2011 American Control Conference
on O'Farrell Street, San Francisco, CA, USA
June 29 - July 01, 2011

A Novel Stochastic Agent-based Model of Building Occupancy

Chenda Liao and Prabir Barooah

***Abstract*— 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 (occupant) over time, from which a time-series of occupancy (number of people in each zone) can be determined. The model is meant to provide realistic simulation of occupancy dynamics in non-emergency situations, which can be used for extraction of reduced order models of occupancy dynamics for estimation and control purposes. Comparison of the model's prediction of mean occupancy, and distributions of random variables such as first arrival time, are provided against those estimated from measurements in a commercial building.**

I. INTRODUCTION

There is an increasing demand on developing methods to design and operate smart buildings that have high energy efficiency, high level of thermal comfort and higher safety and security features. Modeling occupancy dynamics in buildings is going to be important in achieving this vision. A model of occupancy dynamics is a mathematical tool to predict occupancy (number of people) in a building (or zone) as a function of time given some initial condition. Such predictions can serve as inputs to various types of building energy simulation tools and models in the commissioning and recommissioning phase. For instance, heating and cooling loads experienced by the building HVAC (heating, ventilation and air-conditioning) system strongly depend on time variation of building occupancy. HVAC equipment schedules can be optimized based on the relevant statistics, such as means, variances, and max values of occupancy-driven energy loads computed from the model's predictions. Another use of such models is real-time estimation/prediction 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 [1]. Certain control techniques designed to reduce HVAC energy use, such as demand controlled ventilation, also need real-time occupancy estimation/prediction capability. There are several types of sensors that can provide information on occupancy indirectly, such as CO2 sensors, video cameras, and PIR motion detectors. However, sensor measurements alone may not be enough for accurate estimation since they suffer from large measurement error [2], [3]. Filtering techniques can be used to compensate for measurement error by fusing noisy sensor measurement with prediction from a model [3], [4]. This requires a model of occupancy dynamics.

Constructing a mathematical model of occupancy dynamics of a building is a challenging problem because of the high uncertainty of people movement that governs occupancy evolution. On the lower end of the spectrum of complexity – as well as predictive capability – are models with low temporal and spatial resolution that only seek to predict the whole-building (or a single zone) occupancy at an hourly rate (see [5], [6], [7], [8], [9]). On the high-resolution end of the spectrum of modeling 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 behaviors that are designed to mimic behavior 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, which 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 [10] and reference therein. However, almost all the work on agent-based modeling of occupants in buildings have been designed to study emergency situations such as fire and explosions [11], [12], [13]. The number of works that seek to model building occupancy dynamics during normal, day-to-day operations using agent-based models is limited with the exception of a few studies such as [14], [15]. The most relevant one among these is the work by Page et al. [14]. They model the dynamics of a single person in a single-occupancy room by a time-inhomogeneous Markov chain with two states (in/out or occupied/unoccupied). The model requires as input the sequence of 2×2 transition probability matrices $\mathbf{P}(k), k = 1, \dots, K$, where K is the number of time periods for which the simulation is to be conducted. 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 each transition probability matrix $\mathbf{P}(k)$ becomes an $(n+1) \times (n+1)$ matrix, which is not-trivial to determine from sensor measurements. The paper by Erikson et al. [15] introduces an agent-based model of occupants in 4 zones. The model is constructed from measurements of people's trajectories obtained from cameras. This method of constructing models is not feasible for a large building with a large number of occupants.

In this paper 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

C. Liao and P. Barooah are with the Dept. of Mechanical and Aerospace Engineering, Univ. of Florida, Gainesville, FL; {cdliao,pbarooah}@ufl.edu.
This work has been supported by the National Science Foundation by Grants CNS-0931885 and ECCS-0955023.

model, named *Multiple Modules* (MuMo) model, decides the location of an agent over 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 [14]; the latter is denoted by "Page model" in the sequel. The model is constructed from information obtained from survey of the building occupants. For greater accuracy, the model need to be calibrated for each building, which requires a limited amount of sensor data. Note that for applications such as real-time estimation, an agent based model such as the MuMo model is not appropriate. However, reduced order models that are more appropriate for real-time applications can be extracted from Monte Carlo simulations of the MuMo model; which is described in [16].

We address three distinct scenarios: single-occupant single-zone, multi-occupant single-zone and multi-occupant multi-zone. The verification of single-occupant single-zone scenarios has been previously reported in [4], where the measured data were provided by author in [14]. Therefore, in this paper we only report performance evaluation of the model in the multi-occupant single-zone and multi-occupant multi-zone scenarios. Verification data for these scenarios were collected in a building in the University of Florida campus by using a number of video cameras for a few months. The model's prediction of mean occupancy as a function of time is compared with that determined from measured data. We also compare the model's prediction of the distribution of key random variables such as first arrival time and last departure time of occupants in a zone with those estimated from measurements. The model predicts the mean occupancy quite well. The predictions of the distributions are mixed in the sense that a few variables are predicted well, but not all.

The rest of the paper is organized as follows. Section II describes the proposed agent-based model. Section III describes the calibration procedure and the verification of the model based on comparison with sensor measurements. The paper concludes with a discussion in Section IV.

II. Agent-based model of building occupants

Consider a building with n *zones* that is occupied by m individuals, called *agents*. Time is measured with a discrete time index $k = 1, \dots, K$, where K is maximum time index, with a sample period of T (measured in 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). The *state* $z_i(k) \in \{1, \dots, n+1\}$ of agent i at time index k refers to the node that the agent occupies during time interval $[(k-1)T \;\; kT]$. The occupancy $x_j(k)$ of node j at time k is defined as the number of entries of the set $\{i | z_i(k) = j\}$ and the occupancy of a n-zones building at time k 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 index. 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 ℓ^{th} module is denoted by $z_i^{(\ell)}(k)$, and the output of the last module is $z_i(k)$.

A. Description of the MuMo model

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

0) **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 *pmf* (probability mass function) matches the nominal presence probability profile, i.e., $\Pr(z_i^{(0)}(k) = j) = P_{i,j}(k)$.
1) **Damping and acceleration modules:** Each agent has an associated *primary zone* that corresponds to the zone in the building where the agent spends most of time. People in primary zone (or outside building) tend to stay there for relatively long periods, while people in hallways or restrooms tend to leave quickly. A damping and an acceleration module are used to mimic this behavior by utilizing *transition probability parameters* p_d and p_a. The implementation of the damping module is as follows: 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 $1-p_d$. In acceleration module, if $z_i^{(0)}(k) = z_i(k-1)$ is either restroom or hallway, then $z_i^{(0)}(k)$ is recomputed with probability p_a by running preliminary state generator module again, and the output is assigned to $z_i^{(1)}(k)$. If both the modules are not applicable, $z_i^{(1)}(k) \leftarrow z_i^{(0)}(k)$. The primary zones of the agents as well as the parameters p_d and p_a are specified as inputs to the model.
2) **Scheduled activity module:** This module takes care of hard constraints on the individuals' locations that may arise from scheduled activities, e.g., the meetings, classes, etc. Specifically, if an agent i has to attend an activity located in node j during a particular time interval, $z_i^{(2)}(k) = j$. Otherwise, $z_i^{(2)}(k) \leftarrow z_i^{(1)}(k)$. Those scheduled activities are inputs to the model.
3) **Access module:** Each agent has an access profile associated with it that specifies which zones the agent has access to. If $z_i^{(2)}(k) = j$ where j is a node that agent i does not have access to, then $z_i^{(3)}(k) \leftarrow z_i(k-1)$. Otherwise, $z_i^{(3)}(k) \leftarrow z_i^{(2)}(k)$. This module is also invoked for zones that have a maximum occupancy limit, such as classrooms and restrooms, with the same fashion. The occupancy of those zones are constantly

tracked during simulation. 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 $z_i(k)$ 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 damping and acceleration modules. The damping module is a key element of the model. 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 behavior 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 Section I), 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 [14].

B. MuMo model construction

Constructing a MuMo model with m agents requires specifying for each agent its nominal presence probability profiles, scheduled activities, and access profiles. In addition, damping and acceleration parameters p_d and p_a, and maximum occupancy limits of rooms in the building, if any, need to be specified. Because of the challenge in tracking each individual over time using sensors, conducting a survey of the occupants' behavior by asking them to fill out a questionnaire is a feasible - albeit less accurate - way of collecting this information. The most time-consuming part in constructing the model is specifying the nominal presence probability profile for each agent. An algorithm for computing the nominal presence probability profiles and the parameters from a survey of the buildings' occupants is described in [16]. We refrain from giving the details here due to lack of space.

III. 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 – some of 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. The parameters and variables mentioned above are the following:

1) *Mean occupancy of zone/building*: The mean occupancy of zone j at each time index k is defined as $\mathbf{E}[x_j(k)]$, where $\mathbf{E}[.]$ 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, for each day, 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 that day.
3) *Last departure time (in a day)*: the last time during a day at which the zone or building becomes unoccupied.
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 θ_{occp}, not necessarily continuously. More precisely, it is the number of elements of the set $\{k \mid x_j(k) \geq \theta_{\text{occp}}, 1 \leq k \leq 24 \times 60/T\}$ for each day.
5) *Number of occupied/unoccupied transitions (in a day)*: the number of transitions between "occupied" and "unoccupied" status in a day for a zone or building. Specifically, it is the number of elements of the set $\{k \mid x_j(k) \geq \theta_{\text{empty}}, x_j(k+1) < \theta_{\text{empty}}\} \bigcup \{k \mid x_j(k) < \theta_{\text{empty}}, x_j(k+1) \geq \theta_{\text{empty}}\}$, for $1 \leq k < 24 \times 60/T$ for each day.

Monte-Carlo simulations of the model are conducted, and the resulting multiple time-series (each one-week long) are used to estimate the pmfs of these random variables. Those pmfs are also estimated from the repeated segments of one-week-long processed sensor data (measurements). Comparison between the two provides an idea of how well the agent-based model can predict such zone-level or building-level phenomenon.

To quantitatively compare the time series of mean occupancy of between the model predictions and measurements, we simply use normalized root mean square deviation (NRMSD). Let $x(k)$ and $y(k)$, $k = 1, \ldots, K$ be two time sequences, the NRMSD between x and y is defined as

$$\text{NRMSD}(x, y) = \frac{\|x - y\|/\sqrt{K}}{\max z - \min z}, \tag{1}$$

where $x = [x(1), \ldots, x(K)]^T$, $y = [y(1), \ldots, y(K)]^T$, $z = [x^T, y^T]^T$ and $\|\cdot\|$ is Euclidean norm. To compare the predicted distributions of those variables 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 [17]

$$d(p\|q) = \sum_i p_i \log(\frac{p_i}{q_i}). \tag{2}$$

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^{Mumo} and p_X^{meas} denote the pmfs of X predicted by the Mumo model and that estimated from measurements.

A. Model calibration procedure

Calibrating the parameters of the MuMo model becomes necessary when the information used in model construction may not be accurate. 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. The statistics of the random variables described

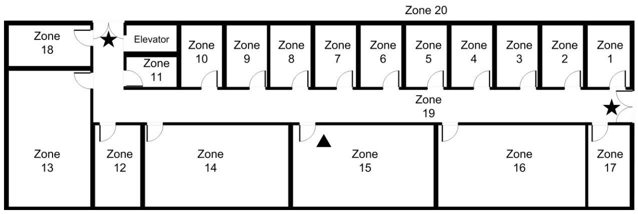


Fig. 1. Floor plan of the 3rd floor of MAE-B building in the University of Florida campus used in verification of the MuMo model for the MOMZ (multi-occupant multi-zone) scenario. Zone 15 is the room that is used for verification of the model in the MOSZ (multi-occupant single-zone) scenario. The triangle and stars indicates the cameras using in MOSZ and MOMZ scenario respectively.

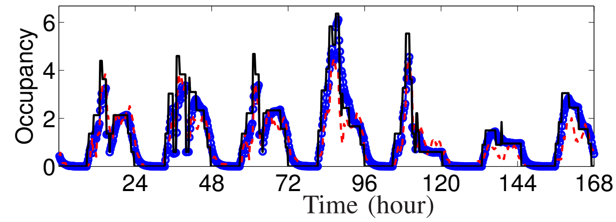
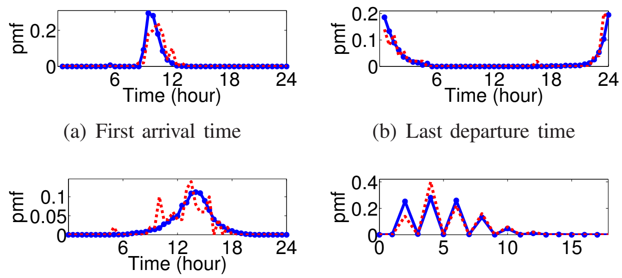


Fig. 2. Mean occupancy estimate in a MOSZ scenario (for zone 15 in Figure 1) from three sources: measurements (dashed red), MuMo model prediction (dotted blue), and survey (black).



(a) First arrival time

(b) Last departure time

(c) Cumulative occupied duration

(d) No. of occupied/unoccupied transitions

Fig. 3. Verification of the MuMo model in the MOSZ scenario: comparison between the predictions (dotted blue) 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 hour.

in the beginning of Section III are first estimated by using only the calibration data. Calibration is then performed by changing the agent based model – from the baseline constructed from the survey – 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 process tractable, we modify only a few parameters, such as the arrival and departure times of the "early bird" and the "night owl", and the transition probability parameters p_d and p_a. The calibration process is described in [16].

B. Model verification

We consider two scenarios: one in which the building consists of a single zone and the other with multiple zones.

1) Model verification: the MOSZ (multi-occupant single-zone) scenario: The MOSZ scenario studied in this paper corresponds to a room in a building in the University of Florida campus, shown as zone 15 in Figure 1. The room housed 5 graduate students who worked there regularly and 3 undergraduate research assistants who used it intermittently. Apart from these, the model also contains 7 additional agents (visitors), who were used to simulate students who would occasionally visit the room to meet with a few of the graduate students (teaching assistants). The MuMo model was constructed by conducting a survey of the occupants to determine the subset of the parameters that are relevant to the MOSZ scenario. We collected occupancy data for this room by using a wireless video camera to monitor the entrance to the room. Data was collected for a period of about four months (during January - April 2010). A motion detection algorithm was used to save only those frames when motion was detected. Manual counting of the number of people was performed to ensure that measurements obtained were of high accuracy. Due to technical problems of video capturing, only 70 days' data could be collected from a total 16 weeks of video feed.

One thousand Monte-Carlo simulations (each of one week duration) are conducted with the proposed model. The statistics of variables other than the mean occupancy, which are described in the beginning of Section III, are estimated from data (measurements and simulation time traces). Here we provide comparison of statistics of variables between the two only for weekdays. The thresholds used in computing first arrival times etc. are: $\theta_{\text{empty}} = \theta_{\text{occp}} = 0.5$. Two weeks of measurements are used as training data to calibrate the model, while the remaining data are used for verification. Calibration of the model led to a change of the transition probability parameter p_d to 0.8 except for visitors. 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 2 compares the mean occupancy predicted by the proposed model with that computed 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. The prediction errors for mean occupancy are NRMSD$(x_{MuMo}, x_{meas}) = 0.0877$, while NRMSD$(x_{survey}, x_{meas}) = 0.1202$, where x_{MuMo}, x_{meas} and x_{survey} are the mean occupancy of the zone over one week computed from the model's prediction, measurements and survey. From the figure, 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 behavior in the survey.

Figure 3(a) and (b) show the pmfs of the first arrival

$d(p_X^{\text{meas}} \| p_X^{\text{MuMo}})$	MOSZ	MOMZ
First arrival time	0.2388	0.5667
Last departure time	0.1900	0.7496
Cumulative occupied duration	0.3261	0.8814
Number of occupied/unoccupied transitions	0.0810	0.0248

TABLE I

K-L DIVERGENCE BETWEEN p_X^{MuMo} AND p_X^{MEAS} IN THE MOSZ AND MOMZ SCENARIO.

times and last departure times of a day as predicted by the MuMo model as well as that estimated from measurements, respectively. We see from Figure 3(a) that the shape of the distribution is predicted correctly, but there are a few late first arrivals around noon that the model does not capture. Similarly, Figure 3(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 3(c). The mismatch is larger in case of the cumulative occupied duration: there are several peaks in the measured pmf that the model does not predict. The overall trend of the distribution is predicted correctly. Figure 3(d) 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. The K-L divergences between the models' predictions and the measured data are shown in Table I, which confirms that the largest difference between the model and measured data is cumulative occupied duration.

We believe part of the reason for the mismatch between model's prediction and that estimated from measurements, as well as that for the non-smoothness in the measured pmfs, is the limited amount of verification data. Specifically, there are only 50 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. Therefore, the estimates from the measured data may have larger error.

2) Model verification: the MOMZ (multi-occupant multi-zone) scenario: The MOMZ scenario studied here corresponds to the third floor of the MAE-B building in the University of Florida campus (see Figure 1). For the remainder of this paper, we'll refer to the 3rd 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. Measurements were collected by two video cameras targeting on each of the two entrances of the building and processed by motion detection algorithm and manual counting. 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. Measurements presented in this study were collected during a period of about 7 weeks during May - July, 2010.

Data for model construction was collected by a survey

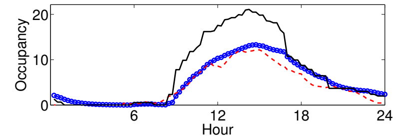
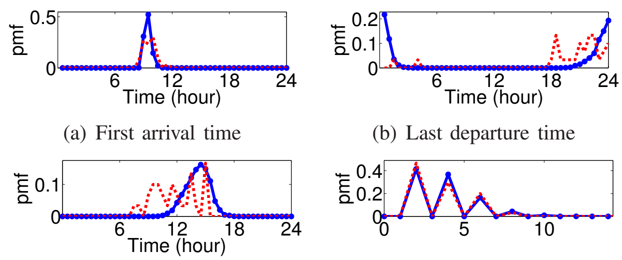


Fig. 4. Estimate of mean occupancy (for the building shown in Figure 1) in a MOMZ scenario from three sources: measurements (dashed red), MuMo model prediction (dotted blue), and survey (black).

of the occupants of the building. The survey was not as extensive as that in the MOSZ scenario (see [16] for the details). Two weeks of measurements were used as training data for calibration, while the remaining data were used as verification data. Model calibration led to the following values of parameters: $p_d = 0.8$, $p_a = 0.5$, $\alpha = 0.1$. An early bird and a night-owl were identified from the survey, whose arrival time and departure time were changed during calibration. Due to the simplicity of visitors' behavior, only $p_d = 0.5$ was used in the model of visitors.



(a) First arrival time

(b) Last departure time

(c) Cumulative occupied duration

(d) No. of occupied/unoccupied transitions

Fig. 5. Verification of the MuMo model in the MOMZ scenario: comparison between predictions (dotted blue) and measurements (dashed red). All variables correspond to total building occupancy. Comparison is for weekdays only, with binsize=1/2 hour.

Figure 4 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-hour 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 III-B.1. The prediction errors are NRMSD$(x_{MuMo}, x_{meas}) = 0.0995$, while NRMSD$(x_{survey}, x_{meas}) = 0.2118$. The NRMSD value of survey prediction is much larger than that of the model prediction, which is interesting since the model is generated from survey data as well. However, the agent-based model mimics various aspects of people's behavior, 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 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 5(a)-(e) show the pmfs 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: $\theta_{\text{empty}} = \theta_{\text{occp}} = 3$. A larger threshold is used here compared to the previous scenario since we are dealing with a building with more than 50 occupants. We see from Figure 5(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 much smaller 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 5(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 late (past midnight) last departures. Since the last departure time of a building is determined by the behavior of a few critical occupants, 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 5(c) shows the distributions of the cumulative occupied duration in a day. As in the multi-occupant single-zone scenario, the prediction of this variable is poorer than the rest. Figure 5(d) shows the distribution of the number of transitions between occupied and unoccupied status, which is predicted by the model quite accurately.

Overall, while the MuMo model predicts the general trend of the distributions of these variables, it does not seem to predict the values of the probabilities accurately. The K-L divergences between the model's predictions and measured data are shown in Table I as well. In the table, we see that the model's prediction in the multi-zone case is poorer than that in the single-zone scenario. A higher error in the multi-zone scenario is expected since survey-based data introduces more inaccuracies in an agent-based model as the number of agents increases. Another reason for the mismatch may be the limited amount of measured data. In fact, this factor may be playing an even stronger role in the multi-occupant multi-zone scenario since the verification data were collected from measurements of only five weeks. Therefore, a significant share of the difference may come from the measured data and not the model.

IV. Summary and future work

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. In the model verification, 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 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 well. The inputs that have to be provided to the model usually have to be collected from the occupants by conducting a questionnaire-based survey. 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.

References

[1] R. Tomastik, S. Narayanan, A. Banaszuk, and S. Meyn, "Model-based real-time estimation of building occupancy during emergency egress," in *Pedestrian and Evacuation Dynamics*, 2008, pp. 215–224.

[2] J. Hutchins, A. Ihler, and P. Smyth, "Modeling count data from multiple sensors: a building occupancy model," in *IEEE International Workshop on Computational Advances in MultiSensor Adaptive Processing*, December 2007.

[3] S. Meyn, A. Surana, Y. Lin, S. M. Oggianu, S. Narayanan, and T. A. Frewen, "A sensor-utility-network method for estimation of occupancy distribution in buildings," in *48th IEEE Conference on Decision and Control*, December 16-18 2009.

[4] C. Liao and P. Barooah, "An integrated approach to occupancy modeling and estimation in commercial buildings," in *American Control Conference*, 2010, pp. 3130–3135.

[5] B. Abushakra and D. E. Claridge, "Modeling office building occupancy in hourly data-driven and detailed energy simulation programs," in *ASHRAE Transactions*, July 2008.

[6] I. Richardson, M. Thomson, and D. Infield, "A high-resolution domestic building occupancy model for energy demand simulations," *Energy and Buildings*, vol. 40, no. 8, pp. 1560 – 1566, 2008.

[7] D. Wang, C. C. Federspiel, and F. Rubinstein, "Modeling occupancy in single person offices," *Energy and buildings*, vol. 37, no. 2, pp. 121–126, 2005.

[8] Y. Yamaguchi, Y. Shimoda, and M. Mizuno, "Development of district energy system simulation model based on detailed energy demand model," in *Eighth International IBPSA Conference*, 2003.

[9] B.Dong and B.Andrews, "Sensor-based occupancy behavioral pattern recognition for energy and comfort management in intelligent buildings," in *7th International IBPSA conference*, 2009.

[10] D. Helbing, "Traffic and related self-driven many-particle systems," *Review of Modern Physics*, vol. 73, pp. 1067–1141, 2001.

[11] S. M. V. Gwynne and E. D. Kuligowski, "Simulating a building as a people movement system," *Journal of Fire Sciences*, 2009.

[12] N. Pelechano and A. Malkawi, "Evacuation simulation models: Challenges in modeling high rise building evacuation with cellular automata approaches," *Automation in Construction*, vol. 17, no. 4, pp. 377 – 385, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/B6V20-4PG2KS6-1/2/48873dcf2cb4b8c1f286f7bef3c62d36

[13] A. Shendarkar, K. Vasudevan, S. Lee, and Y.-J. Son, "Crowd simulation for emergency response using bdi agent based on virtual reality," in *WSC'06: Proceedings of the 38th conference on Winter simulation*. Winter Simulation Conference, 2006, pp. 545–553.

[14] J. Page, D. Robinson, N. Morel, and J.-L. Scartezzini, "A generalised stochastic model for the simulation of occupant presence," *Energy and Buildings*, vol. 40, pp. 83 – 98, 2008.

[15] V. L. Erickson, Y. Lin, A. Kamthe, R. Brahme, A. Surana, A. E. Cerpa, M. D. Sohn, and S. Narayanan, "Energy efficient building environment control strategies using real-time occupancy measurements," in *Conference On Embedded Networked Sensor Systems*, 2009.

[16] C. Liao, Y. Lin, and P. Barooah, "Agent-based and graphical modeling of building occupancy," *Journal of Building Peformance Simulation*, 2011.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, ser. Wiley Series in Telecommunications. New York: Wiley Interscience, 1991.