Building Electricity Use at Carnegie Mellon University

John Carchi (jcarchi)
Courtney Chambers (cchamber)
Samuel Kim (samuelk1)

ABSTRACT

Optimizing electricity usage in buildings is one of the main challenges for facility managers. This becomes extremely difficult when multiple buildings are involved, like on college campuses. A good way to increase energy efficiency is by knowing when energy is required in certain areas. We use a linear-regression model to make predictions about the electricity demand in buildings on campus. This model was created by some members of the Institute of Electrical and Electronics Engineers. In this paper, we compare the regression models for four buildings (Baker Hall, Porter Hall, Scaife Hall, and Doherty Apartments) on Carnegie Mellon University's Campus.

One dataset, called "campusDemand.csv," gives the electrical power consumption measurements for different buildings on campus during varying times in 2014. The second dataset, called "temperature.csv," gives the temperature on campus at 5-min-intervals during the years of 2014. First, we harmonize the two datasets into 15-min-intervals. Next, we separate the data into occupied and unoccupied times and then separate the data into train data and test data. Linear-regression analysis on the train data provides a predictive model for the electricity load for the four buildings. We compare the electricity load from the predictive model to the electricity load from the test data to evaluate the models effectiveness.

The R^2 values for Baker Hall, Porter Hall, Scaife Hall, and Doherty Apartments are 0.91, 0.87, 0.83, and 0.52, respectively. After calculating the R^2 values for each building's model, our analysis shows that the academic buildings (Baker, Porter, and Scaife) are better representations of the actual electricity loading than the residential building (Doherty). One reason for this difference is that the electricity is more regulated and controlled by the university while students have more control over electricity consumption in residential halls and have less predictable habits.

Some observations can also be made from the confidence intervals of the alpha and beta coefficient. A 95% confidence interval for these coefficients shows that the alpha interval is larger than the beta intervals. This could be due to the fact that there are 480 alpha coefficients and only six beta coefficients. Another observation is that the Baker Hall alpha interval has a larger and higher range and the smallest beta interval than the other buildings. The reason for a large alpha range may come from the fact that Baker Hall consumes the most electricity out of the four buildings. The small beta range may be due to the lack of change in electricity use due to temperature changes.

Lastly, the sum of the electricity usage from these four buildings are compared to the total main campus electricity usage. The results show that these four buildings only make up 7% of the total campuses electricity usage. This means that any energy savings in these four buildings will have little impact. In conclusion, meters should also be placed in other buildings to get a better sense of which buildings contribute the most to the electricity consumption on campus.

Table: 95% Confidence Intervals for each building

95% Confidence Intervals		
Building	Alpha Coefficient Interval	Beta Coefficient Interval
Baker Hall	[111, 276]	[154, 155]
Doherty Apt.	[-6, 12]	[3, 8]
Porter Hall	[0, 86]	[24, 31]
Scaife Hall	[56, 98]	[67, 72]

Keywords
Energy efficiency, regression analysis, linear-regression

1. INTRODUCTION
Understanding electricity usage in buildings is an important topic of discussion for many building owners and managers. This understanding can help improve energy efficiency and reduces energy cost. Energy efficiency has posed a challenge for many professionals in the Building Energy Management field. Many data acquisition and analysis techniques are needed to help building managers optimize the performance their facilities. Optimizing electricity usage is not an easy task. There are many different obstacles to face when collecting data, harmonizing data, analyzing data, and testing the models.

In "Quantifying Changes in Building Electricity Use, With Application to Demand Response" written by members of Institute of Electrical and Electronics Engineers (IEEE), they present methods for analyzing 15-min–interval electric load data from commercial and industrial facilities. They created a regression-based model to predict the electricity load of a building during "occupied" and "unoccupied" case. [1]

In our paper, we will be using the methods previously stated in the IEEE paper and apply them to some buildings on the Carnegie Mellon University campus. The purpose of our project is to compare the regression models for Baker Hall, Porter Hall, Scaife Hall, and Doherty Apartments and consider reasons for why there are similarities or differences. We will also observe how the electricity load in these buildings relate to the electricity load for the entire campus.

2. DATASETS
The datasets used for this project are "campusDemand.csv" and "temperature.csv." The campusDemand.csv file contains electrical power consumption measurements for different buildings on campus during the year 2014. The temperature.csv file contains temperature measurements for the same period of time as the campusDemand.csv dataset, collected by a weather station in the Margaret Morrison building.

The campus demand dataset collected data for electrical power consumption from several different buildings. The electrical power consumption used for the project are from buildings on Carnegie Mellon University campus. The main campus data provides the overall electrical power consumption in units of (kW). Electrical power consumption data was also collected in (W) for Porter Hall, Baker Hall, Scaife Hall, and Doherty Apartments. The data for Baker Hall was collected from 2014-02-12 to 2014-11-10 for 271 days. The data for Doherty Apartments was collected from 2014-10-10 to 2014-11-10 for 31 days. The data for Porter hall was collected from 2014-09-10 to 2014-11-10 for 61 days. The data for Scaife Hall was collected from 2014-10-10 to 2014-11-10 for 31 days. The data for the main campus was collected from 2013-11-10 to 2014-11-10 for 365 days.

The data for each building was not collected for the same amount of time, so data analysis is done for a month where data collection overlaps for the different buildings. Power consumption was collected every minute.

The temperature dataset collected the temperature measurements during the same period of time; however, the data collected was in timestamp intervals of 15 minutes. The temperature was collected for the overall campus temperature in Fahrenheit.

The overall data analysis uses the measurements at 15 minute intervals to align the data points."

3. Results

The four buildings we chose to model were Baker Hall, Doherty apartments, Porter Hall, and Scaife Hall. The data we obtained measured the power consumption in Watts for each building. Data points were taken from the time period of October 13th, 2014 to November 10th, 2014. Approximately a month was examined across all four buildings. The later sections further describe the models generated that would approximate the energy consumption for each building.

3.1 Baker Hall

Figure 1 shows the power consumption for Baker Hall from October 13th to November 10th in 2014. The maximum amount of energy consumed appears on October 28th, 2014 at about 308 kW. The minimum amount of energy consumed appears on October 18th, 2014 at about 156 kW.

Figure 1: Baker Hall Power Consumption

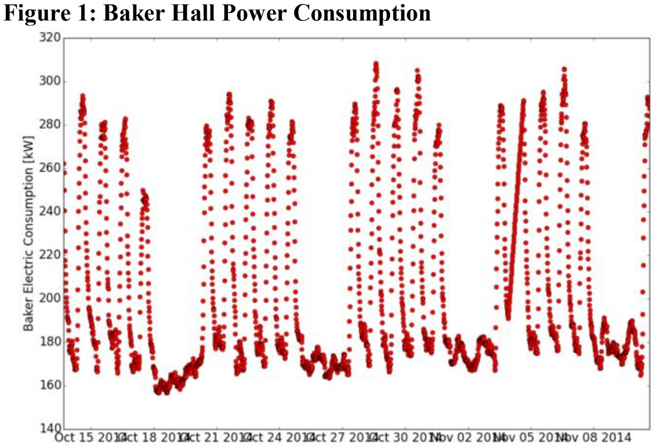


Figure 2 and figure 3 shows the generated model against the training data and the test data, respectively. The model is a good fit for the testing data, measuring an R^2 value of 0.911. The confidence interval for the alpha coefficients is [110.70, 275.88]. The confidence interval for the beta coefficients is [153.55, 154.81].

Figure 2: Baker Hall Generated Model vs. Testing Data

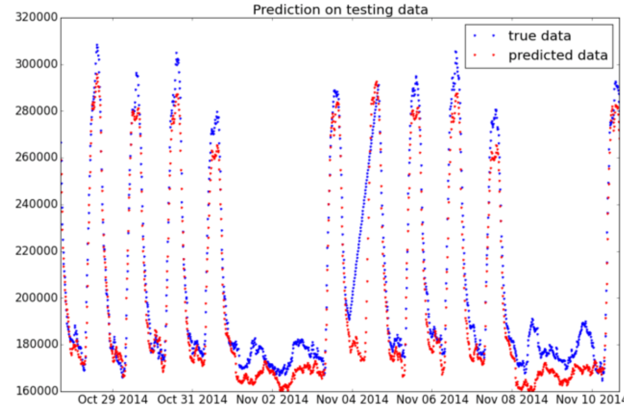
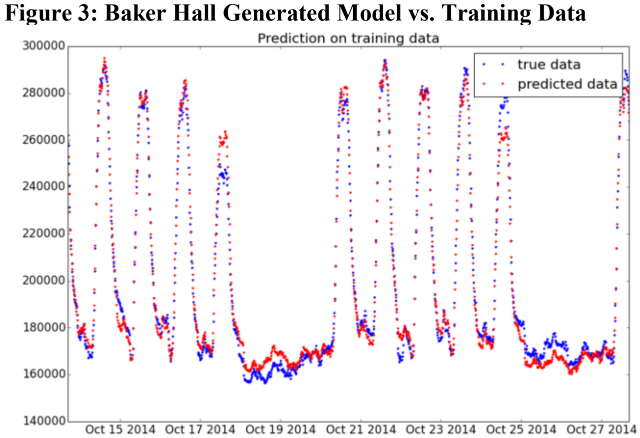


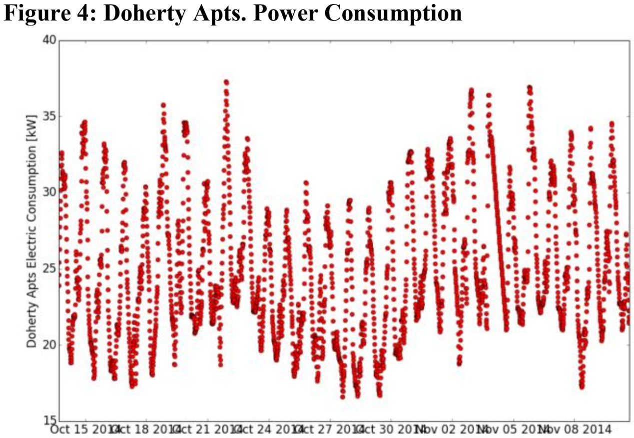
Figure 3: Baker Hall Generated Model vs. Training Data



3.2 Doherty Apartments

Figure # shows the power consumption for Doherty Apartments from October 13th to November 10th in 2014. The maximum amount of energy consumed appears on October 21th, 2014 at about 37.3 kW. The minimum amount of energy consumed appears on October 27th, 2014 at about 16.6 kW.

Figure 4: Doherty Apts. Power Consumption

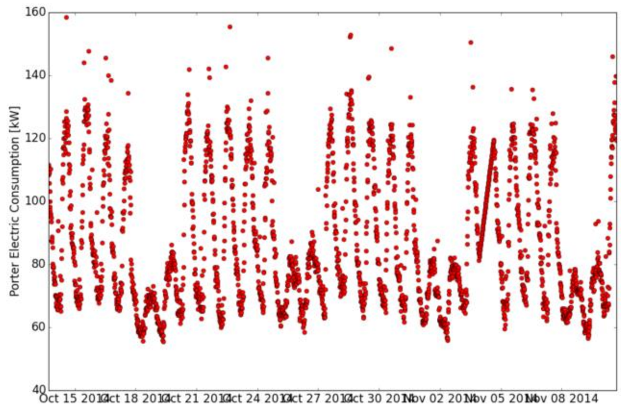


The graphs showing the generated model against the training data and the test data are shown in the python notebook attached. The model is not a good fit for the testing data, measuring an R^2 value of 0.521. The confidence interval for the alpha coefficients is [-5.617, 11.85]. The confidence interval for the beta coefficients is [3.19, 7.65].

3.3 Porter Hall

Figure 5 shows the power consumption for Porter Hall from October 13th to November 10th in 2014. The maximum amount of energy consumed appears on October 14th, 2014 at about 158.37 kW. The minimum amount of energy consumed appears on October 19th, 2014 at about 55.41 kW.

Figure 5: Porter Hall Power Consumption

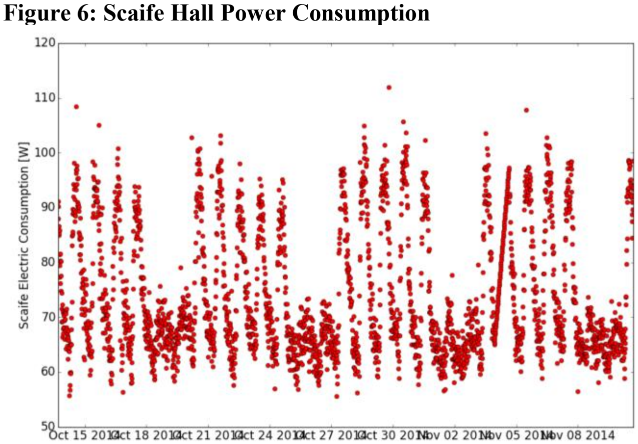


The graphs showing the generated model against the training data and the test data are shown in the python notebook attached. The model is a good fit for the testing data, measuring an R^2 value of 0.868. The confidence interval for the alpha coefficients is [0.259, 85.98]. The confidence interval for the beta coefficients is [24.39, 31.04].

3.4 Scaife Hall

Figure 6 shows the power consumption for Scaife Hall from October 13th to November 10th in 2014. The maximum amount of energy consumed appears on October 29th, 2014 at about 111.98 kW. The minimum amount of energy consumed appears on October 27th, 2014 at about 55.56 kW.

Figure 6: Scaife Hall Power Consumption

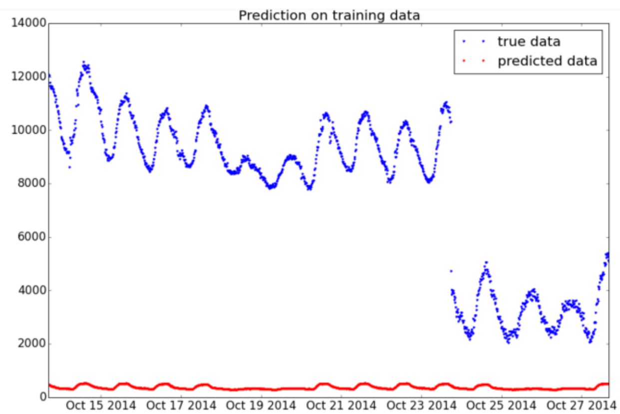


The graphs showing the generated model against the training data and the test data are shown in the python notebook attached. The model is a good fit for the testing data, measuring an R^2 value of 0.828. The confidence interval for the alpha coefficients is [55.59, 97.50]. The confidence interval for the beta coefficients is [66.57, 71.90].

3.5 Comparison with Main Campus

The estimated power consumption for Main campus was derived from the summation of the predicted power values from all four models. Figure 7 shows the comparison between the main campus testing data and the generated model results. The main campus power consumption is much larger than the combined model of all four buildings in the month of interest. There is a significant drop in the end of October and for the month of November that is closer to the model results but still, much higher in terms of power consumption.

Figure 7: Four Building Power Consumption vs Main Campus



4. ANALYSIS

4.1 Separate Building Models

The model for Baker is the best fit for its respective data. The model for Doherty is the worst fit. The model for Baker has more timestamps within the month of interest, requiring little to no data interpolation to fit the measured power values. The model for Doherty has less timestamps, taken at varying times within the day, requiring more interpolation and introducing more error. The models for Porter Hall and Scaife Hall were below Baker but still represent a good fit, averaging at about a 0.85 R^2 value.

The model for Baker has the smallest confidence interval range for the beta coefficients and the largest confidence interval range for the alpha coefficients. This could mean that the temperature change does not affect electricity consumption as much in Baker. The higher interval range may be due to the fact that Baker consumes the most electricity out of the four buildings. A wider confidence interval indicates that there is less certainty in estimating the range of coefficients. Doherty Apt. had narrow ranges for both the alpha and beta coefficients, which means if this test was conducted again, then it is very likely to have similar results.

4.2 Comparison with Main Campus

The model generated from all four buildings represents a small percentage of the power consumption of the entire campus, about 7% of the total consumption. Looking at Figure 7, the rapid dip in power still remains well above the predicted values from our model. The dip could be from buildings shutting down or an error in the acquisition of the data for that time period.

5. CONCLUSION

The results of our data analysis show that most of the model of power energy consumption fits well into the test data over the period of a month.

Table: R^2 Values for Campus Buildings

Building	R^2 Value
Baker Hall	0.91
Porter Hall	0.87
Scaife Hall	0.83
Doherty Apartments	0.52

The R^2 values are close to a value 1.0 for the academic buildings. However, the R^2 value for Doherty Apartments is 0.52, which is a low value. The model data for the academic buildings can be used to predict future energy consumption graphs around the same temperature months. However, Doherty Apartments shows to have a poor fit for the model. This is likely due to the fact that power energy consumption is more difficult to predict for residential buildings than an academic building. An academic building is much more likely to have energy consumption levels controlled on a more regular schedule. Since individual residents can have different energy usage on a daily basis, this is more difficult to predict.

The data for the four buildings, Porter, Baker, Scaife, and Doherty apartments, combined show a resulting electrical power consumption that is much less than the overall campus power consumption. This result shows that the power consumption in these four buildings are not a significant part of the power consumption throughout campus.

An analysis of the main campus data shows a significant drop in power consumption around October 23, 2014. There are several potential reasons for the drop in power consumption. One potential reason could be a technical error in the sensors. A large group of sensors possibly malfunctioned and were not managed. In conclusion, increased monitoring of data collection systems can help facility managers optimize electricity usage in buildings.

6. ACKNOWLEDGMENTS

Our thanks to Mario Berges for teaching the course 12-752 Special Topics: Data-Driven Building Energy Management and allowing us to modify code he had developed

7. REFERENCES

[1] Mathieu, Johanna L., Phillip N. Price, Sila Kiliccote, and Mary Ann Piette. "Quantifying Changes in Building Electricity Use, With Application to Demand Response." IEEE Trans. Smart Grid IEEE Transactions on Smart Grid 2.3 (2011): 507-18. Web.

[2] Carnegie Mellon University. (2014). *campusDemand.csv* [Comma Separated Values File]. Available from 12752 Blackboard site Web site: https://blackboard.andrew.cmu.edu/webapps/blackboard/content/listContent.jsp?course_id=_6577658_1&content_id=_971002_1

[3] Carnegie Mellon University. (2014). *temperature.csv* [Comma Separated Values File]. Available from 12752 Blackboard site Web site: https://blackboard.andrew.cmu.edu/webapps/blackboard/content/listContent.jsp?course_id=_6577658_1&content_id=_971002_1

[4] Berges, Mario. (2015) *12-752_Assignment_2_Sol.ipynb* [IPython Notebook] Available at https://blackboard.andrew.cmu.edu/webapps/blackboard/content/listContent.jsp?course_id=_6577658_1&content_id=_973022_1 (Accessed 8 December 2015)