Ali Akhondzada

Modelling the Occupancy Profile Deterministically, Probabilistically and Stochastically

Literature Review

Metropolia University of Applied Sciences

Bachelor of Engineering

Sustainable Building Engineering

Bachelor's Thesis

30 May 2017

Author Title Number of Pages Date	Ali Akhondzada Modelling the occupancy profile deterministically, probabilistically and stochastically 38 pages 30 May 2017
Degree	Bachelor of Engineering
Degree Programme	Civil Engineering
Specialisation option	Sustainable Building Engineering
Instructor	Jarek Kurnitski, Adjunct Professor Jorma Säteri, Head of Department Kaiser Ahmed, MSc.

This bachelor's thesis aimed at analyzing and comparing the properties and principles of deterministic, probabilistic and stochastic occupancy approaches using various models. For each model, the results of this analytical assessment were compared against either the outcomes of similar models or the results of measured databases. The purpose was to evaluate the performance of the different occupancy models.

Literary sources, such as articles and books were reviewed. For the assessment of the functionality of the models functionality, algorithms presented in each model were followed step by step and their formulas and equation sets tested with random numbers.

The final results showed that the models based on probabilistic and stochastic approaches could simulate the occupancy rate more accurately compared to the models based on the deterministic approaches. The results of the methodological review also showed that each model can capture a certain number of diversity factors of occupancy rate. The findings from this thesis can be used for developing a new occupancy model based on a combination of both stochastic and probabilistic approaches. Furthermore, the developed model can be integrated in developing a building simulation tool.

Keywords	occupancy profile, stochastic approach, probabilistic approach

Acknowledgment

This study was carried out at the Department of Civil Engineering, Aalto University. Firstly, I am heartily thankful to my supervisor, Professor Jarek Kurnitski, for this great opportunity, which enabled me to develop a deep understanding of the subject. My sincere thanks go also to my other supervisor Dr. Kaiser Ahmed for his introduction to the subject and his guidance during writing of the thesis. I would also like to extend my thanks for the constant support of the Head of Civil Engineering Department at Metropolia University of Applied Sciences Mr. Jorma Säteri. My greatest thanks go to my mom, who always supported and encouraged me to study.

Contents

Abbreviations

ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers
CDF	Cumulative Distribution Function
HAC	Hierarchical Agglomerative Clustering
HBS	Household Budget Survey
HMC	Homogeneous Markov Chain
HVAC	Heating, Ventilation and Air Conditioning
HW	Hallway
MLE	Maximum Likelihood Estimation
MOMZ	Multi-Occupants Multi-Zone
MOSZ	Multi-Occupants Single-Zone
MUMO	Multiple Modules
PDF	Probability Distribution Function
PMF	Probability Mass Function
PZ	Primary Zone
RR	Restroom
SOSZ	Single-Occupant Single-Zone
SZ	Secondary Zone
TUS	Time-Use Survey

1 Introduction

Buildings account for nearly 40% of the total energy usage in European countries [1]. The energy consumption of buildings is influenced by several factors including the physical properties of the building, heating, ventilation and air conditioning (HVAC) systems, lighting, geometry, occupancy and the behaviour of the occupants. Also, a large percentage of energy is used to maintain a comfortable and healthy indoor environment for the occupants. [2.] To provide an acceptable indoor thermal condition and adequate ventilation in a building, it is necessary to know the occupancy rate in the building. Occupancy rate is one of the key factors in HVAC system sizing, and in most cases the maximum occupancy rate is used in calculations due to a lack of information about the actual occupancy information. Vieira reported an oversizing of HVAC systems due to overestimating the peak occupancy rate. [3.] This overestimation, which is the result of insufficient information of occupancy rate, causes more energy consumption in buildings. Also, the internal heat gain from lighting and appliances, together with the occupancy rate may change the heating and cooling demand. Therefore, realistic characterization of a building's occupancy is required for a reliable prediction of the energy performance of a building. Up to 30% of the energy used by HVAC systems can be saved by implementing occupancy-driven HVAC control strategies [4]. Different models have been developed to simulate the occupancy rate in both residential and commercial buildings. These models are used in simulation tools such as DeST, EnergyPlus and ESP-r [5;6;7].

In this thesis, deterministic, probabilistic and stochastic approaches of occupancy modelling are introduced and the methodologies used in the occupancy models are discussed in detail. The formulas and equation sets used in the methodologies are tested with random numbers and the results are compared either against similar models or measured databases. The study focuses on the stochastic and probabilistic approaches, due to several diversity factors that they can capture.

2 Occupancy Modelling

To simulate the building occupancy profile, different methods have been introduced. In each, certain characteristics of occupancy have been emphasized. The deterministic approach is one of the first methods used in occupancy modelling. Later, using the recorded behaviour of occupancy, probabilistic approaches were introduced. To minimize the discrepancy between real and simulated occupancy modelling, stochastic approaches have been introduced, which can capture more unpredicted occupancy behaviours.

2.1 Deterministic Approach

In deterministic models, occupancy rate is determined based on a regular daily schedule varying between zero for unoccupied and one for occupied spaces. The models use the occupancy schedules of existing buildings to simulate the occupancy rate for similar buildings. The presence of occupants in a specific zone depends on many factors, and further a factor that keeps the occupant in a specific zone does not guarantee the presence of him/her at the same place next time. [8.] For example, cooking keeps the occupant in the kitchen for Saturday, but next Saturday he/she may go out to eat at the same time. These uncertainties are the results of the stochastic nature of human behaviour. The deterministic models, therefore, cannot take all diversity factors for occupant presence into account.

2.1.1 Deterministic Occupancy Profile in ASHRAE

The American society of Heating, Refrigerating and Air-conditioning Engineers (ASHRAE) is a society that focuses on energy efficiency, building systems, indoor air quality, sustainability technologies and refrigeration. ASHRAE standard 90.1-2004 provides guidance for energy-efficient new building design required in the deterministic approach. [9.] Studies show that there are some differences between ASHRAE reference data and measured data. As an example, Duarte et al. revealed a 45% and a 12% reduction in daily average occupancy peaks for a single-person office and shared office room, respectively, compared to the ASHRAE reference data. A comparison of the ASHRAE reference data to the data collected from 223 single-person offices shows similarities between two simulation models, especially in capturing diversities factors. However, there are two significant differences as well. In the ASHRAE 90.1 2004 guidelines, the diversity factors peak at 95%, while in the measured single-office data they peak at 50%. [10.]

2.1.2 Discrete Deterministic Occupancy Profiles

In another study, Aerts et al. developed a set of occupant profiles based on the deterministic approach that took a number of variables, including employment, income, household size and age into consideration, as illustrated in figure 1. The occupant profiles determine how long the occupants spend their time at home on weekdays and weekends. The data was then categorized in seven profiles according to the similarities, to check what factors are more effective for the occupancy rate of residential buildings. [11.]

The profiles show three possible states: at home and awake, at the home and sleeping, or absent. They were constructed from a cluster analysis on the 2005 Belgian time-use survey (TUS). The Belgian survey included the activity data of 6400 individuals from 3474 households. To determine the variations in the Belgian behaviour, a hierarchical clustering on the TUS data was used. The seven developed profiles (for weekdays, Saturdays and Sundays) show that there is a relationship between the employment of occupants and the occupancy rate. The full-time employed occupants are largely represented by two profiles with low occupancy levels during day time. Conversely, the unemployed or retired occupants are mainly situated in the profiles 6 and 7, which show a high occupancy rate. The occupant profiles show the full-time employed occupants, with an average age between 25-39 years and an average income between $ 1000 to 1500 spend their daytime outside and probably at their work place. The unemployed occupants, with the same age and almost the same monthly income, spend most of their weekdays at home.

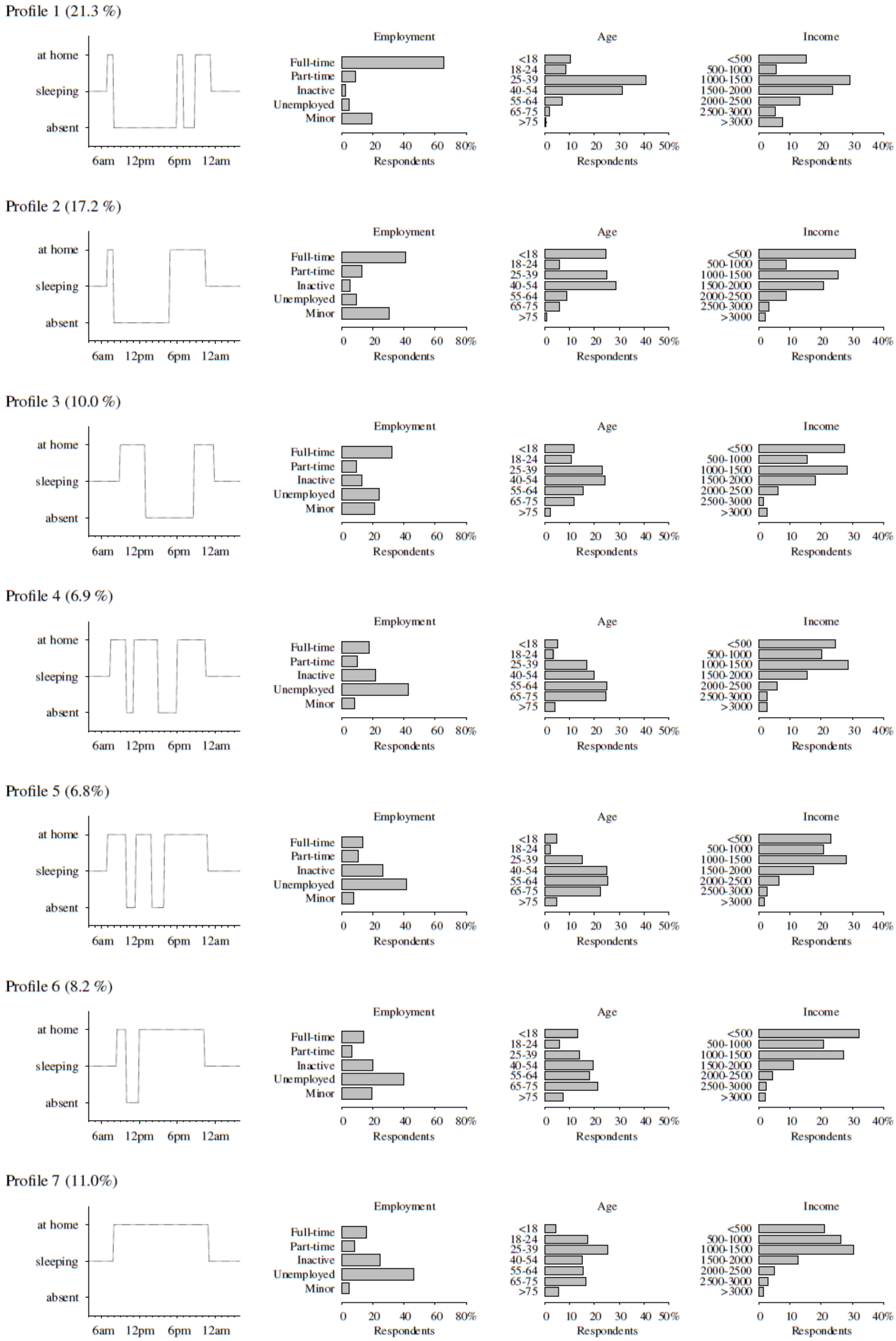

Figure 1. Occupancy profile for weekdays [11].

In order to minimize the discrepancy between the results of a simulated model and the real performance of buildings, a more realistic model is needed to take the uncertainties of occupant presence, randomness of behaviour, as well as time variation of behaviour into account. [11.]

2.2 Probabilistic Approach

The probabilistic approach uses statistical data to predict the occurrence of certain actions in buildings. Probabilistic models use the observed data from the past to provide models which can predict the occupancy in the future. [12.] Compared to the deterministic approach in which the direct correlation of certain action or time and occupancy rate determines the occupancy rate, probabilistic models can capture more variations in the occupancy rate.

2.2.1 Probabilistic Occupancy Model for Single-Person Offices

Wang et al. presented a probabilistic model for the occupancy of a single-person office. The study examined the statistical properties of occupancy obtained from 35 single- person offices during one year. The occupancy logs in this model are taken from motion sensors in each office. The space is recorded occupied when the office is vacant initially and the sensors detect motion before 15 minutes, and the space is recorded vacant when the sensors do not detect any motion for a time interval of 15 minutes. The model is used a heterogeneous Poisson process in which two different exponential distributions explains the occupancy and vacancy of a single-person office. Occupancy and vacancy intervals are the two exponential distributions of the Poisson process. It is assumed that the occupancy and vacancy intervals are independent and sequential random variables. [13.] The results of simulation using the Wang model are compared against the results of measured data, which have been collected from the actual time use data of office spaces. This comparison is illustrated in table 1.

The probability of a transition between the states (e.g. occupied to vacant) in a specified time interval can be determined using the Poisson process as

$$p(k\ in\ T) = \frac{\lambda^k . e^{-\lambda}}{k!}, \tag{1}$$

where k is the number of events (e.g. occupied to vacant) in a time interval T (e.g. one day). The parameter λ is the average number of events in a unit of time.

Table 1. Comparison of simulation and measurement [13].

	Measurement	**Simulated**
Total number of analysed days per year	171	171
Average occupied hours per day	6.17	6.47
Average vacancy ratio per day	0.20	0.33
Average departures and arrivals per day	**4.93**	**5.51**

For instance, using the data provided in table 1 the average number of departures per day in the simulation is 5.51 (λ=5.51) and whereas in the measured data the number is 4.93. [13.] Using equation (1), the probability of having 1 to 12 departures, which is assumed to be the maximum number of departures per day can be calculated as,

$$p(k = 1,2, \ldots ,12 \; in \; 24h) = \frac{5.51^{1,2,\ldots,12}.\, e^{-5.51}}{1,2, \ldots ,12!}$$

For example, the probability of having exactly 1, 4, 7 or 11 departures per day would be calculated as below, respectively

$$p(1 \; in \; 24h) = \frac{5.51^{1}.\, e^{-5.51}}{1!} = 0.022$$

$$p(4 \; in \; 24h) = \frac{5.51^{4}.\, e^{-5.51}}{4!} = 0.155$$

$$p(7 \; in \; 24h) = \frac{5.51^{7}.\, e^{-5.51}}{7!} = 0.123$$

$$p(11 \; in \; 24h) = \frac{5.51^{11}.\, e^{-5.51}}{11!} = 0.014.$$

Using the average number of departures per day provided in table 1 (bolded) and the equation (1), the distribution of the number of departures per day or transition from occupied state to vacant state, for the whole office is illustrated in figure 2 for both simulated and measured data. Additionally, Wang provided the distribution of number of departures per day for a randomly selected office. The distribution is provided based on the measured data. [13.] To make an easier comparison the number of departures from the randomly selected office is also illustrated in figure 2. Since the exact occupancy values for randomly selected office were not available the numbers are taken from the figure from, the Wang model therefore there might be small assumption error.

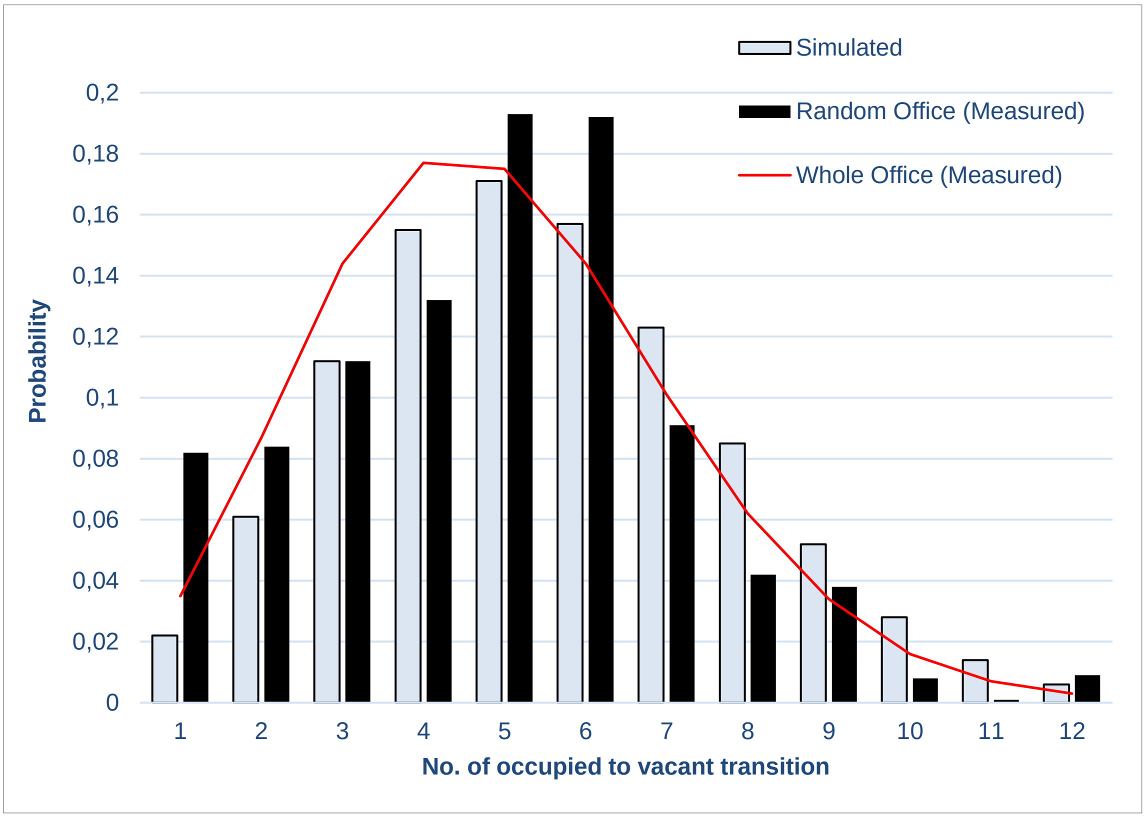

Figure 2. Probability distribution of the number of departures per day.

The results display that the most probable number of transitions (departures) is five per day for the whole office building according to the simulation, while the measured data shows that four transitions per day is more probable. The random office also shows five departures per day as the most probable number.

The lengths of the intervals are considered to be an exponentially distributed random variables. The parameters of the exponential distribution are determined using the Maximum likelihood estimation (MLE). The probability distribution function (PDF), is defined using alternative parameterization

$$f(x;\beta) = \begin{cases} (1/\beta)e^{-x/\beta}, & x \geq 0 \\ 0, & x < 0. \end{cases} \tag{2}$$

The parameter's mean (β) and variance are estimated using Maximum likelihood estimation (MLE). To examine the exponential distribution, the dispersion parameter $\emptyset$ should be equal to 1, and the $\widehat{\emptyset}$ must be correct,

$$\widehat{\emptyset} = \frac{1}{(n-1)} \sum_{i=1}^{n} \left((x_i - \hat{\beta})/\hat{\beta} \right)^2. \tag{3}$$

For example, the length of the occupancy and vacancy intervals of the random office is calculated to be $\hat{\beta}_{occupied} = 72.8\, min$ and $\hat{\beta}_{vacancy} = 42.6\, min,$ respectively. [13.] A scaled deviance is used to evaluate the goodness of fit of the model

$$D * (x; \hat{\beta}) = 2 \sum_{i=1}^{n} \frac{\log\left(\frac{\hat{\beta}}{x_i}\right)}{\hat{\emptyset}}. \tag{4}$$

The exponential distribution model is accepted if the above scaled deviance is smaller than $X^2_{n-1;1-\alpha}$, where α is 0.05. The results from the Wang models show that the occupancy intervals of randomly selected is not exponentially distributed because the $D * (x; \hat{\beta})_{occu} = 2853.5 > X^2_{n-1;1-\alpha} = 859.6$. The scaled deviance of vacancy intervals has a chi-squared distribution with $n-1$ degree of freedom, therefore the model of exponential distribution for vacancy intervals is accepted, or $D * (x; \hat{\beta})_{vac} = 516.8 < X^2_{n-1;1-\alpha} = 797.0$. [13.]

The PDFs of the randomly selected office show that the length of occupancy intervals is not exponentially distributed while the lengths of vacancy intervals are exponentially distributed as shown in figure 3.

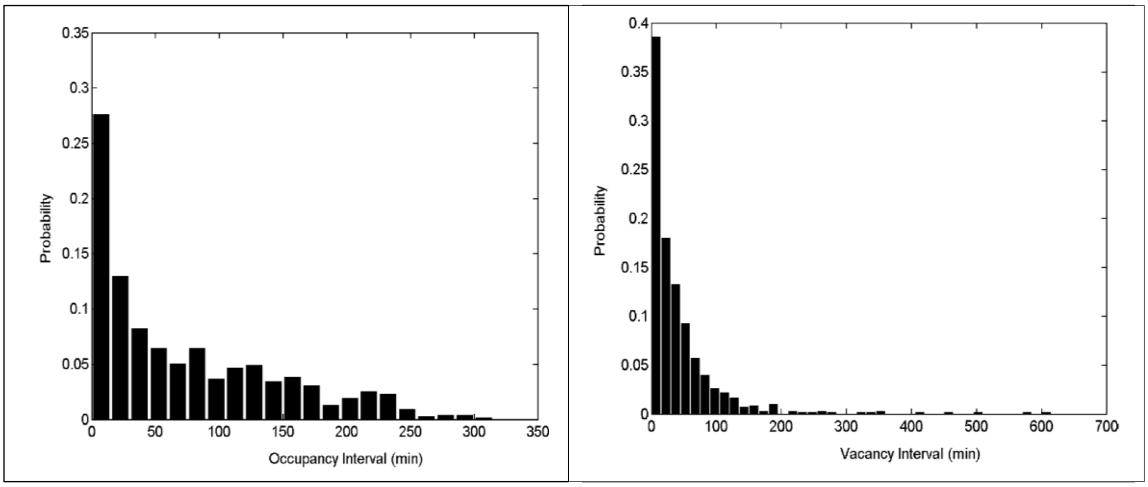

Figure 3. Probability distribution of the vacancy intervals for a randomly selected office based on the observed data [13].

The frequencies of the lengths of both occupancy and vacancy intervals for the randomly selected office are plotted in figure 4 according to the observed and fitted data. As figure 4 illustrates, exponential distributions fit better for longer intervals. Both distributions underestimate the frequencies for intervals that last less than 15 minutes.

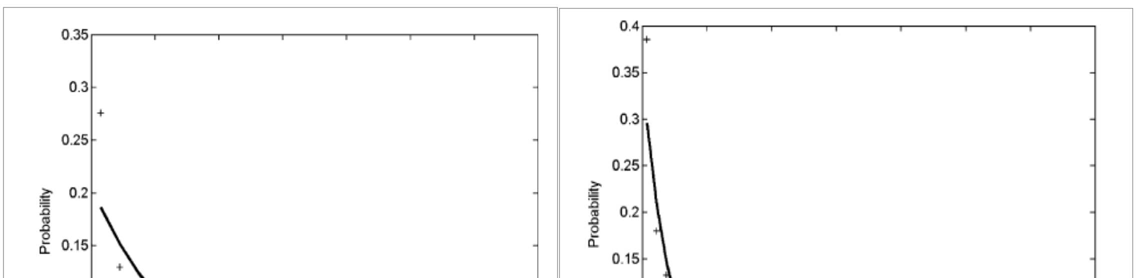

Figure 4. Occupancy and vacancy intervals resulted from fitted and observed data [13].

This non-homogeneous model has been tested on 35 office rooms and the results indicate that the vacancy intervals are exponentially distributed while for the occupancy intervals the exponential distribution is rejected. This deficiency might be caused by the presence of more than one occupant in a room. In this case, a single motion sensor cannot precisely tell the arrival time and duration of stay for each occupant individually. [13.]

2.2.2 Three-states Probabilistic Based Model

In another probabilistic model, Aerts et al. developed a probabilistic occupancy model which has three possible states; (1) at home active (awake), (2) at home inactive (asleep), and (3) absent [14]. The model predicts the probability of occupancy sequences for each occupant using statistical data collected from two surveys, the Belgian Time-Use Survey (TUS) and the Belgian Household Budget Survey (HBS) conducted in 2005 [15]. The average occupancy profile is illustrated in figure 5. The three states of occupancy which are present, absent and sleeping are shown in respect to the time of day. The vertical axis represents the fraction of being present in each of these states (at home, sleeping, absent) and the horizontal axis shows the time. Between midnight and 04:00 AM the majority of people are at home and asleep, after 05:00 AM the number of people asleep decreases gradually. They are either at home and awake or they are absent. Between 07:00 AM to 04:00 PM the probabilities of being active and absent are distributed equally. The number of active occupants increases after 04:00 PM which is the time people usually come back from work. [14.]

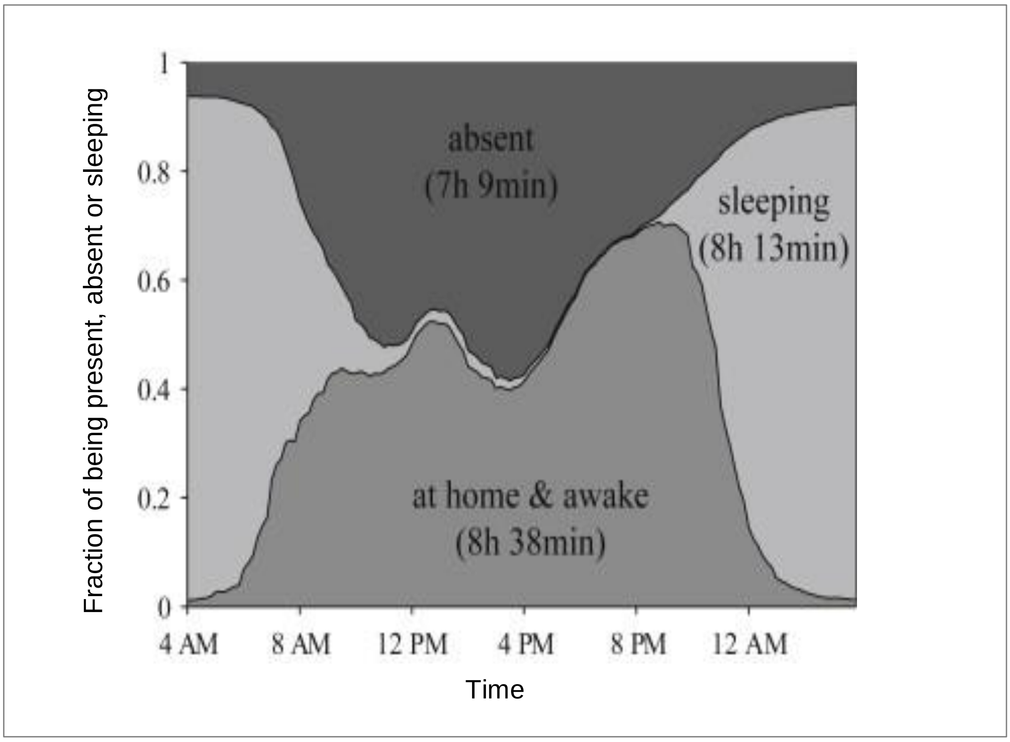

Figure 5. The average occupancy profile based on TUS dataset [14].

Transition probability together with duration probability are the two main variables in the model, they are both time dependent. The initial state in the chain is determined using the presence proportion observed in the TUS data. From the first state onward a 3*3 matrix (5) explains the probabilities

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}. \quad (5)$$

In the transition probability matrix, each element demonstrates the probability of a transition to the next state given the current state. The transition probability for the previous state “at home” to the next state “absent” increases during the day time for 08:00 AM to 08:00 PM while the probability of transition to the next state “Sleeping” with the same previous state is at the minimum level during the day time (08:00 AM-80:00 PM). Based on the observed data, the transition between the states at home and absent or sleeping is time dependent, while the probability of transition between the state sleeping to the state at home is almost one during the day time. [14.]

In addition to the transition probability in this model, duration probability is another important factor. It depends on the current state and the time a person stays in that state. Based on the TUS data, the duration of at home is influenced strongly by the starting time. [14.] For instance, in the figure 6, it is illustrated that the duration of at home in the afternoon (after 02:00 PM to 04:00 PM) is longer than in the evening, when individuals go from being active to being inactive or sleeping.

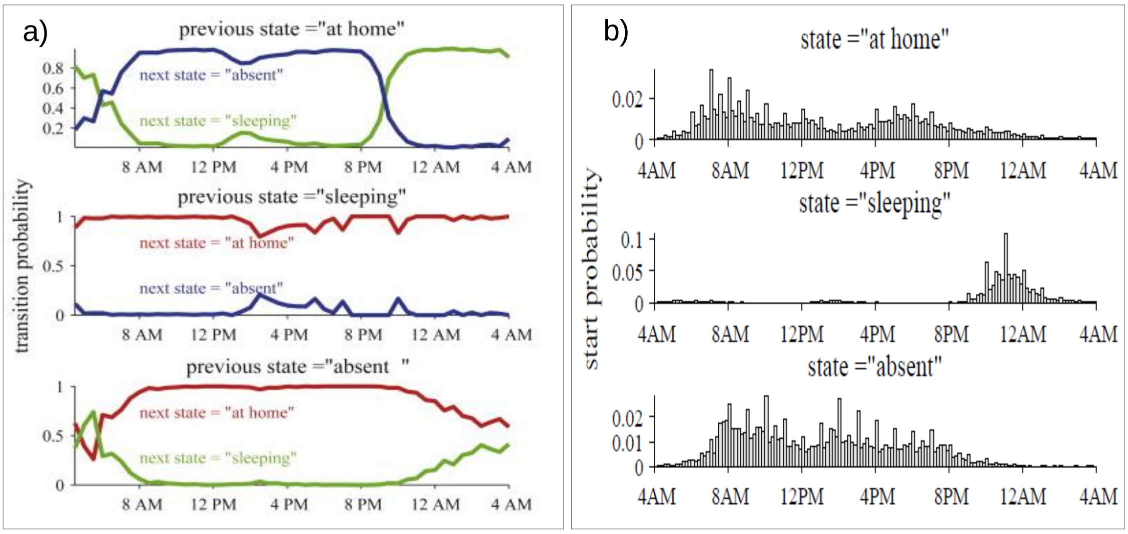

Figure 6. (a) The transition probability as a function of the previous state and the time, and (b). The start probability of an occupancy state as a function of time [14].

In order to increase the accuracy of the model, seven categories have been defined for occupancy using the hierarchical agglomerative clustering (HAC) method. HAC categorizes the elements in dataset hierarchically based on their similarities. Clusters with similarities are merged into pairs using linking method and this process continues till making one cluster. The categories produced by the HAC method are “mostly absent”, “mostly at home”, “very short daytime absence”, “night-time absence”, “daytime absence”, “afternoon absence” and “short daytime absence”. The main differences between the categories are the time spent at home, the time of transition between states, and the number of transitions in a day. Aerts’ model is calibrated using the Belgian Time Use Survey (TUS) database and the Belgian Household Budget Survey (HBS) [15]. The database is extracted from the results of 6400 respondents of 3455 households. In the time- use survey weekdays and weekends are considered separately. A time slot of 10 minutes is determined with at least one activity in each. [14.]

2.2.3 Event Driven Framework for Occupancy Simulation

In an agent based model, Gunathilak et al. proposed a generalized event-driven framework for the simulation of building occupancy. Unlike other agent based models, this model can simulate a variable number of agents. In the framework, the authors used categorized events as inputs of an algorithm, the outcomes are a user log and a building occupancy list. The first output provides information about the events that each individual is involved in, and the second one presents information about the number of occupants present in a specified time slot. The study claims that the model can capture sudden peaks and drops in occupancy rate with a good accuracy for both office and laboratory environments [16.]

2.3 Stochastic Approach

In order to simulate the building occupancy more accurately, recent studies have focused on the modelling of occupancy profile with stochastic approaches. In a stochastic model, the occupancy status is considered as a random variable, and at a certain time the prediction is done according to the previous status, not the long-term historical data. [17.] The main principle behind this method is the Markov chain. The Markov chain is memoryless, which makes the method independent form other occupancy statuses but the previous one. Therefore, the initial state and the transfer probability are the important parameters in a stochastic process. [18.] However, each stochastic model emphasizes different parameters. In some models the limited occupancy states are used to predict the occupancy rate in buildings, others consider more states to cover all possible events. [19;20.]

2.3.1 Movement Based Stochastic Occupancy Model

Wang et al. developed a stochastic model for building occupancy. It sees the occupancy as a result of occupant movement inside and outside a building. The homogeneous Markov chain (HMC) method is used to simulate the occupancy of the building. The main properties in the HMC are the current state and the transition probability from one state to another. To generate the location of an occupant at each time slot, a module called “movement process” is used. Further, the final module to complete the Markovian property is “events”. The “events” module is used to determine the transition probabilities. In the Wang model, three assumptions are made in order to simulate the occupancy status: 1) the occupant location has a Markovian property, 2) each time step is long enough for changing location and 3) occupant movement is an independent process. The movement of an occupant is driven by a number of events and in some cases the movements are driven by multi-events. Therefore, a priority system has been defined to consider the events according to their importance and effectiveness. [21.]

To simulate the occupancy in an office building the movement process has been defined as “walking around”, “going to and coming back from lunch”, “going to and getting off work”. This process is illustrated in figure 7.

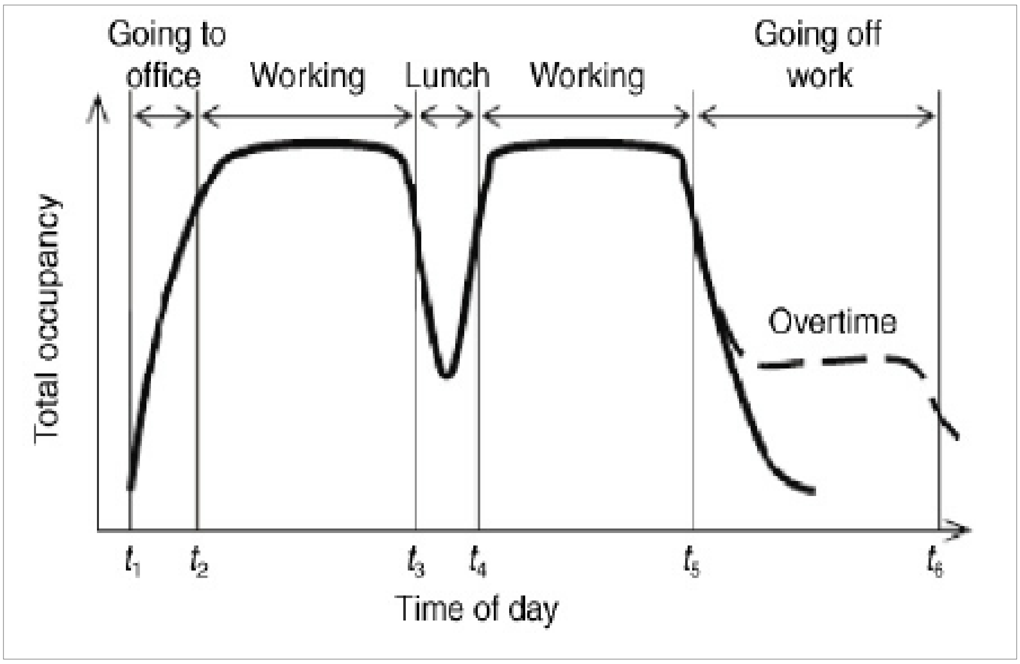

Figure 7. The occupancy rate of the sample office building in a working day [21].

The total occupancy increases between t_1 and t_2, the morning arrival time to the office, and decreases due to night departure between t_5 and t_6. In between the total occupancy is almost constant during working hours except for the lunch break which is between t_3 and t_4. The event of “walking around” includes moving inside the office and going outside during business time (8:00 AM- 05:00 PM). “Leaving for launch” is excluded from the “walking around” as it is described in a separate category. [21.]

Each of the events is determined with different formulas and equation sets. The stochastic process can be explained using an ergodic Markov chain with a stationary distribution.

The transition probability matrix for walking around between spaces is defined as

$$P = \begin{bmatrix} P_{00} & P_{01} & \cdots & P_{0n} \\ P_{10} & P_{11} & \cdots & P_{1n} \\ \vdots & \vdots & & \vdots \\ P_{n0} & P_{n1} & \cdots & P_{nn} \end{bmatrix}. \tag{6}$$

The model uses an ergodic Markov chain and the stationary distribution of the ergodic Markov chain is denoted by $\pi = (\pi_1, \pi_2, ..., \pi_n)$ where π_i is the long-run proportion of time that the Markov chain remains in state i. The vector π can be specified with the following equations

$$\sum_{t=0}^{n} \pi_i = 1 \tag{7}$$

$$\pi = \pi P\ . \tag{8}$$

By giving the transition probability matrix P to the equation (7), the stationary distribution vector will be determined. The Equation (7) shows that the sum of all probabilities in a vector π is equal to one. The Equation (8) tells that after long a run the vector π will give steady state and the multiplication of the transition matrix P does not affect the results. Using equation (7) and the fact mentioned in equation (7), given the transition matrix P, the vector π can be specified. In this study the π is determined using a transpose vector denoted by π^T. [21.]

$$A = \begin{bmatrix} P_{00} & P_{01} & \cdots & P_{0n} \\ P_{10} & P_{11} & \cdots & P_{1n} \\ \vdots & \vdots & & \vdots \\ P_{n0} & P_{n1} & \cdots & P_{nn} \end{bmatrix} \qquad b = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \pi^T = A^{-1} b\ . \tag{9}$$

The time that an occupant stays at a certain state, e.g. i is distributed geometrically, and the expected value for this distribution is determined as

$$E(ST_i) = \frac{1}{1 - P_{ii}} \qquad \rightarrow \qquad P_{ii} = 1 - \frac{1}{E(ST_i)} \qquad (10)\,(11)$$

As the geometric distribution determines the number of failures before the first success, here a failure is defined as the number of time slots during which the occupant is out of the state i and shown by k.

The vector for the sojourn time can be written as $Est = (Est_0, Est_1, \cdots, Est_n)$. The two key elements for determining the transition matrix P, which are the long run proportion and the expected sojourn time, can be specified using equations (7) and (11). [21.]

Using a two state homogeneous Markov chain together with an absorbing state, the morning arrivals can be expressed

$$P_{Arrival} = \begin{bmatrix} P_{00} & P_{01} \\ 0 & 1 \end{bmatrix} \quad . \tag{12}$$

The expected arrival time in the morning can be determined by

$$E(FA) = \frac{1}{1-P_{00}} \quad P_{00} = 1 - \frac{1}{E(FA)} \quad . \tag{13}$$

The departure in the evening is shown by a 2*2 matrix equation 14, and the expected departure time can be expressed by

$$P_{departure} = \begin{bmatrix} 1 & 0 \\ P_{10} & P_{11} 1 \end{bmatrix} \tag{14}$$

$$E(LD) = \frac{1}{1-P_{11}} \quad P_{11} = 1 - \frac{1}{E(LD)} \quad . \tag{15}$$

Leaving for lunch and coming back from the lunch break are considered separately. It is assumed that occupants have their lunch outside of the office, thus, similarly to the morning arrival and evening departure, the equations (16), (17), (18), (19) are used to determine the lunch break along with the expected time of leaving for and returning from the lunch break. [21.]

$$P_{lunch\_out} = \begin{bmatrix} 1 & 0 \\ P_{10} & P_{11} \end{bmatrix} \tag{16}$$

$$E(LL) = \frac{1}{1-P_{11}} \quad P_{11} = 1 - \frac{1}{E(LL)} \tag{17}$$

$$P_{lunch\_back} = \begin{bmatrix} P_{00} & P_{01} \\ 0 & 1 \end{bmatrix} \tag{18}$$

$$E(LB) = \frac{1}{1-P_{00}} \quad P_{00} = 1 - \frac{1}{E(LB)} . \tag{19}$$

Using a Markov chain, the movement patterns of occupants in a typical office building are modeled. During different periods of a day (morning arrival, lunch break, and evening departure) the probabilities of events are distinguished by statistical indices of events. To validate this model, an illustrative case study is used to check the model capacity.

Theoretically, the introduced validation system is not accurate enough because the outcome can be totally different than the real life situation. [21.]

In a case study, the simulation is applied to an office building to demonstrate the accuracy of the proposed model. The building has 4 office rooms, 1 corridor and 1 restroom, together with the “outside” makes a total of 7 spaces, indexed from 0 to 7, as it is illustrated in figure 8.

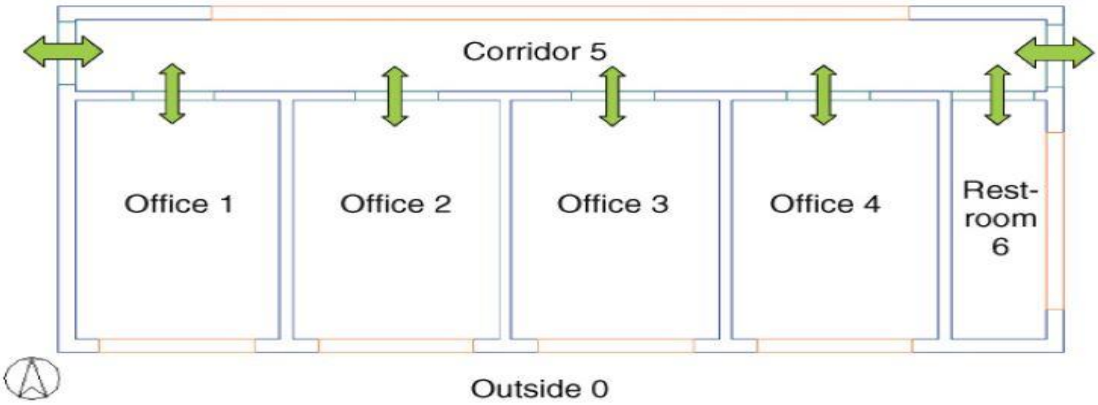

Figure 8. Plan of the office building used in simulation [21].

Rooms 1 and 2 belong to ordinary employees, room 3 is occupied by a secretary and the manager is in room 4. There are 6, 6, 2 and 1 occupants in offices 1 to 4, respectively.The working schedule is 08:00 AM to 05:00 PM and the lunch break is from noon to 01:00 PM. The detailed description of events and their time are illustrated in table 2.

Table 2. Events and their time [21].

Events	Valid Time
Arrival	07:00- 08:30
Leave for lunch	12:00- 12:30
Return from lunch	12:30- 13:30
Departure	17:00- 21:00
Walk around	08:00- 12:00 & 13:00- 17:00
Meeting	10:00- 11:30

Wang has provided the probability of each event for office number 1. [21.] For walking around, the transition probability matrix is denoted by matrix and the initial state P^0 is the night time. As shown in equations (7) and (8), where a transition matrix is created for the next time step, the current transition matrix shall be multiplied to the vector matrix. Therefore, with one transition matrix and the stationary distribution matrix, it is possible to generate a transition matrix for 1, 2 …n step. This can be done step by step by multiplying the results of previous step to the stationary distribution matrix, or it is possible to directly calculate the transition probability matrix in the certain time step using the matrix power. [21.]

$$P = \begin{bmatrix} 0.5000 & 0.3724 & 0.0247 & 0.0247 & 0.0250 & 0.0260 & 0.0272 \\ 0.0042 & 0.9583 & 0.0023 & 0.0023 & 0.0023 & 0.0042 & 0.0263 \\ 0.0244 & 0.2120 & 0.6667 & 0.0234 & 0.0237 & 0.0243 & 0.0255 \\ 0.0242 & 0.2131 & 0.0233 & 0.6667 & 0.0235 & 0.0241 & 0.0253 \\ 0.0247 & 0.2110 & 0.0237 & 0.0236 & 0.6667 & 0.0246 & 0.0258 \\ 0.0257 & 0.3747 & 0.0242 & 0.0243 & 0.0245 & 0.5000 & 0.0267 \\ 0.0055 & 0.4734 & 0.0052 & 0.0052 & 0.0053 & 0.0055 & 0.5000 \end{bmatrix} \qquad P^0 = [1\ 0\ 0\ 0\ 0\ 0\ 0]$$

Using the dot product of matrices and equations (7) and (8), the stationary distribution of the Markov chain can be determined. To examine the calculation an online calculator is used [22].

$$\pi = [0{,}01 \quad 0{,}9 \quad 0{,}01 \quad 0{,}01 \quad 00{,}01 \quad 0{,}01 \quad 0{,}05]$$

The expected staying time in a certain state can be specified using equation (10). For instance in state P_{ii}, the $E(ST_i)$ is,

$$E(ST_{00}) = \frac{1}{1 - 0{,}5000} = 2\ (each\ time\ slot\ is\ 5\ min)\ the\ total\ length\ at\ a\ certain\ state\ is\ 10\ min$$

$$E(ST_{11}) = \frac{1}{1 - 0{,}9583} = 24\ (each\ time\ slot\ is\ 5\ min)\ the\ total\ length\ at\ a\ certain\ state\ is\ 120\ m$$

$$E(ST_{22}) = \frac{1}{1 - 0{,}6667} = 3\ (each\ time\ slot\ is\ 5\ min)\ the\ total\ length\ at\ a\ certain\ state\ is\ 15\ min$$

$$E(ST_{33}) = \frac{1}{1 - 0{,}6667} = 3\ (each\ time\ slot\ is\ 5\ min)\ the\ total\ length\ at\ a\ certain\ state\ is\ 15\ min$$

$$E(ST_{44}) = \frac{1}{1 - 0{,}6667} = 3\ (each\ time\ slot\ is\ 5\ min)\ the\ total\ length\ at\ a\ certain\ state\ is\ 15\ min$$

$$E(ST_{55}) = \frac{1}{1 - 0{,}5000} = 2\ (each\ time\ slot\ is\ 5\ min)\ the\ total\ length\ at\ a\ certain\ state\ is\ 10\ min$$

$$E(ST_{66}) = \frac{1}{1 - 0{,}5000} = 2\ (each\ time\ slot\ is\ 5\ min)\ the\ total\ length\ at\ a\ certain\ state\ is\ 10\ min$$

Thus, the expected time that an occupant stays in office 1 is

$$Est = [10min, 120min, 15min, 15min, 15min, 10min, 10min]$$

The valid period of morning arrival is assumed to be between 7:00 AM to 8:30 AM. The expected value for this event is 7:45 AM, resulted from the 9 time slots (each time slot is 5 min). The 9 time slots totally gives 45 minutes and the from earliest time of arrival which is 07:00 AM the expected time of arrival is 07:45 AM, thus;

$$P_{Arrival} = \begin{bmatrix} 0.8889 & 0.1111 \\ 0 & 1 \end{bmatrix}$$

The expected time of first arrival would be,

$$E(FA) = \frac{1}{1 - 0.8889} \;\; = 9\, unit\ of\ time\ slot, each\ slot\ is\ 5\ min$$

Similarly, the scheduled time of leaving for lunch and coming back to the office are 12:00 AM-12:30 AM and 12:30 AM-13:30 AM, respectively. The expected values would be

$$P_{lunch\_out} = \begin{bmatrix} 1 & 0 \\ 0.5000 & 0.5000 \end{bmatrix}, \qquad E(LL) = \frac{1}{1 - 0.5} \; = 2\, unit\ of\ time\ step\ .$$

The expected time of leaving for lunch is 12:10 PM. It means the earliest scheduled time for lunch plus the two units of time (2*5= $10\, min$), driven from equations (16) and (17) above. Similarly for coming back from lunch break which is 12:30 PM to 01:30 PM, the expected time would be

$$P_{lunch\_back} = \begin{bmatrix} 0.7368 & 0.2632 \\ 0 & 1 \end{bmatrix}, \qquad E(LB) = \frac{1}{1 - 0.7368} = 3.8\, unit\ of\ time\ step\ .$$

Thus, the expected time of coming back to the office is 12:50 PM driven from equation (18) and (19), or the earliest scheduled time for coming back from lunch plus the 3.8 units of time (3.8*5= $20\, min$).

Finally, the time for leaving the office is assumed to be between 05:0 PM and 09:00 PM. The expected time can be driven from equations (14) and (15).

$$P_{departure} = \begin{bmatrix} 1 & 0 \\ 0.0833 & 0.9167 \end{bmatrix}$$

The expected time of departure is 06:00 PM or 12*5 min from the earliest departure at 05:00 PM. this can be written

$$E(LD) = \frac{1}{1 - 0.9167} = 12\, unit\ of\ time\ step$$

The number of occupants during a working day is illustrated in figure 9. The occupancy gradually increases after the morning arrival and reaches to maximum between 9:00 AM to 10:00 AM. [21.]

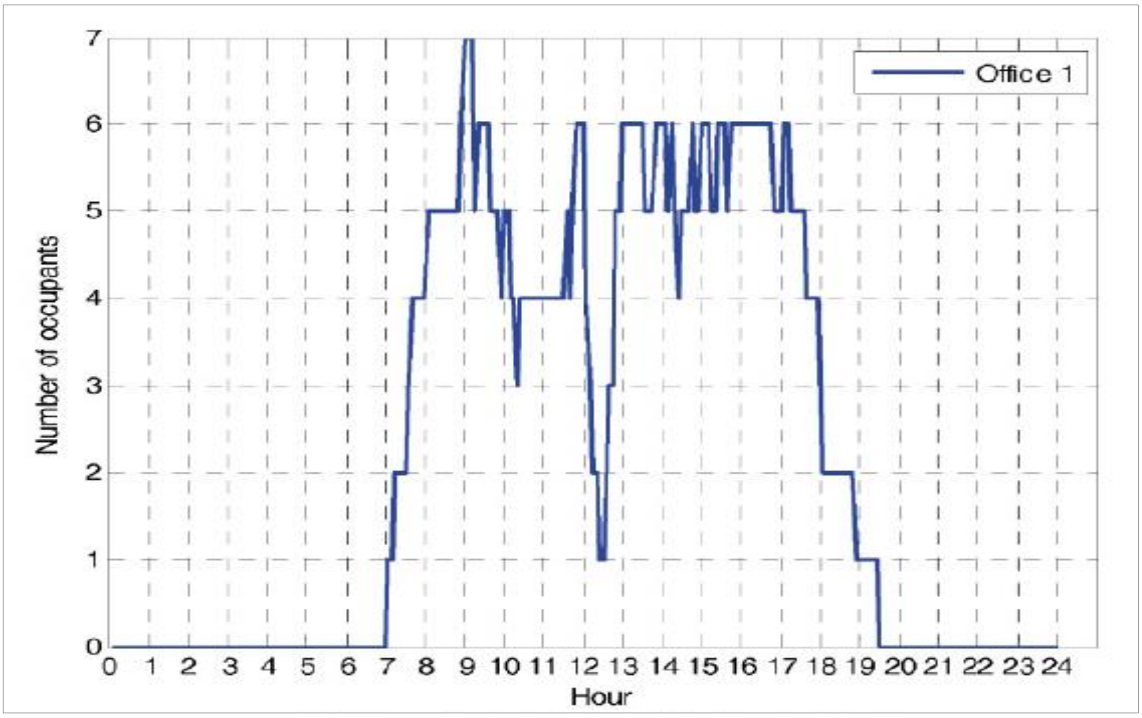

Figure 9. Occupancy rate at office room [21].

Two or three occupants leave the room between 10:00 AM and 11:30 AM for meetings. After lunch time the office is fully occupied until night departure, which starts at 05:00 PM. After 07:30 PM the office is unoccupied. [21.]

2.3.2 Agent-Based Stochastic Model

Liao et al. proposed a stochastic agent based model which is extendable for an arbitrary number of zones and occupants. The model, which is a extend version of the model provided by Page et al. incorporates the presence of multiple occupants in a zone. [23;24] Therefore, the model generates a time series for the location of each occupant. Then the occupancy rate in different zone is produced using this information.

The occupancy of a building with n zones can be calculated with,

$$x(k) := \sum_{j=1}^{n} \mathrm{x}_j\,(k), \qquad (20)$$

where k is time discrete, individuals (agents) are indexed as *I,* and building zones, together with the outside, are indexed as j. Since the proposed model consists four modules, it is named Multiple Modules (MuMo) which together determine the state of an agent at every time step. [24.] For instance, in a sample building with 4 zones $j = \{1,2,3,4\}$, and each zone occupied by 1,0,2,1 agents respectively. For the a weekday the occupancy of the building for the $k = 41$, (10:00 AM- 10:15 AM) using equation (20) can be calculated as

$$\left.\begin{array}{l} for\ j = 1 \rightarrow\ x\_1\ (1) = 1 \\ for\ j = 2 \rightarrow\ x\_2\ (1) = 0 \\ for\ j = 3 \rightarrow\ x\_3\ (1) = 2 \\ for\ j = 4 \rightarrow\ x\_4\ (1) = 1 \end{array}\right\} \rightarrow\ x(41) := \sum_{j=1}^{4} \mathrm{x}_j\,(1) = 1 + 0 + 2 + 1 = 4$$

The equation (20) shows that the building is occupied with four agents during time interval 10:00 AM to 10:15 AM. In the simulation, the number of occupants in each zone at each time step will be specified using the four modules below,

- **Preliminary state generator module** is used to determine the initial state of occupants by giving the presence probability of each agent to this module.
- **Acceleration and damping modules** are used to simulate the agent behavior using transition probability parameters. The inputs of this module are the primary state of each agent driven from first module together with the transition probability parameters which are p_a and p_d . The node of the agent i at time $(k-1)$ is indexed by $z_i(k-1)$. If the state of an agent at current time (k) is the same as the state at the previous time $(k-1)$, i.e. $z_i^{(0)}(k) = z_i(k-1)$
- , the preliminary state determiner runs again to compute the state of the agent i using p_a . If the equation $z_i^{(0)}(k) \neq z_i(k-1)$ is true then the damping module is used $z_i(k-1)$ either outside node or a primary zone. In this case $z_i^{(1)}(k) \leftarrow z_i(k-1)$ is true with the probability $1 - p_d$.
- **Scheduled activity module** is used in order to determine the constraints on the agent's locations according to the preset schedule, and denoted by $z_i^{(2)}(k)$.

- **Access module** is used to ensure that the agents are not presence inside zones they do not have access to. Thus if $z_i^{(2)}(k) = j$, where j is the node which is unavailable for an agent i, then $z_i^{(3)}(k) \leftarrow z_i(k-1)$ would be true, and the output of this module is the state of the agent i at time $(k), z_i^{(3)}(k) \leftarrow z_i(k)$. [23.]

In addition to the modules mentioned above, three scenarios have been defined to cover all possible cases: (1) single-occupant single-zone (SOSZ), (2) multi-occupants single-zone (MOSZ) and (3) multi-occupants multi-zone (MOMZ). In order to use the model, the nominal presence probability profile for each occupant must be determined. This determines the initial state of each agent, the occupant presence schedule, and the access of each occupant. In addition, the acceleration and damping parameters must also be considered. [23.] To compute the nominal presence probability profiles, the following procedure can be used:
Primary and secondary zones are denoted as

$$PZ_i(Primary\ zone)\colon\ p_{i\,PZ_i}^{(d)} = RP_i^{(d)} + \alpha, \tag{21}$$

where $RP_i^{(d)}$ is the ratio of the presence of the agent i in PZ_i at the d-th day of week to the time between agent's arrival and departure time for that day. The probability of presence in the primary zone of all the agents is determined with α. [23.]

For example, if an agent x occupies his primary zone for 5:30 hours, and his presence interval in an office (from first arrival to last departure) is 8 hours. Thus, the $RP_i^{(d)}$ is 0.687. If the other agents spend half of their time in their primary zone, then the presence probability of all agents (α) in their primary zones is 0.5, and the nominal presence probability of the primary zone is 0.687+ 0.5= 1.187 which is more than 1. [23.]

$$SZ_i(Secondary\ zone)\colon\ p_{i\,SZ_i}^{(d)} = \begin{cases} FOSZ_i \dfrac{AVT_i}{DT_i^{(d)} - AT_i^{(d)}} & d \in \{1, \ldots, 5\} \\ 0 & otherwise \end{cases}, \tag{22}$$

where $FOSZ_i$ is the frequency of occupying secondary zone, AVT_i is the average duration of visiting a secondary zone, $DT_i^{(d)}$and $AT_i^{(d)}$ are the arrival and departure time respectively. It is assumed that during weekends ($d \in \{6,7\}$), the secondary zones are occupied only by agent i. For the restroom (RR) and hallway (HW), it is assumed that an agent visits the restroom 3 times a day and each visit lasts 5 minutes. For the hallways the same assumption has been done. [23.] Thus,

$$p_{i\,HW_i}^{(d)} = p_{i\,RR_i}^{(d)} = \frac{3 \times 5}{(DT_i^{(d)} - AT_i^{(d)}) \times T}. \qquad (23)$$

The initialization of nominal presence probability profiles can be done as;

$$p_{i,j}^{(d)}(k) = \begin{cases} p_{i,j}^{(d)} & k \in \left\{AT_i^{(d)}, DT_i^{(d)}\right\}, j \in \{PZ_i, SZ_i, RR, HW, n+1\} \\ 0 & otherwise \end{cases}. \qquad (24)$$

Therefore, the probability of an agent occupying any zone other than the primary zone, secondary zone, restroom, hallway, and outside (n+1) is zero. [23.]

The model verification is carried out by comparing the parameters such as the mean occupancy of a zone, the first arrival time and the last departure driven from measured data and the results of the model simulation. The results of the model verification for the single occupant- single zone (SOSZ) against measurement data are summarized in table 3. [23.]

Table 3. Kullback–Leibler (K-L) divergence between p_x^{meas} and p_x^{MuMo} in the SOSZ scenario [23].

Variable (X)	$\boldsymbol{d(p_x^{meas} \| p_x^{MuMo})}$
First arrival time	0.4968
Last departure time	0.6388
cumulative occupied duration	0.3215
Continuously occupied duration	0.0229
Number of occupied/unoccupied transitions	0.2421

Table 3 indicates that the Multiple Modules (MuMo) model can accurately simulate the continuously occupied duration as it has a 0.0229 K-L divergence with the actual measurements. The data also illustrates the poor prediction of the first and last departure time in the MuMo model which are 0.4968 and 0.6388, respectively. The comparison of the MuMo model and Page's model against measured data is illustrated in figure 10. [23.]

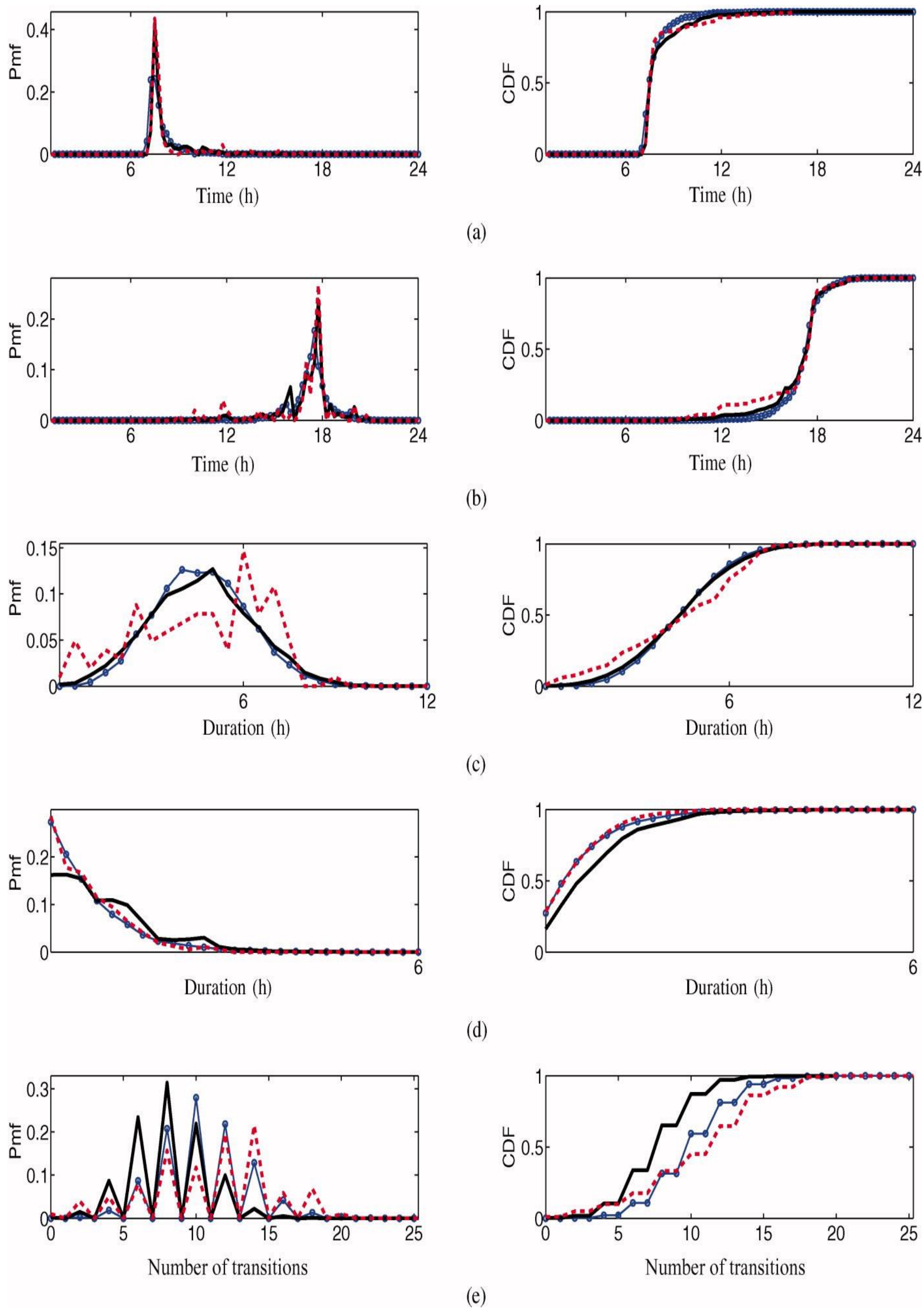

Figure 10. Result comparison of MuMo Model (blue), Page model (black) and actual measured data (red) for SOSZ scenario. (a) First arrival, (b) Last departure, (c) Cumulative occupied duration, (d) Continuously occupied duration, (e) Number of occupancy frequency. The bin-time in (c) is 30 min, and for the rest it is 15 min [23].

2.3.3 Occupancy Model for Regular Occupancy of Office Building

Chen et al. developed a stochastic model for regular occupancy in commercial buildings. The model introduces two scenarios, multi occupant- single-zone (MOSZ) and multi occupant multi zones (MOMZ). The Markov chain theory is used as the basis of this model. The MOSZ considers the increment of occupancy in a zone as a state of the Markov chain. The maximum number of people moving in or out of a zone is assumed to be one within each short interval. According to this assumption, a simple 3×3 matrix can be used to explain the transition probability regardless of the maximum number of occupants in a single-zone (figure 11). Thereby, it is possible to calculate the transition probabilities for different numbers of occupants, e.g. 10, 30, 50, and 100 people using a 3×3 matrix. But the problem with this model comes out when calculating the transition probabilities of a larger zone (open office), occupied with e.g. 50 or 100 employees. The office has more than one transition in each time interval, which means more than one occupants might change their states at the same time. [25.]

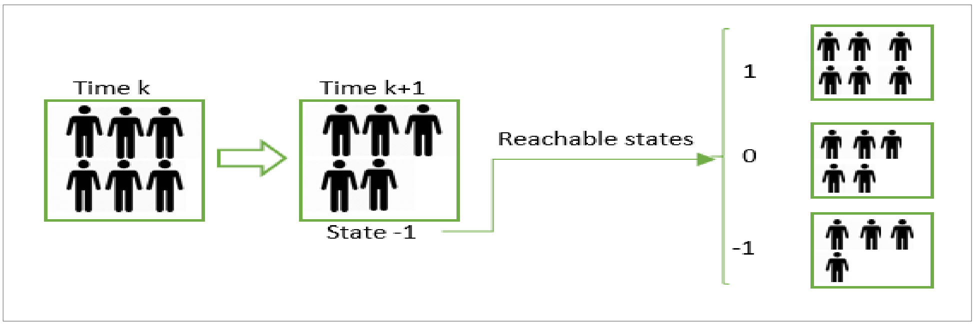

Figure 11. Occupancy states in the MOSZ scenario [25].

In the MOMZ, the state in the Marko chain is illustrated with a vector in which each component is the increment of occupancy in each zone. For instance, the total number of states in four zones, with the assumption of one movement in each time step, would be $3^4 = 81$. The relationship between the matrix dimension and the number of zones is defined as d=3^m, where m is the number of zones. Thus, a corresponding transition probability matrix has the dimension of $3^4 \times 3^4$. In the MOMZ scenario just like MOSZ the dimensions of the transition probability matrix is independent of the number of occupants, thus the matrix dimension is determined by the number of zones. To determine the parameters of the transition probability matrix, a maximum likelihood estimation (MLE) is used. Using The MLE the maximum likelihood of an action can be estimated if have the probability of the action. The binomial trail is used to present the occupancy in a zone, thus 0 and 1 are the two states for an occupied and unoccupied zone respectively. [25.]

The number of present occupants within a specific zone at a particular time step is defined as

$$F(x_i|P) = p^{x_i}(1-p)^{1-x_i}$$
$$x_i = \begin{cases} 1, & present \\ 0, & absent \end{cases}, \tag{25}$$

where F is the probability distribution function, and p^{x_i} is the probability an occupant is present in the zone. In a zone with only one occupant, the equation can be simplified as

$F(1|P) = p^1(1-p)^{1-1}$ → *P* which is the probability of presence
$F(0|P) = p^0(1-p)^{1-0}$ → *1-P* which is the probability of absence

For a multi- occupant single-zone the situation is more complex

$$F\left(x_1, x_{2,} \dots, x_n \,|P\right) = p^{x_1}(1-p)^{1-x_1}.\, p^{x_2}(1-p)^{1-x_2}. \dots\, p^{x_n}(1-p)^{1-x_n} \tag{26}$$

The simplified equation can be presented as a likelihood product equation,

$$L = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}\ . \tag{27}$$

For $x_k = 1$ the probability is denoted by p and $1-p$ is used to determine the observation when $x_k = 0$, thus the likelihood function using the observation x_k can be written as,

$$F(p|X_k) = \begin{cases} p, & X_k = 1 \\ 1-p & X_k = 0 \end{cases}. \tag{28}$$

To simplify the above equation, the log-likelihood function is used. The first derivative of ln $F(p|X_k)$, which is the likelihood equation, is

$$\frac{\partial\ \ln F(p|X_k)}{\partial p} = 0\,. \tag{29}$$

Thus, the solution is $p = p_{MLE}$ .To ensure that the given value maximizes the likelihood, the second derivative of the likelihood function must be less than 0,

$$\frac{\partial^2\ \ln F(p|X_k)}{\partial p^2} < 0\,. \tag{30}$$

A binomial trial with the probability of $\Pr(X_{k+1} = s_{k+1}|X_k = s_k)$ is used to explain the transition from state S_k to the state S_{k+1}. For example, if M is the days where $X_k = s_k$ and among these days, W is the days where $X_{k+1} = s_{k+1}$, then the log likelihood function would be

$$\ln L(p|\emptyset) = \ln \frac{M!}{W!\,(M-W)!} p^{w}(1-p)^{M-W} \tag{31}$$
$$= \ln \frac{M!}{W!(M-W)!} + W \ln p + (M-W)\ln(1-p),$$

where the ∅ is the total number of days collected and the likelihood equation would be,

$$\frac{\partial\ \ln F(p|X_k)}{\partial p} = \frac{W}{p} - \frac{M-W}{1-p} = 0. \tag{32}$$

In order to recheck the MLE estimator, the second derivative of log-function must be negative, and can be calculated as

$$\frac{\partial^2\ \ln F(p|X_k)}{\partial p^2} = \frac{W}{p^2} - \frac{M-W}{1-p^2} < 0. \tag{33}$$

The same process is repeated for each transition probability to determine each parameter of the transition probability matrix. The transition probability matrix in both scenarios is independent of the maximum number of occupants. Since this model is based on the transition of one occupant at each time interval for a zone, the transition of more than one occupant at the same time increases the matrix dimensions which means more complexity. [25.]

The evaluation of the performance of this model is based on five variables related to the occupancy properties under two evaluation criteria. The occupancy related variables considered are time of first arrival, time of last departure, mean occupancy, cumulative occupied duration and the number of transition from occupied to unoccupied and vice versa. [25.]

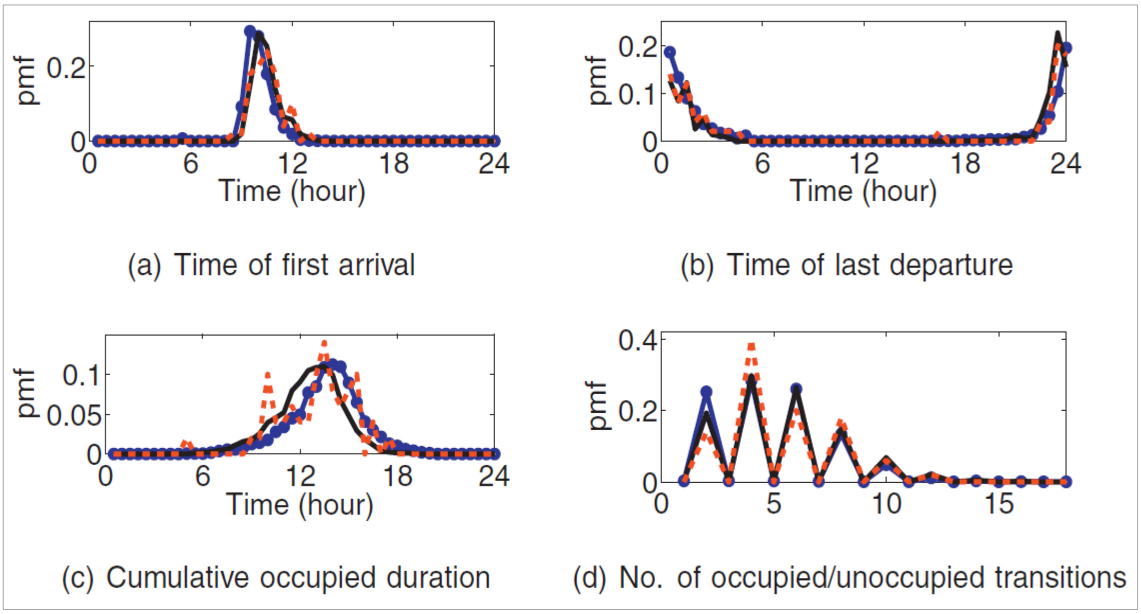

Figure. 12. Comparison of the probability mass functions (pmfs) of four random variables in weekdays from measured data (red), Liao's model [22] agent based model (blue) and the presented model (black). The bin size is 30 m [25].

The result of a comparison of the agent-based model, measurement data and the proposed model shows that the model can predict the time of the first arrival and last departure more accurately than Liao's model (figure 12). Neither Chen's model nor Liao's model could predict the cumulative occupied duration. For the number of transitions between states occupied to unoccupied, both models give good results.

2.3.4 Two-State Stochastic Occupancy Model for Residential Buildings

Richardson et al. constructed a stochastic occupancy model for active occupants based on the Time-Use Survey (TUS) conducted in the United Kingdom in the year 2000. [19;26.] The current state, along with the transition probabilities of changing from the current state to the next one, are the two important properties of this stochastic model. The current state starts at midnight when the occupants (~ 85%) are inactive according to the TUS data. The transition probability determines the probability of change from one state to another. The transition probability matrices are generated using the TUS data. Each day (24 hours) is divided in 10- minute- time slots, which in total the occupancy rate of 144 time slots must be computed. [19.] The process of calculating the transition probabilities for one person households during weekdays is illustrated in table 4.

Table 4. Calculation of the transition probability matrix on weekdays [19].

No. of Active occupants		No. of occurrences in the TUS data		Transition probability
At 00:00	At 00:10			
0	0	1428	1428+8= 1436	1428/1436= 0.994
0	1	8		8/1436= 0.006
1	0	55	55+211= 266	55/266= 0.207
1	1	211		211/266= 0.793

As the calculations in table 4 show, during the first time intervals from 12:00 AM to 12:10 AM, the number of inactive occupants who remain inactive is 1428 in the TUS data, and only 8 occupants change their states from inactive to active in 10 minutes. Thus, the transition probability can be calculated as number of occurrences total number of population (1428/ (1428+8) =0.994). The transition probability for a change in state from inactive to active is 1-0.996= 0.006. [19.]

Further, the transition probability from active state at 12:00 AM to inactive at 12:10 AM according to the TUS data is 0.207, as the number of people who change their state from active to inactive is 55, and the number of people who remain active is 211, thus the total number of population is 266. To represent the occupancy rate for a whole day, the same process must be repeated for all144 time slots. [19.]

The model is provided in the form a computer program and the results of two runs of the simulation for a one- person household in weekdays are shown in figure 13. In each run, a random initial number is picked, with the same transition probability matrix. [27.]

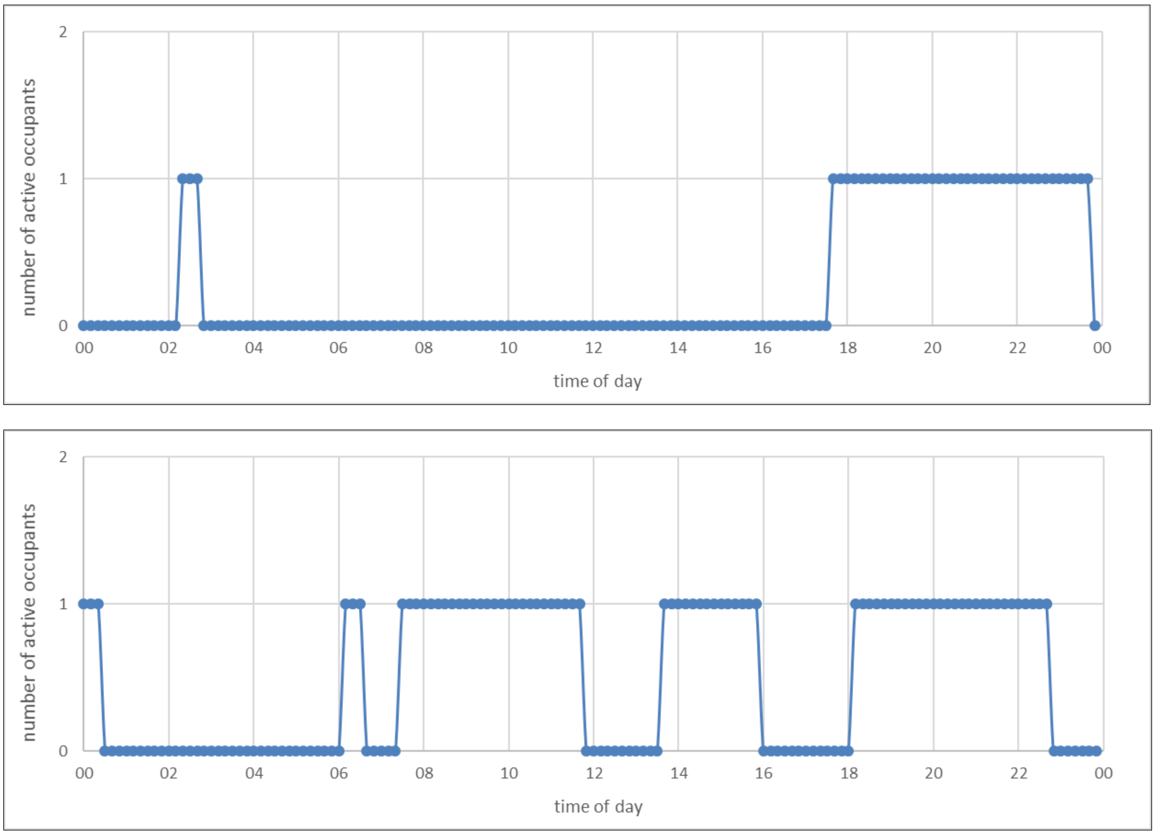

Figure 13. Two runs of model for one resident household in weekdays [20].

In order to validate this model, the model was run repeatedly and the results were compared against the TUS.

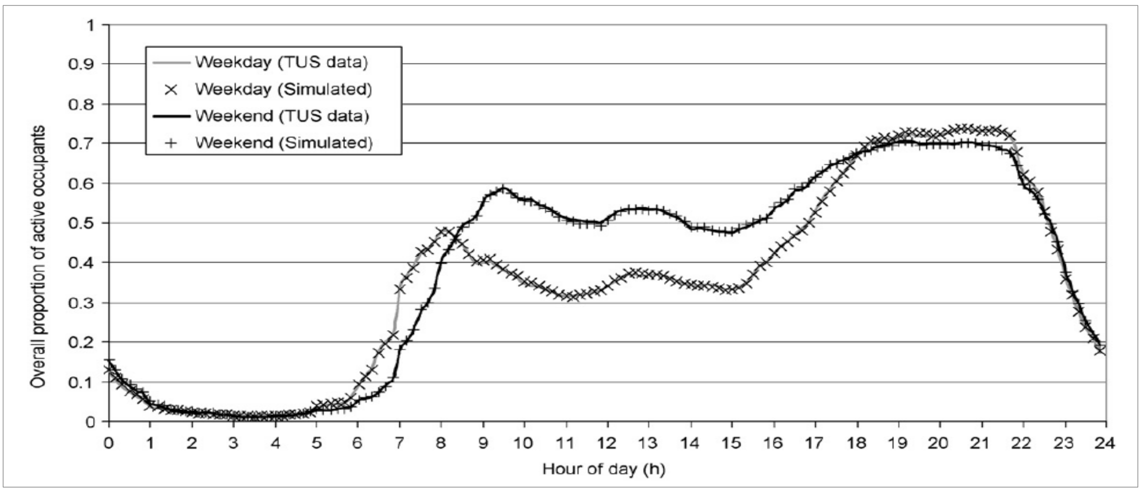

Figure 14. Comparison of simulated model and TUS data for sample 2 resident household [19].

A good correlation between the results from the model and the original data form TUS database was seen, as shown in figure 14. [19.

2.3.5 Four-State Stochastic Occupancy Model for Residential Buildings

In another model, Mckenna et al. presented a four states domestic occupancy model in which the active/inactive and absent/present occupancy states are treated separately [20]. The model provided by Richardson et al, consideres only two states of occupancy (active and inactive), the sates of being absent and inactive are merged together [19]. McKenna's model considers the whole cycle of occupancy in dwelling. Four states of occupancy have been defined as not at home and not active (00), not at home and active (01), at home and not active (10) and at home and active (11). The first digit represents the states of being absent (0) or present (1), where the second digit describes the state of being active (1) or inactive (0). A first order inhomogeneous Markov chain method is used in this model to precisely simulate the occupancy rate of a residential building. [20.]

Compared to Richardson's model, the size of transition probability matrices are larger, as this model describes two more states. Thus, the size of matrices can be defined as

$$[(n+1)\times(n+1)]\times[(n+1)\times(n+1)] = (n+1)^4, \qquad (34)$$

where n represents the number of occupants. Therefore, a 4*4 matrix is used to describe the transition probability for a single-resident household. [20.]

$$P = \begin{bmatrix} T_{0000} & T_{0100} & T_{1000} & T_{1100} \\ T_{0001} & T_{0101} & T_{1001} & T_{1101} \\ T_{0010} & T_{0110} & T_{1010} & T_{1110} \\ T_{0011} & T_{0111} & T_{1011} & T_{1111} \end{bmatrix} \rightarrow P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.0227 & 0.9091 & 0 & 0.0682 \\ 0 & 0 & 0.9985 & 0.0015 \\ 0 & 0 & 0.1949 & 0.8051 \end{bmatrix}. \qquad (35)$$

The TUS data is categorized in an excel file. Based on the categorized data provided in the excel file, the transition probability matrix for the first time step or first 10 minutes can be written as in matrix (35) for one resident house on weekdays. [28.] For instance, the probability of remaining in the states, outside and inactive in the 10 min time interval is T_{0000} is 1. The occupants who are outside and active will either change their states to the outside and inactive with the probability of 0.0227, or remain in their states with the probability of 0.9091. The probability of returning home and becoming active for an occupant in the state is 0.0682. To generate the transition probability matrix for a whole day, the same process is repeated for the other 143 time slots.

The starting states are selected on the base of the probability distribution of the states in observed data. For the first time step it is described as

$$SPD = \begin{bmatrix} 0.0311 & 0.0517 \\ 0.8802 & 0.1145 \end{bmatrix}.$$

Using the starting time together with the transition probability matrix for each time slots, the occupancy state for any time interval will be generated. As an example, a simplified version of the model is provided in an Excel file To see the outcomes of the Makenna's model, the Excel file is used and the results are illustrated in figure 15. [28.]

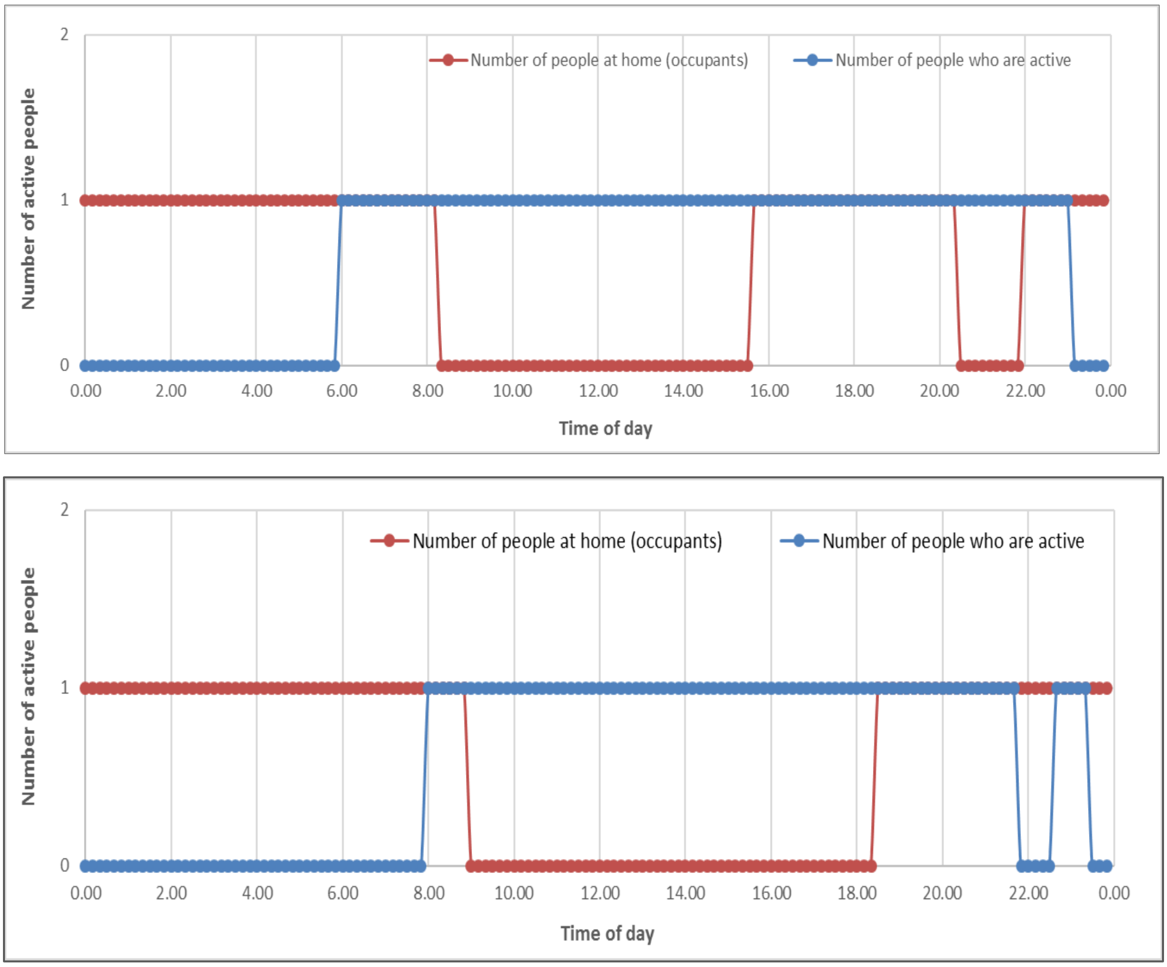

Figure. 15. Three example runs of model for 1 resident household on weekdays [28].

When the results of the model are compared to the results of the TUS data, it can be seen that they are in close agreement, this is illustrated in figure 16.

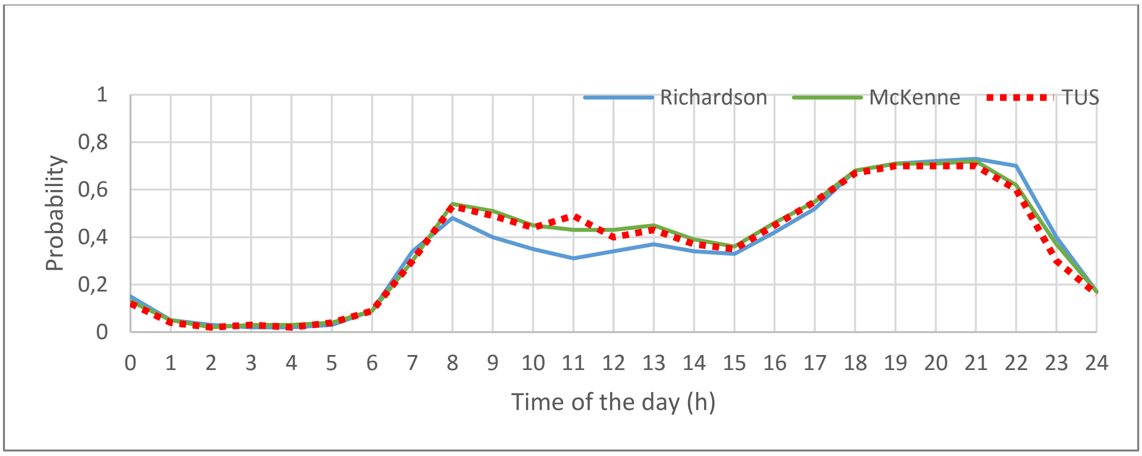

Figure 16. Comparison of the active state probability of Richardson and McKenna models against TUS data.

As illustrated in figure 16, the probability of the state at home/active increases between 06:00 PM to 09:00 PM. When the occupants are at home, they are mostly inactive from 12:00 AM to 06:00 AM.During the daytime 07:00 AM- 06:00 PM the majority of the occupants are outside at work. The probability of being in the state not home/inactive decreases during the daytime (07:00 AM- 09:00 PM).

3 Conclusion and Discussion

Occupancy has a great influence on the energy consumption of a building. This factor is not only the key to occupant's behaviour modelling but also to building simulation tools. Therefore, to predict a reliable energy performance in a building using energy simulation tools, realistic characterization of a building's occupancy is required. Different models have been introduced to simulate the occupancy rate. This study has reviewed the deterministic, probabilistic and stochastic occupancy models.

In the deterministic approach, the presence or absence of an occupant is directly linked to a specific time of a day or building types. For instance, in typical residential buildings the occupancy rate during daytime (09:00 AM - 05:00 PM) is usually modeled as unoccupied while for an office building the situation is normally considered to be occupied during this time interval. The deterministic approach is simple to use but it has some limitations, such as a lack of accuracy due to not considering the stochastic nature of occupancy in different types of buildings, or un-scheduled activities which lead to a sudden presence or absence in a building. The probabilistic approach is introduced to simulate the occupancy rate more precisely. The probabilistic approach is based on probability and statistics. Using statistical data, the likelihood of occupant being present or absent is predicted. However, the foundation of the probabilistic approach is observed data collected from different sources. Still in many cases, the probabilistic models cannot capture all possible conditions of occupancy. In order to solve the problem, stochastic approaches have been applied in occupancy modelling. Stochastic approaches allow more probable states than the other two approaches. The Markov chain technique is usually used to trace the stochastic nature of the occupancy rate. The current state along with the transition probability of changing states are the two important features of the Markov chain technique. The limitations of the stochastic approaches vary from model to model, depending on the variables and the method utilized.

In Wang's probabilistic model, the number of transitions between two states (occu-pied-vacant) is described using a Poisson process technique. Wang showed that the occupancy intervals are not exponentially distributed, while the vacancy intervals are distributed exponentially. The comparison of the results from observed data and the model simulation show an ignorable discrepancy in the number of transitions. [13.] The Wang model cannot capture the long absence of occupants, which decreases the accuracy of its results.

Aert et al. in a probabilistic model defined three states of occupancy for residential buildings. The TUS data was used for generating the model that is based on two parameters, “transition probability” and “duration probability”. To capture even more variations, the authors categorized the database in seven groups. Each group has a different occupancy pattern. This makes the model easier to use and increases its accuracy. [14.]

In a stochastic model, Wang et al. used a homogeneous Marko chain method to explain the movement of occupants inside and outside of the buildings. “Movement” and “events” are the two modules applied in this model to describe the occupancy rate. The movement determines different sorts of movement in each zone, and the module of events explain the transition probabilities between events. The movement of each individual is divided in five segments, which increases the accuracy of a simulation. The model is also applicable for an arbitrary number of occupants. Since this model uses time independent probabilities, it is easy to use. [21.] The disadvantage of the Wang model is that some events, such as short visits or unscheduled departures, are not taken into account, which can affect the simulation results.

In agent-based models, certain behaviors are dedicated to an agent, and then the behavior of the agent is studied in a specific situation. In case of occupancy modelling, agents are designed to mimic the human behavior in a specific zone or building. The results are a time series of each individual’s location and, consequently, this can generate time traces of occupancy in a whole building. A common problem with an agent-based model is its high level of complexity. The thesis compered the results of the MuMo model with those of the Page model. The comparison shows that both simulate the time of first arrival and last departure with good accuracy. The two models show poor results in a simulation of cumulative occupied duration compared to the measured data. The MuMo model simulates the continuously occupied duration with a higher accuracy than the Page Model. Generally, the MuMo model can predict the distribution of some variables fairly well, such as mean occupancy, continuous occupancy duration and the number of transitions from occupied to unoccupied. The model cannot predict the probabilities of some variables, such as last departure and cumulative occupancy duration precisely. In the MOMZ scenario this mismatch is even larger due to more inaccurate data when the number of agents increases. [21;23.]

Chen et al. in their stochastic model used almost the same methodology as the one used by Liao [23;25]. Chen defined two scenarios instead of three. A 3×3 matrix is used to explain the transition probabilities. Since the transition probability matrix in both scenarios is independent of the maximum number of occupants, the same 3×3 matrix can used for an extended number of occupants, which makes the computation much easier.[25.] The comparison of Chen’s model against Liao’s model revealed that the former model has less errors and can produce more accurate results.

The occupancy modelling for residential buildings is different from the one for office or commercial buildings. Office buildings are typically occupied during daytime (08:00 AM to 06:00 PM) on weekdays, which narrows down the variations and shortens the period of simulation. In residential buildings, the time circle is 24 hours as people might stay in their homes all day long. Richardson et al. explained the occupancy rate in residential buildings with a stochastic model based on the two states of active and inactive occupants. The transition probability from one state to another in short time intervals is driven from the British TUS data. A time trace of active or inactive occupants is produced for each 10 minutes and by repeating the process for 24 hours, the occupancy rate for a day is determined. [19.] Later, McKenna et al. provided a four-state occupancy model for residential buildings which includes more occupancy states. McKenna’s model, like Richardson’s model used TUS database to determine the parameters of transition prob-abilities. [20.] The results of the two models are in close agreement.

The results comparison of different occupancy models revealed that the occupancy models based on the deterministic approaches could capture less diversity factors than the models based on the probabilistic and stochastic approaches. The findings also show that both the probabilistic and stochastic models capture diversity factors such as first arrival, last departures, mean occupancy and continuous occupancy duration with good accuracy. All models still have problems with capturing uncertainties such as short visits or long absences. For future studies, the results from this thesis can be used to develop a model based on a combination of both stochastic and probabilistic approaches, which can capture maximum diversity factors with minimum requirements.

References

1 Berardi U. Building energy consumption in US EU and BRIC countries. International Conference on Sustainable Design Engineering and Construction 2015;118:128-136.

2 Trčka M, Hensen JL. Overview of HVAC system simulation. Automation in Construction 2010,19(2):93-99.

3 Vieira RK, Parker DS, Klongerbo JF, Sonne JK, Cummings JE. How Contractors Really Size Air Conditioning Systems. Florida Solar Energy Center; 1996.

4 The European Standard EN 15232 A key contribution to worldwide energy efficiency. ABB; 2010.

5 Yan D, Xia J, Tang W, Song F, Zhang X, Jiang Y. DeST An integrated building simulation toolkit Part I: Fundamentals. Building Simulation 2008,1(2):95-110.

6 Crawley DB, Lawrie LK, Pedersen CO, Liesen RJ, Fisher DE, Strand RK, Taylor RD, Winkelmann RC, Buhl WF, Huang YJ, Erdem AE. A New-Generation Building Energy Simulation Program. Proceedings of Building Simulation 2001,1:81-88.

7 The ESP-r System for Building Energy Simulation. User Guide Version 10 Series. University of Strathclyde; 2002.

8 Abushakra B, Sreshthaputra A, Haberl JS, Claridge DE. Compilation of Diversity Factors and Schedules for Energy and Cooling Load Calculations ASHRAE Research Project 1093-RP. Texas: Texas Engineering Experiment Station; 2001.

9 Energy Standard for Buildings Except Low-Rise Residential Buildings ISSN 1041-2336. ASHRAE; 2004.

10 Duarte C, Wymelenberg KVD, Rieger C. ASHRAE. 2013,119(2):1.

11 Aerts D, Minnen J, Glorieux I, Wouters I, Descamps F. Discrete occupancy profiles from time-use data for user behaviour modelling in homes. 13th Conference of International Building Performance Simulation Association; 2013.

12 Chen Z, Soh Yeng Chai. Comparing occupancy models and data mining approaches for regular occupancy prediction in commercial buildings. Building Performance Simulation 2016. 1-9.

13 Wang D. Federspiel CC. Rubinstein F. Modeling occupancy in single person offices. Energy and Buildings 2004,37:121-126.

14 Aerts D, Minnen J, Glorieux I, Descamps F. A probabilistic activity model to include realistic occupant behaviour in building simulations. Ottawa: eSIM; 2014.

15 Glorieux I, Minnen J, Tienoven TP, et al. Belgisch tijdsbudgetonderzoek [Online]. Brussels: Statistics Belgium; 2015.
URL: http://www.time-use.be/. Accessed 4 April 2017.

16 Gunathilak G, Prasannakumar A, Nazarian N, Naeimi H. A generalized event driven framework for building occupancy (WIP). Symposium on Simulation for Architecture & Urban Design. San Diego: California; 2013.

17 Pinsky MA, Karlin S, Taylor HM. An introduction to stochastic modelling. Oxford: Academic Press; 2011.

18 Meyn S, Tweedie RL. Markov chain and stochastic stability. Cambridge: Cambridge University Press; 2009.

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

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

21 Wang C, Yan D, Jiang Y. A novel approach for building occupancy simulation. Building Simulation 2011,4(2):149-167.

22 Kemeny J, snell L. Calculator for Finite Markov Chain [Online]. New York: Springer-Verlag; March 2008.
URL:https://kilin.clas.kitasato-u.ac.jp/software/markov/FiniteMarkovChain.html. Accessed 04 April 2017.

23 Liao C, Lin Y, Barooah P. Agent-based and graphical modelling of building occupancy. Journal of Building Performance *Simulation* 2012,5(1):5-25.

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

25 Chen Z, Xu J, Soh YC. Modelling regular occupancy in commercial buildings using stochastic models. Energy and Buildings 2015,103:216-223.

26 Ipsos-RSL and Office for National Statistics. United Kingdom Time Use Survey 2000. [data collection]. 3rd Edition. UK Data Service.
URL: http://doi.org/10.5255/UKDA-SN-4504-1. Accessed 4 April 2017.

27 Richardson I, Thomson M, Infield D. A high-resolution domestic building occupancy model for energy demand simulations. Energy and Buildings 2008. URL: https://www.dropbox.com/s/tva2oam6m0ngvwh/Active_Occupancy_Simulation_D ata_Sheet_1_0c.xlsm?dl=0. Accessed 04 April 2017.

28 McKenna E, Krawczynski M, Thomson M. Four-State Domestic Occupancy Model [computer programme]. Loughborough: Loughborough University; February 2005.
URL: https://www.dropbox.com/s/8dmaf8wv8h95dpe/Copy%20of%20Fourstate _ Domestic_Occupancy_Model_v1_1.xlsm?dl=0. Accessed 04 April 2017.