Energy and Buildings 86 (2015) 349–355

Contents lists available at ScienceDirect

Energy and Buildings

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

Predicting people's presence in buildings: An empirically based model performance analysis

Ardeshir Mahdavi, Farhang Tahmasebi*

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

ARTICLE INFO

Article history:
Received 28 July 2014
Received in revised form 6 October 2014
Accepted 14 October 2014
Available online 23 October 2014

Keywords:
Occupancy models
Probabilistic methods
Non-probabilistic methods
Predictive building systems control
Model evaluation
Evaluation statistics

ABSTRACT

Building performance is influenced by occupants' presence and actions. Knowledge of occupants' future presence and behaviour in buildings is of central importance to the implementation efforts concerning predictive building systems control strategies. Specifically, prediction of occupants' presence in office buildings represents a necessary condition for predicting their interactions with building systems. In the present contribution, we focus on the evaluation of a number of occupancy models to explore the potential of monitored past occupancy data towards predicting future presence of occupants. Towards this end, we obtained long-term high-resolution monitored occupancy data from a number of workplaces (in open, semi-open, and closed office settings) in a university building. Using this data, we trained two existing probabilistic occupancy models and an original non-probabilistic occupancy model to predict the occupancy profiles of the same workplaces on a daily basis. The predictions were evaluated via comparison with monitored daily occupancy profiles. To conduct the model evaluation in a rigorous manner, separate sets of data were used to train and evaluate the models. A set of five specific evaluation statistics was deployed for model comparison. In general, the obtained level of predictive accuracy of all models considered was found to be rather low. However, the proposed non-probabilistic model performed better in view of short-term occupancy predictions. The results thus facilitate a discussion of the potential and limitations of predicting building occupants' future presence patterns based on past monitoring data.

1. Introduction

Occupants influence buildings' energy and indoor environmental performance due to their presence (via releasing sensible and latent heat) and actions (operation of devices such as windows, shades, luminaries, radiators, and fans) [1]. To quantify such influences, both empirical and simulation-aided studies have been deployed. For instance, Azar and Menassa [2] observed that energy models of office buildings' in different climatic zones in USA are highly sensitive to occupancy-related behavioural parameters. More specifically, Yang et al. [3] showed that application of HVAC schedules that use observation-based personalized occupancy profiles in a three-story office building test bed could save up to 9% energy compared to the conventional (default) schedules.

Most commonly, representation of occupancy in building performance simulation relies on libraries of standardized diversity factors and schedules. These diversity profiles, which represent typical presence probability profiles, are derived from long term monitored data in different classes of buildings. In this context, multiple efforts are being undertaken to derive more reliable building occupancy profiles. For example, Davis and Nutter [4] used data from different sources (building security cameras, doorway electronic counting sensors, semester classroom scheduling data, and personal observations) to derive occupancy diversity profiles for six types of university buildings. Another study [5] used data obtained from 629 occupancy sensors mounted in an 11-story commercial office building to detail occupancy diversity factors for private offices, open offices, hallways, conference rooms, break rooms, and restrooms. The authors point out that the resulting diversity profiles differ as much as 46% from those published in ASHRAE 90.1 2004 [6], which is referenced by many energy modellers regarding occupancy diversity factors for simulations.

Notably, there have also been a number of efforts in the pertinent scientific and professional communities to develop probabilistic models of occupants' presence in buildings. As one of the first attempts, Newsham et al. [7] considered the probabilistic nature of occupancy while developing a stochastic model to predict lighting profiles for a typical office. Their model deployed the

* Corresponding author at: Department of Building Physics and Building Ecology, Vienna University of Technology, Karlsplatz 13, 1040 Vienna, Austria.
Tel.: +43 1 58801 270 31.
E-mail address: farhang.tahmasebi@tuwien.ac.at (F. Tahmasebi).

http://dx.doi.org/10.1016/j.enbuild.2014.10.027
0378-7788/

probability of first arrival and last departure as well as the probability of intermediate departures from and returning back to the workstations. Reinhart [8] further developed this model by using the inverse transform sampling method [9] 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. The statistical properties of occupancy in single person offices were further examined by Wang et al. [10]. They proposed a probabilistic model to predict and simulate occupancy in single-occupancy offices. In another effort, Page et al. [11] proposed a generalized stochastic model for the occupancy simulation 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. Richardson et al. [12] presented a similar method for generating stochastic weekday and weekend occupancy time-series data with the aid of a matrix of transition probabilities derived from a 10-min resolution monitoring occupancy data for UK households.

Note that in 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 agreement of the generated profiles with the monitored data (one-to-one correspondence between generated and monitored daily profiles) is not taken into consideration. Even while modelling long absences [11], the unoccupied days are scattered randomly across the year and do not necessarily match the dates of absences in the measured data. Hence, the models' performance cannot be said to have been documented (let alone validated), if the actual day-to-day prediction of occupancy and control action probabilities is relevant. Such a short-term predictive function is not a theoretical construct. Rather, it represents an essential scenario in the increasing run-time use of simulation models in the building operation phase, as practiced, for example, in model-predictive and simulation-based predictive building systems control approaches [13–15]. In these scenarios, short-term predictions of occupancy and weather are incorporated in the simulation model to predict the near-future performance of the building towards optimizing its operational regime. Thus, the level of achievable agreement between the predicted and real short-term (e.g., daily) occupancy profiles is of utmost importance.

A further issue regarding the existing probabilistic occupancy models pertains to the provided associated "validation" information. With few exceptions (see for example [16]), most of the work on evaluating 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 assessment. This is especially important while evaluating the predictive potential of an occupancy model, which is intended to be used for operational purposes (i.e., predictive control) in buildings.

Given this background, we pursue in the present study a systematic empirically based assessment of two previously developed probabilistic occupancy models with regard to their short-term predictive performance. To put the predictive performance of these models in context, we compare it with the performance of a simple original non-probabilistic model that also relies on past observation-based aggregated occupancy data to generate daily Boolean patterns of people's presence in buildings. The latter model was developed within the context of two ongoing EU projects [17,18] to be deployed in run-time simulations incorporated in the control logic of existing buildings.

To assess predictive potential of the two probabilistic models and the non-probabilistic one, we used data from a university campus office area, which is equipped with a monitoring infrastructure and includes a number of open and closed offices. Thereby, long-term high-resolution monitored occupancy data from eight workspaces were obtained. To train and evaluate the models, separate sets of monitored data were used. The comparative assessment of the models' predictive performance was accomplished with the aid of a number of pertinent statistics. Thus, the results of the study facilitate a fact-based discussion of the potential and limitations of models for the prediction of people's presence in buildings. Specifically, the results provide a proper basis towards assessing the fitness of probabilistic occupancy models in view of their incorporation potential in the building operation phase (i.e., predictive building systems control).

2. Approach

2.1. Overview

In the present paper, we evaluate the predictive potential of two existing probabilistic occupancy models. Moreover, we compare their performance with a simple original non-probabilistic model of occupants' presence, which was developed to be deployed in simulation-powered predictive building systems control [15]. These models were used to generate predictions of daily occupancy profiles using the past monitoring occupancy data obtained from eight (individually monitored) workplaces in an office area in a building of the Vienna University of Technology. One workplace is within a closed single-occupancy office, two are within semi-closed individual offices, and the rest are within an open-plan office area.

The main objective was to discern how well the models performed towards predicting daily occupancy profiles for the aforementioned eight workplaces. Model training and model evaluation were based on two separate data sets. Once trained, the models were used to predict the daily occupancy profiles of the eight workplaces for 90 working days. To evaluate the two probabilistic methods, multiple predictions were generated via a 100-run Monte Carlo execution. Model comparison was conducted using a set of five statistics and their distribution.

2.2. 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 min 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-min interval data, using stored procedures of the MySQL database [19]. 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 train and compare the occupancy models.

2.3. First probabilistic occupancy model

The occupancy model developed by Reinhart [8] has been primarily used in Lightswitch-2002 model [20] for the purpose of