Title
A Statistical Process Control Approach to Energy Efficiency in Buildings

Author
Trager, Jason Scott

Publication Date
2014

A Statistical Process Control Approach to Energy Efficiency in Buildings

by

Jason Scott Trager

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Mechanical Engineering

and the Designated Emphasis

in

Energy Science and Technology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Paul K. Wright, Chair
Dave Auslander
Daniel Kammen

Fall 2014

A Statistical Process Control Approach to Energy Efficiency in Buildings

Abstract

A Statistical Process Control Approach to Energy Efficiency in Buildings

by

Jason Scott Trager

Doctor of Philosophy in Engineering - Mechanical Engineering

and the Designated Emphasis

in

Energy Science and Technology

University of California, Berkeley

Paul K. Wright, Chair

Statistical Process control is a technique often used in manufacturing in order to determine whether a process is changing based upon most recent measurements. In this thesis, I will cover how this is applicable to energy efficiency in buildings, and construct the framework that allows for this viewpoint to be effective in understanding how buildings consume energy. I will discuss how to effectively baseline a buildings energy performance, predict that performance in basic and advanced ways, and then use that prediction to determine when building parameters are out of line. I outline a methodology for determining when building operation parameters have changed, attempt to identify them, and suggest potential problems, given a top-down monitoring methodology combined with a bottom-up fault diagnostics approach. The approach explored in this talk will combine Shewhart and other control charts with modern monitoring techniques in order to explore the efficacy of this approach. I conclude by developing new modifications on the idea of control charts and by using those modifications to produce an analysis of the campus energy use at the University of California, Berkeley.

Keywords: Energy Efficiency, Statistical Process Control

For Art Rosenfeld

I hope that I can make even a fraction of the impact that you have.

For John-Fred Crane

The best frame of reference is the one that makes your life easiest.

For Mr. Napoli

I will never forget the importance of Confidence.

For Mrs. Watson

For believing in me before I even had a chance to believe in myself.

Contents

Contents **ii**

List of Figures **iv**

List of Tables **ix**

1 1. An exploration of building power use data **1**
1.1 An Overview of this Thesis . . . 1
1.2 The magnitude of energy use in buildings . . . 3
1.3 Metrics of energy efficiency in buildings . . . 4
1.4 Converting standard building metrics into control charts . . . 8
1.5 Energy Efficiency in Buildings: Policy and implementation . . . 11
1.6 Marketplace evaluations, solutions, and certifications for energy efficiency . . . 15
1.7 SPC as a framework in buildings . . . 16

2 Time series analysis and prediction in the context of buildings **17**
2.1 Introduction . . . 17
2.2 Examples of time series data in commercial buildings . . . 17
2.3 Exploratory data analysis on building full and sub-meters . . . 18
2.4 Building Load Profiles . . . 20
2.5 Better prediction through time series analysis . . . 22
2.6 Efficacy of several prediction models on a single building . . . 26
2.7 Implications and conclusions . . . 33

3 Better prediction models through automation and combination **34**
3.1 Introduction . . . 34
3.2 Evaluating models across multiple buildings and timescales . . . 35
3.3 Adaptive Combination models . . . 40

3.4 Experimental Setup of Combination Algorithm testing 42
3.5 Results of combination model production 43
3.6 Conclusions and Implications 47

4 Anomaly Detection **48**
4.1 Introduction 48
4.2 Example: Prediction for anomaly detection 49
4.3 Control charts: an application to anomaly detection 50
4.4 Example Root Cause Analysis 56
4.5 Implications and Conclusions 57

5 An exploration of additional parameters and metrics in SPC based anomaly detection **58**
5.1 Introduction 58
5.2 Transforming control charts: Looking for scheduling faults 59
5.3 Variation of control - Daily differences 60
5.4 Load Duration Curves: A snapshot of building condition 62
5.5 Diagnostic Results 68
5.6 Conclusions and implications 71

6 Applying data mining techniques to SPC based anomaly detection in buildings **72**
6.1 Introduction 72
6.2 data set 73
6.3 LDC Analysis 73
6.4 Mining the Results 81
6.5 Diagnostic results 87
6.6 Conclusions and Implications 91

7 Implications of Statistical Process control in Buildings **92**
7.1 Research Overview 92
7.2 Research Challenges 93
7.3 Research Vision 93

Bibliography **94**

List of Figures

1.1 Load Duration Curve For one week of Leconte Hall's power use 2
1.2 Energy use breakdown of US Economy, 2012 4
1.3 Energy use breakdown within buildings 5
1.4 A weekend and a weekday of power use, annotated with metrics 7
1.5 Electrical EUI clustered by climatic zone and building type 7
1.6 Energy Use Intensity averaged every week over the metered data timeframe for a building in LeConte Hall. As we can see, seasonality plays an important role in EUI when it is not simply captured as a snapshot. . . . 8
1.7 Energy Use Intensity residual in LeConte hall using the LBNL prediction method. Outliers whole weeks where there was difficulty predicting EUI. 9
1.8 Weekly Load Factor variation in LeConte Hall. 10
1.9 Weekly Load Factor residual in LeConte Hall using LBNL model. 10
1.10 The Rosenfeld Curve - showing energy consumption in California bending downward 12
1.11 A Demand Response event in Sutardja Dai Hall across 14 different sub meters 13
1.12 Load duration curve for California, 2005-2006 13
1.13 Building Power Draw and outside air temperature of a building that has undergone a significant energy efficiency retrofit. 14

2.1 Electricity and Steam Use in two buildings 18
2.2 Autocorrelation of a building's power use with max lag 168 hours. Original power trace of the building can be found in fig. 1.13 19
2.3 Partial Autocorrelation of a building's power use with max lag 168 hours. Original power trace of the building can be found in fig. 1.13 20
2.4 Demand response baseline by comparing similar day. As we can see, sometimes performing a demand response event involves shifting load earlier in the day. 21

2.5 Rolling window cross validation allows us to more effectively evaluate prediction methods on time series data. 26
2.6 Seasonal Trend Loess (STL) decomposition is one way to de-decompose a time series for better prediction. 31
2.7 Hourly Power Draw of the UC Berkeley Art Museum. 31
2.8 Results of BLP generation using Cross Validated time series methodology for a single building: In this case, the Art Museum on the UC Berkeley Campus. 32

3.1 Flow chart diagram of multiple models leading to a superior baseline load profile. 34
3.2 Box-plot of errors in twenty building's load profiles 36
3.3 Box-plot of by errors by forecast horizon 37
3.4 Distribution of errors across each of the forecast horizons 38
3.5 Comparisons of model performance against other models in first 24 hours of forecasting 39
3.6 Comparisons of model performance against other models in 168 hours of forecasting 39
3.7 Flow chart diagram showing how model combinations can improve predictions. 40
3.8 Box and whisker diagram of multiple performance measures for combination models after removing models that take excessive time to run. . . . 43
3.9 Box and whisker diagram of relative mean absolute error for different horizons of combination models after removing models that take excessive time to run. 44
3.10 Percentile plot of errors of combination models over all horizons. The dotted lines represent the 5th and 95th percentile of error, the solid line represents the median, and the gray line represented the error of the Seasonal Naive model. 45
3.11 Table of performance of methods against other methods including combinations over the first 24 hours of prediction. In each row, the percentage indicates how many times that method performed better than the comparable in the column. 46
3.12 Table of performance of methods against other methods including combinations over all 168 hours of prediction. In each row, the percentage indicates how many times that method performed better than the comparable in the column. 46

4.1 MAPE of Seasonal Trend Loess method error function for two buildings using regression on temperature for prediction of residual. 49
4.2 Moving average Control Chart for Building 1 52
4.3 Rolling Average Control Chart for Building 1 52
4.4 Combined measure control chart projected onto actual building energy use 53
4.5 Moving average Control Chart for Building 2 54
4.6 Rolling Average Control Chart for Building 2 54
4.7 Combined measure control chart projected onto actual building energy use 55
4.8 Temperature vs power use in building 2 55
4.9 Overnight Cycling picked up by the algorithm - This occurred during the commissioning phase of Building 1 56
4.10 The algorithm picked up this point in the power usage of Building 2 that showed the building not shutting down correctly at night. 57

5.1 A Daily Mean Control Chart demonstrating poor scheduling on an example building. From this chart we can see that, on an average work-day, the building varies more in off-hours than in on-hours 59
5.2 A Daily Mean Control Chart demonstrating poor scheduling on an example building. From this chart we can see that, on an average work-day, the building varies more in off-hours than in on-hours 60
5.3 A weekly demand control chart for a building in Chicago, with day 1 being Monday and day 7 being Sunday. 61
5.4 A weekly demand control chart for a building in Chicago, with day 1 being Monday and day 7 being Sunday. 61
5.5 A weekly residual control chart for a building in Chicago, with day 1 being Monday and day 7 being Sunday. 62
5.6 Load Duration Curve For LeConte Hall on UC Berkeley Campus 63
5.7 Load Duration Curve For one week in Leconte's power use 64
5.8 Six Parameter control chart based on each of the parameters outlined for a load duration curve. The purple lines represent the fit of the LDC over the entire timeframe, while the green lines represent the average fit over all the weeks computed. 66
5.9 Six Parameter control chart based on the residuals parameters fit to the actual load duration curve minus the parameters fit to the expected. . . 67
5.10 Diagnostic chart for LeConte Hall, with week 40 presented in magenta and week 43 presented in blue. 69
5.11 Power use in LeConte hall, with previously flagged weeks highlighted in colors matching the above chart. 70

5.12 Diagnostic Weekly Residual Control Chart, Leconte hall showing unpredictable behavior in off hours. 70

6.1 Schedule Depth for all meters in the dataset, UC Berkeley campus. Each building meter is represented by a different color. It should be noted that there are two distinct peaks in the scheduling parameters for buildings. . 73
6.2 Mean schedule depth for meters in our data set. This group was divided into two cluster that represent two different operational modes, corresponding to the previous two peaks in fig. 6.1. 74
6.3 Box Plot showing schedule depth for each UC Berkeley building in the data set 75
6.4 Box Plot showing residual of schedule depth for each UC Berkeley building in the data set 76
6.5 Box Plot showing the Off-Schedule Slope for each UC Berkeley building in the data set 77
6.6 Box Plot showing the Off-Schedule Slope residual for each UC Berkeley building in the data set 77
6.7 Box Plot showing the Peak Height for each UC Berkeley building in the data set 78
6.8 Box Plot showing the Peak Height residual for each UC Berkeley building in the data set 79
6.9 Box Plot showing the Peak Depth for each UC Berkeley building in the data set 80
6.10 Box Plot showing the Peak Depth residual for each UC Berkeley building in the data set 80
6.11 Daily Residual Control Chart for Doe Library's 480V feed. As we can see, the building is hard to predict at night and at off hours, with a dip in predictability in the mid-morning. This is likely caused by a control problem. 82
6.12 Daily Residual Control Chart for LSA's Main Substation B South feed. The building is generally hard to predict at any hour of the day, with significant unpredictability in off hours. This building requires a high level of scheduling evaluation 82
6.13 Daily Mean Control chart for California Hall, a well scheduled building on campus. 83
6.14 Daily Residual Control chart for California Hall, a well scheduled building on campus. 83
6.15 Daily Mean Control chart for Haas Pavillion. 85
6.16 Daily Residual Control chart for Haas Pavillion. 85

6.17 Daily Mean Control chart for Wheeler Hall. 86
6.18 Daily Residual Control chart for Wheeler Hall. 86
6.19 Power Use Trace of LSA Substation B. Operation changes and unpredictable features are shown clearly. 87
6.20 A density distribution of residuals for LSA Main Substation B. The non-normal distribution of errors indicates that the building is not easily predictable. 88
6.21 Power Use Trace of Wheeler Hall. It is easy to see some operational changes in this building which would are indicative of inefficient behavior. 88
6.22 A density distribution of residuals for Wheeler Hall. This distribution is closer to normal than the one presented above, indicating greater predictability in operations. 89
6.23 Power Use Trace of California Hall. This building displays regular scheduling an operations. 89
6.24 A density distribution of residuals for California Hall. This distribution is very close to normal, and points of interest in the building operations could be extracted by normal SPC methods. 90

List of Tables

3.1 Computation times on a single building 42

4.1 Two buildings were used in this study: The buildings are anonymized for privacy purposes 50

5.1 Table displaying fitted coefficients for the example LDC given above . . . 65
5.2 BSPC Parameters for Week 40 for Leconte 68
5.3 Residual BSPC Parameters for Week 40 for Leconte 68
5.4 BSPC Parameters for Week 43 for Leconte 68
5.5 Residual BSPC Parameters for Week 43 for Leconte 69

Acknowledgments

Many people contributed to this dissertation in a myriad of ways. My friends and family have been incredibly supportive, and plenty of people have put up with the side-effects of graduate student stress on my behalf. I am grateful to have such a community. In terms of those that have contributed specifically, I'd like to acknowledge Jeremy Coyle for his knowledge of statistics and Devan Johnson for his knowledge of buildings. Without these two individuals, this dissertation would never have been completed. I would also like to acknowledge the support and wisdom of one who has far more experience than myself: Carl Blumstein has been a supportive mentor throughout this whole experience, encouraging me when I have been down and providing inspiration by way of his excellent beard. Finally, I'd like to acknowledge my communities: The Berkeley Student Cooperative and the East Bay Scorchers - living cooperatively and riding bikes will forever be lifestyles I will rejoice in.

Chapter 1

1. An exploration of building power use data

1.1 An Overview of this Thesis

Statistical Process Control (SPC) is a heuristic method used to gather and visualize data with the end goal of reducing variation in a system and thus increasing the level of its performance. This methodology has been applied to manufacturing processes [8], healthcare outcome improvement [56], and in evaluating building operations from a high level[25, 27, 28]. This thesis extends the state of the art of the last application - building energy analysis - to include methodologies for evaluating and benchmarking buildings at small time scales in order to facilitate systems diagnosis and anomaly correction. Prior art shows how mathematically transforming data enables anomaly detection and feature extraction in complex SPC [47] This work will conclude by showing how transforming data from whole-building meters into physically meaningful Load Duration Curves (LDCs) based upon physical variables can be used to rapidly filter building behaviors on the UC Berkeley campus for examination. Figure 1.1 shows an example LDC, which can be modeled on six variables, which will be explained further in Chapter 5:

- - a - Off-schedule Slope - How the building varies in off-hours.
- - b - Peak Height - How much normalized power the building used.
- - c - Peak Depth - How long the building used a high amount of power.
- - d - Schedule Shift - How much difference is there between scheduled and unscheduled hours.

- - f - Shift Slope - How fast the building switches to unscheduled more.
- - g - Schedule Depth - What fraction of the week that the building is scheduled.

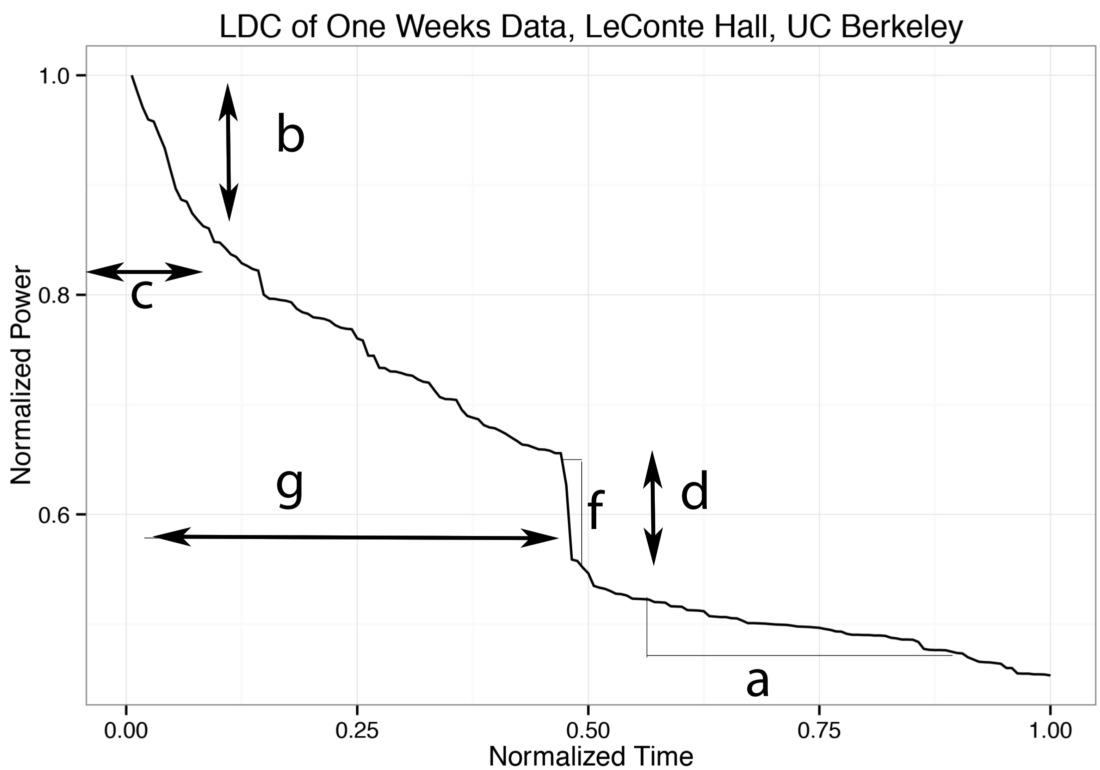


Figure 1.1: Load Duration Curve For one week of Leconte Hall's power use

This framework is built up in several steps: the remainder of Chapter 1 describes energy use in buildings in a societal framework, Chapter 2 provides an intro to properly predicting short term power use in buildings, Chapter 3 extends the work on short term prediction by automating the application and producing optimal forecast model selection, Chapter 4 applies the work of prediction to anomaly detection in sample buildings, Chapter 5 introduces math and visualization tools for engineers in this space, and Chapter 6 uses the prediction and the tools introduced in previous chapters to show how to analyze a campus for rapid energy consumption evaluation. Chapter 7 concludes and states further implications.

1.2 The magnitude of energy use in buildings

World primary energy use in 2011 totaled to approximately 500 quadrillion BTUs (Quads) [24]. Of that use, roughly 20% occured in the United states, with two recent estimates stating annual consumption of 97.1 Quads (2011) and 95.1 Quads (2012) [10, 24].[Figure 1.2 / EIA 2011] shows the economy used a total of 97.3 quads in 2011 to produce 41.7 quads of energy services equating to an overall economic energy efficiency (OEEE) of 42.857%. As a sub-category of this total energy use, consumption in buildings is broken down between consumption of electricity produced on the grid, and direct use of solar, geothermal, natural gas and petroleum. From this diagram, we can approximate that electricity production on the grid has an overall efficiency of 32.1% and can infer that a significant savings in energy use in any major sector will result in a direct reduction of consumption along with a larger reduction in primary fuel use. While this reduction will not correspond exactly to the efficiency of the grid, if a large enough reduction is made, an approximately similar fuel savings will occur, and the carbon footprint of this can be computed by whatever fuel mix is powering these applications.

To address the magnitude of such possible fuel savings, and the associated level of impact on the carbon life cycle of our building stock, an in depth analysis of how buildings use power must be made. This analysis starts with a breakdown of end use of consumption within these buildings. Buildings use 41% of end use energy in the US and commercial buildings consume 19% of the end use total [10]. Commercial buildings consume energy through a variety of systems, with the vast majority being used for space conditioning and lighting. Figure 1.3[10] shows a breakdown of both the overall energy consumption in the US and that of commercial and residential buildings. Lighting and space/water conditioning comprise 2/3 of all energy use in these spaces, and as such will be reviewed in more detail later in this document. Given that building energy usage is projected to grow in the future [41] and that this energy use is a major contributor to global climate change through greenhouse gas emissions [52], it is prudent to devise strategies that will allow consumption to be nipped in the bud.

Further breakdown of energy use in commercial buildings is of interest to this discussion, and so we will begin there. The most commonly referred to studies of commercial floorspace are the Commercial Buildings Energy Consumption Survey (CBECS) from 2003 [11] and 2007 [12]. In 2003, there were approximately 4.8 million commercial buildings in the US with an estimated floor space, that accounted for approximately $107.9 billion in energy expenditure. This number can be evaluated in the light of how much energy is lost every year as the result of faulty systems, leaks, and lack of advanced control systems, which is estimated at $6 billion a year

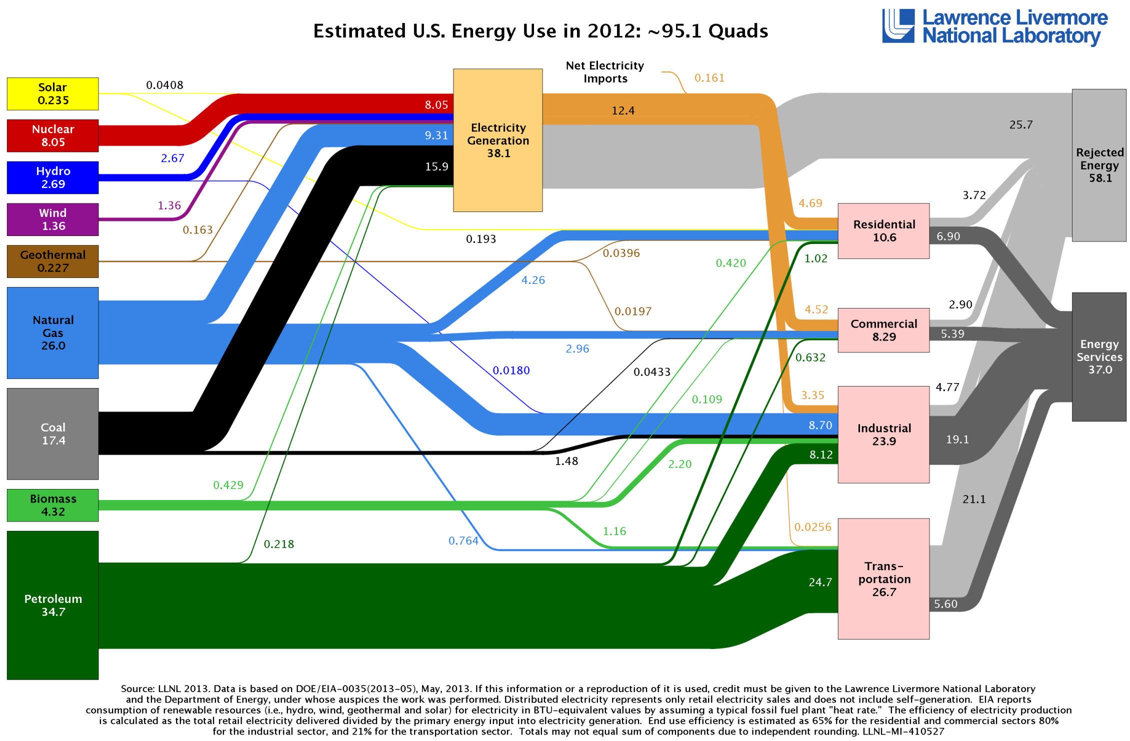


Figure 1.2: Energy use breakdown of US Economy, 2012

for waste and between \$13.6 and \$38.5 billion for lack of advanced controls [46].

1.3 Metrics of energy efficiency in buildings

Comparison of energy and power use in buildings is an important way to understand how a particular built environment may be made more efficient, and is also pertinent in understanding how much energy does a building and its constituent parts use. These metrics typically fall into either a magnitude of use (MOU) or a time of use (TOU) framework. From the MOU perspective, a building will exert some sort of average draw characteristics on the grid as it is designed. From the TOU perspective, not all appliances or devices within a building are on at any given time. It is from the combination of these perspectives that the notion of building load factor is introduced. Load factor (LF) is defined as:

$$LF = \frac{Average\, load}{Maximum\, load\, within\, time\, frame} \tag{1.1}$$

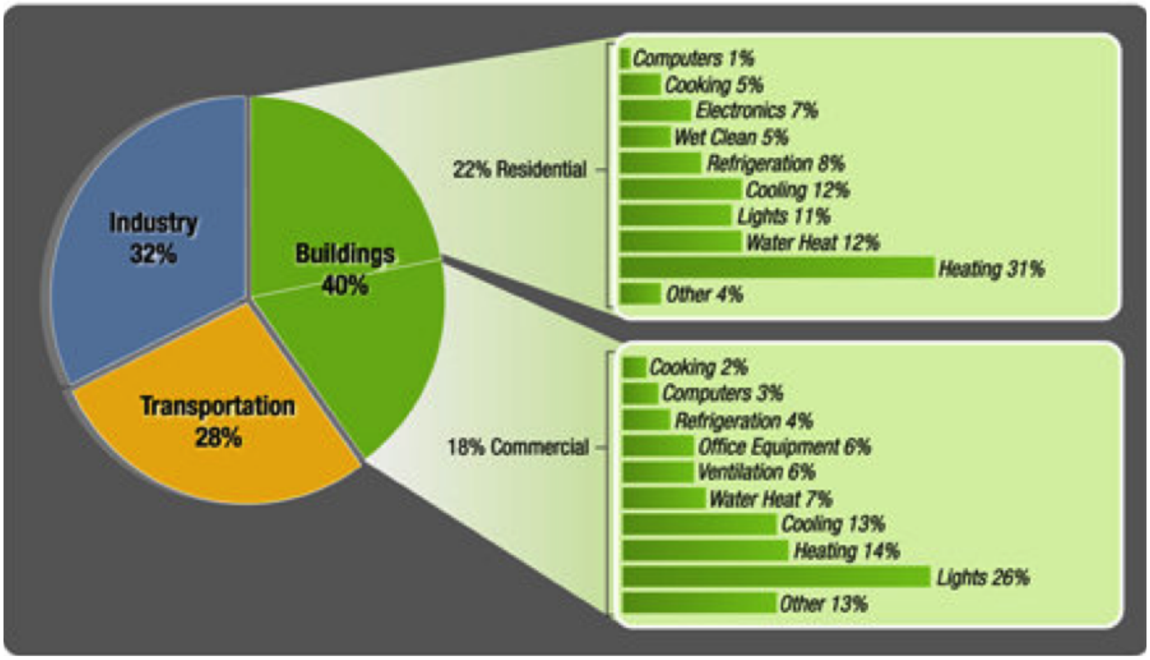


Figure 1.3: Energy use breakdown within buildings

This calculation, from a 1915 paper [59] captures one aspect of time varying power use - that buildings do not draw the same amount of power at all times. Similarly, whole circuits and individual devices within a building draw varying amounts of power depending on circumstances within their environment. In addition, There are other measures of building power-use data that transform time series elements into easy to parse numbers for practitioners of energy efficiency in buildings. While these are typically generated as one number for the entire timeframe in which the building is being monitored, for the remainder of this, we will also consider these metrics as they evolve in a building over time. First, we take metrics proposed by LBNL and combine them with several other metrics of energy use in buildings. Price et al proposed the following series of metrics in their 2011 paper [44]

- Near-Peak Load (kW) - 96.5 percentile of daily load
- Near-Base Load (kW)- 3.5 percentile of daily load
- High-Load Duration (Hr) - Time-period of greater than median load.
- Rise Time (Hr) - Time to go from Near Base to start of High-Load Duration
- Fall Time (Hr) - Time to go from High Load Duration to Near-Base Load.

In addition to these metrics, it is possible to calculate the energy use intensity (EUI) of a building over time, where EUI is defined as:

$$EUI = \frac{Total\,Energy\,Use\,In\,Building}{Floor\,Space\,In\,Building} \frac{8760}{Number\,of\,hours\,in\,measurement\,period} \tag{1.2}$$

This metric benchmarks the level to which a building is operating relative to equivalent buildings in similar climactic zones. Information on what is typical EUI can be found in [11, 12], and two relevant examples are given below. It is important to note that EUI is typically a combined measure of both electricity consumption and natural gas consumption within a building, and looking at each individually does not yield the whole picture. To help conceptualize what these numbers really mean, fig. 1.4 is a graph of a building's power use from a Sunday to a Monday, annotated with each of the descriptors above. For reference, the building pictured below has an EUI of 13.81$kWh/sqft$ in electrical draw only. It is a 100,000 square foot office space win sunnyvale and is below the average EUI for the west coast zone for offices, specified infig. 1.5 [11].

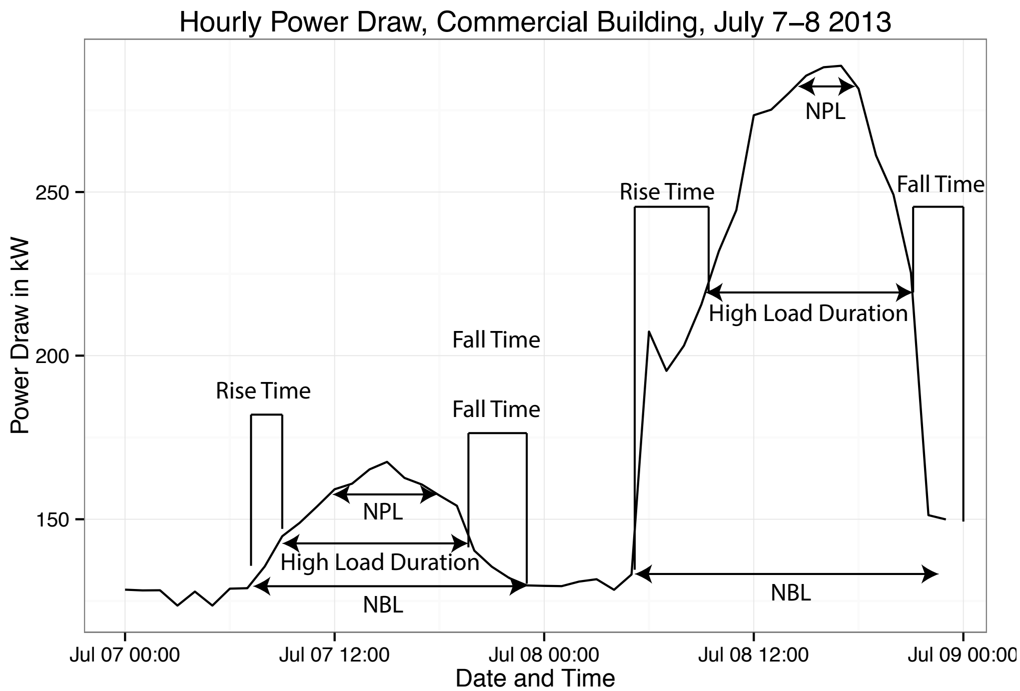
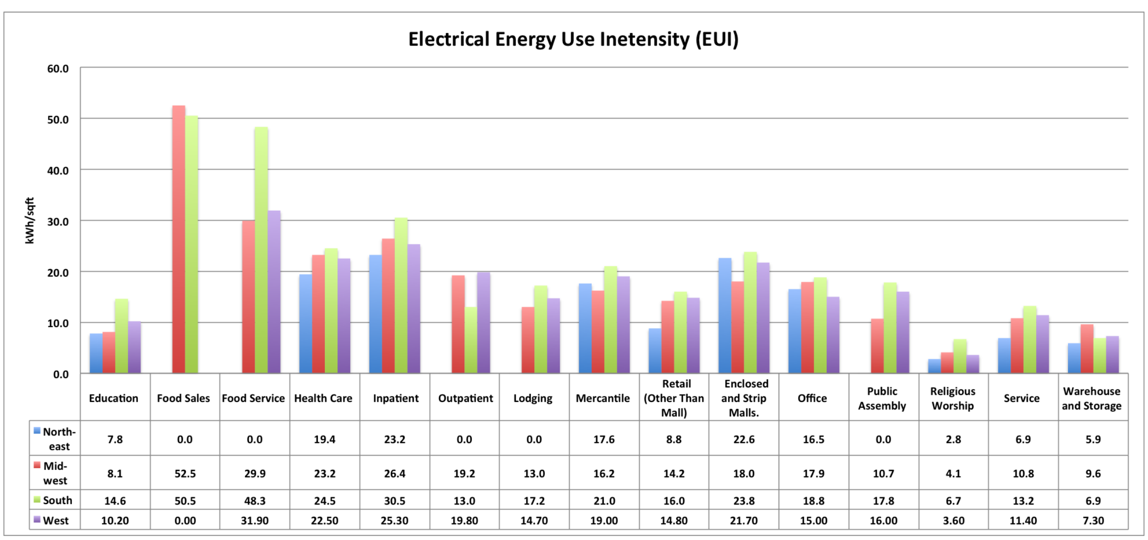


Figure 1.4: A weekend and a weekday of power use, annotated with metrics



	Education	Food Sales	Food Service	Health Care	Inpatient	Outpatient	Lodging	Mercantile	Retail (Other Than Mall)	Enclosed and Strip Malls.	Office	Public Assembly	Religious Worship	Service	Warehouse and Storage
North-east	7.8	0.0	0.0	19.4	23.2	0.0	0.0	17.6	8.8	22.6	16.5	0.0	2.8	6.9	5.9
Mid-west	8.1	52.5	29.9	23.2	26.4	19.2	13.0	16.2	14.2	18.0	17.9	10.7	4.1	10.8	9.6
South	14.6	50.5	48.3	24.5	30.5	13.0	17.2	21.0	16.0	23.8	18.8	17.8	6.7	13.2	6.9
West	10.20	0.00	31.90	22.50	25.30	19.80	14.70	19.00	14.80	21.70	15.00	16.00	3.60	11.40	7.30

Figure 1.5: Electrical EUI clustered by climatic zone and building type

1.4 Converting standard building metrics into control charts

Buildings are analyzed as snapshots - the overall building characteristics are taken into some sort of calculation in software, and then computations are done in order to determine if the building meets a certain benchmark or efficiency grade. This process typically devolves into benchmarks that happen once a year [42]. It is not typical to analyze buildings against themselves and to have that analysis lead to efficiency measures being put into place in the building. In order to illustrate the possibility of this, presented here are benchmarking methods that are framed as control charts, a method of quality control first outlined by Walter Shewart.

EUI

EUI provides good information about how the building is behaving at a particular time and can be useful to evaluate how the energy use intensity fluctuates. This is valuable, but must be weather normalized in order for it to be meaningful. in fig. 1.6, we can see the energy use intensity plotted as a one week moving average. In order

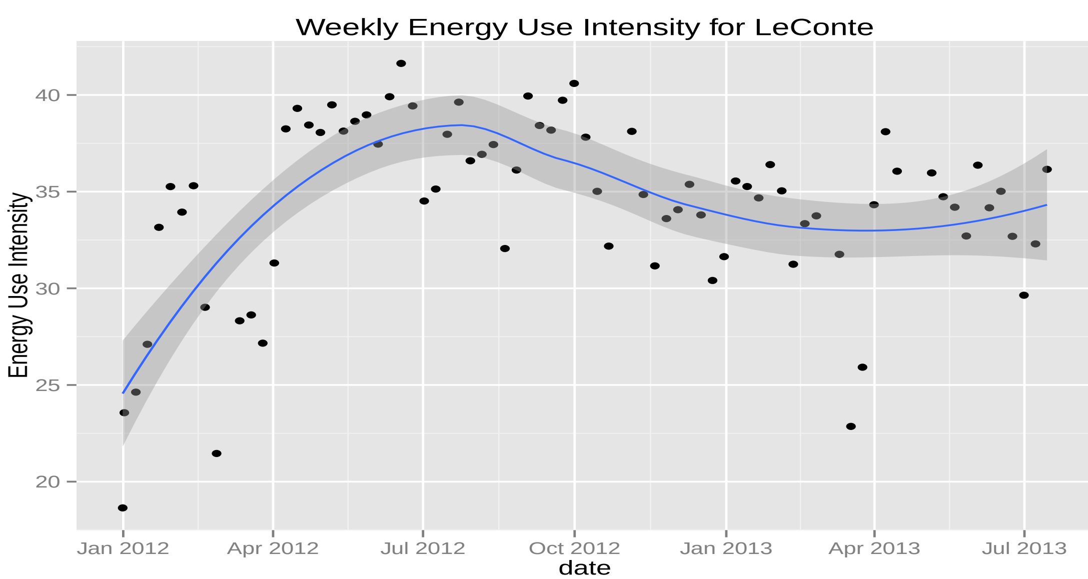


Figure 1.6: Energy Use Intensity averaged every week over the metered data time-frame for a building in LeConte Hall. As we can see, seasonality plays an important role in EUI when it is not simply captured as a snapshot.

to convert Energy Use Intensity into a useful, chartable metric, we must normalize for weather and other variability in the building. We start by constructing a control chart of the predicted EUI using the LBNL method which will be detailed later. Then a 95% confidence interval is built in for the data so that it can be plotted as a control chart.

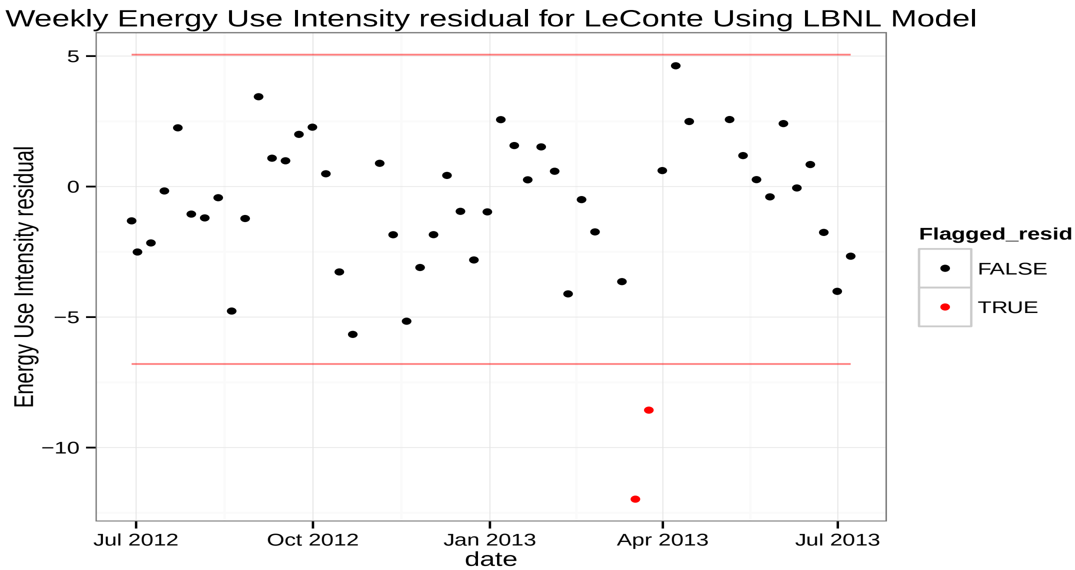


Figure 1.7: Energy Use Intensity residual in LeConte hall using the LBNL prediction method. Outliers whole weeks where there was difficulty predicting EUI.

Load Factor

The same level of data mining can be done for the load factor, which can show us how predictable the peak load over the average load for a building is in a given period of time.

In fig. 1.8, we can see the Load Factor of LeConte hall over the course of a year and a half, as was done with the EUI of the building in fig. 1.6. Flagging points when the load factor is outside of the norm for the building should indicate when the building's scheduling has changed significantly or when the weather has shifted substantially in a short time frame. Once again, we can make control charts analogous to common cause and special cause by using models that perform differently on varying time scales. Because of the similarity of the metrics, fig. 1.9 appear similar to fig. 1.7. We notice however, that different points are flagged. This will become relevant upon the

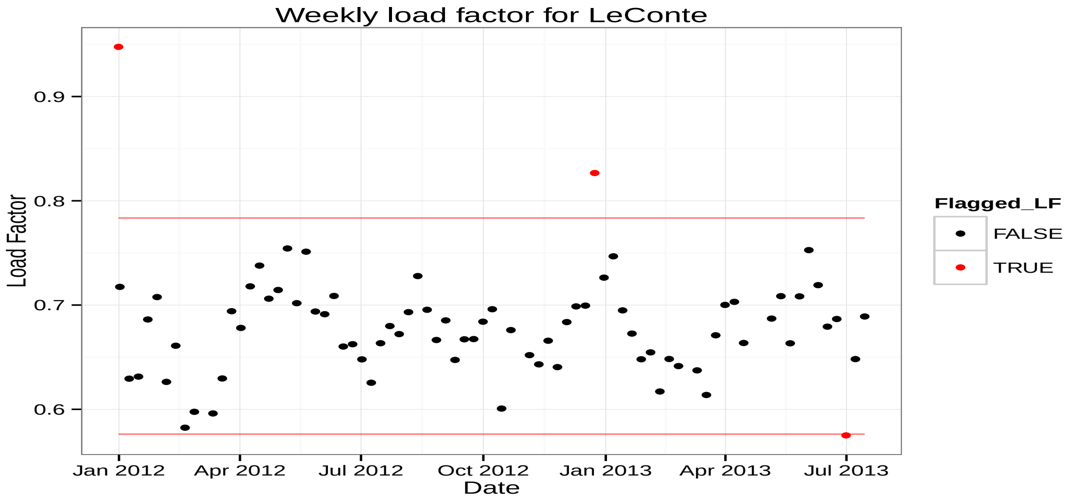


Figure 1.8: Weekly Load Factor variation in LeConte Hall.

initiation of data mining this information in order to draw conclusions about what malfunctions the building is experiencing.

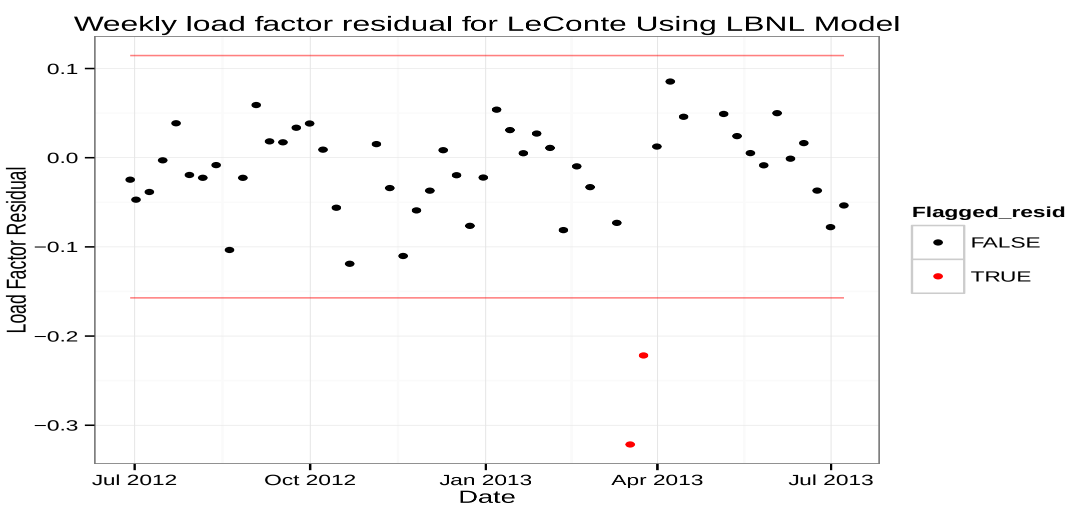


Figure 1.9: Weekly Load Factor residual in LeConte Hall using LBNL model.

1.5 Energy Efficiency in Buildings: Policy and implementation

Energy efficiency policies in buildings are a highly critical area of research at the moment, with both anecdotal past policies [9] and hypothetical future ones being looked at [50]. One of the important aspects of implementing an energy efficiency policy on one or many buildings is to target metrics toward the desired goal and type of consumption that is to be reduced. While the goal of reducing energy use is noble (We could get 1/8 of the way to solving our carbon problem by upgrading all building energy efficiency to best practices [39]), it is important to evaluate how energy efficiency scales if we are to understand how to deploy it in a widespread manner effectively. Policy for energy efficient must come from some sort of regulatory body, which typically is divided into three levels - Federal, State, and Local. Each of these different policy levels has advantages and disadvantages, but it can be generally summarized that the larger the body regulating the energy efficiency policy, the more uniform it is but also the less tailored it is to specific environmental and market forces in the areas which it must address. While Federal regulations may be overreaching and over regulate markets, it is a risk that local policies may be limited by funding and geographical influence. States however, have jurisdiction over utilities and therefore the fees and regulatory mandates that these companies must comply with. This is beneficial in implementing a successful energy efficiency policy [9]. As an example of how a state can push the edge of policy in a successful direction, the Rosenfeld curve - showing the difference between California and US per capita energy consumption, illustrates this point in fig. 1.10, sourced from [45].

Energy use reduction in buildings takes the form of either short term reductions to alleviate strain on the power grid or long term savings that will produce sustained energy savings over the lifetime of a the building. We can evaluate both from the context of a load duration curve (LDC), which is a frequentist method for documenting how often the hourly load on a grid is at a certain level. If we look at **??**, we can see that in California in September 2005-2006, the hourly load on the grid does not exceed 27,500MW more than approximately 50% of the time - we refer to this as the base load, mostly powered by nuclear and coal fired power plants. The next level of load, from 27,500 MW to 40,000 MW occurs for approximately 46.8 percent of the time, and we refer to this as intermediate load - mostly powered by natural gas power plants. As we move up the curve, we enter the "peak demand" period - in which the grid exceeds 40,000 MW only 3.2% of the time, and is typically powered by gas fired power plants which are held in reserve on days when this is anticipated. Finally, we enter the "Ultra-peak" periods of power demand, in which

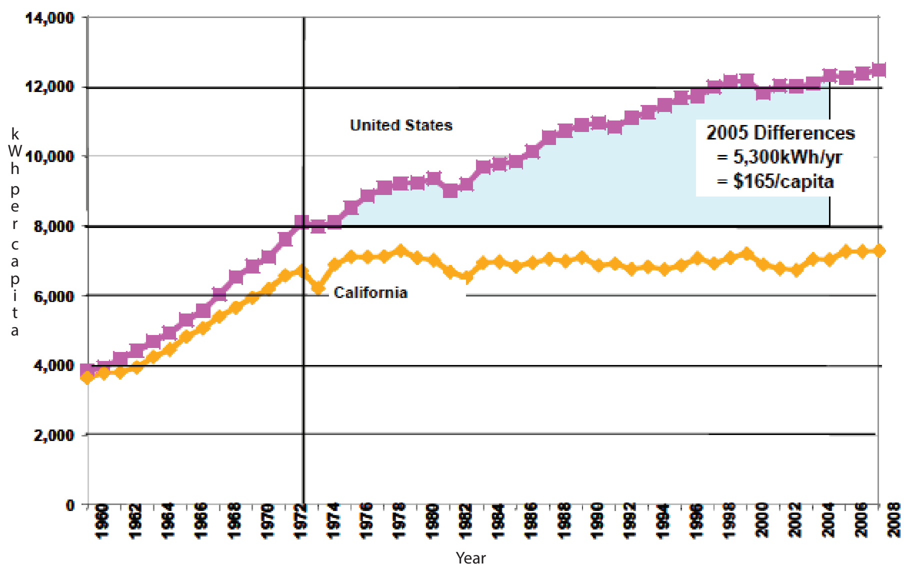


Figure 1.10: The Rosenfeld Curve - showing energy consumption in California bending downward

the grid exceeds 45,000 MW only 0.65% of the hours in the year[18]. If we want to think of energy efficiency vs demand response (AKA: demand side load management (DSLM)), then we can do so intuitively by thinking that demand response shaves the peak off of the LDC, while energy efficiency shifts the entire curve downward. We can further illustrate this distinction by showing specific examples in buildings - fig. 1.11 shows an example of a Demand Response (DR) event - a short term power saving event designed to alleviate strain on the electrical grid, typically on a high temperature day. If we are to look at an energy efficiency measure, we can observe the power trace of the building shown in fig. 1.13, which features a large efficiency upgrade toward the end of the power trace. Note that this figure shows building power color coded by outside air temperature such that it becomes easier to see the relationship between the outside air temp and the consumption in the building.

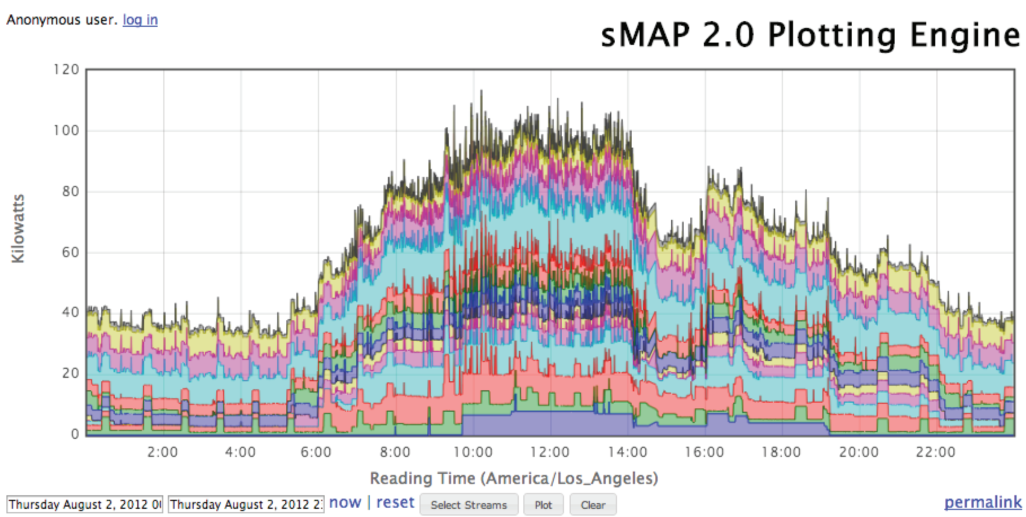


Figure 1.11: A Demand Response event in Sutardja Dai Hall across 14 different sub meters

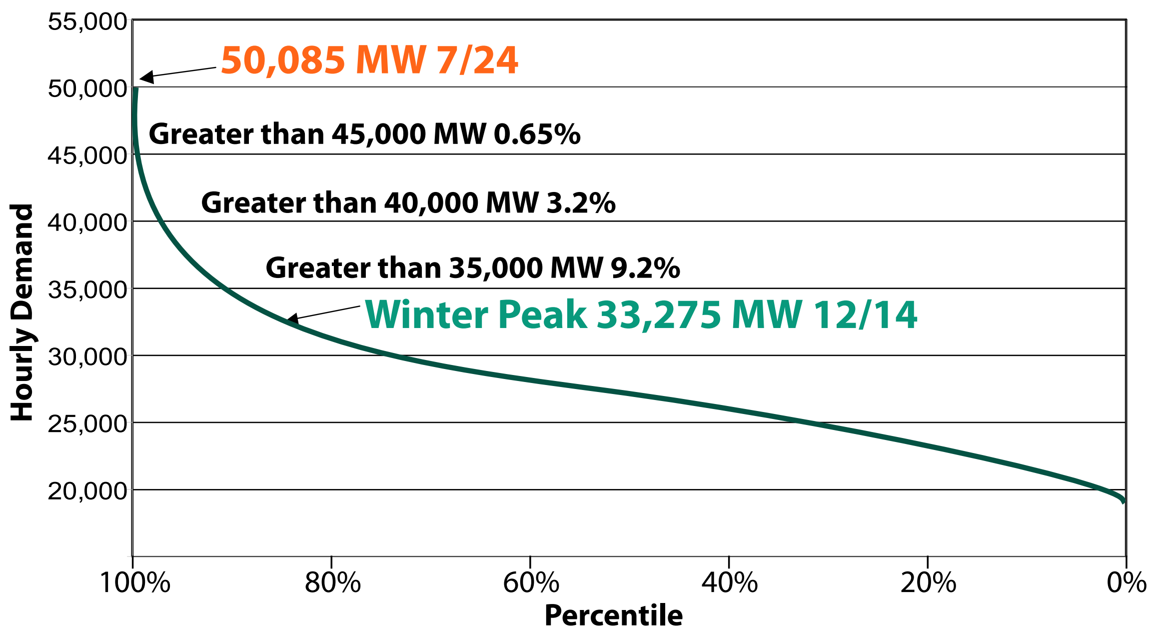


Figure 1.12: Load duration curve for California, 2005-2006

Finally, we would be remiss not include a concrete example and to mention the application of "usability" and behavioral mechanisms as they relate to policies of implementing energy efficiency. As an example, we will look at the deployment of programmable thermostats, which were granted the government policy designation of "Energy star" (Energy star is an efficiency labeling program run by the US federal government. More information can be found at www.energystar.gov). With the efforts to curb the 9% of total energy controlled by these devices in the united states, the energy saving label was applied to these devices, but failed to account for the fact that further user action was needed after purchasing in order to successfully reduce energy use in a home. Upon study, it was determined that approximately 90% of users never programmed the thermostat for scheduling after purchasing it for a variety of reasons, resulting in no energy use reduction at that source [35]. With approximately 30% of homes having already purchased and installed programmable thermostats as of 2011, these statistics represent an example of how a widespread energy efficiency policy can fail because of a lack of proper prior planning, leading to poor performance. At the end of the day, the Energy Star label was removed from many of these thermostats and they were subject to further study in order to improve their function[40].

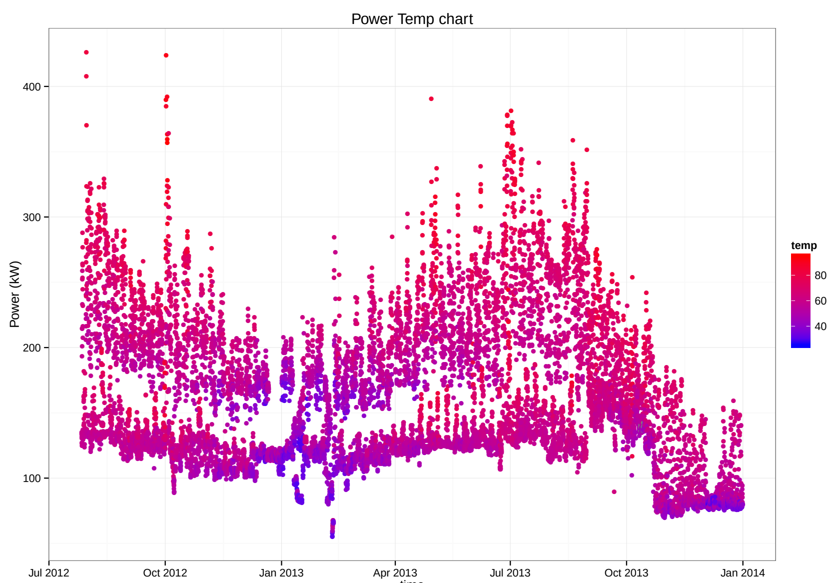


Figure 1.13: Building Power Draw and outside air temperature of a building that has undergone a significant energy efficiency retrofit.

1.6 Marketplace evaluations, solutions, and certifications for energy efficiency

In order for any energy efficiency implementation to make financial sense, it must be characterized against the energy that it is offsetting that would be used instead. Typically, a measure called the cost of conserved energy CCE, originally proposed as an important measure in [34] is used to evaluate this. Simply put, if the CCE is less than the price of the energy that it displaces, then that measure should be put into place, as it is economically rational. CCE is defined:

$$CCE = \frac{Investment}{Annual\,Energy\,Savings} * \frac{d}{1-(1+d)^{-n}}$$

where d is the discount rate of the investment and n is the number of compounding periods (in this case, years) that the investment occurs over. For a good overview of this and other investment formulas, please refer to [37].

If decision makers made all the economically rational choices for changing the energy efficiency of their buildings an processes, then we would be in a much better place as a species on this planet. In fact, many of the energy efficiency measures that could be made yet have not would occur at a negative cost of conserved energy - that is, a decision maker would essentially get paid to make these decisions. This is presented well in the McKinsey Abatement Curve, which is discussed in [14], but will not be replicated here for the sake of brevity. People need different motivations for saving energy than just saving money, or else they would already be doing it.

In the context of buildings, there have been various rating systems that have been introduced that can certify a building as "green" or "sustainable." The most commonly referred to of these systems is the leadership in energy efficiency and design (LEED) rating system by the United States Green Buildings Council (USGBC). LEED has several designations that are available for buildings and building operators, including the Exiting Building Operations and Maintenance (EBOM), Interior design and construction (ID+C),and Building design and construction (BDC). Information about this system can be found at http://www.usgbc.org/leed/certification. While this seems all well and good, further analysis reveals that point based systems do not always achieve the goals of reducing energy use, and one study says that while on average, LEED buildings have 18-39% less EUI than their counterparts, 28-35% of those buildings actually use more energy than their conventional counterparts. As an additional damning statement, it was noted that the energy performance of LEED buildings has little to do with their certification level (the system goes: certified, silver, gold, platinum in order of increasing performance [38]. Another study finds that these measures depends on a particular definition of mean energy intensity that allows these buildings to save energy, and does not account for energy generated off site and does not offer any evidence that LEED buildings actually save energy [49].

1.7 SPC as a framework in buildings

This dissertation will explore a method for performing statistical process control work in buildings with the end goal of rapidly identifying finding root causes of anomalies in performance. The key parts in this method are:

- Exploratory data analysis - the data must be analyzed using time series methods in order to find key variables.
- Prediction - A predictable building is a necessary but not sufficient condition for a well behaved building.
- Feature Extraction - Physical characteristics of the building must be portrayed into real features which can define the building behavior
- control chart construction - Out of the ability to predict both actual and extracted features of the building, an analysis can be made as to the quality of the building's performance.

Chapter 2

Time series analysis and prediction in the context of buildings

2.1 Introduction

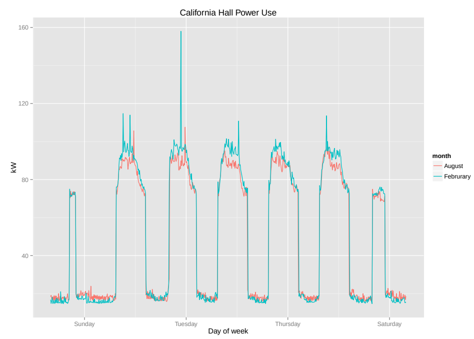
Time series analysis is a systematic technique for exposing underlying trends and information within data that is temporally correlated and therefore cannot be considered independently and identically distributed [51]. The first step in this approach is an exploratory data analysis (EDA) that involves several steps:

- De-trending and differencing data
- Computation of the signal's autocorelation (how much the signal relates to itself)
- Smoothing and simple forecasting in order to understand the predictability of the time series

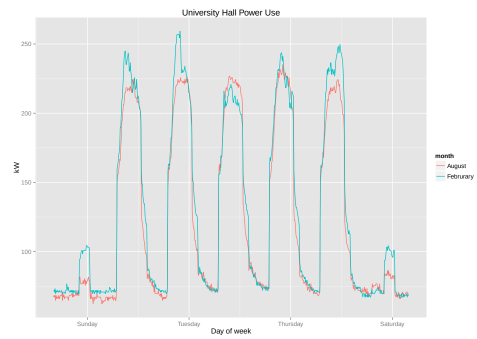
These steps underpin the framework can be set up for more advanced prediction of the time series,a although it should be noted that time series prediction is sometimes a place where simpler models outperform more complex ones [32].

2.2 Examples of time series data in commercial buildings

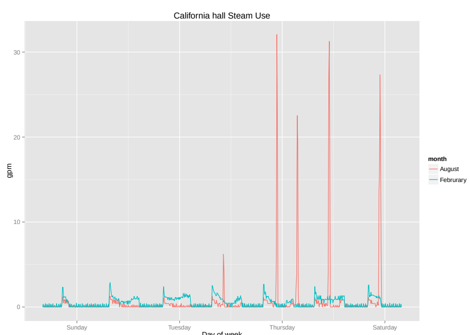
In order to apply statistical techniques to the data of a building's power use, we must first have an appropriate description and understanding of the data sets that we will be dealing with. an example of some power data is shown below in fig. 2.1 - this figure shows electricity and steam usage in two buildings at two different points in the year of 2012: February and August. It should be noted that some parts of the signal have large sudden spikes, and this could be because of a real issue with building performance or because of an error in the equipment doing metering in the buildings.



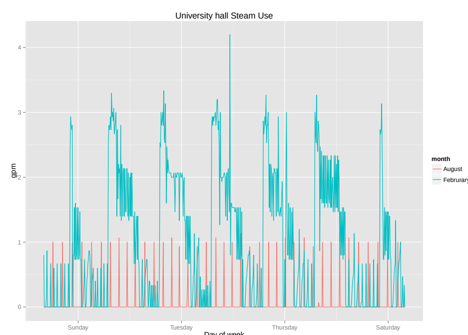
(a) California Hall Power Use



(b) University Hall Power Use



(c) California Hall Steam Use



...

(d) University Hall Steam Use

Figure 2.1: Electricity and Steam Use in two buildings

(a) Blue for Februrary, Red for August

2.3 Exploratory data analysis on building full and sub-meters

The first step in doing exploratory data analysis on a time series is determining if the signal is autocorrelated - that is, whether present values of the data are correlated with past values of the data. This is treated similarly to correlations between data sets that are related without time variation, where we would graph a scatterplot for each data set and compute an r value in order to compute the relationship between the data. For a time series, however, this would compute a scatter plot for every lagged time step, which would result in a great many plots that would be hard to interpret. Instead, the autocorrelation plot is used - the r value of said scatter plot at each time lag that we are interested in. These r values are computed as:

$$r_k = \frac{\sum\limits_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum\limits_{t=1}^{T} (y_t - \bar{y})^2} \tag{2.1}$$

Figure 2.2 is an example ACF plot and more information on this concept can be found in [23].

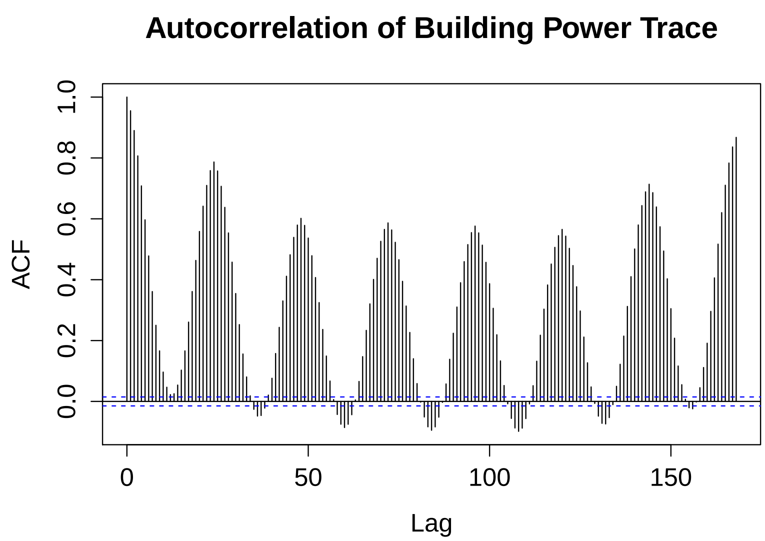


Figure 2.2: Autocorrelation of a building's power use with max lag 168 hours. Original power trace of the building can be found in fig. 1.13

It is also useful to understand what times are specifically correlated to what other time lags. "Partial" autocorrelation, or PACF, is a tool to do this. Where the autocorrelation determines if a point is correlated with every point between itself and some time in the past, pACF only determines if it is correlated with that one particular point, thus revealing information about structural linkages in the lagged data.

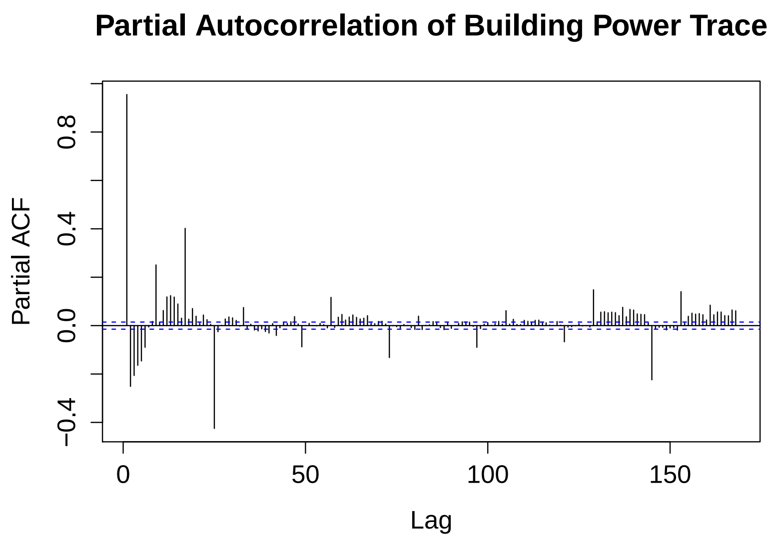


Figure 2.3: Partial Autocorrelation of a building's power use with max lag 168 hours. Original power trace of the building can be found in fig. 1.13

2.4 Building Load Profiles

This framework allows for critical examination of the state of the art in building energy use modeling. As an initial application, demand response baselining is examined - this is computing a counterfactual for what energy use would have been if action had not been taken to reduce it, as in fig. 1.11. The simplest method of this is to evaluate what the load was like on a similar day, and just compare that to the actual, as in fig. 2.4. This estimation of how much power would have been used if no intervention had been taken is the building's "Baseline Load Profile" or BLP.

BLPs are often used to determine the amount of incentive that can be compensated to an electrical customer for load curtailment after a demand response event, and for this purpose should contain several qualities which are enumerated in [26]. The core principles state are:

- Accuracy - The BLP can correctly reflect the actual avoided usage.

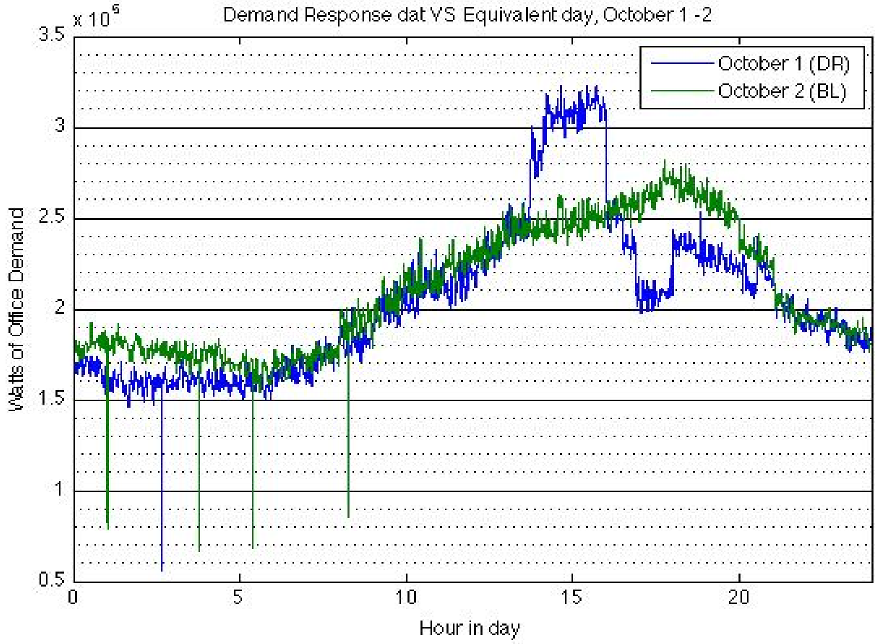


Figure 2.4: Demand response baseline by comparing similar day. As we can see, sometimes performing a demand response event involves shifting load earlier in the day.

- Integrity - The BLP cannot be easily gamed by adjusting consumption before the event.
- Simplicity - Anyone should be able to calculate and understand the baseline
- Alignment - The goals of the baseline program should be considered in BLP generation.

An effective way to predict building energy use is a multivariate regression model developed by Johanna Mathieu in conjunction with LBNL. The model, reproduced here from [33], has several very important features regarding building power use which will be useful for discussion. In the LBNL model, the load in a building is predicted by:

$$\hat{L}_o(t_i, T(t_i)) = \alpha_i + \sum_{j=1}^{6} \beta_j T_{c,j}(t_i) \tag{2.2}$$

when the building is occupied and

$$\hat{L}_u(t_i, T(t_i)) = \alpha_i + \beta_u T(t_i) \tag{2.3}$$

when the building is unoccupied.

This model accounts for several factors when predicting building power use including occupancy, varying behavior at different temperatures, time of day and day of week. It is important to draw distinctions between predicting the building in an occupied mode ($\hat{L}_o$ and an unoccupied mode $\hat{L}_u$). If we recall from fig. 1.4, the load shape characteristics differ significantly between a Sunday (unoccupied), and a Monday(Occupied). Each T_i refers to a temperature block which is regressed on - there are six blocks in this model, representing the fact that buildings react different at different temperature ranges. α_i represents the time of week that the model is predicting at, while β_j is an indicator variable used to denote whether that temperature block is currently being predicted on.

This method of predicting buildings is sound but it is not standard practice for industry. The value of a building's baseline on previous days for utilities is often a simple linear combination of previous day's peak hours. Con Edison uses a formula to average peak hours of the previous 5 comparable workdays, excluding holidays, weekends, and special days [13]. PG&E defines their baseline method by a "10 day baseline" that averages the three hottest days out of the previous ten [17]. Both of these methods have an option for weather normalization of loads, but display fundamental weaknesses through their lack of accounting for how buildings respond at higher temperatures, lack of clear inclusions of time series effects, and extrapolating lower load days in predicting the most load intensive days.

2.5 Better prediction through time series analysis

There are many methodologies for making BLP predictions for buildings, and they all have different strengths and weaknesses. Specific methods are reviewed later in this chapter and the science to formulating and selecting an appropriate forecast model for predictions is applied. Appropriate models are chosen using a loss-based, cross validated estimation procedure.

Loss-Based Estimation and performance evaluation

In order to properly evaluate any time series forecasts that are made in our predictions, we must choose the appropriate loss functions to measure our error. A loss function is a measure of how far the estimates of our model are from the actual results, described as some function $L(\hat{y}_t, y_t)$, where $\hat{y}_t$ is the predicted value at some time t and y_t is the actual value at that time. In order evaluate how well our loss functions apply to the goals we are trying to achieve, we must understand several features of "loss" as it applies to time series forecasts. Firstly, we want to observe that the relative performance (and thus utility) of a forecast is dependent upon the forecast horizon - the amount of time that the forecast predicts into the future from the time point that the forecast is made from [30, 32]. Secondly, we will be forecasting, so our evaluation must be made out of sample [54]. Finally, not all loss functions are created equal - some are scale invariant, while others vary with the size of the prediction - typically scale invariant functions are normalized in some manner - either by themselves or by another reference loss function [22].

After properly framing how we evaluate predictions, we must then choose with what measures to evaluate our predictions with. The following loss functions are of note:

- Mean Absolute Percent Error (MAPE) - a standard of forecasting for a long time, this method is scale invariant by virtue of dividing by the actual value being predicted on. When there are zeros in the data, MAPE will be undefined by dividing by zero, and this value can also go toward infinity without restriction.

$$MAPE = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \tag{2.4}$$

- Mean Squared Error (MSE) - a quadratic loss function, minimizing MSE is equivalent to minimizing the variance if the estimator is unbiased.

$$MSE = \frac{1}{n} \sum_{t=1}^{n} (y_t - \hat{y}_t)^2 \tag{2.5}$$

- Root Mean Squared Error (RMSE) - The square root of the MSE, this shrinks large errors relative to the previous loss function.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2}{n}} \tag{2.6}$$

- Mean Absolute Error (MAE) - A scale invariant error function, this is good for judging how well algorithms perform on a single signal, but not across signals of multiple magnitudes.

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |y_t - \hat{y}_t| \tag{2.7}$$

- Root Mean Squared Percentage Error (RMSPE) - A scale invariant measure that is less sensitive to large errors than some of the above measures.

$$RMSPE = \sqrt{\frac{100\%}{n}\sum_{t=1}^{n}\left(\frac{y_t - \hat{y}_t}{y_t}\right)^2} \tag{2.8}$$

- relative Root Mean Squared Error (rRMSE) - This is an excellent way to take a scale-variant loss function and turn it into a scale-invariant loss function by comparing it to a known method. Across many signals, this is a good way to get an honest understanding of the power of some prediction model against another.

$$rRMSE = \frac{RMSE}{RMSE_{naive}} \tag{2.9}$$

- relative Mean Absolute Error (rMAE) - This measure of loss is normalized by the error of a naive loss function in order to asses the behavior of more advanced models.

$$rMAE = \frac{MAE}{MAE_{naive}} \tag{2.10}$$

- Mean Absolute Scaled Error (MASE) - In order to use relative error terms as above, we need multiple forecasts on the same time series. When multiple forecasts are not present, there must be a way to evaluate the out of sample forecast accuracy of the method, and in this case, MASE is a good loss function. There are additional terms in this case, where e_t denotes the error of a function at that step. In this case, $e_t = y_t - \hat{y}_t$

$$MASE = mean\left(\frac{\sum_{t=1}^{n} |e_t|}{\frac{n}{n-1}\sum_{i=2}^{n} |y_i - y_{i-1}|}\right) \tag{2.11}$$

Now that we have gone over loss functions to make sense to evaluate time series forecasts, we must set up our forecasting methods in such a way as they make sense to evaluate. In order to do this, we will delve into the basics of cross validation techniques, and how they apply to time series.

Cross Validation

Cross validation is a technique that is used to understand how well a prediction methodology will perform in practice. There are several methods of cross validation, but all of them depend upon using testing sets and training sets in order to assess the efficacy of a statistical prediction method. There are many techniques of cross validation, and we will review a few here, with other examples that can be found in [3, 23].

- Leave-out-one cross validation - take a test set of data observations $y_1...y_n$, remove a data point at index i, and make that the test set. Compute the error $e_i = y_i - \hat{y}_i$ for each of $i = 1, ..., n$ and enter this error into a loss function that achieves the goals of the cross validation. The value of that loss function is the the cross-validated measure of performance of the forecast method.

- Leave-out-k cross validation - this method is exactly as above, but instead of leaving out one at each step, k samples are left out to be cross-validated.

- k-fold cross validation - in this method, the population sample is randomly partitioned into k sub-samples, and cross validation is done while leaving one out at each step.

A variation on k-fold cross validation is is *rolling window forecast evaluation,* which is sometimes referred to as *time series cross validation.* We will use this framework to do our evaluation of building load profiles in this work. In this chapter, all prediction methods used were trained on 180 days of data (4320 hours), and then tested on 1 week (168 hours) worth of data. In order perform cross-validation, these test sets are then analyzed by computing various loss functions across each of the M test sets produced through this process, with M varying by the amount of data that is available for a building. fig. 2.5 illustrates this process.

Two important nuances play into the cross-validation procedure we use in this case:

1) Some forecasts perform differently depending on how far out the forecast origin is from the point is that is being forecast - that is, the forecast has a different efficacy at different horizons.[30, 32]

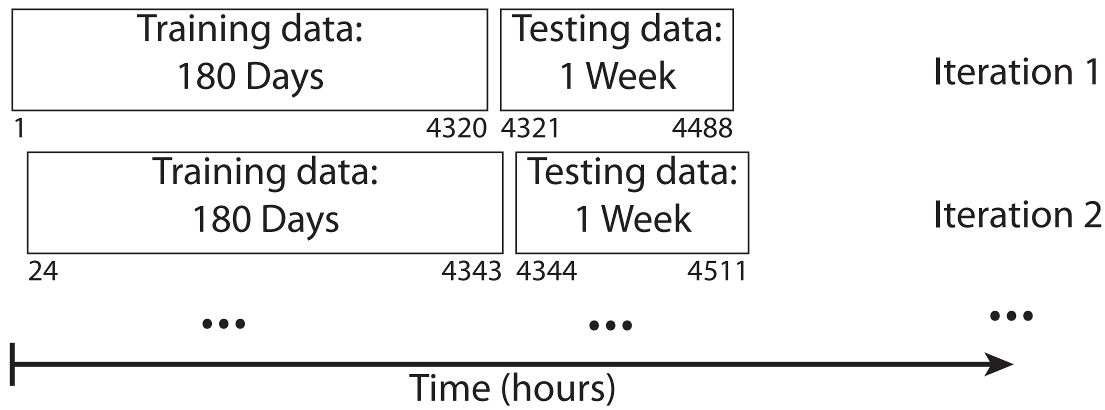


Figure 2.5: Rolling window cross validation allows us to more effectively evaluate prediction methods on time series data.

2) Some hours of day are easier to forecast than others. That means that forecasting what will be happening at 2:00 PM may have a different level of efficacy than forecasting what will be happening at 3:00 AM. [53]

The rolling forecast window model that we use deals with these issues and also has the advantage of making the forecast models *flexible* - that is, they can adapt to new training conditions over time as the rolling window moves along. This ability to make our models adaptive is important and will be explored even more in the next chapter.

2.6 Efficacy of several prediction models on a single building

In order to test our ability to predict buildings effectively, several BLPs were programmed in R. Each of our prediction methods falls into one of three categories: Time Series, Regression, or Naive. Time series method explicitly account for autoregressive errors in their methodology. Regression models use statistics that assume the data are i.i.d., an assumption that may not be valid in the case of time series data. Naive models assume that what is happening in the future is some simple linear combination of what has happened in the past. Methods were all written in R. Further information about the performance of each of the primal methods may be found in [5].

Naive Models

Naive models are this are those that essentially extend what is happening out into the future without advanced mathematical methodologies. We classify two naive models in this case.

- Mean-week model: This model assumes that every day in the data is an average day - that is, every Tuesday is the average of all Tuesdays at each hour. As an example, to predict the load at 4:00 PM on Friday, May 9 2014, one would average the load over all 4:00 blocks on all Fridays in the data set. This model is a common benchmark for forecasting studies in buildings, and tends to work well in mild environments. This model has the weakness that it assumes stationary in the data - that the data varies constantly around the mean. The mean-week model can be written as: [19].

$$\hat{y}_t = \beta_0 + \beta_1(\text{hour}) + \beta_2(\text{weekday}) + \beta_3(\text{hour}) \cdot (\text{weekday}) \tag{2.12}$$

- Seasonal Naive model: This model assumes what happened last week will happen this week - thus it extracts the weekly seasonality in a naive manner, and projects it onto the next week. A data set of building energy use at hour apart points that is differenced by 168 hours is the Seasonal Naive Model.

Regression Models

We evaluated two different regression models, which assume that any future event, regardless of time horizon from prediction hour, are a weighted linear combination of previous events.

- Day-Time-Temperature (DTT) model: The DTT models is a regression model that assumes that "heating days" and "cooling days" should be modeled with separate coefficients, and includes factors for what time of day and day of week is being modeled. The DTT model is widely used, and is the basis for many simplistic studies of modeling buildings for incentive programs and efficiency evaluations. The DTT model can be written as:

$$\begin{multline}\hat{y}_t = \beta_0 + \beta_1(\text{hour}) + \beta_2(\text{weekday}) + \beta_3(\text{hour}) \cdot (\text{weekday})) \\ + \beta_4(\text{temp50}) + \beta_5(\text{temp65}) \tag{2.13}\end{multline}$$

  Where

$$\text{temp50} = \begin{cases} \text{temperature} - 50 & \text{if temperature} < 50 \\ 0 & \text{if temperature} \geq 50 \end{cases}$$

and

$$\text{temp65} = \begin{cases} 65 - \text{temperature} & \text{if temperature} > 65 \\ 0 & \text{if temperature} \leq 65 \end{cases}$$

- LBNL Model: The LBNL regression model - described in eqs. (2.2) and (2.3) is similar to, but more advanced than the DTT model [19].This model includes a piecewise regression over many temperature blocks, not just "heating" and "cooling" days, accounts for occupancy, and can factor in the time of day and day of the week, just like the DTT model. The LBNL model is widely used in industry, and has been cited as useful for such purposes as determining incentives for demand response events [33]. If we expand the terms given above, the LBNL model is written as:

$$\hat{y}_t = \beta_0 + \beta_1(\text{hour}) + \beta_2(\text{weekday}) + \beta_3(\text{hour} \cdot (\text{weekday}) + \sum_{i=1}^{6} \alpha_i (T_i \cdot \text{occupied}) + \gamma(\text{temp} \cdot (\text{1-occupied})) \quad (2.14)$$

Time Series Models

A time series method accounts for the characteristics of time series data: trend, seasonality, and cyclic behavior. In addition to this, it is important to note that time series models account for autoregressive errors, that is - the error in the current step caused by the error in the previous step of the forecast.

- Autoregressive integrated moving average (ARIMA) - This model, written in the equation below, is a common method for time series forecasting. This is often written as ARIMA(p, d, q), and the variables refer (respectively) to the level of autoregression, differencing, and moving averaging of the forecast that is done. The forecast is a function of these three parameters, with autoregression aiding the correction of previous errors (sometimes from a previous model), differencing lending stationary to the signal, and smoothing allows the signal to be more accurately represented.

$$y_t = \underbrace{(y_{t-1} + ... + y_{t-d})}_{\text{Differencing}} + \underbrace{(\phi_1 y_{t-1} + ... + \phi_p y_{t-p})}_{\text{Autoregressive lags}} + \underbrace{(\theta_1 w_{t-1} + ... + \theta_q w_{t-q})}_{\text{Moving average}} + w_t \quad (2.15)$$

In this model, w_t represents white noise that is included in the prediction model in order assume that there is some remainder that is unpredicted on in the residual.

- (BATS) - BATS stands for Box-Cox Transformatiom, ARMA (ARIMA$(p, 0, q)$ Errors, seasonal and Trend components. As it sounds, this model transforms the data, auto-regresses on errors, and accounts for trends in the data, predicting through exponential smoothing. BATS is a six parameter model and requires quite a bit of tuning to get to work properly, but can display superior results given some adjustment. more information on BATS can be found in [7]

- Trigonometric BATS - This is a version of BATS that includes trigonometric terms in the transformation of data in order to aide the computational efficiency of the transformation.

- Holt-Winters method - This is a forecasting method that is based upon exponential smoothing of the data for looking forward while accounting for seasonality. More information can be found in: [20]. The equations to predict using Holt-Winters method are:

$$\hat{y}_t(k) = (s_t + kT_t)\, I_{t-s+k}, \tag{2.16}$$
$$s_t = \alpha(y_t/I_{t-s}) + (1-\alpha)(s_{t-1} + T_{t-1}), \tag{2.17}$$
$$T_t = \gamma(s_t - s_{t-1}) + (1-\gamma)T_{t-1}, \tag{2.18}$$
$$I_t = \delta(X_t/S_t) + (1-\delta)I_{t-s} \tag{2.19}$$

  Where s_t is the local level smoothing term, T_t is the trend, and I_t is the seasonality of the time series.

- Double Seasonal Holt-Winters method - This is a forecasting technique that has been shown in [55] to be effective in forecasting grid operations on data sets from european operators. It is computationally expensive but accounts for many parameters of a double seasonal model. The model is detailed in these equations:

$$\hat{y}_t = l_{t-1} + b_{t-1} + s_t^{(1)} + s_t^{(2)} + d_t, \tag{2.20}$$
$$l_t = l_{t-1} + b_{t-1} + \alpha d_t, \tag{2.21}$$
$$b_t = b_{t-1} + \beta d_t, \tag{2.22}$$
$$s_t^{(1)} = s_{t-m_1}^{(1)} + \gamma_1 d_t, \tag{2.23}$$
$$s_t^{(2)} = s_{t-m_2}^{(2)} + \gamma_2 d_t \tag{2.24}$$

where m_1 and m_2 are the seasonal periods of both seasonal cycles and d_t represents the white noise of the prediction error. l_t and b_t respectively represnet the level and trend components of the time series. in addition to all these, $\alpha, \beta, \gamma_1 \, and \, \gamma_2$ are smoothing parameters that can be tuned in order to adjust the model.

- Seasonal Trend Loess (STL) - STL models decompose the data into a seasonal, trend, and remainder component in order to aide in forecasting into the future. The seasonality and trend are extracted, and prediction is done by regressing on the remainder of component. It should be noted that this method can take other input factors as covariates in the regression function, and this is demonstrated in our modeling of a building shown below, and in this particular case, we denote STL.LBNL and STL.TEMP as models with the inputs form the LBNL models and the outside air temperature, respectively. This convention is held for other methods of prediction from STL methods. For purposes of brevity, please refer to [4] for additional details on the STL method, which is used extensively going forward. This model can also simply extend the seasonality or trend over a given period and extend it in much the way that a naive model operates in order to project into the future. An example of data decomposed by the STL method is given in fig. 2.6 for illustrative purposes.

The results of applying these methods to a single building can be seen below in fig. 2.8. This particular building that is being predicted on is the Art Museum on the UC Berkeley campus, for which predictions were done over 13039 hours of power draw data, with the original meter data shown in fig. 2.7. In this framework, the first 4320 hours of data were used for training, and thus the predictions were done over 8719 hours for this building. As we can see, the LBNL regression performed very well overall in this case, scoring more than 10% better than the seasonal naive model by measure of rMAE.

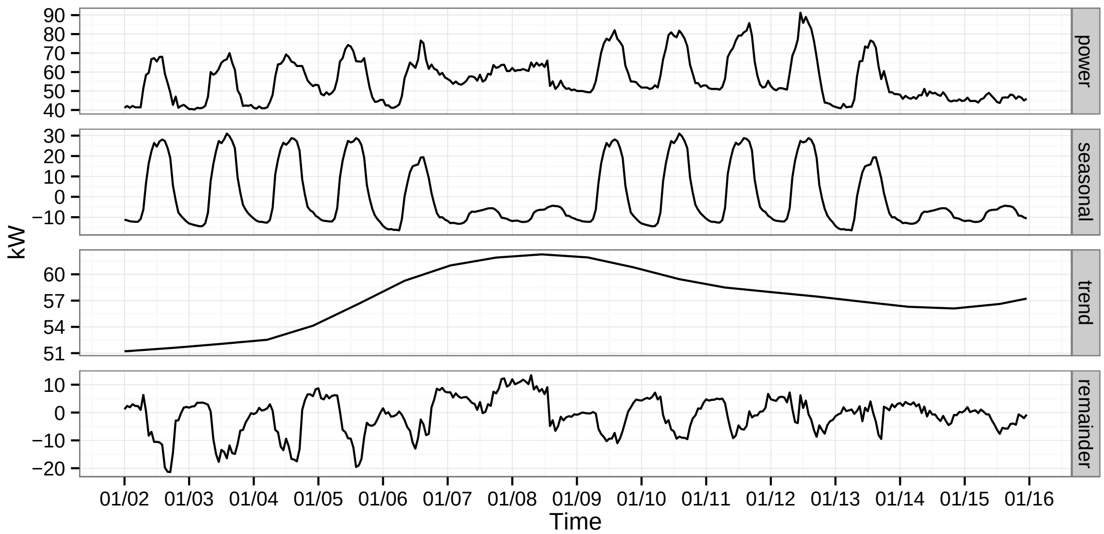


Figure 2.6: Seasonal Trend Loess (STL) decomposition is one way to de-decompose a time series for better prediction.

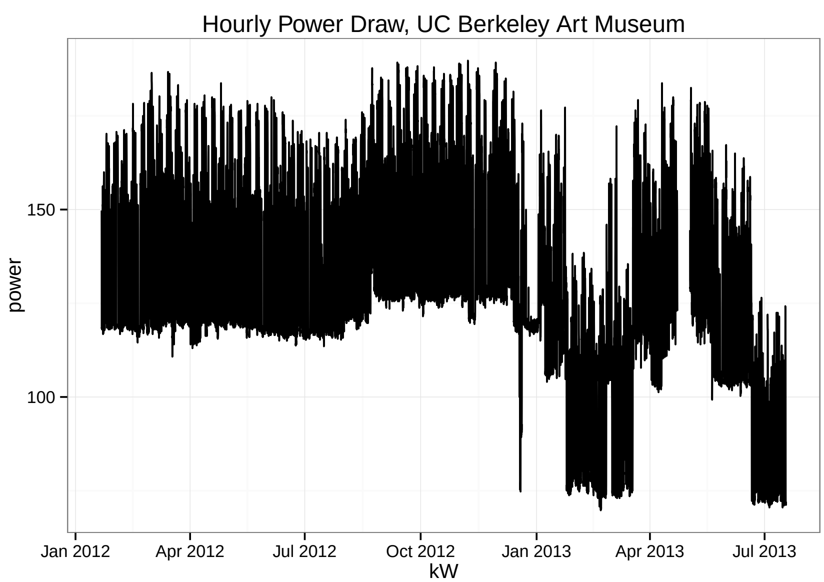


Figure 2.7: Hourly Power Draw of the UC Berkeley Art Museum.

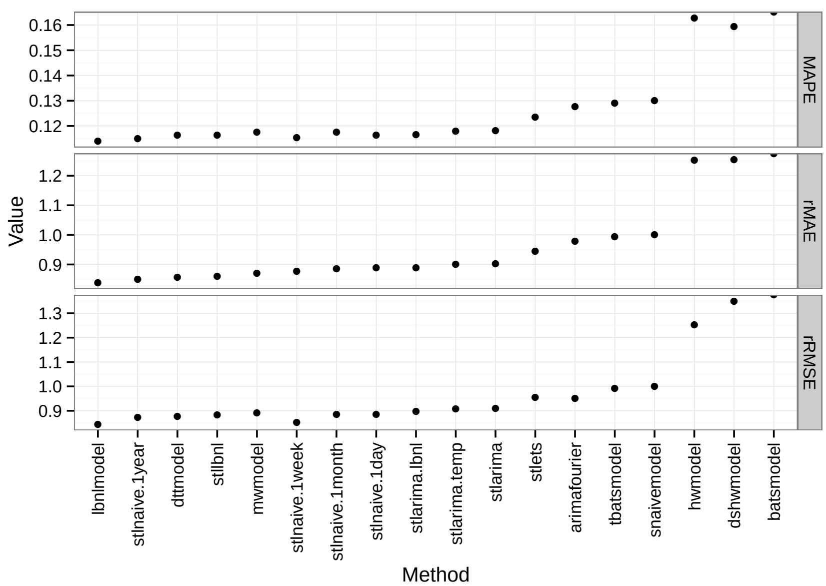


Figure 2.8: Results of BLP generation using Cross Validated time series methodology for a single building: In this case, the Art Museum on the UC Berkeley Campus.

2.7 Implications and conclusions

A framework has been constructed to properly set up baseline prediction models in commercial buildings. Several steps must be taken in order to properly produce BLP models.

- Step 1: Identify goals for the production the prediction model and baseline load profile.
- Step 2: do exploratory data analysis to examine autocorrelation and structural elements of the data that is being predicted on. This can also include doing STL decomposition on the data in order to examine its structure.
- Step 3: Identify loss functions that align with the goals. Different loss functions have different strengths. It might be advantageous to use rMAE to compare multiple models that are run many times on a time series, or MASE for ones that are not run multiple times. Other loss functions such as MSE exaggerate large errors and allow for models with several points that have large errors to stand out from ones that have many points with small errors.
- Step 4: Do predicting using cross-validated models. It is important to construct models to be cross-validated in order for meaningful model selection to be done. Doing this reduces the chance that a model is over-fitted and only performs well in-set.
- Step 5: Find the best model for an application. After going through these steps, the best model for any application chosen should stand out from the field of candidates.

In the next chapter, this framework will be extended in order to understand how to improve model robustness by adaptively combining models, as well as how to compare models across buildings and to automate this deployment process. Different models have different strengths, but it is possible to leverage all of those at once to allow for multiple goals to be met at the same time for a single prediction forecast.

Chapter 3

Better prediction models through automation and combination

3.1 Introduction

Any engineering solution that cannot be applied at scale is essentially useless, and as such, BLP production must be scalable across buildings quickly and with few obstacles. As described in the previous chapter, baseline production requires comparison of several methods for energy prediction follows the layout given in fig. 3.1. While this is simple in execution for a single building, it is complicated as a scalable problem by the multiple protocols that buildings use to communicate data.

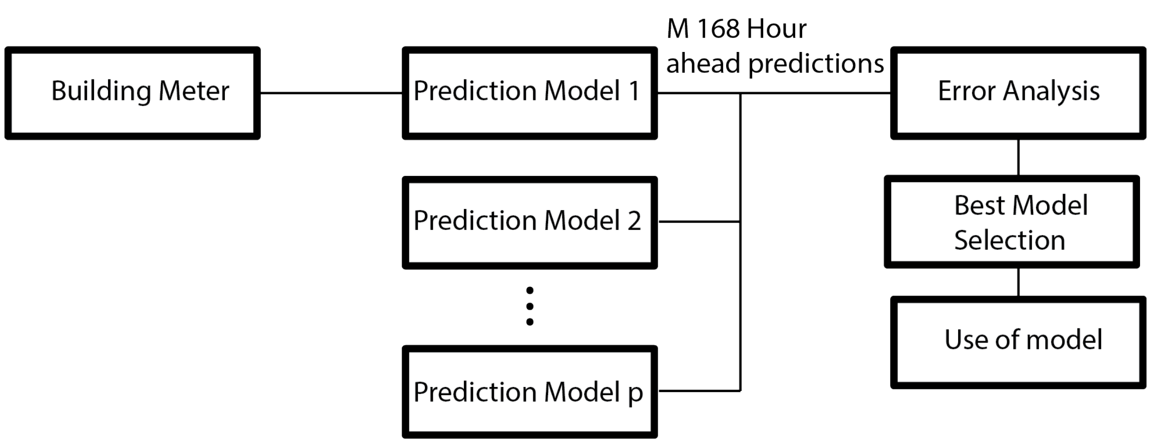


Figure 3.1: Flow chart diagram of multiple models leading to a superior baseline load profile.

3.2 Evaluating models across multiple buildings and timescales

using a simple monitoring and actuation profile (sMAP) for BLP production

Experimental validation of BLP methods on a large number of buildings is tested by employing a system that can be used to predict power usage in many buildings with different control systems. For this purpose, the data in our study was drawn from meters across the UC Berkeley campus, accessed via the Simple Monitoring and Actuation Profile (sMAP). For this application, sMAP bindings developed for R found here: `http://www.cs.berkeley.edu/~stevedh/smap2/R_access.html` were used [21] . All data in this chapter's study occurs between the time frames of 1/1/2010 and 7/24/2013.

Evaluation of results of BLP production on 20 buildings

The methods of prediction outlined in chapter 2 are presented as tested on 20 different buildings on the UC Berkeley campus. Metrics of interest include accuracy and time of computation.. All computations in this experiment were run a machine located in Sutardja Dai Hall at UC Berkeley. The computations of model time were run a Ubuntu system operating with an Intel(R) Core(TM) i7-3930K CPU clocked at 3.20GHz. The system has 32GB of DIMM memory. Each of the 20 buildings had more than 77 available week of whole building meter data available, up to 184 weeks. 50% of the buildings had more than 128 weeks of meter data, and each of the 20 buildings has more than 94.2% of their data intact, with the average percentage of observed points being 97.3% When data was missing was missing within the time periods specified for training, it was imputed by using the mean-week model as a reasonable approximation of what energy draw should be at that time. For the weather data that was used in predictive models, the California Irrigation Management Information System's (CIMIS) was used. CIMIS data can be drawn off of their website at `http://wwwcimis.water.ca.gov`. Data was cleaned to remove outliers and to identify any anomalies associated with daylight savings time before being used for modeling.

Results of the initial study

Figure 3.2 shows each model tested against three different loss functions: MAPE, rMAE, and rRMSE, with each method being tested across all cross-validation folds and all forecast time horizons. The first thing we notice from this figure is that there is not much variety across the methods in the median mean absolute error - indicating that across all horizons, there is not much ability to differentiate the absolute difference between models, except that some models (holt-winters, stets), have a more significant spread of errors. This pattern recurs again with the rMAE and rRMSE measurements, but here is it particularly important to note that the relative measure - the seasonal naive method. With twenty buildings predicted upon, only four models performed marginally better than the model that forecasts that what happened last week happens again this week.

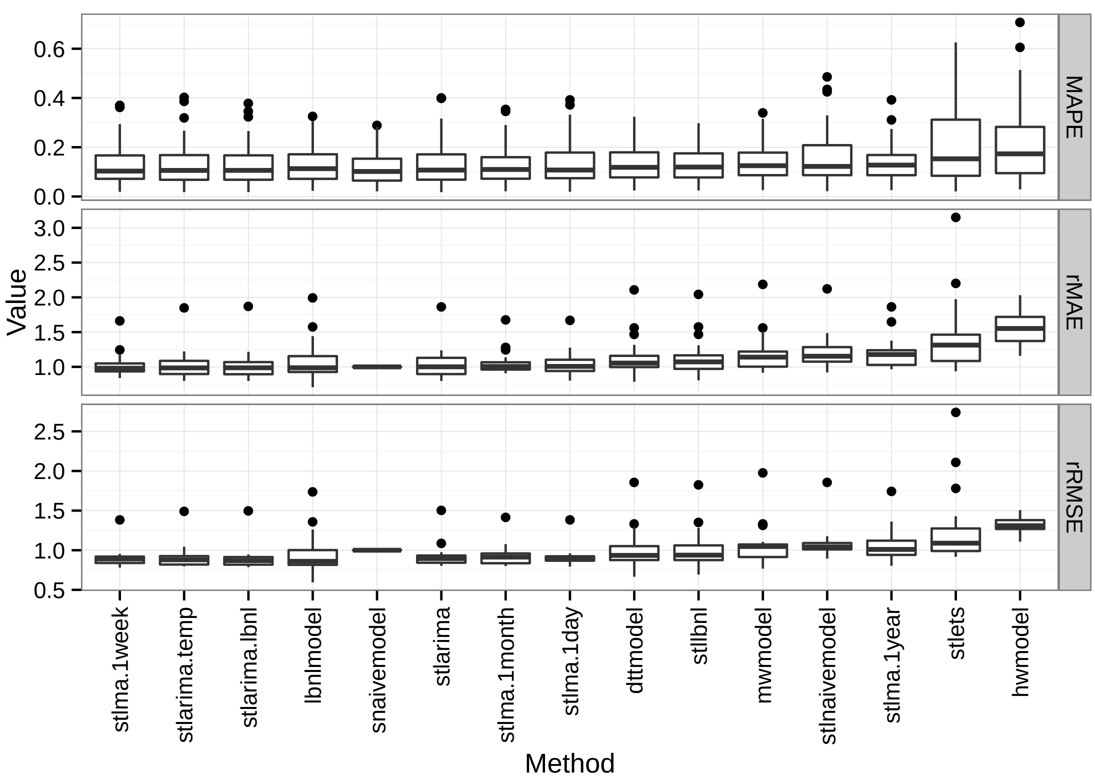


Figure 3.2: Box-plot of errors in twenty building's load profiles

Of course, given that many of the models tested are time-series specific, we can expect some differentiation between models as we separate the errors by time scales that they are specific to. Figure 3.3 separates the results of each model's performance in this sample of predictions by three time-horizon groups: zero to twenty-four hours, twenty-four to forty-eight hours, and forty-eight to one hundred sixty-eight hours, the latter hours all being inclusive.

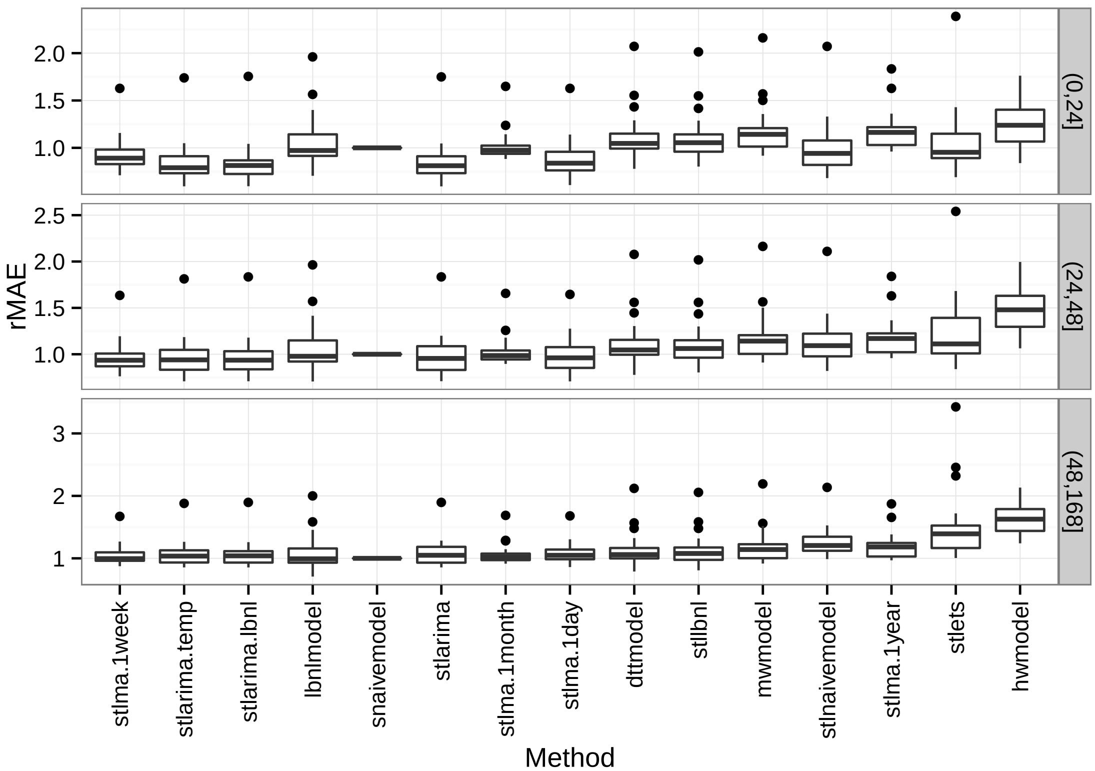


Figure 3.3: Box-plot of by errors by forecast horizon

The evaluation of errors of models with longer prediction horizons, it is difficult to separate out if any model actually outperforms any other model in the long run because all of the models appear to perform to almost the same level. As we move to the shorter prediction horizons, however, the benefits of using time-series appropriate models emerge clearly with most of the time series methods outperforming the regression models (with the exception of the holt-winters methods). In order to further understand the benefits of using time series methods, lets us further examine the benefits of using them vs regression and naive methods. Figure 3.4 compares the relative mean absolute error of three methods vs the seasonal naive method. In this graph, the 95th percentile of errors at a specified horizon is presented as the upper dotted line, while the 5th percentile is presented as the lower dotted line - the median is presented a solid line. As we can see, the LBNL method performs exactly as we would expect - being an intelligently design regression method, it maintains a relatively low amount of error across all time scales, but does not improve with information about the most recent time point. Model differentiation between predic-

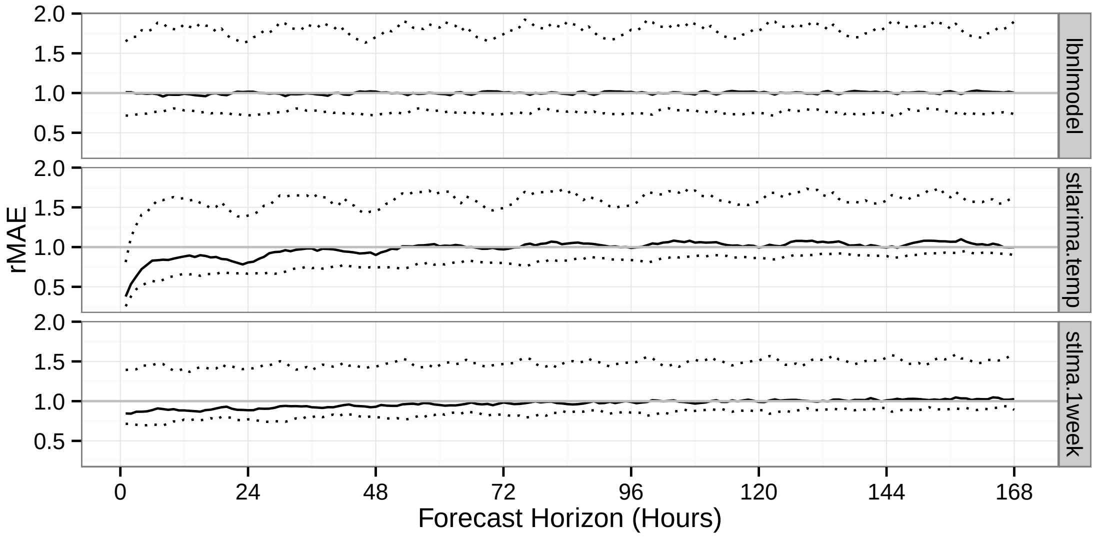


Figure 3.4: Distribution of errors across each of the forecast horizons

tions twenty-four hours ahead and one-hundred-sixty-eight hours ahead can be quite significant, as shown in figs. 3.5 and 3.6. The models presented in these figures for comparison all performed as well or better than the seasonal naive model over these time scales. Time-series models outperform regression models on the twenty-four hour prediction scale - an STL decomposition with prediction on LBNL regression covariates outperform the two benchmark models (seasonal naive and LBNL regression) 90% of the time. While this is great, it should be noted that none of the models (regression or time series) are a clear winner in the longer prediction horizon scheme, with many models outperforming others in at best a random fashion. This is an area of weakness in prediction that we will address in the next section of this chapter - improving model performance across all time scales.

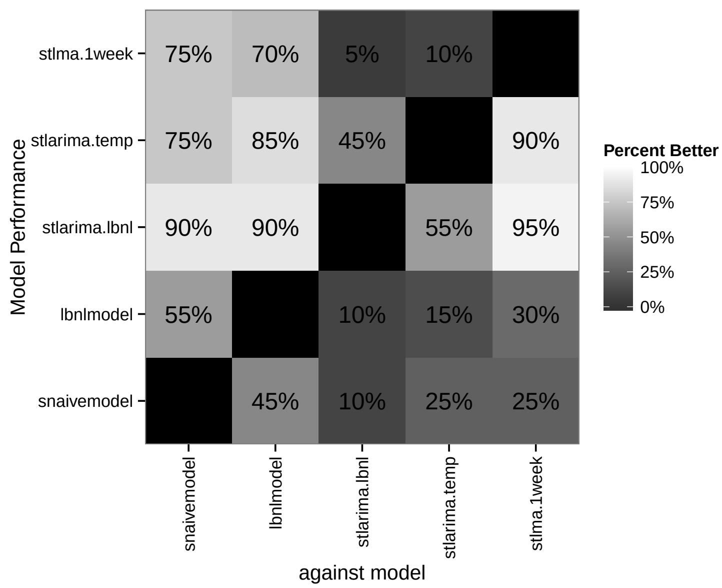


Figure 3.5: Comparisons of model performance against other models in first 24 hours of forecasting

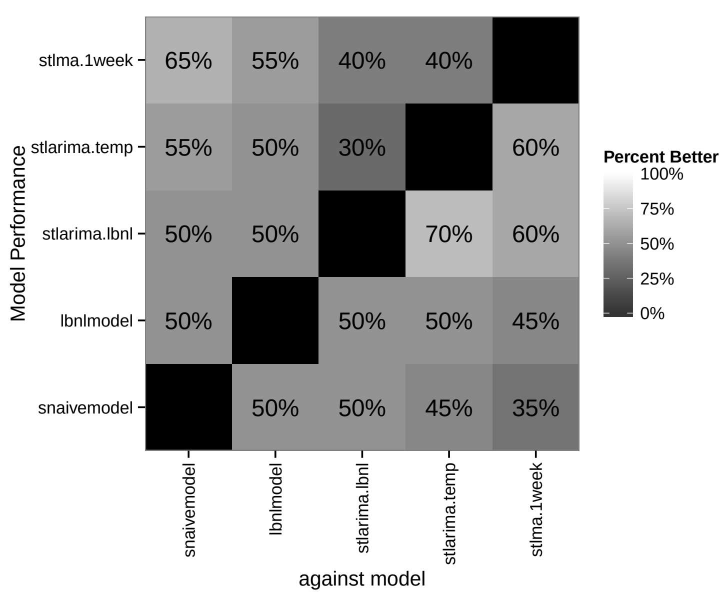


Figure 3.6: Comparisons of model performance against other models in 168 hours of forecasting

3.3 Adaptive Combination models

To more effectively produce predictions for our purposes, we must reduce model error at each step and also each horizon for each prediction model. One method of doing this is to produce an adaptive automated model combination - an ensemble method for combining forecasts into a better forecast model. These methods all work on a rolling window technique that adapts the predication at every time step that is predicted on. Figure 3.7 shows the extension of the process flow for producing these combination models. As we can see, the primal prediction models feed a combination scheme that we will demonstrate in this section.

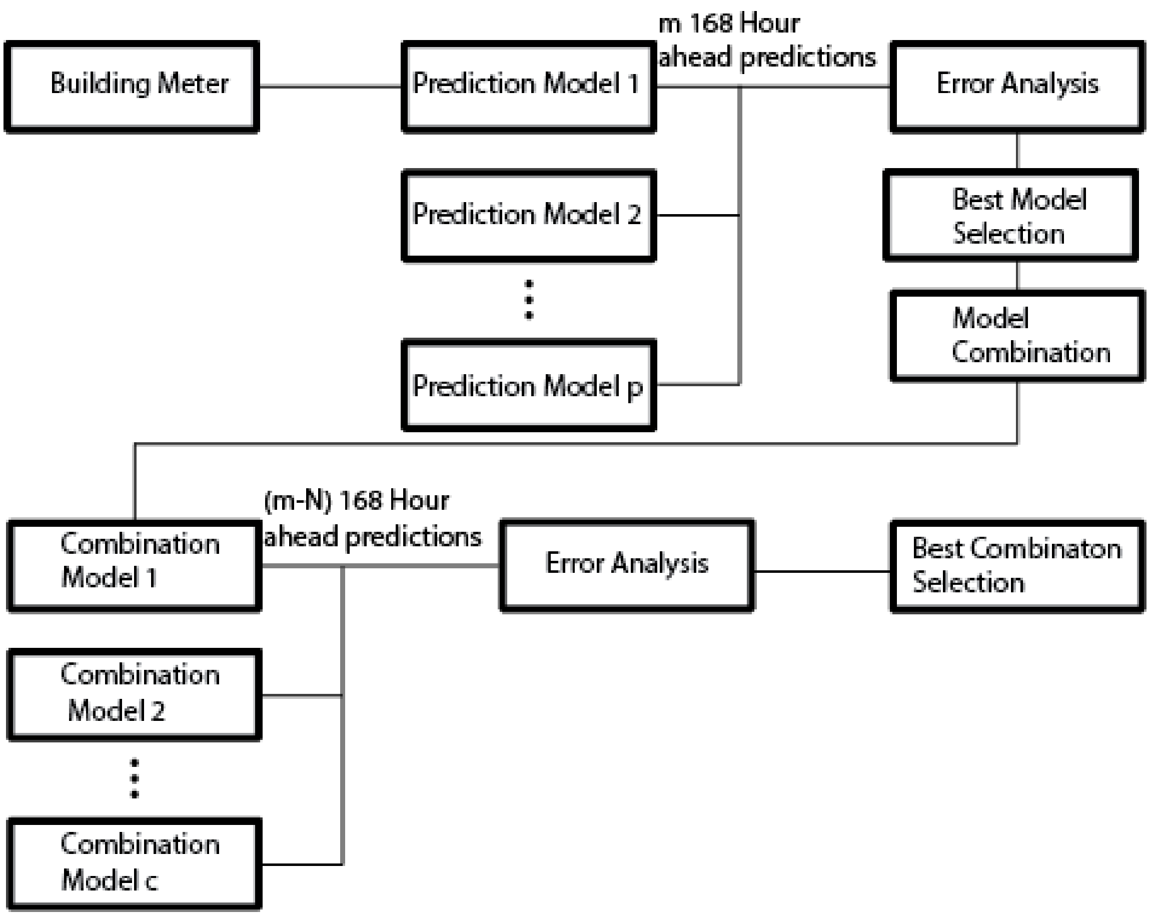


Figure 3.7: Flow chart diagram showing how model combinations can improve predictions.

Adaptive Model I: Best model at each horizon

In first iteration of adaptive combination models, each of the models described in section 2.6 has added to it a sliding forecasting combination scheme as outlined in [31]. This involves breaking the data into sets, in our case consisting of M groups of 168-hour ahead forecast sets, each set containing N samples (In this case, we have chosen $N = 30$, while M varies based on the amount of data available for each building. Each of the sets up to M is indexed as m, and within each m, the index starts at one and goes to J - the maximum horizon predicted in each m). Each m is shifted 23 hours from the last to avoid introducing daily seasonality for forecasting (this would occur if we forecasting from the same time each day), thus overlapping with the previous M for a total of 145 hours.. This scheme is illustrated in **??** below for clarity. After breaking the data up into these sets, MSE is computed for each model in each m set over all J forecasting horizons (with j as the index between 1 and J = 168, in this case), and the best model is selected for each horizon j in each m. For each of $M - N$ groups of training sets, the average best model is then chosen as the forecaster for that horizon in the next m. This forecast is then evaluated in terms of other loss functions, such as rMAE and MAPE in order to ensure the quality of the adaptive model.

Adaptive Model II: Best model confidence interval method

The robustness of forecasting by computing the best forecast method in each set m and forecasting on set $m + 1$ can be improved by computing a confidence interval of methods at each m. In this case, the same procedure is employed as in section 2.2.1, but at each of J forecasting horizons, a confidence interval is computed as $CI_j = m_j + z * \sqrt{MSE_j}$, with z being the z-score corresponding to the level of of confidence required ($z = 1.96$ for a 95% CI in this case) and m_j being the prediction in that m at horizon j. The prediction at set $m + 1$ is then the average of all the models within the confidence interval for each j in that m. This estimated is updated for each of M time blocks in our estimate.

Adaptive Model III: Weighted model by Non-Negative Least Squares Regression

This combination model involves computing a weighted combination of methods at each time step j in each set m. This is done by collecting the predictions of all the models at each time horizon into a matrix A, and fitting them to a vector of the actual power use in the building p with coefficient vector r, as in the equation:

$Ar = p$ fit with constraint $r \geq 0$. For each step m, and new r is selected for A and p and used to predict on step $m + 1$.

Adaptive Model IV: Hybrid Confidence interval weighted Least Squares Regression

Methods II and III are combined in order to produce a weighted average of methods that fall within the confidence interval produced by the method described in section 2.2.2. This is the the same procedure as in 2.2.3., but only those selected by method II are viable for selection and weighting.

3.4 Experimental Setup of Combination Algorithm testing

Inferior combination BLP models are culled according to the results in section 3.2 by noting the amount of time that each model took to run on twenty buildings. Any models that took more than five minutes per building were excluded from the production of combination models in order to produce better models at a lower time cost. The model run time for each of the primal models is presented in table 3.1.

With "long models" were excluded, the remaining models were run on 85 whole building meters and sub meters that were available on the UC Berkeley campus. The results of this are presented below.

Method	Run Time
snaivemodel	0.03 seconds
mwmodel	15 seconds
dttmodel	17 seconds
stlnaive.1month	18 seconds
stlnaive.1year	18 seconds
stllbnl	19 seconds
stlnaive.1week	19 seconds
stlnaive.1day	20 seconds
lbnlmodel	23 seconds
hwmodel	58 seconds
stlarima	91 seconds
stlarima.temp	126 seconds
stlets	4.2 minutes
stlarima.lbnl	7.7 minutes
arimafourier	42 minutes
dshwmodel	91 minutes
tbatsmodel	3.3 hours
batsmodel	11.0 hours

Table 3.1: Computation times on a single building

3.5 Results of combination model production

The first figure of merit to pay attention to is the overall error measurements for each of the three loss functions that were used in the original analysis. Figure 3.8 shows this set of analysis over all time frames in the prediction set (up to 168 hours ahead). As we can see, the mean absolute percent error of the combination models has less spread than the non-combination models, while the relative mean absolute error and relative root mean squared error improve significantly (¡10%) over the seasonal naive method.

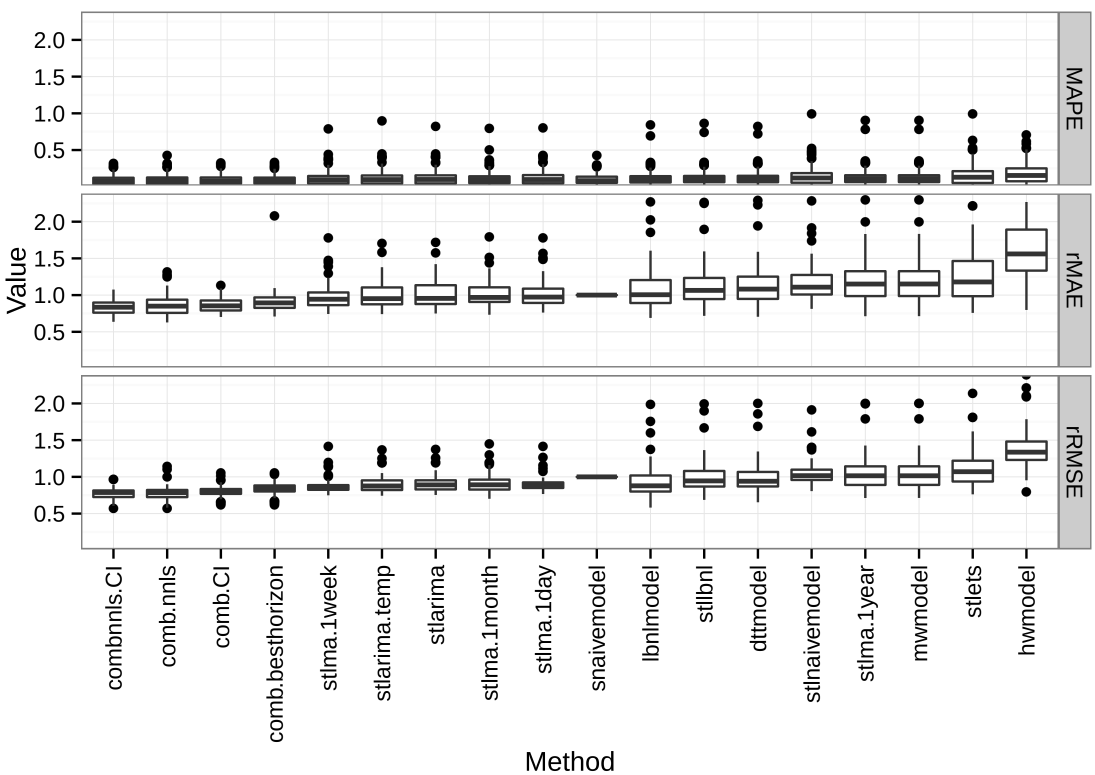


Figure 3.8: Box and whisker diagram of multiple performance measures for combination models after removing models that take excessive time to run.

Digging further into the performance of each method, we continue to break down the models' performance by using rMAE over varying time scales in the same analysis as was done before. We can see that in the zero to twenty-four hour time scale, combination methods show at least a 20% improvement over the seasonal naive method, with at least 15% improvement given in the twenty-four to forty-eight hour and forty-eight to one-hundred-sixty-eight hour timescales. This analysis is given in fig. 3.9, and shows quite clearly the superiority of combination models. Figure 3.10

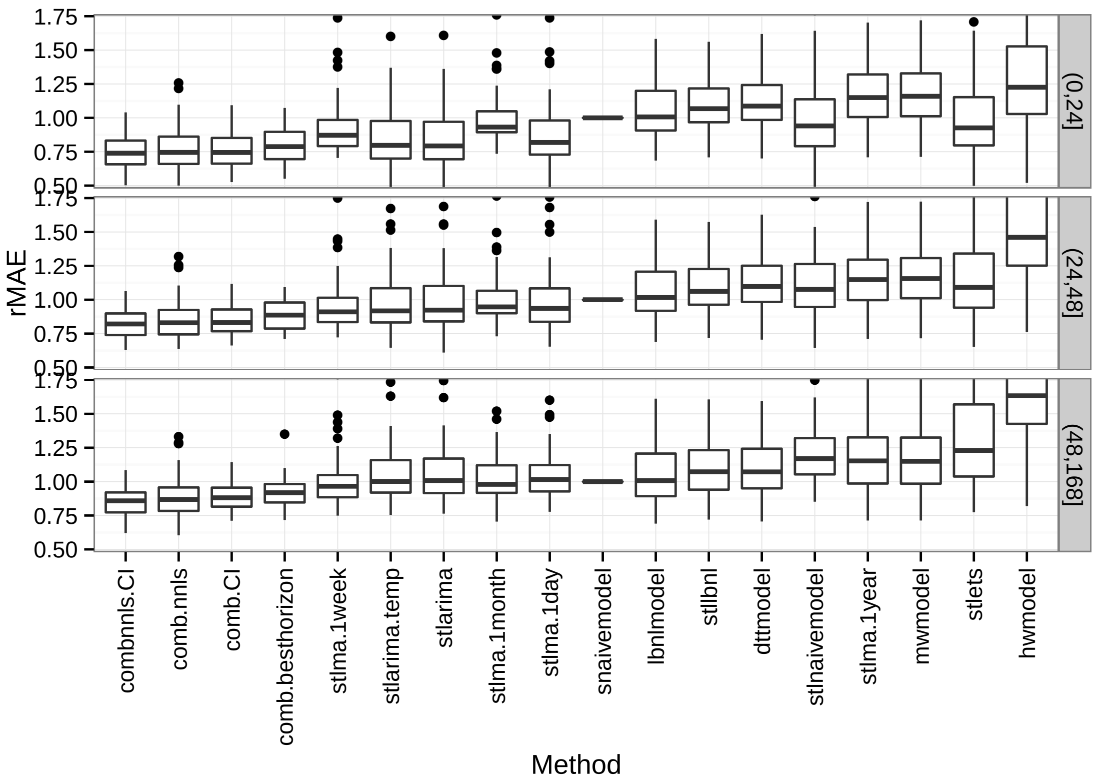


Figure 3.9: Box and whisker diagram of relative mean absolute error for different horizons of combination models after removing models that take excessive time to run.

compares different combination models over all the time horizons available for each of the combination models as well as the LBNL model, STL-Temperature, and Seasonal naive models. The NNLS / CI method produces results across all time scales that are superior to all the other models, while the STL-Temperature model produces excellent results in the short term forecasting arena. Using the combination models, the bounds of errors across the entire timeframe of prediction horizons is much tighter than the original methods. This is valuable for evaluating whether a buildings is malfunctioning, and can be used to anomaly detection as well as for demand response policy evaluation.

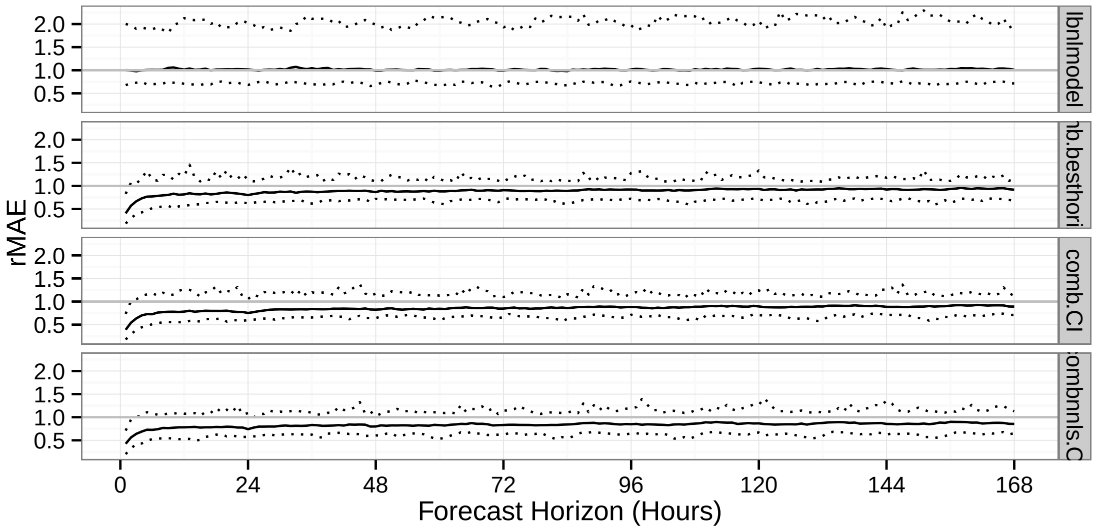


Figure 3.10: Percentile plot of errors of combination models over all horizons. The dotted lines represent the 5th and 95th percentile of error, the solid line represents the median, and the gray line represented the error of the Seasonal Naive model.

Method performance against each of the other models in both day ahead forecasting and across all time horizons is compared.Figures 3.11 and 3.12 illustrates the analysis on a sample of 85 buildings with combination predictions applied to them. A differentiation of the combination models against all of the primal models is clear - the most advanced combination model (NNLS / CI) outperforms the STL arima-temperature model 80% of the time.

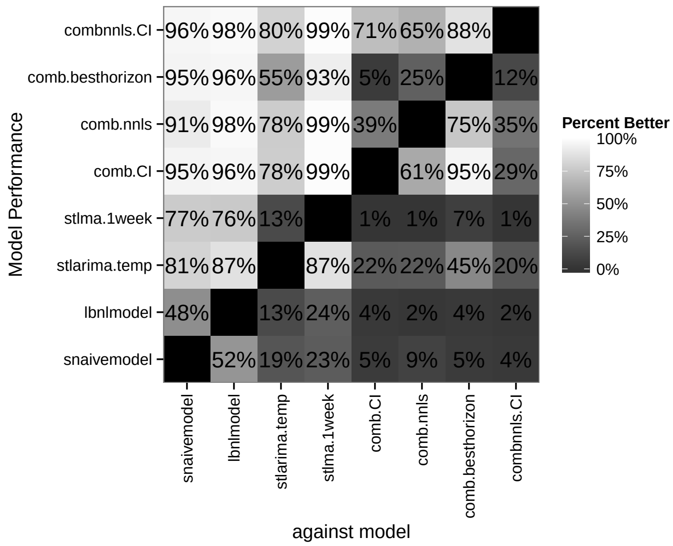


Figure 3.11: Table of performance of methods against other methods including combinations over the first 24 hours of prediction. In each row, the percentage indicates how many times that method performed better than the comparable in the column.

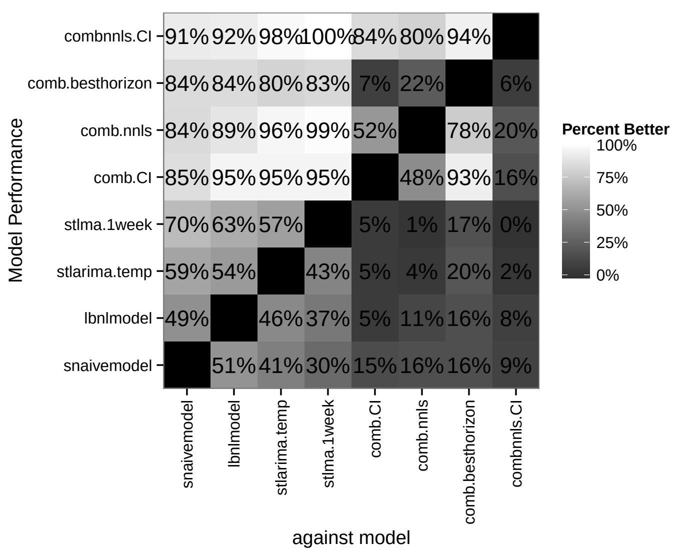


Figure 3.12: Table of performance of methods against other methods including combinations over all 168 hours of prediction. In each row, the percentage indicates how many times that method performed better than the comparable in the column.

3.6 Conclusions and Implications

In this chapter, a method for computing BLP predictions on buildings is outlined, tested, and the result is reported. Several conclusions are very pertinent to the next chapters of this work:

- Time series models are superior to standard regression models for short time scales of prediction horizon in energy modeling for buildings. This should be a trend that is widely adopted by the industry and academia for future studies, as it makes sense to model buildings this way given statistical analysis.

- Adaptive combination outperform known models in the literature across all time horizons tested on our sample of 85 meters. This provides a proof of concept for this application of our combination based adaptive building baseline load profiles. This initial application provides information to recommend that both industry and research focus on future models that evolve from combinations of simpler models.

- Automated combinations provide a quick way to baseline buildings that have data available either online in real time or in CSV format. This should be utilized in order to advance the state of the art, which involves one-off baselining of buildings that are not often compared side by side. A good application of this would be to provide an automated, on line platform for building baseline production that can be accessed freely in order to promote open source baselining and model production from simple smart meter data.

- Combination models bring in the strengths of all the models that we used in this study - low error at small forecast horizons, consistent results at longer horizons, all while adding more robustness to prediction - if there is a time that a model is highly variable and imprecise, the combination algorithm allows for other models to dominate and thus produce better results. This is powerful in the building analysis space, where different models need to be used for different purposes. In our case, instead of one model needing to be used to model short time horizon predictions while another is used for longer time scale predictions, we can employ a combination that takes the best parts of both models.

- The results here are consistent with previous work in the forecasting space, namely the M-competition that concluded that combinations of models are typically better than individual complex prediction methods.

Chapter 4

Anomaly Detection

4.1 Introduction

In the Total Quality Revolution of the 1950s, Edward Deming developed and helped implement a statistical quality improvement system in Japans auto industry. Demings work built upon the work of Walter Shewart, who introduced control charts as a statistical approach to improving manufacturing at Bell Labs [8]. Previous work has looked at how to attach metrics to building energy use that are chartable, but not necessarily statistically derived in a process control framework [44]. While building commissioning is sometimes referred to as quality assurance for the purposes of ensuring energy consumption is minimized buildings [36], is at the point where it needs both a mass production method to accompany that Q&A methodology. We present a method of statistical process control (SPC) that offers the basis of these requirements for a commercial deployment of SPC technology on buildings.

Residual based SPC is a technique for monitoring and improving temporally autocorrelated processes in industrial processes [2] and the concept can be extended to deal with time series data processes [1] . For this procedure, at every time point, a processes outcome is predicted by some means that properly accounts for time-series effects, and then compared to the result acquired when the process is measured at that time step. In order to measure the statistical significance of the prediction deviating from the actual measurements, a control chart methodology is applied to this problem. In this first iteration, we will demonstrate the applicability of two different kinds of control chart variants to this problem- a moving average control chart and a rolling outlier control chart. We choose these charts because they display an ability to detect deviance in processes quickly and effectively given prior information. The rolling average control chart is not a control chart per-se, but a variant on the concept of charting outlier in order to detect process deviance.

4.2 Example: Prediction for anomaly detection

We evaluate the use of short term prediction for anomaly detection buildings. As figs. 3.5 and 3.11 show, using shorter time scales for energy consumption prediction in buildings produces lower errors over time. Figure 4.1 shows the error over all prediction horizons for two buildings using the STL method regressed on temperature. The two buildings are described in table 4.1, and have been anonymized to protect data privacy.

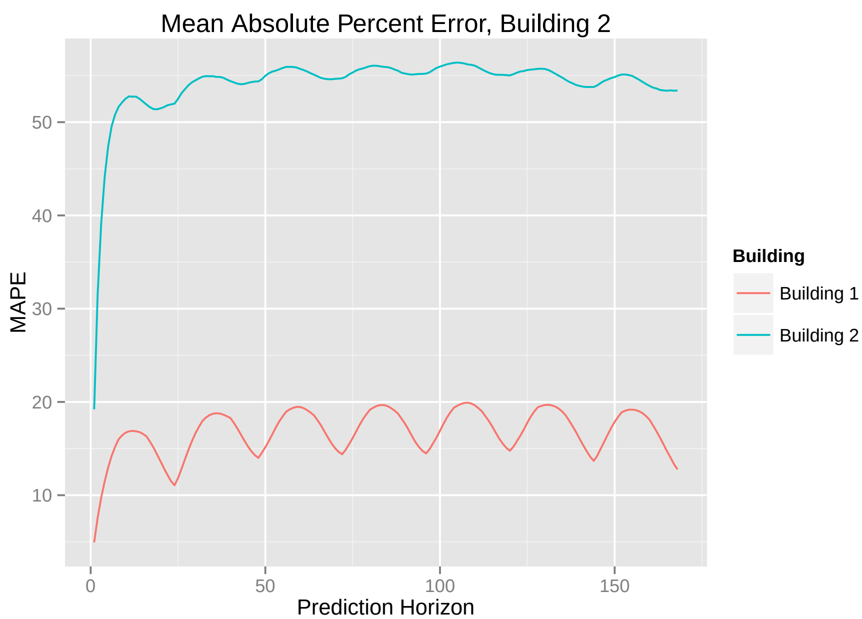


Figure 4.1: MAPE of Seasonal Trend Loess method error function for two buildings using regression on temperature for prediction of residual.

All of the data in this section were analyzed using the R code developed for the previous chapter on prediction with each building being predicted 168 hours ahead and each cross validation step rolling by one-hour at each time step - similar to creffig:cvdiag, but with the overlap between cross-validation folds being one hour each. In this framework, we choose the STL.TEMP model because of the low amount of error experienced at one hour ahead prediction in order to attempt to do error prediction over the short term. The framework embodied here is that we know that

Building Name	Description
Building 1	100,000 sq.ft 1990s vintage office building in California thermal zone 3. This building is a high tech sector 3-story office building with significant lab loads. The building is primarily a VAV Reheat system, with additional cooling via fan coils serving high load areas.
Building 2	200,000 sq.ft 1980s vintage office building near Chicago, IL. The 12-story all-electric office building is mostly tenant occupied office space with a detached non-enclosed three-level parking garage.

Table 4.1: Two buildings were used in this study: The buildings are anonymized for privacy purposes

the model performs to a certain specification an hour ahead, so if it is outside that specification, either the model or the building is malfunctioning, and we can use statistics to determine which it is to some degree of precision.

4.3 Control charts: an application to anomaly detection

It is a necessary but not sufficient condition for a building to be predictable in order for it to be well tuned. This is in line with the literature on statistical process control for time series methods [1]. In order to test this theory, a residual based control chart technique was employed. Following this predictive method, each hour of prediction is subtracted from the actual power use to yield the residual power use, or error in prediction. This error term denotes how far off at each hour our prediction is from the actual behavior of the building. This information of error is then fed into our control chart methodology, which allows us to diagnose buildings at an accelerated rate. In this case, we use three control charts - a moving average (MA) control chart, a rolling outlier (RO) control chart, and a dual measure control chart - combining the metrics of the MA and RO charts. The moving average control (figure of merit designated as: Mt), assumes that at the start of the process, the mean of the process should be stationary and computes the mean of the prediction process residual (actual minus predicted) as a two week rolling average - This is updated as the process moves along, and each point in the control chart is then developed using a two-week (336 hour) rolling window. In SPC, upper and lower control limits (UCL and LCL, respectively) are designated to specify when a process is out of control - for the MA chart, the control limits that we use are:

$$UCL/LCL = Process\,Mean +/- \frac{3\sigma}{\sqrt{n*w}} \tag{4.1}$$

Where σ is the standard deviation of the process up to time within window w, n is the number of samples averaged at each time point, and w is the total number of samples in the rolling window. For the purposes of this examination, we used $n = 4$ and $w = 336$, with measurements in hours. The RO chart is constructed in a manner that allows us to see which points are hard to predict on a weekly basis. For every hour in the data set after the first week, the 99% percentile of residuals in the preceding week is calculated. Following this calculation, a point is flagged as hard to predict if its residual is above the 99% percentile for the previous week, giving us the hardest points to predict in each weeks rolling window. In short, the MA chart shows us when the mean of the prediction process shifts, and the RO chart shows us when it produces extreme outlier. We view the construction of control charts as a convenient method for data mining. The two charts previously constructed are therefore combined to allow for a vote on when the building is malfunctioning. The combined chart is a dual measure (DM) chart, which tells us when the prediction process is producing outliers at the same time that the prediction process mean shifts. This methodology could be extended to any number of possible rule sets, but we only demonstrate this one here as a means of data mining the prediction process.

In figs. 4.2 and 4.3, the moving average control chart and the rolling average outlier detection is demonstrated for Building 1. As can be seen from the graphs below, the algorithms produce significant flags on data points over the time period of prediction. In this particular building, a chiller replacement occurred in November of 2013, and the commissioning of that device appears as a drop in the power use, demonstrated in fig. 4.4, and occurs as a "cone" of energy prediction residuals that can be observed in figs. 4.2 and 4.3. As the new operation of the building becomes more predictable, the residuals of power prediction shrink, and the ability to predict this building becomes greater.

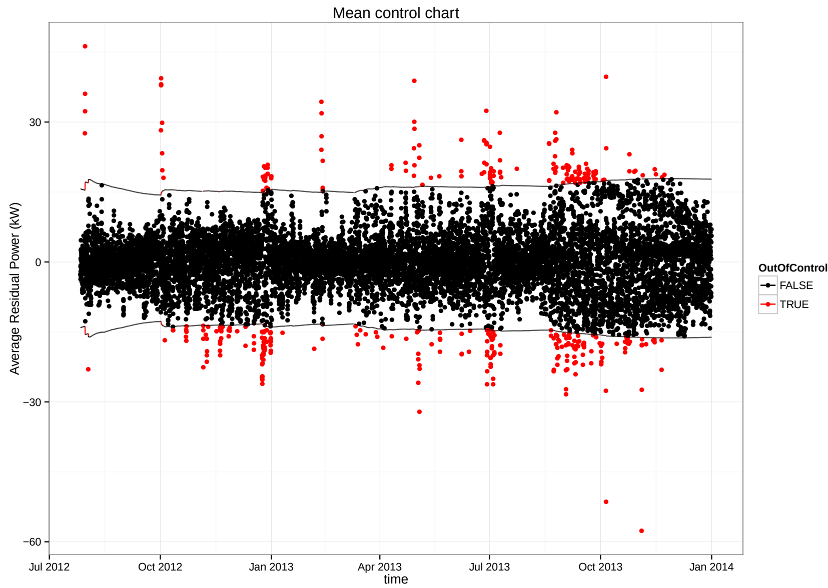


Figure 4.2: Moving average Control Chart for Building 1

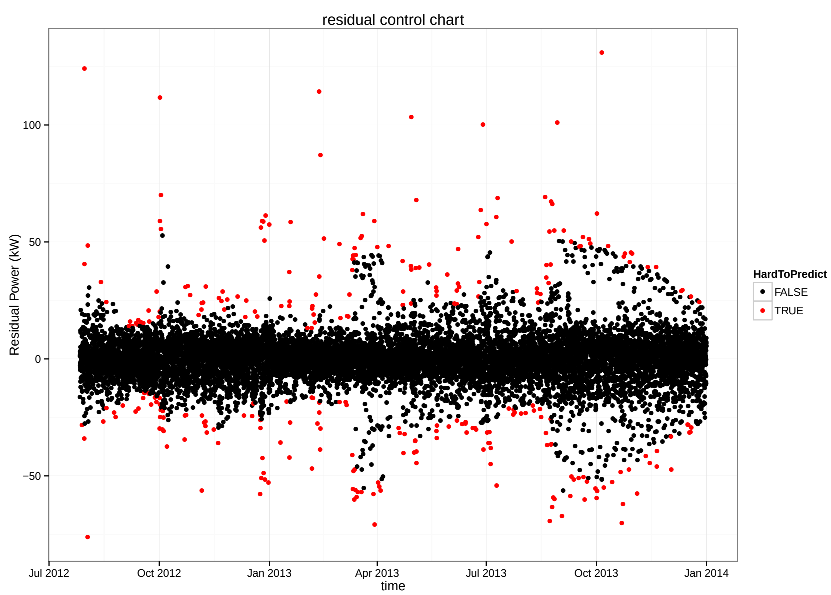


Figure 4.3: Rolling Average Control Chart for Building 1

Observing fig. 4.4, we can see that there are a number of flags in the building - when these flags were inspected, not all of them led to any significant errors in building performance, and many of them yielded no results at all. While this is to be expected of an untuned algorithm, at least one result that was picked up was of interest and will be reviewed in fig. 4.9, showing a chiller that was cycling over night being detected when root cause analysis was done on one of the data points.

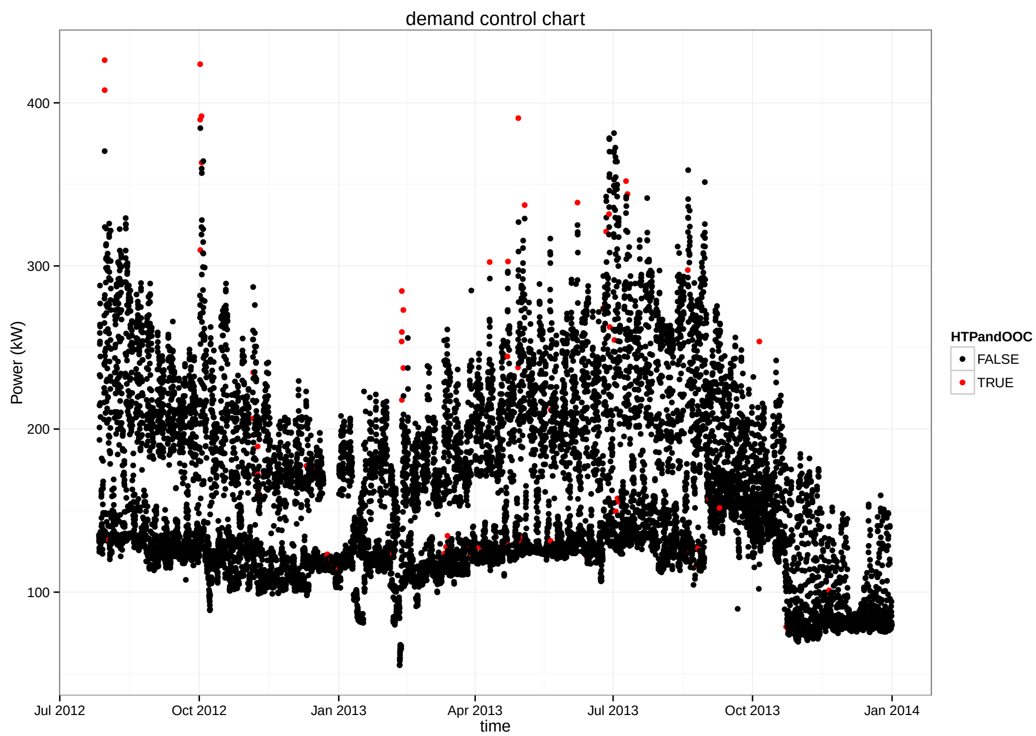


Figure 4.4: Combined measure control chart projected onto actual building energy use

Evaluating this method of anomaly detection on a building that is not as easy to predict is also an interesting exercise. In building 2, the MAPE at one hour of prediction is 19%, yielding a much wider control limit range than in the previous building - this is exemplified in figs. 4.5 and 4.6, where the hour control limits hover at approximately 16% of the building's total energy use. While this is the case, the voting algorithm only flagged 13 points out of the total of 8661. The point evaluated in fig. 4.10 represented the building not shutting down correctly overnight

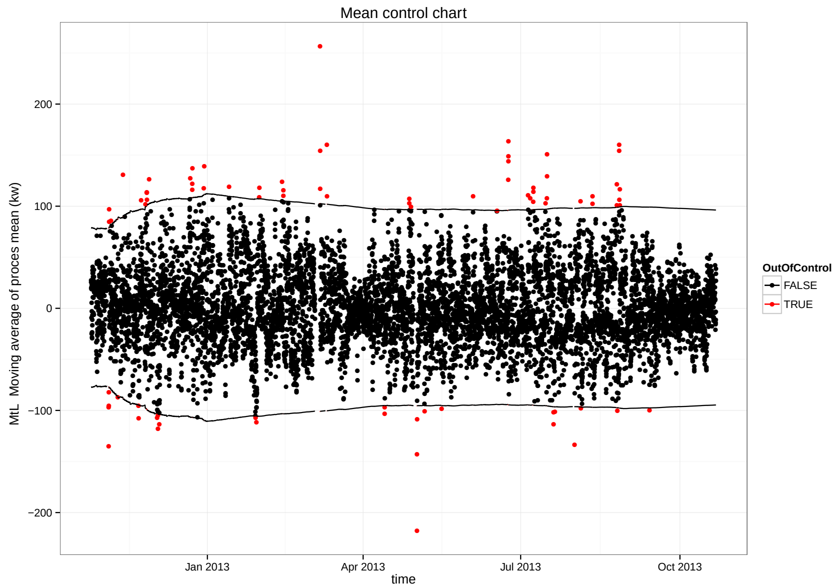


Figure 4.5: Moving average Control Chart for Building 2

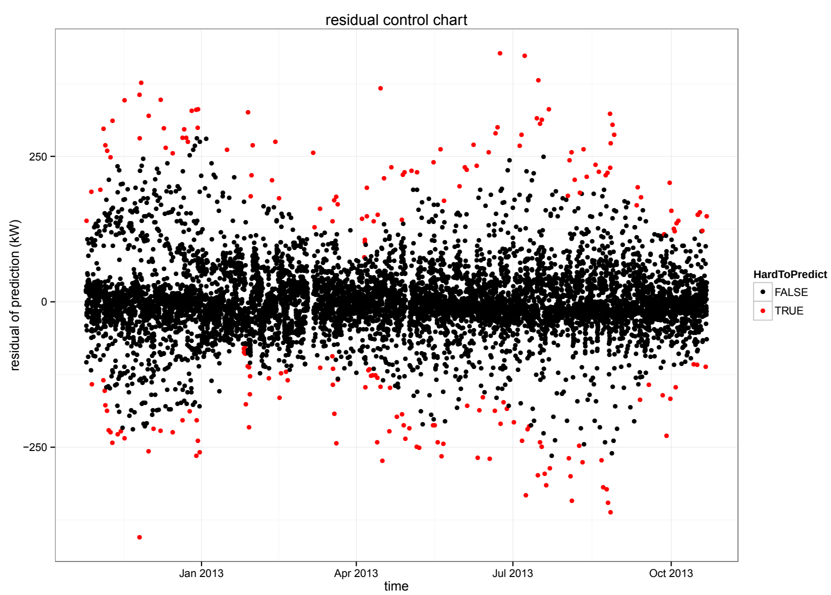


Figure 4.6: Rolling Average Control Chart for Building 2

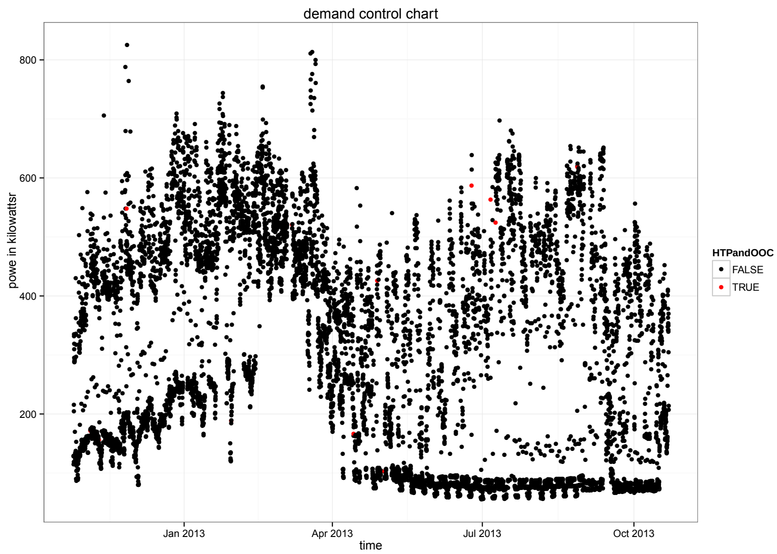


Figure 4.7: Combined measure control chart projected onto actual building energy use

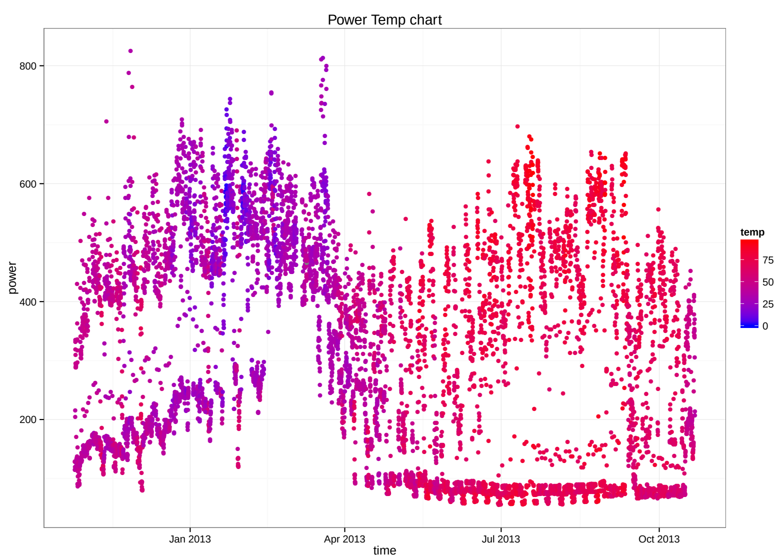


Figure 4.8: Temperature vs power use in building 2

For the purposes of further analysis of power use efficacy in building 2, the power use is plotted against a heat map of temperature in fig. 4.8. As we dig down into the causes of energy use consumption in buildings, it will be helpful to keep this figure in mind.

4.4 Example Root Cause Analysis

An important tool of statistical process control methodologies is root cause analysis - evaluating what causes data points to deviate into the zone where they get flagged by the algorithm. In order to do so in this context, the analysis of the flagged points was handed off to a building engineer familiar with the project to evaluate the SPC charts. In fig. 4.9, we see cyclical behavior of the building's energy use picked up by the algorithm, something that would normally take the building engineer many hours of sifting through data to find. In the case of Building 2, there was a point

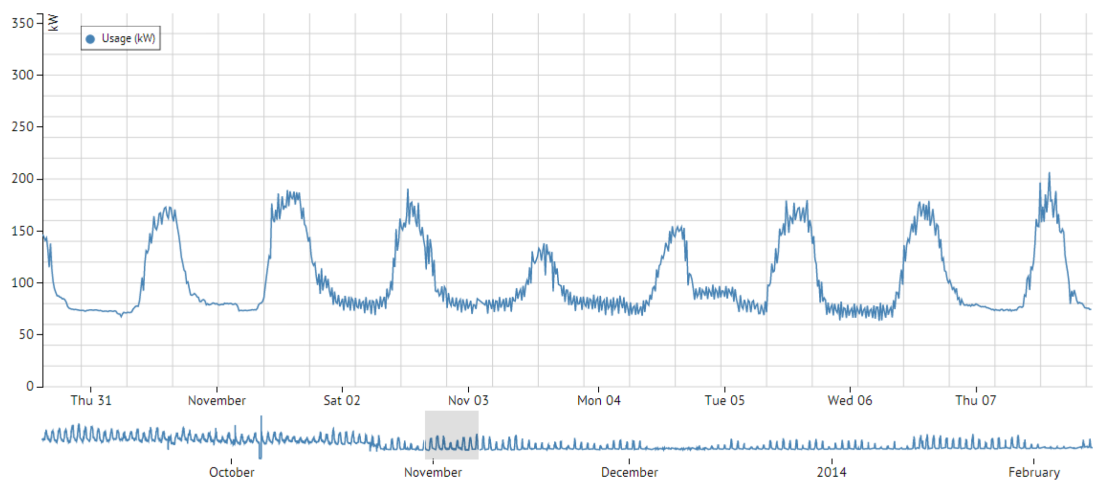


Figure 4.9: Overnight Cycling picked up by the algorithm - This occurred during the commissioning phase of Building 1

where the building operations sequence was faulty and caused the building to not shut down properly at night - leading to the pattern shown in fig. 4.10. Once again the algorithm picked up this problem with no input from the engineer a priori, and it was something that they had not seen in advance, allowing them to detect a problem that had been missed before.

Figure 4.10: The algorithm picked up this point in the power usage of Building 2 that showed the building not shutting down correctly at night.

4.5 Implications and Conclusions

It is possible to determine when some energy efficiency flaws are happening in a building by using statistical process control methods. Using these early stage methods, we have managed to find errors in a building that would take an engineer many hours to pull out by hand. We have also demonstrated a method for determining if there are scheduling faults in a building by simply running a predictive algorithm on whole building meter data. This is promising and points to a good future for the development of these algorithms. This is just the start of predictive analytics based commissioning.

There were more alert flags than errors that could be ascertained. The system could benefit from a rule based expert systems approach to eliminating possible false flags, and this can be added in as additional buildings are analyzed. While these present technical challenges, the overall analysis of these methods indicates that they are sound and can be examined for deployment at scale. This is a promising new direction for building commissioning as an industry and represents the power of bringing new computational methods to bear on a problem that has been studied in as a manual procedure in the past.

There is additional work to be done in this area by incorporating more types of control charts as well as charted building performance metrics - by evaluating existing and known methods of diagnosing buildings from a chartable framework, we will attempt to uncover new methods of rapid building diagnosis. This will form the basis of work for the next two chapters of this work.

Chapter 5

An exploration of additional parameters and metrics in SPC based anomaly detection

5.1 Introduction

In the evaluation of energy efficiency in building, (retro) commissioning is sometimes referred to as the quality control part of making buildings work. Previous work has been done on graphically analyzing this process - [44] goes into detail about how building data is to be used in order to graphically represent the amount of energy use in buildings for retrocomissioning work. [6] shows how predicting energy usage in buildings can be used to identify outlier points and start to do analysis on this using heating degree days (HDD), while [57] shows how these analysis can be converted into a very simple control chart for the purposes of energy management. CUSUM charts have been used in multiple contexts in order to detect problems in Variable Airflow Volume HVAC systems and other building systems [29, 48]. Different charting metrics are presented that can be used to visualize problems in energy usage in buildings, with a focus on the presentation of the data and how to extract information from it.

5.2 Transforming control charts: Looking for scheduling faults

It is possible to transform the traditional control methodology into a more useful form for looking at building operation issues. In order to do this, we introduce two new types of charts: the daily mean control chart (DMCC) and the daily residual control chart (DRCC). Because control charts are essentially a heuristic technique, it

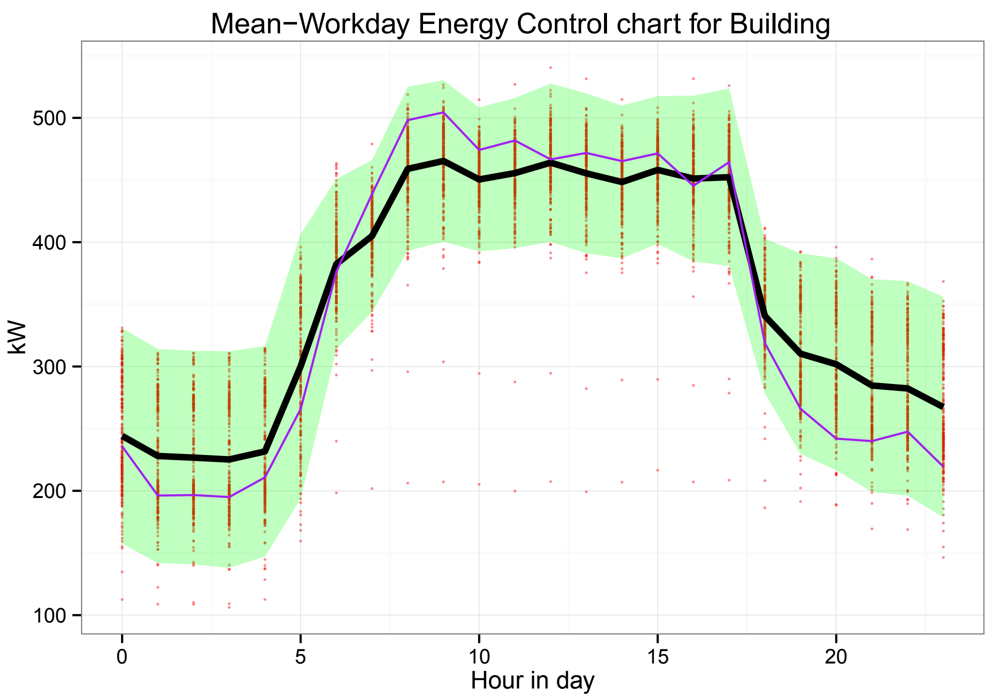


Figure 5.1: A Daily Mean Control Chart demonstrating poor scheduling on an example building. From this chart we can see that, on an average work-day, the building varies more in off-hours than in on-hours

is worth remembering that they should capture the sources of variance in whatever process is being diagnosed. In the case of a building, that process is often the actual scheduling of the building, along with factors such as occupancy, weather, and the history of maintenance in the building. Because of these time of day and weather effects, it is once again worth looking at the residuals of prediction for the building, but in this case from an "average day perspective." This is presented below in fig. 5.2.

By doing this, we can determine when the building is predictable and present this information to a building engineer who will be able to correctly analyze this data.

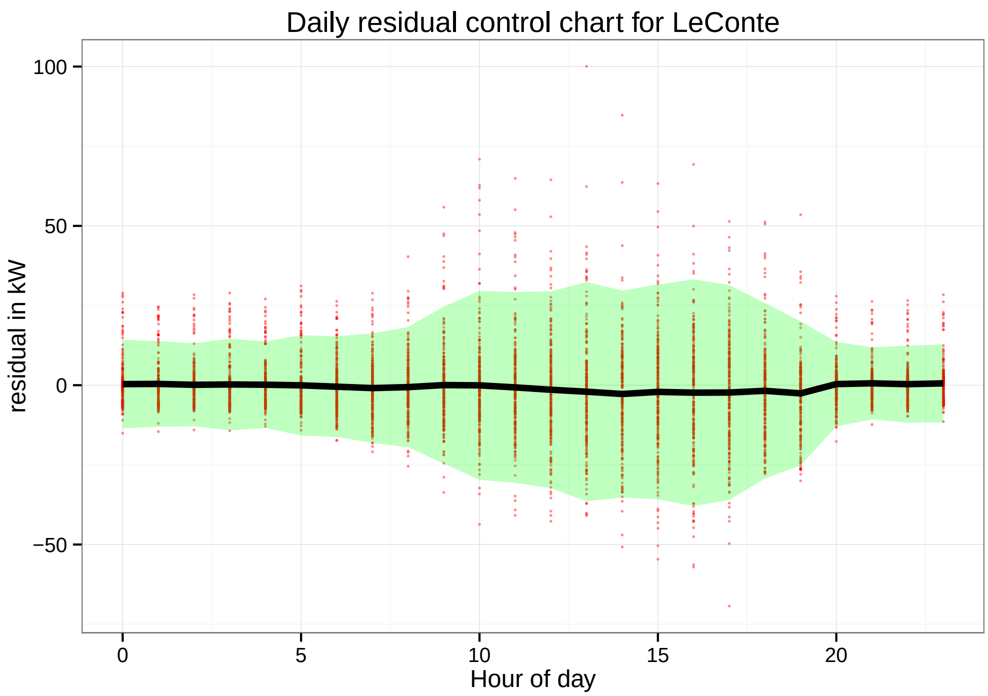


Figure 5.2: A Daily Mean Control Chart demonstrating poor scheduling on an example building. From this chart we can see that, on an average work-day, the building varies more in off-hours than in on-hours

5.3 Variation of control - Daily differences

In order to further extend the analogy between traditional control charts and daily load charts, we develop a weekly control chart, which can be functionalized to produce actionable information. In this case, we demonstrate a weekly control chart that has upper and lower specification limits that are the lowest average usage weekly at every point during the day. Figures 5.3 and 5.4 show two examples of this in example buildings from Chicago, both with different scheduling problems which will be automatically extracted in the next chapter.

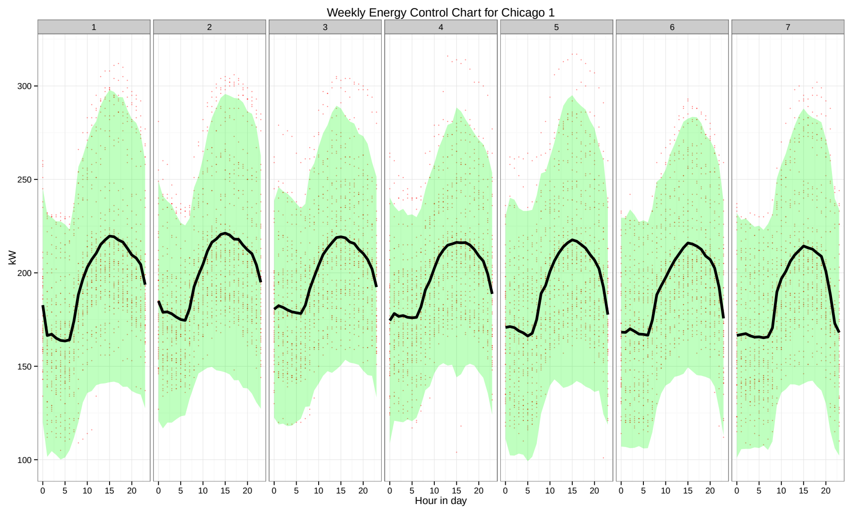


Figure 5.3: A weekly demand control chart for a building in Chicago, with day 1 being Monday and day 7 being Sunday.

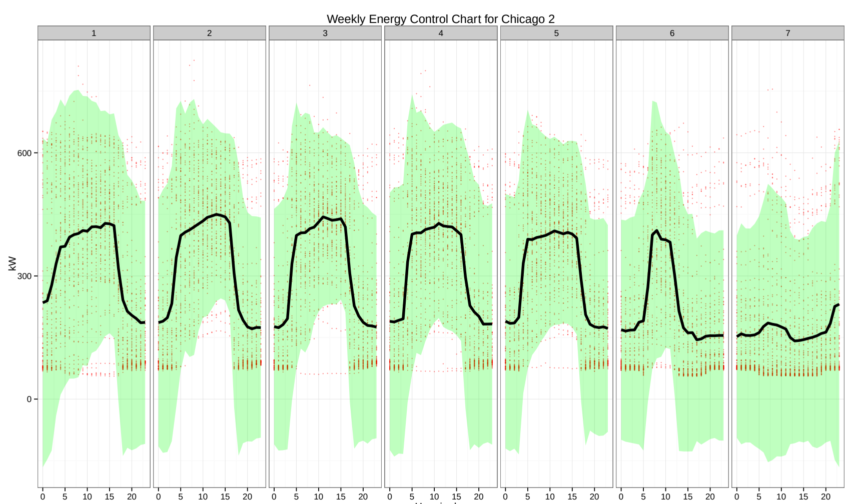


Figure 5.4: A weekly demand control chart for a building in Chicago, with day 1 being Monday and day 7 being Sunday.

As the logical next step in using modified control charts to examine the data, we look once again at a weather normalized prediction chart for daily power use. In this case, we have broken out the chart into Monday through sunday again, and examine it for variability. In this case, we notice that the days have different levels of predictability, and we will derive control limits off of the most predictable day, and examine all points that are less predictable than that.

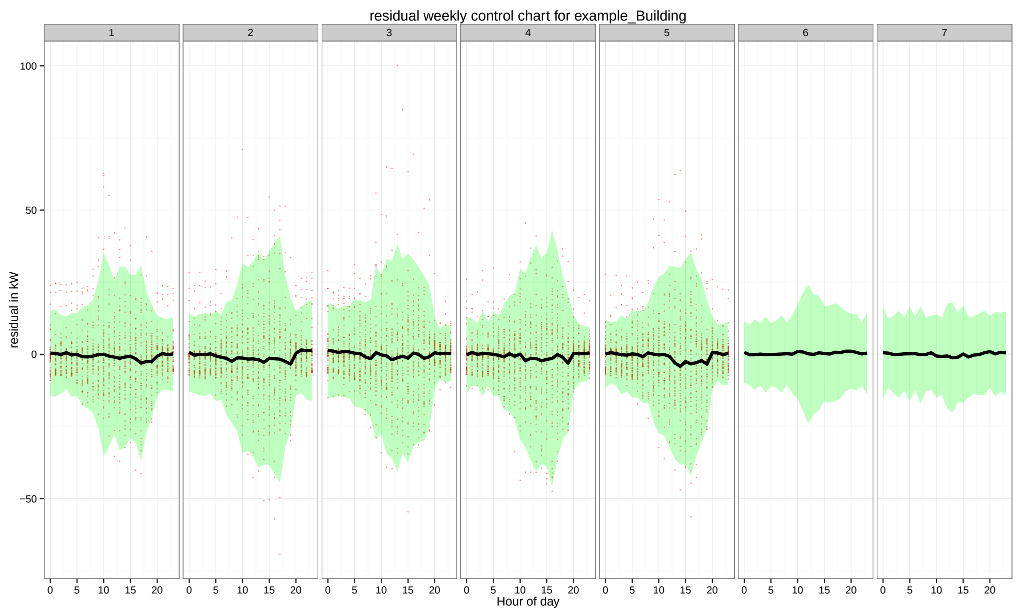


Figure 5.5: A weekly residual control chart for a building in Chicago, with day 1 being Monday and day 7 being Sunday.

5.4 Load Duration Curves: A snapshot of building condition

If we recall from fig. 1.12, the grid can essentially be looked at as a probability distribution of how often certain magnitudes of power draw occur due to loads in the grid such as buildings - this is known as the Load Duration Curve (LDC). If we consider this, then we can also picture each subsystem on the grid to have a LDC, such as that of a building. As an example, the LDC of LeConte Hall on UC Berkeley's campus is given in fig. 5.6 for a period of 13,299 hours, or 544.125 days.

This curve represents the average behavior of the building, yet only gives us this information as a snapshot of overall information.

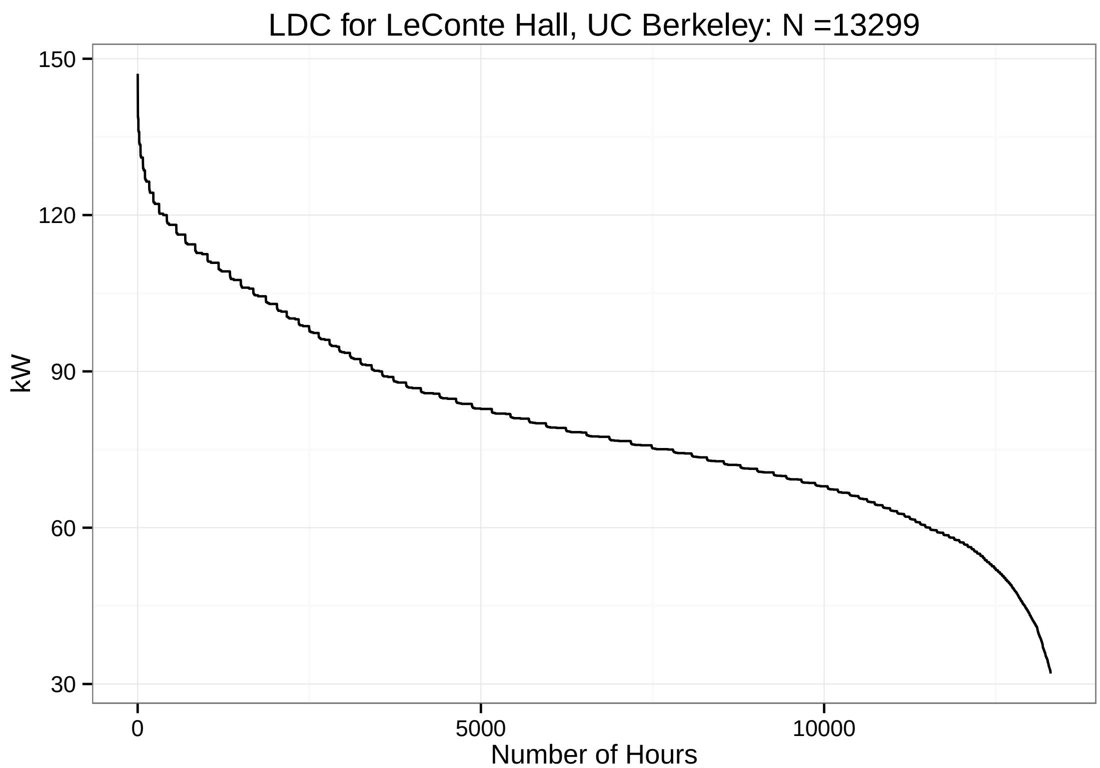


Figure 5.6: Load Duration Curve For LeConte Hall on UC Berkeley Campus

This curve has several pieces of information in it, but typically is only calculated for a whole year or more worth of data. In order to analyze the data at a more granular level, we need to extract specific features and parameters from the LDC in order for it to be useful.For this procedure, we draw on the work from [43], where the LDC is modeled as a mathematical function that can be fit as a function of certain parameters. The equation we use for the LDC is:

$$P = 1 - aT - bT^c + \frac{d}{1 + e^{f(T-g)}} - \frac{d}{1 + e^{-fg}} \tag{5.1}$$

We can see how each of these parameters is assigned when we fit them to a load duration curve shown below. table 5.1 shows the parameter values for $a, b, c, d, f \& g$, while fig. 5.7 shows the actual curve, annotated with the features that the parameters affect on the plots.

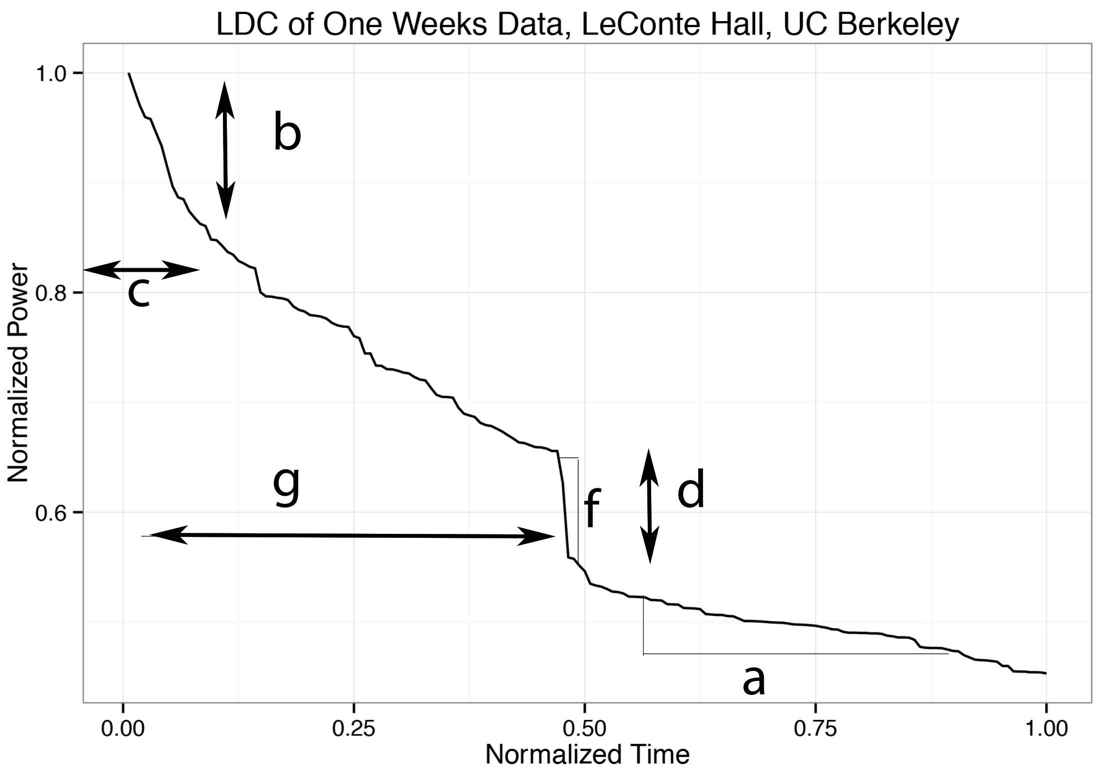


Figure 5.7: Load Duration Curve For one week in Leconte's power use

These variables can be specifically interpreted to indicate certain things about a building's behavior during a given week. Here is a little more description about each:

- a- Off-Schedule Slope - This is the slope of the "off-schedule" period of the load duration curve. This indicates the amount that the base load varies during the week. This number should be as small as possible - it is lower bounded at .05.

- b- Peak Height - The Peak height indicates the maximum power draw that is achieved during the week. This height is correlated with the temperatures experienced during the week.

- c- Peak Depth - The Peak Depth indicates how much of the week was spent in the "peak period" - that portion of time when the building uses the most power. This is typically correlated with the duration of extreme temperatures during the week.

- d- Schedule Shift - The Schedule Shift indicates how much the building's power consumption shifts between "scheduled" and "unscheduled" hours - this is indicative of how well the buildings major power consuming systems are programmed to turn off when the building is not in use.

- f- Shift Slope - The Shift Slope indicates how distinguishable the difference between schedule and unscheduled hours is. A large shift slope indicates that the building shuts down rapidly and effectively, while a longer one means that it is slow and more ineffective.

- g- Schedule Depth - The Schedule Depth indicates how much of the time the building is "On." This is useful for determining if the building has a scheduling problem - this number should not exceed .5 due to the fact that the building should be "on" less than 50% of the time.

Coefficient	Description	Value in Example
a	Off-Schedule Slope	.05
b	Peak Height	.62
c	Peak Depth	.23
d	Schedule Shift	.11
f	Shift Slope	554.34
g	Schedule Depth	.48

Table 5.1: Table displaying fitted coefficients for the example LDC given above

This parameterization lends itself naturally to becoming a control chart - that is, variability in these parameters indicates variability in the building. We can further introduce definition into our charting of these parameters by observing the difference between the mean-week behavior of the building (the behavior at every week, averaged over all weeks), and the overall building behavior - defined by fitting the LDC curve to the whole building power usage over all times observed.

Much like we did before with EUI and LF, we can compute metrics on the expected values and the observed values, and construct control charts based upon the residuals of the computed parameterization, and derive conclusions based upon this. This will be useful for constructing rule sets in order to do fault detection in the following chapter's data mining exercise. For the purposes of illustration, we present the six parameters as a control chart of LeConte hall on UC Berkeley campus in fig. 5.8, and the residuals of fit versus predicted LDC fit in fig. 5.9. This fit is done with the LBNL model, and thus can be considered a reflection of weather normalized, long established behavior in the building.

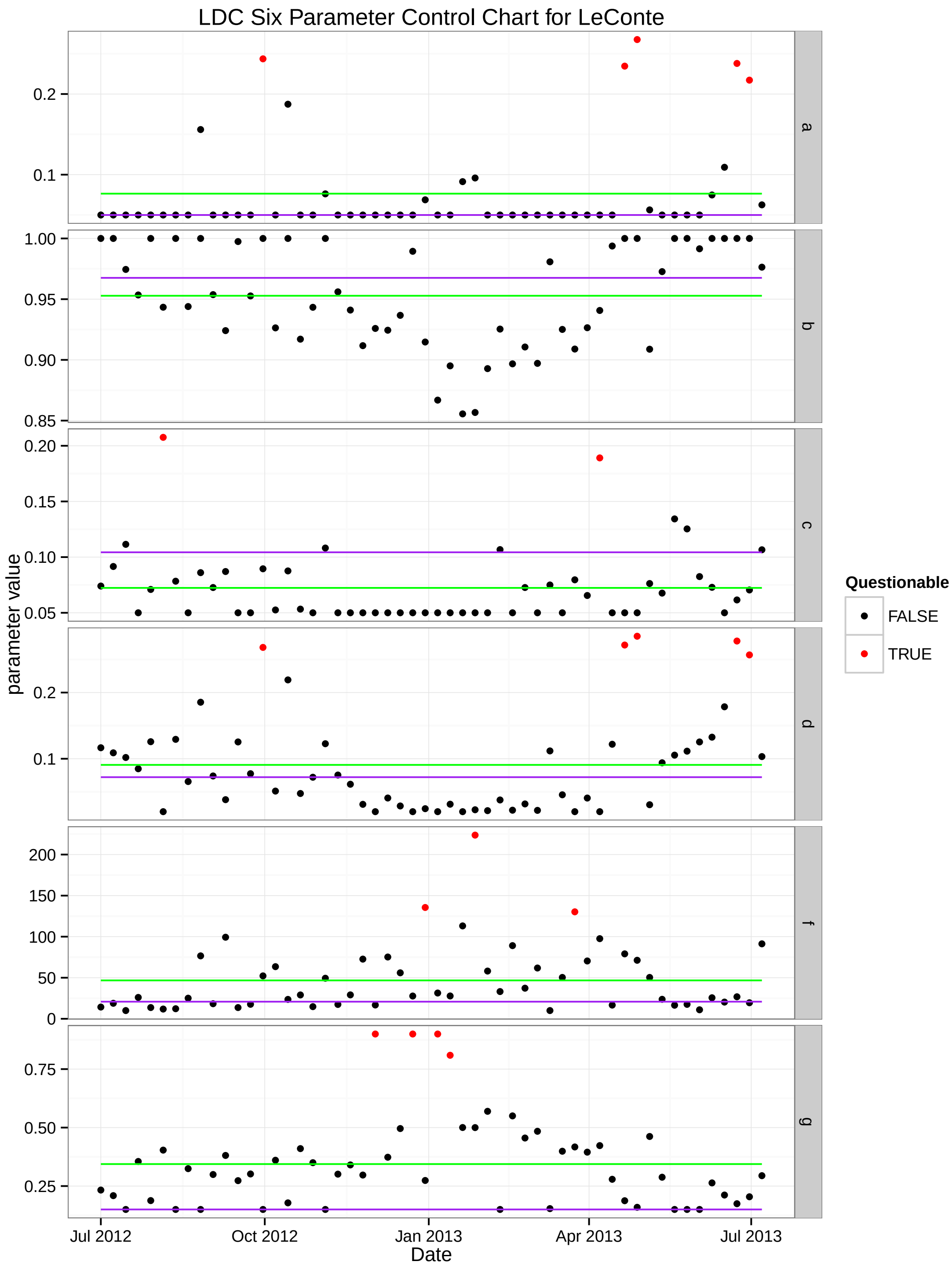


Figure 5.8: Six Parameter control chart based on each of the parameters outlined for a load duration curve. The purple lines represent the fit of the LDC over the entire timeframe, while the green lines represent the average fit over all the weeks computed.

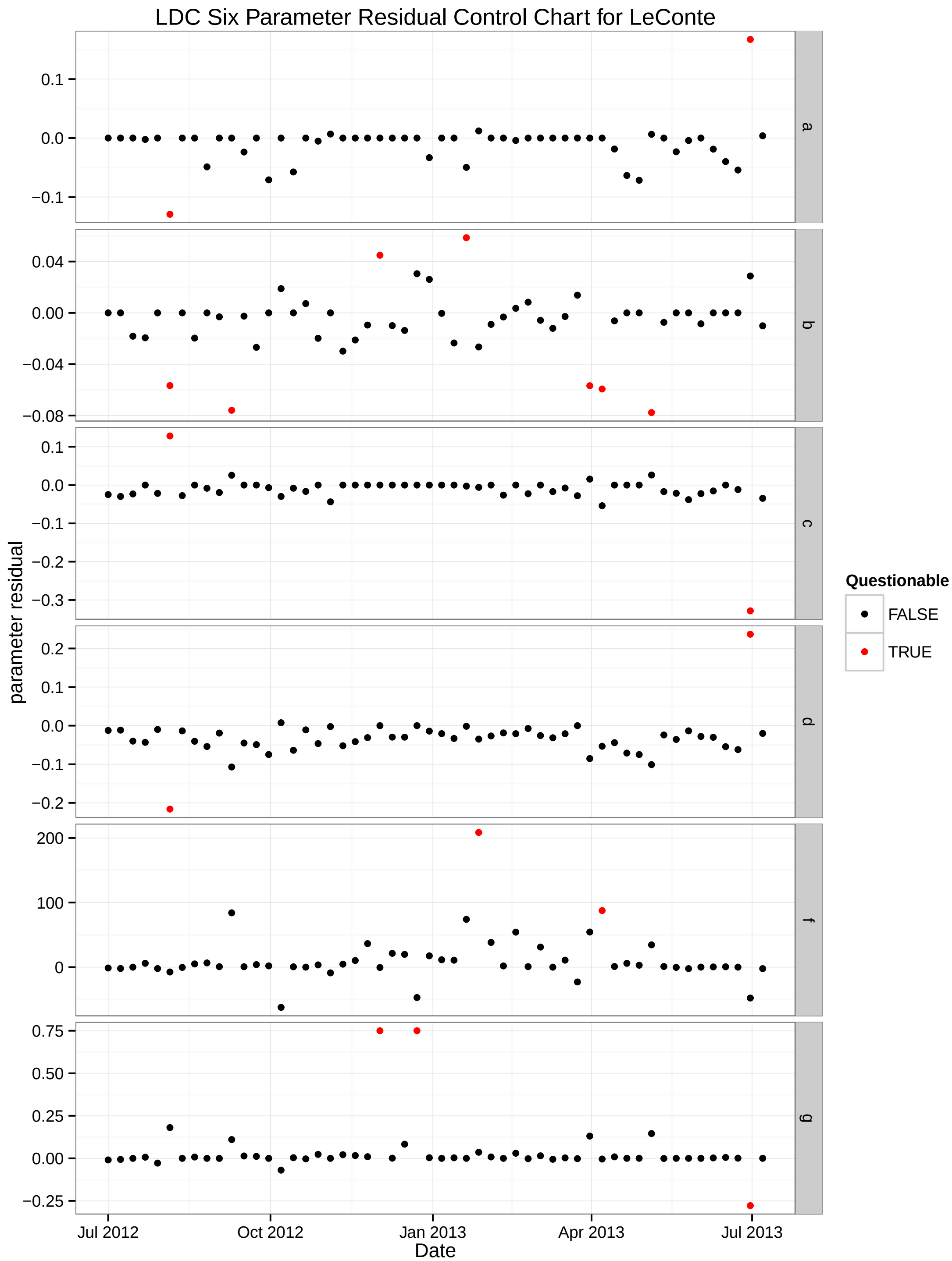


Figure 5.9: Six Parameter control chart based on the residuals parameters fit to the actual load duration curve minus the parameters fit to the expected.

5.5 Diagnostic Results

Results of the six parameter control chart can be used to to further diagnose the state of a building through exploratory analysis of it's behavior and predictability. Two weeks, 40 and 43 are chosen for further analysis due to receiving 3 and 4 flags respectively.

week	variable	value	mean	SD	Questionable
40	a	0.33	0.14	0.08	TRUE
40	b	0.34	0.54	0.10	TRUE
40	c	0.05	0.10	0.07	FALSE
40	d	0.04	0.10	0.05	FALSE
40	f	82.24	32.70	25.10	TRUE
40	g	0.23	0.38	0.28	FALSE

Table 5.2: BSPC Parameters for Week 40 for Leconte

variable	weekno	value	mean	SD	Questionable
a	40	0.05	0.08	0.09	FALSE
b	40	1.00	0.97	0.04	FALSE
c	40	0.04	0.08	0.06	FALSE
d	40	0.15	0.13	0.09	FALSE
f	40	14.27	47.40	134.49	FALSE
g	40	0.25	0.35	0.21	FALSE

Table 5.3: Residual BSPC Parameters for Week 40 for Leconte

weekno	variable	value	mean	SD	Questionable
43	a	0.33	0.14	0.08	TRUE
43	b	0.35	0.54	0.10	FALSE
43	c	0.05	0.10	0.07	FALSE
43	d	0.10	0.10	0.05	FALSE
43	f	147.57	32.70	25.10	TRUE
43	g	0.23	0.38	0.28	FALSE

Table 5.4: BSPC Parameters for Week 43 for Leconte

Table 5.2 and **??** show the six parameter SPC values for week 40 and their residuals, respectively - table 5.4 and **??** do the same for week 43. The algorithm

variable	weekno	value	mean	SD	Questionable
a	43	0.34	0.08	0.09	TRUE
b	43	1.00	0.97	0.04	FALSE
c	43	0.02	0.08	0.06	FALSE
d	43	0.39	0.13	0.09	TRUE
f	43	59.92	47.40	134.49	FALSE
g	43	0.19	0.35	0.21	FALSE

Table 5.5: Residual BSPC Parameters for Week 43 for Leconte

flags a,b and f in week 40, and a & f for week 43, along with the residuals of a & d. From these flags, we should expect unusual off-schedule behavior in week 40, specifically relating to how fast or completely the building turns off, as well as unusual peak-to-base ratio. Week 43 should display unusual scheduling with specific problems of predicting incomplete shutdown.

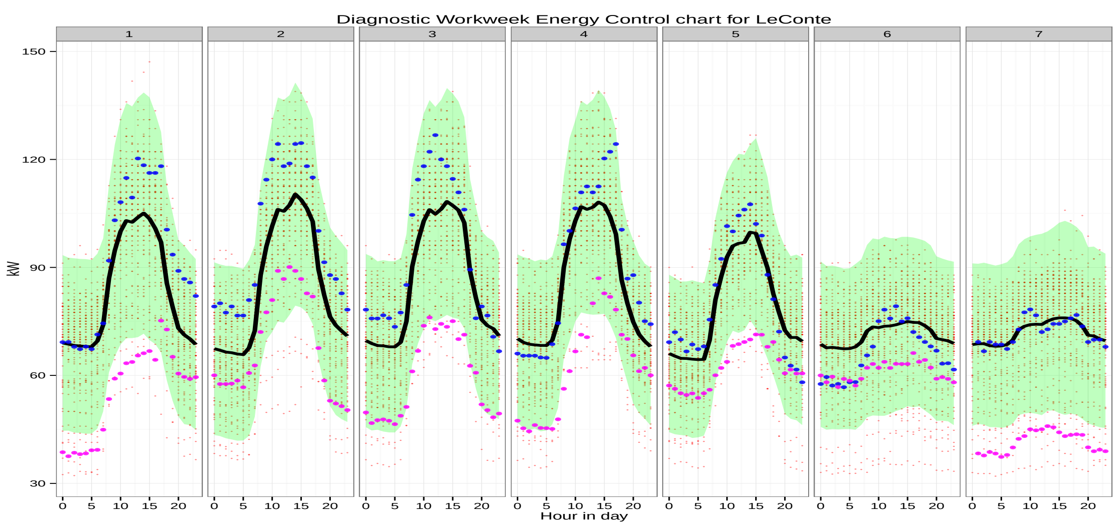


Figure 5.10: Diagnostic chart for LeConte Hall, with week 40 presented in magenta and week 43 presented in blue.

Figure 5.10 highlights these weeks, showing unusual scheduling for week 40, highlighted in magenta, and week 43, highlighted in blue. The issues predicted by the algorithm for week 40 appear naturally. Observing fig. 5.11, the problems with week 43 are not as apparent. Further diagnostic material can be obtained from fig. 5.12, showing that the building is hard to predict in off hours for much of the week.

The power trace in fig. 5.11 also shows this problem upon closer examination - the building does not return to a steady low state during this week, indicating improper and possibly energy wasting shutdown of the building during off hours. The astute observer should also notice that weeks 37 and 38 in this example also display behavior that is not consistent, and were flagged for inconstant shutdown behavior and unusual peak-to-base ratio, respectively.

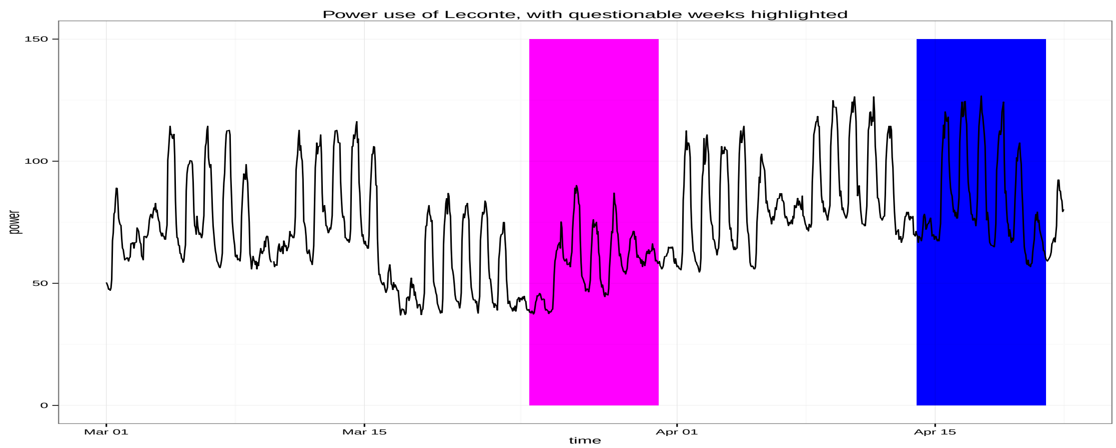


Figure 5.11: Power use in LeConte hall, with previously flagged weeks highlighted in colors matching the above chart.

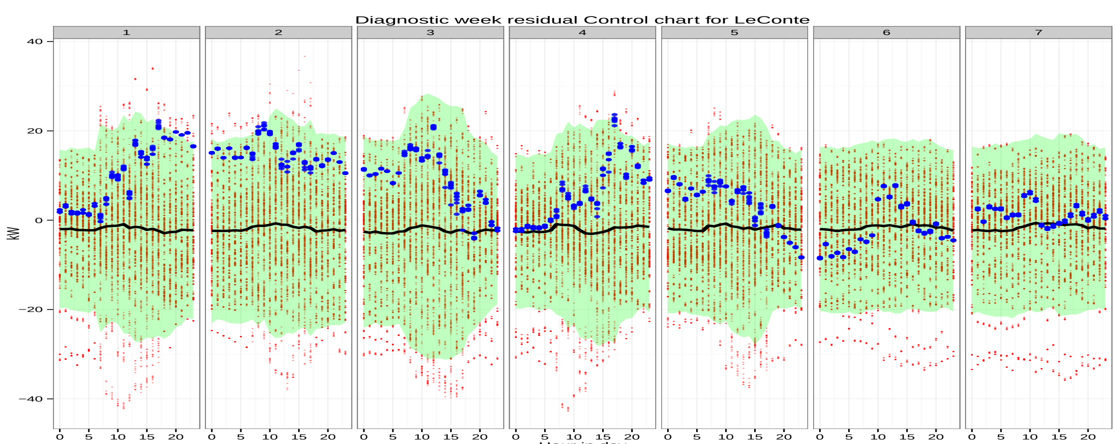


Figure 5.12: Diagnostic Weekly Residual Control Chart, Leconte hall showing unpredictable behavior in off hours.

5.6 Conclusions and implications

Treating building metrics as control charts is a powerful visualization tool which can be used to find points in time when the building controls require further examination. This is incredibly useful, but needs to be paired with an ability to rapidly screen buildings to determine which days they are underperforming on and what the conditions of those days are. These screening processes will depend on rule setsIn order to do so, several conditions must be met:

- Rule sets must be constructed according to the qualities of each model: If we with to evaluate how the building has performed over the long term, such as in a common cause chart, then we need to use a model that normalizes for weather and occupancy in a reasonable manner. Such models are the LBNL regression, day time temperature, and in some cases the mean-week model.

- As a corollary to the previous point, if we wish to evaluate ”special causes” - the sources of error that are from one time faults in the building, then we wish to use models that are extremely good at predicting in a short time span ahead based on adaptation from recent points of interest for the buildings. Models that are useful for this are any of the time-series models, including STL, ARIMA, or any of the combination models that include these.

- Visualizations must be analyzed by professionals in order to effectively tease out the important information contained within them. In order to do this, I am collaborating with building engineers familiar with the buildings in question in order to determine what parameters mean in relation to the the behavior of the buildings.

The next chapter will conclude by data mining a large set of buildings based upon rule sets generated from these visualizations. Actionable information that can be used to inform building engineers in fixing buildings, will be presented that in an intuitive format. This information will then be used to build a prioritized list of possible actions to take on a set of buildings on the UC Berkeley campus.

Chapter 6

Applying data mining techniques to SPC based anomaly detection in buildings

6.1 Introduction

Statistical Process Control is, at the end of the day, an intelligent method for mining the data of processes so that they can be brought in line before and after they go out of control. SPC is often applied to boost manufacturing yield but also to reduce waste or re-work in factories [8]. In recent times, there have been a number of of attempts to apply the methods of SPC to work in buildings, including IBM's smatter cities initiative [25], and various papers by the same research group [27, 28]. Control charts, a key tool of SPC, have also been used to detect errors in HVAC operations, including in variable air-flow volume boxes [48] and diagnosing problems in these systems overall [58], and as early as the 1990s for the purposes of testing performance [15, 16].

While the aforementioned methods of using SPC to detect anomalies in buildings apply to either subsystems of buildings or to whole buildings metered at intervals of one month, there is no mention in the literature of using SPC methods including prediction and statistical data analysis in order to mine the data of whole buildings on the hourly level, as I have demonstrated in Chapter 4, or on the weekly basis using LDC fit and other chartable metrics as was proposed in chapter 5. I will further demonstrate the practicality of flagging hourly data that is outside of the normal bounds for the purposes of data mining using mean-workday and mean-week control charts as demonstrated in figs. 5.1 to 5.5

6.2 data set

The data presented in this analysis comes from 51 whole building and substation meters on the UC Berkeley campus, many of which were presented in the analysis in chapters 2 and 3. Specifically, a combination method of NNLS + CI of prediction will be used for any control charts that need to be residual based, while the same range of data will be fitted to meter data as was generated in predictions - in this way some data on the measured buildings is lost at the expense of homogenizing results with the predicted values.

6.3 LDC Analysis

For this analysis, all the buildings on the UC Berkeley campus were analyzed with the data available for those meters between the dates of July 4th, 2010 and July 7th, 2013. For each week available, starting on the first Sunday in the data set, an LDC was computed to fit parameters $a, b, c, d, f \& g$ as described in eq. (5.1).

Scheduling

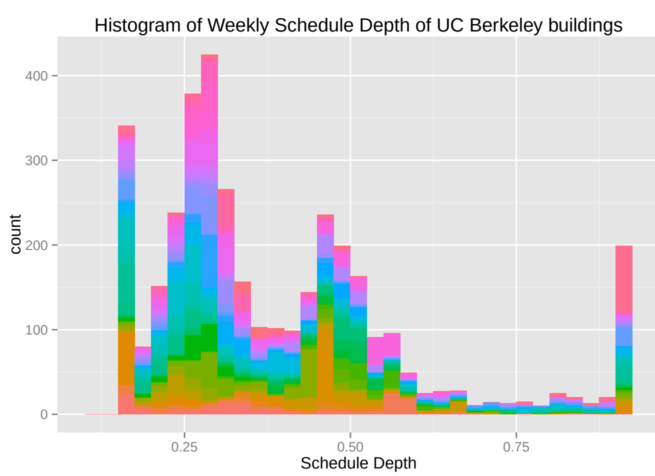


Figure 6.1: Schedule Depth for all meters in the dataset, UC Berkeley campus. Each building meter is represented by a different color. It should be noted that there are two distinct peaks in the scheduling parameters for buildings.

Parameter g in the LDC fit refers to the "schedule depth" that occurs during the week of the LDC fit for a particular building. In the case of the UC Berkeley campus, the schedule depth of buildings on campus is presented in fig. 6.1, with each building being represented by a different color. As one can see, there are two distinct peaks for the campus with respect to the weekly schedule depth, and I can use this classify the buildings involved as well as to find outliers to focus on in order to bring the buildings more in line. Figure 6.2 shows the average of the schedule depths for all the meters in our sample. The meters have been clustered to show the two peaks previously identified visually in the histogram - clustering in this case occurs by using the k-means clustering algorithm which is built into the R stats package.

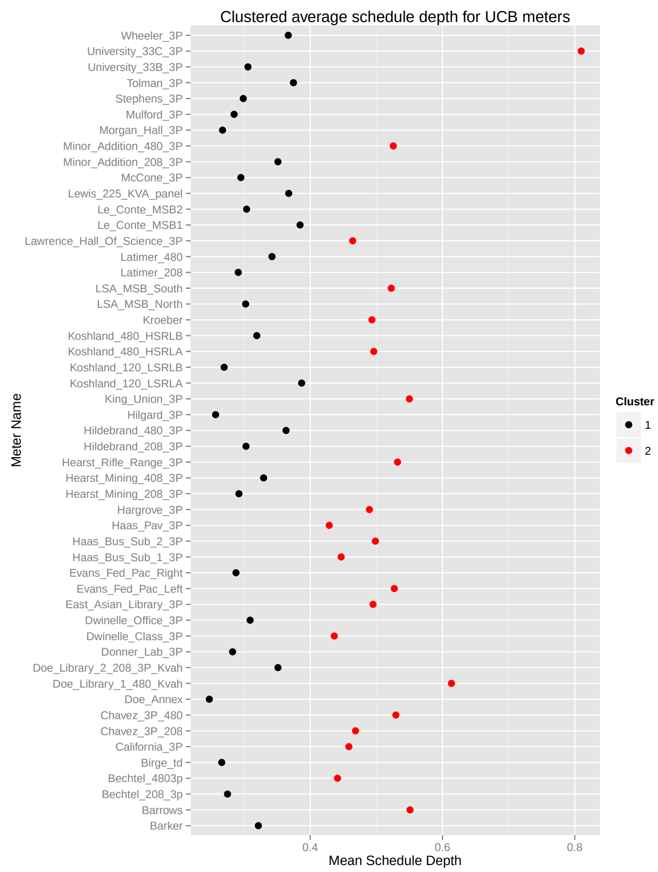


Figure 6.2: Mean schedule depth for meters in our data set. This group was divided into two cluster that represent two different operational modes, corresponding to the previous two peaks in fig. 6.1.

From this figure, we can determine that meters on the campus fall into these two different clusters, with one significant outlier - University Hall's 33kV 3 phase substation, which appears to be scheduled in a rather poor manner. In order to further dig into the scheduling on campus and how it varies from week to week, we can plot each meter's distribution of scheduling depth (g), as a box plot in one graph. This graph is shown in fig. 6.3, and we can clearly see the variation across the campus. It is important to note that the fitting method is normalized across all buildings, so these measures are scale invariant with respect to buildings' behavior and purely depend on the timing of the building shutting down at night.

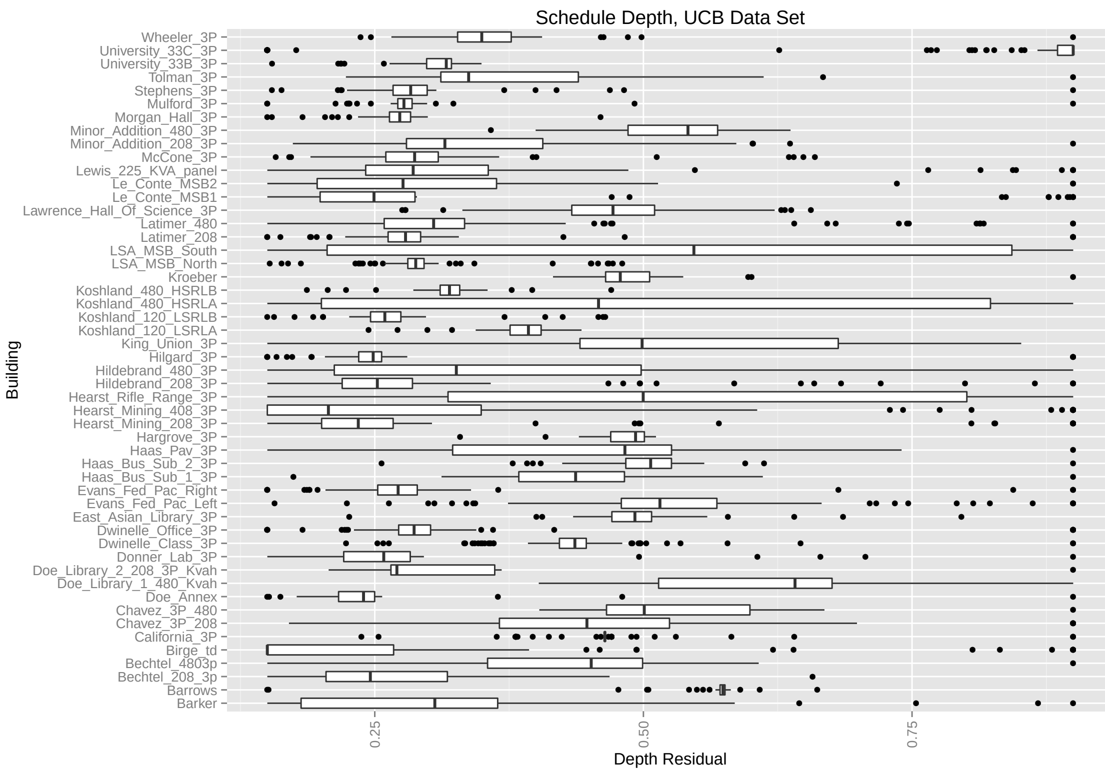


Figure 6.3: Box Plot showing schedule depth for each UC Berkeley building in the data set

Any model can only be ever as good as the fit of that model, so it is important that the model that we use to fit it, so we must be sure that the fit actually matches that which was predicted in the first place. Figure 6.4 shows the residuals of the fit to the LDC curve given in eq. (5.1) for variable g. As we can see, the residuals for most buildings are tightly clustered, with the maximum range for any residual's inter-quartile range (IQR) being 0.02709 and the standard deviation of that residual

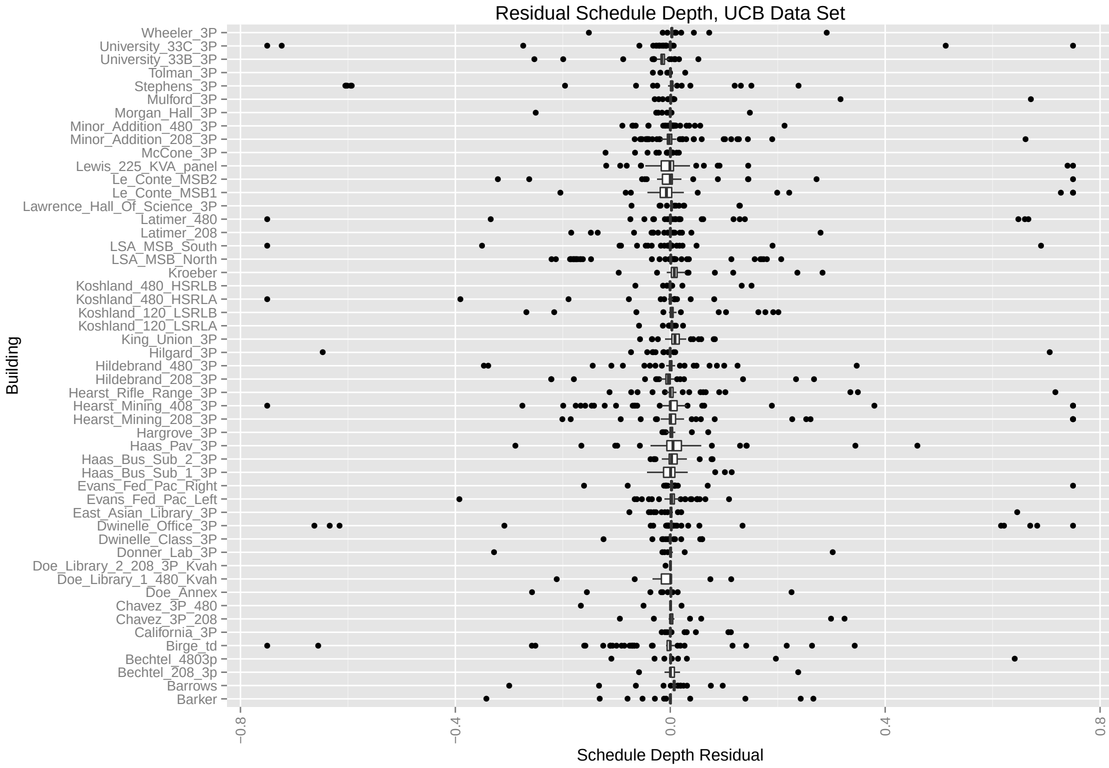


Figure 6.4: Box Plot showing residual of schedule depth for each UC Berkeley building in the data set

being .006926. This is a strong indication that the model as fit by the equation is strongly matched by the reality of the situation in the buildings.

This analysis can apply to any of the parameters in the data set, and so to continue with the evaluation of campus scheduling, the next plot evaluates parameter a - the off schedule slope, graphed as a box plot in fig. 6.5. We can further determine that in some buildings, the off schedule slope is more variable than in others, indicating that some buildings have a less constant load in the off hours than others - which could be beneficial or not to building performance, depending on what that load is. For the purposes of determining this, it might useful to know the weekly peak to base ratio of the building - to choose a building that is both hard to predict (based on fig. 6.6) and has a reasonably wide range of off schedule slope, we find that peak to base ratio of Stephens hall is, on average 4.187301. indicating that there is a significant difference between the average scheduled hour in the building on peak and the average off-schedule hour in the building, presumably at night.

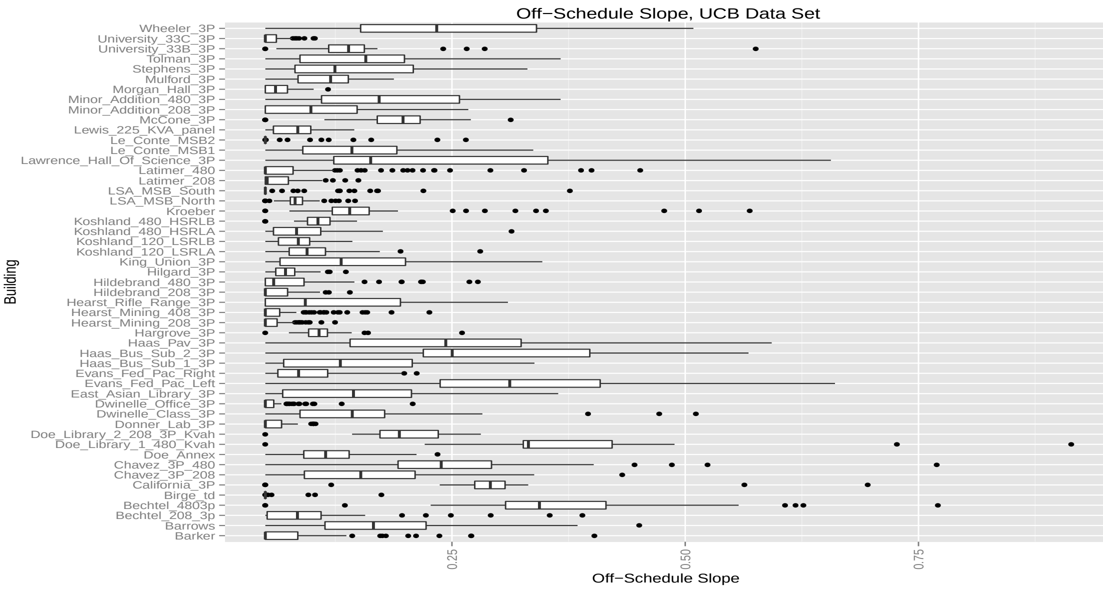


Figure 6.5: Box Plot showing the Off-Schedule Slope for each UC Berkeley building in the data set

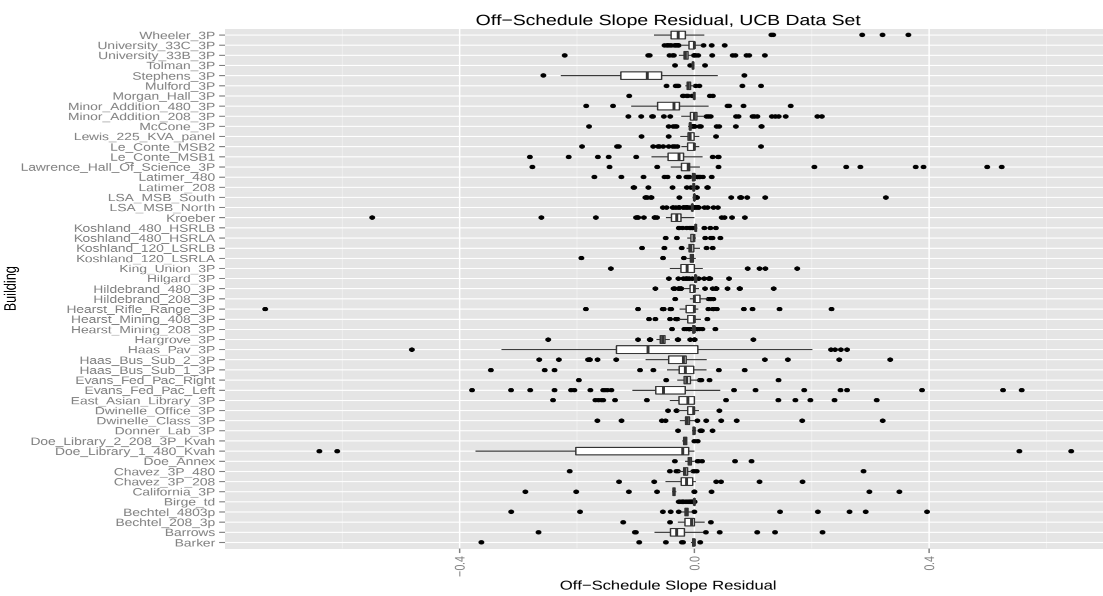


Figure 6.6: Box Plot showing the Off-Schedule Slope residual for each UC Berkeley building in the data set

Peak Variability

In addition to the scheduling of the building, peak variability is also important to asses how variable the peak parameters are with respect to the campus as a whole. In general, these parameters should vary with the weather - the peak height b should be variable depending on how hot the day is (or cold, in another climate), and the peak depth c should be variable depending on the amount of days that the extreme weather persists for. Given that the buildings in this sample are normalized across the campus, the behavior of each building with respect to the weather can be contrasted to each of the other buildings in the sample. The weekly peak height, b for each building is displayed as a box plot in fig. 6.7.

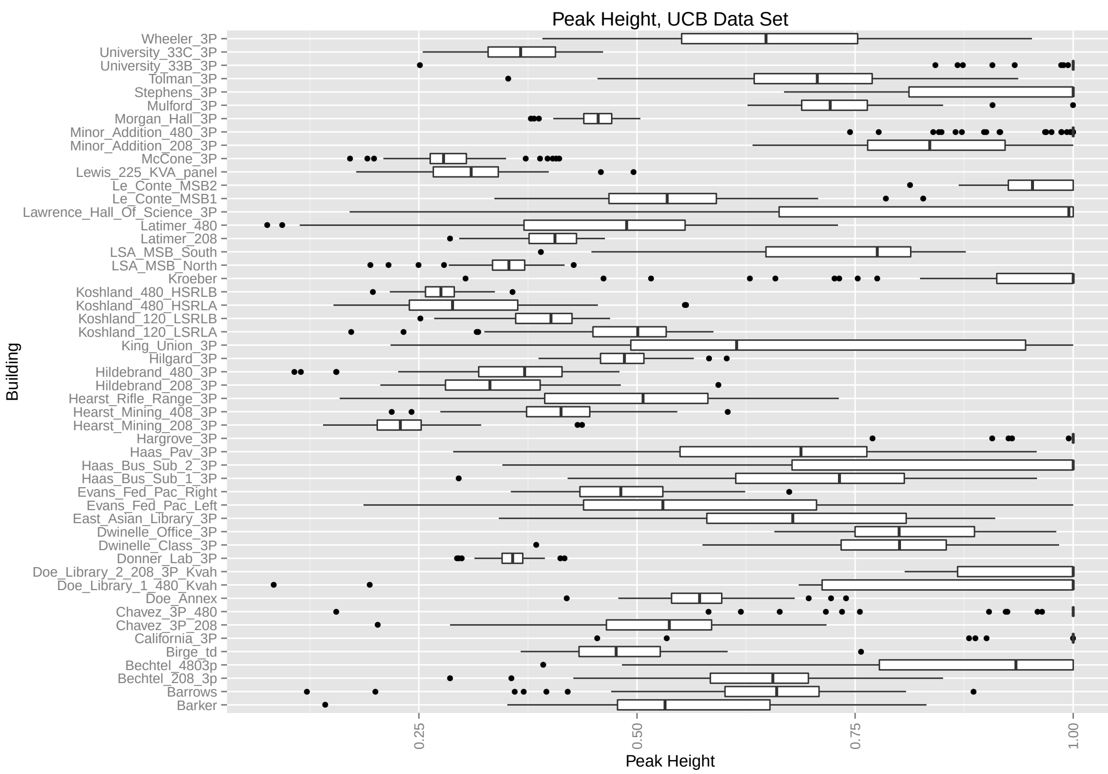


Figure 6.7: Box Plot showing the Peak Height for each UC Berkeley building in the data set

From the analysis of peak height for all UC Berkeley meters in the data set, it becomes apparent that the specific meters have different peak heights on average - if we recall from chapter 5, this peak height is normalized against the maximum peak over the entire year, so bars in this chart that cluster near the range of 0.75 – 1.00 demonstrate that the building is less responsive to weather events than other buildings in the data set - that is, they are more constantly near the peak of their power consumption.

In addition to seeing which buildings are constantly near peak, it is important to understand which buildings are unpredictable in terms of their peak height. Once again, we plot the residual of the predicted fit versus the actual fit - this is shown in fig. 6.8. In particular, Doe Library's 480V feed (probably powering lights + HVAC), Haas Pavillion, and Stephens hall stand out in this case. These buildings will be worth investigating as to whether there is a reason that they are unpredictable on peak. The second metric to evaluate in whether a building responds well to

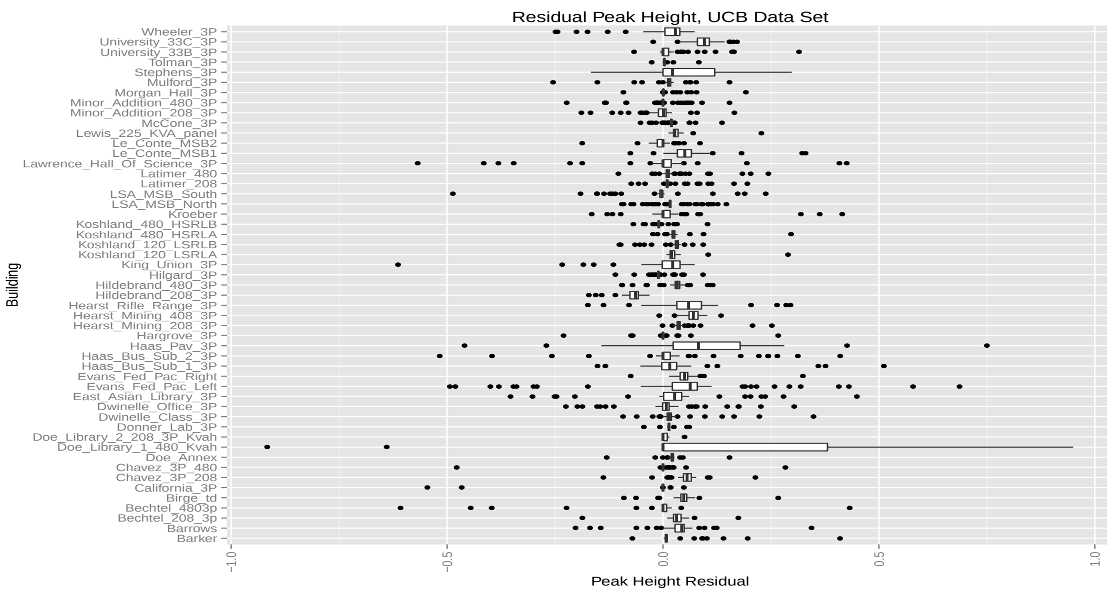


Figure 6.8: Box Plot showing the Peak Height residual for each UC Berkeley building in the data set

temperature changes that influence peak consumption is the peak depth, denoted by c in the fit equation. fig. 6.9 shows box plots of the peak depth on campus, and fig. 6.10 shows the residual of said prediction. In the first graph, Minor addition's 208 feed seems to be highly variable, while in the second graph, Doe library's 408 kVah seems to be a suspect for investigation. We will examine these further in the data mining section

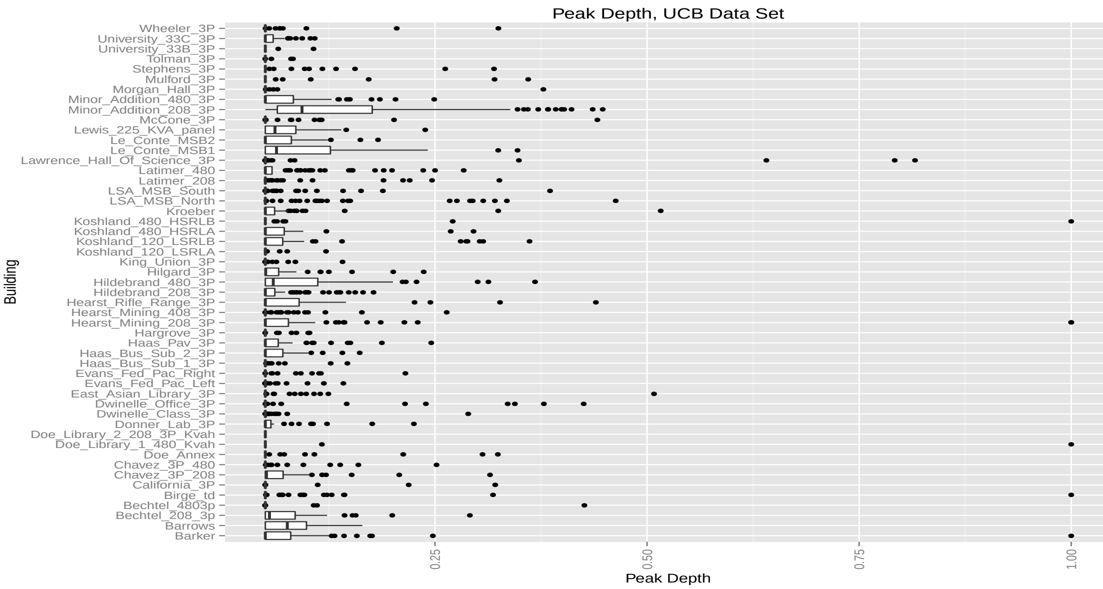


Figure 6.9: Box Plot showing the Peak Depth for each UC Berkeley building in the data set

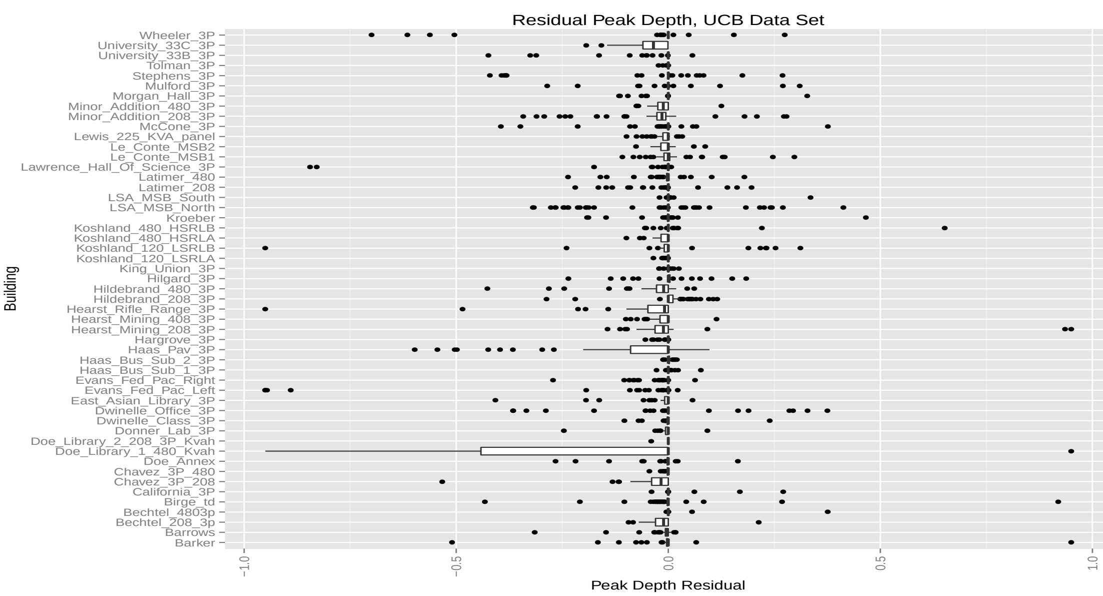


Figure 6.10: Box Plot showing the Peak Depth residual for each UC Berkeley building in the data set

6.4 Mining the Results

It is important to be able to automatically pick out problems in a campus of buildings. For this diagnostic step, the data that generated by the previous graphs is used in order to automatically pick out problem buildings on the campus. A table of the buildings that have the highest inter-quartile range for Schedule Depth and off-schedule slope residual, is generated, indicating unpredictable behavior in the buildings. Then a Daily Residual Control Chart, a six parameter control chart are made.These each show the characteristic of the worst performers in the table. From

Meter	g Residual IQR	Meter	a IQR
Doe_Library_1_480_Kvah	0.19230010	LSA_MSB_South	0.6372847
Haas_Pav_3P	0.13868136	Koshland_480_HSRLA	0.6226659
Stephens_3P	0.06944395	Hearst_Rifle_Range_3P	0.4828450
Evans_Fed_Pac_Left	0.05021513	Hildebrand_480_3P	0.2855318
Minor_Addition_480_3P	0.03721965	King_Union_3P	0.2403630

section 6.4, we can see that the Doe Library meter has the highest inter-quartile range of the residual for g - the schedule shift. We can then automatically produce the associated Demand Residual Control Chart and examine whether the building is behaving oddly in off hours, as we should expect it to be with this result. Figure 6.11 shows the graph automatically generated from this result, and we can see that in the hours from midnight to five in the morning the building is oddly unpredictable. This is cause for further examination of the facility.

The Life Science Addition Main Substation B (South) feed is even more mysterious - this feed displays unusual unpredictability at all hours of the day, an the building itself seems to be relatively unscheduled in this manner, according to fig. 6.12. It is worth evaluating why this is and contrasting this meter with a properly behaving building such as California hall, whose Demand Mean control chart is shown in fig. 6.13, and Demand Residual Control Chart is fig. 6.14. It should be noted that on average in Califorina hall, the off-hours involve a base load that is much lower and much more predictable than the on-peak hours, which is what we should expect in a well conditioned building - controls that shut the building down in off hours. This is supporting evidence for the hypothesis that in order for a building to be well behaved it is a necessary but not sufficient condition for it to be predictable.

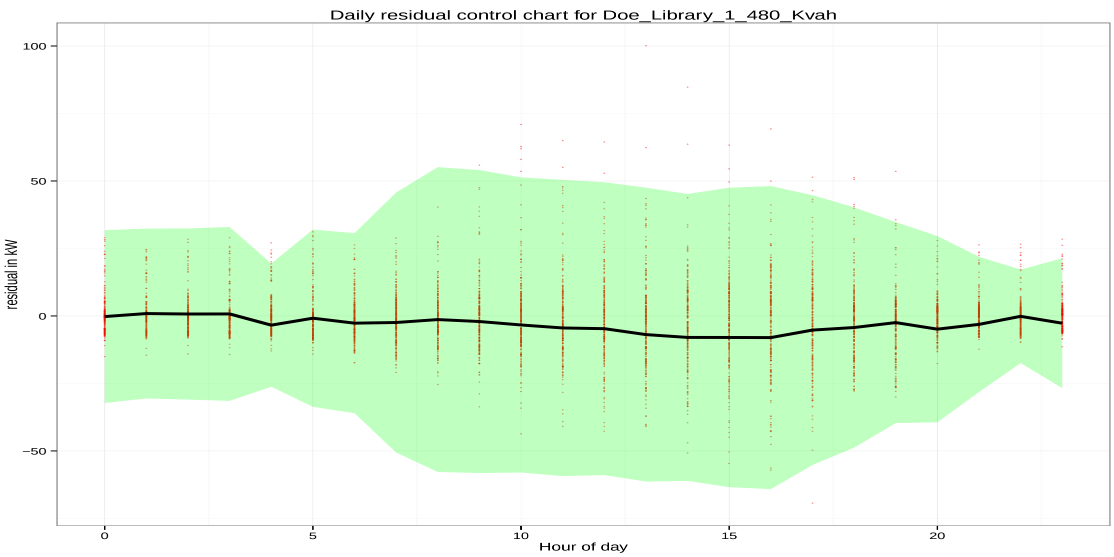


Figure 6.11: Daily Residual Control Chart for Doe Library's 480V feed. As we can see, the building is hard to predict at night and at off hours, with a dip in predictability in the mid-morning. This is likely caused by a control problem.

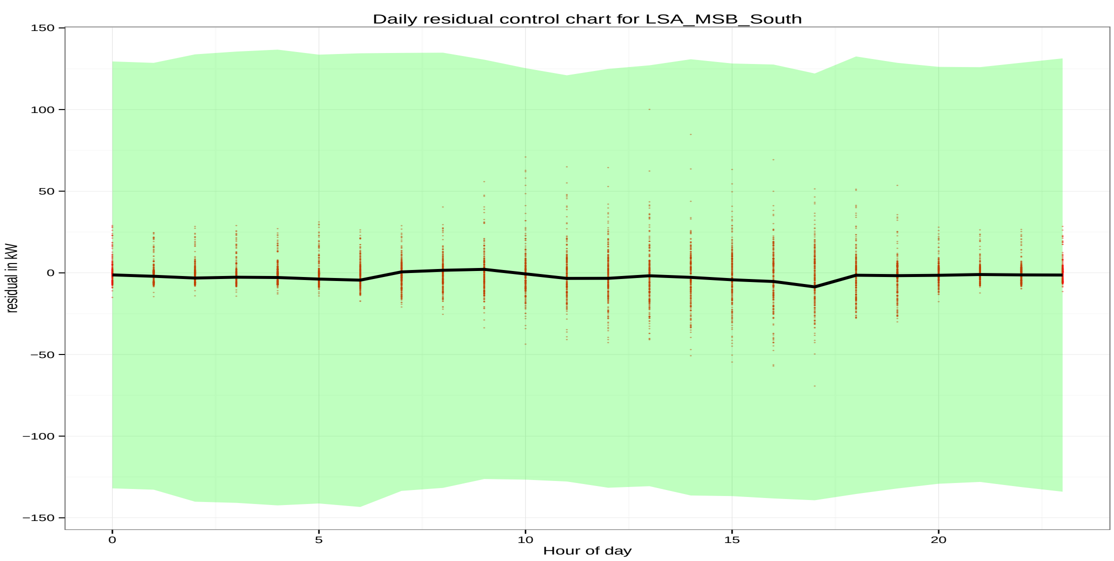


Figure 6.12: Daily Residual Control Chart for LSA's Main Substation B South feed. The building is generally hard to predict at any hour of the day, with significant unpredictability in off hours. This building requires a high level of scheduling evaluation

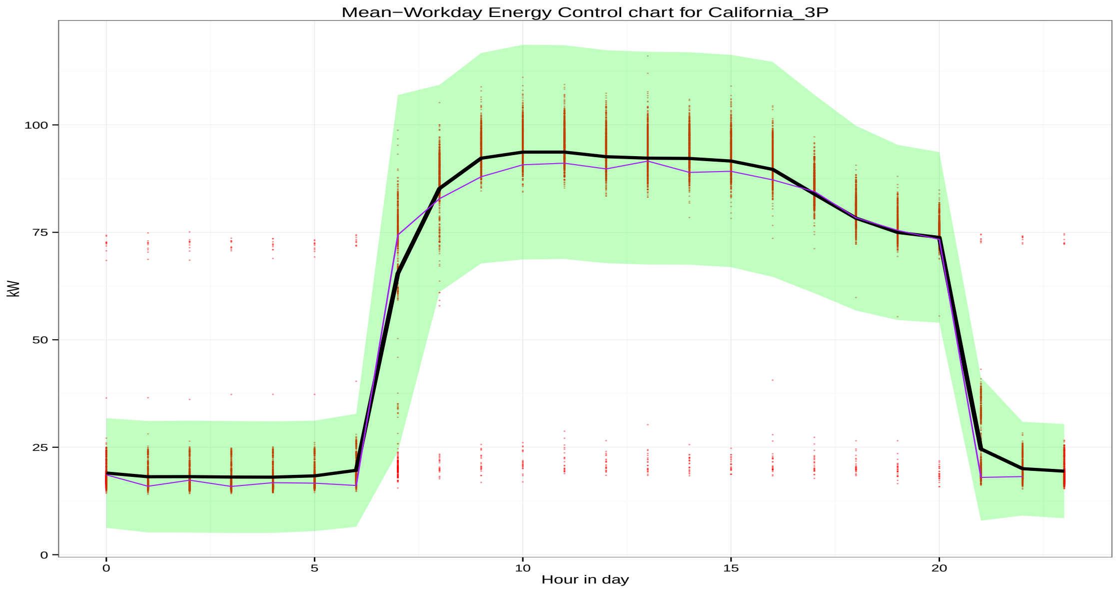


Figure 6.13: Daily Mean Control chart for California Hall, a well scheduled building on campus.

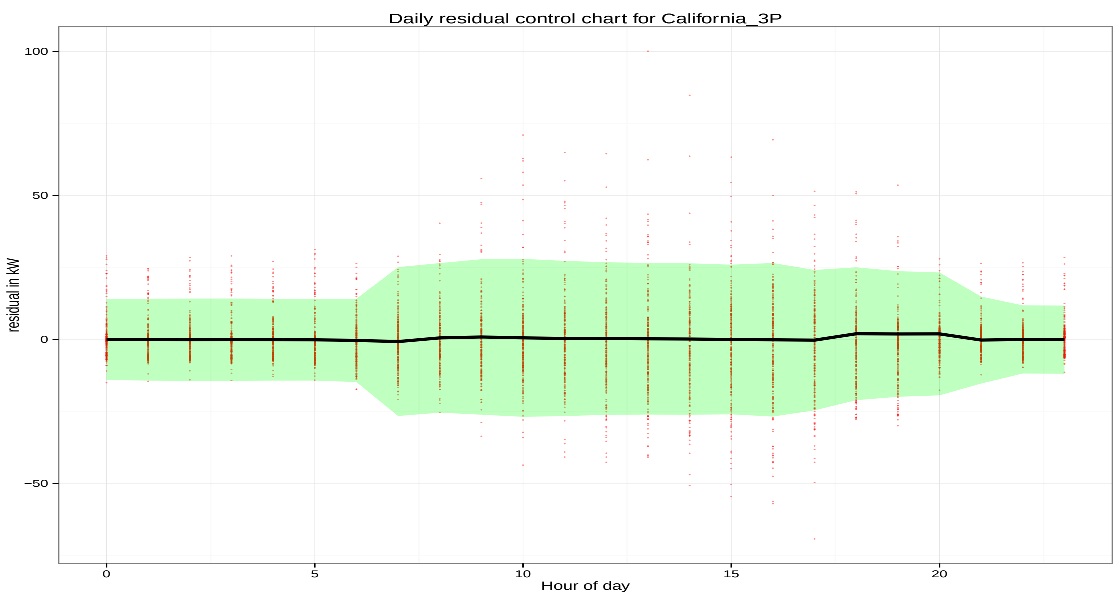


Figure 6.14: Daily Residual Control chart for California Hall, a well scheduled building on campus.

Exactly as the data for the scheduling can automatically be sorted and control charts produced for building engineers to examine, we can do the same for the peak variability of the buildings. For this examination, we will use the residuals of parameters b and c from our curve fit - the peak heigh and peak depth. This will give us insight on how well buildings can be predicted as to their peak power use based upon previous power use. The table appropriate here is given by section 6.4.

Meter	b Residual IQR	Meter	c Reisdual IQR
Doe_Library_1_480_Kvah	0.38094344	Doe_Library_1_480_Kvah	0.44165437
Haas_Pav_3P	0.15512817	Haas_Pav_3P	0.08886970
Stephens_3P	0.11971847	University_33C_3P	0.05899582
Evans_Fed_Pac_Left	0.05785709	Hearst_Rifle_Range_3P	0.04798973
Hearst_Rifle_Range_3P	0.05696093	Chavez_3P_208	0.03924812

Doe Library's 480V feed appears as a problem area in this analysis. Since this control chart has already been produced above, the next analysis to be done is on Haas Pavillion, which has been identified as a problem area from our analysis before. From fig. 6.15, it can be determined that this building has a large load on average toward the end of the day, and has an unusual pattern of shutdown. Even further evidence of the need to examine this building stems from fig. 6.16, where it can be determined that the building is in fact highly unpredictable at night, something that should be corrected as an examination is made into the building's control systems.

Finally, we once again want to look at a building that is "well behaved" according to this metric, so we choose Wheeler Hall, with peak height residual 0.034429050 and a peak depth residual that is so low that is exceeds the resolution of the model. This building's control charts are shown in figs. 6.17 and 6.18. These graphs are interesting because as we can tell, the building does not shut down quickly at night, with a smoother transition to an off state than California hall, for instance. However, the building is relatively predictable throughout all hours of the day, following the pattern that we have seen in California hall and being less predictable in the middle of the day and more predictable at night. This is what a well behaved building should perform as, and the analysis bears this out.

Additional parameters could be analyzed in order to determine whether the buildings are performing poorly according to the feature extraction used in the LDC technique. In the authors opinion, the parameters examined in this analysis are the most pertinent to discovering whether a building is behaving in an in-control manner. The schedule slope and schedule shift should also be examined in situations where similar buildings are being compared, but for the moment, we leave the formulation of other parameter examinations as a template for how to do this analysis.

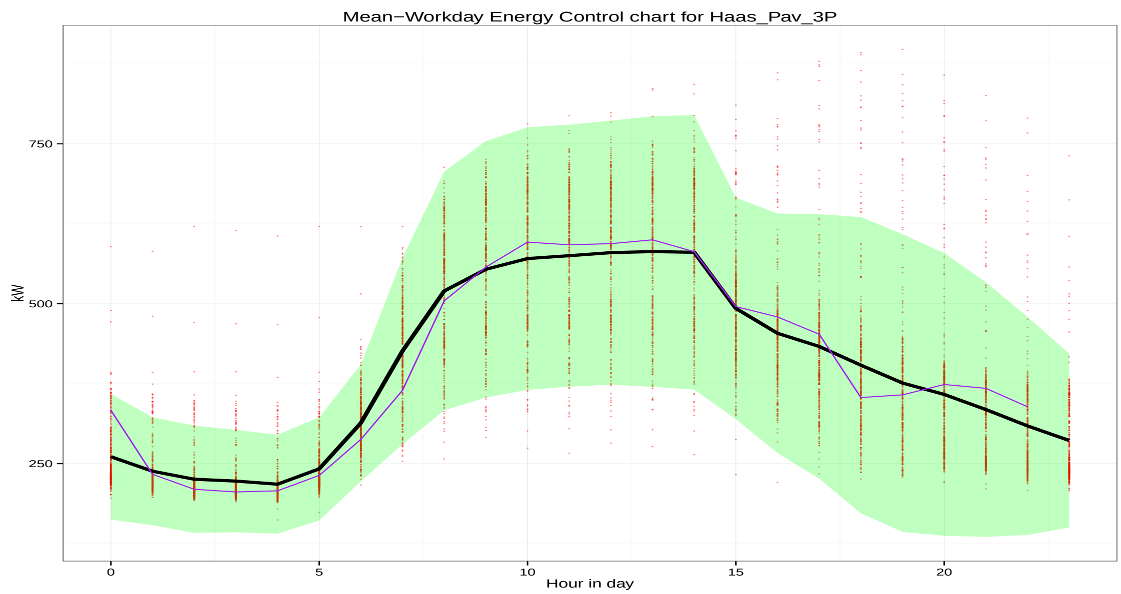


Figure 6.15: Daily Mean Control chart for Haas Pavillion.

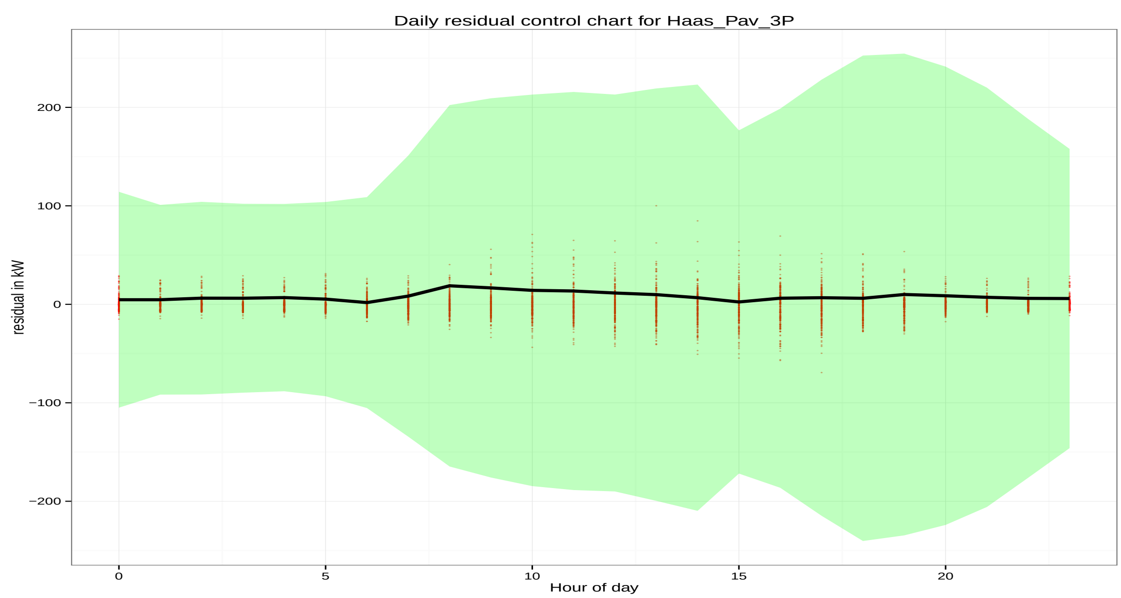


Figure 6.16: Daily Residual Control chart for Haas Pavillion.

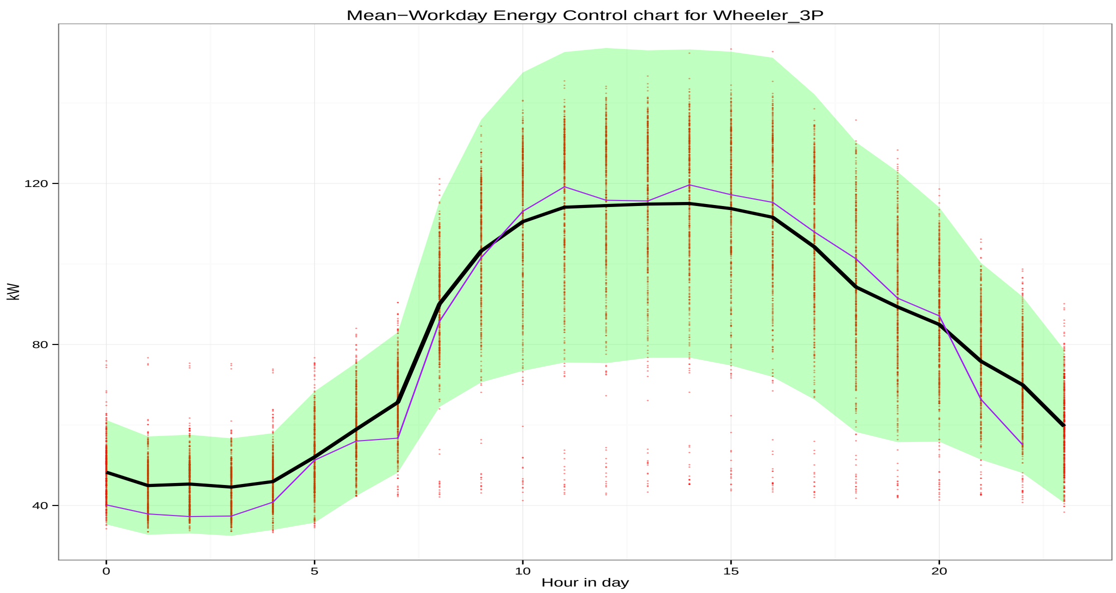


Figure 6.17: Daily Mean Control chart for Wheeler Hall.

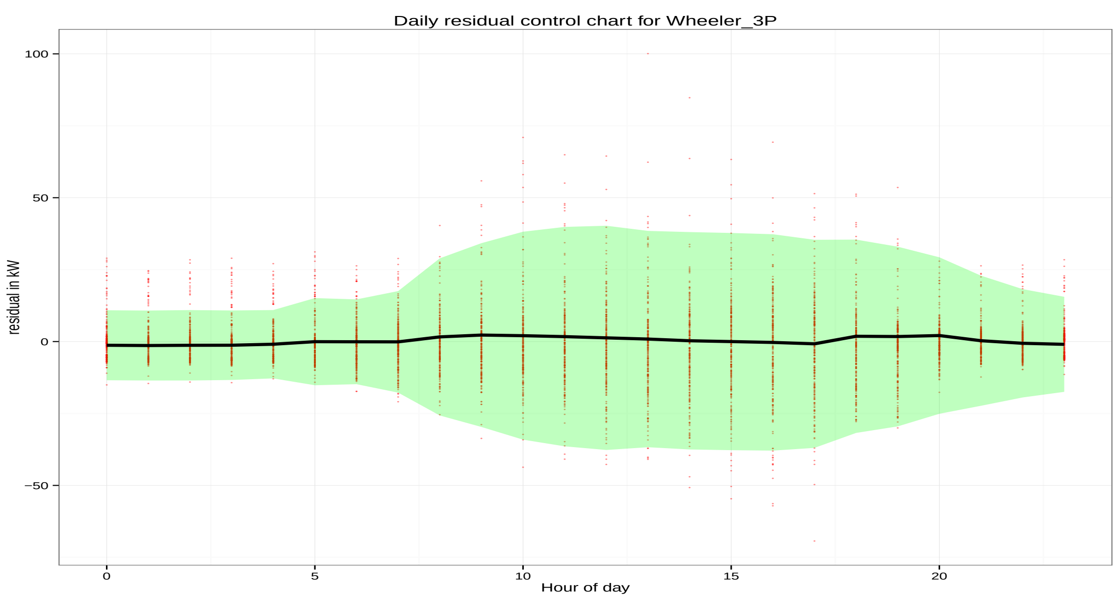


Figure 6.18: Daily Residual Control chart for Wheeler Hall.

6.5 Diagnostic results

Sorting building behaviors in the manner presented quickly shows which buildings on a campus are behaving irregularly. Below, fig. 6.19 shows the power trace of LSA substation MSB - with fig. 6.20 showing the distribution of the residuals for predicting power use in that building.

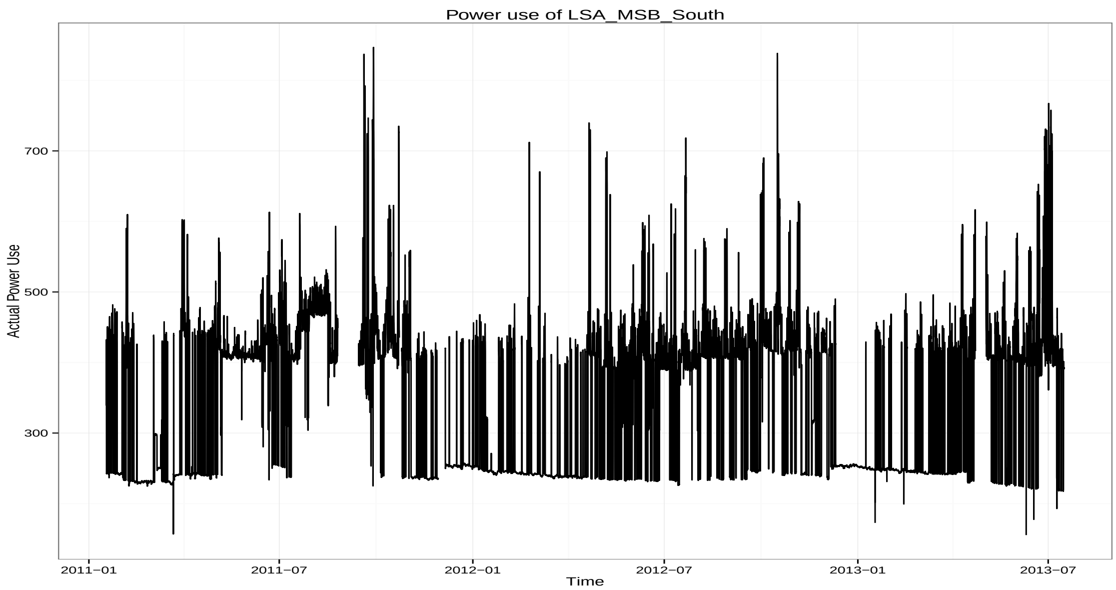


Figure 6.19: Power Use Trace of LSA Substation B. Operation changes and unpredictable features are shown clearly.

This building can be contrasted with other buildings presented in the analysis, such as Wheeler Hall, which is one of the more well behaved buildings according to the algorithm. The power trace presented in fig. 6.21 clearly shows more consistent behavior than the previous building, but depicts some operational changes that indicate inefficiencies. Again, the distribution is presented in fig. 6.22, and this is more normal than the previous measurement. In this context, we take normally distributed errors to indicate that our prediction algorithms are stationary - and therefore that the building is "misbehaving in a predictable fashion." To further illustrate, a well behaved building, California Hall is presented in fig. 6.23, with the density distribution shown in fig. 6.24. This building, which is well scheduled, displays a normal error distribution and has (largely) regular operations.

Just as in the case of LeConte Hall in Chapter 5, Anomalies in building operations could be detected by transforming the data and subsequently highlighting outliers.

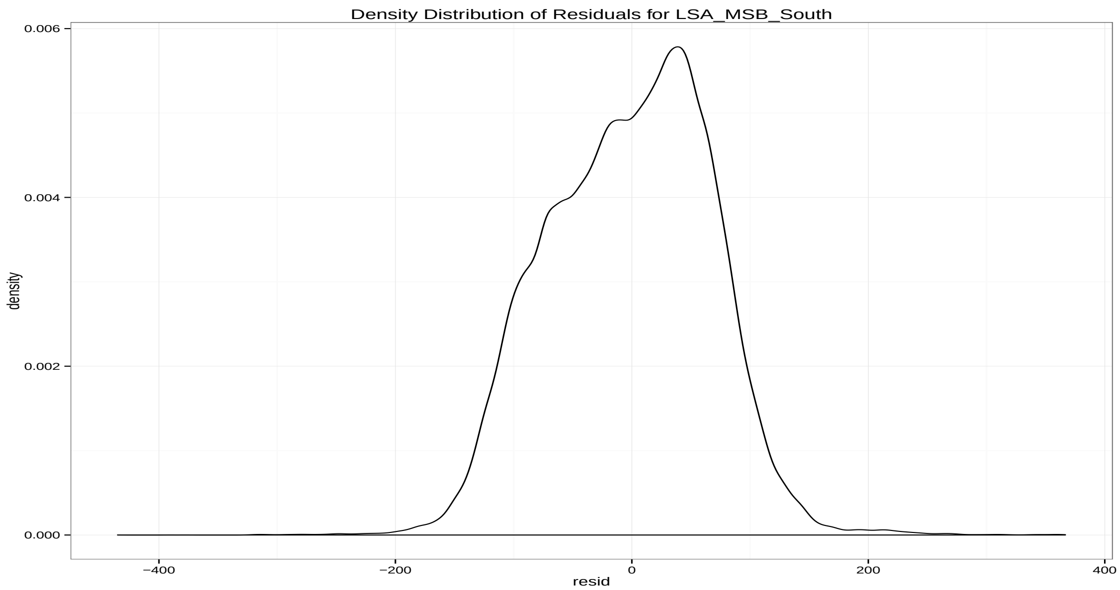


Figure 6.20: A density distribution of residuals for LSA Main Substation B. The non-normal distribution of errors indicates that the building is not easily predictable.

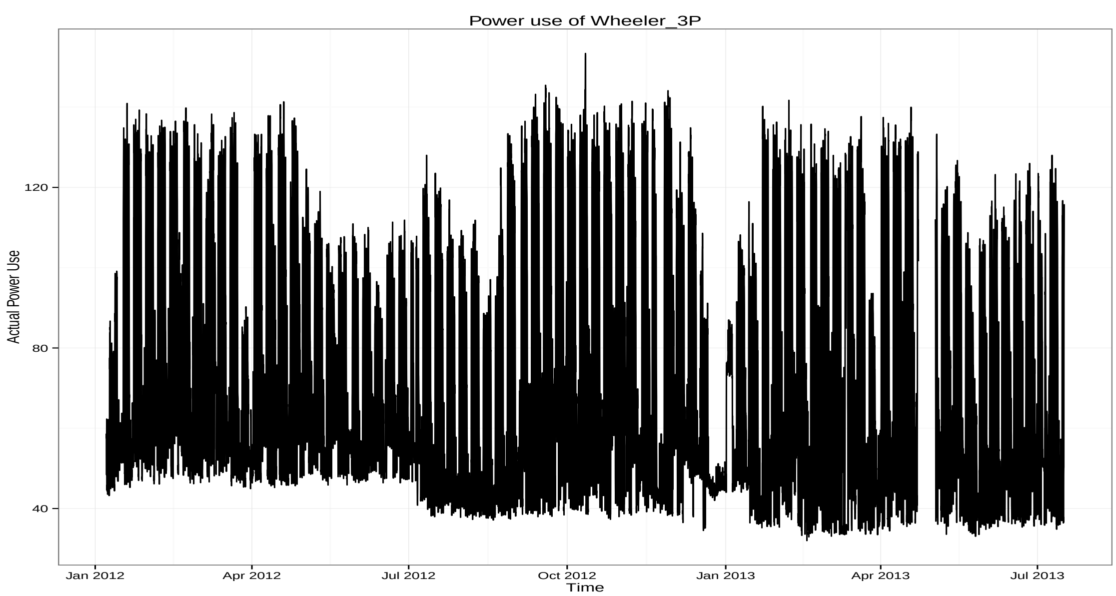


Figure 6.21: Power Use Trace of Wheeler Hall. It is easy to see some operational changes in this building which would are indicative of inefficient behavior.

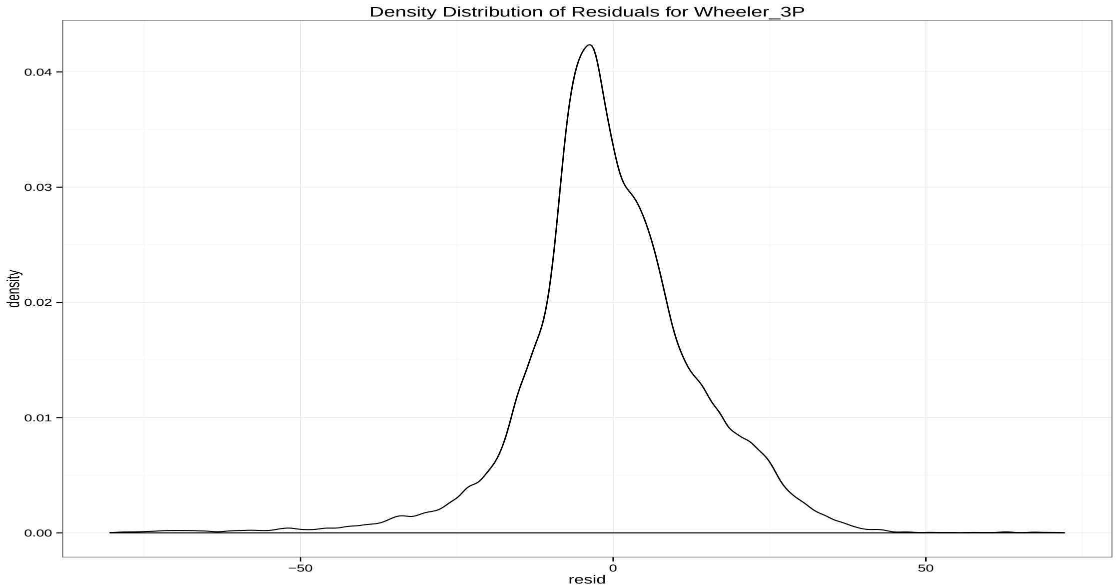


Figure 6.22: A density distribution of residuals for Wheeler Hall. This distribution is closer to normal than the one presented above, indicating greater predictability in operations.

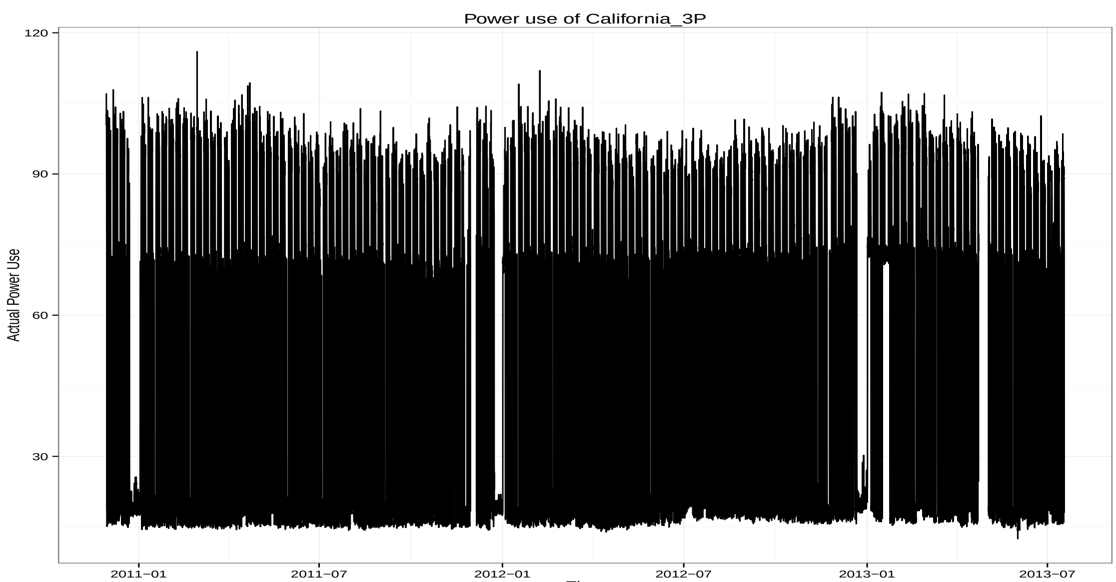


Figure 6.23: Power Use Trace of California Hall. This building displays regular scheduling an operations.

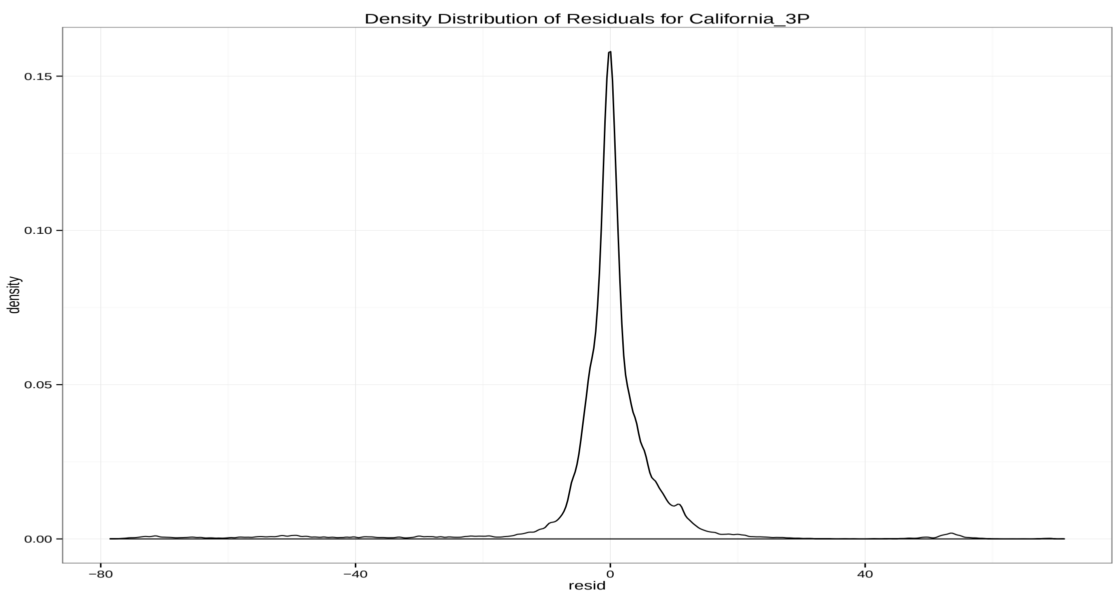


Figure 6.24: A density distribution of residuals for California Hall. This distribution is very close to normal, and points of interest in the building operations could be extracted by normal SPC methods.

6.6 Conclusions and Implications

The data analyses in this section build up an ability to sort through a campus's worth of data for energy use in order to prioritize that data for examination by a building engineer. This work builds upon the previous work in chapter five of load duration curve analysis, and shows how not only can we find outliers within specific buildings, but we can find which buildings are the worst behaving on an entire campus. This leads us to several specific points:

- Load Duration Curve analysis is a new and important way to categorize building behaviors within a campus, and lends itself to quickly and specifically being able to pick out characteristics of buildings that are out of the normal range for buildings within a specific locale. These buildings may be clustered to show differing behaviors in various ways, including k-means clustering and sorting by behavior over time.

- Prediction is an important part of being able to sort through building behavior. While we can sort through specific behaviors of a building without predicting on previous behavior (example schedule depth), there is no other way to compare the building's actual behavior to what we would expect without predicting expected behavior of the systems against previous behavior of the systems. This is a novel technique for building analytics, and is an extension of techniques used for time series analysis in statistical process control.

- As with SPC in factories, the methods presented here cannot solve the problems on the campus, and serve merely to aide the engineer responsible for resolving problems and magnify their effectiveness. This approach, which is applicable to campus wide energy use, is best combined with an intelligent building commissioning agent to design a sensible approach to fixing the buildings.

Chapter 7

Implications of Statistical Process control in Buildings

7.1 Research Overview

In this work, I first took a look at buildings, their energy consumption, and why they it is important to curtail their energy use as a societal need. With buildings using 41% of primary energy in the United States and a comparable amount of energy being used worldwide, it is imperative that we cut our building energy usage as a method of managing our carbon consumption. Key metrics of building energy use were discussed, presented, and used for basic data analysis. Chapter two presented methods for analyzing buildings as time series data, evaluating how those methods are performing, and shows how to view time series data. Chapter three goes over prediction in the form of time series, how to effectively automate this for buildings, and how to use best practices in prediction methodologies for ensemble based, automated predictions of buildings for a week ahead. This is followed up by chapter four, showing how predictions can be used to produce control charts in the manner of statistical process control which can lead to anomaly detection in commercial buildings. Chapter five goes into some depth on how previously used "snapshot" images of buildings can be converted into moving pictures of the building life cycle, and chapter six goes into how these moving pictures can be used to analyze an entire campus quickly and effectively to find buildings that are not behaving as expected.

7.2 Research Challenges

The research conducted in this field has many challenges, but the biggest one is the difficulty in obtaining data. I was fortunate enough to have a working relationship with Devan Johnson of KW Engineering, who helped me through the challenges of getting data on real buildings and analyzing them intelligently. For future work however, this area is very difficult to approach in a university setting because privacy concerns with data with energy in buildings. In addition, there is a large learning curve assorted with buildings systems in order to learn about the interactions between components that may not be intuitive to an individual before they understand the greater picture of energy in buildings.

7.3 Research Vision

Future work on Statistical Process Control in buildings will involve much more machine learning and data mining than what was presented here. While it takes an experienced engineer or building scientist to understand the implications of each of the control graphs or prediction methods that I have presented in this thesis, it should be perfectly possible to produce an automated diagnostic system that continuously runs on a building based upon SPC charts that continuously run in the building after it has been retrocomissioned, identifying outliers automatically calling professionals to fix these problems. Additionally, more metrics will be visualized as control charts for buildings than listed here - it should be possible to transform the data in such a way that reveals even more information that was given here for the load duration curve transformation of data on a weekly basis.

Finally, sensor technology will improve in the future, and an intelligent manner for using these sensors would be to install them, have them automatically update with information relevant to their location, usage, and other relevant metadata, and then being a process control method right at that moment based on prediction and machine learning. This will be the future of this research - learning what to do with ubiquitous sensing and almost unlimited sensor information in a computationally effective manner.

Bibliography

[1] Layth C Alwan and Harry V Roberts. "Time-series modeling for statistical process control". In: *Journal of Business & Economic Statistics* 6.1 (1988), pp. 87–95.

[2] Daniel W Apley. "Time series control charts in the presence of model uncertainty". In: *Journal of manufacturing science and engineering* 124.4 (2002), pp. 891–898.

[3] Sylvain Arlot, Alain Celisse, et al. "A survey of cross-validation procedures for model selection". In: *Statistics surveys* 4 (2010), pp. 40–79.

[4] Robert B Cleveland et al. "STL: A seasonal-trend decomposition procedure based on loess". In: *Journal of Official Statistics* 6.1 (1990), pp. 3–73.

[5] Jeremy Coyle et al. "The application of Time Series Methodology to improve Prediction of Building Power Demand up to 1 week". In: *In Submission* (2014).

[6] Tony Day et al. "Degree-days: theory and application". In: *TM41* (2006).

[7] Alysha M De Livera. "Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing". In: *Journal of the American Statistical Association* 106.496 (), pp. 1513–1527.

[8] R E DeVor, T Chang, and J W Sutherland. *Statistical quality design and control: contemporary concepts and methods*. Pearson/Prentice Hall, 2007.

[9] Elizabeth Doris, Jaquelin Cochran, and Martin Vorum. "Energy efficiency policy in the united states: overview of trends at different levels of government". In: *National Renewable Energy Laboratory of the US Department of Energy, Colorado, US* (2009).

[10] EIA. "November 2013 Monthly Energy Review". In: (). DOI: `DOE/EIA-0035(2013-05)`. URL: `http://www.eia.gov/totalenergy/data/monthly/pdf/mer.pdf`.

[11] Energy Information Agency (EIA). *Commercial Buildings Energy Consumption Survey (CBECS)*. Tech. rep. U.S. Department of Energy, 2003.

[12] Energy Information Agency (EIA). *Commercial Buildings Energy Consumption Survey (CBECS)*. Tech. rep. U.S. Department of Energy, 2007.

[13] *ENERGY EFFICIENCY AND DEMAND MANAGEMENT PROCEDURE GENERAL- CALCULATING CUSTOMER BASELINE LOAD*. 0th ed. CONSOLIDATED EDISON COMPANY OF NEW YORK, INC. 4 IRVING PLACE NEW YORK, NY 10003, Feb. 2013.

[14] P Enkvist, Tomas Nauclér, and Jerker Rosander. "A cost curve for greenhouse gas reduction". In: *McKinsey Quarterly* 1 (2007), p. 34.

[15] Paul S Fasolo and Dale E Seborg. "Monitoring and fault detection for an HVAC control system". In: *HVAC&R Research* 1.3 (1995), pp. 177–193.

[16] Paul Fasolo and Dale E Seborg. "An SQC approach to monitoring and fault detection in HVAC control systems". In: *American Control Conference, 1994*. Vol. 3. IEEE. 1994, pp. 3055–3059.

[17] Pacific Gas and Electric Company. *Automated Demand Response (ADR) Customer Update*. Powerpoint. Aug. 2013. URL: `http://pge-adr.com/wp-content/uploads/2013/09/PGnE-Customer-ADR-Training-V2.pdf`.

[18] Charles Goldman. "Coordination of energy efficiency and demand response". In: (2010).

[19] J. Granderson and P.N. Price. *Evaluation of the predictive accuracy of five whole-building baseline models*. Tech. rep. LBNL-5886E. Berkeley, CA: Lawrence Berkeley National Laboratory, 2012.

[20] CC Holt. "Forecasting trends and sesonals by exponentially weighted moving averages". In: *ONR Memorandum* 52 (1957).

[21] Tyler Hoyt. *RSmap: sMAP client bindings and utilities for R*. R package version 1.0. 2013.

[22] Rob J Hyndman. "Another look at measures of forecast accuracy". In: *International Journal of Forecasting* 22.4 (), pp. 679–688.

[23] Rob J Hyndman and George Athanasopoulos. "Forecasting: principles and practice". In: *2014, http://otexts. com/fpp/, Accessed on May 2014* (2014).

[24] IEA. "Key World Energy Statistics 2013". In: (). DOI: `http://dx.doi.org/10.1787/key_energ_stat-2012-en`. URL: `/content/book/key_energ_stat-2012-en`.

[25] Mike Kehoe et al. "Smarter cities series: A foundation for understanding ibm smarter cities". In: *Redguides for Business Leaders, IBM* (2011).

[26] D Kozikowski, A Breidenbaugh, and M Potter. *The Demand Response Baseline, v. 1.75*. 2006.

[27] Young M Lee et al. "Analytics for Smarter Buildings". In: *International Journal of Business Analytics (IJBAN)* 1.1 (2014), pp. 1–15.

[28] Young M Lee et al. "Building energy performance analytics on cloud as a service". In: *Service Science* 5.2 (2013), pp. 124–136.

[29] Zhengwei Li et al. "A rule augmented statistical method for air-conditioning system fault detection and diagnostics". In: *Energy and Buildings* 54 (2012), pp. 154–159.

[30] S Makridakis. "The accuracy of extrapolation (time series) methods: results of a forecasting competition". In: *Journal of Forecasting* ().

[31] Spyros Makridakis. "Note-Sliding Simulation: A New Approach to Time Series Forecasting". In: *Management Science* 36.4 (1990), pp. 505–512.

[32] Spyros Makridakis and Michele Hibon. "The M3-Competition: results, conclusions and implications". In: *International journal of forecasting* 16.4 (2000), pp. 451–476.

[33] J L Mathieu. "Quantifying Changes in Building Electricity Use, With Application to Demand Response". In: *IEEE Trans. Smart Grid.* 2.3 (), pp. 507–518.

[34] A Meier. "The cost of conserved energy as an investment statistic". In: *Proceedings of the Sixth Annual Industrial Energy Conference*. Vol. 2. 1984.

[35] Alan Meier et al. "Usability of residential thermostats: Preliminary investigations". In: *Building and Environment* 46.10 (Oct. 2011), pp. 1891–1898.

[36] Evan Mills. "Building commissioning: A golden opportunity for reducing energy costs and greenhouse-gas emissions". In: (2010).

[37] Donald G Newnan. *Engineering economic analysis*. Vol. 2. Oxford University Press, 2004.

[38] Guy R Newsham, Sandra Mancini, and Benjamin J Birt. "Do LEED-certified buildings save energy? Yes, but..." In: *Energy and Buildings* 41.8 (2009), pp. 897–905.

[39] Stephen Pacala and Robert Socolow. "Stabilization wedges: solving the climate problem for the next 50 years with current technologies". In: *science* 305.5686 (2004), pp. 968–972.

[40] Therese Peffer et al. "How people use thermostats in homes: A review". In: *Building and Environment* 46.12 (2011), pp. 2529–2541.

[41] Luis Perez-Lombard, Jose Ortiz, and Christine Pout. “A review on buildings energy consumption information”. In: *Energy and buildings* 40.3 (2008), pp. 394–398.

[42] Luis Pérez-Lombard et al. “A review of benchmarking, rating and labelling concepts within the framework of building energy certification schemes”. In: *Energy and Buildings* 41.3 (2009), pp. 272–278.

[43] Alain Poulin et al. “Load duration curve: A tool for technico-economic analysis of energy solutions”. In: *Energy and buildings* 40.1 (2008), pp. 29–35.

[44] Phillip N. Price et al. “Using Whole-Building Electric Load Data in Continuous or Retro-Commissioning”. In: *19th National Conference on Building Commissioning*. Cincinnati, OH, Aug. 2011.

[45] Arthur H Rosenfeld et al. “Opportunities in the Building Sector: Managing Climate Change”. In: *Aip Conference Proceedings*. Vol. 1044. 1. 2008, p. 3.

[46] KW Roth et al. “Energy impact of commercial building controls and performance diagnostics: market characterization, energy impact of building faults and energy savings potential”. In: *Prepared by TAIX LLC for the US Department of Energy. November. 412pp (Table 2–1)* (2005).

[47] Noa Ruschin-Rimini, Oded Maimon, and Roni Romano. “Visual analysis of quality-related manufacturing data using fractal geometry”. In: *Journal of Intelligent manufacturing* 23.3 (2012), pp. 481–495.

[48] Jeffrey Schein and John M House. “Application of control charts for detecting faults in variable-air-volume boxes”. In: *TRANSACTIONS-AMERICAN SOCIETY OF HEATING REFRIGERATING AND AIR CONDITIONING ENGINEERS* 109.2 (2003), pp. 671–682.

[49] John H Scofield. “Do LEED-certified buildings save energy? Not really...” In: *Energy and Buildings* 41.12 (2009), pp. 1386–1390.

[50] Michael J Scott et al. “Evaluating sub-national building-energy efficiency policy options under uncertainty: Efficient sensitivity testing of alternative climate, technological, and socioeconomic futures in a regional integrated-assessment model”. In: *Energy Economics* 43 (May 2014), pp. 22–33.

[51] Robert H Shumway and David S Stoffer. *Time series analysis and its applications: with R examples*. Springer, 2011.

[52] D Qin Stocker. “Climate change 2013: The physical science basis”. In: *Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Summary for Policymakers, IPCC* (2013).

[53] Norman R Swanson and Halbert White. "Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models". In: *International journal of Forecasting* 13.4 (1997), pp. 439–461.

[54] Leonard J Tashman. "Out-of-sample tests of forecasting accuracy: an analysis and review". In: *International Journal of Forecasting* 16.4 (), pp. 437–450.

[55] James W Taylor. "Short-term electricity demand forecasting using double seasonal exponential smoothing". In: *Journal of the Operational Research Society* 54.8 (2003), pp. 799–805.

[56] Johan Thor et al. "Application of statistical process control in healthcare improvement: systematic review". In: *Quality and Safety in Health Care* 16.5 (2007), pp. 387–399.

[57] The Carbon Trust. *Degree days for energy management.* Tech. rep. The Carbon Trust.

[58] Daniel A Veronica. "Automatically detecting faulty regulation in HVAC controls". In: *HVAC&R Research* 19.4 (2013), pp. 412–422.

[59] GEORGE P Watkins. "A third factor in the variation of productivity: The load factor". In: *The American Economic Review* 5.4 (1915), pp. 753–786.