# Building Energy Prediction with Occupancy

Kenan Zhang
School of Architecture
MS AECM
kenanz@andrew.cmu.edu

Xiaowen Gu
School of Architecture
MS SD
xiaoweng@andrew.cmu.edu

## ABSTRACT
In this paper, we investigate the impact of occupancy pattern on building electricity consumption with the use of multivariable linear regression.

## Keywords
Building Electricity Consumption, Occupancy Pattern, Linear Regression, Markov Chain Monte Carlo.

## 1. INTRODUCTION
With the rapid increase of energy consumption in recent decades, concerns have been raised over building energy use as it accounts for up to 40% of primary energy consumption in US. The majority of building energy consumption is attributed to electricity use [1]. Occupancy schedule has an impact on building energy consumption as it is related to the time period when equipment is on to maintain a healthy and comfortable indoor working environment. Therefore, it is interesting to study the relationship between the electricity consumption and occupancy patterns with the consideration of building energy management.

We had our study focused on Gates-Hillman Center (GHC) at Carnegie Mellon University (CMU). The occupancy sensors in GHC provided us with reliable occupancy data to study occupancy patterns on the campus during semesters. In Assignment 2, we built linear regression model to predict the electricity consumption of campus buildings with time-interval indicators and outdoor temperatures. In this study, we firstly applied this model to GHC and then added real occupancy to analyze whether the new variable improve the model performance.

To explore the feasibility of occupancy patterns in campus building energy prediction, we continued to build Markov model based on GHC occupancy and simulated an occupancy pattern using Monte Carlo experiments. After comparing the simulation result with the ground truth, we replaced real occupancy in the energy prediction model with simulated occupancy.

## 2. METHODS
### 2.1 Linear Regression
Following the method used in [2], we built linear regression model of electricity consumption as follows:

$$L(t,T(t)) = \alpha_t + \sum \beta_j T_j^c(t) + \lambda Occ_t$$

where $\alpha_t$ is the time-interval indicator at time $t$; $\beta_j$ is the coefficient of piecewise-temperature model variable $T_j^c(t)$; $\lambda$ is the coefficient of occupancy level $Occ_t$ at $t$

### 2.2 Markov Chain Monte Carlo
#### *2.2.1 Markov Chain*
Markov model is often used in occupancy prediction and simulation[3][4]. One of its important assumption is that the present state is depended and only dependent on the previous state, which is represented as follows:

$$P(X_t \mid X_{t-1}, X_{t-2}, ..., X_1) = P(X_t \mid X_{t-1})$$

Therefore, the transition probability from state $i$ to state $j$ could be estimated by from the normalized counts as follows:

$$p_{ij} = n_{ij} / \sum_{k=1}^{N} n_{ik}$$

where $N$ is the size sample set; $k$ represents each possible state.

In this study, the occupancy level of each timestamp was firstly calculated as the percentage of occupied offices. The discrete states of occupancy were defined based on daily pattern and assigned to each timestamp.

#### *2.2.2 Blended Transition Matrix*
However, one single transition matrix is not appropriate for the building occupancy. For instance, it is more likely to have the occupancy of an office transfer from 0 to 1 than 1 to 0. Therefore, following the method used in [4], we define multiple transition matrices based on real occupancy data and use blended transition matrix for each time slot. This method also help solve the issue of sink state.

Specifically, the blended transition matrix for time interval t is calculated as:

$$\overline{T_t} = \sum_{s=1}^{M} \theta_{ts} T_s$$

where $s$ represents each time slot; $M$ is the total number of time slots; $\theta_{ts}$ is the weight of each transition matrix $T_s$, which is defined as follows:

$$\theta_{ts} = \frac{\varphi(c_t - c_s)}{\sum_{s'=1}^{M} \varphi(c_t - c_{s'})}$$

where $c_t$ and $c_s$ are indicators of time slots; $\varphi$ is defined as slot function

$$\varphi(x) = \sigma(\frac{2a}{d}(x + \frac{d}{2})) - \sigma(\frac{2a}{d}(x - \frac{d}{2}))$$

where $\sigma(x) = 1/(1 + e^{-x})$ is the sigmoid function.

Therefore, the blended transition matrix $\overline{T_t}$ at time slot $t$ integrate all transition matrix $T_s$ while heavily weighting $T_t$. We also used default values in [3] ( $a = 10, d = 3$ ).

### 2.2.3 Monte Carlo Simulation

After building the Markov model, we used Monte Carlo simulation to explore the occupancy pattern. In brief, at each timestamp a float was randomly generated to determine the occupancy state. The probability distribution at present stage is dependent on the prior stage. After a large number of simulations, a stable occupancy pattern could be generalized.

## 3. Data Processing

### 3.1 Dataset

Four datasets were used in this study, including the outdoor temperatures collected on CMU campus[1], the occupancy data of GHC, the electricity consumption of GHC and other campus building. The general information of these datasets are summarized in Table 1.

**Table 1. Dataset Description**

| Dataset | Duration | Window |
|---|---|---|
| Outdoor Temperature (temp) | 2014.11.01 - 2015.12.01 | 5 min |
| Electricity Consumption of GHC (electGates) | 2014.12.10 - 2015.12.10 | 15 min |
| Occupancy of GHC (occGates) | 2014.09 -2015.12 | ~20 min |

### 3.2 Data Filtering

As is shown in Table 1, durations and time windows are different among datasets. Additionally, when preparing the data, we noticed that some data was missing, duplicated or incorrect. For instance, we got the occupancy data of 308 offices in GHC, however only 206 were valid. Therefore, we firstly cleaned each dataset and interpolate them with same time interval. The time window used in our study is 15min. And the first interval starts from 12:00am of each day. Specifically, the data filtering of occupancy data was done in Matlab due to the data source.

### 3.3 Data Preparation

The occupancy data of GHC is collected by the occupancy sensor in each office. Therefore, the raw data is binary. To obtain overall occupancy level of GHC, we harmonized the occupancy data with same time series and calculated the percentage of occupied offices.

Since we were interested in the impact of occupancy patterns on building energy consumption, we chose to study a time period with regular occupancy. Figure 1 shows the occupancy level of GHC from Nov.3 2014 to Dec.4 2015.

[1] Collected by Center for Building Performance and Diagnostics (CBPD)

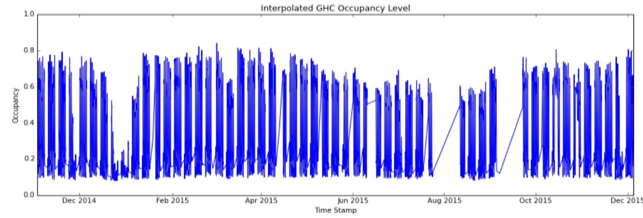


**Figure 1. GHC Occupancy Level.**

As is shown in Figure 1, GHC had regular occupancy patterns during the majority of fall and spring semesters. Therefore, we chose the months with full data in both semesters and split the data into training set and test set equally week by week. Table 2 summarizes the final dataset used in this study. Figure 2 shows the electricity consumption of GHC.

**Table 2. Final Dataset**

| Dataset | Duration | Window |
|---|---|---|
| Outdoor Temperature (temp) | 2015.01.12-2015.04.03<br>2015.04.20-2015.05.15<br>2015.06.22-2015.07.10<br>2015.09.28-2015.10.23 | 15 min |
| Electricity Consumption of GHC (electGates) | | |
| Occupancy of GHC (occGates) | | |

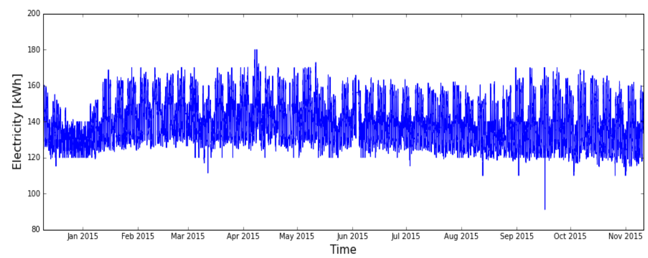


**Figure 2. GHC Electricity Consumption.**

## 4. RESULTS

### 4.1 Energy Prediction of GHC without Occupancy

As is shown in Figure 3, the predicted electricity consumption according to the linear regression model in Assignment 2 [2] fits well with the actual values. The corresponding coefficient of determination ($R^2$) is about 0.72, which indicates a linear correlation. After a t test for a null hypothesis for each coefficient, the null hypothesis is rejected for all of the coefficients at a significance level of 0.025. This demonstrates the impact of and outdoor temperature on GHC electricity consumption. To further improve the model, another variable occupancy is added to the model.

(Dec 10 2014 – Oct 28 2015)

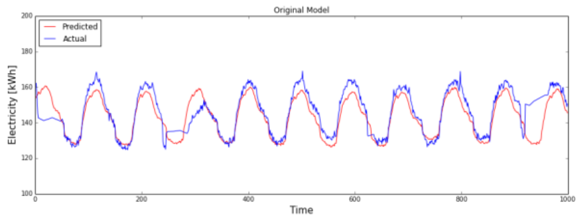


**Figure 3. GHC Energy Prediction without Occupancy.**

## 4.2 Occupancy Patterns and Simulation

### *4.2.1 Occupancy States*

We chose sample week and sample day to explore the occupancy pattern of GHC. As is shown in Figure 4, the occupancy has a regular daily pattern. Base on this patter, we assumed 4 occupancy states ([0.1,0.3,0.6,0.8]). Figure 4 also shows assumed occupancy states with respect to the same time period.

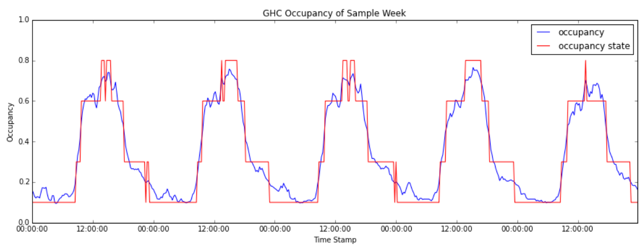


**Figure 4. Comparison of Real Occupancy and Assumed Occupancy State.**

### *4.2.2 Blended Transition Matrix*

36-week occupancy data was used to estimate the 96 basic transition matrices. These matrices were then combined with the method introduce in the section 2.2. Figure 5 shows the transition probabilities between occupancy states. Since the occupancy is more likely to change continuously, here we only plotted the probabilities between adjacent states.

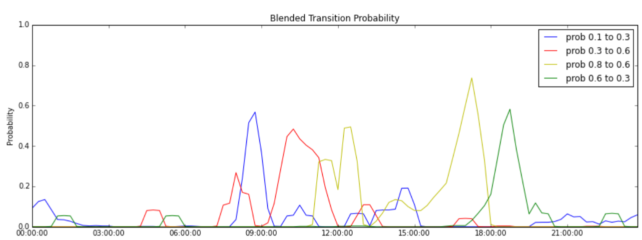


**Figure 5. Blended Transition Probabilities.**

As is shown in Figure 5, the transition probabilities of increasing occupancy are high in the morning with a smaller peak early in the afternoon. Correspondingly, those of decreasing occupancy are high in the evening and the probability moving from 0.6 to 0.3 follows that from 0.8 to 0.6. These patterns well interpret the daily schedule of campus building. People come to campus around 9am to 10am, leave for lunch at noon and come back soon, and leave the building in the evening. The small peak of increasing occupancy represents people who work at night, which is also common at CMU.

### *4.2.3 Markov Chain Monte Carlo*

We applied 100 Monte Carlo simulations and calculated the average occupancy to be the simulated occupancy. Figure 6 shows the simulated occupancy and the real occupancy of a sample week.

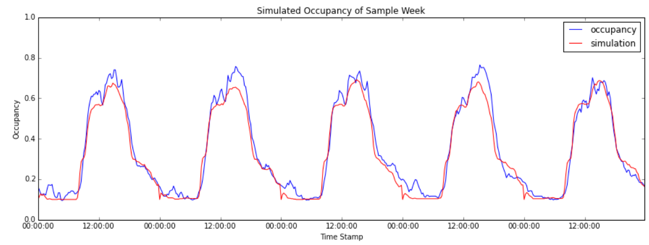


**Figure 6. Comparison of Real Occupancy and Simulated Occupancy.**

As is shown in Figure 6, the simulated occupancy well interpreted the occupancy pattern of GHC. This simulated occupancy could be used as the occupancy schedule for building energy management. Compared to manually predefined occupancy schedules, this schedule may perform better since it is generated from real data.

To explore the feasibility of using simulated occupancy in building energy prediction, we replaced the real occupancy in the regression model and analyzed the performance once again of the new model.

## 4.3 Energy Prediction with Simulated Occupancy

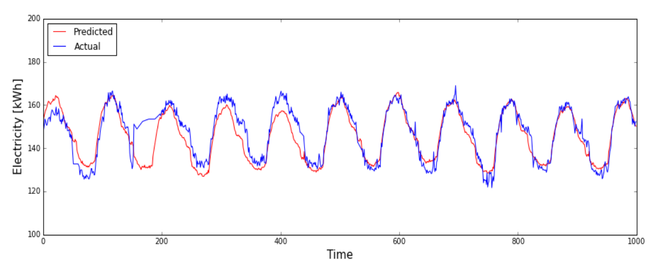


**Figure 7 GHC Energy Prediction with Simulated Occupancy**

As is shown in Figure 7, the predicted electricity consumption including simulated occupancy in addition to time and outdoor temperature fits well with the actual values. The corresponding coefficient of determination ($R^2$) is about 0.79, which is higher than that without the occupancy variable. After a t test for a null hypothesis for each coefficient, the null hypothesis is rejected for all of the coefficients except the last coefficient at a significance level of 0.025. This indicates the effect of occupancy on GHC electricity consumption prediction is insignificant.

## 5. SUMMARY & FUTURE WORK

In this study, we followed the method of Assignment 2 and used data from GHC to predict the building energy consumption. Different from Assignment 2, we added occupancy into the linear regression model. Although there are missing data and other problems with raw data sources, we were able to interpolate full dataset of 23 weeks.

In sum, adding occupancy improved the model performance of building energy prediction. However, the occupancy variable did not t-test. Nevertheless, as is shown in Figure 8, there is strong correlation between energy use and occupancy. Therefore, future work will focus on exploring the reason.

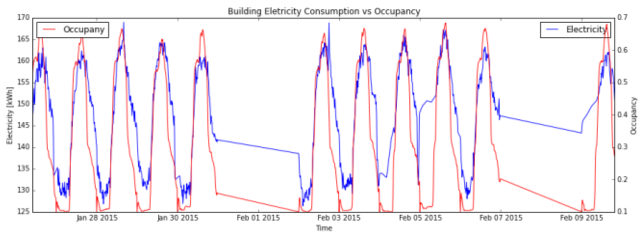


**Figure 8. Occupancy vs Building Energy Consumption.**

## 6. ACKNOWLEDGMENT


We give our sincere thanks to Professor Mario Berges, who offered us suggestions and instructions on this study. Besides, we want to thank Jingkun, Jerry and Joseph, who provided us with great help to get access to different data sources.


## 7. REFERENCES


[1] DOE, Building Energy Data Book, 2012. [Online]. Available: http://buildingsdatabook.eren.doe.gov/ChapterIntro1.aspx

[2] Mathieu J L, Price P N, Kiliccote S, et al. Quantifying changes in building electricity use, with application to demand response[J]. *Smart Grid, IEEE Transactions on*, 2011, 2(3): 507-518.

[3] Page J, Robinson D, Morel N, et al. A generalised stochastic model for the simulation of occupant presence[J]. *Energy and buildings*, 2008, 40(2): 83-98.

[4] Erickson V L, Carreira-Perpiñán M Á, Cerpa A E. OBSERVE: Occupancy-based system for efficient reduction of HVAC energy[C]//*Information Processing in Sensor Networks (IPSN)*, 2011 10th International Conference on. IEEE, 2011: 258-269.