BUILD SIMUL (2023) 16: 483–497
https://doi.org/10.1007/s12273-022-0950-8

Research Article

Architecture and Human Behavior

Occupancy of rooms in urban residential buildings by users in cold areas of China

Qi Dong[1,2], Zikai Ma[1,2], Cheng Sun[1,2] (✉)

1. School of Architecture, Harbin Institute of Technology, Harbin 150006, China
2. Key Laboratory of Cold Region Urban and Rural Human Settlement Environment Science and Technology, Ministry of Industry and Information Technology, Harbin 150006, China

Abstract

Occupancy is used to represent the movements and locations of users among various zones of buildings, and it is the basis of all other daily energy consumption behaviors. This study investigated eight families in cold areas of China based on occupancy measurements obtained in four main rooms, i.e., living room, bedroom, kitchen, and bathroom. In particular, we analyzed the duration of user occupancy and hourly mean occupancy, and characterized their regular and random features. According to the results, we developed an event-based occupancy model using an inhomogeneous Markov chain, where the rooms were modeled and daily events were divided into three categories according to their randomness. We established a new method for conversion between event characteristic parameters and a transition probability matrix, as well as an overlap avoidance method for active events. The model was then validated using real data. The results showed that the model performed well in terms of two evaluation criteria. The model should improve the accuracy of simulations of occupancy.

Keywords

event;
occupancy;
urban residential building;
simulation;
cold areas of China

Article History

Received: 06 June 2022
Revised: 01 September 2022
Accepted: 05 October 2022

1 Introduction

At the end of 2021, the total carbon emissions from urban construction accounted for about 20% of the total carbon emissions in China (Jiang and Hu 2021). Therefore, in order to achieve the goal of carbon neutrality (Wang and Gu 2021), the construction industry needs to vigorously promote energy conservation and reduce emissions. According to IEA Annex-53, the factors that affect the energy efficiency of buildings include the meteorological conditions, building structure, and user behaviors (Yan et al. 2015). Previous studies have shown that simply improving the performance of a building might not effectively enhance the building energy conservation level (Blight and Coley 2013), mainly due to the impacts of user behaviors on building energy consumption (Wang 2014). User behaviors can vary greatly among different buildings, thereby resulting in differences in energy consumption (Zhang et al. 2018). The energy consumed by residential buildings is the main component of building energy consumption (Zhang et al. 2019). And studies have shown that occupant behavior is commonly considered as a major contributor to result in a "gap" between actual and simulated energy consumption of buildings (Yan et al. 2015; Hong et al. 2016). Among the various user behaviors, occupancy is generally used to describe the locations and movements of users in various zones in buildings, and it is the basis of all other energy consumption behaviors. Occupancy is beginning to draw the attention of researchers because of the "gap" between actual and predicted energy consumption caused by the incorrect simulation of occupancy.

1.1 Literature review

In recent years, numerous efforts have been made to develop the following different types of occupancy models to predict occupants' energy-related behaviors and presence schedules.

1) Deterministic model

This type of model considers that the activities of users in a

E-mail: suncheng@hit.edu.cn

room are fixed during a day, so the time that the room is occupied is also fixed. For example, EnergyPlus includes the "*Schedules*" module to set the hourly user activities and lighting usage in a building (Crawley et al. 2001).

2) Stochastic model

The three subtypes of stochastic models are described as follows. The first subtype comprises time series forecasting models, which can be divided into two levels. (1) The first level of the model is for a single user. In particular, Wang et al. (2005) used a Poisson distribution to describe the hourly mean occupancy. Richardson et al. (2008) used the UK Time-Use Survey data to predict occupancy by using a first-order inhomogeneous Markov chain. Page et al. (2008) introduced the "move parameter-μ" to improve the first-order Markov chain for predicting the hourly mean occupancy. And Wang et al. (2015) proposed an occupancy model based on a Markov chain and daily user events to simulate the daily movements of users with only a few parameters from the daily events. (2) The second level of the time series forecasting model subtype is related to multiple users. In particular, Aerts et al. (2014) conducted cluster analysis based on occupancy patterns and proposed a series of diverse occupancy models. In addition, Flett and Kelly (2021) used statistical methods to develop a model for simulating the diversity of user behaviors. The second stochastic model subtype comprises multi-factor estimation models used for predicting the mean occupancy based on multiple factors, such as time, environmental parameters, and user behavior. In particular, Dong et al. (2011) collected real time data regarding the changes in the number of users in a specific room, and then used a neural network and hidden Markov chain to predict the number of users. Fajilla et al. (2021) used the law of total probability (LTP), naïve Bayes classifier (NB), and classification and regression tree (CART) to predict the mean occupancy, and compared the performance of the three models. The third statistical model subtype is characterized by the statistical modeling of indicators such as the number of users and occupancy time in a single day rather than predicting the hourly mean occupancy (Sun et al. 2014).

Most models can also be divided into two categories by whether they used discrete-time approaches or discrete-event approaches, as summarized in the second column of Table 1 (Li et al. 2022). A discrete-time approach can directly generate the occupants' movement at each time-step of the day (Page et al. 2008; Richardson et al. 2008; Aerts et al. 2014; Diao et al. 2017), and a discrete-event approach can predict an ordered event sequence of the day, which can generate the occupants' movement later. Wang et al. (2011) proposed a stochastic model based on daily events and the Markov chain, which took the time parameters of typical events as the model input, and can easily simulate the actual occupancy in the room. Wilke et al. (2013) presented a bottom-up modelling approach to predict occupants' time-dependent events in residential buildings. Similarly, Yamaguchi and Shimoda (2017) developed a stochastic

Table 1 Modelling approach, building type and scope of previous studies

Literature	Modelling approach	Building type	Scope
Wang et al. 2005	Discrete-time	Office building	The whole room
Richardson et al. 2008	Discrete-time	Residential building	The whole apartment
Page et al. 2008	Discrete-time	Office building	Different zones
Chang and Hong 2013	Discrete-time	Office building	Different zones
Wilke et al. 2013	Discrete-event	Residential building	The whole apartment
Aerts et al. 2014	Discrete-time	Residential building	The whole apartment
Wang et al. 2015	Discrete-event	Office/residential buildng	Different zones
Chen et al. 2015	Discrete-time	Commercial building	Different zones
McKenna et al. 2015	Discrete-time	Residential building	The whole apartment
Diao et al. 2017	Discrete-time	Residential building	The whole apartment
Yamaguchi and Shimoda 2017	Discrete-event	Residential building	The whole apartment
Salimi et al. 2019	Discrete-time	Office building	Different zones
Flett and Kelly 2021	Discrete-time	Residential building	The whole apartment
Fajilla et al. 2021	Discrete-time	Office building	The whole room
Rueda et al. 2021	Discrete-event	Residential building	The whole apartment
Jeong et al. 2021	Discrete-time	Residential building	The whole apartment
Malekpour Koupaei et al. 2022	Discrete-time	Residential building	The whole apartment

model to predict occupants' daily events such as working and eating meals for community-/urban-scale energy demand modelling. By using discrete-event approach or taking occupancy as event, we can simplify the occupants' movement as events or activities, and indirectly determine the occupants' location by simulating the occurrence time of events.

From the perspective of building type, the review findings suggest that previous researches into occupancy in different building zones had mainly focused on office buildings, whereas almost all studies (10 of 11 reviewed studies) that have considered residential buildings investigated whether users were occupying anywhere in a whole apartment, as shown in the third and fourth columns of Table 1.

Previous studies had also concluded that the relationship between different types of occupants (e.g., full-time workers, stay-at-home parents, retired individuals etc.) and occupancy was complex (Flett and Kelly 2016). Yao and Steemers (2005) proposed that the type of employment and the working hours were the most significant occupancy effect. Similarly, Alerts et al. (2014) identified seven typical occupancy patterns by using hierarchical clustering. The results showed that different patterns occurred for different occupant types.

Thus, there is a trend to improve the accuracy of occupancy modeling and reduce its complexity. Wang's event-based model can reduce the complexity by setting daily events for users and by applying a Markov chain to simulate the randomness of occupancy (Wang et al. 2015). However, this model is still insufficient of simulating the randomness of occupancy in residential buildings. For example, the times of occurrence for daily events rarely follow a geometric distribution within the effective time period. In addition, definitions are not provided in the model for periodic events and there is no method for processing overlapping fixed events.

1.2 Aims of this study

In order to address the problems which are "previous researches into occupancy in different building zones had mainly focused on office buildings" and "the deficiencies of Wang's model" (Wang et al. 2015), we conducted field measurements for eight families in Harbin and Shijiazhuang, which are representative cities in cold areas of China. According to the studies mentioned above, we selected two main types of families: stay-at-home family and full-time work family. The collected data were analyzed to determine the differences and similarities in terms of the occupancy of four main rooms, i.e., living room, bedroom, kitchen, and bathroom. An improved event-based occupancy model was then developed based on an inhomogeneous Markov chain, and the accuracy of the model was validated. A flowchart illustrating the research process is shown in Figure 1.

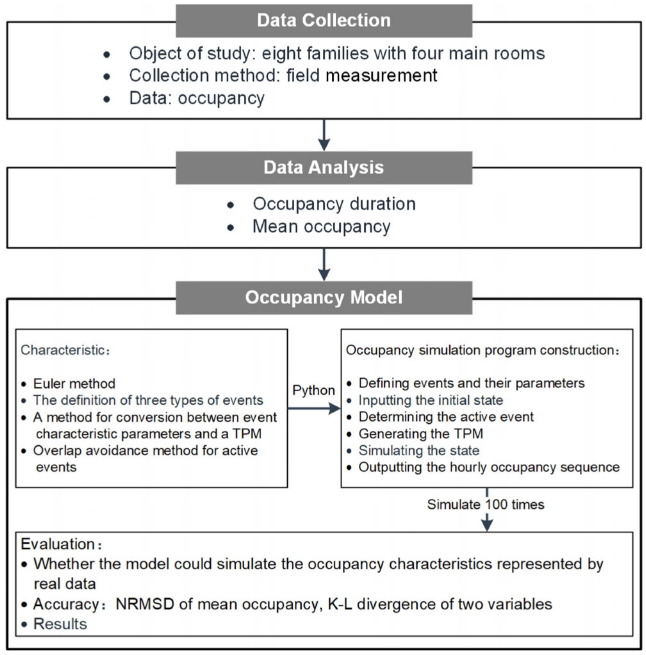


Fig. 1 Schematic illustrating the flow of the research process

2 Methods

2.1 Site survey and measurements

Eight families were selected in Harbin and Shijiazhuang, China, which comprised five stay-at-home families and three full-time work families. Basic details regarding the apartments and families before the field measurements were collected are shown in Table 2.

Families A, B, C, D, and E were all stay-at-home families characterized by the adults staying at home for a long time without going out to work. Most were working from home during the COVID-19 pandemic or retired elderly families. Families F, G, and H were full-time work families characterized by the adults going out to work during the day and returning home at night.

HOBO UX90-006x occupancy loggers (Figure 2) were used to collect occupancy data for each room. This instrument has a long self-record function and it collects data every 2 min. The measurement cycle was from November 25, 2021 to April 10, 2022, and the average measurement period per household was 2–3 weeks.

The occupancy loggers were directed toward the zones where people often occurred in the test rooms at a height of about 1.5 m. For example, as shown in Figure 3, the instrument was placed on the wall in the room in household

Table 2 Basic information for the eight families and their apartments

	City	Family type	Family size	Family members	Apartment type	Area (m^2)	Occupied since	Floor
A	Harbin	Stay-at-home	2	Middle-aged couple	3 rooms, 1 hall	80	2000 s	3rd
B	Harbin	Stay-at-home	3	Young couple + 1 child	2 rooms, 1 hall	70	2010 s	14th
C	Harbin	Stay-at-home	3	Middle-aged couple + 1 young man	3 rooms, 2 halls	92	2010 s	8th
D	Shijiazhuang	Stay-at-home	1	1 Young man	2 rooms, 1 hall	66	1990 s	5th
E	Harbin	Stay-at-home	6	2 old people + young couple + 2 children	3 rooms, 1 hall	90	1990 s	1st
F	Harbin	Full-time work	1	1 young man	2 rooms, 1 hall	45	2000 s	3rd
G	Harbin	Full-time work	1	1 young man	1 room, 1 hall	31	1990 s	1st
H	Shijiazhuang	Full-time work	2	Middle-aged couple	2 rooms, 1 hall	72	1990 s	5th

Notes: The age ranges for "child", "young man/couple", "middle-aged couple", and "old people" are <18 years old, 18–44 years old, 45–65 years old, and >65 years old respectively.

Fig. 2 HOBO UX90-006x occupancy logger

(a)

(b)

Fig. 3 Instrument layout in (a) a living room, (b) a bedroom

E, and Figure 4 shows the measurement points and the layout of the apartment for household B.

2.2 Occupancy modeling

Based on Wang's model (Wang et al. 2015), we developed an improved occupancy model. Wang's model contains two main parts, as shown in Figure 5. The first part uses an inhomogeneous Markov chain to represent the random movements of users. A Markov chain $\{X_t\}$ is a set of discrete random variables with the Markov property. The conditional probability of random variable X_t satisfies the following relationship: $p(X_{t+1}|X_t, \ldots, X_1) = p(X_{t+1}|X_t)$. If the transition probability $p_{ij}(t) = p\{X_{t+1} = j|X_t = i\}$ is related to the time t, $\{X_t\}$ is called an inhomogeneous Markov chain (Serfozo 2009). The second part is the event method. Wang's model defines a series of daily events that cause changes in the user's position (Wang et al. 2015). Each event can modify the state transition probability matrix of the Markov chain.

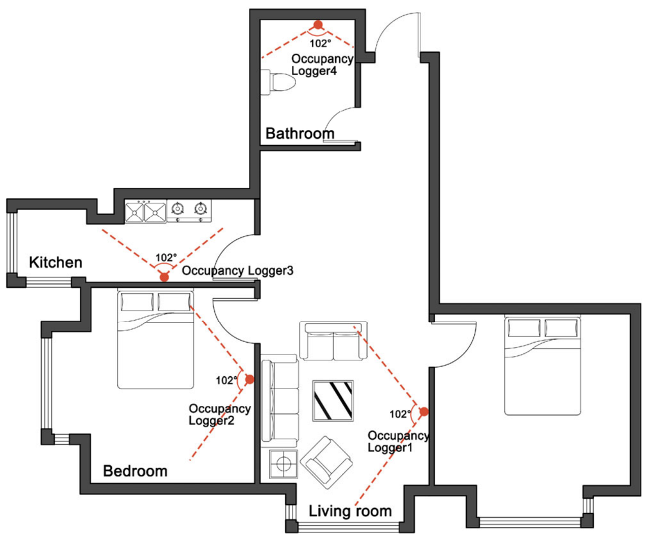


Fig. 4 Measurement points and apartment layout

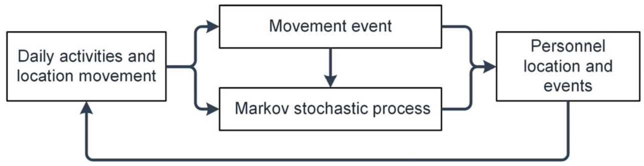


Fig. 5 Wang's model (Wang et al. 2015)

The model can generate the hourly user positions and room states according to the user's initial position and the state transition probability matrix at each time step.

2.3 Occupancy simulation program construction

Based on the occupancy model described above, an occupancy simulation program was constructed. The program simulated a one-day state sequence for one room at each time according to the following steps (see Figure 6).

(1) Creating the event set: The occupancy data was collected and preprocessed. Three types of events were defined and their characteristic parameters in the four rooms were determined based on real data, before simulating whether each periodic event occurred using the characteristic parameters. Then the event set was created.

(2) Determining the active event at time *i*: The active event was determined by using the occurrence time, priority method, and last state review method. Firstly, use the occurrence time to filter each event of the event set created in step (1). The characteristic parameters "*s_time*" and "*e_time*" were used during filtering, which was defined in Table 5. Secondly, if the number of events in the active event set, i.e., "*ongoing_events*" generated in the first filtering was greater than 1, then the priority of events was used for the second filtering, which was also defined in Table 5. Thirdly, if the active event set

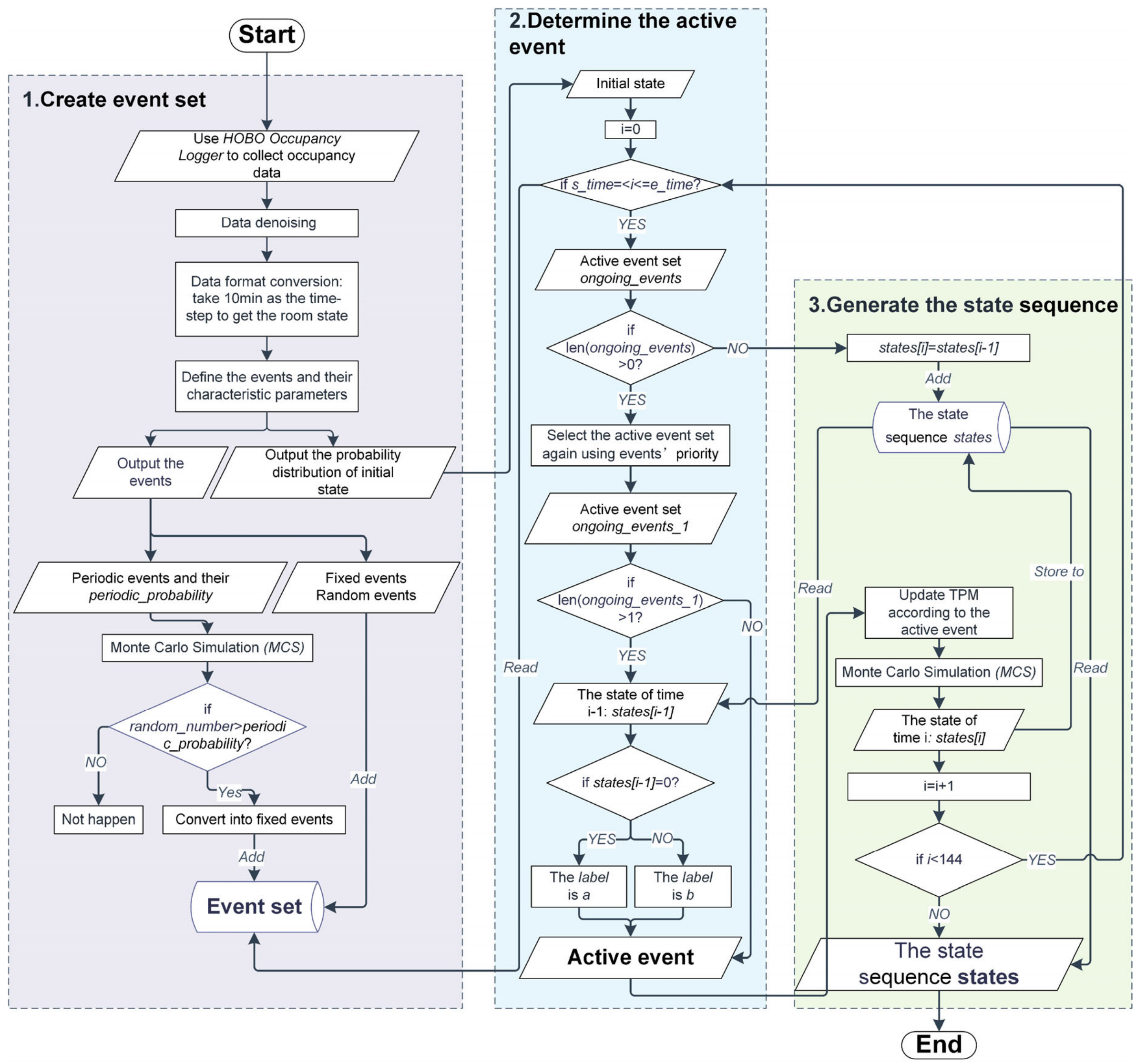


Fig. 6 Occupancy simulation process

"*ongoing_events_1*" generated in the second filtering still contained more than one event, use the "last state review" method to determine the active event at time *i*. The "last state review" method was introduced in detail in the fourth part of Section 3.3.

(3) Simulating the occupancy state at time *i*: According to the active event determined in step (2), the occupancy state at time *i* was simulated by using the state transition probability matrix generated using the characteristic parameters and the occupancy state at time *i* − 1. The procedure to generate the transition probability matrix, i.e., TPM was explained in the third part of Section 3.3.

(4) Steps (2) and (3) were repeated to obtain the hourly state sequence for a room for one day.

2.4 Occupancy model evaluation

We evaluate the performance of the proposed model as follows.

(1) We evaluated whether the proposed model could simulate the occupancy characteristics represented by the data measurements for the main four rooms. The performance evaluation was conducted by analyzing the one-day state sequence simulated by the program (Figure 6), before comparing the similarity between the simulated data and actual measurements.

(2) Accuracy of the proposed model: The data measurements showed that the occupancy state sequence varied for the same room on different days, but it was not useful to compare the measured and simulated values in an occupancy state sequence for one day. Thus, three variables related to the occupancy properties were defined: mean occupancy, cumulative occupied duration, and number of occupied/unoccupied transitions. These variables were defined in a previous study (Liao et al. 2012).

The following two evaluation metrics were used to quantify the performance of the proposed model in terms of the three variables described above.

(1) The normalized root mean square deviation (NRMSD) was used to measure the error between the mean occupancy for a simulated sequence ($\hat{y}$) and the mean occupancy for a measured sequence (y). The following formula was used to calculate the NRMSD, where n represents the length of the sequence (Chen et al. 2015).

$$\mathrm{NRMSD} = \frac{\sqrt{\dfrac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}}{y_{\max} - y_{\min}} \tag{1}$$

(2) The Kullback–Leibler (K–L) divergence is used to measure the difference between two probability distributions for random variables (Cover and Thomas 2006). Thus, it was used to evaluate the difference between the probability mass functions (pmfs) for the measured data and simulated data for two variables comprising the cumulative occupied duration and number of occupied/unoccupied transitions. In the following formula, P_k represents the pmf for the measured data, Q_k represents the pmf for the simulated data, and $D(P||Q)$ denotes the K–L divergence between P_k and Q_k.

$$D(P||Q) = \sum_k P_k \log \frac{P_k}{Q_k} \tag{2}$$

3 Analysis and results

3.1 Occupancy duration

In this study, HOBO UX90-006x occupancy loggers were used to collect and preprocess the occupancy data. We analyzed the data to obtain the daily total occupancy duration and daily occupancy duration for each room for "stay-at-home families", "full-time work families on weekdays" and "full-time work families on weekend" and the results are shown in Table 3 and Figure 7.

The daily total occupancy durations differed greatly among the three family types. The daily total occupancy duration for "stay-at-home families" was about 21.5 h per day, which accounted for 89% of the day. The daily total occupancy duration for "full-time work families on weekdays" was 12.3 h per day, which comprised 51% of the day. The daily total occupancy duration for "full-time work families on weekend" was 21.9 h, i.e., about 91% of the day. Thus, the total occupancy duration for "stay-at-home families" was approximately the same as that for "full-time work families on weekend," and about 78% more than that for "full-timework families on weekdays".

The occupancy durations in the four main rooms by the three family types are shown in Figure 7(b). The occupancy

Table 3 Occupancy duration results

Conditions	Total occupancy duration per day (h)	Living room (h)	Bedroom (h)	Kitchen (h)	Bathroom (h)
Stay-at-home families	21.5	5.7	15.6	4.2	2.9
Full-time work families on weekdays	12.3	1.6	11.3	1.0	1.2
Full-time work families on weekend	21.9	4.4	17.8	3.0	2.4

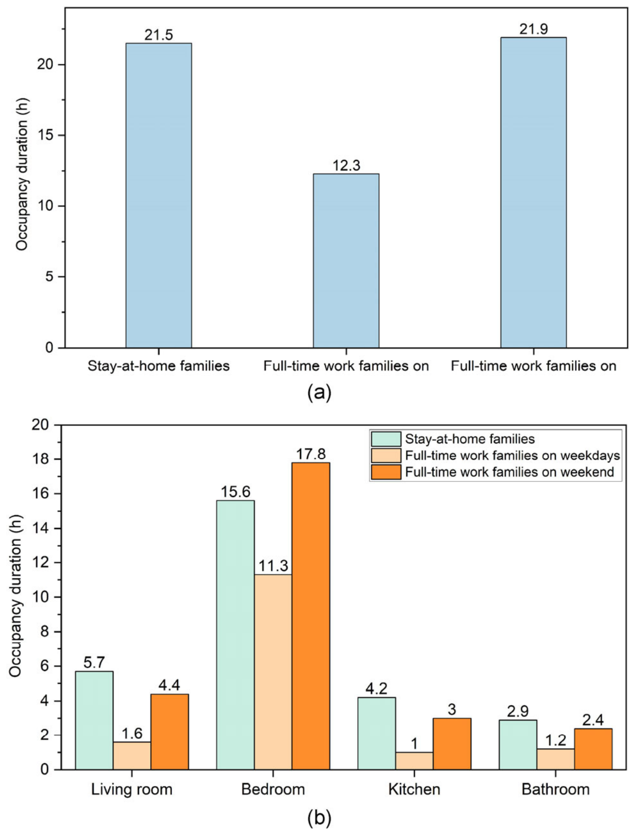


Fig. 7 Comparison of occupancy duration: (a) total occupancy duration per day, (b) occupancy duration for each room per day

durations in the four rooms decreased in the following order: bedroom > living room > kitchen > bathroom. The total occupancy duration in the bedroom was 91.9% of the total occupancy duration for “full-time work families on weekdays” and longer than those for “stay-at-home families” and “full-time work families on weekend”. The occupancy duration in the bedroom was 12% less for “stay-at-home families” than “full-time work families on weekend,” and the occupancy durations were longer in the living room, bathroom, and kitchen.

3.2 Mean occupancy

The occupancy durations in the whole apartment and four main rooms differed little between “stay-at-home families” and “full-time work families on weekend”. The daily activities of “full-time work families” at the weekend were diverse, such as traveling and overtime, with high complexity and no typical patterns. Therefore, the mean occupancy was only analyzed for “stay-at-home families” and “full-time work families on weekdays” (as shown in Table 4).

1) Stay-at-home families

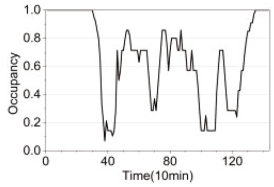
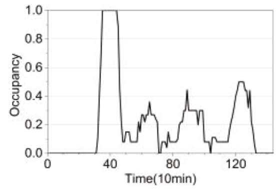
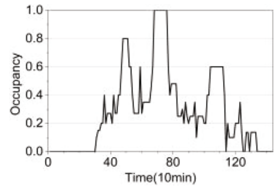
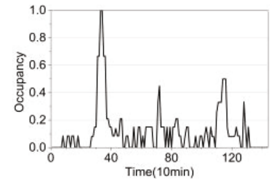
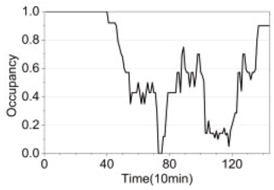
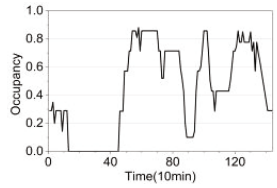
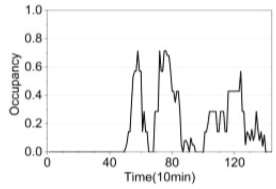
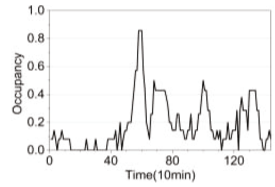
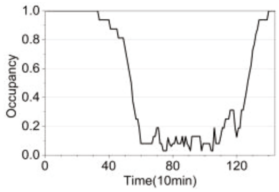
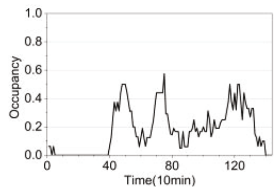
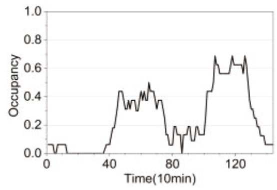
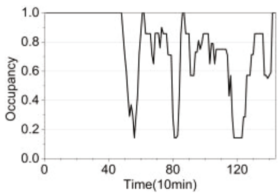
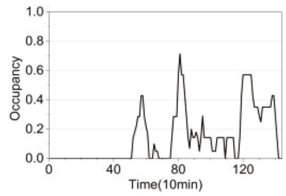
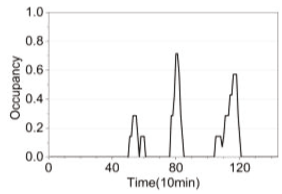
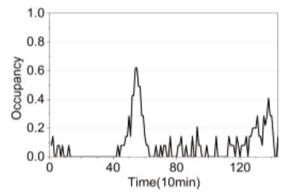
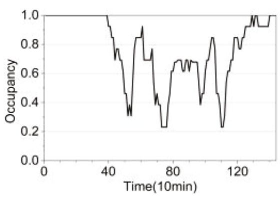
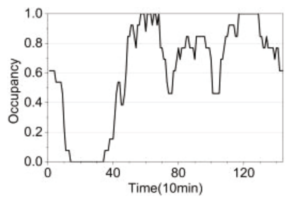
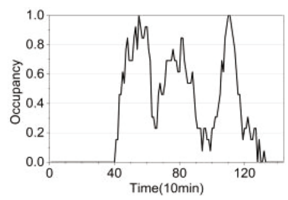
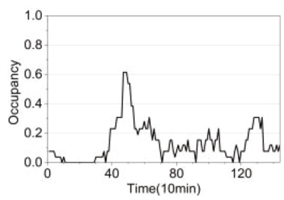
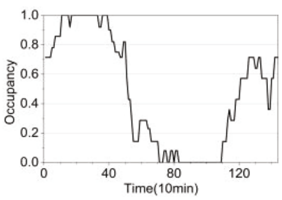
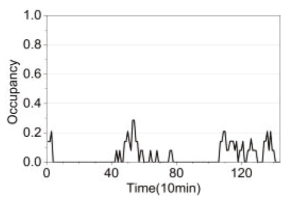
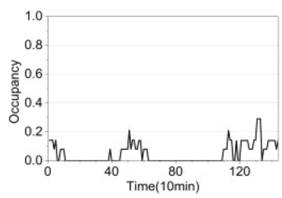
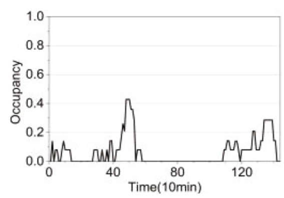
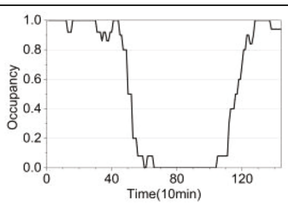
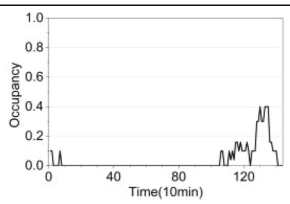
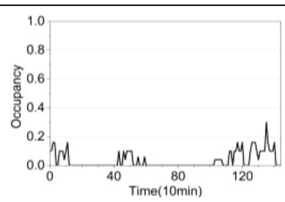
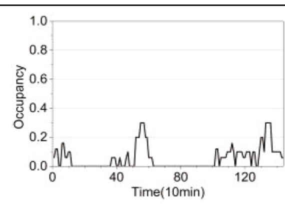
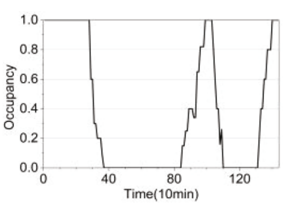
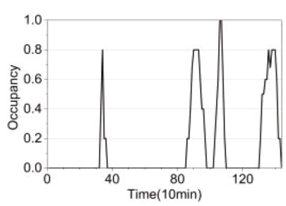
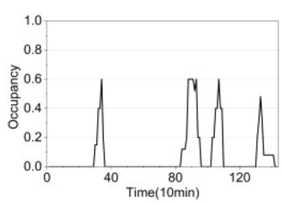
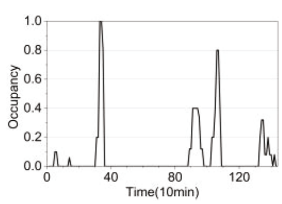
The hourly mean occupancy in the bedroom followed a highly regular pattern. Except for family C, several obvious troughs in the mean occupancy curves were found for each family type. In general, the troughs in the morning, noon, and evening corresponded to lunch, dinner, and breakfast, respectively. However, the times of occurrence and durations of the troughs varied among the different family types, where they occurred in the morning between 7:30 and 9:30 for about 1 h, at noon between 11:00 and 13:00 for about 1.5 h, and in the evening between 16:00 and 20:00 for 1–3 h. The users typically slept from 23:00 to 5:00 and the occupied state was maintained during this period, with a mean occupancy of 1.

The mean occupancy in the living room was characterized by a common pattern but also individual differences. The mean occupancy curve for each family contained multiple peaks and individual differences were found in the duration and height of these peaks. The highest peaks were found for families B and E because they had children, where they mostly resided in the living room during the day and the peak height reached about 0.8. The peak height was the lowest for family C because the members of this family were middle aged with no children. This family spent less time in the living room every day and the peak was about 0.35. The mean occupancy in the kitchen was similar to that in the living room, with three peaks that corresponded to the three main meal times. The mean occupancy in the bathroom had two peaks, with the high mean occupancy peaks in the morning and evening corresponding to waking up and washing before going to bed, respectively. The mean occupancy was low and stable for the rest of the time because the family used the bathroom according to a random pattern.

2) Full-time work families on weekdays

The mean occupancy pattern was simpler for “full-time work families on weekdays” than “stay-at-home families”. For each of the families, the mean occupancy patterns in the living room and bedroom were regular but individual differences were also found, which were strongly related to the working hours of the families. The working hours for families F and G ranged from 9 a.m. to 5 p.m., whereas family H worked from 6 a.m. to 2 p.m. and 4 p.m. to 9 p.m. The commuting times for the three families were 5–10 min. Therefore, the mean occupancy patterns in the bedroom and living room were very similar for families F and G because they had the same working hours, which were consistent with the occurrence times for commuting events. Compared with families F and G, family H had very different working hours. Similarly, the mean occupancy of the bedroom was 1.0 at night and during the rest period in the afternoon. The mean occupancy of the living room was 0.7 at other times because the middle-aged couple in family H preferred to reside in the living room when at home on working days unlike families F and G who had children.

Table 4 Mean occupancy in the four main rooms

No.	Bedroom	Living room	Kitchen	Bathroom
A				
B				
C				
D				
E				
F				
G				
H				

The mean occupancy characteristics of the families described above were strongly related to typical daily events, such as waking up, having three meals, washing, and going to bed. These events determined the duration and height of the peaks and troughs in the mean occupancy curves. In addition, the families had different behavioral habits, but regular patterns and individual differences were found in terms of their mean occupancy characteristics.

3.3 Occupancy modeling

We developed an improved occupancy model based on Wang's model (Wang et al. 2015). The model can output the occupancy state sequence after inputting the initial state of the room, event set, and characteristic parameters (as shown in Figure 8).

The model mainly comprises the following parts.

1) Selecting the rooms for modeling

The status of the occupancy logging equipment in the apartments was only related to whether users were in the room rather than who was in the room and the number of people. In addition, the area of the apartments was small and the frequent random movements of users could lead to the equipment registering the occupation of two rooms at the same time. Therefore, instead of the Lagrange scheme that tracks individual position changes, we selected the Euler method that focuses on specific positions and space as more suitable for our study. We selected the four main rooms for modeling and used (0, 1) to represent (unoccupied, occupied), so each room had a corresponding 2 × 2 state transition probability matrix at each time step. The Euler method reduces the dimension of the matrix from n to 2 (as shown in Figure 9).

2) Definitions of three types of daily events

The function of events is to generate a state transition probability matrix for each time step. According to their randomness, occupancy events can be divided into fixed events (as shown in Figure 10(a)), periodic events (Figure 10(b)), and random events (Figure 10(c)). Fixed events occur during a fixed time period every day. Periodic events also occur during a fixed time period but they are periodic (e.g., three times each week). Random events clearly occur at completely random times, but their total duration each day and the duration of each event are relatively similar (Wang et al. 2015). As shown in Figure 10(c), the total duration of random events in the three days is relatively close, which is 140 min,

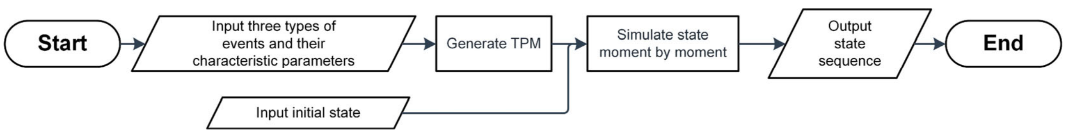


Fig. 8 Overview of the occupancy model

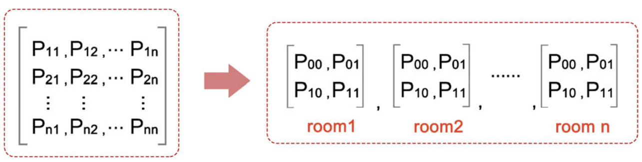


Fig. 9 Reducing the dimensionality of the state transition probability matrix

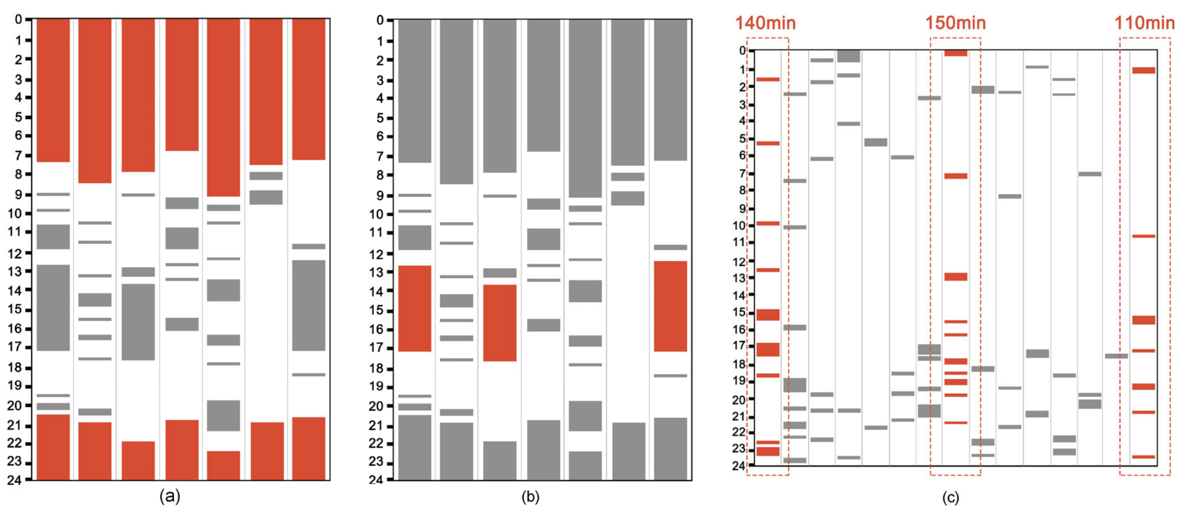


Fig. 10 Daily events: (a) fixed events, (b) periodic events; (c) random events

150 min and 110 min respectively. Thus, each type of event was defined by its characteristic parameters, as shown in Table 5.

Table 5 Characteristic parameters for events

Type	Characteristic parameters
Fixed event	*s_time*, *e_time*, *ranking*, *label*
Periodic event	*s_time; e_time*, *ranking*, *label*, *periodic_probability*
Random event	*s_time*, *e_time*, *ranking*, *label*, *x*, *y*

Notes: "*s_time*" is the earliest time of occurrence for an event, whereas "*e_time*" is the latest time of occurrence. "*ranking*" represents the event priority and 0 is the highest priority. The "*label*" parameter has three types: *a* denotes "*entering the room*" as a fixed event with a status from 0 to 1, *b* denotes "*leaving the room*" as a fixed event with a status from 1 to 0, *c* represents a periodic event, and *d* denotes a random event. "*periodic_probability*" refers to the probability of the daily occurrence of a periodic event. "*x*" is the proportion of the total occupancy duration for a random event per day and "*y*" represents the average duration of each event.

3) Conversion between event characteristic parameters and the corresponding transition probability matrix

The model uses the characteristic parameters for events at each time step to generate the corresponding state transition probability matrix, and matrix is generated for different types of events according to different methods, as follows.

(a) Fixed events: The two types of fixed event comprise the event type with a status from 0 to 1, i.e. "*entering the room*," and the other event type ranges from 1 to 0, i.e., "*leaving the room*".

"*Entering the room*" occurs at a certain time with *c_time* in [*s_time*, *e_time*], and the transition probability matrix within this period can be expressed using Eq. (3). In the effective time period of [*s_time*, *e_time*], the probability of occurrence $p_{0\text{-}1}$ for *c_time* does not change and it is uniformly distributed, where $p_{0\text{-}1}$ is equal to 1/(*e_time* – *s_time* +1). Therefore, in order to ensure that $p_{0\text{-}1}$ remains unchanged, the transition probability matrix is time varying and it can be calculated using Eq. (4) and Eq. (5).

$$\boldsymbol{P}_a = \begin{bmatrix} p_{00} & p_{01} \\ 0 & 1 \end{bmatrix} \tag{3}$$

$$p_{01} = \frac{1}{e\_time - c\_time + 1} \tag{4}$$

$$p_{00} = 1 - p_{01} \tag{5}$$

Similarly, "*leaving the room*" also occurs at a certain time of *c_time* in [*s_time*, *e_time*], and the matrix within this period can be expressed using Eq. (6). The transition probability matrix is time varying and it can be calculated using Eq. (7) and Eq. (8).

$$\boldsymbol{P}_b = \begin{bmatrix} 1 & 0 \\ p_{10} & p_{11} \end{bmatrix} \tag{6}$$

$$p_{10} = \frac{1}{e\_time - c\_time + 1} \tag{7}$$

$$p_{11} = 1 - p_{10} \tag{8}$$

(b) Periodic events: First, it is necessary to obtain the characteristic parameter "*periodic_probability*" to determine whether the periodic event occurs. If the periodic event occurs, it is converted into a fixed event and the method used for calculating the transition probability matrix is the same as that for fixed events.

(c) Random events: The method used for calculating the transition probability matrix for random events is as described by Wang et al. (2015).

4) Overlap avoidance method for active events

At time *c_time*, the proposed model determines the active event set using the effective time period of [*s_time*, *e_time*] for all events. If the number of elements in the active event set is greater than 1, the model applies two methods to avoid overlapping events. In the "priority method," the model conducts screening based on priority by using the event characteristic parameter "*ranking*". In the "last state review" method, if two events have the same priority and the characteristic parameters for "*label*" comprising *a* and *b* are in the active event set at the same time, then their effective time periods of [*s_time*, *e_time*] and [*s_time*′, *e_time*′] overlap (as shown in Figure 11). Therefore, the active event cannot be determined in the overlapping time period of [*s_time*′, *e_time*]. In this case, we need to determine the active event based only on the status *j* in the previous time step of *c_time* – 1. If *j* = 0, the label for the active event is *a*; otherwise, the label is *b* (as shown in Figure 12). These two methods can avoid overlapping to ensure that only 1 event or 0 occurs at all time steps during the day.

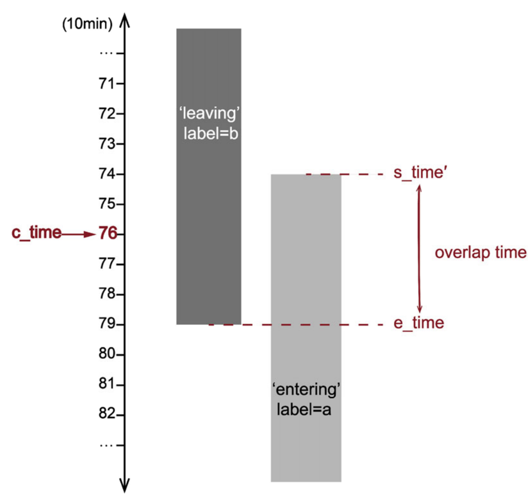


Fig. 11 Schematic diagram illustrating overlapping events

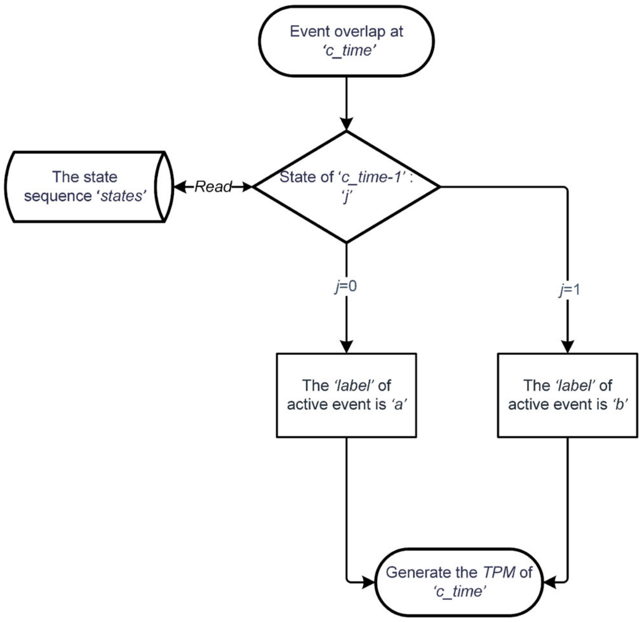


Fig. 12 Last state review method

3.4 Evaluation

Real data were used for simulation in order to evaluate the performance of the proposed model. The occupancy of each room by a family was simulated using the simulation program and evaluated as described above.

3.4.1 Data

Occupancy data covering weeks for family D were used for the evaluation. The data were collected from January to February 2022 using the instruments and methods described above. Occupancy data for the four main rooms comprising the bedroom, living room, kitchen, and bathroom were used.

3.4.2 Results and discussion

Each of the four main rooms occupied by family D was simulated 100 times and the duration of each simulation was one day.

1) Performance of the proposed model

Figure 13 shows the simulated occupancy results obtained in the bedroom by family D for two days. The simulation results for the two days were different because the proposed model is random. In particular, the "*leaving the room*" event (*label* = *b*) occurred later in simulation 2 than the "*entering the room*" event (*label* = *a*) in simulation 1, i.e., *e_time* for *b* was later than *s_time'* for *a*, thereby demonstrating that the model could handle situations where the effective time periods of fixed events with the same priority overlapped and the active event could not be determined (as shown in Figure 11). The times of occurrence and durations of random events (2), (3), (4), (6), (10), (12), (13), and (15) reflected the randomness of random events in the real data. Periodic event (8) occurred in simulation 1 but not in simulation 2, and thus the model could simulate the periodicity of periodic events in the real data.

2) Accuracy of the proposed model

Figure 14 compares the measured and simulated mean occupancy sequences for the four rooms. The simulated sequence was highly consistent with the measured sequence, especially in terms of the simulations of fixed events and periodic events. Table 6 shows the NRMSD values for the

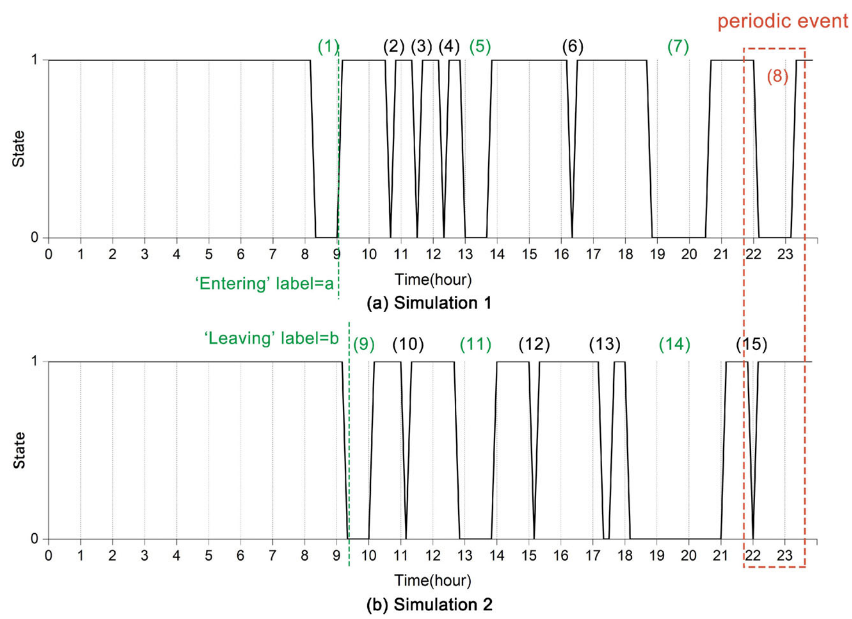


Fig. 13 Results obtained from simulations 1 and 2

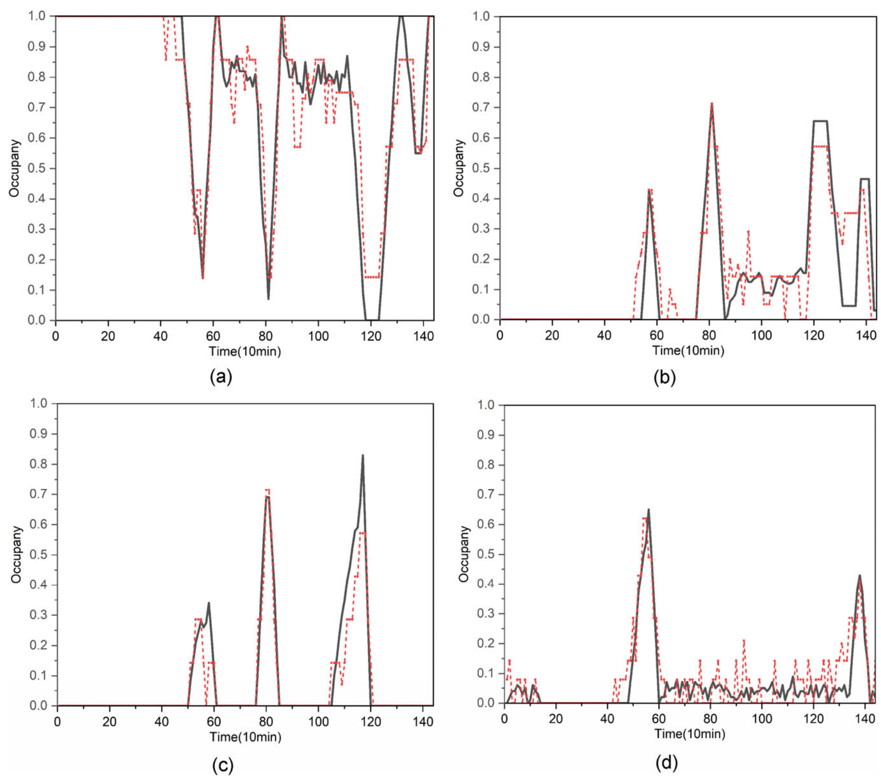


Fig. 14 Measured (red line) and simulated (black line) mean occupancy values in (a) the bedroom, (b) the living room, (c) the kitchen, (d) the bathroom

Table 6 NRMSDs for mean occupancy values

	Bedroom	Living room	Kitchen	Bathroom
NRMSD	0.098	0.135	0.076	0.105

mean occupancy of the four rooms. The proposed model performed well at simulating the mean occupancy in the four rooms, although its effectiveness was relatively poor for the living room. Figure 14(b) shows that the performance of the model was affected by the occurrence of random events at night.

It should be noted that as the simulation time increased, the mean occupancy curve for the period with random events tended to be flat, such as in time steps 80–110 for the bedroom, time steps 85–120 for the living room, and time steps 60–135 for the bathroom (Figure 14). Thus, the times of occurrence and durations of the random events were complex, such as in the kitchen for family A, the living room for family C, and the bathroom for family D (Table 4). The frequent changes in the mean occupancy during the periods with random events could not be fitted well by the proposed model, but the results reflected the average mean occupancy in those periods.

Figure 15 compares the differences between the pmfs for the measured and simulated data for two variables comprising the cumulative occupied duration and number of occupied/unoccupied transitions. The model fitted well to these two variables. Table 7 compares the K–L results for the two variables. Figures 15 shows that the model performed very well in terms of the two variables in all four rooms. The K–L divergence values for the cumulative occupied duration were relatively large in the bedroom, as shown in Figure 15(ai), where the simulated values were slightly smaller than the measured values. It is possible that the duration of fixed events was fairly rigid for this family, whereas the model lacked this constraint.

3.5 Preliminary application

As shown in previous studies, building performance are highly dependent on the occupant behavior (Ren and Yan 2014; McKenna et al. 2015; Jeong et al. 2021; Mitra et al. 2021). Thus, the proposed occupancy model can influence the predictions of other occupant behaviors by accurately predicting the occupancy, and then influence the building performance simulation indirectly. The simulation results of lighting energy consumption are taken as an example to illustrate this point.

Firstly, we constructed a simulation program of lighting behavior by using the coupling relationship between occupancy and lighting behavior (as shown in Figure 16), and used it to simulate switching lights. We focused on explaining that the proposed model should influence building

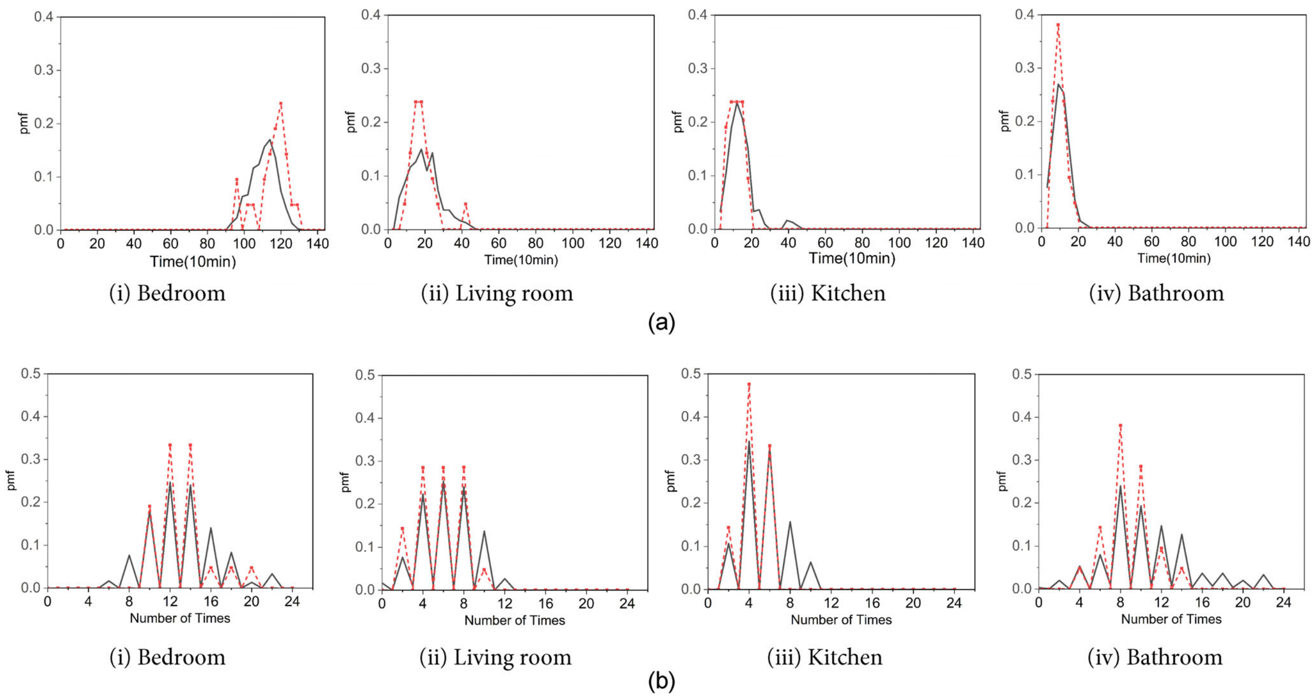


Fig. 15 Measured (red line) and simulated (black line) values for (a) cumulative occupied duration in different rooms, (b) the number of occupied/unoccupied transitions

Table 7 K–L divergence results for cumulative occupied duration and number of occupied/unoccupied transitions

Variable		Bedroom	Living room	Kitchen	Bathroom
K–L divergence	Cumulative occupied duration	0.256	0.130	0.070	0.060
	Number of occupied/unoccupied transitions	0.088	0.106	0.063	0.120

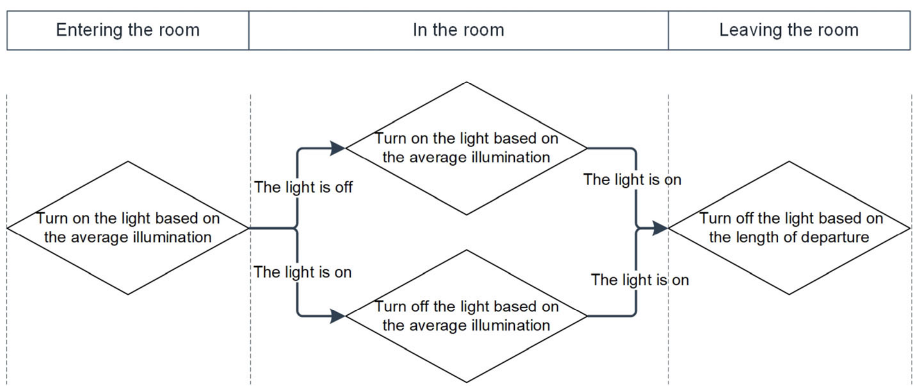


Fig. 16 The coupling relationship between occupancy and lighting behavior

performance simulation in this study, so the method in a previous study was used to construct the simulation program of lighting behavior (Ren and Yan 2014).

We have used three kinds of occupant data. The first one was the lighting utilization rate of four main rooms, i.e., bedroom, living room, kitchen and bathroom, as specified in the existing Chinese standard Design standard for energy efficiency of residential buildings in severe cold and cold zones. The second one was the occupancy data of four main rooms for "stay-at-home families" simulated by the proposed occupancy model, and the third one was the occupancy data of four main rooms for "full-time work families" simulated by the proposed model as well. The lighting energy consumption of the first situation can be calculated directly by the lighting utilization rate. The second and the third situation used the simulation program of lighting behavior mentioned above to predict the lighting behavior of one day, and then calculated the energy consumption.

We performed 365 simulations for both situations 2 and

3 to represent the lighting energy consumption throughout the year. The results showed that there were significant differences in simulations for the three kinds of occupant data, as shown in Figure 17. In terms of the energy consumption of the bedroom, the “full-time work families” was the highest, and the existing standard had the highest energy consumption of the living room, which was 216% of that for the “full-time work families” and 129% of that for the “stay-at-home families”. While the energy consumption for the kitchen and bathroom under all three situations was less according to the results, with the existing standard being higher than the remaining two. Thus, we can see that there are differences in building performance by using the existing standard and the proposed occupancy model.

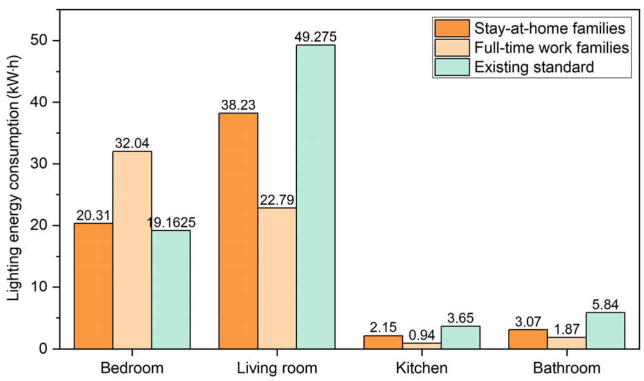


Fig. 17 The lighting energy consumption for three kinds of occupant data

4 Conclusion

In this study, we investigated eight families in cold areas of China and determined the occupancy patterns for their four main rooms. The results showed that the total occupancy durations were approximately the same for the “stay-at-home family” and “full-time work family on weekend” types, and the mean duration was about 9.6 h more than that for the “full-time work family on weekdays” type. The occupancy durations in the four rooms followed the order of: bedroom > living room > kitchen > bathroom. The “full-time work family on weekdays” type spent more time in the bedroom, and the “stay-at-home family” type spent less time in the bedroom than the “full-time work family on weekend” type. Analysis of the hourly mean occupancy showed that the occupancy characteristics were strongly related to daily events, such as commuting, waking up, and eating three meals, where these events could be divided into three categories, and the occupancy characteristics of different family types had similar patterns and individual differences.

We also developed an improved event-based occupancy model using an inhomogeneous Markov chain. An overlap avoidance method for active events with the same priority and periodic events was defined, which enhanced the ability of the model to simulate reality compared with the existing model. A uniform distribution was used to fit the probability distribution for the times of occurrence for events in the effective time period, thereby obtaining more accurate simulations. Finally, the model was evaluated using real data and the results showed that the model performed well in terms of simulating the mean occupancy, cumulative occupied duration, and number of occupied/unoccupied transitions.

The study mainly focused on two types of families comprising “stay-at-home family” and “full-time work family” types. In future research, the types of families investigated should be further subdivided and also expanded in order to more accurately and comprehensively describe the occupancy of residential buildings by users. And this study is aimed to propose an accurate occupancy model, the application of the model in building performance simulation and the accuracy improvement of building performance simulation should be further verified and studied.

Acknowledgements

The authors gratefully acknowledge the funding support from the National Natural Science Foundation of China (No. 52008129), the Postdoctoral Science Foundation of China (No. 2019M651289), and the National Natural Science Foundation of Heilongjiang Province (No. LH2020E051, No. GZ20210211).

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

Author contribution statement

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Qi Dong and Zikai Ma. The first draft of the manuscript was written by Zikai Ma and all authors commented on previous versions of the manuscript. Cheng Sun guided and revised the manuscript. All authors read and approved the final manuscript.

References

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

Blight TS, Coley DA (2013). Sensitivity analysis of the effect of occupant behaviour on the energy consumption of passive house dwellings. *Energy and Buildings*, 66: 183–192.

Chang WK, Hong T (2013). Statistical analysis and modeling of occupancy patterns in open-plan offices using measured lighting-switch data. *Building Simulation*, 6: 23–32.

Chen Z, Xu J, Soh YC (2015). Modeling regular occupancy in commercial buildings using stochastic models. *Energy and Buildings*, 103: 216–223.

Cover TM, Thomas JA (2006). Elements of information theory, 2nd edn. New York: John Wiley & Sons.

Crawley DB, Lawrie LK, Winkelmann FC, et al. (2001). EnergyPlus: Creating a new-generation building energy simulation program. *Energy and Buildings*, 33: 319–331.

Diao L, Sun Y, Chen Z, et al. (2017). Modeling energy consumption in residential buildings: A bottom-up analysis based on occupant behavior pattern clustering and stochastic simulation. *Energy and Buildings*, 147: 47–66.

Dong B, Lam KP, Neuman CP (2011). Integrated building control based on occupant behavior pattern detection and local weather forecasting. In: Proceedings of the 12th International IBPSA Building Simulation Conference, Sydney, Australia.

Fajilla G, Austin MC, Mora D, et al. (2021). Assessment of probabilistic models to estimate the occupancy state in office buildings using indoor parameters and user-related variables. *Energy and Buildings*, 246: 111105.

Flett G, Kelly N (2016). An occupant-differentiated, higher-order Markov Chain method for prediction of domestic occupancy. *Energy and Buildings*, 125: 219–230.

Flett G, Kelly N (2021). Modelling of individual domestic occupancy and energy demand behaviours using existing datasets and probabilistic modelling methods. *Energy and Buildings*, 252: 111373.

Hong T, Taylor-Lange SC, D'Oca S, et al. (2016). Advances in research and applications of energy-related occupant behavior in buildings. *Energy and Buildings*, 116: 694–702.

Jeong B, Kim J, de Dear R (2021). Creating household occupancy and energy behavioural profiles using national time use survey data. *Energy and Buildings*, 252: 111440.

Jiang Y, Hu S (2021). Paths to carbon neutrality in China's building sector. *Journal of HV&AC*, 51(5): 1–13. (in Chinese)

Li Y, Yamaguchi Y, Shimoda Y (2022). Impact of the pre-simulation process of occupant behaviour modelling for residential energy demand simulations. *Journal of Building Performance Simulation*, 15: 287–306.

Liao C, Lin Y, Barooah P (2012). Agent-based and graphical modelling of building occupancy. *Journal of Building Performance Simulation*, 5: 5–25.

Malekpour Koupaei DM, Cetin KS, Passe U (2022). Stochastic residential occupancy schedules based on the American Time-Use Survey. *Science and Technology for the Built Environment*, 28: 776–790.

McKenna E, Krawczynski M, Thomson M (2015). Four-state domestic building occupancy model for energy demand simulations. *Energy and Buildings*, 96: 30–39.

Mitra D, Chu Y, Cetin K (2021). Cluster analysis of occupancy schedules in residential buildings in the United States. *Energy and Buildings*, 236: 110791.

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

Ren XX, Yan D (2014). A study of lighting energy consumption model for office buildings based on occupant behavior. *Building Science*, 30(6): 1–9. (in Chinese)

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

Rueda L, Sansregret S, Le Lostec B, et al. (2021). A probabilistic model to predict household occupancy profiles for home energy management applications. *IEEE Access*, 9: 38187–38201.

Salimi S, Liu Z, Hammad A (2019). Occupancy prediction model for open-plan offices using real-time location system and inhomogeneous Markov chain. *Building and Environment*, 152: 1–16.

Serfozo R (2009). Basics of Applied Stochastic Processes. Heidelberg, Germany: Springer.

Sun K, Yan D, Hong T, et al. (2014). Stochastic modeling of overtime occupancy and its application in building energy simulation and calibration. *Building and Environment*, 79: 1–12.

Wang D, Federspiel CC, Rubinstein F (2005). Modeling occupancy in single person offices. *Energy and Buildings*, 37: 121–126.

Wang C, Yan D, Jiang Y (2011). A novel approach for building occupancy simulation. *Building Simulation*, 4: 149–167.

Wang C (2014). Simulation research on occupant energy-related behaviors in building. PhD Thesis, Tsinghua University, China. (in Chinese)

Wang C, Yan D, Feng X, et al. (2015). A Markov chain and event based model for building occupant movement process. *Building Science*, 31(10): 188–198. (in Chinese)

Wang Y, Gu BH (2021). New process of China's sustainable development: Exploring the road towards carbon neutrality. *China Sustainability Tribune*, 2021(6): 15–20. (in Chinese)

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

Yamaguchi Y, Shimoda Y (2017). A stochastic model to predict occupants' activities at home for community-/ urban-scale energy demand modelling. *Journal of Building Performance Simulation*, 10: 565–581.

Yan D, O'Brien W, Hong T, et al. (2015). Occupant behavior modeling for building performance simulation: Current state and future challenges. *Energy and Buildings*, 107: 264–278.

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

Zhang Y, Bai X, Mills FP, et al. (2018). Rethinking the role of occupant behavior in building energy performance: A review. *Energy and Buildings*, 172: 279–294.

Zhang L, Huang X, Chen YY (2019). Literature review on residential energy behavior. *Urbanism and Architecture*, 16(2): 160–164. (in Chinese)