

This article was downloaded by: [University of Florida], [Prabir Barooah]

On: 08 July 2011, At: 06:33

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

## Journal of Building Performance Simulation

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/tbps20>

### Agent-based and graphical modelling of building occupancy

Chenda Liao<sup>a</sup>, Yashen Lin<sup>a</sup> & Prabir Barooah<sup>a</sup>

<sup>a</sup> Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, Florida, USA

Available online: 20 January 2011

To cite this article: Chenda Liao, Yashen Lin & Prabir Barooah (2011): Agent-based and graphical modelling of building occupancy, Journal of Building Performance Simulation, DOI:10.1080/19401493.2010.531143

To link to this article: <http://dx.doi.org/10.1080/19401493.2010.531143>



PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan, sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

## Agent-based and graphical modelling of building occupancy

Chenda Liao, Yashen Lin and Prabir Barooah\*

*Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, Florida, USA*

*(Received 30 July 2010; final version received 7 October 2010)*

We propose a novel stochastic agent-based model of occupancy dynamics in a building with an arbitrary number of zones and occupants. Simulation of the model yields time-series of the location of each agent (a software representation of an occupant). The model is meant to provide realistic simulation of occupancy dynamics in non-emergency situations. Comparison of the model's prediction of distributions of random variables such as first arrival time of a building is provided against those estimated from measurements in commercial buildings. We also propose a lower complexity graphical model of occupancy evolution in multi-zone buildings. The graphical model captures information on mean occupancy and correlation among occupancy at various zones in the building. The agent-based model can be used in conjunction with building performance simulation tools, while the graphical model is more suitable for real-time applications, such as occupancy estimation with noisy sensor measurements.

**Keywords:** agent-based model; building occupancy; occupancy model; graphical model; building energy performance.

### 1. Introduction

There is an increasing emphasis on developing methods to design and operate buildings in a way to make them smart and energy efficient. In addition to reducing energy use while maintaining high indoor air quality and thermal comfort levels, the smart buildings of the future are expected to provide services such as controlled egress in case of emergencies and targeted advertisement. Modelling occupancy dynamics in buildings is going to be increasingly important in achieving this vision. We use the word *occupancy* to mean the number of people in a zone (or building) at any given time. A model of occupancy dynamics is a mathematical tool to predict occupancy in a building as a function of time given some initial conditions. Such predictions can serve as inputs to building energy simulation tools in the design and recommissioning phase. Cooling and heating loads experienced by the building HVAC (heating, ventilation and air-conditioning) system depend on the sensible heat gains, which strongly depend on occupancy. A tool capable of realistic simulation of occupancy evolution over time is therefore useful in simulation and analysis of building energy use. Several building energy simulation programmes such as EnergyPlus and ESP-r can incorporate occupancy information in computing loads. Currently, the most popular method of incorporating information about time-variations in occupancy to energy simulations is through schedules and diversity

factors (Abushakra *et al.* 2001). However, these methods are not meant to capture occupancy dynamics. Diversity factors, for example, are used to provide a correction to mean heat gain calculations from occupancy schedules so that the mean heat gain so computed are more representative of the true mean value. A model of occupancy dynamics can provide more realistic sample time-traces of occupancy over time from which, peaks, means and variances of occupancy driven heat gains can be computed.

Models of occupancy dynamics can be useful in commissioning and re-commissioning of buildings as well. The number and behaviour of occupants change over time, and can be quite different from what was expected during the design phase. A model of occupancy dynamics can be constructed for a specific building by surveying the building occupants (and calibrated based on limited number of measurements) to provide accurate occupancy statistics for that specific building. HVAC equipment schedules can be optimized based on the the relevant statistics, such as means, variances and max values of occupancy-driven heat gains computed from the model's predictions.

Such models can also be useful in developing advanced building control algorithms such as those based on model predictive control (MPC). Researchers are working on developing MPC algorithms that seek to minimize HVAC energy use based on weather forecasts and occupancy predictions (Gyalistras and

---

\*Corresponding author. Email: pbarooah@ufl.edu

Gwerder 2009, Oldewurtel *et al.* 2010). Such algorithms can utilize models of occupancy dynamics for obtaining occupancy predictions. Another use of such models is real-time estimation of zone-level occupancy in a building from limited number of sensors. Real-time occupancy estimates are useful in providing information to first responders and in performing controlled egress in the event of an emergency (Tomastik *et al.* 2008). Certain control techniques designed to reduce HVAC energy use, such as in demand controlled ventilation, also need real-time occupancy estimation capability. There are several types of sensors that can provide information on occupancy indirectly, such as CO<sub>2</sub> sensors, video cameras and Passive infrared (PIR) motion detectors. However, sensor measurements alone may not be enough for accurate estimation as they suffer from large measurement error (Hutchins *et al.* 2007, Meyn *et al.* 2009). Filtering techniques can be used to compensate for measurement errors by fusing noisy sensor measurements with predictions from a model (Meyn *et al.* 2009, Liao and Barooah 2010). This requires a model of occupancy dynamics.

The requirements of an occupancy dynamics model may differ depending on the intended application of the model. For use in building design, an occupancy model should be able to predict statistics of occupancy related variables, e.g. mean and variance of occupancy, distribution of the first arrival time of occupants in a zone or building, etc. Since the number of occupants directly impact the sensible heat gains and that from lighting and equipment, fluctuations in the building load can be predicted accurately only if fluctuations in occupancy can be modelled accurately. It is estimated that building HVAC systems are frequently over-designed (Vieira *et al.* 1996). Sizing HVAC equipment for worst case peak load conditions does improve the building's ability to handle such loads. However, this leads to poor energy efficiency of the HVAC system while such worst-case loads are rarely encountered. Ability to model stochastic fluctuations in occupancy realistically can translate into the ability to determine the most efficient sizing of equipment so that the HVAC system operates at maximum possible efficiency most of the time, even though it may perform poorly for those rare occasions when peak demand is experienced. If design for peak load handling is the goal, then the model has to be able to predict the peak occupancy, since the maximum load that a building experiences depends on peak occupancy. For real-time control and estimation, the models have another requirement: that of simplicity. An overly complex model with high computational requirement is not suitable for real-time applications.

Constructing mathematical models of occupancy dynamics in a building is a challenging problem

because of the high uncertainty of people movement that governs occupancy evolution. On the high-resolution end of the spectrum of modelling possibilities lie the so-called agent-based models. An agent-based model consists of agents (encoded in software) in which each agent is endowed with a set of behaviours that are designed to mimic behaviour of humans under situations that the model is meant to study. Computer simulation with an agent-based model can be used to generate time-traces of each occupant's location. These can then be aggregated to yield time traces of occupancy of each zone or of the entire building. An extensive literature exists on agent-based models for a diverse set of applications during the last 40 years; see the review article (Helbing 2001) and reference therein. However, almost all the work on agent-based modelling of occupants in buildings have been designed to study emergency situations such as fire and explosions (Shendarkar *et al.* 2006, Pelechano and Malkawi 2008, Gwynne and Kuligowski 2009). To the best of our knowledge, little work has been done on modelling building occupancy dynamics during normal, day-to-day operations using agent-based models.

On the lower end of the spectrum, both in terms of complexity as well as predictive capability, are models with low temporal and spatial resolution that only seek to predict for the whole-building at an hourly rate. For example, Abushakra and Claridge (2008) propose a method for predicting total building occupancy based on estimated correlation between occupancy and lighting/equipment load. A stochastic model of occupancy in UK households was proposed in Richardson *et al.* (2008) that can be used to generate statistically representative sample paths of household occupancy. A probabilistic model to predict and simulate occupancy in a single person office was proposed by Wang *et al.* (2005). Vacant and occupied intervals are modelled as random variables with exponential distributions. The results of the study were mixed. Yamaguchi *et al.* (2003) proposed a model to simulate the "working states" of a single agent, where the working states could take one of four possible values (using one PC, using two PCs, not using PCs and being out). The transition between states was described in terms of a Markov chain.

Among models of a single person's occupying (or not occupying) an office, the most recent and comprehensive model is the one proposed in Page *et al.* (2008). They model the dynamics of a single person in a single-occupancy room by a Markov chain with two states (in/out, or, occupied/unoccupied). The model also incorporates periods of long absence. The Markov chain is time-inhomogeneous and is described by a sequence of  $2 \times 2$  transition probability matrices  $\mathbf{P}(k)$ , where  $k = 1, 2, \dots$  represents a discrete time

index. Each increment of the time index represents a 15 min interval. The model requires as input the sequence  $\mathbf{P}(k)$ ,  $k = 1, \dots, K$ , where  $K$  is the number of time periods for which the simulation is to be conducted. The model has been extensively validated with 2 years worth of measurements collected in a number of single-person offices. Validation is performed by comparing statistics of variables such as first arrival time, number of changes between occupied and unoccupied in a day, etc. predicted by the model with that extracted from sensor measurements. Tanimoto *et al.* (2008a–c) have developed an agent-based model for predicting demand profiles in residential households, and has performed extensive comparisons with measured demand data. Their model uses an available database of statistical information on people's behaviour profile in Japan. Such information is not available for commercial building in most countries. Even if such datasets were available, since occupancy dynamics may vary widely from one commercial building to another depending on the type and size of the building, location, climate, etc., relying on such data to generate occupant behaviour may not lead to accurate models. We believe a methodology is needed that takes information about occupants in a specific building to construct a model for that particular building.

The agent-based model we propose here is inspired by the model by Page *et al.* (2008). Extending the model in Page *et al.* (2008) to the case of multiple occupants is straightforward when the behaviour of occupants are independent. However, extending the model to multiple zones is much more challenging. For a building with  $n$  zones, an occupant can be in any one of  $n + 1$  states (the  $n + 1$ -th state corresponding to outside the building), so the transition probability matrix  $\mathbf{P}(k)$  becomes an  $(n + 1) \times (n + 1)$  matrix. Determining the entries of the transition probability matrix, even for a specific  $k$ , is not trivial. The problem gets worse when one has to specify a sequence of matrices  $\mathbf{P}(k)$ ,  $k = 1, 2, \dots$ . Even in the case with a single zone, specifying the  $2 \times 2$  transition probability matrices is not straightforward. The authors of (Page *et al.* 2008) computed the four entries of the matrix in terms of a time-varying “profile of probability of presence” and “parameter of mobility” that the user had to specify. This parameter of mobility had to be changed from its nominal constant values at certain times, so that the constraint that probability cannot be negative or larger than one is maintained at all times.

In this article, we propose a stochastic agent-based model that is easily scalable to arbitrary number of zones and arbitrary number of individuals, or *agents*. The proposed model, named *Multiple Modules* (MuMo) model, decides the location of an agent at a

given time through a set of rules specified by a number of modules. The modules are designed to maintain a Markov-like property of the agent dynamics so that the location of an agent at a given time depends on its location in the previous time. The MuMo model is thus inspired by that in Page *et al.* (2008), which we denote by “Page model” in the sequel.

The MuMo model is intended to be used for predicting occupancy in non-emergency situations. Simulation of this model produces a time-series of each occupants' location, which can then be collected to generate time-series of zone-level occupancy. Since the model is stochastic, each simulation will produce non-identical time-series. The information needed to specify the inputs to the model can be obtained from survey of occupants, measured sensor data or a combination of the two. The predictions of the model have been compared against measured occupancy data in commercial buildings for three distinct scenarios: single-occupant single-zone, multi-occupant single-zone and multi-occupant multi-zone. The data used for verification in the single-occupant single-zone scenario are the same as that used in Page *et al.* (2008), and was provided to us by the authors of that paper. The prediction accuracy of the MuMo model in the single-occupant single-zone case was found to be quite good (measured by how well it reproduces distributions of variables such as first arrival time), and comparable to that of the by Page model. Measurements for the multi-occupant single-zone and multi-occupant multi-zone scenarios were obtained from video data gathered in a building in the University of Florida campus for a number of months. The proposed model's prediction compares quite well with that from measured data in the multi-occupant single-zone scenario, but is poorer in the multi-occupant multi-zone scenario.

A weakness of agent-based models is their high-degree of complexity that makes them unsuitable for certain applications such as real-time estimation. The second contribution of this article is a low-complexity model of occupancy in a multi-zone building based on the so-called covariance graphical model framework. The graphical model captures information on mean occupancy and correlation among occupancy at various zones in the building in the form of the covariance matrix. For instance, the occupancy in a number of offices during lunchtime is likely to be inversely correlated to the occupancy in a cafeteria located inside the building. A graphical model can be used to represent such spatial correlations by non-zero entries in the covariance matrix. The zones of the building corresponds to the nodes of a graph and the sparsity pattern of the covariance matrix determine correlations among the nodes. Due to the simplicity of



the graphical model, it is suitable for real-time occupancy estimation (Liao and Barooah 2010). Multiple time-series data of occupancy are needed to construct a graphical model. Such time-series data can be obtained from an agent-based model, or in some cases from sensor measurements. Graphical models have been widely used in a variety of disciplines such as spatial statistics, image analysis and bioinformatics (Lauritzen 1996, Rue and Held 2005). Comparison of predictions from the graphical model and the agent-based model shows that it does reproduce the mean and variance of occupancy predicted by the agent-based model.

For the proposed agent-based model to predict occupancy patterns accurately for a given building and a set of occupants, some amount of calibration of the model is necessary. Information of the agents gathered from a survey of the occupants is usually not completely accurate. A survey may miss a few occupants, or certain occupants may provide inaccurate information in the questionnaire, whether unintentionally or intentionally. Certain parameters of the model therefore needs to be calibrated based on measurements of occupancy in a few locations in the building. This requires the building to have such sensors in place. It should be noted that surveying of occupants for model construction, and model calibration based on occupancy measurements, can only be carried out for existing buildings. If one wants to conduct simulations for a building that is in the design phase, one will have to specify the number of agents and specify their behaviours. This can potentially be done by picking agents from a database of agents and associated behaviour patterns, which is then used to specify the behaviour of the agent when she is placed in the fictitious building. The question of how to specify building-specific behaviour of an agent from limited information about her general behaviour patterns, is a topic of additional research and is not addressed here. The work of Tanimoto *et al.* (2008a–c) in which limited statistical information on individuals are used to construct detailed behaviour profiles may be useful in this direction. At this juncture, we expect that the proposed MuMo model and its reduced-order counterpart will be more immediately useful for building operation rather than building design, such as building (re-) commissioning, real-time predictive control of HVAC equipment, occupancy estimation, etc.

The rest of the article is organized as follows. Section 2 describes the proposed agent-based model. Section 3 describes the calibration procedure and the verification of the model based on comparison with sensor measurements. Section 4 describes the covariance graph models of occupancy evolution and identification of graphical models from occupancy

time-traces. This section also presents comparison of the graphical model with the agent-based MuMo model. The article concludes with a discussion in Section 5.

## 2. Agent-based model of building occupants

Consider a building with  $n$  zones that is occupied by  $m$  individuals, called *agents*. Time is measured by a discrete time index  $k$ , and the maximum time index for which a simulation is conducted is denoted by  $K$ , so that  $k = 1, \dots, K$ . Each increment in the time counter corresponds to an equal interval of time, i.e. a *sampling period*, that is denoted by  $T$  minutes. The agents are indexed as  $i = 1, \dots, m$ . An  $n$ -zone building has  $n + 1$  nodes that are indexed as  $j = 1, \dots, n, n + 1$  ( $n + 1$ -th node refers to the outside of the building). A zone can be a room, a hallway, or any location inside the building, while a node can also cover the outside.<sup>1</sup> The state  $z_i(k)$  of agent  $i$  at time index  $k$  refers to the node that the agent occupies during  $k$ , so that  $z_i(k) \in \{1, \dots, n + 1\}$  for every  $i$  and  $k$ . Note that the time index  $k$  is really a discrete representation of the time interval  $[(k-1)T, kT]$ , and the state  $z_i(k)$  is an aggregate measure of the agent  $i$ 's location during that time interval. The occupancy  $x_j(k)$  of a node  $j$  is the number of agents that are in  $j$  at time  $k$ , i.e.  $x_j(k)$  is the number of entries of the set  $\{i | z_i(k) = j\}$ . The occupancy of a building with  $n$  zones is  $x(k) := \sum_{j=1}^n x_j(k)$ .

The proposed agent-based model, named *Multiple Modules (MuMo) Model*, consists of a number of modules that together determine the state of an agent at every time. The state of an agent is initialized by the first module, and each module after the first modifies the state determined by the previous module. The output of the  $\ell$ th module is denoted by  $z_i^{(\ell)}(k)$ , and the output of the last module is the state of the agent at that time. If the  $\ell$ th module is not applicable, then  $z_i^{(\ell)}(k) \leftarrow z_i^{(\ell-1)}(k)$ . We now describe these modules.

### 2.1. Description of the MuMo model

The model consists of the following modules that govern the behaviour of each agent:

- (1) *Preliminary state generator module*: An agent-specific *nominal presence probability profile*  $\{P_i(k), k = 1, \dots, K\}$  is specified as input to this module for every agent  $i$ , where  $P_i(k) = [P_{i,1}(k), \dots, P_{i,n+1}(k)]^T$  and  $P_{i,j}(k)$  is an approximation of  $\Pr(z_i(k) = j)$ , the probability that agent  $i$  occupies node  $j$  at time  $k$  ( $\Pr(\cdot)$  denotes probability). During simulation,  $z_i^{(0)}(k)$ , i.e. the initial guess for the  $i$ -th agent's

state at time  $k$ , is generated using a pseudo-random number generator so that its probability mass functions (pmfs) matches the nominal presence probability profile, i.e. so that  $\Pr(z_i^{(0)}(k) = j) = P_{i,j}(k)$ .

- (2) *Acceleration and damping modules*: Each agent has an associated *primary zone* that corresponds to the zone in which the agent spends most of her time while inside the building. A person who is in a hallway or a restroom tends to leave that location quickly, while a person in her primary location (or outside the building) tends to stay there for long periods of time. An acceleration module and a damping module are used to mimic this behaviour by utilizing *transition probability parameters*  $p_a$  and  $p_d$ . The implementation of acceleration module is as follows. Suppose  $z_i(k-1) = j_a$ , where  $j_a$  is any node, such as a hallway and restroom, from which agents tend to transition (accelerate) out soon. If the tentative state at the current time is the same as that in the previous time, i.e. if  $z_i^{(0)}(k) = z_i(k-1)$ , then the state of agent  $i$  is recomputed with probability  $p_a$  by running preliminary state generator module again, and this new value is assigned to  $z_i^{(1)}(k)$ . Since the probability of staying in the node  $j_a$  is usually small, this forces the agent to transition out of that node soon. The damping module is defined as: if  $z_i^{(0)}(k) \neq z_i(k-1)$  and  $z_i(k-1)$  is either a primary zone or the outside node, then  $z_i^{(1)}(k) \leftarrow z_i(k-1)$  with probability  $1-p_d$ . The primary zones of the agents as well as the parameters  $p_a$  and  $p_d$  are specified as inputs to the model.
- (3) *Scheduled activity module*: This module takes care of any hard constraints on the individuals locations that may arise from scheduled activities, e.g. meetings of office workers, classes for students and teachers, etc. Specifically, if an agent  $i$  has to attend an activity located in node  $j$  during a particular time interval, the agent's state will be assigned to node  $j$  during that time. The output of this module is  $z_i^{(2)}(k)$ .
- (4) *Access module*: Each agent has an access profile associated with it that specifies which zones the agent has access to. The access module ensures that agents do not occupy zones that they do not have access to. If  $z_i^{(2)}(k) = j$  where  $j$  is a node that agent  $i$  does not have access to, then  $z_i^{(3)}(k) \leftarrow z_i(k-1)$ . This module is also invoked for zones that have a maximum occupancy limit, such as classrooms. The occupancy of those zones are constantly tracked during simulation. If the occupancy of such a zone is

found to reach the maximum allowed value, the state of any agent that was assigned to this room thereafter is assigned back to its previous state  $z_i(k-1)$ . The access profiles have to be specified as inputs.

The state of agent  $i$  at time  $k$  in the MuMo model is the output of the last module, i.e.  $z_i(k) \leftarrow z_i^{(3)}(k)$ .

For the sake of concreteness, we set the initial condition  $z_i(0) = n+1$  for every  $i$ . The model determines the states starting from time  $k=1$ . Note that although the state  $z_i(k)$  at time  $k$  is generated according to  $i$ 's nominal presence probability profile, it does depend on its previous state  $z_i(k-1)$  due to the effect of the acceleration and damping modules. The damping module is a key element of the model. People's movement inside a building is not independent over time, since a person's location at time  $k+1$  does depend on where he was at time  $k$  to a large extent. For instance, during regular working hours except early morning or evening, if a person is in his office at a particular time, he is likely to remain there with high probability in the next time instant. An appropriate stochastic model to capture this behaviour is a Markov chain. Although we did not specify a Markov chain model due to the difficulty in identifying the parameters of such a model (see the discussion in Section 1), the damping module endows the agents with a Markov-like property. In this regard, the proposed model is similar in spirit to the model in (Page *et al.* 2008).

We distinguish between two kinds of agents, *primary agents* and *secondary agents*. A primary agent is an individual who occupies a building on a regular basis for most of the working hours, while a secondary agent occupies the building for brief periods of time and does not do so regularly. Examples of secondary agents are students attending lectures and customers visiting a place of business. The behaviour of a secondary agent is governed by the same modules that govern primary agents. However, the nominal presence probability profile for a secondary agent is usually simpler than that of a primary agent. For instance, secondary agents used to simulate students who come into a building to attend a lecture in a particular room during a specific time may be assigned zero probability of being inside the building at all other times, and during that time period they have zero probability of occupying any zone other than the lecture hall or the hallways/restrooms. The transition probability parameters  $p_a$  and  $p_d$  of the primary agents may be different from those of the secondary agents.

Although the MuMo model is designed to simulate behaviour of an arbitrary number of individuals in a

building of arbitrary size, it is convenient to consider three distinct scenarios separately that together cover all possible cases: (i) single-occupant single-zone (SOSZ), (ii) multi-occupant single-zone (MOSZ) and (iii) multi-occupant multi-zone (MOMZ). The model is considerably simpler in the SOSZ scenario compared to the other two. Since there is only one agent and only two possible nodes (in and out, denoted by 1 and 2) in the SOSZ scenario, the nominal presence probability profile requires specifying only two number at each time index:  $P_{1,1}(k)$  and  $P_{1,2}(k)$  for  $k = 1, \dots, K$ . In both SOSZ and MOSZ scenarios in which the building consists of a single zone, only  $p_d$  needs to be specified since the accelerating module that uses  $p_a$  is not applicable. In addition, the access module is only applicable to the MOMZ scenario.

**Remark 1:** In this article, we choose the sampling period to be 15 min (i.e.  $T = 15$ ). One reason for the choice is convenience of comparison with the Page model and use of the data reported in (Page *et al.* 2008), which used a sampling period of 15 min. In addition, since our intended use of the model is assistance in building energy simulation and control, which have slow dynamics, a time period shorter than 15 min seems unnecessary.

In this study, we model the motion of agents for 1 week, assuming that their behaviour does not change in a statistically significant way from week to week. For the same reason, we did not incorporate a long-absence module in the MuMo model, which will simulate periods of long absence due to vacations, sickness, etc. In principle, long-absences can be incorporated into the model in the same way it is done in the Page model. However, we feel that incorporating a long-absence module is appropriate only if the model is meant to simulate occupants' behaviour over very long periods of time, say, of the order of several months and years. Although the MuMo model is capable of such a simulation, in that case slow variations in behaviour that occur during long time periods will have to be incorporated in specifying the nominal presence probability profiles. However, the data collected for verification in the present study for scenarios involving multiple occupants, which will be presented in Section 3.2, are not suitable for comparison involving very long time durations. The issues involved in simulating behaviour over long periods of time is a topic for further investigation. We therefore did not incorporate a long-absence module in the MuMo model in this study.

## 2.2. MuMo model construction

Constructing a MuMo model with  $m$  agents requires specifying for each agent its nominal presence

probability profile, agents' schedules and access profiles. In addition, damping and acceleration parameters  $p_d$  and  $p_a$  (one pair of values for primary agents and another pair for secondary agents), and maximum occupancy limits of rooms in the building, if any, need to be specified. For a single occupant in a single room, most of this information can be collected by deploying sensors, as was done by Page *et al.* (2008). However, this becomes challenging in the case of multiple occupants, since sensors have to track each individual over time. Conducting a survey of the occupants' behaviour by asking them to fill out a questionnaire is a more feasible – albeit less accurate – way of collecting this information. Due to the inaccuracy inherent in a survey, a model constructed from survey data may require more effort in calibration than one from sensor measurements.

The most time-consuming part in constructing the model is specifying the nominal presence probability profile for each agent. Figure 1 describes the algorithm we used to compute the nominal presence probability profile. The algorithm needs a number of parameters as input, which are described in Table 1. In this study, these parameters are determined from a survey of the occupants. It should be noted that for a building with a distinct usage pattern and occupant demographic, the information needed to construct the nominal presence probability profiles, and therefore the kind of survey to be conducted, may be different from what was done in this study. The parameters such as arrival time of an agent that have units of time are gathered in “minutes”. They are then converted to units of “time index” by scaling with the sampling interval, since the algorithm described in Figure 1 requires these parameters in the unit of time index. In addition, the time index  $k$  in Figure 1 spans only the duration of 1 day,  $1 \leq k \leq 24 \times 60/T$ , while the days of the week are indexed by  $d$ , where  $d = 1, \dots, 7$ , with 1 denoting Monday. This is done for ease of description of the algorithm, especially to highlight the day-to-day variation in the nominal presence probability profiles. During each day,  $k = 1$  corresponds to the time interval from midnight (0: 00 h) to 15 min past midnight (00: 15 h). For scenarios involving a single zone, information on the secondary zone and parameters related to the secondary zone (see Table 1) need not be gathered. After the nominal presence probability profiles are determined, we have to specify the parameters  $p_a, p_d$ . These are initially assigned to  $p_a = 1$  and  $p_d = 0.5$ , and may be changed during model calibration. Information on scheduled activity is also obtained from the survey of the occupants, which serves as the input to the schedule module of the MuMo model. At this point the model is completely specified.

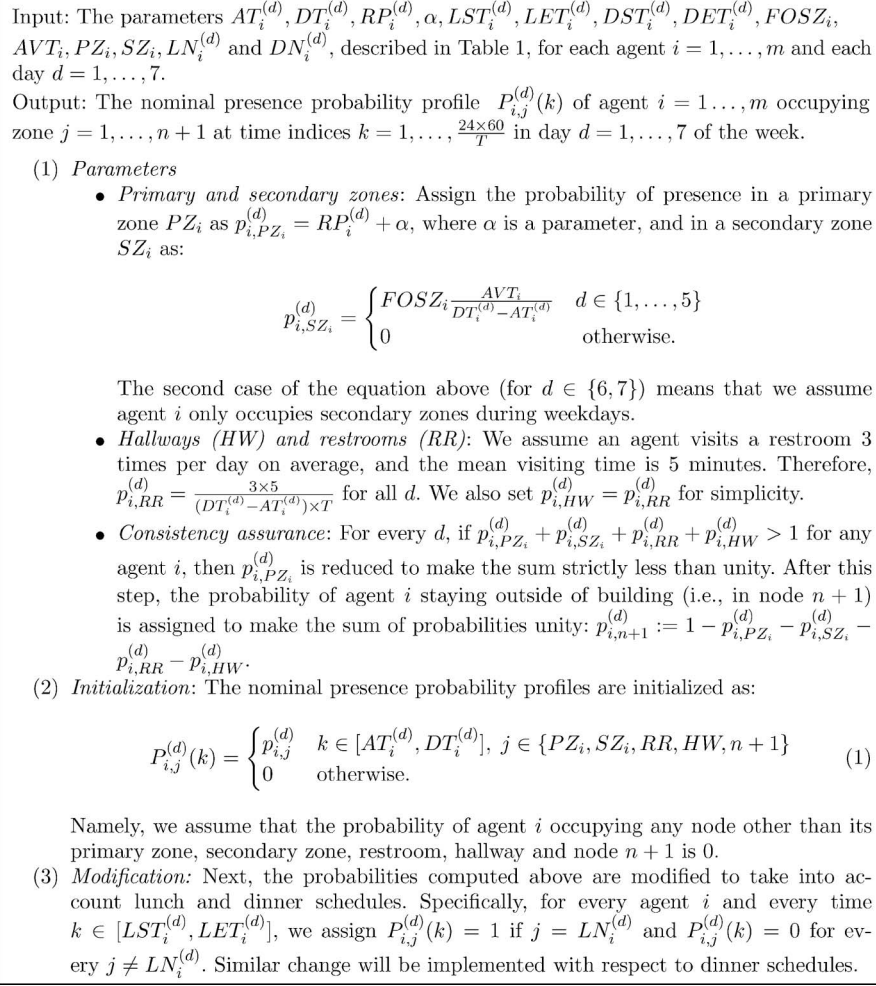


Figure 1. Procedure for computing the nominal presence probability profiles for a one-week long period from information obtained from a survey.

### 3. Model calibration and verification

Since the parameters that have to be specified in the model may be difficult to determine accurately – especially when the model is constructed from survey data – these parameters may need to be calibrated. Calibration is performed by comparing parameters and distributions of certain zone-level or building-level random variables predicted by the model with that estimated from measurements.

Model verification is also conducted similarly: by comparing the statistics of these variables predicted by the model with that estimated from measurements. Since the model is designed to produce realizations of a stochastic process, comparing individual time traces generated by the model with measurements is not appropriate. The data used in calibration are distinct from that used in verification. The parameters and variables mentioned above are the following:

- (1) *Mean occupancy of zone/building*: This is the (ensemble) average value of occupancy at each time  $k$ . Specifically, the mean occupancy of zone  $j$  at  $k$  is defined as  $\mathbf{E}[x_j(k)]$ , where  $\mathbf{E}[\cdot]$  denotes expectation.
- (2) *First arrival time (in a day)*: The time when the zone or building gets occupied for the first time during a day. More precisely, assuming the time index starts at 1 when day  $d$  starts, if  $x_j(k) \geq \theta_{\text{empty}}$  and  $x_j(\ell) < \theta_{\text{empty}}$  for all  $\ell < k$ , where  $\theta_{\text{empty}} > 0$  is an appropriately chosen parameter, then  $k$  is the first arrival time of zone  $j$  in day  $d$ .
- (3) *Last departure time (in a day)*: The last time during a day at which the zone or building becomes unoccupied. More precisely, limiting the time index  $k$  to those values it takes within a specific day  $d$ , if  $x_j(k) \geq \theta_{\text{empty}}$  and  $x_j(\ell) <$



Table 1. Parameters to be determined from a questionnaire-based survey of the building occupants for constructing the MuMo model.

Content	Symbol	Remark
Primary zone	$PZ_i$	The zone in the building in which agent $i$ spends most of its time, e.g. the room in which $i$ 's office is located.
Secondary zone	$SZ_i$	The zone in the building in which agent $i$ spends most of its time, after the primary zone. For example, a cafeteria in an office building or a meeting room.
Arrival/departure time	$AT_i^{(d)}, DT_i^{(d)}$	The time when agent $i$ arrives at the building/ leaves the building, in day $d$ of the week, where $d = 1, \dots, 7$ .
Lunch/dinner schedule	$LST_i^{(d)}, LET_i^{(d)} / DST_i^{(d)}, DET_i^{(d)}$	The starting time and ending time of lunch/dinner of agent $i$ is denoted as $LST_i^{(d)}, LET_i^{(d)} / DST_i^{(d)}, DET_i^{(d)}$ respectively. We only consider the lunch/dinner schedule in time range between $AT_i^{(d)}$ and $DT_i^{(d)}$ .
Lunch/dinner nodes	$LN_i^{(d)}, DN_i^{(d)}$	$LN_i^{(d)}$ and $DN_i^{(d)}$ represent the nodes (location) that agent $i$ eats lunch and dinner, which can be either the outside node $n + 1$ or any zone in the building.
Ratio of presence	$RP_i^{(d)}$	Ratio of the duration that agent $i$ occupies $PZ_i$ in the $d$ -th day of the week to the duration between its arrival and departure times for that day.
Frequency of occupying secondary zone	$FOSZ_i$	The average number of times that agent $i$ occupies $SZ_i$ per day.
Average visit duration of secondary zone	$AVT_i$	The average duration that agent $i$ spends in its secondary zone when it visits that zone

$\theta_{\text{empty}}$  for all  $\ell > k$  then  $k + 1$  is the last departure time of zone  $j$  in day  $d$ .

- (4) *Cumulative occupied duration (in a day)*: The total length of time in a day during which the occupancy in a zone or building is above a threshold  $\theta_{\text{occ}}$ , not necessarily continuously. More precisely, assuming the time index starts at 1 when day  $d$  starts, the cumulative occupied duration during that day is the number of elements of the set  $\{k | x_j(k) \geq \theta_{\text{occ}}, 1 \leq k \leq 24 \times 60/T\}$ .
- (5) *Continuously occupied duration*: The length of a maximal time interval during which the occupancy in a zone or building is above a threshold  $\theta_{\text{occ}}$  continuously. To define it precisely, we say an interval  $S_{k_1, k_2} := [k_1, k_2]$  is a maximal occupied interval if  $x_j(k) \geq \theta_{\text{occ}}$  for  $k_1 \leq k \leq k_2$  and  $x_j(k_1 - 1) < \theta_{\text{occ}}, x_j(k_2 + 1) < \theta_{\text{occ}}$ . The lengths of all maximal occupied intervals are realizations of the random variable continuously occupied duration.
- (6) *Number of occupied/unoccupied transitions (in a day)*: The number of transitions between ‘‘occupied’’ and ‘‘unoccupied’’ status in a day for the zone or building. Assuming that the time index starts at 1 when day  $d$  starts, the number of elements of the set  $\{k | x_j(k) \geq \theta_{\text{empty}}, x_j(k + 1) < \theta_{\text{empty}}\} \cup \{k | x_j(k) < \theta_{\text{empty}}, x_j(k + 1) \geq \theta_{\text{empty}}\}$ , for  $1 \leq k < 24 \times 60/T$  is the number of occupied/unoccupied transitions of zone  $j$  in day  $d$ .

Monte-Carlo simulations of the model are conducted, and the resulting multiple time-series (each 1-week long) are used to estimate the pmfs and cumulative distribution functions (CDFs) of these random variables. The pmfs and CDFs of these variables are also estimated from the repeated segments of 1-week long processed sensor data. All the variables that are defined in terms of a time index, such as first arrival time, are expressed in units of time (usually hours) when presented, for ease of visual interpretation. The variable ‘‘number of occupied/unoccupied transitions’’ is the same as ‘‘number of changes’’ in Page *et al.* (2008) in the single-occupant single-zone scenario.

For ease of quantitative comparison of the model's predictions and measurements, we use the normalized root mean square deviation (NRMSD) and Kullback-Leibler divergence. Let  $x(k)$  and  $y(k)$ ,  $k = 1, \dots, K$  be two time sequences, the NRMSD between  $x$  and  $y$  is defined as

$$\text{NRMSD}(x, y) = \frac{\|x - y\| / \sqrt{K}}{\max z - \min z}, \quad (2)$$

where  $x = [x(1), \dots, x(K)]^T$ ,  $y = [y(1), \dots, y(K)]^T$ ,  $z = [x^T, y^T]^T$  and  $\|\cdot\|$  is the Euclidean norm. To compare the predicted distribution of a random variable by the model with that estimated from measurements, we use the Kullback-Leibler (K-L) divergence. The K-L divergence is frequently used to compute distances between two densities  $p$  and  $q$ , and

is defined as (Cover and Thomas 1991)

$$d(p||q) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right), \quad (3)$$

with the standard convention that  $0 \log \frac{0}{q} = 0$  and  $q \log \frac{q}{0} = \infty$ . In this article, we used the convention  $q \log \frac{q}{0} = 0$ . Note that the K–L divergence is only a pseudo-metric since  $d(p||q) \neq d(q||p)$  in general. For a random variable,  $X$ ,  $p_X^{\text{MuMo}}$  and  $p_X^{\text{meas}}$  denote the pmfs of  $X$  predicted by the MuMo model and that estimated from measurements.

### 3.1. Model calibration procedure

Calibrating the parameters of the MuMo model becomes necessary when the information used in model construction may not be accurate. This is especially true in scenarios involving multiple occupants, since in these cases the proposed model is constructed from information obtained from occupants' response to a questionnaire, which is likely to suffer from large uncertainty. We choose a fraction of the measured occupancy data for calibration and call it the *training data*. The rest of the data, called *verification data*, are not used for calibration. For the SOSZ scenario, we used half the data (6 months) for calibration and the other half for validation. In both the MOSZ and MOMZ scenarios, calibration was performed with 2 weeks of data, while the validation data was of 10 weeks and 7 weeks duration, respectively. The choice of the length of the calibration data vis-a-vis the validation data was arbitrary. The statistics of the random variables described in the beginning of Section 3 are first estimated by using only the calibration data. Calibration is then performed by changing the agent based model – so that the difference between the model predictions and measurements, as measured by the values of NRMSD and K–L divergence, is small. To keep the calibration tractable, we modify only a few parameters. The calibration process, and the sequence in which it is performed, is described below.

- (1) *The arrival and departure times of the “early bird” and the “night owl”*: An “early bird” is an agent whose arrival time is smaller than or equal to that of the rest of the agents. A “night owl” is an agent whose departure time is greater than or equal to that of the rest. The first arrival and last departure time of a zone or a building are mostly dependent on these occupants. While these agents can be identified from the information that agents provide in the survey, the actual arrival and departure times

that they provide may not be accurate. We adjust the arrival time  $AT$  of early birds (departure time  $DT$  of night owls) so that K–L divergence  $d(p_X^{\text{meas}} || p_X^{\text{MuMo}})$  is small, where random variable  $X$  is the first arrival (last departure) time of the zone or building. Every time  $AT$  and/or  $DT$  of an agent is changed, the procedure described in Section 2.2 is executed to re-compute the nominal presence probability profiles of these agents. Although there may be more than one early bird (night owl), we only calibrate the arrival and departure times of one early bird (night owl) to perform model calibration in this study.

- (2) *The parameter  $\alpha$* : Since  $p_{i,PZi}^{(d)} = RP_i^{(d)} + \alpha$ , the parameter  $\alpha$  determines the probability of presence in the primary zone of all the agents. We modify  $\alpha$  so as to make NRMSD ( $x_{\text{MuMo}}, x_{\text{meas}}$ ) small, where  $x_{\text{MuMo}}$  and  $x_{\text{meas}}$  is the mean occupancy (zone or building) of the model's prediction and measured (training) data over 1 week. Every time  $\alpha$  is changed, the nominal presence probability profiles of all the agents are recomputed.
- (3) *The transition probability parameters  $p_d$  and  $p_a$* : The parameter  $p_d$  is calibrated so as to make  $d(p_X^{\text{meas}} || p_X^{\text{MuMo}})$  small, where random variable  $X$  is number of occupied/unoccupied transitions of the zone or building. Since the parameter  $p_a$  determines how fast an agent moves away from zones such as a hallway or a restroom in which people do not typically spend most time, the choice  $p_a$  depends on the sampling period  $T$ . If  $T \geq 15$ , we set  $p_a$  be 1 so that the agent “transitions out” of such a zone with high probability in the next time index. However, if  $T \leq 5$ ,  $p_a$  should be decreased to reduce the acceleration effect.

In the SOSZ scenario, the MuMo model can be constructed from either survey or sensor data. If a survey is used, the model calibration procedure is essentially the same as in the multiple occupants case. If the model is constructed from sensor data, steps (1) and (2) described above are unnecessary since the nominal presence probability profile generated from data is accurate. Step (3) is still needed.

### 3.2. Model verification

#### 3.2.1. Model verification: SOSZ (single-occupant single-zone) scenario

In this scenario, data reported in (Page *et al.* 2008) were used for model construction, calibration and

verification, which was provided to us by the authors of Page *et al.* (2008). The sensor measurements consist of series of change of occupancy states (occupied/unoccupied) with time stamps collected in a single occupant office for a duration of about 1 year. We processed the data and obtained the time-series of occupancy states in a 15-min sampling interval. We use half of the processed data to estimate the nominal presence probability profile, which is called the “probability of presence” in Page *et al.* (2008). The method of computing the nominal presence probability profile is the same as computing the “probability of presence” as described Page *et al.* (2008). The interested reader is referred to Page *et al.* (2008) for details on the data collection and processing. After calibration, the transition probability parameters  $p_d$  in the MuMo model is set to 0.5.

The Page model reported in Page *et al.* (2008) contains a model of long absences, which are defined as absence of more than 1 day not including weekend. In contrast, the proposed MuMo model does not; see Remark 1. In order to maintain consistency, we simulate a variant of the Page model, by removing the long absence part from the model. Similarly, we remove the periods of long absence from the sensor data obtained from the authors of (Page *et al.* 2008) before using it for model calibration and verification. The same data are used for construction and calibration of both the Page model and the MuMo model.

One thousand Monte-Carlo simulations (each of 1 week duration) are conducted with both the proposed model and the Page model. The distributions of variables (other than mean occupancy) described in

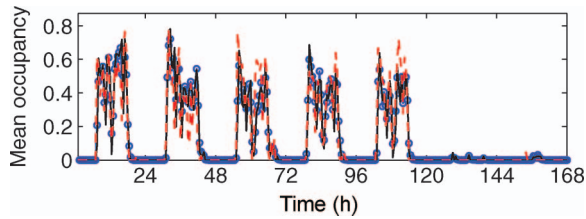


Figure 2. Mean occupancy in the SOSZ(single-occupant single-zone) scenario for one week: comparison between predictions of the Page model (black), the MuMo model (blue with circles), and measurements (dashed red).

the beginning of Section 3 are estimated from measured data and simulation time traces. Since the statistics of variables such as first arrival time in weekdays are going to differ from that in weekends, comparison in weekdays should be done separately from that in weekends. In the interest of space, we only present comparison for weekdays.

Figure 2 shows that both the MuMo model and Page model predict the mean occupancy well, and with similar degree of error. The NRMSD between the mean occupancy predicted by the MuMo model and measured values is 0.0995; for Page model is it slightly higher: 0.1371.

The distributions of variables such as first arrival time predicted by the two models (MuMo and Page) and from measurements are shown in Figure 2, while the K–L distances are shown in Table 2. Figure 3(a) and (b) suggests that both the models predict the distributions of first arrival time and last departure time well. While the proposed MuMo model captures the location of the peak of first arrival time better than the Page model, the latter captures the peak value of the pmf better. A similar observation holds for the last departure time. We see from Table 2 that the Page’s prediction of these two variables is more accurate according to the K–L divergence metric. None of the two models predicts the cumulative occupied duration very well, as seen from Figure 3(c), though their predictions are quite close to each other. This is also seen from the K–L divergences for this variable in Table 2. Figure 3(d) shows that the distributions of continuously occupied duration. The Page model slightly underestimates the frequency of short period of presence compared to the MuMo model, but the difference with the measured value is small. The distribution of the number of occupied/unoccupied transitions, which reflects the mobility of the agent, is shown in Figure 3(e). The MuMo model underpredicts small and large values of this variable, while predicting the intermediate values accurately. The error in the prediction of both the models seem to be similar, though they occur at different values of the variable. Based on this comparison, we conclude that the MuMo model has the same level of accuracy as that of the Page model. The K–L divergence values

Table 2. K–L divergence between  $p_X^{\text{MuMo}}$ ,  $p_X^{\text{Page}}$  and  $p_X^{\text{meas}}$  in the single-occupant single-zone scenario.

Variable( $X$ )	$d(p_X^{\text{meas}} \parallel p_X^{\text{MuMo}})$	$d(p_X^{\text{meas}} \parallel p_X^{\text{Page}})$	$d(p_X^{\text{MuMo}} \parallel p_X^{\text{Page}})$
First arrival time	0.4986	0.2081	0.1826
Last departure time	0.6388	0.2945	0.1903
Cumulative occupied duration	0.3215	0.3416	0.0201
Continuously occupied duration	0.0229	0.0986	0.0751
Number of occupied/unoccupied transitions	0.2421	0.8717	0.3954

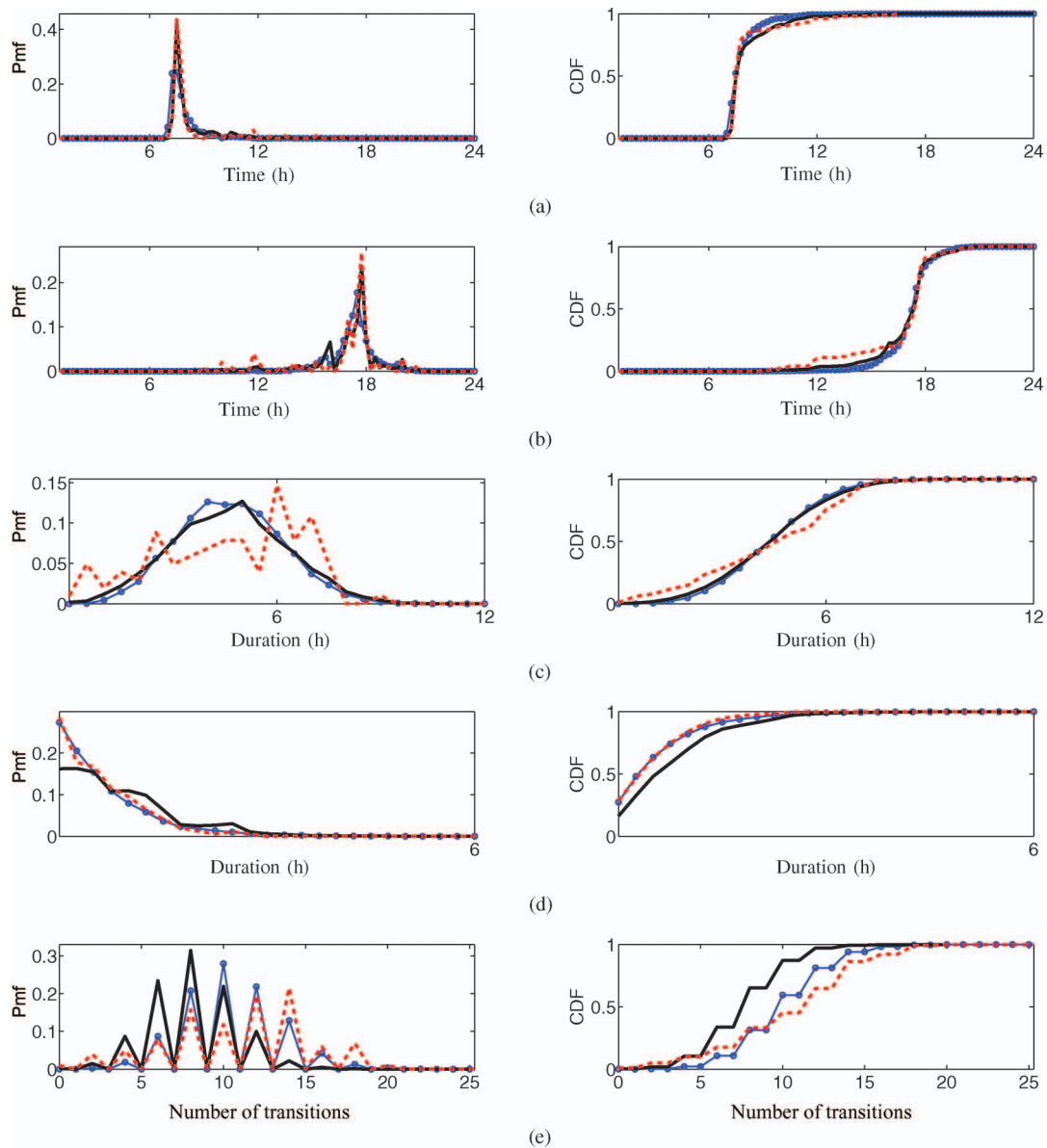


Figure 3. Comparison the statistics predicted by the MuMo model (blue with circles), Page model (black) and measured data (dashed red) in the SOSZ scenario. The times 0 and 24 h correspond to midnight. The bin-size in (c) is 30 min, while for (a),(b),(d), it is 15 min. (a) First arrival time. (b) Last departure time. (c) Cumulative occupied duration. (d) Continuously occupied duration. (e) Number of occupied/unoccupied transitions.

shown in Table 2 indicate that the MuMo model predicts the distribution of the continuously occupied duration the best and that of the last departure time the poorest. The two models seem to predict most of the occupancy statistics sufficiently well, except that of cumulative occupied duration.

### 3.2.2. Model verification: the MOSZ (multi-occupant single-zone) scenario

The MOSZ scenario studied in this article corresponds to a room in a building in the University of Florida

campus, shown as zone 15 in Figure 4. The room housed five graduate students who worked there regularly and three undergraduate research assistants who used it intermittently. The MuMo model constructed for this room had eight primary agents. The model also had seven secondary agents, who were used to simulate students who would occasionally visit the room to meet with a few of the graduate students who were teaching assistants. The MuMo model was constructed by conducting a survey of the occupants to determine the subset of the parameters listed in Table 1 that are relevant to the MOSZ scenario. All the



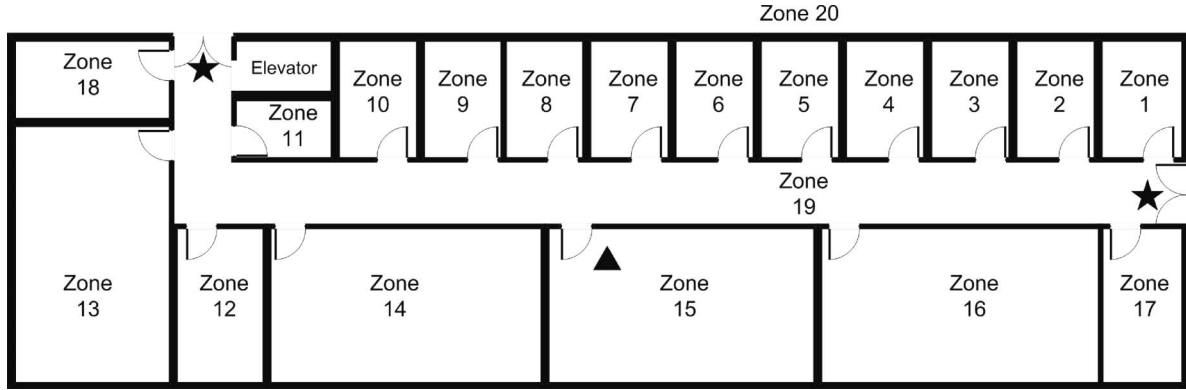


Figure 4. Floor plan of the 3rd floor of MAE-B building in the University of Florida campus, where the measurements used in the verification of the MuMo model for the MOSZ (multi-occupant single-zone) and MOMZ (multi-occupant multi-zone) scenarios were collected. Verification of the model in the MOSZ (multi-occupant single-zone) scenario was done using data collected from the camera (shown as a triangle) in Zone 15. Verification in the MOMZ (multi-occupant multi-zone) scenario was done with data collected with the cameras (shown as stars) near the entrances to the floor.

time-related answers in the survey were rounded off to have a 15 min time-resolution.

We collected occupancy data for this room by using a wireless video camera to monitor the entrance to the room. The data were collected for a period of about 4 months (during January–April 2010). The video camera captures grey-scale images at a rate of 30 frames per second. A motion detection algorithm was used to save only those frames when motion was detected. Each saved frame was time stamped. Tests were conducted by manually comparing the saved frames with fully recorded video, which established that the motion detection algorithm was capable of detecting anyone moving in the field of view of the camera under all lighting conditions. The saved frames were analysed manually to count the flow rate of occupants (number of people coming into and exiting the room). Manual counting was performed to ensure that measurements obtained were of high accuracy. Room occupancy at a particular time was then calculated by computing the cumulative sum of the flow rate measurements, and then adding it to the room's initial occupancy. Since the data were collected with a high frame-rate video camera, the occupancy time-series so obtained had a time resolution of seconds. This raw time-series data was converted to one with a sampling interval of 15 min using a weighted average. For instance, if the occupancy value is 1 from 1 to 10 min and 2 from 11 to 15 min, then the value of occupancy in the time step that corresponds to this 15-min interval is computed as  $1 \times 10/15 + 2 \times 5/15 = 1.33$ . Although data were collected for 16 weeks, due to technical problems with the video cameras we had measurements available for 12 Mondays, 10 Tuesdays, 7 Wednesdays, 10 Thursdays, 11 Fridays, 11 Saturdays and 9 Sundays. Of

these, about 2 weeks of data were used for calibration and the rest for verification. We do not take into account the janitorial staff in the model construction because (s)he usually occupies the room for less than 1 min in the time range from 5:30 am to 7:30 am. To maintain consistency, we also ignore the janitorial staff in counting occupancy from the sensor (video) measurements. It is important to do so since otherwise the statistics of first arrival time as measured by the video sensor will be significantly affected by the janitorial staff while the model will not be able to predict that.

As in the previous scenario, 1000 Monte-Carlo simulations (each of 1 week duration) are conducted with both the proposed model and the Page model. The statistics of variables other than the mean occupancy, which are described in the beginning of Section 3, are estimated from data (measurements and simulation time traces). Comparison between the two is provided only for weekdays, for the reasons described in Section 3.2.1. The thresholds used in computing first arrival times, etc. are:  $\theta_{\text{empty}} = \theta_{\text{occ}} = 0.5$ . Calibration of the model according to the procedure described in Section 3.1 led to a change of the transition probability parameter  $p_d$  to 0.8 for primary agents (instead of the default value 0.5). The transition probability parameter for the secondary occupants is kept at  $p_d = 0.5$ . The nominal presence probability profiles of one early bird and one night owl were also adjusted during calibration. These two agents were identified from the survey of the occupants of the room.

Mean occupancy at each time was computed by averaging over all the measurements available for that time. Figure 5(a) compares the mean occupancy predicted by the proposed model with that computed

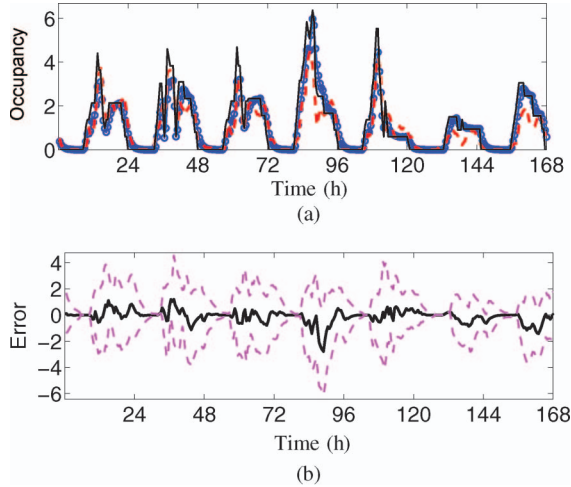


Figure 5. (a) Mean occupancy estimate in a MOSZ scenario (for zone 15 in Figure 4) from three sources: measurements (dashed red), MuMo model prediction (dotted blue) and survey (black). (b) shows the mean (black) and  $\pm 3 \sigma$  (dashed magenta) of the error between the model prediction and the measured values.

from measurements, and that computed from survey. The mean occupancy estimated from survey is simply the sum of the probability of each agent being “inside”, where these probabilities are determined from the nominal presence probability profiles. To see why it is so, let  $\mathcal{I}_A(\cdot)$  be the indicator function of the set  $A$ , so that  $\mathcal{I}_A(x) = 1$  if  $x \in A$  and 0 otherwise. Now,

$$x(k) = \sum_i \mathcal{I}_{\{1, \dots, n\}}(z_i(k)).$$

Since  $\mathbf{E}[\mathcal{I}_{\{1, \dots, n\}}(z_i(k))] = \Pr(z_i(k) \neq n + 1)$ , we have that the mean occupancy is  $\mathbf{E}[x(k)] = \sum_i \Pr(z_i(k) \neq n + 1)$ . The error in the model’s prediction of mean occupancy is NRMSD  $(x_{\text{MuMo}}, x_{\text{meas}}) = 0.0877$ , while that from the survey is NRMSD  $(x_{\text{survey}}, x_{\text{meas}}) = 0.1202$ , where  $x_{\text{MuMo}}$ ,  $x_{\text{meas}}$  and  $x_{\text{survey}}$  are the time-series of mean occupancy of the zone over 1 week computed from the model’s prediction, measurements, and survey. The error in mean occupancy predication, expressed as a fraction of the mean occupancy, is largest during the weekends. We believe the reason is that since the occupants have a greater variability in using the building during the weekends, they are not able to provide accurate description of their own behaviour in the survey.

Figure 6 compares the distributions of random variables predicted by the model and empirically estimated from measured data. Table 3 provides the K–L divergences between the predicted and measured distributions. We see from Figure 6(a) that the shape of the distribution of first arrival time is predicted

correctly, but there are a few late first arrivals around noon that the model does not capture. Similarly, Figure 6(b) shows that the overall trend of last departure time is predicted correctly by the model, though it does not capture all the peaks in the pmf. There is a small peak in the measured pmf at around 5 pm that correspond to occupants leaving the room in the evening that the model does not predict. This may be due to the night owl occasionally leaving earlier than usual. The distributions of cumulative occupied duration are shown in Figure 6(c) and those of continuously occupied duration are shown in Figure 6(d). There are several peaks in the measured pmf of the cumulative occupied duration that the model does not predict, though the overall trend of the distribution is predicted correctly. In case of the continuously occupied duration, the model does accurately predict that multi-modal nature of the distribution, but its prediction of the values of the probabilities is less accurate. Figure 6(e) shows the distribution of the number of occupied/unoccupied transitions in a day. The MuMo model predicts the distribution quite well, especially the probabilities of the number of transitions greater than 5.

We see from Table 3 that among all the variables, the model’s prediction of the distribution of cumulative occupied duration is the poorest, and of the number of occupied/unoccupied transitions is the best. In addition, it is seen from Figure 6 that, in general, the model’s predictions of distributions are smoother compared to the measured ones. We believe part of the reason for the difference between prediction and measured values, as well as of the non-smoothness of the measured pmfs, is the limited amount of verification data. Specifically, there are only 250 samples of the variable cumulative occupied duration that the measured distributions are estimated from, since measurements from each weekday leads to only one sample. In contrast, the pmfs from the model are estimated from 5000 samples. Because of this, the estimates from the measured data may have larger error. For further evidence in support of the hypothesis, notice that the model predicts the multiple peaks in the pmf of continuously occupied duration shown in Figure 6(d), but not that in several other variables. For a given time interval in which sensor measurements are collected, there are many samples of continuously occupied duration in that time-series while there is only one sample of cumulative occupied duration in a time-series of 1 day. So, for the same duration of data collection, we obtain more samples of the variable continuously occupied duration than others, which is likely to make the estimate of the pmf of continuously occupied duration more accurate than that of the others.

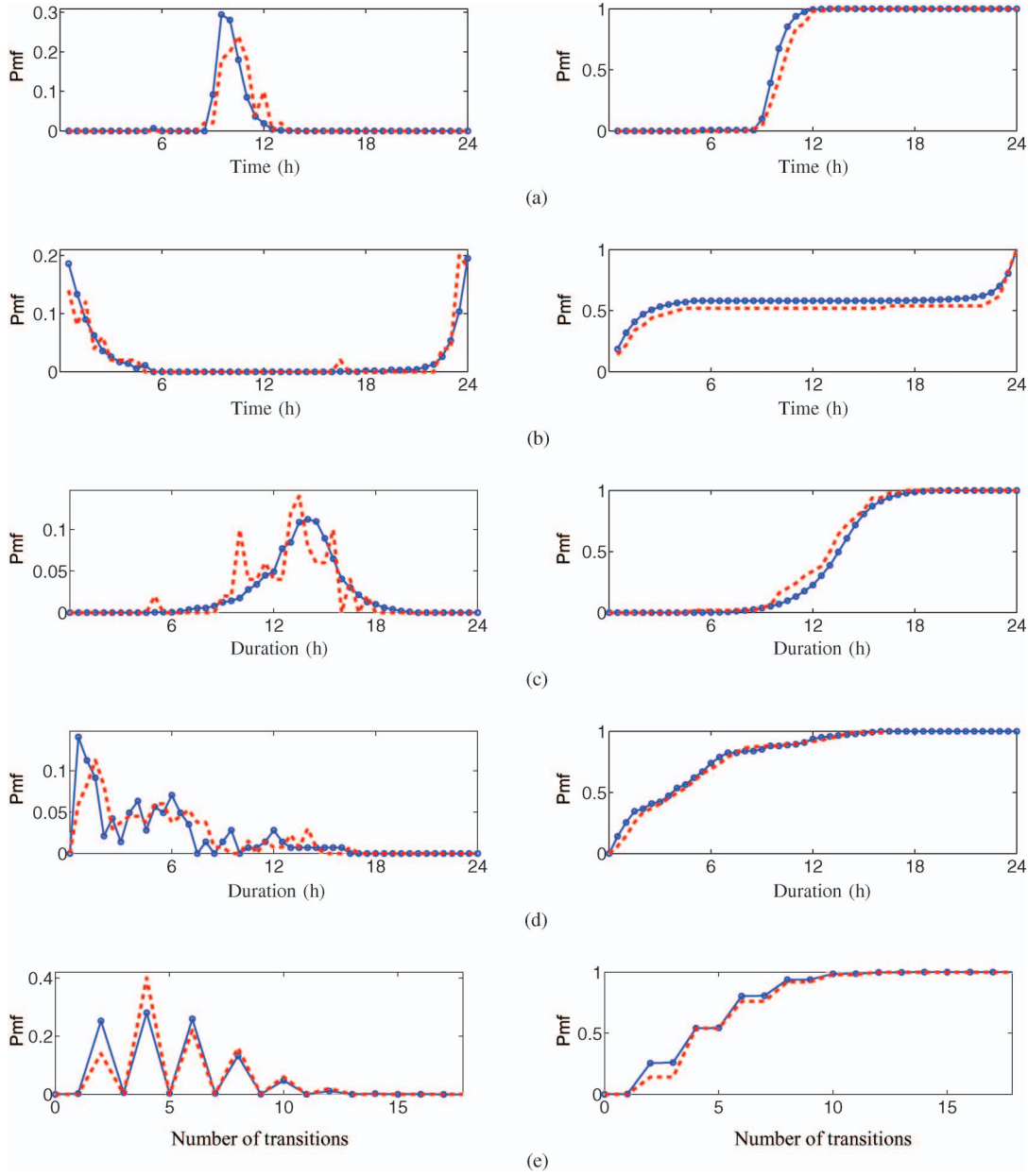


Figure 6. Verification of the MuMo model in the MOSZ scenario: comparison between the predictions (blue with circles) and measurements (dashed red). The distributions are estimated from 1000 Monte-Carlo simulations of the model and from about 12 weeks of measurements. Comparison is for weekdays only, with binsize = 1/2 h. (a) First arrival time. (b) Last departure time. (c) Cumulative occupied duration. (d) Continuously occupied duration. (e) Number of occupied/unoccupied transitions.

Table 3. K–L divergence between  $p_X^{\text{MuMo}}$  and  $p_X^{\text{meas}}$  in the multi-occupant single-zone scenario.

Variable( $X$ )	$d(p_X^{\text{meas}} \  p_X^{\text{MuMo}})$
First arrival time	0.2388
Last departure time	0.1900
Cumulative occupied duration	0.3261
Continuously occupied duration	0.1881
Number of occupied/unoccupied transitions	0.0810

The reader will notice that the pmfs in Figure 6(a)–(e) are with “binsize = 1/2 h”, which means the data have been aggregated to a time-resolution of 30 min (interval covered by two time indices) in computing the probabilities. The reason for doing so is again the limited number of measured data, which is not enough to cover all the bins in a reasonable range of values with a 15-min resolution. For example, there is no conceivable reason for the value of  $\{\text{Pr}(\text{Cumulative occupied duration}) = 12 \text{ hr } 30 \text{ min}\}$

to be significantly different from  $\{\text{Pr}(\text{Cumulative occupied duration}) = 12 \text{ hr } 15 \text{ min}\}$ , while this is what the measurements seem to say if a time-resolution smaller than 30 min is used. A more conclusive comparison will require measurements collected for a longer duration.

Overall, we see from the comparison that the proposed model does predict the distributions of several key occupancy-related variables with limited error. In fact, the comparison with measured values is better in this scenario than in the single-occupant single-zone scenario.

### 3.2.3. Model verification: the multi-occupant multi-zone (MOMZ) scenario

The MOMZ scenario studied here corresponds to the third floor of the MAE-B building in the University of Florida campus (see Figure 4). For the remainder of this article, we will refer to the third floor of the MAE-B building as the “building”. About 51 people (faculty, staff, graduate and undergraduate research assistants and visitors) used the building at the time of survey. The model includes 40 primary agents and 11 secondary agents. Measurements were collected by high resolution video cameras mounted on the walls. The cameras have auto back light compensation to provide stable video quality under all lighting conditions. A motion detection algorithm of the kind described in Section 3.2.2 is used for saving frames only when motion was detected. Each of the two entrances (exits) into (from) the building was monitored by one camera. Net flow rate of occupants into the building was obtained by adding the net flow rate across each of these camera’s field of view. The flow rate into the zone 14 in Figure 4 was also measured from the camera images. As a result, we obtained from these cameras the total occupancy of the building and occupancy in zone 14 (a room with multiple occupants). Measurements presented in this study were collected during a period of about 9 weeks during May–July, 2010.

Data for model construction were collected by a survey of the occupants of the building. Instead of asking questions to each occupant of a room in multi-occupant rooms, we randomly picked one or two occupants and asked them about the nominal behaviour of their colleagues. Only rough estimates were sought. For instance, we divided the range of arrival time into 1-h resolution, namely, “before 8am”, “8–9 am”, “9–10 am” or “after 12 pm”, and the surveyed individuals were asked merely to provide the number of people that fell into each of those ranges. We also collected information on schedules of the faculty members whose offices are in this building.

Furthermore, all professors’ course schedules were collected from the registrar’s office and/or professors’ websites. Note that since this survey did not distinguish between the occupants in a room, we specified the agents by randomly combining the answers provided by the occupants. For instance, we assigned the arrival times of all agents in a room such that the number of people in each time range was consistent with the information collected from the survey. As a result, the agents may not have to map to the real occupants in an one-to-one fashion.

Two weeks of measurements were used as training data for calibration, while the remaining data (7 weeks, with some days unavailable due to technical issues) were used as verification data. Model calibration, following the procedure described in Section 3.1, led to the following values of parameters:  $p_d = 0.8$ ,  $p_a = 0.5$ ,  $\alpha = 0.1$ . An early bird and a night low were identified from the survey, whose arrival time and departure time were changed during calibration. For secondary agents,  $p_d = 0.5$  was used. For the sake of simplicity, the acceleration module was not used for secondary agents.

Figure 7 compares the mean value of occupancy of the entire building estimated from three sources: measurements, prediction by the MuMo model and the survey. Because of the limited number of measurements available, the measured mean occupancy is computed only for a 24-h period by averaging over the measurements obtained for 30 weekdays. Model prediction of mean occupancy is computed by averaging over 5000 samples from Monte-Carlo simulations. The mean occupancy estimated from survey was computed in the manner described in Section 3.2.2. The prediction errors are NRMSD ( $x_{\text{MuMo}}, x_{\text{meas}}$ ) = 0.0995, while NRMSD ( $x_{\text{survey}}, x_{\text{meas}}$ ) = 0.2118. This much higher accuracy of the model compared to the survey’s prediction is interesting since the model is generated from survey data as well. However, the agent-based model mimics various aspects of people’s behaviour, including the fact that they do not remain inside the building for the whole duration between arrival and departure. Therefore it is able to predict the

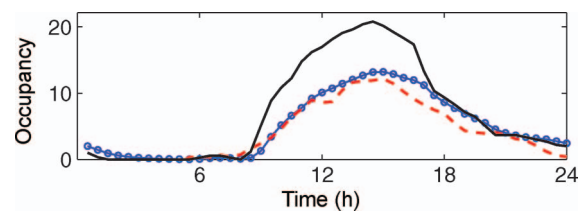


Figure 7. Estimate of mean occupancy (for the building shown in Figure 4) in a MOMZ scenario from three sources: measurements (dashed red), MuMo model prediction (dotted blue), and survey (black).



trend of building occupancy better than the survey. The large over-prediction of mean occupancy by direct processing of survey information shows that *using schedule information, even after accounting for probabilities of presence obtained from a survey, may lead to poor estimation of building occupancy.*

Figures 8(a)–(e) show the pmfs and CDFs of variables such as first arrival time (for the whole building) as estimated from 1000 Monte-Carlo simulations of the MuMo model. They also show the distributions estimated from sensor measurements (verification data) for the same variables. The

thresholds used are:  $\theta_{\text{empty}} = \theta_{\text{occ}} = 3$ . A larger threshold is used here compared to the previous two scenarios since we are dealing with a building with more than 50 occupants. We see from Figure 8(a) that the model does predict the location of the main peak in the pmf of the first arrival time quite well, though it misses a peak corresponding to late first arrivals. The model's prediction of the last departure time is poorer than that for the first arrival time, as seen from Figure 8(b). There is a large probability of the last departure time being close to 6 pm that the model does not reproduce. It also over-predicts the probability of very

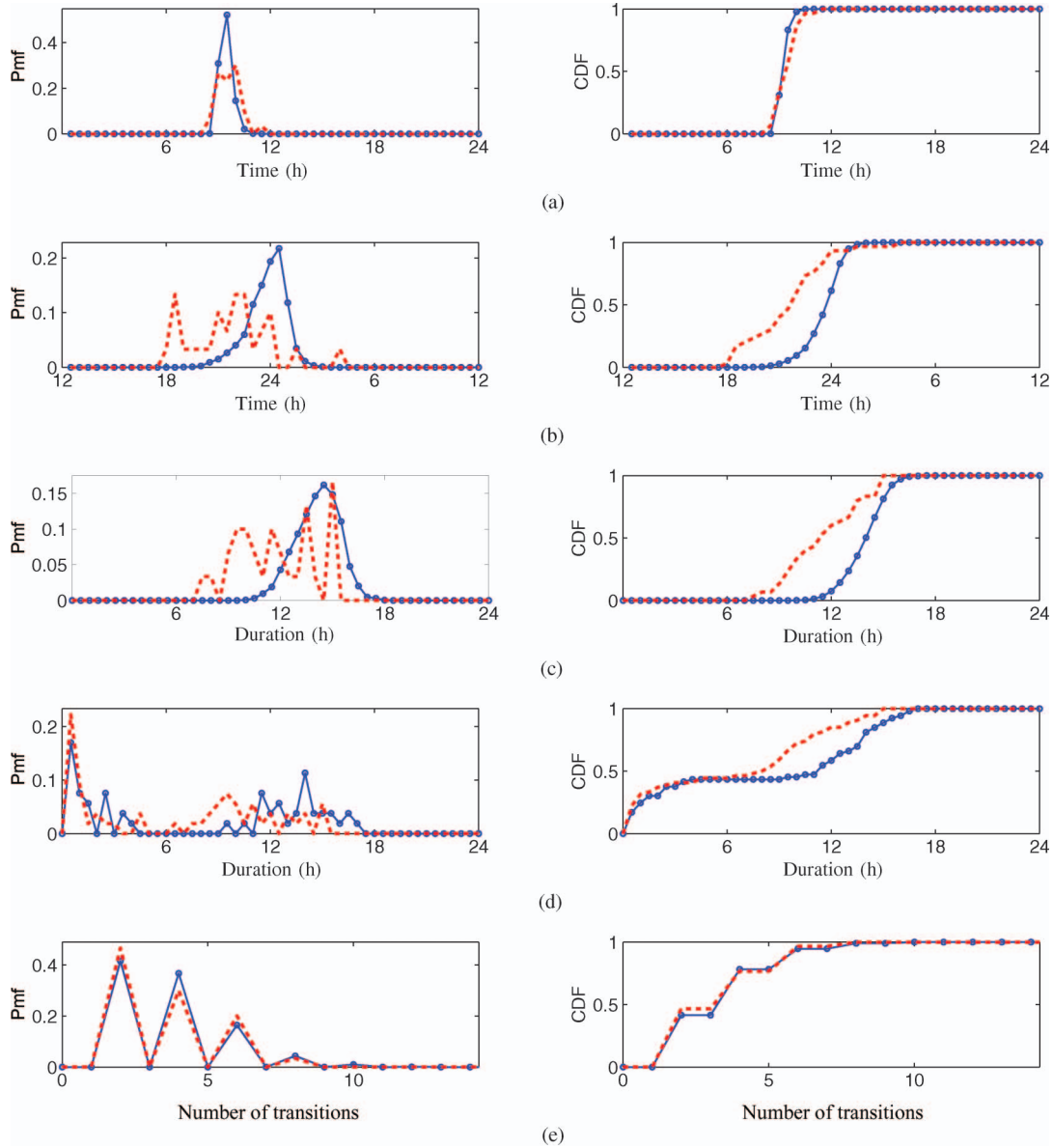


Figure 8. Verification of the MuMo model in the MOMZ scenario: comparison between predicted (dotted blue) and measured (dashed red). All variables correspond to total building occupancy. Comparison is for weekdays only, with binsize = 1/2 h. (a) First arrival time. (b) Last departure time. (c) Cumulative occupied duration. (d) Continuously occupied duration. (e) Number of occupied/unoccupied transitions.

late (past midnight) last departures. Since the last departure time of a building is determined by the behaviour of a few “night owls”, the model’s inability to predict these statistics may come from the inaccuracy of the information obtained from the survey. A possible cause of the mismatch is that the night-owl occupants misjudged how often they leave early when they provided this information in the survey. Figure 8(c) shows the distributions of the cumulative occupied duration in a day. As in the single-occupant single-zone and multi-occupant single-zone scenarios, the prediction of this variable is poorer than the rest. The model predicts the multi-modality of the distribution of continuously occupied duration better, which is shown in Figure 8(d). The model predicts the main peak around 1–2 h quite well, but the other peaks are not predicted as accurately. We see from Figure 8(e) (as well as the K–L divergences in Table 4) that, as in the single-occupant single-zone and multi-occupant single-zone scenarios, among all the variables the model predicts most accurately the number of transitions between occupied and unoccupied status.

Overall, while the MuMo model predicts the general trend of the distributions of these variables, and the prediction is quite accurate for a few variables. However, it fails to accurately predict the values of the probabilities for several variables, especially of last departure time and cumulative occupied duration. The mismatch between the model’s prediction and the measured values in the MOMZ scenario is higher than that seen in the previous two scenarios that involved single zones. A higher error in the multi-zone scenario is expected since survey-based data introduce more inaccuracies in an agent-based model as the number of agents increases, and the survey in this case was not as detailed as in the MOSZ case. Another reason for the mismatch may be the limited amount of measured data, as it was the case in the multi-occupant single-zone scenario and discussed in Section 3.2.2. In fact, this factor may be playing an even stronger role here since the verification data were collected from measurements of only 7 weeks. Therefore, a significant share of the difference may come from the measured data and not the model.

Table 4. K–L divergence between  $p_X^{\text{MuMo}}$  and  $p_X^{\text{meas}}$  in the multi-occupant multi-zone scenario.

Variable( $X$ )	$d(p_X^{\text{mea}} \parallel p_X^{\text{MuMo}})$
First arrival time	0.5667
Last departure time	0.7496
Cumulative occupied duration	0.8814
Continuously occupied duration	0.0981
Number of occupied/unoccupied transitions	0.0248

#### 4. Covariance graph model

In this section, we describe a simpler model of building-occupancy than the agent-based model described in the previous section. Agent-based models are not suitable for real-time occupancy estimation, for which a compact representation of the occupancy dynamics is needed whose predictions can be readily fused with sensor measurements to yield occupancy estimates.

The model we describe now, which is called a *Covariance Graph Model*, or a *graphical model* for short, compactly represents marginal dependencies among the occupancy of various zones. Note that a graphical model does not describe the behaviour of individual occupants. Graphical models have been widely used in spatial statistics, image analysis and bioinformatics; see (Lauritzen 1996; Rue and Held 2005) and references therein. A covariance graph model of an  $n$ -variate distribution is specified in terms of the mean  $\mu$  and covariance matrix  $\Sigma = \{\sigma_{ij}\}$ . It is called a graphical model since the structure of the covariance matrix defines a graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , where  $\mathcal{V} = \{1, \dots, n\}$  is the node set and  $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$  is the edge set with the property that  $(i, j) \notin \mathcal{E} \Rightarrow \sigma_{i,j} = 0$  (Drton and Perlman 2004). For occupancy modelling, the random vector whose distribution we are interested is the occupancy vector  $\mathbf{x}(k) = [x_1(k), \dots, x_n(k)]^T$ , where  $x_i$  is the occupancy of node  $i$ ,  $i = 1, \dots, n$ . Note that the  $n + 1$ -th node that corresponds to “outside the building” is not part of the nodes of the graph. Since occupancy statistics vary with time, the model varies with time as well, so that we have a sequence of covariance graph models  $(\mu(k), \Sigma(k))$ ,  $k = 1, \dots, K$ .

Graphical models can be identified from time-series of the underlying random vector. For the application in hand, we need zone-level occupancy data for graphical model identification. In a building in which each zone is equipped with sensors that measure the zone’s occupancy, such data can be directly obtained from the sensors. Simulations of the agent-based model can also be used to produce the required time-series data. The next paragraph discusses the identification of covariance graphical model from time-series of occupancy of all zones. We briefly note that the graphical models we describe are relevant only in scenarios involving multiple zones.

##### 4.1. Model identification

To describe model identification, we first consider the case when the model does not change with time. The identification of a covariance graph model  $(\mu, \Sigma)$  from samples of the random vector  $\mathbf{x}$  consists of two steps:

(i) model selection and (ii) parameter estimation. Model selection refers to choosing the structure of the graph  $\mathcal{G}$  (or equivalently, the sparsity pattern of  $\Sigma$ ), while parameter estimation refers to choosing the values of those entries of  $\Sigma$  that have been decided to be non-zero in the model selection step. The goal of this two-step identification is to estimate the possibly sparsest graph structure that can still explain the first- and second-order statistics of the data. We follow the methods proposed by Drton and Perlman (2004) and Chaudhuri *et al.* (2007) to carry out the model selection and parameter estimation steps. The first step is the computation of the sample covariance matrix from  $N$  samples of data. Assuming we conduct  $N$  Monte-Carlo experiments, for every time  $k$ , we compute the sample mean

$$\bar{\mathbf{x}}(k) = \sum_{j=1}^N \mathbf{x}^{(j)}(k), \quad (4)$$

where  $\mathbf{x}^{(j)}(k)$  is  $\mathbf{x}(k)$  observed in the  $j$ -th experiment. Similarly, we compute the sample covariance of the state at time  $k$  as

$$W(k) = \frac{1}{N} \sum_{j=1}^N (\mathbf{x}^{(j)}(k) - \bar{\mathbf{x}}(k))(\mathbf{x}^{(j)}(k) - \bar{\mathbf{x}}(k))^T. \quad (5)$$

Model selection is based on hypotheses testing on all edges (i.e. all entries of  $W(k)$ ) at an overall confidence level determined by a designed parameter  $\eta$  (Drton and Perlman 2004). Assuming that the data come from the true graph model  $\mathcal{G}$ , this method leads to an estimated graph model  $\hat{\mathcal{G}}_\eta$  with the following confidence level  $\lim_{n \rightarrow \infty} \inf P(\hat{\mathcal{G}}_\eta = \mathcal{G}) \geq 1 - \eta$ . This means, for fixed  $\eta$ , the correct model is selected with probability at least  $1 - \eta$  for large sample size. Once the structure of the graph model is chosen based on model selection, an iterative conditional fitting algorithm based on maximum likelihood estimation is used for estimating the values of the non-zero entries of  $\Sigma$ . The interested reader is referred to Chaudhuri *et al.* (2007) for the details. Although rigorous results on identification of graphical models require the assumption that the underlying distribution is multi-variate Gaussian, applications of these models to non-Gaussian data are common (Borgelt and Kruse 2002).

For time-varying models, we can follow the method above to identify a distinct model at each time step. Using multiple time-series of occupancy in an  $n$ -room building, we estimate  $K$  covariance graph models ( $\mu(k)$ ,  $\Sigma(k)$ ),  $k = 1, \dots, K$ , one for each time step. There is one difficulty in proceeding in such a straightforward manner. The sample covariance matrix  $W$  in Equation (5) is required to be non-singular for these methods to

be applicable (Drton and Perlman 2004, Chaudhuri *et al.* 2007). At certain times, especially at night, the probabilities of certain rooms being occupied are very small. In this case, we may get zero rows and columns in the sample covariance matrix due to finite number of samples. To address this issue, we eliminate these rows and columns and construct reduced sample covariance matrix  $W(k)_r$ . The reduced covariance matrix  $\Sigma(k)_r$  can now be identified by using the techniques described above, and then  $\Sigma(k)$  can be regained by plugging in the zero rows and columns back to  $\Sigma(k)_r$ .

The model selection part of the graphical model identification process reduces the complexity of the model significantly. For comparison we define a *sample covariance graphical model* in which the sample covariance computed from data is used for the covariance matrix without model selection. Since the latter does not perform model selection, any non-zero entry in the sample covariance matrix, no matter how small, will be retained. As a result, if only the sample covariance is used, the number of edges in the graphical model will be far larger than that if model selection is performed. The reduction in model complexity (as measured by the number of edges in the graph) due to the model selection part can be seen clearly from Figure 9, which shows the number of edges as a function of time in both the models. It should be noted that model selection has minimal effect on the predictive power of the model. As evidence, Figure 10 shows the mean occupancy of the building predicted by the covariance graph model as well as that by the sample covariance graph model, which are seen to be nearly indistinguishable.

#### 4.2. Comparison with agent-based model

The graphical model can be used to generate realizations of the occupancy vector  $\mathbf{x}(k)$ . At every  $k$ ,  $\mathbf{x}(k)$  has to be generated to be multi-dimensional Gaussian with

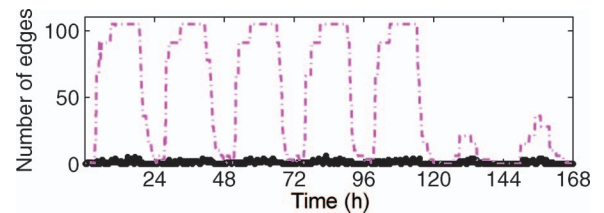


Figure 9. Effect of model selection on model complexity. The number of edges in covariance graph model (dot black) compared to that in the sample covariance graphical model (dashed pink). Model selection is not applied to the sample covariance graph model, which leads to a large number of edges compared to the covariance graph model.

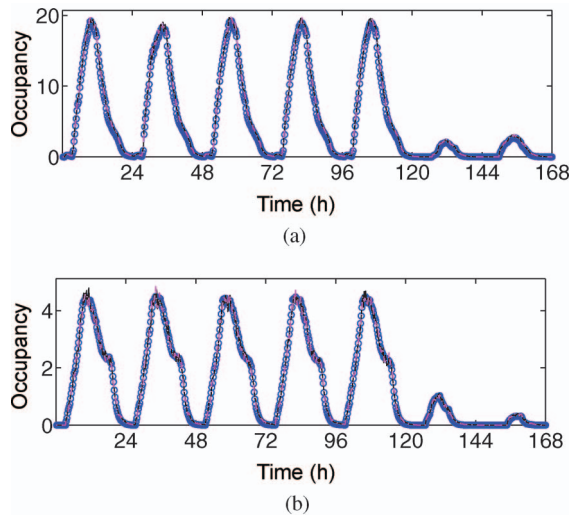


Figure 10. Comparison between mean occupancy predicted by the agent-based MuMo model (dotted blue) as well as the covariance graphical model (CGM) (black), which was identified from simulations of the agent-based model. The prediction by the sample covariance graphical model (SCGM) (dashed pink) is also plotted to show that model selection in the identification of the CGM has little impact on the model's predictive capability. (a) Building (Figure 4). (b) Zone 14 (Figure 4).

mean  $\mu(k)$  and covariance matrix  $\Sigma(k)$ , which can be done by software packages such as MATLAB [copyright]. Building occupancy can then be computed by summing over the entries of  $\mathbf{x}(k)$ . When generated this way, the resulting time series of  $\mathbf{x}(k)$  will be approximately independent over time (depending on the properties of the pseudo-random number generator used) and multi-dimensional Gaussian at every time. The temporal correlations between occupancy at distinct times are not captured by the graphical model, and therefore will not be reflected in the time-series generated by the graphical model.

Figure 10 compares the mean occupancy of the building (the one shown in Figure 4) as estimated by the agent-based MuMo model and the covariance graphical model that was identified from simulation time-traces of the agent-based model. The prediction of the mean occupancy from the two models are almost identical, which is expected since the graphical model is constructed to capture the first and second moments of the agent-based model.

Apart from predicting mean occupancy over time, graphical model is also useful in performing real-time estimation from limited measurements. If sensors are installed in a few zones of a building that provide information on occupancy in those zones, then these measurements can be fused with the mean and covariance information embedded in the graphical model to estimate occupancy in all the zones that are

nodes of the graph. The interested reader is referred to Liao and Barooah (2010) for the details of occupancy estimation using graphical models.

## 5. Summary and future work

This article makes two contributions. First, we presented a novel stochastic agent-based model of occupancy dynamics in a building with an arbitrary number of zones and occupants. The proposed MuMo model can be used to simulate the evolution of occupancy over time during non-emergency situations. Second, we showed that the graphical modelling framework can be used to construct a simpler model of occupancy in a multi-zone building. The graphical model only retains information about mean and covariance of occupancy at various zones over time, but is more suitable for applications such as real-time occupancy estimation due to the reduced complexity compared to the agent-based model. Simulation time-traces from the MuMo model can be used to identify the graphical model.

In the single-occupant single-zone scenario, we found that the MuMo model has similar predictive capability as the model by Page *et al.* (2008) that was proposed for this scenario. In the multi-occupant single-zone and multi-occupant multi-zone scenarios, it was found through comparison with measured data that it predicts certain variables more accurately than others. In general, mean occupancy, and the marginal distributions of the first arrival time, continuously occupied duration, and number of transitions between occupied and unoccupied states are predicted well. However, the distribution of last departure time and cumulative occupied duration are not predicted as well.

In cases involving multiple agents, the inputs that have to be provided to the model usually have to be collected from the occupants by conducting a questionnaire-based survey, since obtaining this information from sensor measurements may be infeasible for even a moderately sized building. For situations involving a large number of agents, gathering enough information to specify the input to the model may become a hurdle in using such a model. The accuracy of the input information is likely to affect model prediction accuracy; but it is not known by how much. Our survey for the multi-occupant multi-zone scenario was not as detailed as in the multi-occupant single-zone scenario, and the prediction accuracy was found to be correspondingly lower. Among the three scenarios, the difference between the prediction and measured values seems to be poorest in the multi-occupant multi-zone scenario. However, the number of measurements available to compare the model's



prediction against was also the smallest in the multi-occupant multi-zone scenario. This makes drawing strong conclusions difficult. The first arrival time and the last departure time are found to depend on a few key individuals, so accuracy of the model depends on obtaining accurate information on their behaviour pattern.

Due to slow variation in a building's usage pattern over time (owing to change in occupants, etc.), the dynamics of occupancy in a building may change with time. A useful attribute of an occupancy model is its ability to adapt to these slow variations automatically based on measurements. One can think of it as the ability to self-calibrate the parameters of the model as the dynamics change. We have not addressed this issue; but it presents an exciting avenue for future research. Due to its lower complexity, the graphical model may be a more appropriate candidate for performing such on-line adaptation compared to an agent-based model.

The graphical model described here is meant to capture only spatial correlations; it loses information on temporal correlations in occupancy. Meaning, the statistical relationship among occupancy at distinct times is lost. However, it should be possible to retain these temporal dependencies using a graphical model in the following way. Define the augmented state vector as  $\mathbf{X}(k) := [\mathbf{x}^T(k), \mathbf{x}^T(k+1), \dots, \mathbf{x}^T(k+\tau)]^T$ , where  $\tau$  is some positive integer. The mean and covariance matrix of  $\mathbf{X}(k)$ , which is a graphical model with  $n\tau$  nodes, will have information on temporal dependencies between occupancy at time less than  $\tau$  indices apart. This is a subject of future study.

### Acknowledgement

The research reported in this work has been supported by the National Science Foundation through grants ECCS-0955023 and CNS-0931885. The authors wish to thank Dr. D. Robinson and F. Haldi for providing us with the data reported in Page *et al.* (2008) that was used in model verification for the single-occupant single-zone scenario.

### Note

1. Another reason for introducing the term “node” is its use in the graphical models, which will be described later.

### References

- Abushakra, B., *et al.*, 2001. *Compilation of diversity factors and schedules for energy and cooling load calculations*. Technical report, Energy Systems Laboratory, Texas A and M University. <http://hdl.handle.net/1969.1/2013>.
- Abushakra, B. and Claridge, D.E., 2008. Modeling office building occupancy in hourly data-driven and detailed energy simulation programs. *ASHRAE Transactions*, July 114(2).
- Borgelt, C. and Kruse, R., 2002. *Graphical models: methods for data analysis and mining*. Wiley. ISBN-10: 0470843373.
- Chaudhuri, S., Drton, M., and Richardson, T.S., 2007. Estimation of a covariance matrix with zeros. *Biometrika*, 94 (1), 199–216.
- Cover, T.M. and Thomas, J.A., 1991. *Elements of information theory*. Wiley Series in Telecommunications. New York: Wiley Interscience.
- Drton, M. and Perlman, M., 2004. Model selection for Gaussian concentration graphs. *Biometrika*, 91, 591–602.
- Gwynne, S.M.V. and Kuligowski, E.D., 2009. Simulating a building as a people movement system. *Journal of Fire Sciences*, 27 (4), 343–368.
- Gyalistras, D. and Gwerder, M., 2010. *Use of weather and occupancy forecasts for optimal building climate control (OptiControl): two years progress report*. Technical report, ETH Zurich, and Siemens Building Technologies Division, Siemens Switzerland Ltd., Zug, Switzerland. <http://www.opticontrol.ethz.ch/Literature.html#top>.
- Helbing, D., 2001. Traffic and related self-driven many-particle systems. *Review of Modern Physics*, 73, 1067–1141.
- Hutchins, J., Ihler, A., and Smyth, P., 2007. Modeling count data from multiple sensors: a building occupancy model. *IEEE International Workshop on Computational Advances in MultiSensor Adaptive Processing*, December.
- Lauritzen, S.L., 1996. *Graphical models*. Oxford: Clarendon Press.
- Liao, C. and Barooah, P., 2010. An integrated approach to occupancy modeling and estimation in commercial buildings. In *Proceedings of the 2010 American Control Conference*, 30 June–2 July, 3130–3135.
- Meyn, S., *et al.*, 2009. A sensor-utility-network method for estimation of occupancy distribution in buildings. *48th IEEE Conference on Decision and Control*, 16–18 December 1494–1500.
- Oldewurtel, F., *et al.*, 2010. Energy efficient building climate control using stochastic model predictive control and weather predictions. In *Proceedings of the American Control Conference*, Baltimore, MD, USA, 5100–5105.
- Page, J., *et al.*, 2008. A generalised stochastic model for the simulation of occupant presence. *Energy and Buildings*, 40, 83–98.
- Pelechano, N. and Malkawi, A., 2008. Evacuation simulation models: challenges in modeling high rise building evacuation with cellular automata approaches. *Automation in Construction*, 17 (4), 377–385.
- Richardson, I., Thomson, M., and Infield, D., 2008. A high-resolution domestic building occupancy model for energy demand simulations. *Energy and Buildings*, 40 (8), 1560–1566.
- Rue, H. and Held, L., 2005. *Gaussian Markov random fields: theory and applications*. Monographs on Statistics and Applied Probability. Vol. 104. London: Chapman and Hall.
- Shendarkar, A., *et al.*, 2006. Crowd simulation for emergency response using BDI agent based on virtual reality. *WSC'06: Proceedings of the 38th conference on Winter simulation*, Monterey, California, Winter Simulation Conference, 545–553. ISBN: 1-4244-0500-9.
- Tanimoto, J., Hagishima, A., and Sagara, H., 2008a. A methodology for peak energy requirement considering actual variation of occupants' behavior schedules. *Building and Environment*, 43, 610–619.
- Tanimoto, J., Hagishima, A., and Sagara, H., 2008b. Validation of methodology for utility demand prediction considering actual validations in inhabitant behavior schedules. *Journal of Building Performance Simulation*, 1, 1–12.

- Tanimoto, J., Hagishima, A., and Sagara, H., 2008c. Validation of probabilistic methodology for generating actual inhabitants' behavior schedules for accurate prediction of maximum energy requirements. *Energy and Buildings*, 43, 316–322.
- Tomastik, R., *et al.*, 2008. Model-based real-time estimation of building occupancy during emergency egress. *Pedestrian and Evacuation Dynamics*, Berlin, Heidelberg: Springer, 215–224.
- Vieira, R.K., *et al.*, 1996. How contractors really size air conditioning systems. *ACEEE Summer Study on Energy Efficiency in Buildings*, American Council for an Energy-Efficient Economy, Washington, DC. [www.fsec.ucf.edu/en/publications/html/FSEC-PF-289-95/index.htm](http://www.fsec.ucf.edu/en/publications/html/FSEC-PF-289-95/index.htm)
- Wang, D., Federspiel, C.C., and Rubinstein, F., 2005. Modeling occupancy in single person offices. *Energy and buildings*, 37 (2), 121–126.
- Yamaguchi, Y., Shimoda, Y., and Mizuno, M., 2003. Development of district energy system simulation model based on detailed energy demand model. *In: Eighth International IBPSA Conference*, Eindhoven, The Netherlands; 1443.