Investigation of A Short-term Prediction Method of Occupancy Presence in Residential Buildings

Zhaoxuan Li[1], and Bing Dong[1]
[1]Department of Mechanical Engineering,
University of Texas at San Antonio, San Antonio, U.S.A

Abstract

Occupancy models are essential parts of the control operations of smart buildings. Developing an appropriate model to simulate and predict the occupancy presence for common buildings will allow a better optimization of energy consumption and utility bill. However, most occupancy models tested by the on-site data only concentrate on the commercial buildings. On the other hand, the residential studies simulate mostly Time Use data. The applicability of those models are largely unknown for the real-time residential environment. This study focuses on providing a unique data set of four residential houses collected from PIR sensors in the U.S. One new inhomogeneous Markov model for predictive building controls is proposed and compared to the homogeneous model. Optimized training periods for the occupancy presence prediction are decided individually from change-point analysis of historical data. Then, various scenarios that utilize 15-minute data to forecast 15-min ahead, 1-hour ahead, and 24-hour ahead occupancy presences are presented. Both room-level and house-level predictions are evaluated. The one-to-one matching between the prediction and the ground truth demonstrates the model performance. The analysis of the residential result validates the effectiveness of the proposed method and provides more realistic analysis on the occupancy forecasting of the residential environment.

Introduction

Energy consumption of buildings accounts for around 40% of the energy used worldwide. Meanwhile, buildings are increasingly expected to meet higher and potentially more complex standards on the sustainability, comfort, grid-friendly, and yet to be built and maintained economically. It has become a widely accepted idea that smart controls during building operations can yield substantial savings on the energy consumed by buildings and help to achieve those high standards (Lazaroiu, 2012). Many complex models (building load dynamics, grid networks, building-grid coupling control, etc) need to be developed and on-site tests need to be addressed as well. The detection and prediction of the occupancy profile of the building indoor environment is one of the critical parts toward the success of the model predictive building controls.

The occupancy in general is less varied in commercial buildings. The occupants usually have strict working schedules and limited environmental controls from morning until evening, Monday through Friday (Gunay,2013). The workers normally stay at a certain place, namely, the assigned cubicle, without much movement. The absence such as the lunch break and group meetings can be treated as regular tasks, which are highly predictable. Meanwhile the occupancy of the residents living in residential houses has larger uncertainty. For example, the indoor climate controls are more accessible to residential users (Dong, 2015). There are also more irregular activities of the residential occupants, such as watching TV, showering, shopping, etc. The diversity of the living habits of the individual resident could contribute more stochastic occupancy pattern during each period of different days with multiple people in a house consisting of multiple rooms.

Nowadays, occupancy sensing is becoming more and more accurate comparison to the early prototypes (Haq, 2014). The smart homes are able to maintain the indoor comforts in the current time step when the occupants are present and save energy when the occupant is absent. However, those homes require a predictive optimization that several time-step ahead operations will need to be scheduled (Halvgaard, 2014). Hence, the occupants' comforts should not only be maintained in the current period but also need to be addressed from a future perspective. Therefore, how to utilize the information from occupancy sensing to forecast the future occupancy states from an occupancy model is a more urgent topic for the occupancy-based smart building controls.

Occupancy model in building simulation tools is conventionally represented in terms of static schedules. Although the static schedule as an occupancy model has many shortages in nature, it is still applied to many simulations owing to the stable performance in the office buildings (Mahdavi, 2015). However, the fixed schedule should be carefully used depending on the problems, since it is unlikely too generic enough to obtain the solutions in a residential building environment. In the contrast, the stochastic occupancy models aiming to simulate the realistic stochastic pattern of the occupancy are suitable for the control of the residential house.

https://doi.org/10.26868/25222708.2017.242

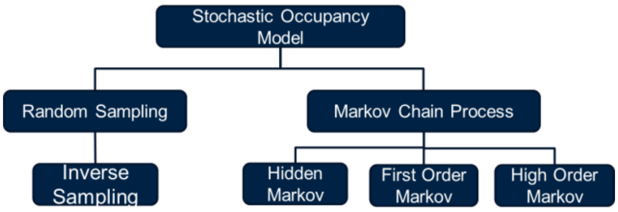


Figure 1: State-Of-the Current Stochastic Occupancy Models..

One popular stochastic model is the random sampling approach (Parys, 2011). The occupants' presence can be generated from a sampling process of the fitted distribution functions from historical data. The cumulative distribution functions (CDFs) and probability distribution functions (PDFs) contain the occupancy information in terms of the average occupied hour per day, the average vacancy ratio per day, the average departure and arrival time per day, and the associated deviations. To predict the daily occupancy from the fitted CDFs and PDFs, the process first identifies the first arrival time and last departure time using the inverse sampling from the CDFs. Then, the time of the intermediate activities to occur and how long the actives will be, can be determined last by comparing a pseudorandom number against to the intermediate departure PDF. There are two issues that need to be addressed if this approach is used for the residential samples. First, the arrival time and last departure time in house level for residential houses are more complex to model than the office ones. The high possibility to produce accurate predictions of the first arrivals and last departures for offices is due to the enforced working schedules of the employees, which usually do not exist for residential residents. People can arrive and leave from their homes, as they desire. The second issue is that there are too many irregular intermediate activities for residential occupants and the routine modelling approach for regular intermediate activities such as the lunch break needs to be modified to guarantee accurate predictions for residential occupancy.

The Markov chain is another stochastic process that has been applied in stochastic occupancy modelling. Different algorithms have been explored using both homogeneous and inhomogeneous Markov chains. Such studies include reproducing the Time-Use Survey data using integrated Markov Montel Carlo technique (Richardson, 2008), estimating the transitional probabilities of presence and absence by utilizing the inverse function method to the inhomogeneous model (Page, 2008), and generating the occupancy using a hierarchical approach that combines homogeneous model with the occupants' moving events at different locations of the offices (Wang, 2014). For all first-order Markov chain models mentioned above, the key assumption is that the only state that matters is the state previous to the predicted state. However, there may also be a habitual sequence that connects several occupancy states together. The work done by Wkike (Wilke, 2013), using a higher-order Markov model, gives an insight of the implementation of this idea. The transitional matrixes were calculated by an integration of the several past occupancy states that are assumed to contribute to the current transition probabilities. Meanwhile, a Markov chain of higher order has more complex calculations and non-bimodal characters than a first-order model. Another type of the Markov model, hidden Markov chains, introduces an additional state (the hidden state) to incorporate the exogenous measurements to enhance the accuracy to predict the occupancy (Ai, 2014; Dong, 2014). Since the model is developed as a double embedded stochastic process, the occupancy states in the underlying stochastic process are hidden. The model treats the observation from measurement as a probabilistic expression of the hidden state. The model can approximate the hidden state, the occupancy, in the hidden state level by the different combination of indoor environmental measurements.

The successful utilization of these models, on one hand, depends on the inputs. Exogenous measurements (e.g. the CO_2 level) other than the occupancy information required by specific model, namely the hidden Markov chain, could be available for certain buildings. However, these measurements may be difficult to get for other buildings. On the other hand, the complexity of the model in terms of theoretical development and computational efficiency could prohibit the popularity of their field applications in building systems, such as higher order Markov chain. Besides the issues with the models themselves, the majority of these models are normally used to generate random non-repeating daily profiles as a specific estimation problem for the building simulations rather than the real-time predictions for the on-site applications. They are "validated" if the occupancy is simulated in a statistically reasonable way. In other words, the models are matching the tendency between the simulated occupancy and the monitored data in terms of the average arriving time, the average occupied durations, and the average leaving time. They are not evaluated by an exact one-to-one correspondence of the predicted occupancy to the ground truth in each time step, which is utterly important in the advanced control application, namely the model predictive controls (MPC).

Given this state of the art in the occupancy presence modelling, authors pursue in this study a systematic approach and assessment of the short term predictions of occupancy presence profiles of the residential buildings for the purpose of better MPC. One new method was developed for application purposes, an inhomogeneous Markov process integrated a moving window forecast optimized by change-point analyses. It is compared to a traditional homogeneous model by using the same input. Diverse occupancy data from different residential houses are tested. Predictive performance is evaluated based on one-to-one matching, also known as the correctness ratio calculated by summing of the true positive (predicted presence while the observations of the room are occupied) and the negative positive (predicted absence while the observations of the room are not occupied).

Methodology

The stochastic models defined in this study are those that predict the occupancy using probabilistic theories. They are expected to provide explicit mathematical explanations for the variation of the presence pattern. They are also designed to quantify the randomness of the data and make predictions solely by the historical occupancy states. Two forms of the first-order Markov models are introduced in this section. The key assumption of the first-order chain is that the imminent state is only influenced by the present state, not the previous past states (called Markov property).

Let a Markov chain X at time step k be a sequence containing variables $x_1, x_2, \ldots, x_k$ and the observed set of occupancy states is $S = \{s_1, s_2, \ldots, s_n\}$ where $n \leq k$. The chance of the chain to move from the state s_i to the state s_j at time step $k+1$ is decided by the transitional probability defined as:

$$P_k^{ij} = p(x_{k+1} = s_j | x_k = s_i) \quad (1)$$

Where
$p(x_{k+1} = s_j | x_1, x_2, \ldots, x_k) = p(x_{k+1} = s_j | x_k = s_i)$.
Usually, there are two types of chains based on the time homogeneity of the transitional probability defined from Equation (1). Time homogeneity implies that there is no change in the underlying probability of the transition between the same pair of the states as time goes on:

$$p(x_{k+1} = s_j | x_k = s_i) = p(x_k = s_j | x_{k-1} = s_i) \quad (2)$$

The inhomogeneity implies that there is a change between the transition as time goes on:

$$p(x_{k+1} = s_j | x_k = s_i) \neq p(x_k = s_j | x_{k-1} = s_i) \quad (3)$$

The transitions between states for more than one time step is more easily calculated by a transition matrix. Let $P = (p_k)_{i,j}$ denotes the matrix where each element at index (i,j) represents the probability defined in Equation (1). Suppose that the probabilities are fixed when they are not influenced by any other factors in the current time step, the transition matrix is defined as:

$$P = (p_{ij})_{n \times n} = \begin{bmatrix} p_{00} & p_{01} & \cdots & p_{0n} \\ p_{10} & p_{11} & \cdots & p_{1n} \\ \vdots & \vdots & & \vdots \\ p_{n0} & p_{n1} & \cdots & p_{nn} \end{bmatrix} \quad (4)$$

Where $\sum_{j=1}^{n} p_{ij} = 1$ for any $0 \leq i \leq n$.
The transition matrix trained for the Markov chain in this study uses the Maximum Likelihood Estimation (MLE). For homogeneous Markov model, given prediction time step k+1 at day n, the transitional probabilities are estimated through all possible changes for each pair of occupancy states at step k in a daily file. The historical training set begins from the first day to day n-1, given $S = \{(x_{t-T}, x_{t-T+1}), (x_{t-2T}, x_{t-2T+1}), (x_{t-3T}, x_{t-3T+1}\}$, …, $(x_{t-n\times T}, x_{t-n\times T+1}\}$ where T is the time scale of the daily file, namely T=24 for 24 hour resolution. The transition probability is estimated by the MLE where n_{ij} pairs of the states' sequence $\{s_i, s_j\}$ exist in all pairs of the sequences of set S is:

$$\hat{p}_{ij} = \frac{n_{ij} + \alpha}{\sum_{l=1}^{k} n_{il} + \alpha} \quad (5)$$

Where α is a smooth factor (0<α<0.1) to avoid an extremely small probability of transition.

All the theoretical developments discussed above have no difference than what is commonly used by other researchers. In this study, the authors propose to integrate a moving window strategy to estimate the transitional probability and utilize the model for prediction with a change-point analysis. In essence, the moving window is a trimmed window covering a training sequence $W = \{x_w, x_{w+1}, \ldots, x_v\}$ before prediction time step t+1 where $t - T \leq w \leq v \leq t$. The window does not fully utilize all the historical data compared to the homogeneous training set S. The approach proposes two new procedures to estimate the transitional probabilities: 1) at which time step should the horizon of the moving window be trimmed according to the pattern change; and 2) how long should the training data be chosen to perform MLE in one horizon of the window. Let $D = \{d_1, \ldots, d_{T\times z}\}$represents the all selectable historical data before the state that needs to be predicted. Here, if the occupancy presence state that is to be predicted is in a working day, the selection of D only contains the available profiles of z working days. Regardless of the occupancy level, D is processed into a data set containing only the presence and absence as 1 and 0. A discrete profile of the presence probability in daily scale is generated by:

$$P_m = \frac{\sum_{j=1}^{z} (\lambda^{z-j} \cdot d_{(j-1)\times T+m})}{z} \quad (6)$$

Where $1 \leq m \leq T$ and λ is an exponential forgetting factor, which is below 1. Without forgetting effect, the data of the all tested period exerting equal but ever-decreasing influence on the distribution of the presence. An exponential forgetting could maximize penalty on the older information from the data and allow the presence probability to retain the most recent information only.

Change-point detection is implemented to check the change point of the presence profile calculated from Equation (6). Four main change-point of the presence pattern can be classified: shifting to long absence, shifting to low presence rate, shifting to high presence rate and shifting to long presence. The assumption is that a change of the moving-window should be trimmed based on the changes of presence pattern in a daily scale. The detection algorithm in this study used relative density-ration estimation with the Pearson divergence as a divergence measure to score the possible change points. For a subsample m selected from the distribution n, the symmetric divergence score is defined as follows (Liu, 2013):

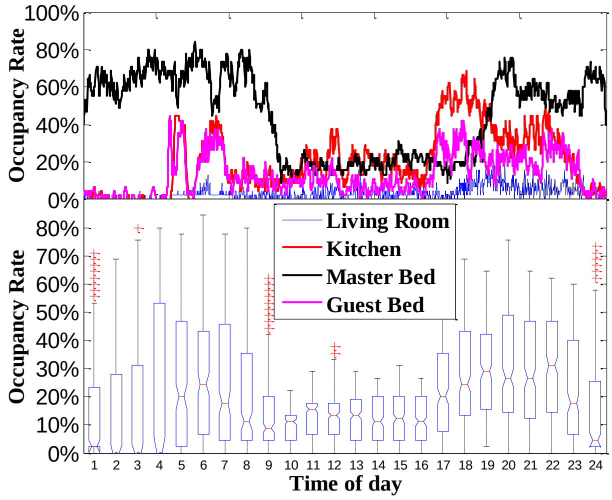
$$\int p_\alpha(m)[\frac{p(n)}{p_\alpha(m)} - 1]^2 d(m) + \int p(n)[\frac{p_\alpha(m)}{p(n)} - 1]^2 d(n) \quad (7)$$

Where $p_{\alpha}(m) = \alpha p(n) + (1-\alpha)p(m)$, p is the density function of the corresponding variables, and the factor $\boldsymbol{\alpha}$ is a smooth factor to the plain density ration.

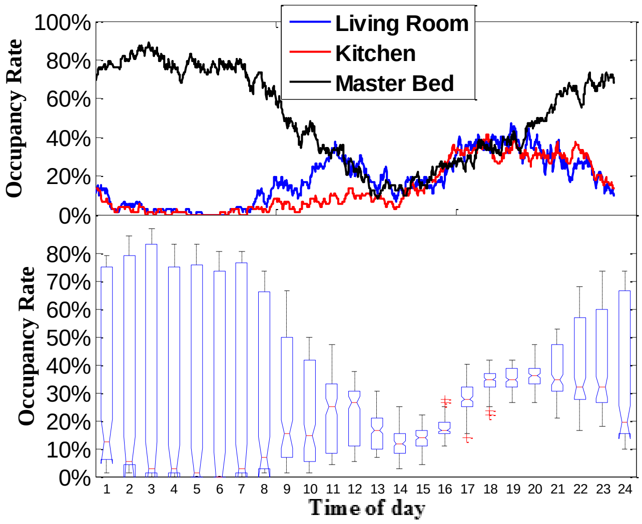
The tuning of the training data in a trimmed window, as an answer to the second question, depends on the prediction horizon. In an extremely short-term forecast case, models are commonly trained within one moving window (e.g. 15-min, 30-min, and 1-hour ahead cases). The size of the window is determined by a contingency table with a leave-one-out validation. The validation is performed on two subsets of data in five most recent working days. Commonly used 10-folder validation is not suitable in this case due to the limited length of window size. Each tuning length within the horizon is assigned a score that added the true positive value and true negative value from the contingency table. The highest score represents a suitable candidate of the tuning length of one moving horizon. In day ahead prediction, there is no possibility to access the intraday information to calibrate the inhomogeneity. Hence, the assumption is that two consecutive days have similar patterns. The predictions thus are simulated in a daily scale where the full horizon of each moving window classified by Eq. (7) is used as the tuning length in that moving window.

Test Beds

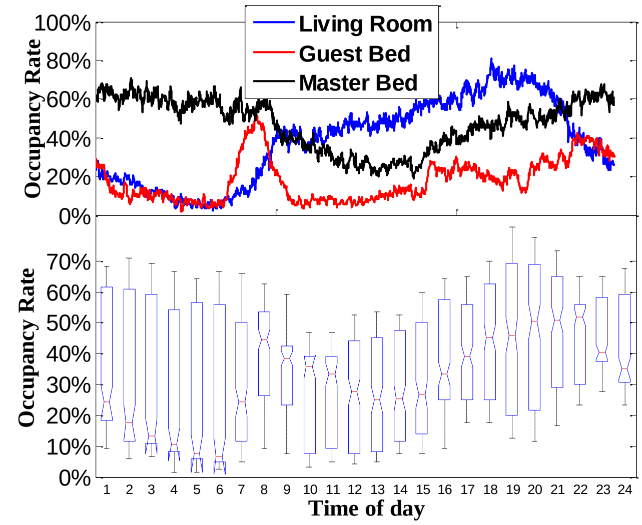
The occupancy data in this research are collected from four houses in downtown San Antonio. The detailed test-bed is described in Dong et al. (Dong, 2015). Houses are named according to construction materials: SIP (Structure Insulated Panels), ACC (Autoclaved Aerated Concrete), Container (Ship Container), and Stick (Wood). They are leased and operated mostly by part-time workers and low-income residents. The presences of occupancy are monitored at 5 min intervals from over 30 PIR sensors for all the rooms including kitchen, bathroom, living and bedroom areas during the year of 2014. The following testing period is selected: ACC's occupancy was modeled from Sep 17th to Oct 31st. Container was modeled from May 21st to July 31st. SIP was modeled from Jan.1st to Apr.30th. Stick was modeled from Jan.1st to March.31st.



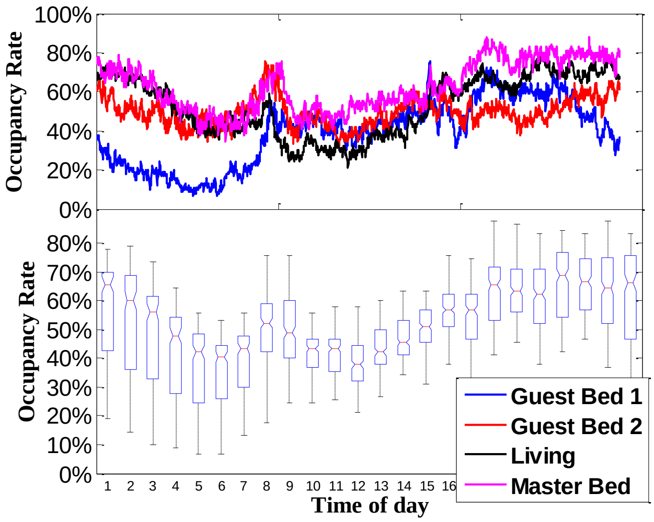
a) ACC House



b) Container House



c) SIP House



d) Stick

Figure 2: The presence rate of the rooms in the four houses.

The average rates of the presence are plotted for the monitored rooms of the four samples, shown in the upper part of each figure in Figure 2. All four houses demonstrate striking differences of presence at the room level. However, the common pattern of presence is revealed for individual rooms in a cross-sectional comparison. Master bedrooms are mostly occupied during nights. Living rooms or kitchens are occupied mostly around afternoons and evenings. The variances from all rooms' presence rate in the individual house are presented in the lower part of each figure in Figure 2 using 1.5 interquartile range (99% confidence interval) in box-and-whisker plots. Large variances are observed in first three houses while the fourth one has a very similar presence patterns. More details can be referred to the study of those samples (Dong, 2015).

Results

The key success of the proposed model depends on the optimization of the moving window as mentioned in the previous section. One example of the detected changing points between the windows is shown in Figure 3 for the prediction of ACC's master room on Oct 15th . The normalized score is calculated based on 0.8 forgetting factor with a span of all the historical records before the date. Based on the analysis of historical data, there should be five windows for prediction of that specific day as one starting from 12 a.m. to 7 a.m., the next one ends around 10 a.m., the third one ends around 6 p.m., the fourth one ends around 8 p.m., and the last one ends until the end of the day. All the predictions using the proposed inhomogeneous Markov chain follow the change rule of the window mentioned above.

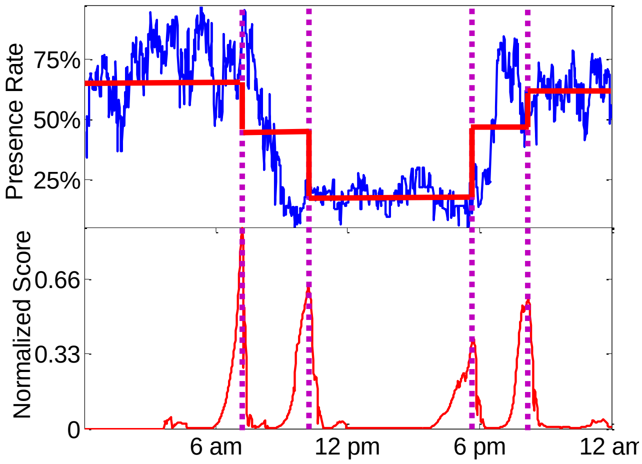


Figure 3: Demonstration of the change-point detections of the presence pattern.

The predictive performance of the models is evaluated based on the correctness of the occupancy predictions in terms of the occupied and unoccupied states. There are only two predicted classes, presence and absence. Of the total *l* predictions, if there are *m* predicted presence while the observations of the rooms are occupied and *n* predicted absence while the observations of the rooms are not occupied, the overall accuracy is thus calculated as a percentage as $(m+n)/l$. All results of the stochastic models' predictions for the individual rooms are presented in Table 1,2,3,4 and 5 for 15-mins ahead, 30-mins ahead, 1-hour ahead and 24-hour ahead respectively.

Table 1: Presence predictions for the residential houses (15-min ahead in 15-min resolution).

Methods		Inhomo Markov	Homo Markov
ACC	Master Room	76.49%	61.38%
	Living Room	96.48%	96.18%
	Kitchen	71.92%	56.02%
	Guest Bed	71.18%	53.68%
Container	Master Room	80.01%	56.01%
	Living Room	80.41%	56.40%
	Kitchen	90.08%	64.39%
SIP	Living Room	72.62%	63.61%
	Master Bed	80.90%	60.53%
	Guest Bed	70.88%	61.18%
Stick	Living Room	68.06%	59.89%
	Guest Bed 1	67.16%	52.33%
	Guest Bed 2	81.85%	54.34%
	Master Bed	79.46%	64.81%

Table 2: Presence predictions for the residential houses (30-min ahead in 30-min resolution).

Methods		Inhomo Markov	Homo Markov
ACC	Master Room	73.91%	57.63%
	Living Room	98.81%	96.86%
	Kitchen	63.99%	53.81%
	Guest Bed	70.63%	52.04%
Container	Master Room	73.21%	53.07%
	Living Room	69.64%	56.71%
	Kitchen	84.42%	64.70%
SIP	Living Room	71.73%	63.17%
	Master Bed	75.50%	60.34%
	Guest Bed	72.02%	61.23%
Stick	Living Room	63.19%	57.27%
	Guest Bed 1	65.87%	54.69%
	Guest Bed 2	78.97%	55.10%
	Master Bed	73.81%	62.40%

Table 3: Presence predictions for the residential houses (1-hour ahead in 1-hour resolution).

Methods		Inhomo Markov	Homo Markov
ACC	Master Room	75.99%	54.71%
	Living Room	99.60%	96.62%
	Kitchen	61.11%	52.94%
	Guest Bed	73.02%	55.81%
Container	Master Room	70.83%	53.35%
	Living Room	67.26%	51.98%
	Kitchen	84.52%	67.88%
SIP	Living Room	67.46%	64.38%
	Master Bed	74.21%	59.02%
	Guest Bed	76.35%	61.31%
Stick	Living Room	61.11%	58.20%
	Guest Bed 1	63.29%	58.32%
	Guest Bed 2	74.40%	56.38%
	Master Bed	67.66%	63.89%

Table 4: Presence predictions for the residential houses (24-hour ahead in 15-min resolution).

Methods		Inhomo Markov	Homo Markov
	Master Room	60.57%	59.25%
	Living Room	96.58%	95.60%

ACC	Kitchen	68.35%	56.45%
	Guest Bed	71.68%	54.09%
Container	Master Room	55.90%	56.11%
	Living Room	73.02%	55.12%
	Kitchen	83.33%	64.23%
SIP	Living Room	63.14%	63.56%
	Master Bed	70.49%	60.24%
	Guest Bed	65.72%	61.15%
Stick	Living Room	59.08%	60.28%
	Guest Bed 1	54.91%	51.46%
	Guest Bed 2	56.80%	55.01%
	Master Bed	69.54%	64.11%

Table 5: Presence predictions for the residential houses (24-hour ahead in 1-hour resolution).

Methods		Inhomo Markov	Homo Markov
ACC	Master Room	58.73%	51.64%
	Living Room	99.54%	96.41%
	Kitchen	62.10%	52.83%
	Guest Bed	77.57%	54.32%
Container	Master Room	52.78%	53.02%
	Living Room	68.25%	50.68%
	Kitchen	78.97%	69.49%
SIP	Living Room	62.50%	64.08%
	Master Bed	77.89%	60.01%
	Guest Bed	74.40%	62.09%
Stick	Living Room	55.16%	59.06%
	Guest Bed 1	57.94%	57.74%
	Guest Bed 2	57.34%	56.20%
	Master Bed	65.48%	63.72%

lower part of each figure in Figure 2). The examples are the living room of ACC (blue line) in Figure 2 a) and the guest bedroom 2 of Stick (red line) in Figure 2 d). However, the living room may be a special case owing to the extremely low presence rate (<20%), representing an absence dominated pattern. In contrast, the guest bedroom 2, with a persistent presence rate (between 40% to 60%), can be interpreted as a stable pattern of the occupancy, in which the resident leaves or enters their room on a regular schedule. The third case is an occupancy pattern balanced with smoothness and variance, as seen with guest room of SIP (red line) in Figure 2 c). The prediction from the proposed Markov model can reach 80% accuracy, where a big variance is observed between 6 am to 9 am which is a summit up to 50%. Although the similar findings can be claimed for the homogeneous Markov model, the average accuracy for each prediction of the room is much lower than the proposed inhomogeneous Markov model. For 24-hour ahead predictions, there is no significant difference in terms of accuracy among the two models. It is mainly because the methods are predicting based on the assumption that each day's presence pattern should be similar. This kind of assumption actually could be a drawback for the more stochastic sample among the rooms of the four residential houses. Only a few exceptions existed in Table 4 and Table 5 where they have more than 75% correct prediction.

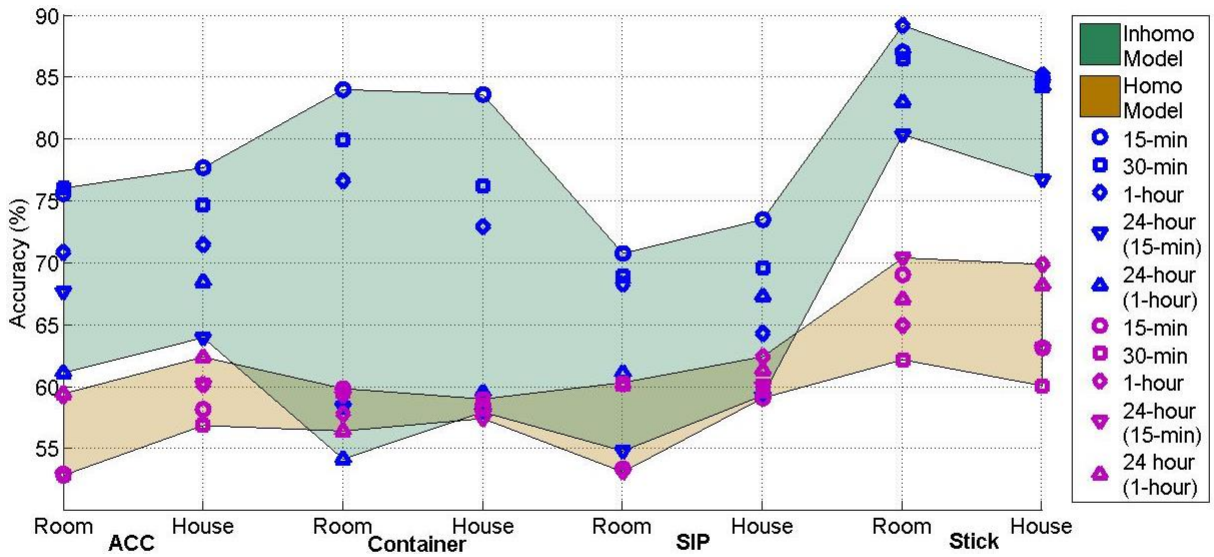


Figure 4: Comparison between models based on the modelling level (House: house-level, Room: room-level).

Comparing the results from Table 1 to Table 5, the presence can be more accurately predicted up to 70% (red in the tables) in the extremely short-term forecast (e.g. 15-min till 1-hour) for the proposed Markov model, if the presence rate is smooth enough. For example, the presence rate of Container compared to other samples does not have the small spikes observed consistently on the curve (comparing all the blue, black, red, and maroon lines). Another observation is that the predictive performance of the model is highly correlated with the variances (the box plots of the variances shown at the

The performances of the models at house level is more important for applications such as smart control of thermostat. The thermostat of the residential houses is usually located at one place (such as the living room) and the cooling and heating usually operated based on house level. For this study, occupancy presence can be predicted and simulated in two ways: 1) aggregates the room-level forecasts to generate the prediction for the house-level, and 2) directly predict the occupancy status by processing the data to a house-level input first. The results of the two ways are presented in Figure 4 for all the samples of all two models. The performance of the homogeneous model is still inferior compared to the proposed inhomogeneous Markov model. Meanwhile, the Markov model is still expected to have a promising performance from 15-min to 1-hour forecast in Figure 4. As noticed in Figure 4, the day ahead forecasts are analyzed based on two different time scales (15-min and 1-hour). The reason to adopt two forecasting intervals is because of the different forecast window of the weather information, electricity prices, and load schedules in predictive control design. In conclusion, for extremely short term forecast from 15-min to 1-hour ahead, the proposed Markov model is highly recommended owing to the high accuracy while the approach needs further improvement for 24-hour ahead forecasts.

Discussions

Currently, only a few of researchers focus on occupancy modelling of residential buildings (Torriti, 2012; Widen, 2012; Lopez-Rodriguez, 2013; Wilke, 2013). Those studies only provide estimates of the domestic occupancy profile in a national scale using the Time Use Survey (TUS) data. Individual occupancy profile at building level can be derived from the national survey and used for single houses (Torriti, 2012). However, studies based on such data represent an averaged occupancy pattern because TUS data are usually reported in terms of the average time used for various activities by occupants in a specific social-economic group of the population (Widen, 2012). In addition, most models used in those studies are usually developed as a homogeneous Markov chain integrating with Monte Carlo technique or Cross Validation to enhance the performances (Lopez-Rodriguez, 2013; Wilke, 2013). Therefore, the authors in this study propose a new method to predict occupancy presence of residential buildings using real-time measured data for the purpose of model predictive controls.

One thing noticed is the temporal influence of different forecast window on the prediction performance of the models. For different temporal resolutions and prediction windows, the stochastic pattern changes dramatically. If a specific temporal resolution or a prediction window has a more regular presence rate, models should be easy to produce accurate predictions, e.g. the more accurate predictions of the 15-minute resolution occupancy data in Table 1. In other research domains, the accuracy of the models' predictions could be improved by changing the window of the forecast (Li, 2016). A more recent study to predict the occupancy level of the office workers has a similar conclusion by increasing the forecast resolution (Chen, 2016). However, in this study, no significant improvements of prediction accuracy are observed for most samples when the prediction horizon and resolution increase to 24-hour ahead and 1-hour interval respectively. The situation could be contributed mostly by the uniqueness of the tested residential data set. There are apparently more varied periods (various months) and samples (4 houses) observed in Figure 2 compared to the previous studies for both residential and commercial samples. A longer period could be collected and tested to further investigate the possibility to improve the forecast ability of the proposed model by changing the forecasting resolution and window.

The last but not the least is that occupancy modelling should be fit-to-purpose, for the smart controls of the buildings. For example, the temporal changes actually have few impacts on the smart controller (e.g. Nest (Peffer, 2011)). Those advance interfaces not only record occupancy presence and the human building interactions from sensors, but also analyze the preference of occupants. The control strategy tries to diminish the stochastic overrides of users and increase the predictive power of the occupancy models. Although an even higher resolution of the occupancy monitoring, such as one minute interval, could be used to try to improve model performance in the smart control environment, the control algorithms will instead have a more frequent track to the occupancy model. Such frequent responses from occupancy-based controller can highly violate the operations of the systems. Unless the occupants are extremely sensitive to the comfort changes, the predictive performances and control difficulties should be equally addressed in a relaxed forecast window, namely 15-min, or even hourly scale.

Conclusion

This paper aims to develop and demonstrate an innovative approach for short-term residential occupancy presence forecasting. By predicting future occupancy presence of different time scales (15-min to 24-hour ahead), the proposed inhomogeneous Markov model demonstrates its predictive power specifically for the purpose of control application. The results are validated through long term measured data from the field tests of the residential houses and compared to a traditional homogeneous modeling approach. The final results show that the proposed Markov model outperforms the traditional Markov in terms of an average 17% correctness with 26% maximum difference in one time step ahead forecast. In day ahead prediction, a higher performance of occupancy forecasting is still observed in the proposed Markov model, but not as competitive as the results from the same model for the shorter prediction windows. Implementing such kind of occupancy model will be a solution for characterizing the large dynamics existing in the real-time energy consumption and help the building to optimally integrate to the electricity grid operations. It will be even more beneficial to advanced building control if more accurate forecasts can be made for longer windows of the occupancy presence.

This study observes a significant lower performance in 24-hour ahead prediction scenario compared to the other prediction window (e.g. 15-min to 1-hour ahead). It is challenging to improve the forecast accuracy in this case even with the changes of temporal resolution (sampling rate) between 15-min and 1-hour resolution. But the results show competitive performances compared to recent studies (Mahdavi, 2015). Further investigation on improvements of the day ahead predictions could be conducted by integrating more advanced time series analyses to accurately detect and predict the dynamics of the occupancy.

Acknowledgement

This material is based upon work supported by National Science Foundation (NSF), under Grant No. CBET-1637249.

References

Lazaroiu, G.C., and M. Roscia. (2012). Definition methodoogy for the smart cities model. *Energy 47*, 326-332.

Gunay, H.B., O'Brien, W., and I. Beausoleil-Morrison (2013). A critical review of observation studies, modeling, and simulation of adaptive occupant behaviors in office. *Building and Environment 70*, 31-47.

Dong, B., Li, Z.X. and G. Macfadden (2015). An investigation on energy-related occupancy behavior for low-income residential buildings. *Science and Technology for the Built Environment 21*.

Haq, A.M., Hassan, M.Y., Abdullah, H., Rahman, H.A., Abdullah, M.P., Hussin, F., and D.M. Said (2014). A review on lighting control technologies in commericial buildings, their performance and affecting factors. *Renewable and Sustainable Energy Reviews 33*, 268-279.

Halvgaard, R. (2014). Model Predictive Control for Smart Energy Systems. *Dissertation*. Department of Applied Mathematics and Computer Science, Technical University of Demark.

Mahdavi, A., and F. Tahmasebi. (2015). Predicting people's presence in buildings: An empirically based model performance analysis. *Energy and Buildings 86*:349-355.

Parys, W., Saelens, D., and H. Hens. (2011). Coupling of dynamic building simulation with stochastic modelling of occupant behavior in offices – a review-based integrated methodology. *Journal of Building Performance Simulation 4*, 339-358.

Richardson, I., Thomson, M., and D. Infield. (2008). A high-resolution domestic building occupancy model for energy demand simulation. *Energy and Buildings 40(8)*, 1560-1566.

Page, J., Robinson, D., Morel, N., and J.L. Scartezzini. (2008). A generalized stochastic model for the simulation of occupant presence. *Energy and Buildings, 40(2)*, 83-98.

Wang, C., Yan, D., and Y. Jiang. (2011). A novel approach for building occupancy simulation. *Building simulation 4(2)*, 149-167.

Wilke, U. (2013). Probabilistic bottom-up modelling of occupancy and activities to predict electricity demand in residential buildings. *Dissertation*. Programme Doctoral En Environnement. Ecole Polytechnique Federale De Lausanne, Suisse.

Ai, B., Fan, Z.Y., and R.X. Gao. (2014). Occupancy estimation for smart buildings by an auto-regressive hidden Markov model. In *American Control Conference*, Portland, Oregon, U.S.A.

Dong, B., and Lam, K.P. (2013). A real-time model predictive control for building heating and cooling systems based on the occupancy behaviour pattern detection and local weather forecasting. *Building Simulation 7(1)*, 89-106.

Liu, S., Yamada, M., Collier, N., and M. Sugiyama. (2013). Change-point detection in time-series data by relative density-ration estimation. *Neural Networks 43*, 72-83.

Torriti, J. (2012). Demand side management for the European supergrid: occupancy variances of European single-person households. *Energy Policy 44*, 199-206.

Widen, J., Molin, A., and K. Ellegard. (2012). Models of domestic occupancy, activities and energy use based on time-use data: deterministic and stochastic approaches with application to various building-related simulations. *Journal of Building Performance Simulation 5(1)*, 27-44.

Lopez-Rodriguez, M.A., Santiago, I., Trillo-Montero, D., Torriti, J., and A. Moreno-Munoz. *Energy Policy 62*: 742-751.

Wilke, U., Haldi, F., Scartezzini, and D. Robinson. (2013). A bottom-up stochastic model to predict building occupants' time-dependent activities. *Building and Environment 60*: 254-264.

Li, Z.X., Mahbobur-Rohman, S.M., Vega, R., and B. Dong. (2016). A hierarchical approach using machine learning methods in solar photovoltaic energy production forecasting. *Energies 9(1)*, 55.

Chen, Z.H., and Soh, Y.C. (2016). Comparing occupancy models and data mining approaches for regular occupancy prediction in commercial buildings. *Journal of Building Performance Simulation*, DOI: 10.1080/19401493.2016.1199735.

Peffer, T., Pritoni, M., Meier, A., Aragon, C., and D. Perry. (2011). How people actually use thermostats: A review. *Building and Environment 46*, 2529-2541.

Mahdavi, A., and F. Tahmasebi. (2015). Predicting people's presence in buildings; An empirically based model performance analysis. *Energy and Buildings 86*, 349-355.