Energy and Buildings 103 (2015) 216–223

Contents lists available at ScienceDirect

Energy and Buildings

journal homepage: www.elsevier.com/locate/enbuild

Modeling regular occupancy in commercial buildings using stochastic models

Zhenghua Chen, Jinming Xu, Yeng Chai Soh*

School of Electrical and Electronics Engineering, Nanyang Technological University, 50 Nanyang Ave., Singapore

ARTICLE INFO

Article history:
Received 31 January 2015
Received in revised form 2 June 2015
Accepted 3 June 2015
Available online 9 June 2015

Keywords:
Building occupancy modeling
MOSZ
MOMZ
Inhomogeneous Markov chain

ABSTRACT

Buildings account for some 40% of the world's total energy usage. One factor that could help to improve energy efficiency in buildings is more accurate modeling of occupancy. In this paper, we propose two novel stochastic inhomogeneous Markov chains to model building occupancy under two scenarios of multi-occupant single-zone (MOSZ) and multi-occupant multi-zone (MOMZ) respectively. In the MOSZ scenario, instead of using occupancy (i.e. the number of occupants in a zone) as the state, we define the state of the inhomogeneous Markov chain as the increment of occupancy in the zone. In the MOMZ scenario, by taking into account interactions among zones, we propose another inhomogeneous Markov chain whose state is a vector in which each component represents the increment of occupancy in each zone. In this way, we can significantly simplify the calculation of transition probability matrix which is a key parameter in Markov chain models. Several simulations with real data have been conducted to evaluate the performance of the proposed models under the two scenarios. To quantify the performance of the proposed models, five variables related to occupancy properties and two evaluation criteria are defined. The results show that our proposed approaches have superior performance over existing ones.

1. Introduction

Energy efficiency in buildings has attracted a great deal of attention because of global energy crisis and greater awareness of sustainable development. Energy consumption in buildings accounts for around 40% of the total energy usage in many countries [1]. This implies that large potential energy saving can be achieved in building sectors if more efficient control and management strategies are implemented. Since occupants directly contribute to the heat gain and trigger demand for heat sources such as lighting and computers, the occupancy dynamics will significantly affect the efficiency of building control systems. A number of works have been done in developing occupancy-based control systems [2–5]. Moreover, occupancy in buildings is an important factor in estimating building energy consumption [6–8], because the behaviors of occupants in buildings directly influence the usage of electronic devices. If an accurate occupancy dynamics model can be constructed, then, based on the occupancy dynamics obtained from the model as well as the relationship between occupancy patterns and energy consumption, accurate forecasting of building energy usage can be achieved [9–11]. Another motivation of building occupancy modeling is to extract occupancy patterns which are the key inputs of energy simulation tools, such as EnergyPlus [12], ESP-r [13], DeST [14,15], and TRNSYS [16]. These tools can produce yearly energy profiles of a building's cooling, heating, and ventilation needs as well as its lighting systems, which can help designers size associated systems. Note that the parameters of an occupancy dynamics model for a building before construction can be obtained from a similar building with the same function.

Several advanced models have been proposed in the literature to model building occupancy under scenarios of single-occupant single-zone (SOSZ), multi-occupant single-zone (MOSZ) and multi-occupant multi-zone (MOMZ). To simulate occupants' behavior in single person offices, Wang et al. [17] have proposed a non-homogeneous Poisson process model with two exponential distributions. These two exponential distributions are applied to describe distributions of vacancy intervals and occupancy intervals respectively. However, the assumption of the exponential distribution of occupancy intervals is rejected in goodness-of-fit test. Therefore, the model cannot simulate occupancy dynamics very well. Another prominent work of SOSZ can be found in [18] where Page et al. presented a two-state (in/out or occupied/unoccupied) inhomogeneous Markov chain model. Moreover, the authors defined a parameter of mobility, which represents the frequency of occupants entering or exiting a zone, to calculate

* Corresponding author. Tel.: +65 6790 5423.
E-mail addresses: chen0832@e.ntu.edu.sg (Z. Chen), XUJI0016@e.ntu.edu.sg (J. Xu), EYCSOH@ntu.edu.sg (Y.C. Soh).

http://dx.doi.org/10.1016/j.enbuild.2015.06.009
0378-7788/

Markov transition probability matrix. In their simulations, several key properties, e.g. the first arrival and last departure of an occupant, are applied to evaluate the performance of their model. However, the extension of the model to the multi-zone situation is non-trivial. For example, for a scenario with n zones (the outside is treated as zone $n+1$), the transition probability matrix will be of dimension $(n+1)\times(n+1)$ which is difficult to solve.

Due to the interactions among occupants or zones, modeling multi-occupant situations can be much more complicated. Gunathilak et al. [19] reported a generalized event-driven and group-based framework for modeling building occupancy. In their model, users are divided into groups based on similar properties. Then, the behavior of a user is fully driven by group and personal events. Wang et al. [20] also proposed an event-driven approach for the simulation of building occupancy. Within an event, a homogeneous Markov chain was employed to simulate occupancy patterns. The authors divided one day into the events of walking around, going to the office, getting off work and lunch break. Then, the parameters of the homogeneous Markov chain, i.e. transition probability matrix, are determined separately according to the properties of each event. These two approaches have some common problems. First, the definitions of events are not clear, and some internal events such as emotion change can be ambiguous to define. Second, the authors only presented a case study using simulation data instead of actual measurement data. Richardson et al. [21] proposed a first-order Markov chain model for simulating occupancy in residential buildings. The state of the Markov chain is the number of active occupants in a ten minutes interval. If the maximum number of active occupants is assumed to be m in a zone, then, the transition probability matrix at one time step is $m\times m$, which means that the dimension of transition probability matrix will increase with increased number of active occupants.

Erickson et al. [22] constructed multivariate Gaussian and agent-based models for capturing occupancy patterns in buildings. In the multivariate Gaussian model, the authors applied multivariate Gaussian distributions to fit room occupancy dynamics at each time step. The model does not consider the previous behaviors of occupants. In the agent-based model, the authors simulated each occupant individually by modeling their behaviors. However, this method does not take into account inter-room correlations and does not have any prediction capability. In order to obtain a better generalized model than their previous work in [22], Erickson et al. [2] proposed an inhomogeneous Markov chain model by considering the inter-room relationship for building occupancy modeling. The Markov chain model is applied in the MOMZ situation, and the states are vectors in which each component represents the number of occupants in each zone. In this case, the transition probability matrix of the Markov chain will increase exponentially when the number of zones or occupants in each zone increases. In addition, the authors mentioned that the model is inhomogeneous in the entire perspective but homogeneous within one hour. This assumption is not accurate in reality.

The last relevant work was done by Liao et al. in [23,24]. They presented an agent-based model with four modules to simulate each agent (occupant) behavior under three different scenarios of SOSZ, MOSZ and MOMZ. The four modules are defined according to the fact that occupants tend to stay at their working place for a long time, which is modeled as a damping process, and tend to leave hallways or restrooms quickly, which is modeled as an acceleration process. The main purpose of the modules is to obtain a Markov-like model without requiring any mathematical calculation of parameters, especially Markov transition probability matrix. During simulations, the model performed well under the SOSZ and MOSZ scenarios, but not as well under the MOMZ scenario which is much more complicated to model. To make up for this drawback of the approach, the authors proposed a graphical model in the MOMZ scenario, which is similar to the multivariate Gaussian model mentioned in [22]. Therefore, the same problem will occur in this graphical model.

In this paper, we propose two novel inhomogeneous Markov chain models under two scenarios of MOSZ and MOMZ for building occupancy modeling, which are based on our preliminary work in [25]. Note that we mainly focus on regular occupancy modeling in commercial buildings. As the SOSZ scenario is relatively easy and has already been very well studied in [18], we will not discuss this scenario in the paper, but we noted that the modeling of this scenario using the proposed method is easy and straight forward. In the MOSZ scenario, we present an inhomogeneous Markov chain whose state is defined as the increment of occupancy in a zone, which is different from the existing methods which use the number of occupants in a zone as the state. We assume that the maximum number of occupants moving into or out of a zone within a short interval is one. Then, the states of the Markov chain model are defined as: “1” which represents one occupant moving into the zone, “0” which represents that the number of occupants in the zone does not change, and “−1” which represents one occupant moving out of the zone. In this manner, the transition probability matrix will be only of dimension 3×3, which is much easier to solve than the existing approaches in the literature. The data for simulation is collected from real buildings in the University of Florida campus by the authors in [24]. In order to evaluate the performance of the proposed model, a comparison has been made between our model and the agent-based model in [24] which is very prominent in this area and uses the same data for model construction. In addition, mean occupancy profile and four random variables, namely time of first arrival, time of last departure, cumulative occupied duration and number of occupied/unoccupied transitions, are defined to test and compare the performance of the models. The simulation results show that the proposed model performs very well and outperforms the agent-based model in the MOSZ scenario. In the MOMZ scenario, considering interactions among zones, we propose another inhomogeneous Markov chain, where the state is a vector and each component represents the increment of occupancy in each zone, instead of the number of occupants in each zone as is usually the case. The same assumption is made in this situation that the maximum number of occupants moving into or out of a zone within a short interval is one. In this way, a greatly simplified model can be obtained as compared to the models in the literature. Simulations have been done using the actual measurement data containing four zones to evaluate the performance of the proposed framework. The proposed inhomogeneous Markov chain model performs very well for each zone in the MOMZ scenario.

2. Methodology

In this section, the basic theory of Markov chain will be introduced first. Then, we will present the modeling processes of the two scenarios, namely MOSZ and MOMZ. Finally, the calculation of Markov transition probability matrix will be presented.

2.1. Markov chain theory

A Markov chain is a sequence following Markov property which declares that the future state only depends on the current state [26]. Assume Ω is the set of all states of X which is the state variable. Then, a discrete time Markov chain $\{X_k\}$ satisfies

$$P(X_{k+1}=s_{k+1}|X_1=s_1,\ldots,X_k=s_k)=P(X_{k+1}=s_{k+1}|X_k=s_k) \quad (1)$$

where X_k is the state variable, $s_k\in\Omega$ is the state, and k is the time step.

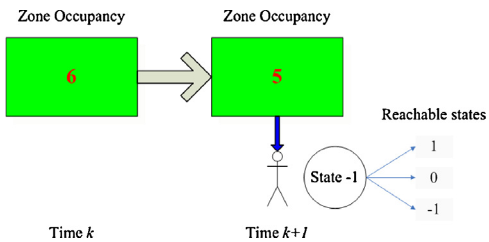


Fig. 1. The occupancy state representation in the MOSZ situation.

2.2. Modeling

In our inhomogeneous Markov chain model for the simulation of building occupancy in the MOSZ situation, the state of the Markov chain is the increment of occupancy in a zone, which is expressed as $N_k^i - N_{k-1}^i$, where N_k^i is the occupancy level in zone i at time step k. One assumption is made, that is, the maximum number of occupants moving into or out of a zone is one within a short interval. This short interval can be determined in terms of real occupancy data. Fig. 1 illustrates a simple example of the proposed model. Precisely, at time step k, there are 6 occupants in a zone. Then, one occupant has moved out of the zone at time step $k+1$, therefore, the current state is -1, and the next reachable state can be 1, 0 or -1. In this manner, we obtain a simple 3×3 transition probability matrix independent of the maximum number of occupants. As a comparison, assuming that the maximum number of occupants in a zone is 20, then with the approach in [21], the state of the Markov chain can be 0–20, resulting in 21 states. The corresponding transition probability matrix will be of dimension 21×21, which is much more complicated.

In the MOMZ situation, our inhomogeneous Markov chain model takes interactions among zones into consideration, and the state of the Markov chain is a vector in which each component is the increment of occupancy in each zone. The same assumption is made that the maximum number of occupants moving into or out of a zone is one within a short interval. Therefore, the component of a state can be 1, -1 or 0. Fig. 2 gives an example illustrating the mechanism of the proposed approach in the MOMZ situation. The occupancies of Zone1, Zone2, Zone3 and Zone4 are 8, 4, 3 and 0 respectively at time step k. At next time step $k+1$, one occupant has moved into Zone2 and Zone4, one occupant moved out of Zone1, and no occupant moved into or out of Zone3 at time step $k+1$.

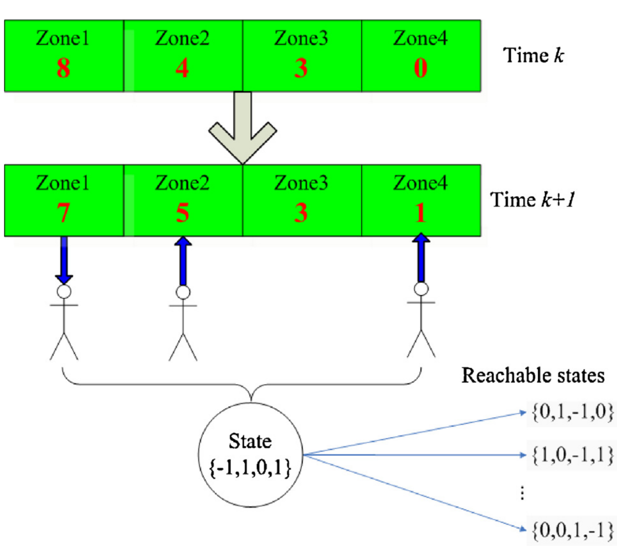


Fig. 2. The occupancy state representation in the MOMZ situation.

Therefore, the corresponding state is $\{-1, 1, 0, 1\}_{k+1}$. Since the total number of states is $3^4 = 81$ in terms of the state definition, the final transition probability matrix has the dimension of $3^4 \times 3^4$ independent of the number of occupants. For the same MOMZ situation, assuming that the maximum number of occupants in each zone is 20, if the advanced model in [2], which defined the state as a vector whose components are the number of occupants in a zone, is employed, then each zone will have 21 states. Therefore, the dimension of the transition probability matrix will be $21^4 \times 21^4$, which is huge and difficult to solve.

2.3. Calculation of Markov transition probability matrix

We adopt maximum likelihood estimation (MLE) to determine the parameters of transition probability matrix in this paper. Let's define two states, 0 and 1 for a state variable Y_k in a Bernoulli experiment. Assume the probability of $Y_k = 1$ is p. Therefore, the likelihood function of p given the observation, y_k, is

$$L(p \mid y_k) = Pr(Y_k = y_k) = \begin{cases} p, & y_k = 1 \\ 1-p, & y_k = 0 \end{cases} \tag{2}$$

Normally, we apply log-likelihood function, $\ln L(p \mid y_k)$, which will simplify the calculation by converting product operation into summation. The MLE estimate can be obtained by maximizing the log-likelihood function [27]. Suppose, the log-likelihood function is differentiable, and a maximum likelihood solution exists. Then, the following partial differential equation called the *likelihood equation* must be satisfied:

$$\frac{\partial \ln L(p \mid y_k)}{\partial p} = 0 \tag{3}$$

where $p = p_{MLE}$ is the solution. The likelihood equation is only a necessary condition for the existence of an MLE estimate. We require an additional condition to make sure that the solution is the maximum or minimum, which can be achieved by using the second derivative of the log-likelihood function. For the purpose of maximizing, the second derivative of the log-likelihood function should be less than zero at the maximum point, and is expressed as:

$$\frac{\partial^2 \ln L(p \mid y_k)}{\partial p^2} < 0 \tag{4}$$

In an inhomogeneous Markov chain model, the transition from the state s_k to the state s_{k+1} at one time step can be treated as a Bernoulli experiment happening with the probability $Pr(X_{k+1} = s_{k+1} | X_k = s_k)$. Note that the likelihood function of this conditional probability is only based on the data $X_k = s_k$ from different days at a particular time step which is assumed to be independent. Let Ω denote the data set we collect, M the total number of days where $X_k = s_k$. Within these days, let W be the total number of days where $X_{k+1} = s_{k+1}$, and p the probability $Pr(X_{k+1} = s_{k+1} | X_k = s_k)$. Then, the log-likelihood function is

$$\begin{aligned} lnL(p \mid \Omega) &= \ln \frac{M!}{W!(M-W)!} p^W (1-p)^{M-W} \\ &= \ln \frac{M!}{W!(M-W)!} + Wlnp + (M-W)ln(1-p) \end{aligned} \tag{5}$$

The corresponding *likelihood equation* is

$$\frac{\partial lnL(p \mid \Omega)}{\partial p} = \frac{W}{p} - \frac{M-W}{1-p} = 0 \tag{6}$$

and the MLE solution is $p_{MLE} = W/M$. To make sure that the MLE estimate is the maximum point, the second derivative of the log-likelihood function at point $p = p_{MLE} = W/M$ is checked, which yields

$$\frac{\partial^2 lnL(p \mid \Omega)}{\partial p^2} = -\frac{W}{p^2} - \frac{M-W}{(1-p)^2} < 0 \tag{7}$$

and it is negative, as expected. Therefore, the solution is the MLE estimate. For each transition probability, $Pr(X_{k+1} = s_{k+1} | X_k = s_k)$, we can calculate it separately using the MLE estimator. Finally, the transition probability matrix can be obtained in this way.

3. Evaluation

In order to evaluate the performance of the proposed models, simulations with real data have been done under two scenarios of MOSZ and MOMZ. First, the data for simulations is introduced. Then, some variables related to occupancy properties and two evaluation criteria are presented. Finally, we will present the simulation results and discussions of the two scenarios.

3.1. Data

For the MOSZ scenario, the data for the simulations is generously provided by the authors in [24], and it was collected by using a wireless camera near the entrance of a zone from January 2010 to April 2010, a span of about four months. The occupancy level is manually determined from the video. Due to certain technical issues, only 70 days of data were collected from 16 weeks of video data. We compare our results with the agent-based model which also used the same data in model construction. For the MOMZ scenario, there are 31 full days of data from 57 cubicles with 2 minutes time step. The data spans from August 2009 to January 2010. According to the detailed analysis in [28], a few cubicle data (6, 20, 26, 56, 57) are considered abnormal with respect to the others. Therefore, these cubicle data are not considered. We uniformly divide the remaining 52 cubicles into 4 zones, namely “Zone1”, “Zone2”, “Zone3” and “Zone4”, for the MOMZ simulation.

3.2. Variables and criteria

To evaluate the performance of the proposed model, five variables related to occupancy properties need to be defined. The definitions of the variables are similar to those in [24]. Suppose $Q_{occupied}$ and $Q_{unoccupied}$ are the thresholds of occupied and unoccupied states respectively, and T is the time interval, the variables are defined are as follows:

- *Mean occupancy*: The mean occupancy of zone i at time step k is defined as $E(X_k^i)$, where $E(\cdot)$ is the expectation operation.
- *Time of first arrival*: The time of first arrival is the first time when the zone becomes occupied. Precisely, for zone i, if $X_k^i \geqslant Q_{unoccupied}$ and $X_t^i < Q_{unoccupied}$ for all $t < k$, where $Q_{unoccupied}$ should be chosen properly, then, k is the time of first arrival.
- *Time of last departure*: The time of last departure is the time from which the zone becomes unoccupied. For zone i, if $X_k^i \geqslant Q_{unoccupied}$ and $X_t^i < Q_{unoccupied}$ for all $t > k$, then, $k + 1$ is the time of last departure.
- *Cumulative occupied duration*: The cumulative occupied duration is the total length of time when the zone is occupied. Precisely, for zone i, it is the number of elements of the set $\{k | X_k^i \geqslant Q_{occupied}, 1 \leqslant k \leqslant 24 \times 60/T\}$ for each day.
- *Number of occupied/unoccupied transitions*: The number of occupied/unoccupied transitions is the total number of transitions between “occupied” and “unoccupied” status of a zone. For zone i, it is the number of elements of the set $\{k | X_k^i \geqslant Q_{unoccupied}, X_{k+1}^i < Q_{unoccupied}, 1 \leqslant k \leqslant 24 \times 60/T\} \bigcup \{k | X_k^i < Q_{unoccupied}, X_{k+1}^i \geqslant Q_{unoccupied}, 1 \leqslant k \leqslant 24 \times 60/T\}$ for each day.

In order to quantify the performance of the proposed models in terms of above-mentioned variables, two evaluation criteria are chosen as follows.

The first Criterion of normalized root mean square deviation (NRMSD) is defined to compare the difference between mean occupancy profiles predicted by models and estimated from measurements. Given two time series, $x(k)$ and $y(k)$, $k = 1, 2, \ldots, K$, the NRMSD between x and y is defined as

$$NRMSD(x, y) = \frac{\|x - y\| / \sqrt{K}}{\max(z) - \min(z)}, \tag{8}$$

where $x = [x(1), x(2), \ldots, x(K)]^T$, $y = [y(1), y(2), \ldots, y(K)]^T$, $z = [x^T, y^T]^T$, T is the vector transpose operation, $\|\cdot\|$ is the Euclidean norm and K is the length of the sequences.

The second criterion is Kullback-Leibler (K-L) divergence which is applied to compare two distributions of the random variables predicted by models and estimated from measurements [29]. K-L divergence between p and q represents the distance of these two densities and is defined as

$$d(p \| q) = \sum_k p_k \log(\frac{p_k}{q_k}), \tag{9}$$

where p_k and q_k are two probability mass functions (pmf). In standard convention, $0 \log(\frac{0}{q_k}) = 0$ and $p_k \log(\frac{p_k}{0}) = \infty$. In this paper, we assume $p_k \log(\frac{p_k}{0}) = 0$, same as in [24].

3.3. Simulation results and discussions

Since the increment information is applied instead of the exact value of occupancy, we can significantly simplify the calculation of the key parameters, i.e. transition probability matrices. But one problem arises because of this increment information. That is, the simulation results may diverge when transferring the increment value into the final occupancy level. To solve this problem, maximal and minimal constraints are employed, which are obtained from real occupancy data. During simulations, if the current occupancy level after transferring violates the constraints, we will manually set the current state into a feasible state to overcome this violation. Precisely, assume $O_k^{\max}$ and $O_k^{\min}$ are the maximal and minimal constraints of occupancy at time step k, the current state is s_k, and the current occupancy is o_k after transferring the increment information into occupancy. If $o_k > O_k^{\max}$, we keep the o_k at the current maximum, $O_k^{\max}$. Similarly, if $o_k < O_k^{\min}$, we keep the o_k at the current minimum, $O_k^{\min}$.

Markov Chain Monte Carlo (MCMC) [30,31] simulations are conducted using the proposed models. The initial state of MCMC is generated in terms of the probability of the initial state in actual measurement data, which is expressed as $P(X_0 = s_0) = N_{s_0}/N_{all}$. In the equation, N_{s_0} is the number of days where the initial state is s_0 and N_{all} is the total number of days.

Since the agent-based model in [24] simulated one thousand weeks, one thousand MCMC simulations, each with a duration of one week, have been conducted using our proposed model in the MOSZ scenario. We compare the mean occupancy profiles that are estimated from the measurements, predicted by the agent-based model and our proposed model by using the criterion of NRMSD. In addition, four random variables of time of first arrival, time of last departure, cumulative occupied duration and number of occupied/unoccupied transitions, of the proposed model and those of the agent-based model are also tested using the criterion of K-L divergence. Note that, only the data for weekdays is chosen for the two criteria, and the threshold is $Q_{unoccupied} = Q_{occupied} = 0.5$.

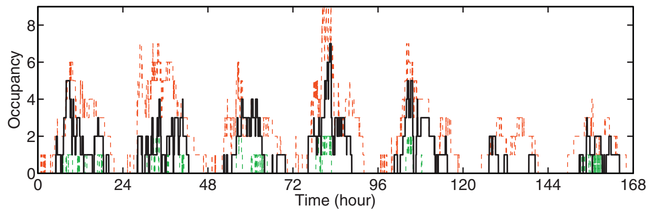


Fig. 3. The output of the proposed model (black) for one week, the maximal occupancy constraint (dashed red) and the minimal occupancy constraint (dashed green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

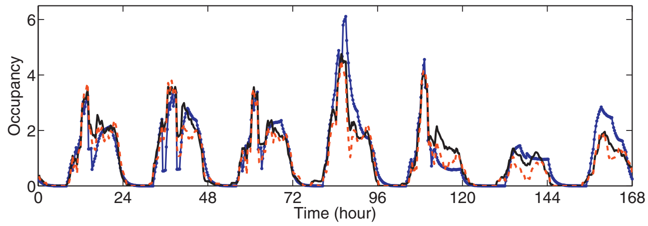
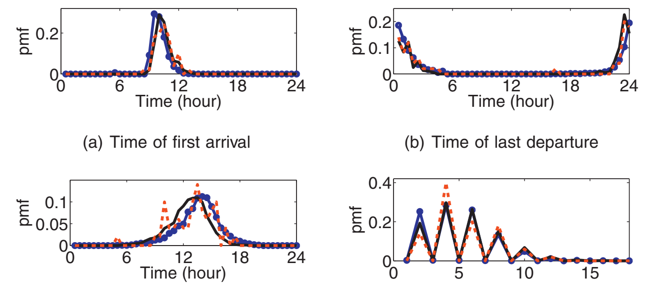


Fig. 4. Mean occupancy profiles estimated from measurements (dashed red), predicted by the agent-based model (dotted blue) and the proposed model (black) in the MOSZ scenario. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) Time of first arrival (b) Time of last departure

(c) Cumulative occupied duration (d) No. of occupied/unoccupied transitions

Fig. 5. The pmfs of the four random variables estimated from the measurements (dashed red), predicted by the agent-based model (dotted blue) and the proposed model (black). Comparison is for weekdays only, and the binsize is 1/2 h. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The output of the proposed model for one week, together with the maximal and minimal occupancy constraints, are shown in Fig. 3. Note that it is a research lab environment in the University of Florida campus. As the model is stochastic, the model output for each run would be different. In order to evaluate the performance of the proposed model, statistical properties of the model output will be tested. The comparison of the mean occupancy profiles estimated from measurements, predicted by the agent-based model and the proposed model is shown in Fig. 4. We can see that the proposed model matches the measurements much better than the agent-based model. Fig. 5a and b shows the pmfs of time of first arrival and time of last departure. It shows that the performance of the proposed model is much better than the agent-based model, and the proposed model can also capture certain peaks of the pmfs of the measurements, while the agent-based model performs much worse at these points. The comparison of cumulative occupied duration estimated from measurements, predicted by the agent-based model and the proposed model, is shown in Fig. 5c. Both models could not capture the peaks of the pmf, but the overall distributions of the pmf are predicted correctly. Fig. 5d gives the pmf of the number of occupied/unoccupied transitions. The agent-based model has already matched very well, but our proposed model still has a better performance, especially at certain points where the number of transitions is 2, 8, 10 and 12 in the pmf. Table 1 shows the NRMSD of the mean occupancy and the K-L divergence of the four random variables under the MOSZ scenario. It can be found that the proposed model performs much better than the agent-based model for the variables of mean occupancy, time of first arrival, time of last departure and number of occupied/unoccupied transitions. In the meantime, both the proposed model and the agent-based model could not predict cumulative occupied duration very well, but our proposed model still performs better than the agent-based model. These conclusions tally with the observations of Fig. 4 and 5. One possible reason for the big fluctuations in cumulative occupied duration could be due to the limited data size of the measurements. The data predicted by the agent-based model and the proposed model involves 1000 weeks. Therefore, the pmfs predicted by both models are much smoother.

In the MOMZ scenario, even though the proposed method can dramatically reduce the dimension of the transition probability matrix as compared to existing approaches, the dimension of the transition probability matrix will still increase exponentially when the number of zones increases. We can represent the relationship as $dimension = 3^m$, where m is the number of zones. According to the observation of the real data, we find that, although the total number of states can be very large, only a small portion of the states are possible in one time step. This phenomenon is mainly caused by the topology constraints of buildings. For example, assume that we have four zones with cascaded connection (shown in Fig. 2), if one occupant wants to go to Zone3 from Zone1, and then he/she will need to go to Zone2 first. Because of this physical constraint, we only consider the states that appear in the real data in simulations for simplification.

For the MOMZ scenario, 52 cubicles are divided into 4 zones. According to the analysis in [28], workdays have similar occupancy profile. And, due to the limited data size, we treat all working days as the same instead of separating them from Monday to Friday in the MOSZ scenario. The analysis in [28] also provides a possible reason for high occupancy after 7 pm, which is caused by a security guard walking around the zones. Thus, we only model occupancy during 4 am to 7 pm. In this case, the random variable of time of last departure cannot be obtained. Only the remaining three random variables are analyzed.

In order to match the number of days in actual measurement data, thirty-one MCMC simulations in which each has a

Table 1
NRMSD of the mean occupancy and K-L divergence of the four random variables in the MOSZ scenario.

Criteria	Variables	Agent-based model	Proposed model
NRMSD	Mean occupancy	0.0867	0.0577
K-L divergence	Time of first arrival	0.2388	0.0484
	Time of last departure	0.1900	0.0411
	Cumulative occupied duration	0.3261	0.2394
	No. of occupied/unoccupied transitions	0.0748	0.0348

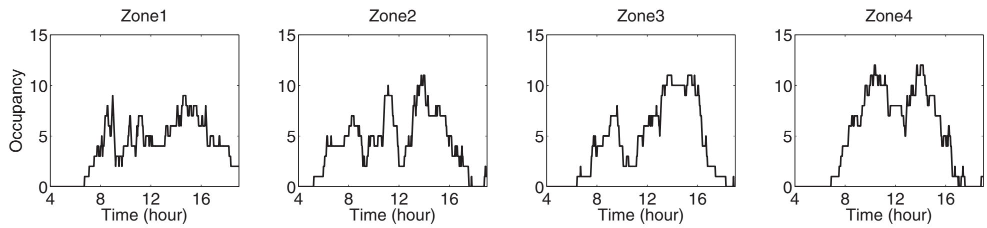


Fig. 6. The output of the proposed model of the four zones for one day.

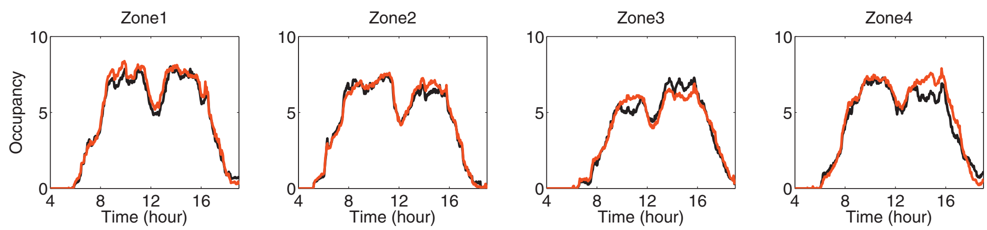


Fig. 7. Mean occupancy profiles of the four zones estimated from measurements (red line) and predicted by the proposed model (black line) in the MOSZ scenario. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

duration of one day have been performed using the proposed model under the MOMZ scenario. The data of the four zones is analyzed separately using the two evaluation criteria. The variables and parameters are the same as the MOSZ scenario, except for the time of last departure due to the reason mentioned earlier.

The output of the proposed model for one day of the four zones is shown in Fig. 6. Since the proposed model is stochastic, the result is different for different days. Fig. 7 shows the mean occupancy estimated from the measurements and predicted by the proposed model. The mean occupancy profile predicted by the model matches that of the measurements very well for the four

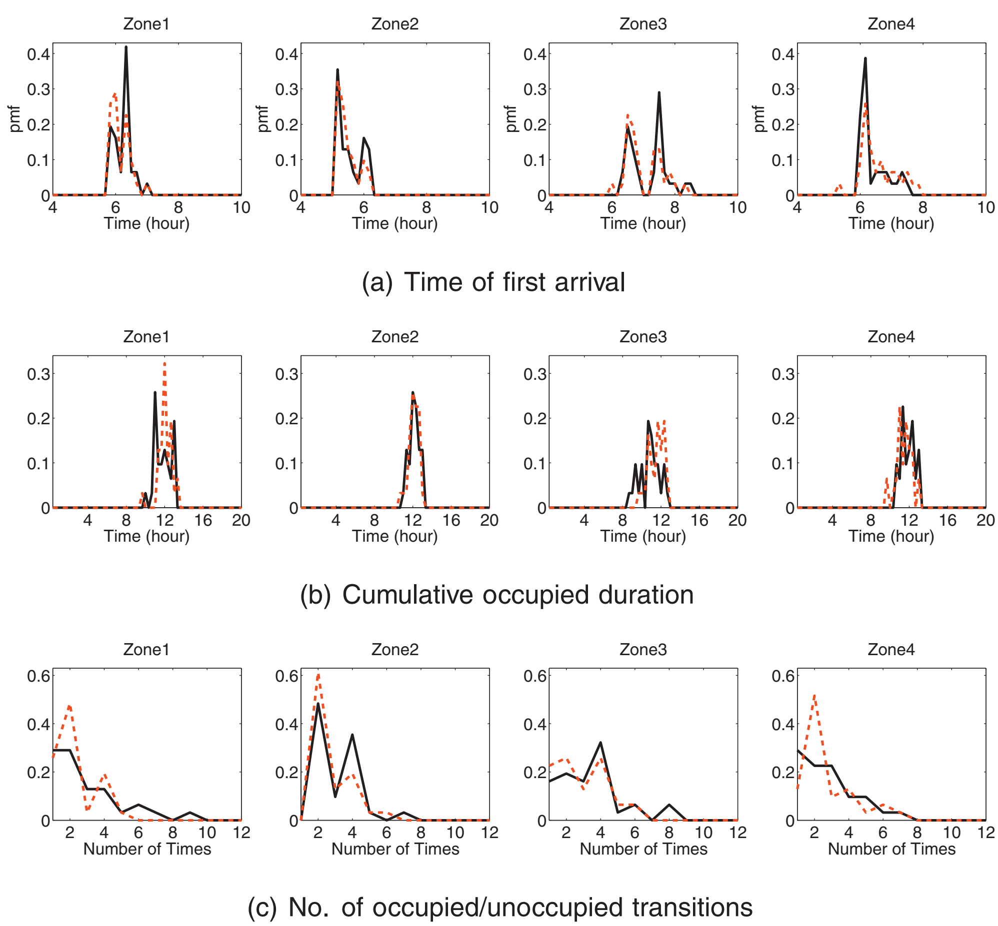


Fig. 8. The pmfs of the three random variables estimated from the measurements (dashed red) and predicted by the proposed model (black). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
NRMSD of the mean occupancy and K-L divergence of the three random variables for the four zones in the MOMZ scenario.

Criteria	Variables	Zone1	Zone2	Zone3	Zone4
NRMSD	Mean occupancy	0.0453	0.0336	0.0573	0.0735
K-L divergence	Time of first arrival	0.1220	0.0932	0.1373	0.1209
	Cumulative occupied duration	0.5247	0.1193	0.3515	0.2118
	No. of occupied/unoccupied transitions	0.2505	0.0647	0.1086	0.2864

zones. Note that the mean occupancy profiles are not as smooth as that in the MOSZ scenario. This is because the number of days for the measurements and simulations is only 31, which is much smaller than the number of days under the MOSZ scenario. Fig. 8 shows the three random variables of time of first arrival, cumulative occupied duration and number of occupied/unoccupied transitions estimated from the measurements and predicted by the proposed model. We conclude that the proposed model captures the properties of real occupancy data very well. Table 2 shows the quantified results with the two evaluation criteria for the four zones. The proposed model performs very well for all the variables under the two criteria in each zone, which is consistent with the observations of Fig. 7 and 8. In addition, we observe that the mean occupancy matches better under the MOMZ scenario than that of the MOSZ scenario, and the span of three random variables are smaller. One possible reason is that the data may come from a more regular place, i.e. company or office, instead of the typical research lab under the MOSZ situation.

4. Conclusion and future works

In this paper, two stochastic models are proposed for building occupancy modeling under multi-occupant single-zone (MOSZ) and multi-occupant multi-zone (MOMZ) scenarios. In the MOSZ scenario, an inhomogeneous Markov chain where the state is the increment of occupancy in a zone is proposed. Under the assumption that the maximal number of occupants moving into or out of a zone is one within a short interval, the transition probability matrix will be of dimension 3×3 which is much smaller than existing approaches in the literature. We remark that the proposed approach can be readily extended to the case where the number of occupants moving into or out of a zone within a short interval can be 2 or 3, with the dimensions of the transition probability matrices increased to 5×5 and 7×7, respectively. In the more complicated MOMZ scenario, taking interactions among zones into consideration, we present another inhomogeneous Markov chain where the state is a vector in which each component is the increment of occupancy in each zone. In this way, significantly simplified transition probability matrices are obtained. Simulations have been conducted to evaluate the performance of the proposed models using actual measurement data. In addition, some variables related to occupancy properties and two evaluation criteria are defined to quantify the performance of the models. In the MOSZ scenario, the proposed model performs very well, and significantly outperforms the agent-based model which models building occupancy very well. In the MOMZ scenario, the proposed model performs very well for all the variables under the two evaluation criteria in each zone.

Our future works will mainly focus on constructing occupancy-based control systems. The potential energy saving will be quantified using energy simulation tools, e.g. EnergyPlus or ESP-r, between the occupancy-based strategy and static rules. Besides energy efficiency, human comfort level is another important parameter that needs to be tested in our future works.

Acknowledgments

The authors would like to thank Dr. Chenda Liao and Dr. Francis Rubinstein for providing us their valuable data. This work is supported by the Building Efficiency and Sustainability in the Tropics (SinBerBEST) program which is funded by Singapore's National Research Foundation and led by the University of California, Berkeley (UCB) in collaboration with Singapore Universities.

References

[1] B. Dong, B. Andrews, K.P. Lam, M. Höynck, R. Zhang, Y.-S. Chiou, D. Benitez, An information technology enabled sustainability test-bed (ITEST) for occupancy detection through an environmental sensing network, Energy Build. 42 (7) (2010) 1038–1046.
[2] V.L. Erickson, M.Á. Carreira-Perpiñán, A.E. Cerpa, Observe: occupancy-based system for efficient reduction of HVAC energy, in: 10th International Conference on Information Processing in Sensor Networks (IPSN), 2011, IEEE, 2011, pp. 258–269.
[3] T. Leephakpreeda, R. Thitipatanapong, T. Grittiyachot, V. Yungchareon, Occupancy-based control of indoor air ventilation: a theoretical and experimental study, Sci. Asia 27 (4) (2001) 279–284.
[4] X. Guo, D. Tiller, G. Henze, C. Waters, The performance of occupancy-based lighting control systems: a review, Light. Res. Technol. 42 (4) (2010) 415–431.
[5] S. Goyal, H.A. Ingley, P. Barooah, Occupancy-based zone-climate control for energy-efficient buildings: complexity vs. performance, Appl. Energy 106 (2013) 209–221.
[6] C. Martani, D. Lee, P. Robinson, R. Britter, C. Ratti, Enernet: studying the dynamic relationship between building occupancy and energy consumption, Energy Build. 47 (2012) 584–591.
[7] B. Abushakra, D.E. Claridge, Modeling office building occupancy in hourly data-driven and detailed energy simulation programs, ASHRAE Trans. 114 (2) (2008).
[8] J. Virote, R. Neves-Silva, Stochastic models for building energy prediction based on occupant behavior assessment, Energy Build. 53 (2012) 183–193.
[9] G.R. Newsham, B.J. Birt, Building-level occupancy data to improve ARIMA-based electricity use forecasts, in: Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building, ACM, 2010, pp. 13–18.
[10] S.S. Kwok, E.W. Lee, A study of the importance of occupancy to building cooling load in prediction by intelligent approach, Energy Convers. Manag. 52 (7) (2011) 2555–2564.
[11] S.S. Kwok, R.K. Yuen, E.W. Lee, An intelligent approach to assessing the effect of building occupancy on building cooling load prediction, Build. Environ. 46 (8) (2011) 1681–1690.
[12] D.B. Crawley, L.K. Lawrie, F.C. Winkelmann, W.F. Buhl, Y.J. Huang, C.O. Pedersen, R.K. Strand, R.J. Liesen, D.E. Fisher, M.J. Witte, et al., Energyplus: creating a new-generation building energy simulation program, Energy Build. 33 (4) (2001) 319–331.
[13] ESRU, Esp-r: a building and plant energy simulation environment, user guide version 9 seriesl.
[14] D. Yan, J. Xia, W. Tang, F. Song, X. Zhang, Y. Jiang, DeST – an integrated building simulation toolkit: part i: fundamentals, in: Building Simulation, vol. 1, Springer, 2008, pp. 95–110.
[15] X. Zhang, J. Xia, Z. Jiang, J. Huang, R. Qin, Y. Zhang, Y. Liu, Y. Jiang, DeST – an integrated building simulation toolkit: part ii: applications, in: Building Simulation, vol. 1, Springer, 2008, pp. 193–209.
[16] S. Klein, W. Beckman, J. Mitchell, J. Duffie, N. Duffie, T. Freeman, J. Mitchell, J. Braun, B. Evans, J. Kummer, et al., TRNSYS 16 – A TRaNsient System Simulation Program, User Manual, Solar Energy Laboratory, University of Wisconsin – Madison, Madison, 2004.
[17] D. Wang, C.C. Federspiel, F. Rubinstein, Modeling occupancy in single person offices, Energy Build. 37 (2) (2005) 121–126.
[18] J. Page, D. Robinson, N. Morel, J.-L. Scartezzini, A generalised stochastic model for the simulation of occupant presence, Energy Build. 40 (2) (2008) 83–98.
[19] G. Gunathilak, A. Prasannakumar, N. Nazarian, H. Naeimi, A generalized event driven framework for building occupancy (WIP), in: Proceedings of the

Symposium on Simulation for Architecture & Urban Design, Society for Computer Simulation International, 2013, p. 26.
[20] C. Wang, D. Yan, Y. Jiang, A novel approach for building occupancy simulation, in: Building Simulation, vol. 4, Springer, 2011, pp. 149–167.
[21] I. Richardson, M. Thomson, D. Infield, A high-resolution domestic building occupancy model for energy demand simulations, Energy Build. 40 (8) (2008) 1560–1566.
[22] V.L. Erickson, Y. Lin, A. Kamthe, R. Brahme, A. Surana, A.E. Cerpa, M.D. Sohn, S. Narayanan, Energy efficient building environment control strategies using real-time occupancy measurements, in: Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings, ACM, 2009, pp. 19–24.
[23] C. Liao, P. Barooah, A novel stochastic agent-based model of building occupancy, in: American Control Conference (ACC), 2011, IEEE, 2011, pp. 2095–2100.
[24] C. Liao, Y. Lin, P. Barooah, Agent-based and graphical modelling of building occupancy, J. Build. Perform. Simul. 5 (1) (2012) 5–25.
[25] Z. Chen, Y.C. Soh, Modeling building occupancy using a novel inhomogeneous Markov chain approach, in: IEEE International Conference on Automation Science and Engineering (CASE), 2014, IEEE, 2014, pp. 1079–1084.
[26] G. Grimmett, D. Stirzaker, Probability and Random Processes, Oxford University Press, 2001.
[27] I.J. Myung, Tutorial on maximum likelihood estimation, J. Math. Psychol. 47 (1) (2003) 90–100.
[28] P.D. Andersen, A. Iversen, H. Madsen, C. Rode, Dynamic modeling of presence of occupants using inhomogeneous Markov chains, Energy Build. 69 (2014) 213–223.
[29] T.M. Cover, J.A. Thomas, Elements of Information Theory, John Wiley & Sons, 2012.
[30] W.R. Gilks, S. Richardson, D.J. Spiegelhalter, Introducing Markov chain Monte Carlo, in: Markov chain Monte Carlo in Practice, Springer, 1996, pp. 1–19.
[31] D. Gamerman, H.F. Lopes, Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, CRC Press, 2006.