Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Room-level occupancy simulation model for private households

To cite this article: Sebastian Wolf *et al* 2019 *J. Phys.: Conf. Ser.* **1343** 012126

Room-level occupancy simulation model for private households

Sebastian Wolf[1], Davide Calì[1], Maria Justo Alonso[2], Rongling Li[1], Rune Korsholm Andersen[1], John Krogstie[2] and Henrik Madsen[1,2]

[1] Technical University of Denmark, 2800 Lyngby Denmark
[2] Norwegian University of Science and Technology

E-mail: sewo@dtu.dk

Abstract. This work presents a novel occupancy simulation model for residential buildings. The main contribution is that occupancy is simulated at room level, as opposed to more course spatial resolutions in previous studies. The model is based on a time-use survey conducted in Denmark including several thousand households. It is formulated as an inhomogeneous hidden Markov model. The simulated occupancy profiles take into account the variables week day, time of day, occupant age and family type. Preliminary results show that they are in good agreement with the measurements.

1. Introduction

In the field of building energy simulation there has been an increased attention in recent years to the modelling of occupant behaviour. The reason for this is that it has been generally acknowledged that the behaviour of the occupants has a significant influence on the energy performance of a building. This influence materialises in substantial differences in energy consumption of equally designed dwellings [1, 2], which becomes a challenge for predictions of the building energy use. The term "building performance gap", which has become a familiar expression in the field of building energy simulation, describes a mismatch between estimated and real energy consumption in buildings. This mismatch has been at least partly ascribed to inaccurate or oversimplified representations of occupant behaviour in the simulation tools [3]. It is therefore crucial to find models that realistically reflect the occupant behaviour and also the diversity among occupants [4, 5]. One aspect of occupant behaviour is the presence or more generally the location of the occupants. In building energy simulations, it is important to represent the occupants' location for two reasons. First, the metabolic heat gain from the occupants has an effect on a room's energy balance. The influence of these heat gains becomes bigger with increasing building insulation standards, as the share of metabolic heat gains on the total spatial heating increases. Secondly, presence is a necessary condition for many occupant actions which have an affect on the energy consumption. Examples in residential buildings are window opening, thermostat adjustment, cooking, use of electronic devices such as ICT appliances or washing machine and tumble dryer. A number of occupant presence (occupancy) models has been presented in recent years. Most of them are probabilistic models based on measurements [6, 7, 8, 9, 10]. The majority of occupancy simulation models in literature can be assigned to either models for office buildings or households. The number of office building models is higher. This can possibly be ascribed to the fact that a monitoring is easier to install in office buildings (e.g. universities) and that occupancy patterns in offices are more straight forward than in households.

The literature of domestic presence models is more scarce. [7] present a domestic occupancy simulation model based on UK time-use data. The model consists of an inhomogeneous Markov chain whose states represent the number of active (not asleep) occupants in the household. The model's configuration includes the distinction of weekend and weekday, and the household size. [11] use a different approach, as in their model each occupant is individually represented by a Markov chain. They present a 3-state occupancy model with the states "active", "inactive", "absent" for the use of simulating domestic lighting demand. The model is based on a Swedish time-use survey (TUS) and comprises four different inhomogeneous Markov chains; one for each combination of weekdays and weekend days, and apartments and detached houses. In [12], this approach is modified by differentiating the "active" state into seven states related to electricity consumption (cooking, dishwashing, washing, TV, computer, audio, other). [13] propose a similar approach as [11] based on a Belgian TUS using the same three states ("active", "inactive", "absent"). In addition to the aforementioned work, the authors apply a hierarchical clustering method to define six distinct behavioural occupancy types, each of which can be simulated separately. [14] present a 3-state occupancy model with some additional features. They use a higher-order Markov model that takes into account the elapsed time of the current state. The authors argue that this improves the simulated activity durations as compared to an ordinary Markov chain model. Moreover, they link occupants that have a high chance of correlation to each other (e.g. spouses, parents, children) by representing the occupants by one combined Markov chain, as done by [7].

None of the aforementioned works attempts to simulate occupancy on a spatial resolution finer than household. In the present work, an occupancy simulation model on room-level is presented. The states of the model represent an occupant's location which is either a room within the household or being not present in the dwelling (out). The model is based on a Danish TUS from 2008/09. Since only activities but no locations are recorded in this data set, a hidden Markov model (HMM) is applied to make statistical inference on the locations based on the observed activities. The transitions between states follow an inhomogeneous Markov chain to respect the diurnal rhythm of the occupants. Furthermore, the model differentiates between weekdays and weekend days, and between six occupant categories which were identified based on demographic features and household characteristics.

2. Data description

The data that was used in this work to train the model stems from a TUS conducted in Denmark between March 2008 and March 2009 [15]. After filtering missing data, the survey included 4344 families with a total number of 11341 individuals, of which 8201 have written an activity diary. Most of the individuals who conducted a diary did this for one weekday and one weekend day. The occupant age varied from 4 years to 85 years (mean: 52.7, median: 56). In the diaries, every 10-minute interval was filled with one of forty predefined activities. Direct information about the location of the occupants was not included in the data set. Furthermore, activities which did not occur at least once in the diary for at least 80% of the individuals were regarded as untypical. Diary-person-days (a sequence of 24 h for one person) that included these untypical activities were removed. After this filtering, 28 activities and 4877 diary-person-days remained.

3. Methods

This section gives a brief overview of the employed methods in this work. In Section 3.1 and Section 3.2, ordinary and hidden Markov chains are defined, respectively.

3.1. Markov chains

For a finite set $M = \{1, 2, \ldots, m\}$, an m-state Markov chain is a sequence of M-valued random variables $\{X_t\}$ with

$$P(X_t|\mathbf{X}^{(t-1)}) = P(X_t|X_{t-1}), \tag{1}$$

where $\mathbf{X}^{(t-1)} = X_{t-1}, X_{t-2}, \ldots, X_0$ is set of variables of $\{X_t\}$ up to time $t-1$. Equation (1) is referred to as the Markov property. In words, the probability of the current time step, conditioned on the entire history, depends only on the previous time step. The matrix of conditional probabilities between two states $i, j \in M$,

$$\Gamma_{i,j}(t) = P(X_t = i|X_{t-1} = j) \tag{2}$$

is called *transition probability matrix* (TPM). If Γ does not depend on the time t, the process is called a homogeneous Markov chain. Otherwise it is called an inhomogeneous Markov chain.

3.2. Hidden Markov chains

A hidden Markov chain is a probabilistic model that consists of two components: An observed sequence $\{Y_t\}$ and an unobserved Markov chain $\{X_t\}$. It is assumed that Y_t only depends on the current state X_t but not on its own history $\mathbf{Y}^{(t-1)}$.

$$P(Y_t|X_t, \mathbf{Y}^{(t-1)}, \mathbf{X}^{(t-1)}) = P(Y_t|X_t) \tag{3}$$

The distribution of $Y_t|X_t$ is referred to as *response distribution*. The parameters in a HMM are given by the set $\{A, B, \pi\}$, where A and B are sets of parameters that correspond to the TPM and the response distribution, respectively. The parameter vector π is the distribution of X_0, that is, the distribution of the unobserved state in the initial time step. The Baum-Welch algorithm, which is based on the maximum likelihood estimation principle, can be used for parameter estimation [16, 17]. In addition to parameter estimation, in the context of HMM, one is often interested in the most likely sequence of unobserved states for a given sequence of observations. This sequence is referred to as *global decoding* and an efficient way to calculate it is the Viterbi algorithm [16].

3.3. Description of the simulation model

In the present model, we define $m = 5$ states that correspond to the occupants' locations, namely {sleeping room, kitchen, bathroom, living room, out}. Further, there are $n = 28$ different activities the occupants can follow, e.g. {sleep, work, relaxing, eating, ...}. Let $\{Y_t\}$ be a sequence of observed activities and $\{X_t\}$ the corresponding unobserved sequence of locations. In order to define a HMM, we need to define a set of parameters corresponding to the transition probabilities, the response distribution and the initial distribution, respectively, i.e. $\{A, B, \pi\}$.

Response distribution We define the conditional distribution of the locations given the activities, $P(X|Y)$ for example as in Table 1. Using Bayes' theorem, the response distribution $P(Y|X)$ can then be obtained by:

$$P(Y|X) = \frac{P(Y)}{P(X)} \cdot P(X|Y), \tag{4}$$

where $P(Y)$ can be calculated as the relative frequency of each observed activity, and $P(X)$ is a normalizing factor.

Transition probabilities The transition probabilities in the presented model are dependent on the time of day. Hence, the underlying Markov chain is inhomogeneous. To represent this time-dependency, the four basis vectors of a B-spline basis with a periodicity of 24 h are used as covariates to the transition probabilities [10]. The use of splines has several merits in comparison to the calculating a TPM for each time step. It results in smooth transition probability curves and avoids unwanted spikes in the transition probabilities that derive from lack of observed transitions in the data rather than from real changes in the probability over time. Moreover, it drastically reduces the number of parameters and with this the calculation costs of the parameter estimation. As is a common approach, we use a logit function as link between the unbounded space of the linear covariates and the bounded space of the probabilities. The transition probabilities are given by:

Table 1. Conditional distribution of States $P(X|Y)$

activity	sleeping room	kitchen	bath room	living room	out
sleep	1.00	0.00	0.00	0.00	0.00
work	0.00	0.00	0.00	0.00	1.00
eating	0.00	0.50	0.00	0.30	0.20
cleaning	0.25	0.25	0.25	0.25	0.00
...	...	...	...	...	...

$$\Gamma_{i,j,t} = \frac{\exp(\boldsymbol{x}_t\,'\boldsymbol{a}_{ij})}{\sum_{h=1}^{N} \exp(\boldsymbol{x}_t\,'\boldsymbol{a}_{ih})}, \quad i,j \in \{1,\dots,m\} \tag{5}$$

with the spline basis $\boldsymbol{x}_t = (x_{t1}, x_{t2}, x_{t3}, x_{t4})$ at time t and parameters $\boldsymbol{a}_{ij} = (a_{ij1}, a_{ij2}, a_{ij3}, a_{ij4})$.

3.4. Categorisation of occupants and days

In order to respect differences in behavioural patterns, the occupants were categorised in the following six groups: (i) children (younger than 18 years), (ii) single adults, (iii) adults living with partner, (iv) adults living with partner and children, (v) single seniors (older than 66 years) and (vi) seniors living with partner. Of course, other constellations are possible, but are disregarded for the sake of simplicity. The gender of the occupants was not considered. Furthermore, the model distinguishes between four categories of days: Fridays, Saturdays, Sundays and weekdays (Monday to Thursday). This categorisation of days was chosen over a more simple distinction in weekdays and weekend, because the days Friday, Saturday and Sunday each have unique occupancy patterns. For instance, on Fridays many people leave work earlier and go to bed later, as they do not have to work next morning. On Sundays, people might go to sleep earlier than on Saturdays, as they have to work on Monday morning. In total, this results in 24 combinations for each of which one TPM was fitted.

4. Results and Discussion

This section describes the simulation results using the above presented model. Figures 1, 2 and 3 are all based on simulations of occupant category (iv), i.e. adult living with partner and children, on a weekday. Figure 1 show the averages time spent in each state based on one thousand simulations. One can see that the occupant gets out of bed no earlier than 6 a.m. and goes to bed no later than 11 p.m. in about 80 % of the cases. Between 9 a.m. and 3 p.m. the probability, that the occupant is out, is higher than 50 %. Around 6 p.m. there is a high chance to find the person in the kitchen, presumably preparing/having dinner. Thereafter, in the evening, the occupant spends most of the time in the living room or out before going to bed. Most of the bathroom visits occur either in the morning or in the evening. Figure 2 visualises the time spent for each activity. Activities that are in the response distribution associated with a probability of 1 to take place inside the house are coloured in red tones. Activities with a probability between 0 and 1 are turquoise, and activities with probability 0 are greyed out. The black reference line in Figure 2 indicates the time spent inside the house based on the simulations shown in Figure 1. Since this line is mainly running through the turquoise area, one can see that the activities and states are in good agreement. Figure 3 shows one hundred simulated days. As opposed to Figures 1, this graph shows the variability (uncertainty) of the created profiles. On almost every day, the occupant visits the bathroom for a short period after getting out of bed and before going to bed. In the mornings, this is mainly followed by going to the kitchen or going out. There is a fairly regular pattern of leaving the house between 8 a.m. and 4 p.m., followed by period in the kitchen and living room in the late afternoon. Figure 4 shows, as an example, an occupancy profile of a four-person household on three weekdays. The household profile consists of four cumulated single profiles, two of category (i) and two of category (iv), i.e. two children and two adults. It is assumed that the children and the adults (parents) share one bedroom, respectively. One can see that the kitchen is mainly occupied in the morning and evening. There is a peak in the living room around 9 p.m. In none of the profiles, the night sleep is interrupted by a transition to another state which indicates that the transition probabilities respect the occupants' diurnal rhythm. In summary: based on the chosen response distribution, the simulation results of the occupant locations are in agreement with the observed activities. For a thorough validation process, ground truth data of room-level occupancy is required.

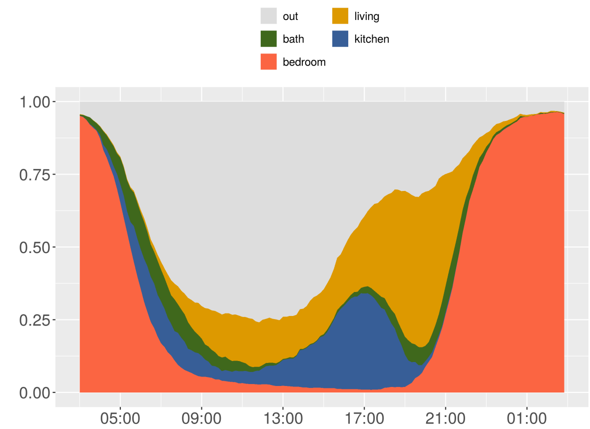

Figure 1. State distribution.

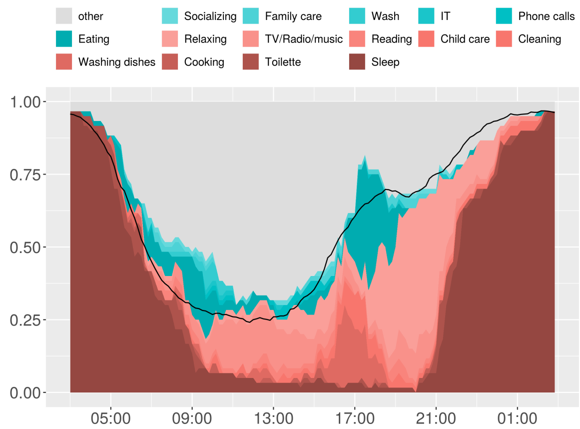

Figure 2. Activity distribution.

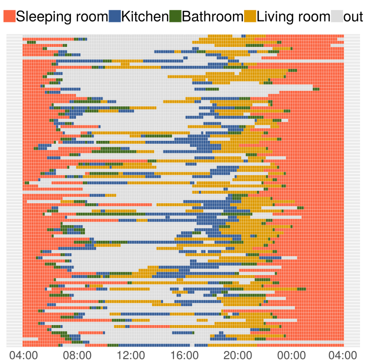

Figure 3. Simulations.

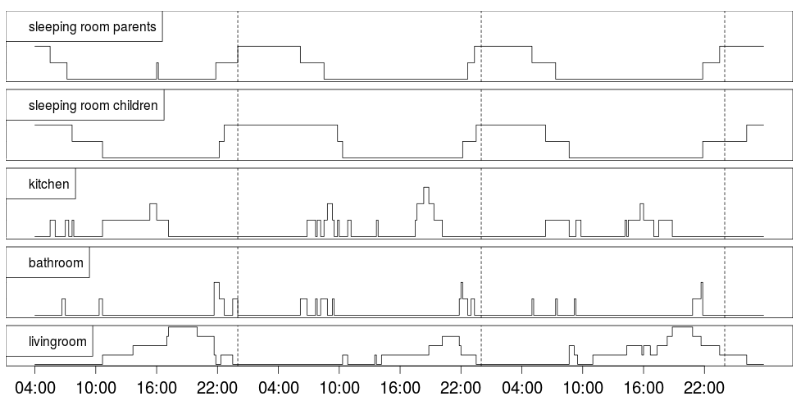

Figure 4. Household profile.

4.1. Limitations
The presented model has a few limitations: a) The model was trained based on Danish TUS data. Hence, it reflects the behavioural patterns of the Danish population. It is not clear to which extent this can be transferred to other countries. b) Ideally a Markov chain is trained on longitudinal data. However, in this case 24-hour diaries were used. Hence there is a cut in the observations after each 24-hour period. Fortunately, this cut occurs at 4 a.m. when almost all occupants sleep leading to a smooth transition between diaries. c) Occupants are simulated independently from each other. However, in reality the behaviour of occupants of the same household are correlated. For instance, the occupants may eat together and go to sleep at the same time. d) The distribution of an activity's location (Table 1) was chosen intuitively. It can be updated once more evidence is available. e) A dwelling can possibly have zero or more than one instances of each room type. For instance, it can have no living room or two bathrooms. This is not reflected by the model.

5. Conclusion and Outlook
To the best of our knowledge, for the first time, a room-level occupancy simulation model for private households was presented. The model is formulated as a hidden Markov model that makes inference of the occupants' location based on observed activities of a Danish time-use survey. The categorisation of occupant types makes it possible to simulate profiles of single dwellings or larger scales such as districts. The simulation results are plausible and reflect the underlying activity data well. However, a thorough validation of the model has not been done yet. As discussed in the previous section, the model has certain limitations which should be addressed. In particular, the correlation between occupants of the same household is left for future research.

6. Acknowledgements
The authors would like to thank Dr. Jens Bonke and The Rockwool Foundation for providing the Danish TUS data. This work has partly been carried out as a part of the CITIES, FME ZEN, and SCA projects. The authors are thankful to Innovation Fund Denmark, EU Interreg V programme, and the Norwegian Research Council and ZEN partners for the support.

References
[1] Andersen R K 2012 The influence of occupants' behaviour on energy consumption investigated in 290 identical dwellings and in 35 apartments *Proceedings of Healthy Buildings, Brisbane, Australia*
[2] Roels S, Bacher P, Bauwens G, Castano S, Jose Jimenez M and Madsen H 2017 *Energy and Buildings* **153** 179–189
[3] Azar E and Menessa C C 2012 *Energy and Buildings* **55** 841–853
[4] Liisberg J, Møller J, Bloem H, Cipriano J, Mor G and Madsen H 2016 *Sust. Cities and Society* **27** 83–98
[5] Haldi F, Cali D, Andersen R K, Wesseling M and Mueller D 2016 *JBPS* **10** 527–544
[6] Page J, Robinson D, Morel N and Scartezzini J L 2008 *Energy and Buildings* **40** 83–98
[7] Richardson I, Thomson M and Infield D 2008 *Energy and Buildings* **40** 1560–1566
[8] Liao C, Lin Y and Barooah P 2012 *Journal of Building Performance Simulation* **5** 5–25
[9] Wang C, Yan D and Jiang Y 2011 *Building Simulation* **4** 149–167
[10] Andersen P D, Iversen A, Madsen H and Rode C 2014 *Energy and Buildings* **69** 213–223
[11] Widen J, Nilsson A M and Wackelgard E 2009 *Energy and Buildings* **41** 1001–1012
[12] Widen J and Wackelgard E 2010 *Applied Energy* **87** 1880–1892
[13] Aerts D, Minnen J, Glorieux I and Descamps F 2014 *Building and Environment* **75** 67–78
[14] Flett G and Kelly N 2016 *Energy and Buildings* **125** 219–230
[15] Bonke J 2016 *International Journal of Time Use Research* **13** 81–108
[16] Zucchini W, McDonald I and Langrock R 2016 *Hidden Markov Models for Time Series - Second Edition* (CRC Press) ISBN 9781482253832
[17] Rabiner L 1989 A tutorial on hidden markov models and selected applications in speech recognition *Proceedings of the IEEE, Vol. 77* pp 257–286