<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper presented at *The IEEE Power & Energy Society General Meeting, July 26-30, 2015* 2015.

Citation for the original published paper:

Sandels, C., Widén, J., Nordström, L. (2015)

Simulating Occupancy in Office Buildings with Non-Homogeneous Markov Chains for Demand Response Analysis.

In:

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-170865>

Simulating Occupancy in Office Buildings with Non-Homogeneous Markov Chains for Demand Response Analysis

Claes Sandels, Joakim Widén, and Lars Nordström, *Senior member, IEEE*

Abstract—Demand Response (DR) is a promising solution to deal with supply/demand imbalances in the power systems. To appreciate the availability of DR, it is important to include end user behavior in the analysis. In this paper, a model that can generate representative occupancy profiles in single office rooms is presented. The used method is non-homogeneous Markov chain modeling, along with exploratory data analysis of occupancy data, and an estimation of occupancy levels for different months and weekdays. The Markov chain model is calibrated with occupancy sensor data from 24 single rooms collected from an office building floor in Sweden. The simulation results have been statistically compared with a test data set consisting of occupancy data in 23 other rooms on the same floor. It was shown that the model could reproduce key occupancy properties of the test data.

Index Terms—Occupancy patterns, Office buildings, Demand Response, Non-homogeneous Markov chains.

I. INTRODUCTION

Due to an expanding integration of renewable energy resources in the power systems, mismatch between short term electricity supply and demand will increase. A promising solution to deal with this issue is Demand Response (DR), which incentivizes end-users to be flexible in their electricity consumption [1]. The key challenge of the future power system is to allow grid operators and end-users to plan and provide DR services.

A nation's building stock is relevant to analyze when it comes to potential in DR volume. In some European countries, the building stock consume half of the total electricity supply for the year 2012 [2]. Further, office buildings could be a suitable building stock class for providing DR. The reason is two-fold. Firstly, office buildings usually have extensive Heating, Ventilation and Air Conditioning (HVAC) installed [3]. These type of loads are suitable for DR as they can be scheduled (ventilation systems), or can take advantage of the thermal inertia in the building (air conditioners). Secondly, office buildings are often equipped with Building Energy Management Systems (BEMS) [4], which provide technical systems to integrate the consumption processes of the buildings with the electric power systems.

Mainly, two factors affect the energy usage in office buildings: (a) weather dynamics (e.g., outdoor temperatures), and

(b) end-user behavior (e.g., occupancy patterns) [5]. Historically, to take account for (b), reference occupancy profiles have been used [6]. This approach is simplified, and can lead to inaccurate results. It is vital to make the profiles more representative as: (i) presence is a necessary condition for electricity consumption (e.g., usage of lighting and office appliances) [5], (ii) the occupants emit heat and pollutants which affect the indoor environment, and (iii) they have indoor comfort requirements which set operational constraints on the HVAC system.

With the increasing usage of sensors in office buildings, it is possible to collect occupancy data via the BEMS. Several models have been proposed which use such data to improve the modeling of occupancy patterns. In [5], a non-homogeneous Markov chain model is presented which simulates times of arrivals and departures, and the duration of intermediate absence and presence for individuals in single office rooms. Ref [7] has used the same sensor data set as the previous study, but used a genetic programming algorithm to estimate the aforementioned behaviors. The two approaches obtain similar accuracies in comparison with the measured data (around 80 to 83%). A statistical approach to simulate the occupancy in single office rooms is presented in [8]. Here, the authors show that the time duration of vacancies follow an exponential distribution. However, this cannot be verified for time durations of occupancies. A stochastic agent-based simulation model is introduced in [9], which simulates individual occupancy patterns between different zones of an office building. The validation is done for one office room where error rates of 1% are achieved. Lastly, in [10], a data mining approach has been applied, to derive aggregated occupancy diversity factors from sensor data located in different room types of a large multi-tenant office building. These profiles are compared with the reference profiles proposed by [6]. They conclude that the measured occupancy profiles are typically more diverse than the reference profiles.

A. Scope of the paper

The main purpose of this paper is to present a non-homogeneous Markov chain model which can reproduce occupancy patterns for a population of single office rooms. The Markov chain model is calibrated with occupancy sensor data from 24 single rooms collected from a BEMS in an office building floor located in Sweden. The model is statistically checked against test data set from 23 other rooms in the same

Sandels and Nordström are with the Department of Industrial Information and Control Systems, Royal Institute of Technology, Stockholm, Sweden, e-mail: {claess, larsn}@ics.kth.se. Widén is with the Department of Engineering Sciences, Uppsala University, Uppsala, Sweden, email: joakim.widen@angstrom.uu.se.

Fig. 1. A diagram depicting the study setup of the paper

building. As opposed to the models described in literature, this model focuses on modeling aggregated occupancy profiles for a population of rooms. Additionally, a systematic validation of simulation results is performed, which is lacking in the related work. A Markov chain methodology is chosen because it can realistically mimic occupancy patterns on short term [5]. Lastly, the model should be able to produce representative occupancy profiles that can be inputted to building simulation platforms for DR analysis.

B. Outline

The remainder of this paper is structured as follows. In Section II, the methodology is described, along with an overview of the collected occupancy sensor data. This is followed in Section II-B by an exploratory data analysis performed on the sensor data. Simulation results, along with the validation of these, are presented in Section III. Section IV contains a discussion of the results, and some concluding remarks are given in Section V.

II. METHODS AND DATA

The steps of the used methodology is depicted in Fig. 1. Below, each step is presented in more detail. From the BEMS, raw room occupancy sensor data is processed into tidy data [11], so that data analysis can be applied in step 1. In the exploratory data analysis, non-relevant observations are removed, and a further processed data set is generated. At this point, the data is randomly split into two data sets - a training data set, and a test data set. The training data will be used for calculating the Markov chain transition probability matrices in step 2. The generated transition matrix has a dimension of $2 \times 2 \times 1440$. The matrix determines the probability for going to a certain state (occupant/vacant), given the current state

(occupant/vacant). The time resolution between two time slots is 1 minute, i.e. 1440 for one day (see [12] for more details). Step 2 only accounts for people that are present at the office, and not the people that are vacant due to sickness, business trips or holidays etc. Thus, an additional step is implemented which estimates the number of people that are expected to arrive (step 3). In essence, step 3 fits a set of probability density functions with respect to the training data set. ANOVA is used to find significant differences between the mean values of the distributions [13]. The results from step 2-3 are input to step 4, where occupancy patterns are simulated for a given time period. The occupancy patterns give information on how many rooms that is occupant/vacant over the day. In step 5, the simulation results are statistically checked on the test data set. The output are model evaluations, along with graphical measures on how well the model can capture key occupancy dynamics, such as arriving/departure times, and occupancy mobility behavior over the day.

A. The unprocessed office room sensor data

The unprocessed data is collected from an office building floor located in Borås, a city in the south west of Sweden. The tenant is a consultancy company which primarily works with energy related services. The floor consists of 47 single office rooms and one meeting room. Each office room is equipped with a supply air diffuser equipped with an occupancy sensor, a temperature sensor and airflow and air pressure sensors. The occupancy sensor controls lighting and activates or deactivates ventilation in the office. There is one air handling unit serving each office room. The sensor data is event based, i.e. triggered when a occupant arrives or leaves the room. This data was converted into time series of occupant/vacant values with a time resolution of one minute. The data has been collected for the time period between November 2012 and April 2014.

Fig. 2. Boxplots of occupancy ratios with respect to different time horizons (type of day, weekday, and months).

B. Exploratory Data Analysis of Occupancy Data

Temporal attributes of the occupancy data set are considered. In particular, three time horizons are analyzed when it comes to variation in occupancy: (i) day type (weekday/weekend/holiday/bridge day), (ii) ordinary working weekdays, and (iii) seasonal (over months). In Fig. 2, the occupancy ratios are depicted in boxplots for the aforementioned time horizons. As observed in the upper subplot, the variation in occupancy is limited for all days except ordinary weekdays. Thus, it was decided to exclude weekends, holidays and bridge days in the upcoming analysis. For the weekday variation (middle subplot), occupancy is highest on Mondays, and is decreasing over the days. Fridays have the lowest occupancy ratios. Furthermore, in the lower subplot, significant variations can be distinguished on a seasonal basis. Here, the levels peak during the winter months, and where notable drops can be detected in July and August (most likely due to summer vacation times). NB: the seasonal occupancy subplot is generated for ordinary weekdays only.

III. SIMULATION RESULTS AND VALIDATION

The Markov transition matrices are calibrated with occupancy data from 24 randomly chosen office rooms (51% of all rooms), for a time period of 327 weekdays (step 2 in Fig. 1). As no difference in occupancy patterns can be distinguished between weekdays, one matrix is fitted for all days. NB: the reminding occupancy sensor data is used as test data (23 rooms).

ANOVA tests have been performed to deduce whether there are any differences in expected number of arriving people to

TABLE I
SUMMARIZING STATISTICS FOR THE MEASURES β AND R^2 WHICH ARE OUTPUT FROM THE FITTED LINEAR MODELS. NB: QU. = QUARTILE.

Statistics/ Measure	Min	1st Qu.	Median	Mean	3rd Qu.	Max
β	-0.05	0.57	0.78	0.80	1.01	3.97
R^2	0.0%	67.7%	75.7%	71.8%	81.1%	92.5%

the office with respect to weekdays and months, respectively (step 3). From the test results, it is concluded that Mondays are significantly different from the other weekdays ($p < 0.001$). Further, July and August are significantly different the other months ($p < 0.001$), and each other ($p = 0.0495$). Thus, normal distributions are fitted with respect to weekday (Monday or other weekday) and month (July, August, or other month), i.e. six distributions in total. The number of arriving people is randomly generated from the distributions, and input to step 4. In step 4, occupancy ratio data is simulated for the 327 day period with a 1 minute time resolution.

A. Linear model evaluation

For model evaluation, linear regression models [13] are fitted between the simulated and the test data, for each respective day. The linear models will output statistical measures for: (i) model slope coefficient value (β), and (ii) coefficient of determination (R^2). Prior the evaluation, the simulated and the test occupancy data are averaged to 15 minutes.

In Table I, summarizing statistics for β and R^2 estimates from the fitted models are listed. These results show to which degree there is regularity in the simulated and monitored aggregated occupancy patterns that cause them to coincide. Overall, both R^2 and β are high, which indicates that the simulation model reproduces the regular variations in the occupancy patterns well. However, for a number of days these measures are close to zero, i.e., there are significant deviations from the test data. Some of the differences are due to the difficulty of estimating the number of people arriving to the office during vacation times in July and August, but most of them are most likely due to the inherent stochastic nature of the profiles. It is important to note that a perfect linear correspondence is not expected unless a very large number of rooms are aggregated. The Markov chain model is a stochastic process, producing randomly fluctuating occupancy patterns for each room that converge more or less to a mean profile depending on the number of rooms that are aggregated. The measured occupancy data also vary randomly during the day and between days. As the results show, both the real occupancy data and the simulated ones show both regularity and randomness, which is exactly the property we want to reproduce.

B. Statistical properties of the simulated and test data

When it comes to simulating occupancy patterns there are a number of statistical properties that are relevant, such as average occupancy profiles, arriving and departure times, etc.

Fig. 3. Plots of the statistical properties of the simulated and test occupancy data, respectively. The blue solid curves are the simulated data, and the red dashed curves are the test data.

These properties are compared between estimated and test data in probability distributions plots.

In subplot A of Fig. 3, averaged daily occupancy profiles for the whole time period are plotted. As seen, the two profiles correlate well. However, the occupancy magnitude is generally higher for the simulated data. Further, a probability distribution plot for quarterly occupancy ratios is shown in subplot B. The simulated data tends to underestimate the probability for low occupancy ratios (below 20%), and overestimates higher occupancy ratios (over 30%). This implies that the real occupancy profiles are more variable than what the Markov model suggests.

The remainder of the subplots shows probability distributions for various behavioral attributes at room level with a 1 minute resolution. For times of arrival (subplot C) and departure (subplot D) to/from the office, the general patterns are covered by the model. However, the model tends to underrate early departures of the occupants. Moreover, subplot E shows the time duration of continuous occupancy. The two curves correlate well, with the exception of higher frequencies of shorter occupancy durations for the test data (under 1 hour). Subplot F depicts the time stamps for when the individuals move from a vacant to an occupant state (arrival times excluded). As seen, the shape of the curves is very similar. Subplot G and H show the same properties as the previous plots, but for vacancies instead. Here, the model is poor on reflecting shorter time durations of vacancy, i.e. under 1 hour (subplot G). However, the time stamps for going from an occupant to a vacant state (departure times excluded) is accurately reflected by the model (subplot H). Essentially, the analysis confirms that the model is not subject to overfitting as it can capture key properties of the test data.

In Table II, summarizing statistics for the distributions shown in Fig. 3 are listed. The statistics are organized ac-

cording to the figure (i.e., from A to H), for the simulated and measured data, respectively.

IV. DISCUSSION

As shown, it is possible to estimate representative occupancy profiles in office rooms with the presented Markov chain model. The estimated profiles could be used for simulating the dynamics of the indoor climate in an office building, along with a comfort requirement of the occupants. NB: different comfort requirements can be used whether the rooms are occupied or not (stressing the importance of reflecting real occupancy behaviors) [3]. By performing such simulations, it is viable to assess how much demand flexibility that is available for DR from the HVAC system. For example, a grid operator might want to decrease the peak load in his network by curtailing the cooling load in an office building located in his network. To execute this operation, an assessment of both the operational status of the cooling load and how the curtailment affects the indoor temperature are required. As occupancy affects the indoor climate and the HVAC control, it is necessary to include this dynamics in such simulations.

Moreover, the model is calibrated with occupancy data from an office building floor in Sweden. Evidently, this leads to questions about the general usability of the results. The authors believe that additional validation must be done with occupancy sensor data from other office buildings to be certain that the actual behavior is captured. However, the presented model is considered to be generic, and can be used for such a validation. The fact that the model only estimates occupancy based on state of the prior time step (occupant/vacant), could be a potential source of deviation in simulation results. In reality, occupancy behavior may be dependent on events that have occurred in earlier time steps as well. In addition, the model assumes that occupancy is independent from the behavior in

TABLE II
SUMMARIZING STATISTICS FOR THE VARIOUS PROPERTIES OF THE SIMULATED AND THE MEASURED OCCUPANCY DATA, RESPECTIVELY.

Statistics/Measure	Min	1st Qu.	Median	Mean	3rd Qu.	Max
A. Simulated daily profiles [%]	0.00	0.23	1.21	12.94	30.71	40.13
A. Measured daily profiles [%]	0.02	0.25	1.12	11.12	26.66	34.33
B. Simulated 15 min occupancy ratios [%]	0.00	0.00	0.00	12.94	26.92	73.08
B. Measured 15 min occupancy ratios [%]	0.00	0.00	0.00	11.08	20.00	88.00
C. Simulated arriving times [hh:mm]	00:09	09:08	09:52	10:08	10:50	19:44
C. Measured arriving times [hh:mm]	00:02	09:11	09:55	10:29	11:17	22:09
D. Simulated departure times [hh:mm]	08:48	16:21	17:29	17:14	18:22	23:58
D. Measured departure times [hrs:min]	01:51	15:22	17:21	16:35	18:13	23:59
E. Simulated time length occupancy [hrs]	0.02	0.62	1.38	1.81	2.57	12.58
E. Measured time length occupancy [hrs]	0.03	0.42	1.28	1.92	2.77	13.58
F. Simulated time stamp occupancy [hh:mm]	00:09	10:26	12:52	12:52	15:11	23:50
F. Measured time stamp occupancy [hh:mm]	00:02	10:29	12:49	12:52	15:09	23:39
G. Simulated time length vacancy [hrs]	0.02	0.50	1.22	1.61	2.28	13.38
G. Measured time length vacancy [hrs]	0.02	0.20	0.67	1.22	1.77	16.22
H. Simulated time stamp vacancy [hh:mm]	00:28	11:59	14:13	14:19	16:53	23:58
H. Measured time stamp vacancy [hh:mm]	00:09	11:56	14:11	14:16	16:47	23:59

the other office rooms. Evidently, the occupants can also have a collective behavior, e.g. arrange joint meetings, lunches etc. These types of dependencies are not covered by the model, and will increase the error rate. However, as we want to capture representative occupancy curves with a simplified model, it is believed to be a disadvantage to include such complex interactions in the model.

By estimating the number of expected arriving employees on a weekly basis, instead of a monthly resolution, the accuracy of the simulations can be increased. This is because a monthly resolution cannot accurately reflect when people go on vacation during the summer time. The simulations for summer time obtain the lowest accuracies. If the timing of vacation could be better covered, the performance of the model will improve. However, two drawbacks are introduced with this approach: (i) the model becomes more complex, and (ii) the results may be overfitted against the studied office building, i.e. the occupancy behavior might not be representative for other office buildings.

The authors are currently developing a model which can simulate electricity consumption profiles in office buildings, where the proposed Markov chain model generates occupancy profiles to these simulations. The model connects simulated occupancies with electricity consumption from appliances (lighting and IT equipment), and their effect on the indoor climate and the operation of the HVAC system. The model simulations will be validated on real power measurements from an aggregation of office buildings.

V. CONCLUSIONS

In this paper, a model used for reproducing occupancy profiles in single office rooms has been presented for further use in electricity demand models. The occupancy profiles were estimated through a non-homogenous Markov chain method. The Markov transition probability matrices were estimated with occupancy sensor data from 24 single rooms in a Swedish office building. The simulation output was statistically compared with occupancy data from 23 other rooms in the same building. It was shown that the model could reproduce key-properties of the real occupancy.

REFERENCES

- [1] R. Belhomme and et al, "Deliverable 1.1 address technical and commercial conceptual architectures," tech. rep., EDF, et al, 2009.
- [2] T. S. E. Agency, "Energiläget 2013," tech. rep., The Swedish Energy Agency, 2013.
- [3] F. Oldewurtel, D. Sturzenegger, and M. Morari, "Importance of occupancy information for building climate control," *Applied Energy*, vol. 101, no. 0, pp. 521 – 532, 2013. Sustainable Development of Energy, Water and Environment Systems.
- [4] N. Djuric and V. Novakovic, "Identifying important variables of energy use in low energy office building by using multivariate analysis," *Energy and Buildings*, vol. 45, no. 0, pp. 91 – 98, 2012.
- [5] 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, no. 2, pp. 83 – 98, 2008.
- [6] A. Standard, "Energy standard for buildings except low-rise residential buildings," tech. rep., American Society of Heating Refrigerating and Air-Conditioning Engineers, 2004.
- [7] T. Yu, "Modeling occupancy behavior for energy efficiency and occupants comfort management in intelligent buildings," in *Machine Learning and Applications (ICMLA), 2010 Ninth International Conference on*, pp. 726–731, Dec 2010.
- [8] 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.
- [9] C. Liao and P. Barooah, "An integrated approach to occupancy modeling and estimation in commercial buildings," in *American Control Conference (ACC), 2010*, pp. 3130–3135, June 2010.
- [10] C. Duarte, K. V. D. Wymelenberg, and C. Rieger, "Revealing occupancy patterns in an office building through the use of occupancy sensor data," *Energy and Buildings*, vol. 67, no. 0, pp. 587 – 595, 2013.
- [11] H. Wickham, "Tidy data," *Journal of Statistical Software*, vol. 59, 9 2014.
- [12] J. Widen and E. Wackelgard, "A high-resolution stochastic model of domestic activity patterns and electricity demand," *Applied Energy*, vol. 87, no. 6, pp. 1880 – 1892, 2010.
- [13] R. Warner, *Applied Statistics - From Bivariate Through Multivariate Techniques*. SAGE, 2007.
- [14] SweGRIDS. <http://www.kth.se/en/ees/omskolan/organisation/central/swegrids>, October, 2013.

ACKNOWLEDGMENT

The work has been carried out under the auspices of SweGRIDS [14]. We want to acknowledge Sofia Stensson and Enar Andersson at SP Technical Research Institute of Sweden for assisting us with the occupancy sensor data.