Analysis of Energy Consumption of Academic & Administrative Buildings at Carnegie Mellon

Joseph Dryer – jdryer@andrew.cmu.edu

Masters Student, Civil and Environmental Engineering, Carnegie Mellon University

12-752 Data Driven Building Energy Management

Professor Mario Bergés – December 13th, 2015

ABSTRACT

This analysis studies the power consumption of Richards Engineering Hall, Scaife Hall, and Warner Hall at Carnegie Mellon University. The goal of the analysis is to compare energy consumption among these buildings as well as create predictive models. One full year of power consumption and outdoor temperature data were available for the analysis (see acknowledgements). The model used to predict power consumption was that of Mathieu et al [1] – a piecewise linear temperature dependent and time of week dependent multivariate regressive model. Through Exploratory Data Analysis (EDA) it is found that modifying this model for time of day as opposed to time of week dependence may be appropriate. Both the original and modified model are applied to the power and temperature data for business hours of weekdays. The results show that both the original and modified model have similar performance.

EDA and the predictive model both show that the power consumption for all three buildings is more dependent on time of day or time of week than outdoor temperature. The model predicts power usage well as it varies by time of day, but could not predict significant variation over the course of a year. In addition, the results show that Roberts had the highest power intensity (normalized power to size of building) followed by Warner and then Scaife. Before making any conclusions about energy or thermal efficiency of the buildings, the use and electric load differences of the buildings was considered. Roberts Engineering hall has a number of research laboratories, while Warner is mostly administrative offices and services, and Scaife is mostly offices and classrooms. Considering the low dependence on temperature for all three buildings and differences in use, no real conclusions could be drawn on energy performance. However, broadly speaking, zero temperature dependence of load may indicate HVAC systems that are not optimized.

Finally, considering the fact that energy use was not a strong function of outdoor temperature but rather time of day, the model was applied to all 24 hours of a day instead of only business hours. It is found that power consumption is predicted well in this different context with a higher percent of output variation explained by the model than when the model is applied to business hours only.

1. INTRODUCTION & METHODS

Building energy consumption can be predicted by regression analysis on temperature, time, and power consumption data. One such regression model is the 'piecewise linear and continuous temperature-dependent load model' described in Mathieu et al:

$$L_o(t_i, T(t_i)) = \alpha_i + \sum_{j=1}^{j} \beta_j T_{c,j}(t_i)$$

This model splits a characteristic time of week period (for example, weekdays during business hours) into *i* time indicators with regression coefficients α_i and into *j* temperature ranges with regression coefficients β_j. These temperature ranges are based on the relationship between power consumption and outdoor temperature. Such a relationship usually has several different linear regions separated by inflection points. These might be from, for example, transition from cooling to heating. In Mathieu et al, 10 F temperature ranges are assigned such that the number of temperature coefficients is equal to the total temperature range divided by 10.

The above model is applied and modified in this analysis.

In terms of implementation, data cleaning and manipulation, exploratory analysis and regression were carried out using the python programming language with the numpy, matplotlib, and scipy packages in an ipython notebook.

Data cleaning involved removing non numerical data points and interpolating the temperature data onto the ten minute resolution power data. Note that while Mathieu et al used a 15 minute time of week time indicator, this analysis uses a 10 minute time of day/week indicator.

Regression is carried out using matrix operations,

$$y = X\beta + \epsilon$$

where y is the load vector (power intensity in this case), X is the design matrix, and β is the vector of regression coefficients. ϵ is the error vector quantifying the difference between predicted and actual load, which we aim to minimize. In matrix form, the Mathieu et al model appears as:

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 & T_{c,j=1}(t_{i=0}) & \dots & T_{c,j=k}(t_{i=0}) \\ 0 & 1 & 0 & \dots & 0 & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & 0 & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & T_{c,j=1}(t_{i=i}) & \dots & T_{c,j=k}(t_{i=i}) \end{bmatrix} \begin{bmatrix} \alpha_{i=1} \\ \vdots \\ \alpha_{i=i} \\ \beta_{j=1} \\ \vdots \\ \beta_{j=k} \end{bmatrix} = \begin{bmatrix} y_{i=0} \\ y_{i+1} \\ \vdots \\ y_{i=i} \end{bmatrix}$$

The solution is then:

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

The load is then predicted as a product of these regression coefficients and the design matrix.

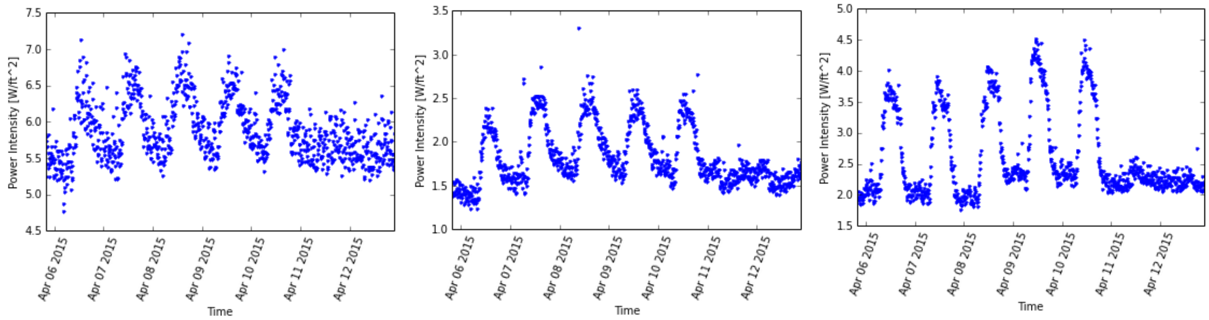

Figure 1: A week of power consumption showing 5 weekdays and then a weekend. From left to right – R, S, and W.

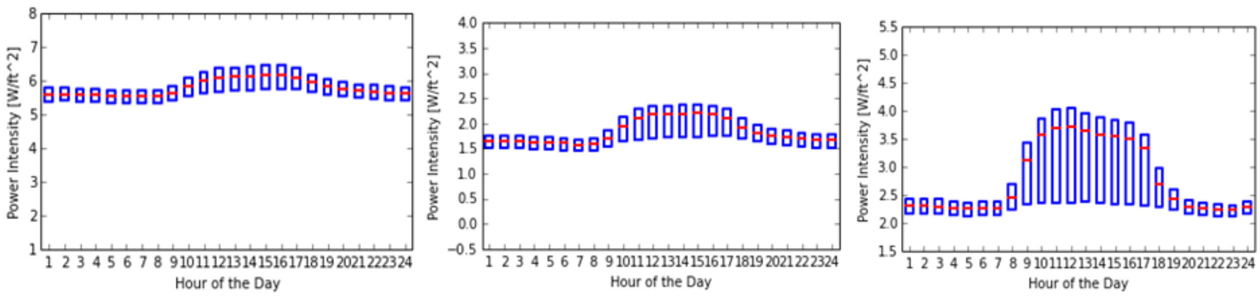

Figure 2: Box plots (median in red and Interquartile Range in blue) for diurnal power intensity. From left to right: R, S, and W.

The model is trained on one of two equally sized, random subsets of the data – the other subset is used for testing. Testing is achieved through graphical comparison of actual to predicted power intensity as well as quantitatively: the coefficient of determination, confidence intervals and tests of significance for regression coefficients are used to asses model performance.

2. RESULTS AND DISCUSSION

2.1 Exploratory Data Analysis

Three buildings were selected to compare: Roberts Engineering Hall ('R'), Scaife Hall ('S), and Warner Hall ('W'). The square footage of the buildings was found [2] so that power data could be normalized by building size – ie power intensity. Plots of power intensity for each building over the course of a week are shown in Figure 1. It is seen that power consumption varies on a daily basis, but that over the business week the pattern of variation is similar. Note that some variation does exist, and so an appropriate course of action with regard to modeling would be to testing both the time of week and time of day indicator based versions such that comparisons can be made.

Examining Figure 2, it can be seen that power intensity is at its peak from 9 AM to 5 PM for the three buildings suggesting these will be the appropriate characteristic business hours of the day. This will be the 'characteristic time of week' in this model. Note that the above figures are for all data, and so a lower bound (bottom of box) remains at the dead band zone corresponding to weekends, nights or other off peak times.

Now looking at the relation between temperature and power consumption for the business hours subset of the data (Figure 3), it is seen that there is no clear relationship between temperature and power intensity. S and R may show a slight increase in power intensity with temperature, while W may show a slight decrease in power intensity with temperature. Since there are not clear inflection points, 10 F bins are used as done in Mathieu et al.

As a final step in exploratory data analysis, mean and standard deviation of power intensity are computed for each building.

Table 1: Power Intensity mean and standard deviation.

	Power Intensity (W/ft^2)	
Building	Mean	Standard Deviation
Roberts	6.2	0.4
Scaife	2.2	0.3
Warner	3.7	0.5

Roberts uses the most power per square foot, on average. Scaife uses the least, while Warner is in between. Warner also shows the largest variance in power usage, while Scaife shows the

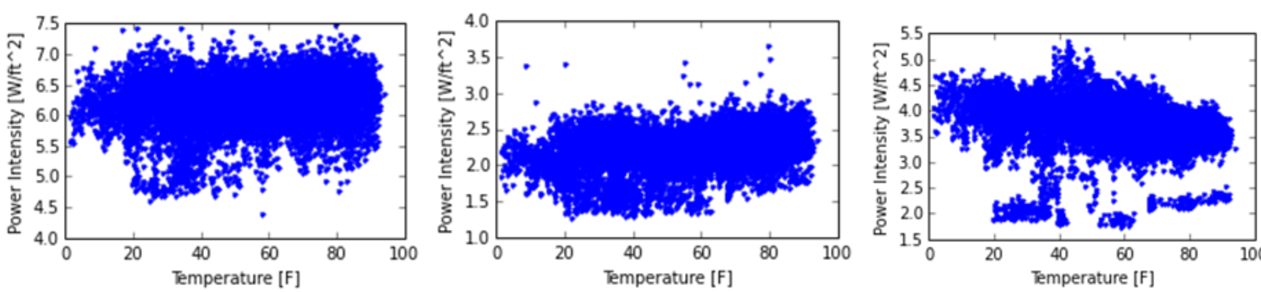

Figure 3: Power Intensity vs Temperature for each building during business hours - from left to right - Roberts, Scaife, Warner.

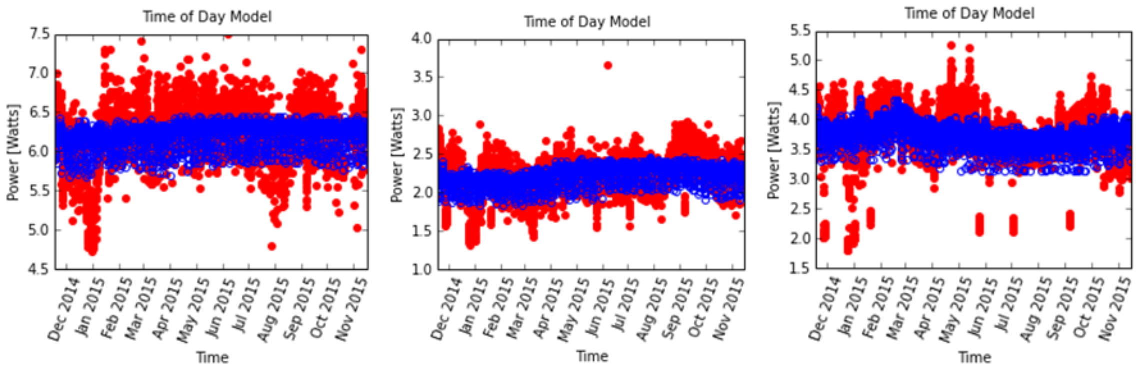

Figure 4: Predicted (blue) vs actual (red) power intensity for - from left to right – R,S, and W.

least. Roberts Engineering hall has a variety of different research labs for the Chemical Engineering, Electrical Engineering, Computer Science, and Materials Engineering departments. Scaife Hall is mostly lecture classrooms and offices, with a few research laboratories. Warner is staff and administration offices and facilities. Based on this, the results for power usage intensity make sense. These three buildings were originally grouped together as administrative and academic buildings into a single category, assuming they would have similar energy use. However, after performing EDA it is seen that engineering buildings with research labs can use nearly three times that of an administrative building or classroom based building.

2.2 Regression

Moving on to the results of the linear regression, actual vs predicted power intensities are examined graphically (Figure 4). Note that the graphs shown are only those for the time of day model – the time of week model is compared to this using the R squared analysis to follow, but had graphical results which were nearly identical.

The prediction shows less variance in values as compared to the actual power. The prediction for the most part fails to follow the dips and peaks of the actual data over the course of a year (although it can be seen doing that in some cases). Temperature is the only factor that can shift the prediction over more than a daily time frame, so the fact that there are changes shows temperature is being used in the model and is significant in some cases (the other dips and peaks are likely non temperature related). This is evident most in W's data.

The model captures the changes over the hours of the working day reasonably well (Figure 5) - although with a higher variability than actual.

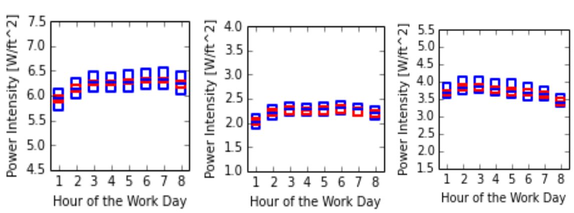

Figure 5: Box plots (in same style as Fig. 2) depicting predicted (blue) vs actual (red) power intensity over the workdays. From left to right – R, S, and W.

Table 2 shows results for the coefficient of determination.

Table 2: R Squared Values

R Squared Values

Building	Time of Week Model	Time of Day Model
Roberts	0.11	0.11
Scaife	0.20	0.20
Warner	0.18	0.19

The percent of output variation explained by the model is in the range 11-20%. Roberts has the lowest value while Scaife and Warner have higher values. This could be a reflection of the fact that Roberts has the research facilities which may consume power in a non-predictable manner (depending on large testing appliances running, clean room HVAC systems, etc). Finally, the overall low values suggest that predicting power based on time of day or time of week and temperature may be inadequate – other important factors influence power consumption, but what those are is unclear.

Additionally, r^2 results are important because they show that the time of week and time of day models performed virtually the same. Since the time of day model affords more data to the coefficient fitting (less indicator variables) this may be beneficial. It also makes sense since there doesn't seem to be much a change in diurnal power consumption over a week. From here on we only use the time of day based model.

For each of the regression coefficients in the coefficient vector $\hat{\beta}_K$, a 95% confidence interval is calculated to assess the variability of the relationship between time of day and power as well as temperature and power. Additionally, a hypothesis test is carried out to find if the null hypothesis $H_0\colon \hat{\beta}_K = 0$ can be rejected at different levels of confidence α.

With a time of day resolution of ten minutes there are 48 different time of day indicator coefficients and 10 temperature coefficients. For each building, the time of day indicator coefficients had confidence intervals that were on average 0.06 W/ft^2 from the predicted coefficient. The probability that each was zero was extremely small (on the order of 10^{-20}).

Looking at probabilities for the temperature coefficients (Table 6), it is seen that many of the coefficients were not significant in the model.

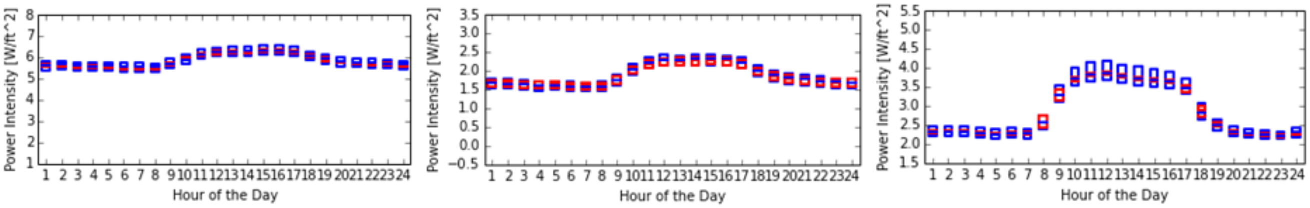

Figure 6: Box plots depicting predicted (blue) vs actual (red) power intensity over the 24 hour day. From left to right – R, S, and W.

Table 3: Temperature coefficient significant test results

	P-values		
Temperature Range	R	S	W
0-10	0.39	0.48	0.44
20-Oct	0.27	0.21	0.5
20-30	0.33	0	0.5
30-40	0.48	0.49	0.43
40-50	0	0.1	0
50-60	0.4	0	0.5
60-70	0.02	0	0.21
70-80	0.32	0.43	0.5
80-90	0.33	0	0.01
90-100	0.3	0.17	0.31
Total > 0.05	8	6	8

In all, between 60% and 80% of the temperature coefficients for each building are not significant in the model. EDA suggested that power and temperature have no strong linear relationship, and this confirms that observation.

2.2.1 Extending to 24 Hours

Since it is seen that the model above was most reliant on time of day as a predictor, it was decided to investigate the effects of extending the model to all 24 hours of each day of the work week (M-F) without having separate regression coefficients to differentiate occupied vs unoccupied as done in Matheau et al. The box plots showing mean and IQR of power intensity vs hour of the day appear in Figure 6.

The box plots show that the predictions follow the actual but with more variance. The R squared values for the new model appear below:

Table 4: R squared values for extended model

Building	**Extended Model R Squared**
Roberts	0.43
Scaife	0.68
Warner	0.81

There is a significant increase in the amount of power intensity variation that is explained by the explanatory variables when the model is applied over the whole day. This makes sense considering the higher variability of the output during a 24 hour period as opposed to an 8 hour working day period.

3. Conclusion

This analysis compared the power intensity in Richards Engineering Hall, Scaife Hall, and Warner Hall. It was found that Richards has the highest power intensity, followed by Warner and then Scaife. Exploratory data analysis showed that the peak hours of power intensity were during business hours (9 AM to 5 PM), and that there was no significant linear relationship between power and temperature during this time. EDA also showed that the power intensity had a strong diurnal pattern which was more or less repeated throughout the weekdays. The regression model described in Matheau et al was modified based on the EDA findings such that time of day was used as an indicator instead of time of week – the resulting model performance was nearly identical. The piece-wise linear multivariate regressive model was able to predict power consumption as it varied during business hours well, but was less able to predict changes that occurred throughout the course of the year. Time of day was significant as a predictor in the model while outdoor temperature was not. Finally, the model was applied to all 24 hours of the weekdays, resulting in a higher coefficient of determination and the ability to predict power as it varies over any hour of the day.

Regarding energy efficiency of the buildings, the only conclusion that can be reasonably drawn is that there is essentially zero power dependence on temperature. This may indicate HVAC systems that are not optimized – one would expect some differences in the significant heating and especially cooling HVAC load as temperature changes.

The model presented here has significant limitations in that temperature and time of day are the only predictors of power. While modifying the Matheau et al model, week of year was considered as an indicator, but was rejected considering the pattern of power consumption over the weeks of a year could vary from year to year. A model that considered other factors contributing to power intensity such as hours of day light, breaks in the academic schedule, etc could greatly improve the models ability to predict power intensity.

Additionally, temperature and time of day are correlated variables – an interaction between them in the model would be appropriate. This would result in a significant increase in the number of explanatory variables, model complexity, and amount of data required. However, it could help isolate the effects of temperature and time of day.

4. ACKNOWLEDGMENTS

The data used in this analysis was supplied by Dr. Bertrand Lasternas and The Center for Building Performance and Diagnostics at Carnegie Mellon. It contains power readings at 10 minute intervals for 31 different buildings on campus from Nov. 17th 2014 to Nov 17th 2015. Also provided was temperature data at 5 minute intervals from Nov. 1st 2014 to Dec. 1st 2015.

I'd also like to acknowledge my Professor, Mario Bergés for teaching the fundamental of data analysis to me and for considerable help outside of class. I'd also like to acknowledge teaching assistant Jingkun Gao for assisting me.

5. REFERENCES

[1] J. L. Mathieu, P. N. Price, S. Kiliccote and M. A. Piette, "Quantifying Changes in Building Electricity Use, With Application to Demand Response," *IEEE Transactions on Smart Grid,* vol. 2, no. 3, 2011.

[2] "Carnegie Mellon Rooms Specification(s)," Carnegie Mellon University, [Online]. Available: https://fst-apps.andrew.cmu.edu/roomspec/roomspec.