropriate sample size in survey research appropriate sample size in survey research.* Information technology, learning, and performance journal, 2001. **19**(1): p. 43.
54. Zimmermann, J.-P., et al., *Household Electricity Survey: A study of domestic electrical product usage.* Intertek Testing & Certification Ltd, 2012: p. 213-214.