Developing an Accurate Baseline for Electricity Consumption, Focusing on University Residence Halls

Richard Turner

Mechanical Engineer
UNC Chapel Hill Energy Management
turnerrwjr@gmail.com

5/1/2012

Abstract

The business of saving energy has begun to truly boom. And in a business of savings, it is important to know where you are and where you would have been. Unfortunately, the most common utility, electricity, is one of the most difficult to predict. This report outlines the approaches taken to develop a better baseline for electricity consumption in UNC Chapel Hill residence halls. Though the approach will be applicable for a wide range of building types, the specific findings will mostly only apply to residence halls. Before beginning this research, electricity consumption was predicted based on the outside air temperature. The typical R^2 value (number describing how closely related two sets of data are, 0 = unrelated, 1 = perfectly identical) between predicted usage and actual usage during the baselining period, where ideally they would be identical, was around 0.4. The common standard R^2 value defining when two sets of data have a statistically viable relationship is 0.75. This meant the current electricity baseline was highly inaccurate. However, this is a typical result for a residence hall type building due to the large usage coming from lights and plug loads which are difficult to predict. The goal of this research was to bring the R^2 value up to or above 0.75. It was decided that the best way to achieve this would be to consider multiple variables instead of just outside air temperature. At first, different combinations of variables were tested until six were decided on: heating degree days, cooling degree days, occupancy (also considering course load), cloud coverage, average wind speed, and hours of day light. Data was analyzed by month and baselines were developed through a programmed Excel spreadsheet. The baselines developed returned an average R^2 value of 0.819 across 19 residence halls. The relevance of each variable was also calculated throughout the process, and it was found that average cloud coverage and hours of daylight were only relevant in a small fraction of the buildings.

Table of Contents

Introduction

The business of saving energy has begun to truly boom. And in a business of savings, it is important to know where you would be and where you are. Unfortunately, the most common utility, electricity, is one of the most difficult to predict. Electricity consumption can be affected by a wide variety of factors that have different levels of impact. Heating and air conditioning consume a large percentage, but lighting and plug loads make up the remaining consumption and are very difficult to predict.

In order to find energy savings after the implementation of any form of energy conservation attempt, there are two things that must be known: what the building uses now and what it would have used had nothing been done. The ladder comes from a baseline which is the predicted energy use. A baseline is formed by finding what factors or variables affect the buildings energy use, such as outside temperature, and using them to predict what the energy usage would be. This is normally done by plotting the consumption versus the variable on a graph and calculating the line of best fit. The graph below shows how it is done with chilled water.

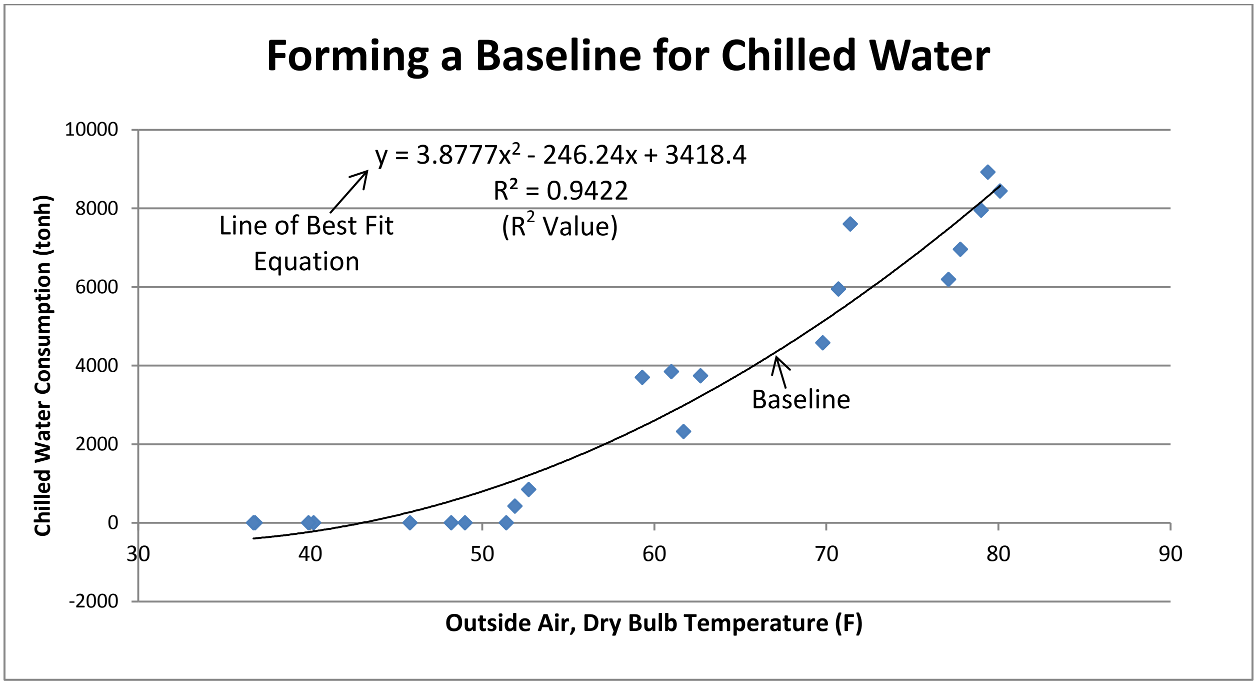


Figure 1: Baselining

By using the line of best fit equation, anyone trying to find out how much chilled water is being saved now can simply input the average outside air temperature as an "x" value, calculate what the chilled water consumption would have been ("y" value), and subtract the current consumption from the baseline consumption to get the savings. The other important variable on the graph is the R^2 value. The R^2 value is a number describing how closely related two sets of data are where a value of 0 = unrelated and a value of 1 = perfectly identical. This means that in the graph above, the baseline is very close to the actual data and is thus a reliable, viable baseline. Generally, a R^2 value of greater than 0.75 is considered statistically viable, or defining a useable relationship.

This study stemmed from a project to find energy savings in college residence halls. The original method of predicting electricity consumption used outside air temperature and did not return good results. The typical R^2 value found was around 0.4 and even though that result is common, it is still highly inaccurate and the goal became to increase the accuracy of the predictions using the best method available. It was decided to consider the affect of multiple variables simultaneously and see how well predictions could be improved and which factors had the most consistent impact on electricity consumption. Because this study focuses on residence hall electricity consumption, the methods discussed in this report could be applied to nearly any building type, but the specific results discussed will only be applicable to residence halls and very similar building types.

Choosing Wisely

The typical approach for measurement and verification of electricity usage at UNC Chapel Hill is to use dry bulb temperature, date, or occupancy. Unfortunately, as mentioned before, though these approaches match that of most others in the industry, they are still unreliable and do not give good dependable data. This fact turned the relatively simple task of finding energy savings in residence halls to a much more challenging task of developing a better way to baseline electricity consumption.

This task was taken on with very little prior knowledge and there was very little help to be found elsewhere. As a result, the first task was to lay out the basic parameters. Electricity consumption data was pulled from monthly billing. This meant that all data being used would be a monthly sum or average. To keep monthly sums from being skewed by moths with fewer days, all monthly sums were divided by the number of days in that month. This gave an average value per day per month and leveled the playing field from month to month. All data was pulled over a four year period. This allowed for 2.5 years of data to be used in developing a baseline that could then be used to find savings over the following 1.5 year period (see Figure 1).

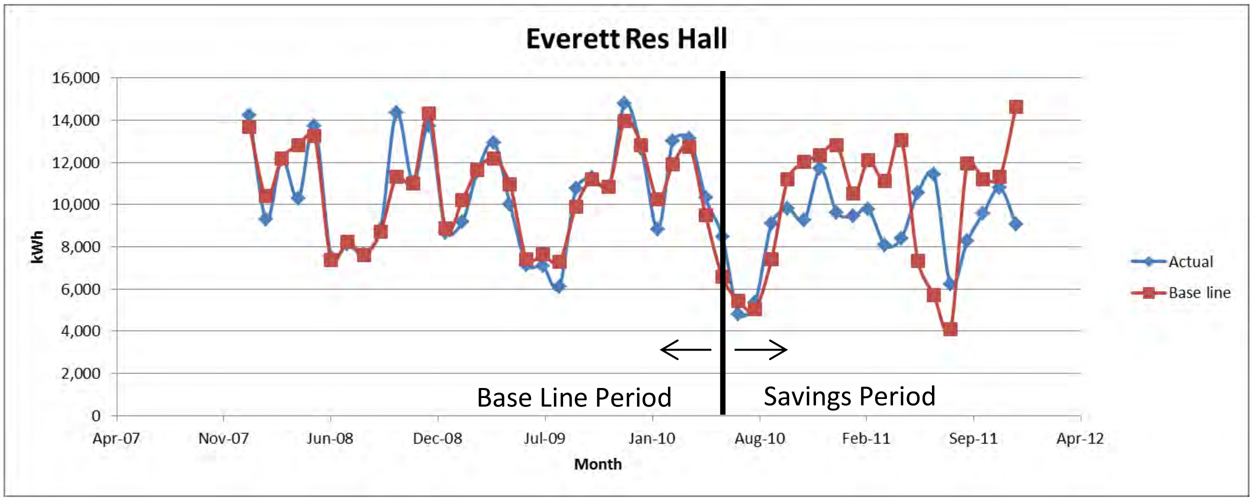


Figure 2: Data Periods

With the parameters established, the next task was to brainstorm possible influences in electricity usage and investigate their potential affect. To begin, the most common factors were included. In order to allow warm and cool days to influence predicted usage separately, it was decided to use monthly total heating degree days (HDD) and cooling degree days (CDD). Another common influence on usage is occupancy (OCC). In an attempt to customize this variable to better fit the task of predicting usage in a residence hall, each month was given a score between 0 and 5 intended to describe the amount of time students spend in the residence hall due to the number of days of the month school is in session and the average level of course load during those months. So, for example, July would receive a score of 0 due to class not being in session, and April would be given a score of 5 due to the heightened work load of exams and final projects at the closing of the semester.

After the common factors for electricity usage were incorporated, some less common factors were investigated. One was the average hours of sunlight (HDL) each month. It seemed to make sense that if it is darker for a larger percentage of the day, then students would need to use artificial lighting for longer periods of time, and would spend more time in doors. Another variable that provided surprisingly significant trends on its own was average wind speed (AWS). The reason for its relevance was thought to be due to increased heat transfer between the building and the outside air on windy days and the tendency of students to stay inside on windy days. Finally, cloud points (CP) are a score derived from the number of cloudy, partly cloudy, and sunny days. Cloud points were calculated based on the credentials of the data provided. The data was based on cloudy days being days of less than 20% sun, sunny days being days of greater than 85% sun, and partly cloudy days being the days in between. The equation used to calculate the cloud points was

$$\frac{(85 * sunny + 20 * cloudy + 55 * partly)}{days\ per\ month} \tag{1}$$

where the constants are the percentage of sun and the partly cloudy constant is the median of the sunny and cloudy constants. The thought process behind these units was that the more sun, the more time students spend outside, and the more heat the building picks up through radiation.

Turning Six to One

After deciding which factors to include, the next, and most important, question was how to combine them. The ultimate goal was to be able to input the value of each variable for each month and get out a baseline usage. In order to do this, it would be necessary to manipulate how much of an affect each variable has on the baseline and quantify that affect. Quantifying the level of impact of each variable should allow some insight as to which variables are the most important for baselining electricity consumption.

The solution developed was actually rather simple. The variables were combined in the equation

$$\left|\frac{HDD}{C1} + \frac{CDD}{C2} + \frac{AWS}{C3} + \frac{CP}{C4} + \frac{HDL}{C5} + \frac{OCC}{C6}\right| = Trend\ Unit \tag{2}$$

where each C# is a different constant. This way, the constants could be made larger or smaller and adjust the impact that each variable has on the total sum. This also alleviated any issues coming from differences in the magnitude of the values of different variables. Factors like CP that have large monthly values, compared to ones like OCC, can be compensated for by using the constant. Only one set of constants was developed for each building and then used for all of the months. By entering each month's value for the six different variables, a "Trend Unit" would be calculated for that month. Once calculated, a line of best fit was calculated using the trend units as the "x" value and the same month's corresponding consumption as the "y" value. When graphed it looks like:

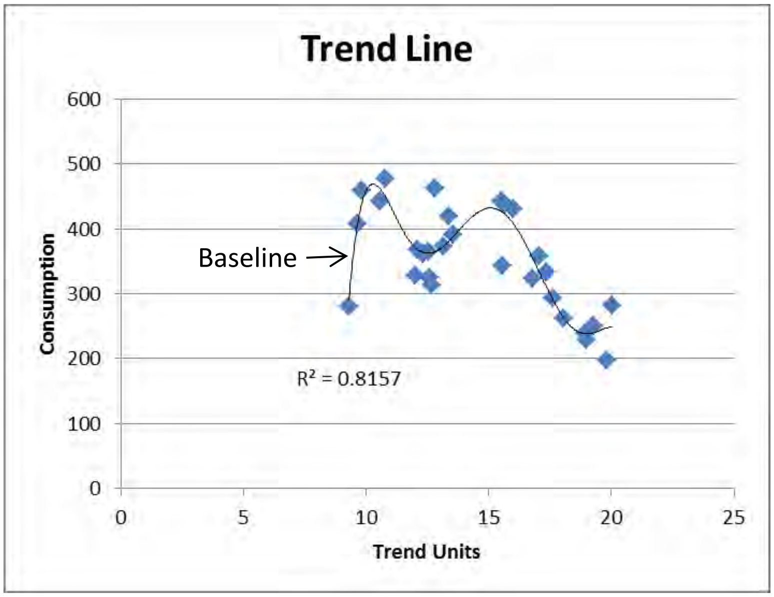


Figure 3: Trend Line

With the equation for the baseline, the trend unit for any month can be entered into the equation as an "x" value, and a very good estimate of that month's consumption will be calculated as the "y" value.

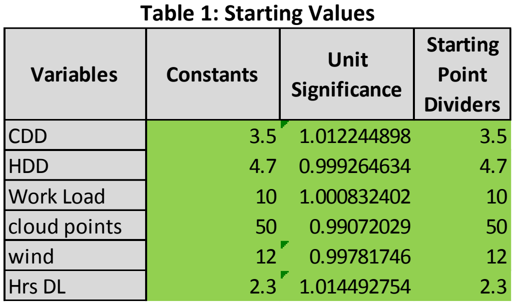
To find the significance of each variable, the average value for that variable over the four year period is divided by that variable's constant. This suggests, on average, which factor is making the greatest impact in equation 2.

Calculating the Constants

In order to calculate the constant, the baseline generator shown in the Appendix was built and programed in Microsoft Excel. The next few steps outline the process of finding the combination of constants that produce the best baseline:

1. The constants for each variable are set to equal the "Starting Point Dividers" (Table 1). The starting points make the significance of each variable equal to one. This is to keep the baseline generator as unbiased towards any one variable as possible. The "Trend Units" column automatically calculates the trend units as the constants are changed.

Table 1: Starting Values



Variables	Constants	Unit Significance	Starting Point Dividers
CDD	3.5	1.012244898	3.5
HDD	4.7	0.999264634	4.7
Work Load	10	1.000832402	10
cloud points	50	0.99072029	50
wind	12	0.99781746	12
Hrs DL	2.3	1.014492754	2.3

2. Next the monthly usage data, in kWh, is entered in its respective column and each month's consumption is automatically divided by the number of days per month in the "Units / Day" column.

3. The baseline generator then uses a built in function to calculate a sixth order, line of best fit equation between the data in the "Units / Day" column and the "Trend Units" column over the baseline period (Figure 3).

4. In the "6th Order Normalize" column, the equation calculated in step three is used to calculate the baselined, total consumption for each month.

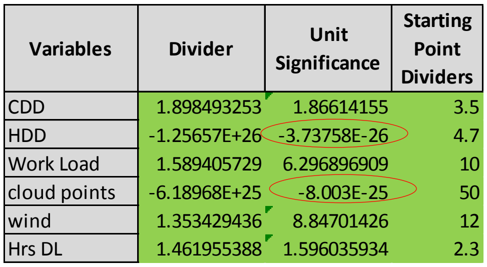
5. Then, looking at the baseline period, an imbedded R^2 function calculates how close the baseline consumption is to the actual consumption.

6. Having this R^2 value, a macro uses the goal seek function to adjust each constant so that it produces the largest possible R^2 value.

7. The macro has to be run multiple times since when one constant is changed; earlier constants may no longer be at their ultimate value. But after 3-6 times of running the macro, which takes about 10 seconds each time, the increase in the R^2 value becomes insignificant.

8. Sometimes at this point it is possible to see some variables that were insignificant in creating the baseline (Table 2). It can also be beneficial to test a very large constant for each variable to find if there are some others that are making an insignificant contribution.

Table 2: Irrelevant Variables



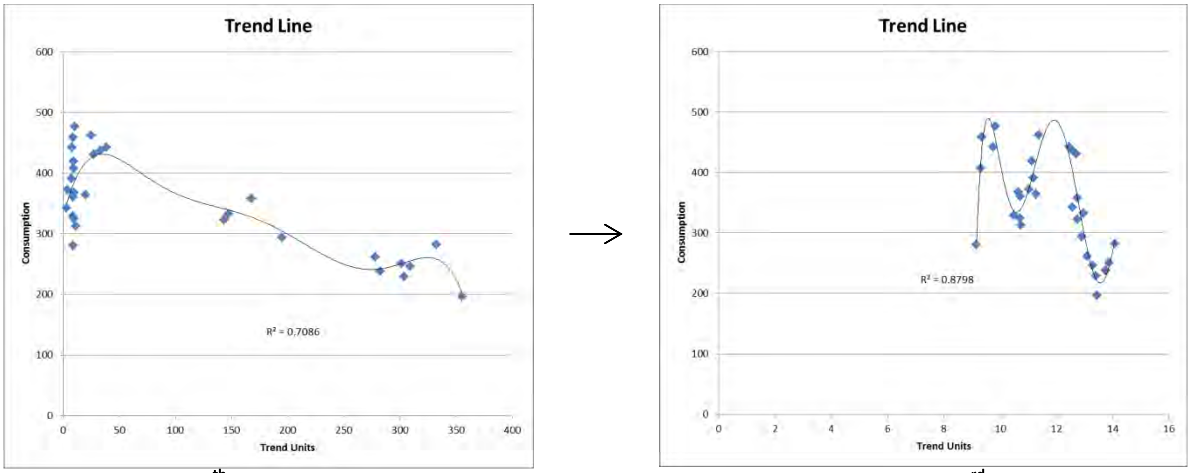
Variables	Divider	Unit Significance	Starting Point Dividers
CDD	1.898493253	1.86614155	3.5
HDD	-1.25657E+26	-3.73758E-26	4.7
Work Load	1.589405729	6.296896909	10
cloud points	-6.18968E+25	-8.003E-25	50
wind	1.353429436	8.84701426	12
Hrs DL	1.461955388	1.596035934	2.3

9. Because of how the line of best fit is formed, there tend to be extreme slopes at either end (see Figure 4 b.). This can lead to any predicted consumption, falling on either extreme, being blown

way out of proportion. For this study, if that occurred, the predicted value was replaced with the median value of the predictions for the month before and after that month.

10. With the best possible combination of constants reached, the baseline energy consumption numbers can be copied from the "6th Order Normalize" column and pasted wherever applicable.

The procedure above only uses the 6th order best fit equation, but the baseline generator also has the capability to use 5th and 3rd order equations. These can be helpful when pulling the equation itself and a 6th order equation is either too complicated or too sensitive. The 3rd and 5th order best fit macros can also be helpful when the 6th order tends to skew data. By running the 3rd or 5th order macro two to four times and then running the 6th order, the distribution and results can be significantly improved like what is shown in Figure 5.



a. **6th order only**

b. Starting with 3rd order

Figure 4: Advantage of 3rd Order Goal Seek

Results

It was with the baseline generator and the approach described above, that this study was successful and achieved some very interesting results. For the 19 buildings investigated, R^2 values ranged from 0.55 - 0.9. When summed together as one group, the buildings attained a total R^2 value of 0.819, surpassing our goal. The graph of the data in Figure 6 clearly shows a statistically viable baseline that stays very close to actual usage during the baseline period.

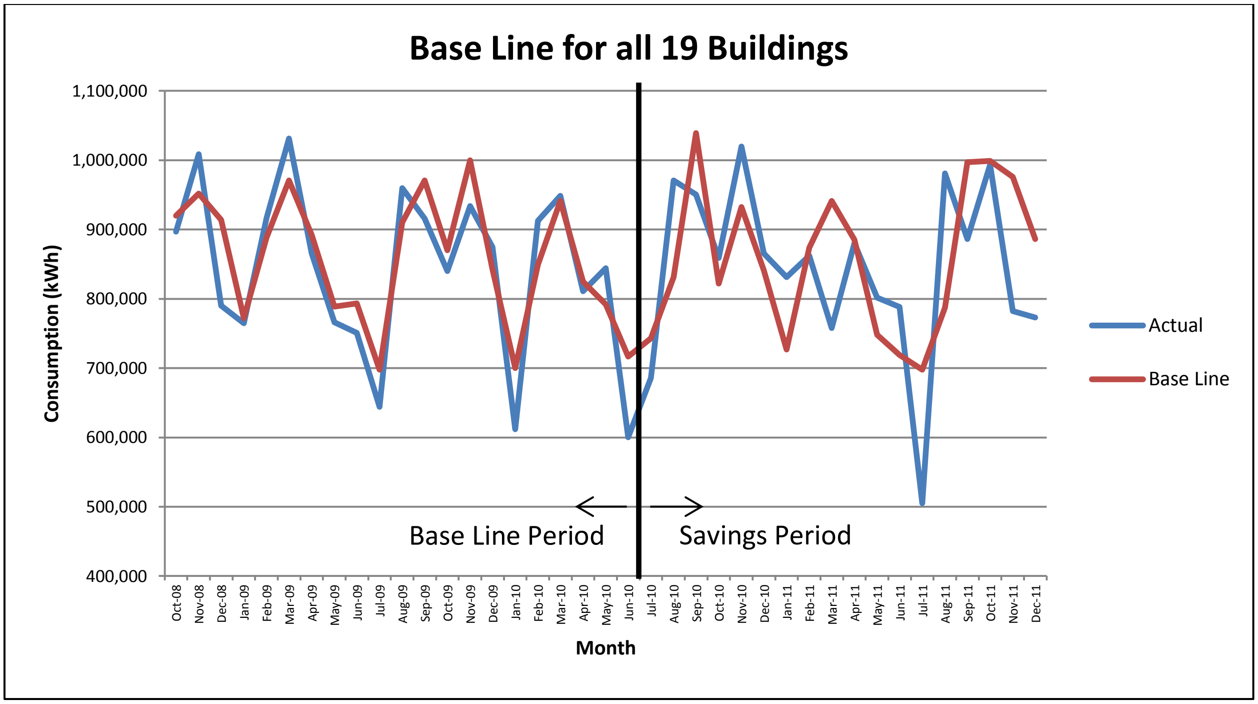


Figure 5: Baseline for All Buildings

Along with actually creating the baseline shown above, the relevance of each of the different variables was also investigated. A variable was considered irrelevant if, after reaching the maximum R^2 value, it's constant could be replaced with 1E10 and the R^2 value was not reduced by more than 0.02. Table 3, shown below, lists how many buildings out of 19 ended up not using the specific factor, and the graph below shows each building, its relevant factors, and how significant each factor was.

Table 3: Irrelevant Variables

Building	CDD	HDD	OCC	HDL	CP	AWS
# Times Irrelevant	5	7	1	11	14	4
% of Buildings	26.32%	36.84%	5.26%	57.89%	73.68%	21.05%

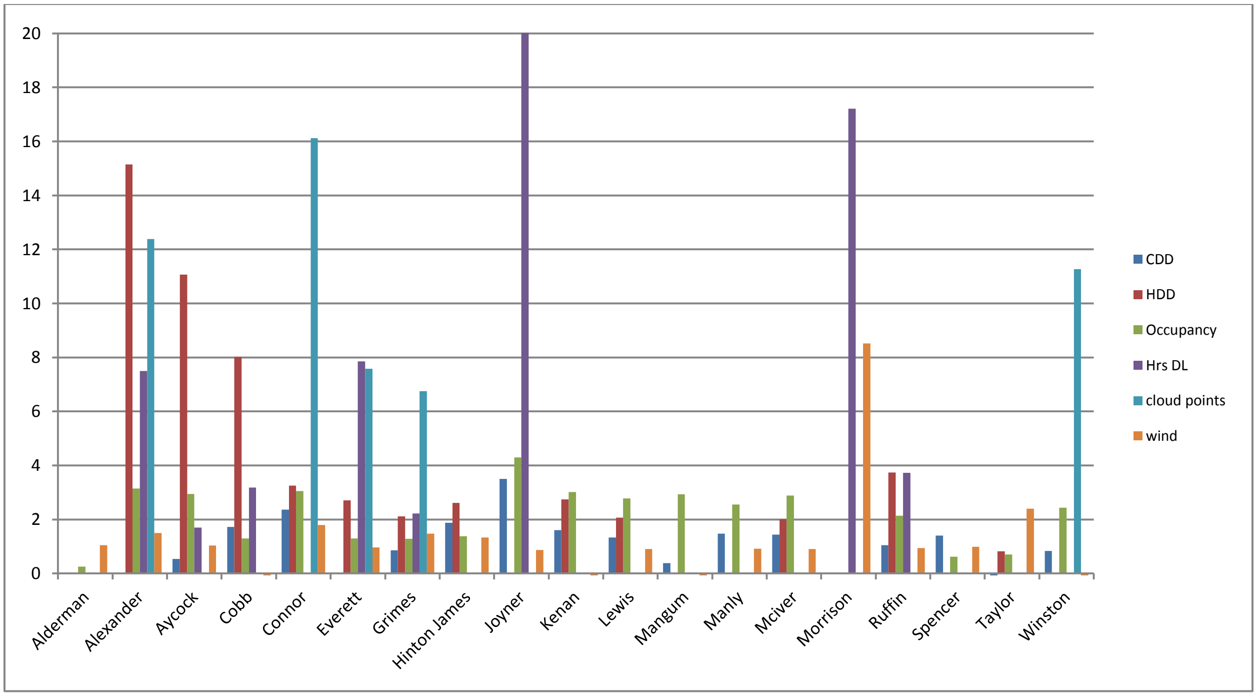


Figure 6: Variable Significance

When looking at the graph and table, there are a few interesting outcomes. For one, the table shows that average wind speed and cloud cover were only involved in less than half of the buildings. Based on that fact, it is likely that the two factors could be omitted when calculating baselines in the future. CDD, HDD, and OCC were no surprise as they are the typical variables in most baselining by measurement and verification companies. It was a surprise how well the occupancy, that was adjusted for course load, worked as one of the variables. In fact, at one point it was decided to change it to purely represent how many days of the month class was in session, but this had a significant negative affect on the baselines being developed. The biggest surprise was the importance of average wind speed. It was usually one of the smaller contributors based on the significance, but it almost always made an impact.

Analysis of Results

Through conducting this study, some very important information was found and some important lessons were learned. When it comes to the baseline itself, numbers don't lie. This study proved that it is possible to have an electricity baseline with an R^2 value greater than 0.75. However, there are some difficulties due to how it was obtained. The baseline generator was a great tool, but needs farther refinement. The results were rather inconsistent and could be significantly different depending on which macro is used and what value is entered as the desired R^2 value. Along with inconsistencies, this also greatly increased the time required to find the best possible baseline. It is the time that actually creates the larger issue for most applications. Even UNC Energy Management department has to baseline over 130 buildings once a year. This would be an extremely time consuming task using the specific method discussed in the paper. However, with the omission of the CP and HDL variables, the process could be much more efficient.

Another important and reliable result is count of the number of times a variable was insignificant. If a variable is removed and it doesn't make a difference, then once again, the numbers don't lie; it wasn't making an impact. However, the baseline generator could also be rather inconsistent as to which variables it made insignificant. But due to the use of 19 buildings, whether or not there are some inconsistencies, it is pretty clear which variables had a tendency to be important and which ones did not.

The result that is the least reliable is the graph of significances (Figure 5). These values had a tendency to spike and change significantly if a baseline was calculated more than once. Also, the significance of a variable turned out to be very loosely related to the variables impact on the R^2 value of the baseline. It was thought that the variable significance would be the best way to show the importance of each variable in developing a base line, but if this study was to be redone, it would be much more useful to show how much the R^2 is changed when each variable is canceled out with a very large constant. However, there is still some information to be observed from Figure 5 such as the consistency of the significance of AWS and OCC and the inconsistency of the impacts of the other variables.

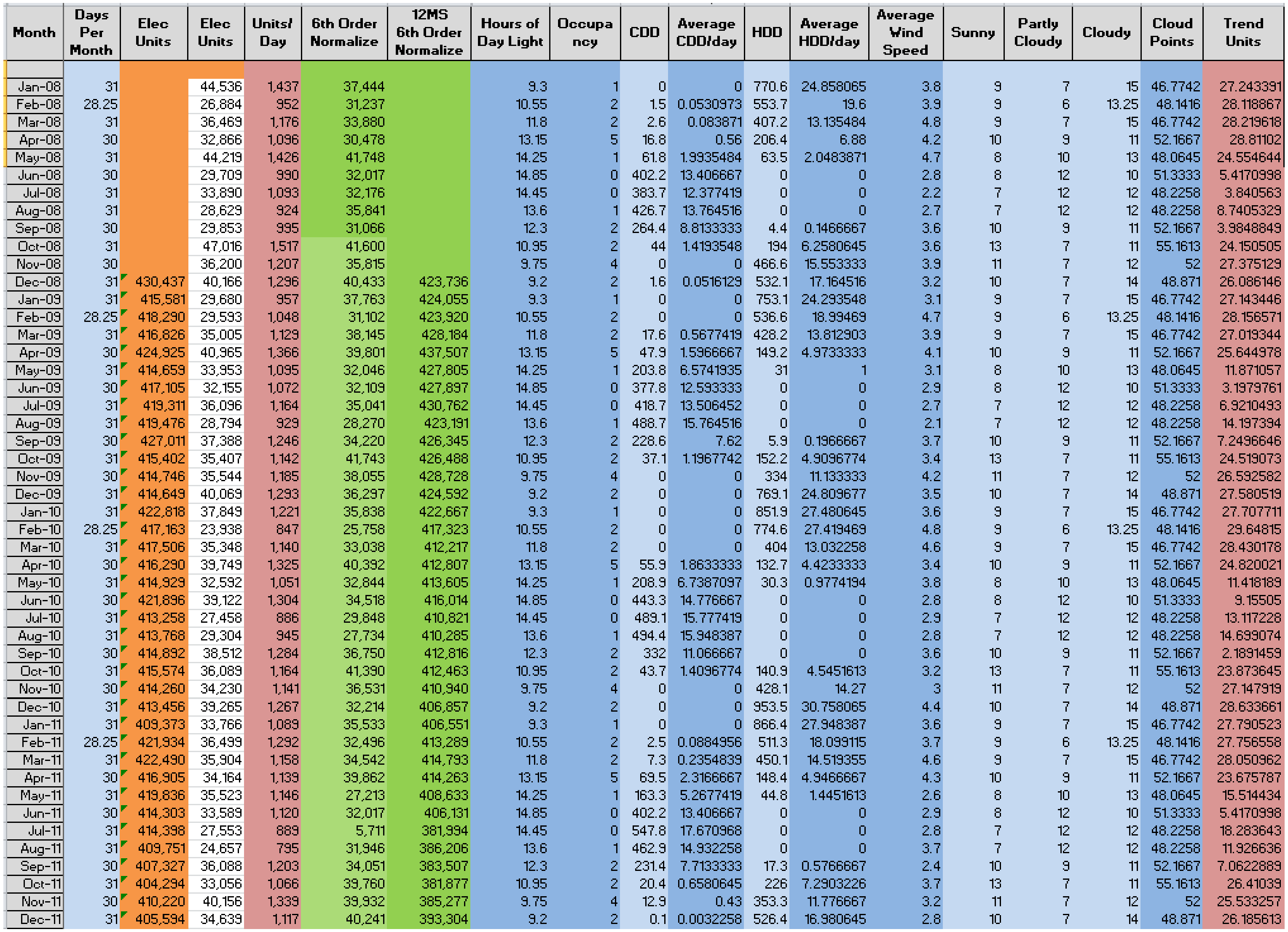
Conclusion

When calculating energy savings it is important to know where energy consumption is and where it would be. Unfortunately, the most common utility, electricity, is also one of the most difficult to predict. Though greater than half of energy consumption comes from heating and cooling in most buildings, the rest comes from extremely inconsistent sources such as lighting and plug loads. This is why this study was conducted to find the combination of factors in college residence halls that would allow for an R^2 value greater than 0.75.

This study produced a lot of truly interesting results. The most important, was that baselines for electricity can be brought into the realm of statistical significance by surpassing a R^2 value of 0.75. It was found that if this task was to be run again, the CP and HDL factors could most likely be eliminated and increase the applicability of this approach by making it simpler and more efficient. It was found that considering school work levels as part of occupancy can improve electricity baselines, and that the month's average wind speed is another very consistent factor in electricity consumption.

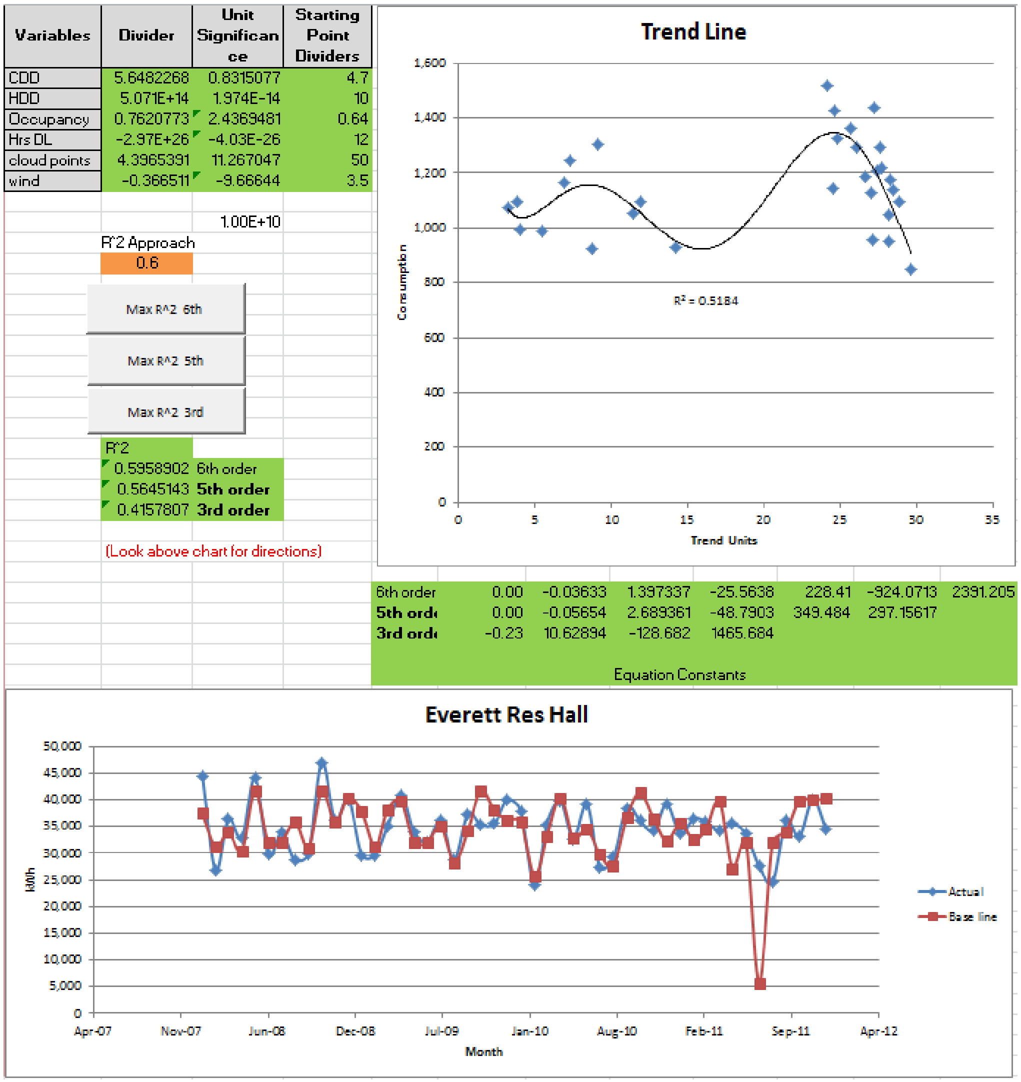
Though this study focused on residence hall electricity consumption and found a few areas of improvement, it is a study and approach that can be applied to any building type. The intent here, is that with the results found and the lessons learned, the next individual intent on finding a better baseline for electricity consumption will have an elevated platform as a starting point and be able to achieve even better results.

Appendix



Month	Days Per Month	Elec Units	Elec Units	Units/ Day	6th Order Normalize	12MS 6th Order Normalize	Hours of Day Light	Occupancy	CDD	Average CDD/day	HDD	Average HDD/day	Average Wind Speed	Sunny	Partly Cloudy	Cloudy	Cloud Points	Trend Units
Jan-08	31		44,536	1,437	37,444		9.3	1	0	0	770.6	24.858065	3.8	9	7	15	46.7742	27.243391
Feb-08	28.25		26,884	952	31,237		10.55	2	1.5	0.0530973	553.7	19.6	3.9	9	6	13.25	48.1416	28.118867
Mar-08	31		36,469	1,176	33,880		11.8	2	2.6	0.083871	407.2	13.135484	4.8	9	7	15	46.7742	28.219618
Apr-08	30		32,866	1,096	30,478		13.15	5	16.8	0.56	206.4	6.88	4.2	10	9	11	52.1667	28.81102
May-08	31		44,219	1,426	41,748		14.25	1	61.8	1.9935484	63.5	2.0483871	4.7	8	10	13	48.0645	24.554644
Jun-08	30		29,709	990	32,017		14.85	0	402.2	13.406667	0	0	2.8	8	12	10	51.3333	5.4170998
Jul-08	31		33,890	1,093	32,176		14.45	0	383.7	12.377419	0	0	2.2	7	12	12	48.2258	3.840563
Aug-08	31		28,629	924	35,841		13.6	1	426.7	13.764516	0	0	2.7	7	12	12	48.2258	8.7405329
Sep-08	30		29,853	995	31,066		12.3	2	264.4	8.8133333	4.4	0.1466667	3.6	10	9	11	52.1667	3.9848849
Oct-08	31		47,016	1,517	41,600		10.95	2	44	1.4193548	194	6.2580645	3.6	13	7	11	55.1613	24.150505
Nov-08	30		36,200	1,207	35,815		9.75	4	0	0	466.6	15.553333	3.9	11	7	12	52	27.375129
Dec-08	31	430,437	40,166	1,296	40,433	423,736	9.2	2	1.6	0.0516129	532.1	17.164516	3.2	10	7	14	48.871	26.086146
Jan-09	31	415,581	29,680	957	37,763	424,055	9.3	1	0	0	753.1	24.293548	3.1	9	7	15	46.7742	27.143446
Feb-09	28.25	418,290	29,593	1,048	31,102	423,920	10.55	2	0	0	536.6	18.99469	4.7	9	6	13.25	48.1416	28.156571
Mar-09	31	416,826	35,005	1,129	38,145	428,184	11.8	2	17.6	0.5677419	428.2	13.812903	3.9	9	7	15	46.7742	27.019344
Apr-09	30	424,925	40,965	1,366	39,801	437,507	13.15	5	47.9	1.5966667	149.2	4.9733333	4.1	10	9	11	52.1667	25.644978
May-09	31	414,659	33,953	1,095	32,046	427,805	14.25	1	203.8	6.5741935	31	1	3.1	8	10	13	48.0645	11.871057
Jun-09	30	417,105	32,155	1,072	32,109	427,897	14.85	0	377.8	12.593333	0	0	2.9	8	12	10	51.3333	3.1979761
Jul-09	31	419,311	36,096	1,164	35,041	430,762	14.45	0	418.7	13.506452	0	0	2.7	7	12	12	48.2258	6.9210493
Aug-09	31	419,476	28,794	929	28,270	423,191	13.6	1	488.7	15.764516	0	0	2.1	7	12	12	48.2258	14.197394
Sep-09	30	427,011	37,388	1,246	34,220	426,345	12.3	2	228.6	7.62	5.9	0.1966667	3.7	10	9	11	52.1667	7.2496646
Oct-09	31	415,402	35,407	1,142	41,743	426,488	10.95	2	37.1	1.1967742	152.2	4.9096774	3.4	13	7	11	55.1613	24.519073
Nov-09	30	414,746	35,544	1,185	38,055	428,728	9.75	4	0	0	334	11.133333	4.2	11	7	12	52	26.592582
Dec-09	31	414,649	40,069	1,293	36,297	424,592	9.2	2	0	0	769.1	24.809677	3.5	10	7	14	48.871	27.580519
Jan-10	31	422,818	37,849	1,221	35,838	422,667	9.3	1	0	0	851.9	27.480645	3.6	9	7	15	46.7742	27.707711
Feb-10	28.25	417,163	23,938	847	25,758	417,323	10.55	2	0	0	774.6	27.419469	4.8	9	6	13.25	48.1416	29.64815
Mar-10	31	417,506	35,348	1,140	33,038	412,217	11.8	2	0	0	404	13.032258	4.6	9	7	15	46.7742	28.430178
Apr-10	30	416,290	39,749	1,325	40,392	412,807	13.15	5	55.9	1.8633333	132.7	4.4233333	3.4	10	9	11	52.1667	24.820021
May-10	31	414,929	32,592	1,051	32,844	413,605	14.25	1	208.9	6.7387097	30.3	0.9774194	3.8	8	10	13	48.0645	11.418189
Jun-10	30	421,896	39,122	1,304	34,518	416,014	14.85	0	443.3	14.776667	0	0	2.8	8	12	10	51.3333	9.15505
Jul-10	31	413,258	27,458	886	29,848	410,821	14.45	0	489.1	15.777419	0	0	2.9	7	12	12	48.2258	13.117228
Aug-10	31	413,768	29,304	945	27,734	410,285	13.6	1	494.4	15.948387	0	0	2.8	7	12	12	48.2258	14.699074
Sep-10	30	414,892	38,512	1,284	36,750	412,816	12.3	2	332	11.066667	0	0	3.6	10	9	11	52.1667	2.1891459
Oct-10	31	415,574	36,089	1,164	41,390	412,463	10.95	2	43.7	1.4096774	140.9	4.5451613	3.2	13	7	11	55.1613	23.873645
Nov-10	30	414,260	34,230	1,141	36,531	410,940	9.75	4	0	0	428.1	14.27	3	11	7	12	52	27.147919
Dec-10	31	413,456	39,265	1,267	32,214	406,857	9.2	2	0	0	953.5	30.758065	4.4	10	7	14	48.871	28.633661
Jan-11	31	409,373	33,766	1,089	35,533	406,551	9.3	1	0	0	866.4	27.948387	3.6	9	7	15	46.7742	27.790523
Feb-11	28.25	421,934	36,499	1,292	32,496	413,289	10.55	2	2.5	0.0884956	511.3	18.099115	3.7	9	6	13.25	48.1416	27.756558
Mar-11	31	422,490	35,904	1,158	34,542	414,793	11.8	2	7.3	0.2354839	450.1	14.519355	4.6	9	7	15	46.7742	28.050962
Apr-11	30	416,905	34,164	1,139	39,862	414,263	13.15	5	69.5	2.3166667	148.4	4.9466667	4.3	10	9	11	52.1667	23.675787
May-11	31	419,836	35,523	1,146	27,213	408,633	14.25	1	163.3	5.2677419	44.8	1.4451613	2.6	8	10	13	48.0645	15.514434
Jun-11	30	414,303	33,589	1,120	32,017	406,131	14.85	0	402.2	13.406667	0	0	2.9	8	12	10	51.3333	5.4170998
Jul-11	31	414,398	27,553	889	5,711	381,994	14.45	0	547.8	17.670968	0	0	2.8	7	12	12	48.2258	18.283643
Aug-11	31	409,751	24,657	795	31,946	386,206	13.6	1	462.9	14.932258	0	0	3.7	7	12	12	48.2258	11.926636
Sep-11	30	407,327	36,088	1,203	34,051	383,507	12.3	2	231.4	7.7133333	17.3	0.5766667	2.4	10	9	11	52.1667	7.0622889
Oct-11	31	404,294	33,056	1,066	39,760	381,877	10.95	2	20.4	0.6580645	226	7.2903226	3.7	13	7	11	55.1613	26.41039
Nov-11	30	410,220	40,156	1,339	39,932	385,277	9.75	4	12.9	0.43	353.3	11.776667	3.2	11	7	12	52	25.533257
Dec-11	31	405,594	34,639	1,117	40,241	393,304	9.2	2	0.1	0.0032258	526.4	16.980645	2.8	10	7	14	48.871	26.185613

Figure 7: Monthly Values and Variables



Variables	Divider	Unit Significance	Starting Point Dividers
CDD	5.6482268	0.8315077	4.7
HDD	5.071E+14	1.974E-14	10
Occupancy	0.7620773	2.4369481	0.64
Hrs DL	-2.97E+26	-4.03E-26	12
cloud points	4.3965391	11.267047	50
wind	-0.366511	-9.66644	3.5

1.00E+10

R^2 Approach

0.6

Max R^2 6th

Max R^2 5th

Max R^2 3rd

R^2	
0.5958902	6th order
0.5645143	**5th order**
0.4157807	**3rd order**

(Look above chart for directions)

6th order	0.00	-0.03633	1.397337	-25.5638	228.41	-924.0713	2391.205
5th ord	0.00	-0.05654	2.689361	-48.7903	349.484	297.15617	
3rd ord	-0.23	10.62894	-128.682	1465.684			

Equation Constants

Figure 8: Baseline generator