EXPLORING THE PREDICTIVE POTENTIAL OF PROBABILISTIC OCCUPANCY MODELS

F. Tahmasebi, A. Mahdavi

Department of Building Physics and Building Ecology
Vienna University of Technology, Vienna, Austria

ABSTRACT

Building performance is influenced by occupants' presence and actions. Knowledge of occupants' future presence and behavior in buildings is of central importance to the implementation efforts concerning predictive building systems control strategies. Specifically, prediction of occupants' presence in buildings represents a necessary condition for predicting their interactions with building systems. In the present contribution, we focus on evaluation of probabilistic occupancy models to explore the potential of using past monitored data in predicting future presence of occupants. Toward this end, we selected a university campus office area, which is equipped with a monitoring infrastructure and includes a number of open and closed offices. For the purpose of this study, we used monitored occupancy data and two previously developed stochastic occupancy models to predict the occupancy profiles on a daily basis. The predictions were then evaluated via comparison with monitored daily occupancy profiles. To conduct the model evaluation in a rigorous manner, a number of specific evaluation statistics were deployed. Thus, the results facilitate a discussion of the potential and limitations of predicting building occupants' future presence patterns based on past monitoring data.

INTRODUCTION

Occupants influence thermal behavior of buildings due to their presence (e.g., via releasing sensible and latent heat), and via operation of control devices such as windows, shades, luminaries, radiators and fans (Mahdavi 2011). Specifically, knowledge of occupants' presence represents a necessary condition for the development of predictive control action models. Performance simulation tool users typically deploy libraries of diversity factors and schedules to represent occupants' presence in buildings. These diversity profiles are derived from long term monitored data in different classes of buildings and are usually included in the simulation packages to facilitate the creation of building performance models. More recently, efforts are being made in the scientific and professional communities to develop probabilistic models that would capture the randomness of occupants' presence. As one of the first attempts, Newsham et al. (1995) considered the probabilistic nature of occupancy while developing a stochastic model to predict lighting profiles for a typical office. Their model deployed the probability of first arrival and last departure as well as the probability of intermediate departures from and returning back to the workstations. Reinhart (2001) further developed this model by using the inverse transform method (Zio 2013) to generate samples from the distribution functions of arrival and departure times. Moreover, days were divided into three phases (morning, lunch, and afternoon) for which the probabilities of start time and length of breaks were computed. Page et al. (2008) proposed a generalized stochastic model for the simulation of occupants' presence using the presence probability over a typical week and a parameter of mobility (defined as the ratio of state change probability to state persistence probability). They also included long absence periods (corresponding to business trips, leaves due to sickness, holidays, etc.) as another random component in their model.

In all these studies, monitored data has been used to derive a probabilistic model that generates random non-repeating daily profiles of occupancy for a long-term (e.g. annual) building performance simulation. Hence, models are suggested to perform well, if the entire set of generated random realizations of the daily occupancy profiles agrees in tendency with the monitored data over the whole simulation period. However, the synchronicity of the generated profiles with the monitored data (one-to-one agreement of the generated and monitored daily profiles) is not taken into consideration. Even in the case of long absences (Page et al. 2008), the unoccupied days are scattered randomly through the year and they do not necessarily match the dates of absences in the measured data. Thus, these practices cannot be said to "validate" the proposed models, if the actual day-to-day prediction of occupancy and control action probabilities is relevant. Specifically, in a run-time use of a simulation model in building operation phase, where short-term predictions of occupancy and weather data are incorporated in the model to predict the future performance of the building, the agreement between the predicted and real future occupancy profiles in each day is of utmost importance.

Moreover, with few exceptions (see Liao et al. 2010), most of the work on validating the probabilistic occupancy models has focused on comparing the model outputs with the very set of data which has been used to derive the model. In our view, a scientifically sound model evaluation approach must clearly separate the data segments used for model development and model validation. This is especially important while evaluating the predictive potential of an occupancy model, which is intended to be used for model-predictive control in buildings.

In the present study, we focus on evaluation of two recently developed stochastic models of occupants' presence to explore the potential of using past monitored data in predicting future presence of occupants. Toward this end, we selected a university campus office area, which is equipped with a monitoring infrastructure and includes a number of open and closed offices. For the purpose of this case study, we deploy long-term monitored occupancy data obtained from eight workspaces. Thereby, separate sets of monitored data are used to train and evaluate the models. The evaluations are accomplished with the aid of a number of occupancy-relevant statistics. Thus, the results facilitate a discussion of the potential and limitations of occupants' presence models intended for incorporation in the control logic of existing buildings.

APPROACH

Overview

In this contribution, we evaluate two existing probabilistic models of occupant's presence, which we use to make predictions of daily occupancy profiles for building systems control purposes. We utilize monitored occupancy data obtained from eight workplaces in an office area in the Vienna University of Technology to train the applied stochastic occupancy model. To evaluate the models, we use a number of key statistics and a separate set of monitored occupancy data. Conducting a Monte Carlo simulation, we evaluate the predicted daily occupancy profiles generated by the stochastic models and obtain distribution of the statistics to discuss the reliability of predictions.

Data collection

To obtain occupancy data, wireless ceiling-mounted sensors (motion detectors) were used. The internal microprocessors of the sensors are activated within a time interval of 1.6 minutes to detect movements. The resulting data log entails a sequence of time-stamped occupied to vacant (values of 0) or vacant to occupied (values of 1) events.

To facilitate data analysis, the event-based data streams were processed to generate 15-minute interval data, using stored procedures in the MySQL database (Zach et al. 2012). This procedure derives the duration of occupancy states (occupied / vacant) from the stored events and returns the dominant occupancy state of each interval. Occupancy periods before 8:00 and after 19:45 were not included in the study to exclude, amongst other things, the presence of janitorial staff at the offices. Occupancy data for a nine-month period (10th of November 2011 to 25th of July 2012) was used to derive and evaluate the occupancy model.

Reinhart probabilistic occupancy model

The probabilistic occupancy model developed by Reinhart (Reinhart 2001) uses the following probability distributions as input to capture the random nature of occupants' presence:

1) The cumulative distribution function of first arrival times (CDF_a);
2) The cumulative distribution function of last departure times (CDF_d);
3) The probability distribution function of intermediate departure times (PDF_{id});
4) The probability distribution of length of intermediate absences for morning, lunch, and afternoon periods.

A daily occupancy profile is then generated by identifying the first arrival time, last departure time, intermediate departure times, and the associated length of intermediate absences as follows:

- Using a random number from the standard uniform distribution in the interval [0, 1] (u_1), the first arrival time (t_a) is derived from CDF_a such that $CDF_a (t_a) = u_1$.
- Using a random number from the standard uniform distribution in the interval [0, 1] (u_2), the last departure time (t_d) is derived from CDF_d such that $CDF_d (t_d) = u_2$.
- To decide if an intermediate departure event occurs at a certain time (t_m), a random number between 0 and 1 (u_m) is compared with the probability of intermediate departure at that time. Once an intermediate departure is identified ($PDF_{id} (t_m) \geq u_m$), the length of the absence is obtained randomly from the corresponding probability function of the length of intermediate absences (morning, lunch time, or afternoon).

Page et al. probabilistic occupancy model

The stochastic occupancy model developed by Page et al. (2008) generates random non-repeating daily occupancy profiles using two inputs: the profile of presence probability, and the parameter of mobility.

The model has been formulated based on the hypothesis that the value of occupancy at the next time step depends on the current occupancy state and the probability of transition from this state to either the same state or its opposite state. This is reflected in Equation 1:

$$P(t+1) = P(t)T_{11}(t) + (1 - P(t))T_{01}(t) \quad (1)$$

Where, P(t+1) and P(t) are the probabilities of presence at the time steps t+1 and t, $T_{11}(t)$ is the transition probability from presence state to the same state at the time step t, and $T_{01}(t)$ is the transition probability from absence to the presence state at the time step t.

In order to derive the transition probabilities based on the presence probabilities, Page et al. defined the parameter of mobility, which should be provided as an input, as the ratio between the probabilities of change of the state of presence over that of no change:

$$\mu(t) = \frac{T_{01}(t) + T_{10}(t)}{T_{00}(t) + T_{11}(t)} \quad (2)$$

Here, $T_{10}(t)$ is the transition probability from presence to the absence state at the time step t, and $T_{00}(t)$ is the transition probability from absence state to the same state at the time step t.

From Equations 1 and 2, and assuming the parameter of mobility as a constant, the profiles of transition probabilities can be obtained as follows:

$$T_{01}(t) = \frac{\mu - 1}{\mu + 1} P(t) + P(t+1) \quad (3)$$

$$T_{11}(t) = \frac{P(t) - 1}{P(t)} \left[\frac{\mu - 1}{\mu + 1} P(t) + P(t+1) \right] + \frac{P(t+1)}{P(t)} \quad (4)$$

Clearly, the other possible transition probabilities can be calculated via the following equations:

$$T_{00}(t) = 1 - T_{01}(t) \quad (5)$$

$$T_{10}(t) = 1 - T_{11}(t) \quad (6)$$

To generate a daily occupancy profile, the procedure starts from the first time step of the day with a vacant state for commercial buildings. Subsequently, for each time step, a random number between 0 and 1 is generated and compared with the transition probabilities to see if a change of occupancy state occurs. This is a simple case of using the inverse transform method, as the cumulative distribution function of transition probabilities is a histogram of two bins. For example, if the current time step has a vacant state and the generated random number is smaller than T01 at that time step, the next time step is assumed to be occupied.

Model training

Implementation of stochastic occupancy models in a continuous running mode in building control system raises a number of questions with regard to occupancy data utilization: What length of past occupancy information shall be considered for model development? Would it be advantageous to differently treat days of the week? Shall the model training occur in static or shifting intervals? In previous publications (Tahmasebi et al. 2014), we evaluated the impact of different model training scenarios on the predictive potential of stochastic occupancy models. For the purpose of the present study, a moving training scheme is applied as follows: To generate a predicted occupancy profile for each working day, the occupancy models are fed with the monitoring occupancy data obtained from the previous 28 days. This 4-week data is used to derive the required inputs for the presented occupancy models, i.e. the probabilities of arrival time, departure time, intermediate departure times, and length of the intermediate absences for the Reinhart model, and the presence probability profile and parameter of mobility for Page et al. model.

Model evaluation

To evaluate the predictive potential of the models, we compared predicted and monitored occupancy profiles of 90 working days between the 1st of April and the 25th of July 2012. As for each run the models are fed with the occupancy data from the prior four weeks, separate sets of data are used for training and evaluating the models. To compare the performance of the models, we used four statistics:

1) First arrival time error [hour]: The predicted minus the monitored first arrival time.
2) Last departure time error [hour]: The predicted minus the monitored last departure time.
3) Duration error [hour]: The predicted minus the monitored daily presence duration. We calculated the presence duration by counting the number of occupied intervals.
4) Number of transitions error [-]: The predicted number of daily occupied-to-vacant transitions minus the monitored number of daily occupied-to-vacant transitions.

Given the stochastic nature of the occupancy models considered, one cannot evaluate the accuracy of the model predictions by comparing the results of a single run with the measurements. Therefore we conducted a 100-run Monte Carlo simulation in order to analyze the distribution of the errors in predictions. The aforementioned statistics are calculated for each individual day during the validation period. Given the length of the validation period (90 working days) and the number of Monte Carlo simulations, we obtained 9000 values for each statistic.

Note that in the present study we do not intend to predict periods of long absences due to business trips, sickness, holidays, etc. Such whole-day absences can be presumably communicated to the building management system and reflected in the predictive building systems control. Therefore, in this contribution, we only included the actual working days in the validation process.

RESULTS AND DISCUSSION

Figures 1 to 4 illustrate the cumulative distribution of the statistics absolute values obtained via comparing the Monte-Carlo predictions of the presented stochastic occupancy models and monitoring occupancy data from eight working spaces. It can be seen from the figures that Reinhart model offers slightly higher accuracy in terms of first arrival time and the number of intermediate transitions. However, the models perform almost similarly with regard to last departure time and daily occupancy duration.

A numeric summary of the results are presented in Tables 1 and 2 to provide a general overview of occupancy prediction errors. Table 1 presents the 80th percentile of the errors. Table 2 shows the percentage of errors below a threshold value, which could be arguably seen as a minimum performance requirement for occupancy models appropriate for deployment in the context of predictive building systems control.

This analysis has important implications. The obtained level of predictive accuracy is simply low. For example, it can be clearly seen from Table 1 and Table 2 that the stochastic models used in this study do not perform satisfactorily in predicting the last departure time and occupancy duration, even though the model training was based on high-quality and high-resolution empirical data (from the same workplaces for which predictions were made). It may be argued that the observed large model errors are due to the poor performance of the specific stochastic models considered in the present study. Consequently, we currently explore other modelling approaches and options. Nonetheless, the obtained results could be also interpreted to the effect that there may be potentially a limit (lower uncertainty threshold) in predicting the occupants' presence by using probabilistic models derived based on past occupancy data.

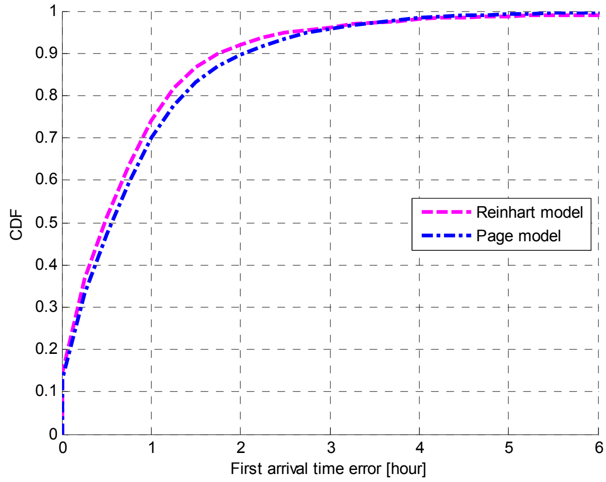

Figure 1: Cumulative distribution of arrival time error absolute values

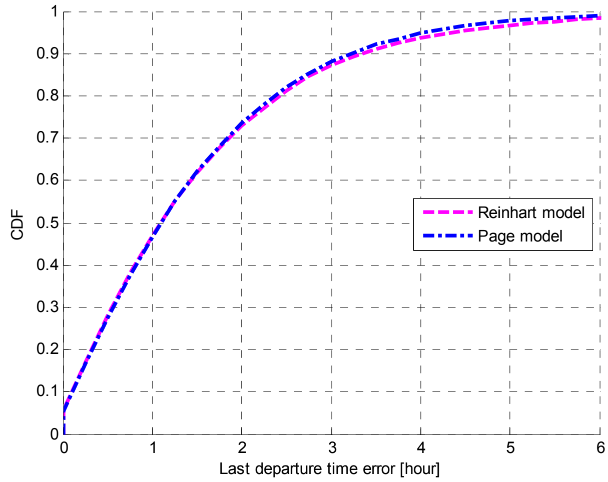

Figure 2: Cumulative distribution of departure time error absolute values

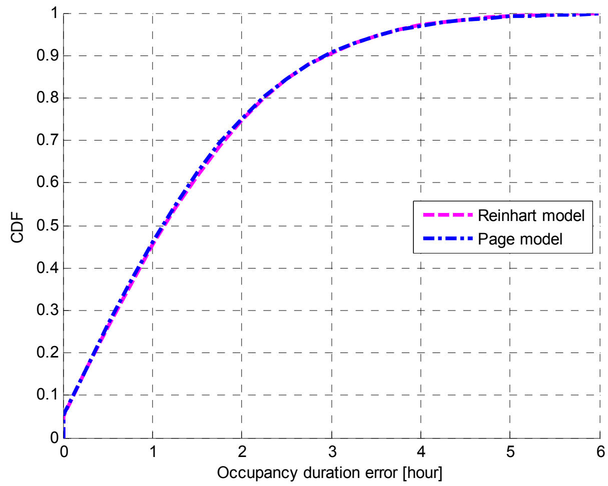

Figure 3: Cumulative distribution of occupancy duration error absolute values

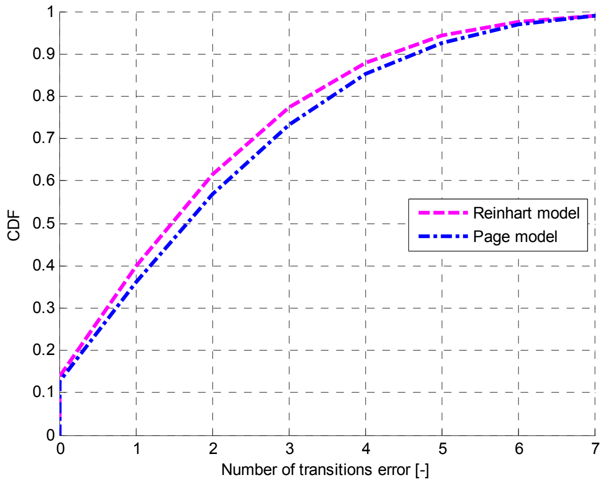

Figure 4: Cumulative distribution of number of transitions error absolute values

Table 1: The 80th percentile of the errors

Evaluation statistics	**Reinhart Model**	**Page Model**
First arrival time error [hour]	1.2	1.4
Last departure time error [hour]	2.4	2.4
Occupancy duration error [hour]	2.3	2.2
Number of transitions error [-]	3.3	3.6

Table 2: Percentage of errors below specific thresholds

Evaluation statistics	**Error threshold**	**Percentage of errors below threshold**	
		Reinhart Model	**Page Model**
First arrival time [hour]	1.0	74.2	70.0
Last departure time [hour]	1.0	46.9	46.7
Occupancy duration [hour]	1.0	45.3	46.1
Number of transitions [-]	2.0	61.5	56.8

CONCLUSION

As noted at the outset of the paper, deployment of stochastic occupancy models in the context of building systems control requires a rigorous standard concerning the evaluation of the models' predictive performance. In such a context, one cannot simply claim that an occupancy model performs well, if it generates occupancy patterns that "resemble" the real ones. We need clearly defined and rigorous statistics to evaluate the predictive performance of an occupancy model. In this context, we suggest that an evaluative approach similar to the one we suggested and applied in this paper is critical for future studies that intend to evaluate and improve stochastic occupancy models.

ACKNOWLEDGEMENT

The research presented in this paper is supported by funds from the project "Control & Automation Management of Buildings & Public Spaces in the 21st Century" (CAMPUS21, Project Number: 285729) as well as the project "Retrofitting Solutions and Services for the enhancement of Energy Efficiency in Public Edification" (RESSEEPE, Project number: 609377) under EU's Seventh Framework Programme.

REFERENCES

Liao C, Lin Y, Barooah P. 2010. Agent-Based and Graphical Modelling of Building Occupancy. Journal of Building Performance Simulation 5(1), p. 5-25.

Mahdavi A. 2011. People in building performance simulation. In: Hensen J, Lamberts R, editors. Building Performance Simulation for Design & Operation. Taylor & Francis; p. 56-83.

Newsham GR, Mahdavi A, Beausoleil-Morrison I. 1995. Lightswitch: a stochastic model for predicting office lighting energy consumption. In: Proceedings of Right Light Three, the 3rd European Conference on Energy Efficient Lighting. Newcastle-upon-Tyne; p. 60-66.

Page J, Robinson D, Morel N, Scartezzini JL. 2008. A generalized stochastic model for the simulation of occupant presence. Energy and Buildings, Volume 40, Issue 2, p. 83–98.

Reinhart CF. 2001. Daylight Availability and Manual Lighting Control in Office Buildings Simulation Studies and Analysis of Measurements. Ph.D. thesis, Technical University of Karlsruhe, Germany.

Tahmasebi F, Schuss M, Mahdavi A. 2014: Monitoring-based dynamically updated occupancy models for predictive building systems control. In: Proceedings of 10th Nordic Symposium on Building Physics, ISBN: 978-91-88722-51-5; 550 – 557.

Zach R, Glawischnig S. Hönisch M, Appel R, Mahdavi A. 2012. MOST: An open-source, vendor and technology independent toolkit for building monitoring, data preprocessing, and visualization. In: Gudnason G, Scherer R, editors. eWork and eBusiness in Architecture, Engineering and Construction. Reykjavik: Taylor & Francis; p. 97-103.

Zio E. 2013. The Monte Carlo Simulation Method for System Reliability and Risk Analysis. 1st ed. London: Springer.