ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/327270356

EDIFES 0.4: Scalable Data Analytics for Commercial Building Virtual Energy Audits

Thesis · August 2016

DOI: 10.13140/RG.2.2.12729.62566

CITATION
1

READS
290

3 authors, including:

Ethan Pickering
California Institute of Technology
13 PUBLICATIONS **18** CITATIONS

SEE PROFILE

Roger H. French
Case Western Reserve University
290 PUBLICATIONS **6,718** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Building Energy Analytics View project

Project Thermal Analysis of Microinverter View project

All content following this page was uploaded by Ethan Pickering on 28 August 2018.

The user has requested enhancement of the downloaded file.

EDIFES 0.4: SCALABLE DATA ANALYTICS FOR COMMERCIAL BUILDING VIRTUAL ENERGY AUDITS

by

ETHAN M. PICKERING

Submitted in partial fulfillment of the requirements

For the degree of Master of Science

Department of Mechanical and Aerospace Engineering

CASE WESTERN RESERVE UNIVERSITY

August, 2016

EDIFES 0.4: Scalable Data Analytics for Commercial Building Virtual Energy Audits

Case Western Reserve University
Case School of Graduate Studies

We hereby approve the thesis[1] of

ETHAN M. PICKERING

for the degree of

Master of Science

Dr. Alexis Abramson

Committee Chair, Adviser — June 29th, 2016
Department of Mechanical and Aerospace Engineering

Dr. Roger French

Committee Member, Adviser — June 29th, 2016
Department of Materials Science and Engineering

Dr. Joseph Prahl

Committee Member — June 29th, 2016
Department of Mechanical and Aerospace Engineering

[1] We certify that written approval has been obtained for any proprietary material contained therein.
Defense Date: June 29th, 2016

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my co-advisors Dr. Alexis Abramson and Dr. Roger French for their continuous support, patience, motivation and guidance in this project and in my future aspirations. I would also like to thank Dr. Joseph Prahl, for years of encouragement, insightful advice, and for his contagious passion for engineering.

I would like to specially thank Mohammad Hossain for successfully guiding me into data science and the language R. Thanks to Jack Mousseau and Elle Zadina, for both of their help in turning EDIFES from a collection of functions into a robust software package.

Additionally, thanks to Ali Ahmed, Eugene Matthews, Erika Welichko, Mohamed Elsaeti, Shariq Ali, Yifan Xu, Timothy Peshek, and Laura Bruckman for their assistance and expert consultation in building energy and/or data science.

EDIFES 0.4: Scalable Data Analytics for Commercial Building Virtual Energy Audits

Abstract

by

ETHAN M. PICKERING

Energy Diagnostics Investigator for Efficiency Savings (EDIFES) has been developed for scalable data analytics to conduct virtual energy audits on commercial buildings. Built as a software package in R, EDIFES ingests building electricity data and readily available weather data, applying various data analytics to determine building markers, characteristics, and operational tendencies. Through these analyses building systems are identified, including Heating Ventilation and Air Conditioning (HVAC), lighting, and plug load or other equipment, with characteristics such as load and system scheduling. Once building systems have been identified, EDIFES conducts virtual energy audits to diagnose efficiency issues, determines the impact (i.e. return-on-investment or payback) of potential retrofit actions (e.g. rescheduling HVAC to occupied hours or conducting a lighting retrofit). After this stage, it can be used for measurement and verification (M&V) or continuous commissioning. Six buildings are presented in this thesis.

1 Introduction

Energy is a global subject capturing the attention of all major countries, as evident from the 2015 United Nation Climate Change Conference in Paris, France. A massive consumer of energy is the building sector where there is also significant opportunity to reduce energy waste[1]. The United States uses approximately 100 quads of energy each year, with about 40% attributed to buildings for uses such as heating, cooling, lighting and electronics[2,3]. The U.S. Department of Energy has recognized this potential and has committed to reducing commercial building energy use 20% by 2020, as well as citing a long term goal of 50% reduction in overall use[4]. The "Prioritization Tool," developed in the Building Technologies Office (BTO) of the DOE has calculated the energy savings that could be achieved from currently available and emerging technologies[2,3]. The analysis revealed that implementation of *cost-effective* technologies available today could lead to a reduction of energy use in buildings by 30% by 2030. Accounting for emerging technologies estimated to become cost-effective within 5 years, the energy savings reaches 55%[2,3]. This is equivalent to saving up to $300 billion per year if investments in energy efficiency are made strategically.

Considering the large lifetimes and slow renewal rate of buildings, energy retrofits are necessary to reduce energy consumption[5]. These retrofits typically are identified

and quantified in terms of cost and payback time through building energy audits, where physical tests are conducted within the building over an hours to days long time period. Building simulations can be employed to have a deeper understanding of a building's retrofit potential, but this requires developing a computational model of a building, which can be time consuming and cumbersome. Various data analytics approaches have also been used to gain insight to buildings but limited progress has been made in this area to date. Although solutions can be cost effective and payoff within several years, building efficiency faces great obstacles to implementation. One might blame the lack of progress on risk aversion and a distrust that after an investment is made, a return-on-investment (ROI) might be a credible possibility[3]. The industry needs a transformational solution to simply diagnose problems, build trust in solutions, accelerate their implementation and validate their future economic return.

1.1 Building Energy Audits

Even in today's technologically advanced world, available energy efficiency diagnostics approaches such as energy audits, building automation systems, equipment monitoring and energy simulation can be expensive, confusing, and unduly complex.

1.1.1 Physical Energy Audits

Current energy audits typically entail a physical walk through of a building, performing leak tests, infrared imaging, blower door tests, equipment sub-metering, extensive sensoring and more. These energy audits often require a team of individuals to survey

an entire building, its tendencies, as well as occupant behavior[6]. The time and cost required can outweigh the potential energy savings, and the recommendations made can vary drastically between audits[6,7]. In fact, an interesting comparison of three on-site independent energy audits on a single commercial building revealed shockingly different findings and recommendations[7]. While each audit recommended prioritizing 5-10 energy conservation measures, only 4 of the same measures appeared on all three auditor's lists and none were prioritized in the same manner. This tendency of physical energy audits has led building managers to question the economic benefit and ultimately refrain from mobilizing their company to conduct energy audits[6–9].

1.1.2 Building Information Modeling and Simulation for Energy Audits

Building Information Modeling (BIM) for energy retrofits has been explored as a more suitable option for efficiency diagnostics[10]. These physics-based models, such as EnergyPlus, BLAST, DOE-2.1E, TRNSYS-TUD, and ESP-r among others[11–13], can assess energy efficiency but require thousands of inputs[14]. Users must define the characteristics of the building envelope, HVAC equipment, interior space, and occupancy levels, as well as the hourly exterior temperature and solar insolation histories to accurately model the energy consumption and operating characteristics of the building[11,15]. Estimating the potential energy savings of a building retrofit using physics-based simulation models is incredibly time consuming, since it requires the building characteristics to be precisely defined. The resulting electricity consumption is calibrated with the historical utility data and fitted through various refinement parameters in the simulation[16,17]. Despite the massive inputs and detailed information, the building simulations have shown large inaccuracies and vary in energy recommendations[18,19]. Various parameters, such as

the solar irradiance gain or the convective heat transfer coefficient, of which can vary up to 30% from actual values, frequently deviate and result in serious accuracy limitations[12,13]. Therefore, trust from building managers and owners, even in the most input intensive energy audit solutions, remains questionable.

1.2 Data Analytics and Virtual Energy Audits

Data analytics applied to buildings have been used in the past to measure the energy savings associated with building retrofits[20] and efficiency programs.[21–23] Most notably, PRISM (PRInceton Scorekeeping Method), released in 1986, was one of the earliest applications of using regression analysis to measure energy savings in commercial buildings.[24] Utilities, companies and government agencies used the statistical procedures of PRISM to analyze monthly utility bills to provide a weather-adjusted analysis of energy consumption before and after building retrofits. The U.S. Department of Energy's EnergyStar Portfolio Manager built on the PRISM approach to include other datasets such as occupancy, plug load and other variables and develop a rating system for buildings associated with their energy efficiency.[25] ASHRAE also developed the Inverse Model Toolkit (IMT) for a similar purpose,[26] which was later extended by Kissock et al. in 1998[27], 2003[26] and 2011[28].

Data analytics is a growing field of interest, and due to advancements in processing, data storage, communication and analytics (such as distributed computing), there are new opportunities to use a more rigorous approach to uncover insights to building energy efficiency[29]. For this work, data streams from buildings and other sources are used to explore relationships among independent and dependent variables, uncovering

correlations, patterns, and anomalies and revealing unique characteristics and behavior associated with specific buildings. This approach assumes that all of the building information is inherently included in the historical energy (and other readily available) data, and consequently, these building characteristics can be uncovered.

Various techniques have been used to study historical data including anomaly detection, clustering, and time series analysis[30–34]. Anomaly detection techniques have been used to determine various irregular behaviors[33] and detect anomalous consumption in real-time[31,32], however much of the energy savings potential does not lie in anomalous behavior, but in systematically poor operation[35]. Clustering techniques, although limited, give some insight into the typical operation and allow for classifications into various usage groups useful for categorizing expected loads in particular conditions[30]. Time series techniques, such as spectral or Fourier analyses, have been implemented to determine underlying components of building energy[36]. Although spectral analysis is useful for determining significant frequencies in the energy data[37], many of these frequencies are already known - season, week, day, etc. Additionally, spectral or Fourier analyses may fail to quantify the behavior within each period, however this can be overcome utilizing a classical time series decomposition approach.

A time series decomposition approach essentially smooths the data to uncover underlying trends that may be useful for prediction, for determining typical behavior, or for developing an initial understanding of building performance. Classical decomposition uses the method of moving averages to examine trend cycles and seasonal behavior[38]. The trend cycles capture the long term behavior, while seasonal aspects are revealed from the periodic fluctuation of the data. Unexplained events and uncertainty must also be assessed to fully represent the behavior of the data as described by an error

term or random component. This work focuses on classical time series decomposition to identify operational characteristics of various buildings and point to opportunities for energy savings. Classical time series decomposition techniques have been used in a variety of applications including to study economics, weather, energy, and other time varying phenomena,[39–42] but with limited application to buildings. Current classical time series decomposition analyses on building energy focus on load forecasting models, but does not assess building systems or efficiency measures[43,44]. These applications of time series decomposition on building energy data give insights into periodic tendencies as well as random behavior that were not otherwise noticed. This work augments this type of insight by conducting classical decomposition to disaggregate the data into trend, seasonal, and random components to further uncover unique building characteristics[38].

Identifying building characteristics further opens the door to building energy disaggregation, arguably the most pivotal barrier to building energy efficiency improvements[45]. Building energy disaggregation is the process of extracting component loads from the aggregate building energy consumption and it continues to be investigated extensively in the literature, providing critical insights to performance for energy efficiency recommendations[46]. Current building energy disaggregation research focuses on residential buildings, using high frequency sub minute to sub second electricity data[46–49]. Commercial buildings have largely been ignored due to their complexity and lack of granular and sub-metered data[45]. These solutions have also focused on using multiple electricity data streams without considering other potentially useful data sets. However, one study has leveraged the use of occupancy behavior data sets along with hourly electricity consumption data to disaggregate energy consumption, though on less complex

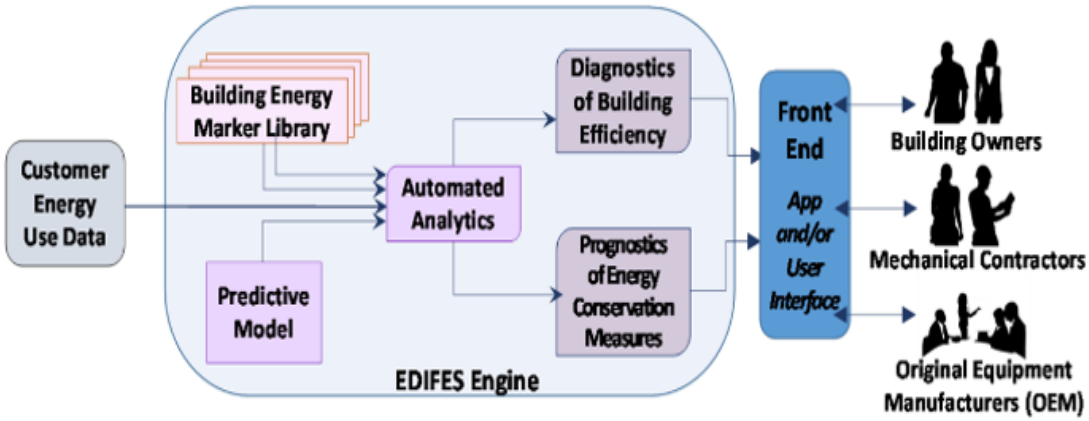

Figure 1.1. *EDIFES Development and Business Model. Consumer Data is ingested, analyzed, and used to enhance EDIFES building energy marker library. This Model appeals to three customers, Original Equipment Manufacturers (OEMs), Mechanical Contractors, and Building Managers.*

residential buildings[45,50]. This work proposes the significant use of independent data sets such as location, temperature, and irradiance data to allow for commercial building disaggregation of coarse, 15-minute interval electricity consumption data.

1.2.1 EDIFES

This thesis focuses on scalable data analytics for virtual energy audits through the development of the software platform EDIFES (Energy Diagnostics for Efficiency Savings). EDIFES uses a rigorous approach to data analytics to demonstrate unique information about building characteristics and operation without setting foot in a building. The development and business model for EDIFES is described in Figure 1.1. EDIFES ingests new building data, applies analytics, and classifies the building through markers contained in the building energy marker library. Following the analytics, diagnostics are determined on the building's energy efficiency and prognostics are computed for energy conservation measures. EDIFES has been developed to appeal to three primary

customers: Building Owners, Mechanical Contractors, and Original Equipment Manufacturers (OEMs).

EDIFES is currently a package developed in the *R* language for statistical computing and comprises over 40 functions. The functions are grouped into various categories, such as cleaning and formating, processing, analytics, and reporting. Cleaning functions are developed to ingest multiple troublesome data sets and merge them into one useful data set primed for analysis. Processing functions perform elementary statistics and create interim predictors to enable further, and more in depth analytics. Analytics functions are the largest and most impactful grouping. These functions compute various identifiers of the buildings through data analysis, which are stored and noted in the building energy marker library. The functions utilize complex analytics, such as anomaly detection, system identification, time series decomposition, and disaggregation. Ultimately, the functions lead to the quantifiable energy savings potential for various energy conservation measures and building retrofits. The sections within this thesis describe the methods of these functions in detail, their structure, and the layers of analysis.

This work was done in collaboration with Mohammad Hossain, Jack Mousseau, and Elle Zadina. Mohammad assisted with this work throughout the entire project and more specifically contributed to early data, derivative, and classical time series decomposition analyses. The structure and work flow of EDIFES was largely to Jack's credit. He had also contributed numerous testing functions and usage capabilities bringing EDIFES to a robust software. Much of the GIS data handling and cleaning was created by Elle Zadina, who also contributed significantly to early data analysis and derivative analysis for system identification.

2 Early Data Analytics

2.1 Data Characteristics

The analysis presented herein describes the operations of six commercial buildings in two different locations: San Jose, California and Richardson, Texas. Table 2.1 describes each building by its location, size, type, climate classification[51], and HVAC characteristics. The building data was collected from utility electricity meters (kWh) taken at 15-minute intervals and measured as a function of time for approximately 2 years. Hourly weather data was collected from publicly available National Oceanic and Atmospheric Administration (NOAA) datasets (within 25 miles from building location), and privately held 30 minute interval GIS datasets (within 200 meters of building location) were also employed[52]. Both NOAA and GIS weather data sets are used interchangeably in this analysis on the premise that EDIFES will employ the use of both data sets depending on building location and data availability. The weather data files include: temperature, dew point, wind speed, global horizontal irradiance, and others.

Building	Location	Size (sqft.)	Purpose	Climate	Electric HVAC
Building 1	Richardson, TX	226000	Office/Lab	Cfa	Heating
Building 2	Richardson, TX	168000	Office/Lab	Cfa	Heating/Cooling
Building 3	Richardson, TX	244000	Office/Lab	Cfa	Heating/Cooling
Building 4	San Jose, CA	109000	Office	Csb	Cooling
Building 5	San Jose, CA	115000	Office	Csb	Cooling
Building 6	San Jose, CA	168000	Office	Csb	Cooling

Table 2.1. Building characteristics of all six buildings. Includes location, size, purpose, climate, and electric HVAC

2.2 Data Preprocessing

The electricity, NOAA, and GIS data files are all received in various formats, requiring substantial formatting processes. Electricity data is received as a comma separated values file (CSV) containing two columns, local timestamp and electricity consumption, in MWh. Table 2.2 shows a sample of the data as read.

Key.Performance.Indicators.	Utility.Electricity.Consumption..MWh.
2012-07-01T00:00:00-04:00	0.04
2012-07-01T00:15:00-04:00	0.04
2012-07-01T00:30:00-04:00	0.04
2012-07-01T00:45:00-04:00	0.04
2012-07-01T01:00:00-04:00	0.04
2012-07-01T01:15:00-04:00	0.04

Table 2.2. Raw Building Electricity Data

These fifteen minute data are computed via time integration. Over a 15-minute interval the power data, or load, of a building changes frequently. As it does, the load is integrated and the integration is reported for the entire interval. Consequently, the power load may take various shapes and values within the 15-minute interval, but result in the same energy consumption value. Figure 2.1 shows an example of these occurrences. This poses a problem in determining when pieces of equipment turn on or off. For example, if one of two equal sized equipment turned on, while the other turned off at

approximately the same time, the corresponding energy consumption would then simply report zero, masking the actual behavior of the building. A second characteristic of the time integration is the *carry-over* of an event from one interval to the next. Consider a 100W piece of equipment that turns on at the seven and half minute of a 15-minute interval. The integration would lead to a 12.5Wh change in energy ($100W * (7.5/60)hr$). The following 15-minute interval will have the same 100W piece of equipment on for the entire interval of 15-minutes, which in turn would result in an increase in consumption of an additional 12.5Wh. Figure 2.1 displays the *carry-over* inherent in the time integration of electricity consumption.

Considering the time integration constraints, knowing the meter resolution helps determine the smallest piece of equipment that could be detected in analysis. The function meter_res() calculates an individual meter resolution. Outputs for all data reports appears in Table 2.3. This table shows, for Buildings 1-3, equipment under 1.8kW may not be detected, and for Buildings 4-6 significantly smaller equipment may be ideniti-fied.

	Meter Resolution (W)
Building 1	1800.00
Building 2	1800.00
Building 3	1800.00
Building 4	40.00
Building 5	120.00
Building 6	90.00

Table 2.3. Meter Resolution for all Six Buildings

The NOAA weather data is received as a text file and is converted to a CSV file containing the columns timestamp, in coordinated universal time (UTC), temperature, dew point, precipitation, and others. Values are recorded every hour at a particular minute,

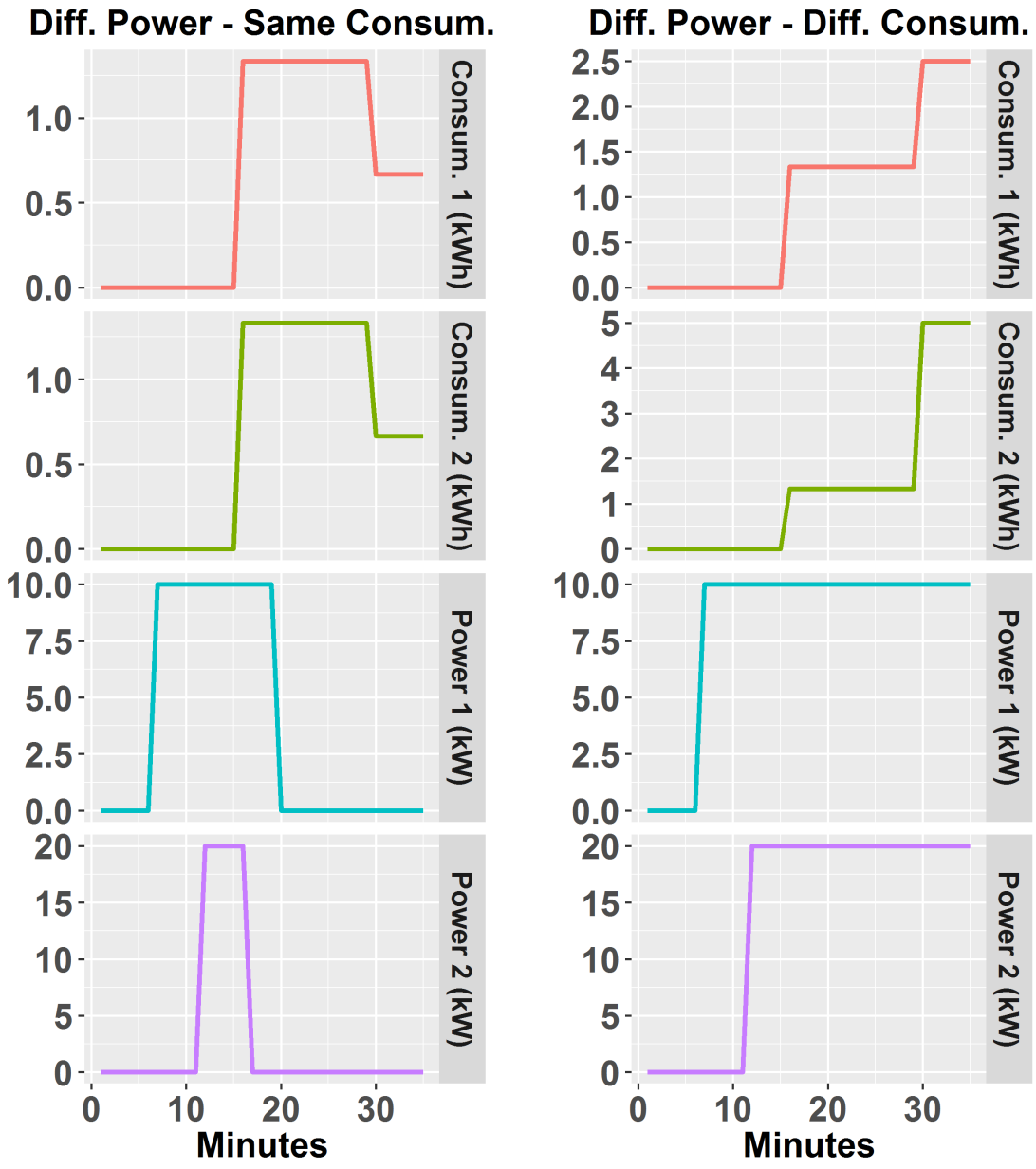

Figure 2.1. *Two Examples of Power and Consumption Curves. Left: Different power loads and same consumption. Right: Different power loads with different consumption.*

although datasets vary their specified recording minute. The datasets also include various daily statistics recorded with the appropriate time stamp. For example, the maximum temperature is recorded in the data at the time it occurred. This is troublesome as

each day will include 96+ data points, recorded at various unpredictable times, therefore requiring code to determine the correct minute for which to look for hourly recordings and omit the rest. There are also other issues with these data sets, including incomplete data, requiring further cleaning prior to analysis. Tables 2.4, 2.5, and 2.6 display some of the raw NOAA data.

USAF	WBAN	YR..MODAHRMN	DIR	SPD	GUS	CLG	SKC	L	M
745090	23244	201201010000.00	***	0	***	722	CLR	*	*
745090	23244	201201010056.00	330	6	***	722	CLR	*	*
745090	23244	201201010156.00	10	9	***	722	CLR	*	*
745090	23244	201201010256.00	360	7	***	722	CLR	*	*
745090	23244	201201010356.00	***	0	***	722	CLR	*	*
745090	23244	201201010456.00	***	0	***	722	CLR	*	*

Table 2.4. Raw NOAA Weather Data - Columns 1-10

H	VSB	MW	MW.1	MW.2	MW.3	AW	AW.1	AW.2	AW.3	W	TEMP	DEWP
*	10	**	**	**	**	**	**	**	**	*	61	39
*	10	**	**	**	**	**	**	**	**	*	59	41
*	10	**	**	**	**	**	**	**	**	*	57	44
*	10	**	**	**	**	**	**	**	**	*	56	44
*	10	**	**	**	**	**	**	**	**	*	54	44
*	10	**	**	**	**	**	**	**	**	*	53	43

Table 2.5. Raw NOAA Weather Data - Columns 11-23

SLP	ALT	STP	MAX	MIN	PCP01	PCP06	PCP24	PCPXX	SD
1021.7	30.16	******	63	51	*****	*****	*****	*****	**
1021.7	30.16	1020.1	***	***	0	*****	*****	*****	**
1021.8	30.17	1020.5	***	***	0	*****	*****	*****	**
1021.7	30.16	1020.1	***	***	0	*****	*****	*****	**
1021.4	30.16	1020.1	***	***	0	*****	*****	*****	**
1021.5	30.16	1020.1	***	***	0	*****	*****	*****	**

Table 2.6. Raw NOAA Weather Data - Columns 24-33

The GIS weather data is received as a CSV file containing the columns timestamp, in local standard time; temperature; global horizontal irradiance; relative humidity; and

others. In comparison, the GIS data is relatively clean and simple. The GIS data is determined using a sophisticated climate model, that employs satellites to determine, through calculation, the temperature, wind speed, solar irradiance, relative humidity, etc. The data is precise within 200m, but may contain uncertainties inherent in the model. In contrast, the NOAA data has a higher accuracy due to direct measurements, but the distance from the building location poses its own uncertainty and error. Tables 2.7 and 2.8 show the raw GIS weather data.

Date	Time	GHI	DNI	DIF	flagR
01.07.2012	00:15	0	0	0	0
01.07.2012	00:45	0	0	0	0
01.07.2012	01:15	0	0	0	0
01.07.2012	01:45	0	0	0	0
01.07.2012	02:15	0	0	0	0
01.07.2012	02:45	0	0	0	0

Table 2.7. Raw GIS Weather Data - Columns 1-6

SE	SA	TEMP	WS	WD	RH	AP
-29.47	-179.08	15.90	1.70	248	92.80	1013.90
-28.97	-171.20	15.80	1.60	249	92.30	1013.40
-27.66	-163.52	15.80	1.50	249	91.90	1013.10
-25.62	-156.15	15.70	1.40	250	91.90	1013.00
-22.89	-149.21	15.30	1.30	250	92.40	1013.00
-19.56	-142.73	14.60	1.20	251	94.30	1013.10

Table 2.8. Raw GIS Weather Data - Columns 7-13

The weather data obtained for this analysis is critical for understanding a building's response to outdoor weather conditions. In order to perform such analysis, the weather and electricity data sets must be merged by timestamp. Remember all three data sets differ in time zones and recording intervals. To create continuity among the data, each

weather observation was imputed (with linear interpolation) across the hour or half-hour to create 15-minute interval data. Due to the extreme inconsistencies between data sets on daylight savings days, they were omitted from the analysis.

The functions developed in EDIFES to automatically perform the above processing tasks include, bee_format(), noaa_addition(), gis_clean(), and gis_bee(). Additionally, the function day_data() creates columns that indicate the day of week, weekend/weekday, business days, and components of the date-time (Day, Month, Year). Function sun_tagging() creates a column indicating the sunrise and sunsets times for each day of the year given the latitude and longitude of the building, the function also creates interim predictors describing the night and day time hours. Lastly, the function fullday_tag() ensures 96 data points are present for each day following the merging and initial tagging functions. This is necessary to ensure accurate time series analyses of constant periods can be conducted.

2.3 Early Data Analysis

2.3.1 Data Cleansing

Initial analysis was performed on all of the datasets to ensure understanding of the data types, as well as to identify potential issues. All datasets, underwent graphical and statistical analysis to identify potential anomalies. In particular, utility electricity data is well known to be incomplete, noisy, and anomalous, recording non-physical quantities or none at all at times. Figure 2.2 shows the raw electricity consumption for the six buildings over two years, 2012-2014, with the exception of Building 6. By visually comparing the data, multiple characteristics of the electricity consumption appear. Buildings

1-3 have relatively compact operation when compared to their large baseload - these buildings also maintain a constant base-usage and peak-usage regardless of the season. Buildings 4 and 6 show higher variability in usage compared to their baseload, but show differences in seasonal operation. Building 4 maintains a constant base, but varies in peak loading, while Building 6 varies in base usage and peak usage throughout the seasons. Building 5, is entirely different, exhibiting large loads that render the visual useless. These many differences among the data sets, even before analysis, demonstrate the utility of a population-based analysis of building energy. Figure 2.3 shows one full week of raw data for each of the buildings. The first five increases in usage indicate the first five weekdays, Monday-Friday, followed by Saturday and Sunday, of which shows little usage when compared to weekdays except in the case of Building 6. The third day in the figure is July 4th, displaying the decrease of energy usage on a holiday as compared to normal workdays, yet still consumes much more than a weekend. Buildings 1-3 show similar behavior in usage, as well as general similarity day to day, while Buildings 4-5 show increased variability. Finally, Building 6 presents a smooth operation unlike any of the other buildings.

In Figure 2.2 Building 5 exhibits anomalous behavior with a large spike toward the end of the dataset, as well as two raised sections of electricity consumption. Using an anomaly detection *R* package[53] developed by Twitter, we can identify and then remove or fix such anomalies in the dataset as necessary. Figure 2.4 shows the percent of anomalies in the Building 5 data set (before and after the cleaning). There are two types of anomalies that appear in these figures. The first is an extreme outlier, seen as the spike, which is simply removed and noted using the function mean_shifts(). This function also detects mean shift anomalies which occur when the data has moved from one trend to

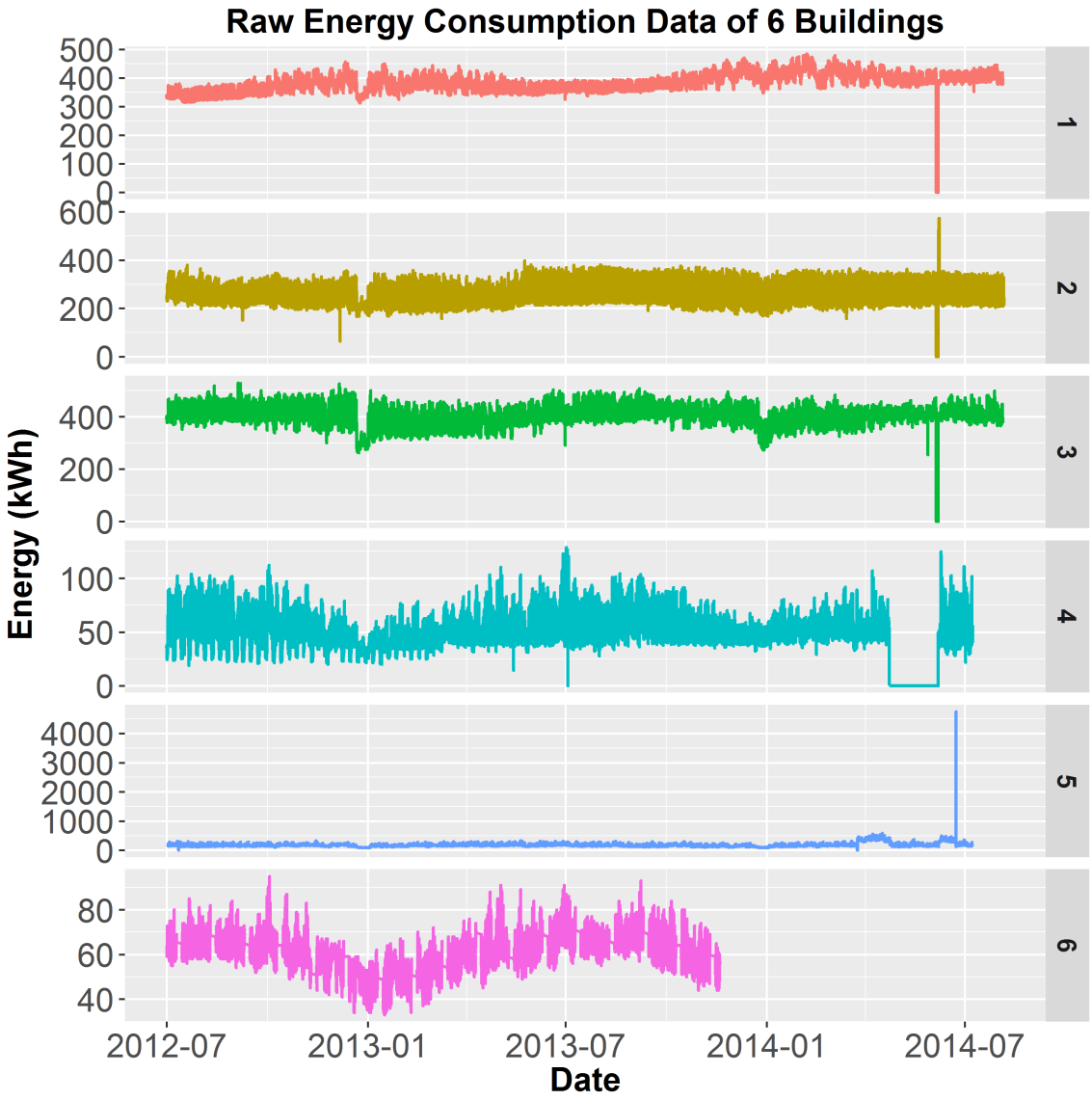

Figure 2.2. *Raw Energy Consumption of full datasets for all 6 Buildings. Buildings 1-3 show high baseload consumption, Building 5 shows the presence of outliers, and all buildings exhibit missing data.*

a new and distinctly different trend. For Building 5 the data has randomly doubled in value, later to be found an error in the meter that double counted the usage over this time period. Using the developed function mean_shifts() the two shifted data ranges are determined and fixed to show correct usage.

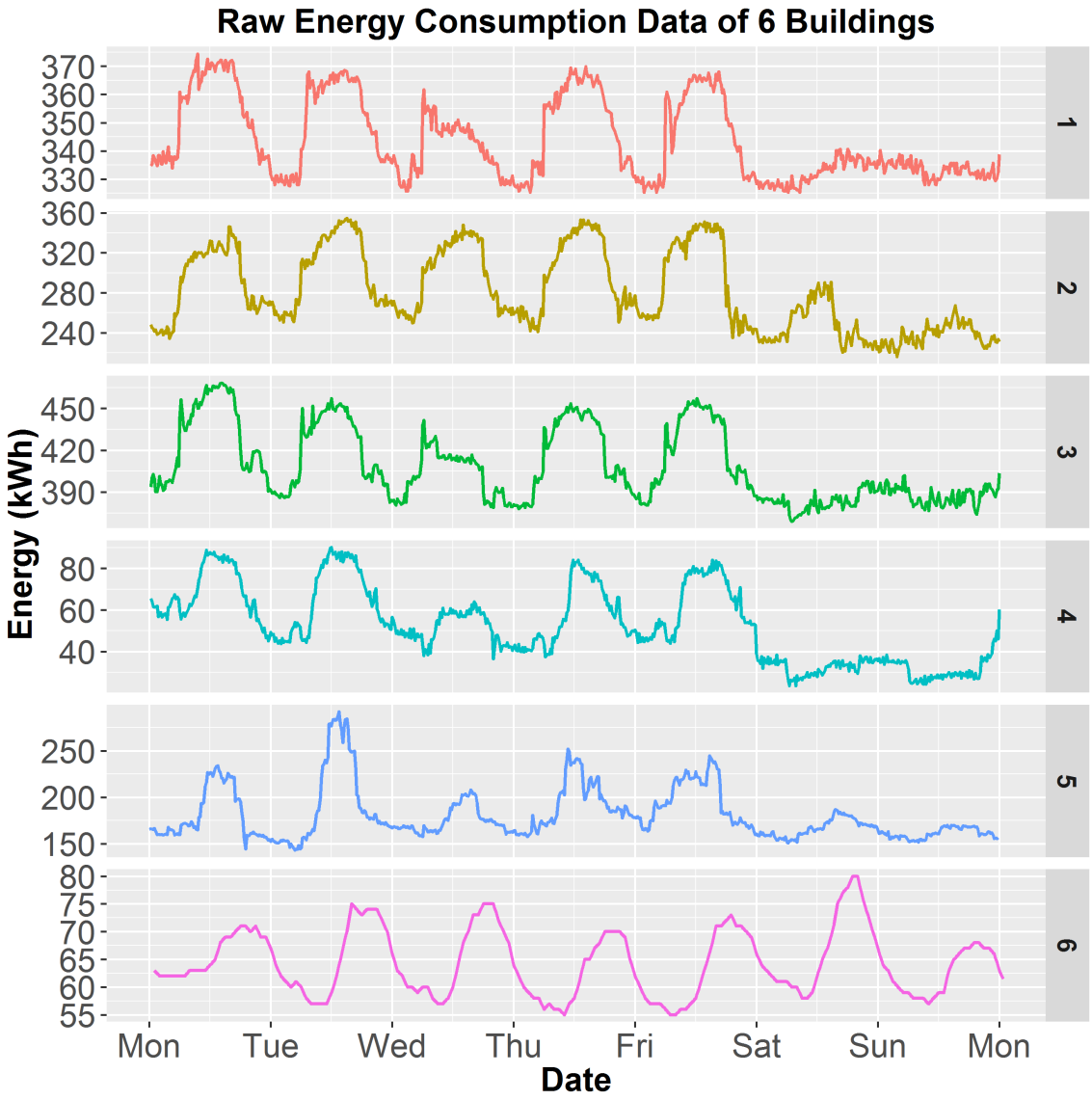

Figure 2.3. *Raw Energy Consumption of one week for all 6 Buildings*

A second building electricity function is used to assess missing data points in electricity consumption. The function missing_data() is used to impute missing data points (with linear interpolation) for intervals less than one hour. In the event that a meter loses connection for an hour, the rest of the day's electricity usage is still worthy of analysis. This step allows additional days to be analyzed, despite minor data losses. Those

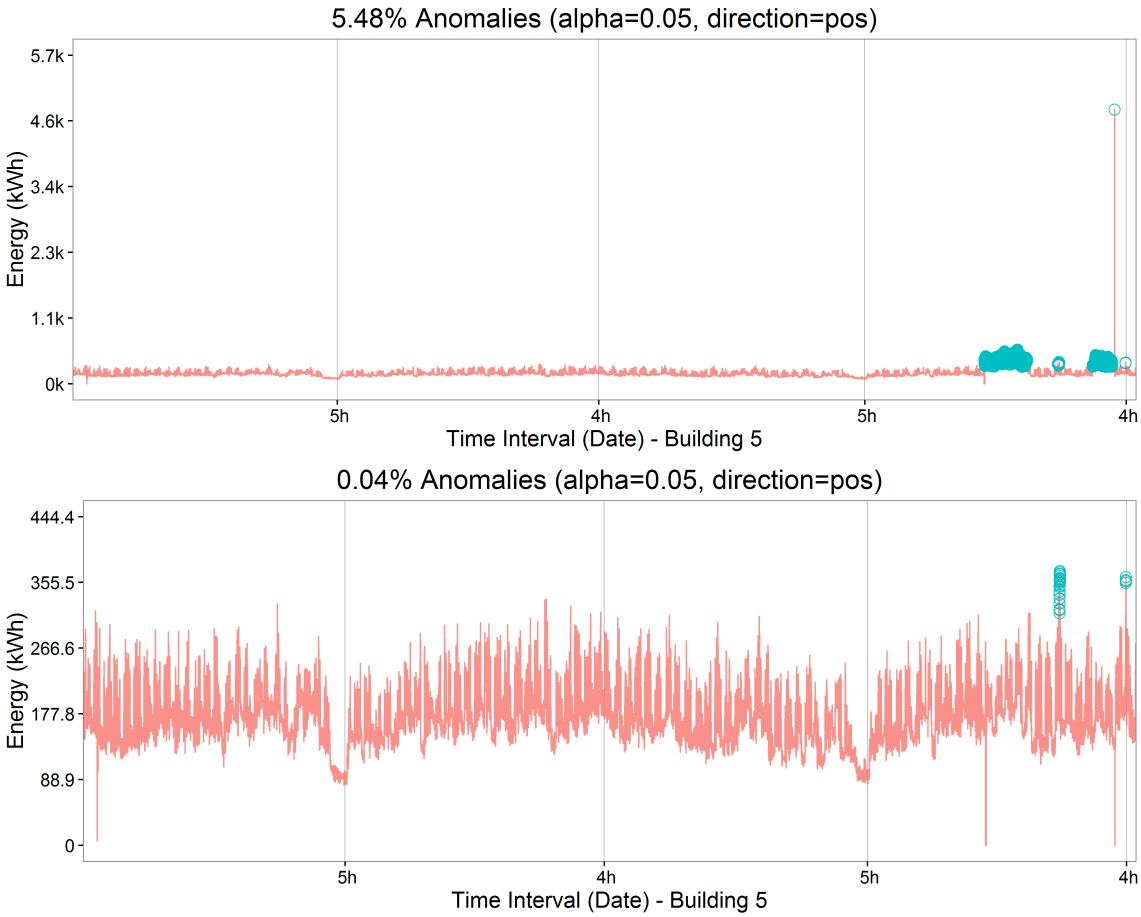

Figure 2.4. *Building 5 Raw Energy Consumption Data (5.48% Anomalies) and Cleaned Energy Consumption (0.04% Anomalies) after missing data and mean shifting functions*

days which exhibit data losses at an hour or larger are omitted from analysis to ensure imputed values are minimal. Figure 2.4 shows Building 5 with 5.48% anomalies before cleaning and with 0.04% anomalies following the use of both cleaning functions. Figure 2.5 shows all six buildings after data cleansing through the functions missing_data() and mean_shifts(). All cleansed data report anomalies of less than 0.5%.

NOAA and GIS weather data are gathered for both locations in Richardson, TX and San Jose, CA. NOAA weather data is collected from Moffet Airfield for San Jose, CA and Dallas Fort Worth International Airport for Richardson, TX. GIS weather data is collected

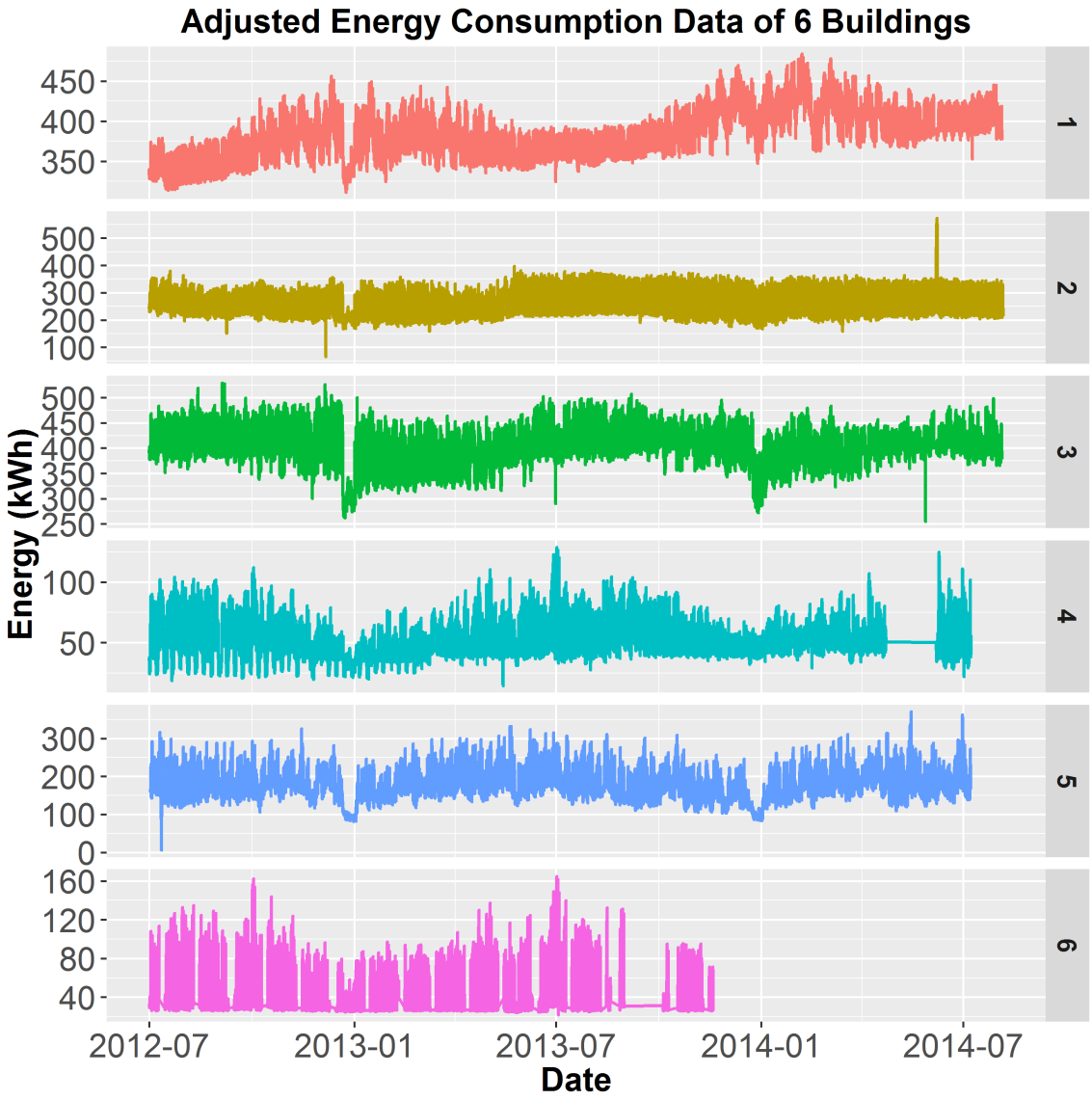

Figure 2.5. *Cleaned Energy Consumption of full datasets for all 6 Buildings*

from exact latitude and longitude coordinates of the buildings to an accuracy of 200m. The weather data includes complete observations for the time interval of interest, July 2012 - July 2014 and requires no additional cleaning. Figure 2.6 displays the two years of temperature data (Fahrenheit) from NOAA for both locations and Figure 2.7 displays the same two year interval temperature data for GIS datasets (Celsius). Comparing the

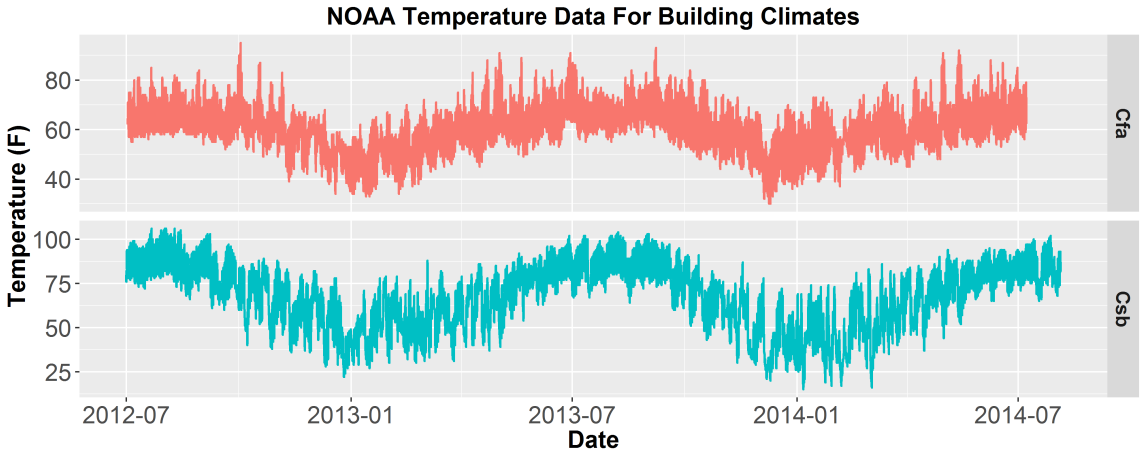

Figure 2.6. *NOAA Temperature Data in Climates: Csb- Richardson, TX (Buildings 1-3) and Cfa-San Jose, CA (Buildings 4-6)*[51]

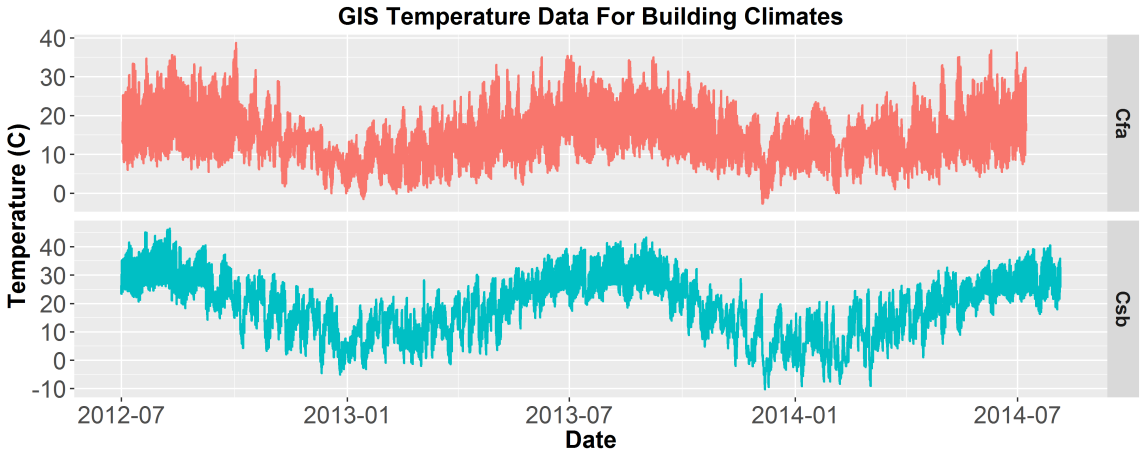

Figure 2.7. *GIS Temperature Data in Climates: Csb- Richardson, TX (Buildings 1-3) and Cfa-San Jose, CA (Buildings 4-6)*

NOAA and GIS data sets the trends of the data are unsurprising similar and indicates the use of either data set as satisfactory, however the amplitude of the data for each is noticeably different. NOAA presents data which has less daily variability overall, while the GIS data shows a much larger fluctuation. The climates are also distinctly different with San Jose showing milder temperatures throughout the data, 95F - 30F (35C - -1.1C), while Richardson varies from 106F - 15F (41.1C - -9.4C).

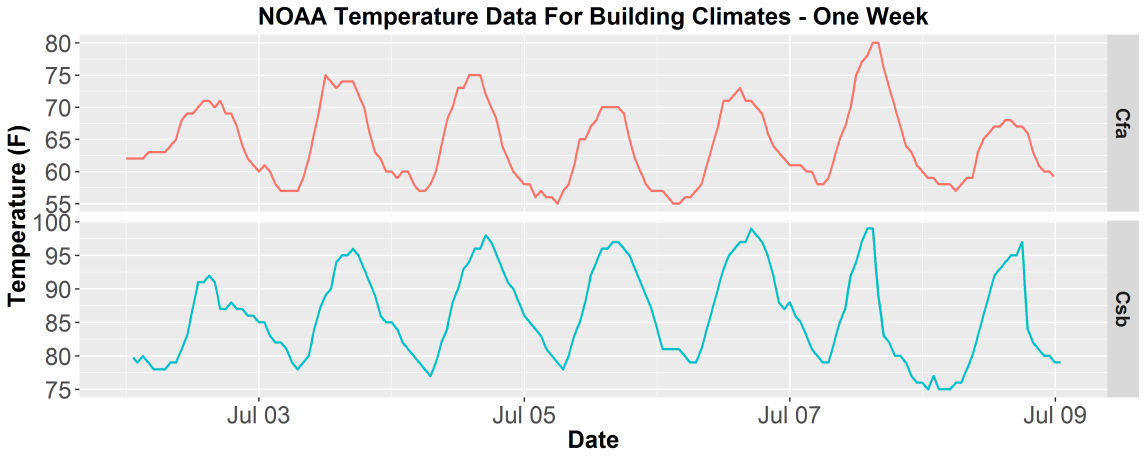

Figure 2.8. *NOAA Temperature Data for One Week in Climates: Csb- Richardson, TX (Buildings 1-3) and Cfa-San Jose, CA (Buildings 4-6)*

Figure 2.8 and Figure 2.9 exhibit a week view of the data to indicate the typical characteristics seen in the temperature data. The amplitude difference between the two data sets is apparent, as well as the difference in climates. Richardson has a larger range in temperatures in an average day from 100F-75F (37.8C - 23.9C), while San Jose ranges from 75-55F (23.9C - 12.8C). There is also a noticeable difference in the shape of each data set between the NOAA and GIS data in Figures 2.8 and 2.9. NOAA data appears sharper and jagged when compared to the GIS data, which is smooth and rounded. This can be attributed to NOAA's direct measurements in temperature compared to the GIS model's calculations. Regardless, both temperature data sets present similar values and trends for each day further justifying the use of either data set in analysis.

Figure 2.10 displays the entire datasets of global horizontal irradiance over two years with peak values corresponding to midday irradiance levels and nighttime is represented by zero values. The peak values of the data change seasonally for both climate zones, with Richardson ranging from 1000 W/m^2 - 600 W/m^2 and San Jose from 1000 W/m^2 - 500 W/m^2. The density of each plot points to the relative amount of sunny and cloudy

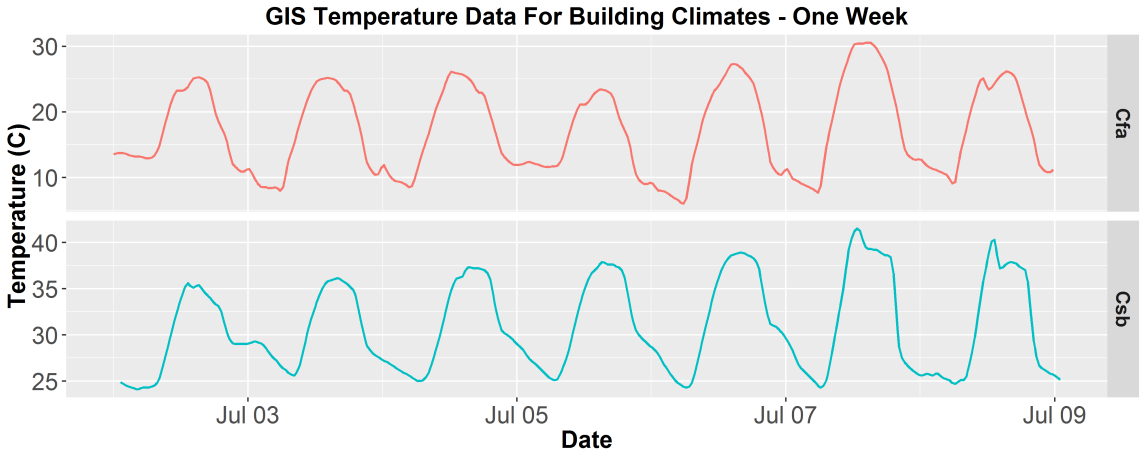

Figure 2.9. *GIS Temperature Data for One Week in Climates: Csb- Richardson, TX (Buildings 1-3) and Cfa-San Jose, CA (Buildings 4-6)*

days in each climate zone. Comparing the two plots, Richardson has more dips from the seasonal peak of the data compared to San Jose, indicating that San Jose's climate encounters fewer cloudy days. Figure 2.11 provides a week long snapshot of irradiance data showing how irradiance changes throughout each 24 hour cycle. Similar findings from the previous figure are further represented. Midday values exhibit the peak values of irradiance while night is shown as zero. San Jose data has virtually the identical daily trend (sunny), while Richardson shows two partly cloudy days.

2.3.2 Exploratory Data Analysis

For most buildings, the largest fraction of energy comes from heating ventilation and air conditioning (HVAC) [1], which responds largely to the outdoor weather characteristics. Therefore, understanding how a building is affected by weather is critical. To determine the relationship between weather and electricity consumption, correlations and scatter plots of electricity versus weather data were examined. Figure 2.12 shows a pair-wise correlation plot corresponding to Building 1, created with the function panel.pairs()

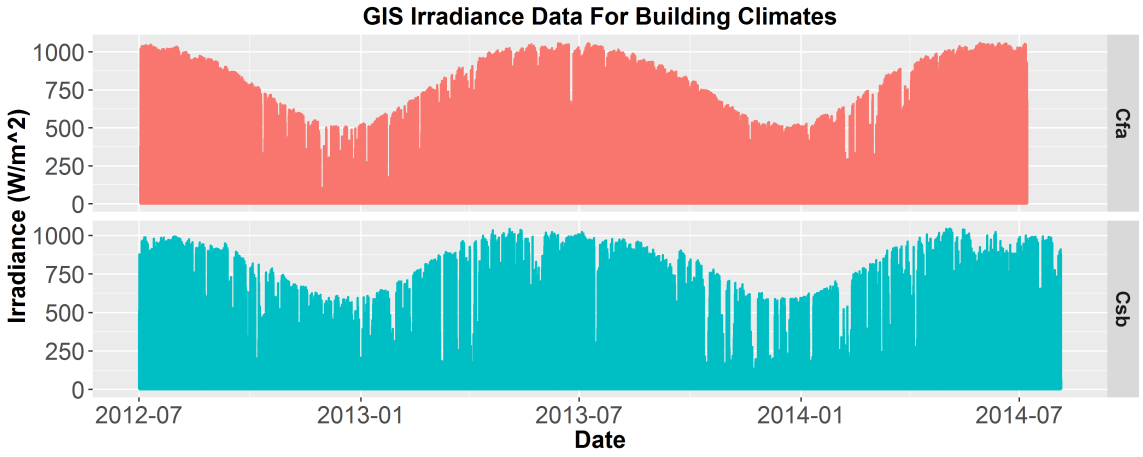

Figure 2.10. *GIS Irradiance Data in Climates: Csb- Richardson, TX (Buildings 1-3) and Cfa-San Jose, CA (Buildings 4-6)*

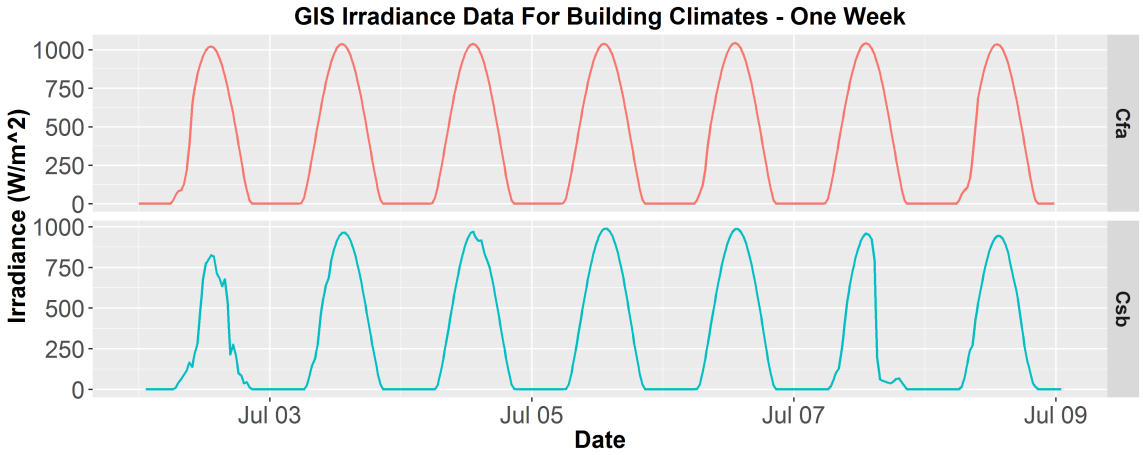

Figure 2.11. *GIS Irradiance Data for One Week in Climates: Csb- Richardson, TX (Buildings 1-3) and Cfa-San Jose, CA (Buildings 4-6)*

from the *R* package *psych*[54]. A pair-wise correlation plot displays scatterplots of two variables and the linear correlation between them. In this case, the data sets are: electricity consumption, exterior temperature, irradiance and relative humidity. Figure 2.12 shows the correlations of energy to temperature as -0.44, energy to irradiance 0.15, and relative humidity 0. All three of these correlations are surprisingly low, with the highest being temperature and energy, of which still shows a weak to moderate (less than

0.67 magnitude[55]) negative correlation, indicating the building electricity load increases with decreasing temperature. Irradiance and relative humidity show very weak correlations at 0.15 and 0 values respectively. Again this is surprising as irradiance values are known to effect the thermal load of the building and humidity varies the specific heat capacity of ambient air, both resulting in responses of electricity loads from HVAC. Note that these values are analyzed at the exact same time interval, i.e. the electricity consumption at a specific time in the day is being compared to the exterior temperature at that same time. A lower than expected linear correlation may be explained by the fact that: 1) plug load use in the building has no relationship to outside temperature (one will use their desk computer regardless of summer or winter temperatures); 2) the building has a large thermal mass and will take time to respond to changes in ambient outdoor temperature; 3) non-electric HVAC systems are present and/or the heating/cooling is provided via another building or a district energy configuration such as is typical for a campus; and 4) building set point temperatures change in occupied and unoccupied states. To account for these issues a cross correlation analysis can be utilized. The function ccf()[56] determines the best lag, or relative shift, of each data set in time to find the highest correlation. Figure 2.13 shows a maximum correlation of -0.5 at a lag of 31 data points, or seven hours and 45 minutes. Despite the higher value, the correlation still exhibits a relatively weaker relationship than expected. The other parameters, irradiance and relative humidity also increased marginally with this analysis.

Considering the lagged response of electricity consumption to weather data, as well as the probability that occupancy plug load consumption hurts the correlation values, a mean day consumption and weather data approach was then considered. For each

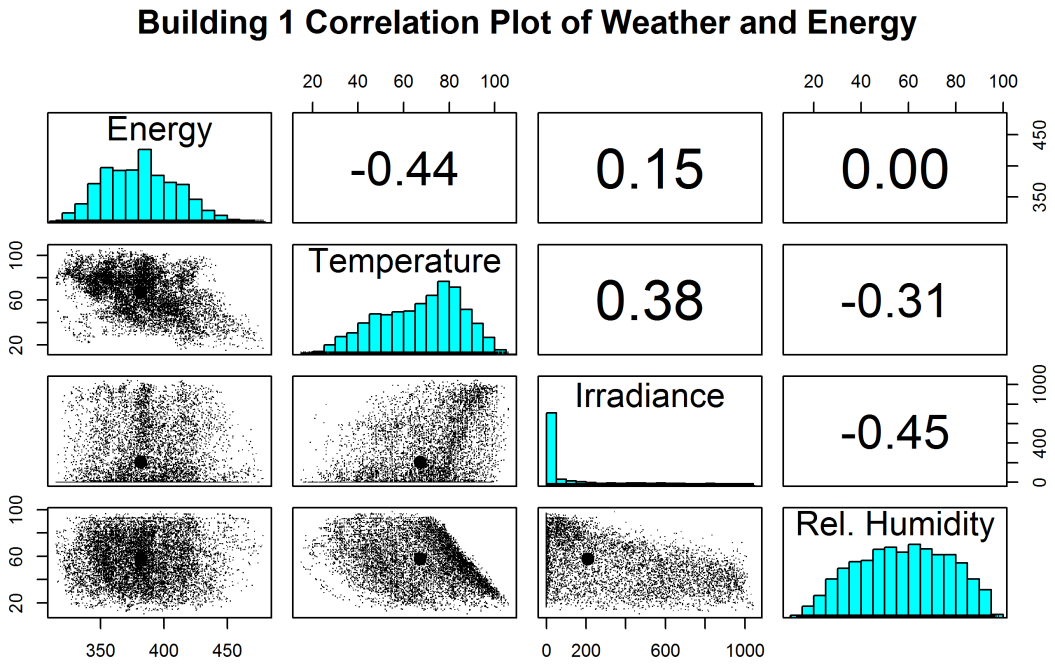

Figure 2.12. *Correlation Plot of Weather Data and Energy for Building 1*

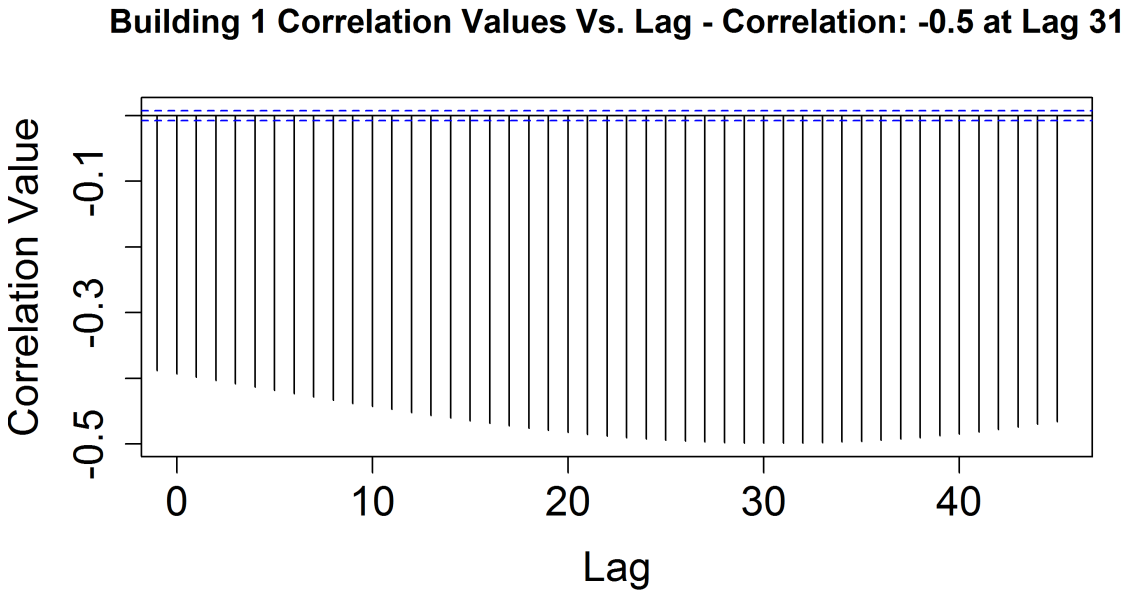

Figure 2.13. *Cross Correlation Plot of Energy and Temperature for Increasing Lag Values of Building 1*

day, the mean electricity consumption was calculated, as well as the mean temperature, irradiance, and relative humidity, along with other variables. To do this EDIFES contains the functions elec_noaa_stats() and gis_stats() to compute the daily mean values, along with other descriptive statistics, or interim predictors. Those statistics include minimum values, maximum values, day ranges, mean values of prior days, energy derivatives, temperature derivatives, and others. Here only the mean electricity consumption, temperature, irradiance, and relative humidity are used. Figure 2.14 shows a pair plot of the mean values of the same variables as previously presented for Building 1. By analyzing the mean values, much higher correlations are present: electricity vs. temperature increases (in magnitude) to -0.56, vs. irradiance to -0.27, and vs. relative humidity to 0.06. It is also worth noting that due to the negative association of energy and temperature, it also fits well that the same is true for irradiance. The negative correlation indicates that as temperature increases (thermal load increases) the electricity consumption decreases, and as irradiance increases (thermal load increases) the electricity consumption decreases. This behavior would be expected if only the heating (i.e. winter) season were being analyzed since building electricity consumption typically increases as temperature rises in summer, but here we know we are analyzing a full two years of data. So we may speculate that there is no cooling system in the building or that perhaps the building is in a climate where cooling is minimal. In this case, it was later determined through EDIFES, and then confirmed by the building manager, that Building 1 is connected to a district supply cooling facility and does not see any electrical loads in summer operation, but is heated electrically in the winter.

Although the daily mean values improved correlations, they were still weakly correlated. To further improve the correlation, further criteria for analysis was applied, such

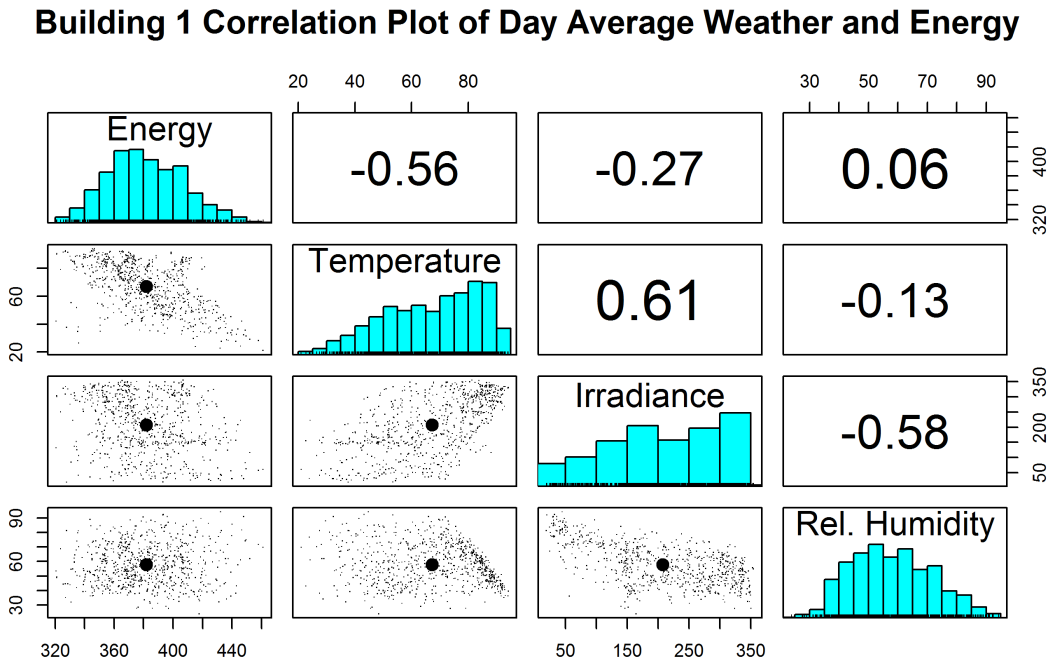

Figure 2.14. *Correlation Plot of Mean Weather and Energy Data for Building 1*

as day of the week. The EDIFES function day_data() was developed to associate each data observation with an interim predictor indicating various attributes such as day of the week, weekday or weekend, and business day. Considering the large difference in usage during weekends or weekdays, the analysis was further refined using only weekdays, Figure 2.15 shows the relationships and resulting correlations for Building 1 again. Through this refinement the temperature-energy correlation increases to -0.61 and the irradiance relationship to -0.28. Again, these values are improved, but still considered weak to moderate correlations. The same analyses were conducted on all buildings with all showing similarly weak correlations, thus demonstrating the complexity of building energy far past just a weather data analysis and the necessity for layers of refinement.

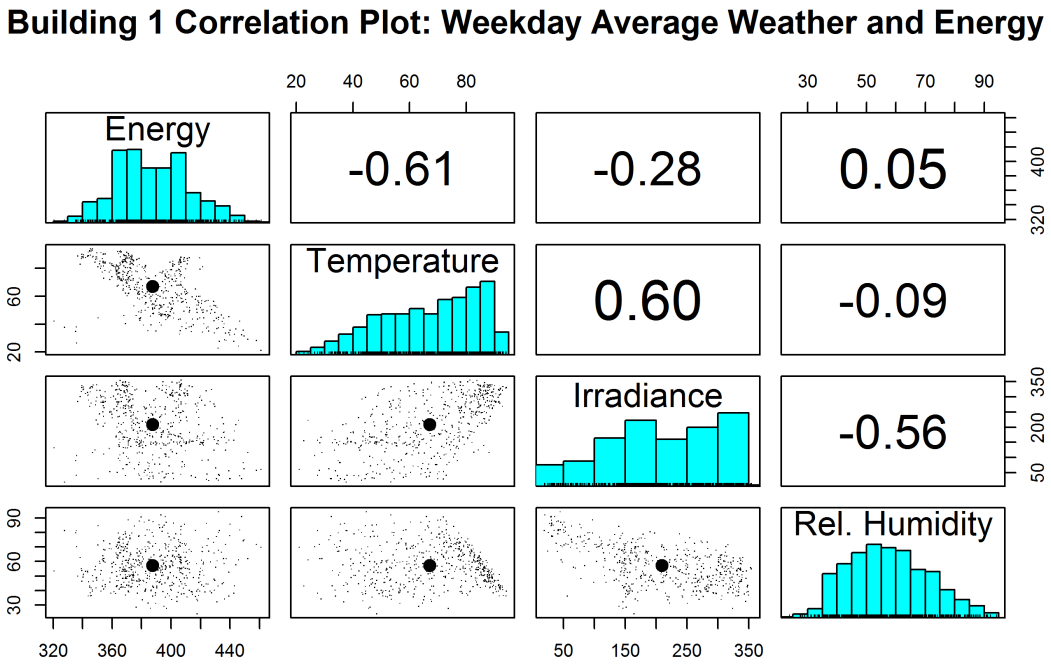

Figure 2.15. *Correlation Plot of Mean Weather and Energy Data of Weekdays for Building 1*

Since most buildings exhibit distinct behaviors in summer and winter, the analysis was further separated into two corresponding data sets. As a result, Building 1 electricity consumption increases with lowering temperatures in the winter, and decreases with increasing temperatures in the summer providing additional evidence that the building likely has electric heating (or there is some significant portion of heating system is electric) and non-electric cooling. The function created to explore this relationship, hvac_elec() plots the building and weather data, and then calculates a smooth spline of electricity consumption vs. exterior temperature, using the function smooth.spline[56] in *R*. The spline function fits a cubic spline to the data and with a smoothing parameter of 1, the spline shown in Figure 2.16 was created. Following the fit, the end points of the spline are taken and a linear line is created in order to determine the change point in the

data, or the point where the spline changes direction. To find this point, the distance from the line to the spline is calculated with the maximum distance indicating the temperature at which the building changes operation. This point is recorded and kept for later analysis. Using the change point the data is split into two, and each analyzed separately. The data is then fit to a linear model of a least squares regression with electricity consumption as the dependent variable (predictor) and temperature as the independent variable (response). The slope of the model is then taken and recorded from the function. The slope below and above the change point indicates the likely type of HVAC system used in the building. In cold weather a negative slope indicates electric heating, while a positive or very low slope likely corresponds to non-electric heating (natural gas or district heating). In hot weather the reverse is true, a positive slope indicates electric cooling, while a negative slope corresponds to non-electric or no direct cooling. Table 2.9 displays all of the outputs of the hvac_elec() function, including the change point temperatures, hot and cold slopes in $kWh/(F * sgft.) * 10^{-5}$, as well as the associated HVAC characteristic. Further pairs plots for each building are located in Appendix B, showing the pairs correlations when the data is additionally split along the change point temperatures.

Building	Change Point (F)	Hot Slope	Hot HVAC	Cold Slope	Cold HVAC
Building 1	70.50	-0.03	Non-Electric	-0.43	Electric
Building 2	54.50	0.70	Electric	-0.53	Electric
Building 3	51.00	0.48	Electric	-0.24	Electric
Building 4	61.00	1.51	Electric	0.38	Non-Electric
Building 5	61.75	2.32	Electric	1.43	Non-Electric
Building 6	54.25	1.06	Electric	0.43	Non-Electric

Table 2.9. HVAC Electricity Function Output for all 6 Buildings

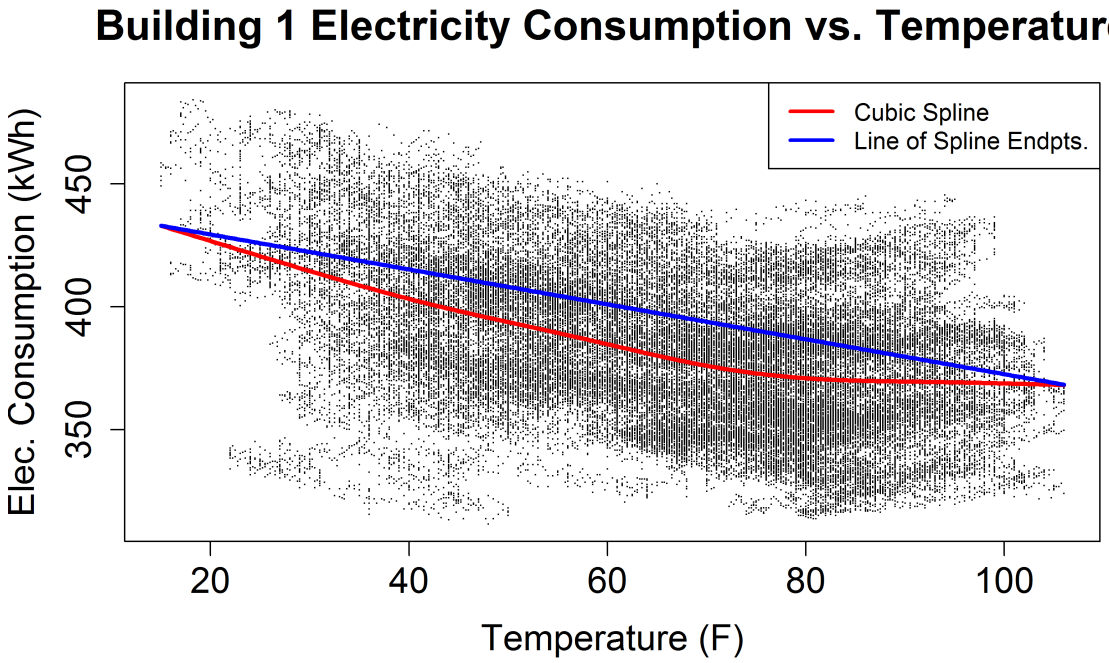

Figure 2.16. *Energy vs. Temperature with a fitted spline and line through endpoints, Building 1*

3 Derivative Analysis for System Identification

Taking the first and perhaps the second derivative of the data can prove insightful for a building's operation. This work focused on computing the standard deviation from the differentiated data and using it to examine exceptionally large temporal changes in energy usage. This analysis allows for building systems, such as HVAC, to be identified and their scheduling determined.

3.1 Method of Derivative Analysis

Due to the constant time interval of 15-minutes, the derivatives taken in this analysis are simply the difference in energy (or temperature, irradiance, etc.) of one time step from the previous. $\Delta Energy = Energy_i - Energy_{i-1}$

Figure 3.1 shows the derivatives of approximately one week of data from Building 1. Most of the data straddles 0 with small variations, but a few data points extend much farther by multiple standard deviations. Taking the standard deviation of the differentiated data and considering specific multiples of the standard deviations from the mean,

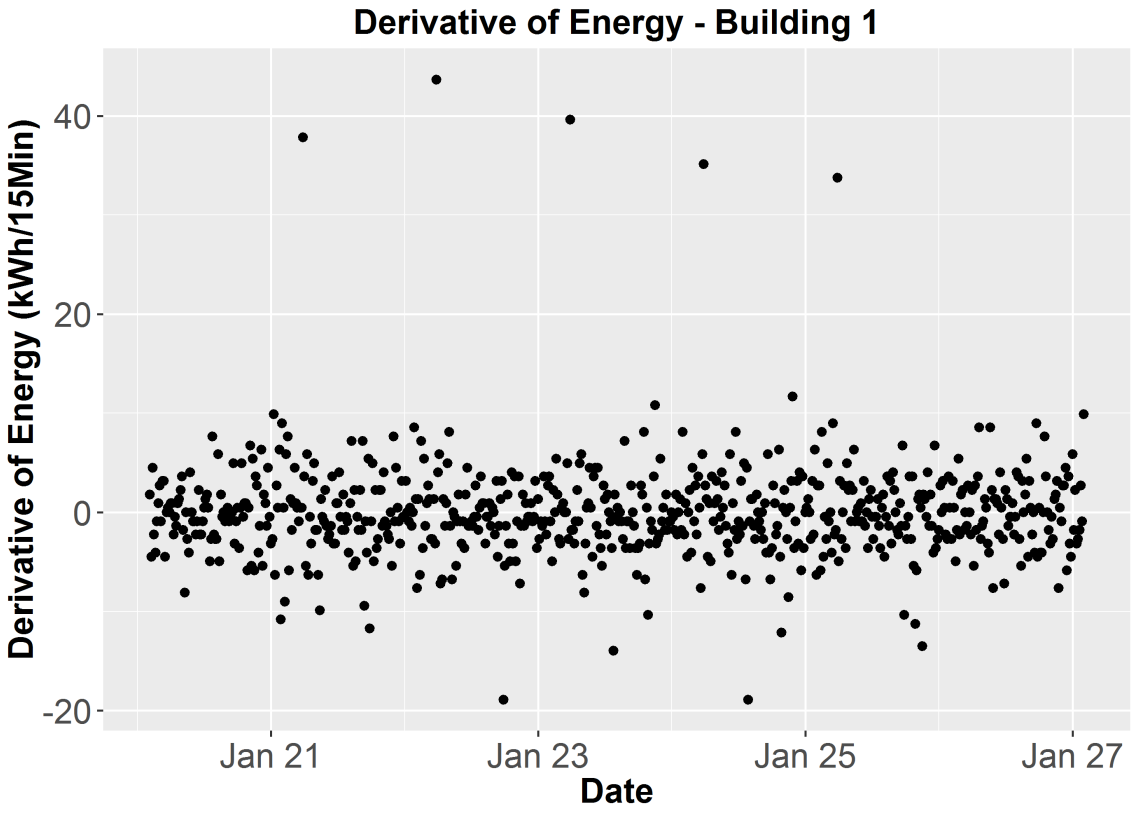

Figure 3.1. *Building 1 Energy Derivatives over One Week*

patterns begin to emerge. Assuming the data to have a normal distribution (true for n > 30 per the central limit theory[57] and n > 70,000 for two years of data), all data exceeding the mean plus one standard deviation comprises of only 15.8% of the data, and similarly for the mean minus one standard deviation[57]. Increasing the criteria to two standard deviations corresponds to 2.2% of the data. The occurrences that lie above or below the one standard deviation mark can be subsetted, or pulled separately into another dataset, and analyzed accordingly.

Pulling the data into two subsetted data sets reveals that the distribution is not quite perfectly normal, with a kurtosis value less than three, as values above the mean (0.0007

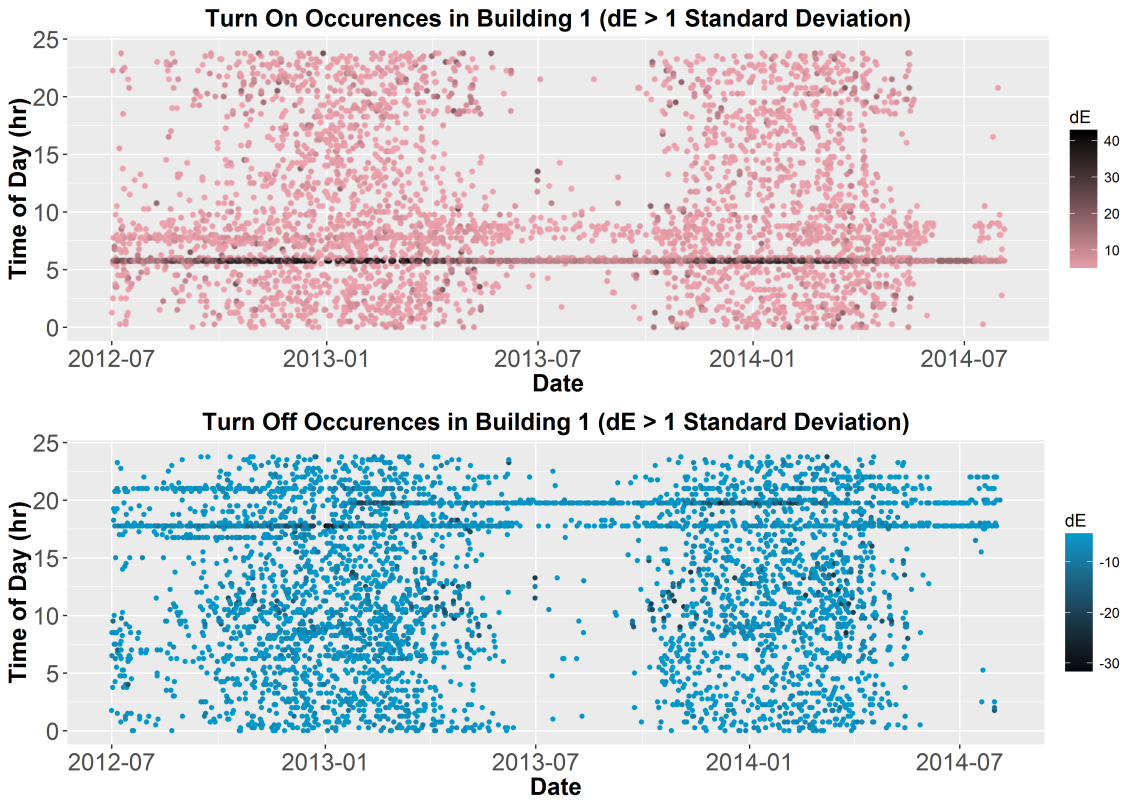

Figure 3.2. *Building 1 Turn On and Off Events, Standard Deviation* ±1

kWh) plus one standard deviation (4.1201 kWh) include just 5.36% of the data and below one standard deviation reports as 5.94%. This ensures that these events are quite rare and meaningful. After subsetting the values, they can be visualized by plotting each occurrence by the time of day it occurred vs. date of occurrence, as well as identifying the intensity of each event by a light to dark color-scale. As shown in Figure 3.2 light corresponds to a minor event and dark to a major event. The Figure depicts both the positive events in red and the negative events in blue. The positive events correspond to the "Turn On" events, due to the nature that they indicate a device, piece of equipment, or entire system has turned on, using a significant amount of energy. Similarly the negative events refer to the "Turn Off" events, or when significant equipment turn off.

The figure indicates one very large "turn on" event at 5:45 a.m. For the "turn off" times there are three distinct events at 5:45 p.m., 7:45 p.m., and 9:00 p.m., although all three of these are smaller in magnitude than the large turn on event in the early morning. These "turn on" and "turn off" events consistently occur over the course of two years and indicate that they are scheduled events of particular systems in the building. Further, the magnitude can be seen, especially in the "turn on" events plot, to vary throughout each year suggesting the systems are related to the HVAC of the building. These findings can help assess the scheduling of a building's large systems and recommend alternative scheduling and the corresponding savings that would result. For example, Building 3 has turn on and off events early in the morning 5:45 a.m. and late in the evening 9:45 p.m., (refer to Appendix B for tables and figures of Building 3 systems). By rescheduling the turn on and off events Building 3 has the potential to save 125 MWh ($8,800 at $0.07/kWh in Richardson, TX) per year by simply changing the scheduling to 7 a.m. - 7 p.m. It is important to emphasize that this energy and cost savings was calculated using the building's actual data over the course of the past year and the method of time series decomposition described in the next chapter. No computational simulations or assumptions were made as might have been required using conventional auditing methods. Certainly, weather will influence the actual savings into the future, but this type of analysis based on the building's actual energy data provides significant confidence in the results.

Though the plots give an insightful visual of the building's operation, this analysis is performed automatically with the EDIFES function system_finder(). The system finder function conducts the derivatives analysis and reports all systems which occur with a

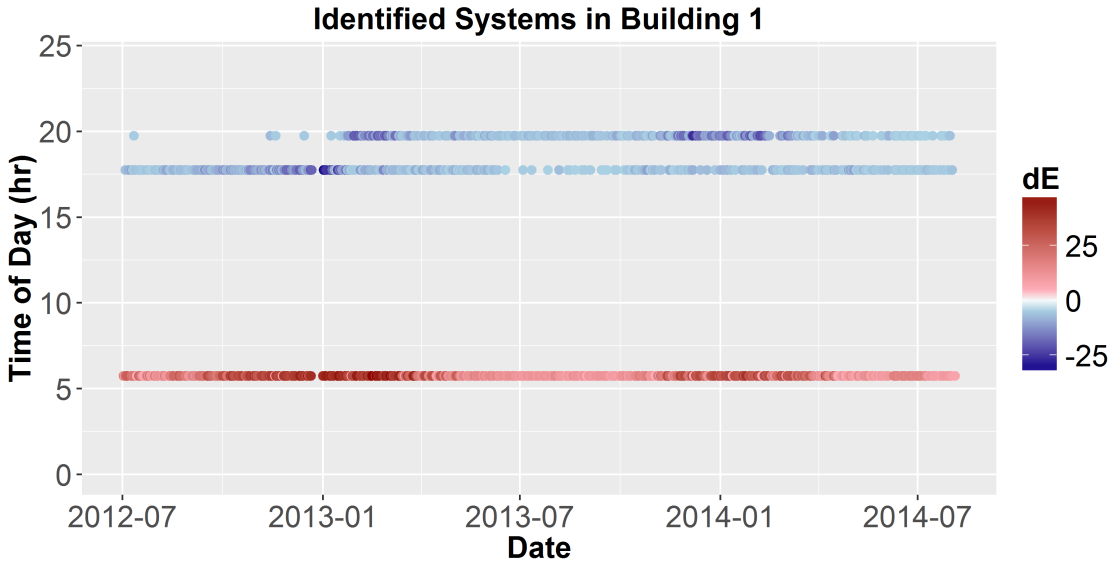

Figure 3.3. *Building 1 On/Off Systems*

frequency over 20%. Frequency is defined as the number of total occurrences of a specific "turn on" or "turn off" time divided by the total days of available data. Table 3.1 and 3.2 display the systems results that exhibit over a 20% frequency. Both tables report the amount of occurrences, time of day the event occurs, average ΔE , frequency, and the energy in Watts required of a piece of equipment or system to create the associated energy change over a 15-minute interval.

	Occurences	Time	Average dE (kWh)	Frequency	Energy (W)
1	553.00	00:05:45	22.09	0.73	12213.45

Table 3.1. Turn On Events for Building 1

	Occurences	Time	Average dE (kWh)	Frequency	Energy (W)
1	371.00	00:17:45	-7.91	0.49	-2933.55
2	317.00	00:19:45	-9.02	0.42	-2858.85
3	162.00	00:21:00	-6.41	0.21	-1038.60

Table 3.2. Turn Off Events for Building 1

All of the "turn on and off" events can subsequently be plotted to show a building manager or user the systems in their building, associated scheduling, and the magnitude of the equipment "turn on/off" as shown in Figure 3.3. This analysis allows one to determine how many systems may exist, whether they are operating in a sensible schedule, and also allow for the detection of certain lighting sensing systems, which will be further discussed in Chapter 5 - Disaggregation.

4 Time Series Decomposition for Building Daily Operation Signatures

Building electricity consumption data is collected via time step intervals, and therefore is classified as a time series data set that may be analyzed using a classical decomposition time series analysis. Classical time series decomposition provides a means for identifying operational characteristics of various buildings and points to opportunities for energy savings. Time series decomposition techniques have been widely used in many fields, as well as in building research, but only in load forecasting models, not in energy efficiency data analytics[43,44]. An application of time series decomposition on building energy data can provide insights into daily and seasonal operation signatures and contribute to further understanding of building operation that a derivative analysis fails to fully capture.

4.1 Method of Classical Time Series Decomposition

Mathematically the decomposition is described as: $X_t = f(M_t, S_t, Z_t)$ where X_t is the observed time series data, M_t is the trend component, S_t is the seasonal component

and Z_t is the error term, or random component. The additive model: $X_t = M_t + S_t + Z_t$ is used here to capture the seasonal variation changes through successive periods[38]. The variables M_t, S_t, and Z_t are developed through moving averages defined by the observed period of the data. Decomposition requires an input of period, P, to first calculate the trend component, M_t, using a 2 x P-moving average, if P is even, and a P-moving average, if P is odd. The detrended series, D_t is calculated as $D_t = X_t - M_t$ and then the detrended series is averaged over the period to compute the seasonal component, S_t. To complete the decomposition, the remainder, Z_t is calculated by $Z_t = X_t - M_t - S_t$. These three data sets represent a disaggregation of the data and each set provides new insights previously undiscovered. The seasonal component provides new information on daily operations, the trend indicates the underlying movement of electricity consumption, and the random component may pin point specific periods in the data displaying uncharacteristic operation. To perform the analysis the function, *decompose*[56], from the *stats* package in *R*, is used to conduct the classical decomposition method described above.

4.2 Classical Time Series Decomposition of Building Energy Consumption

Throughout this thesis classical time series decomposition is performed on two years of building data using a period of 24 hours (i.e. P = 96 points). Utilizing a period of one day allows for the seasonal, or periodic component, to reflect the average electricity consumption for a given day. Figure 4.1 demonstrates the entire decomposition method performed on summer temperature and electricity data from Building 1, displaying the

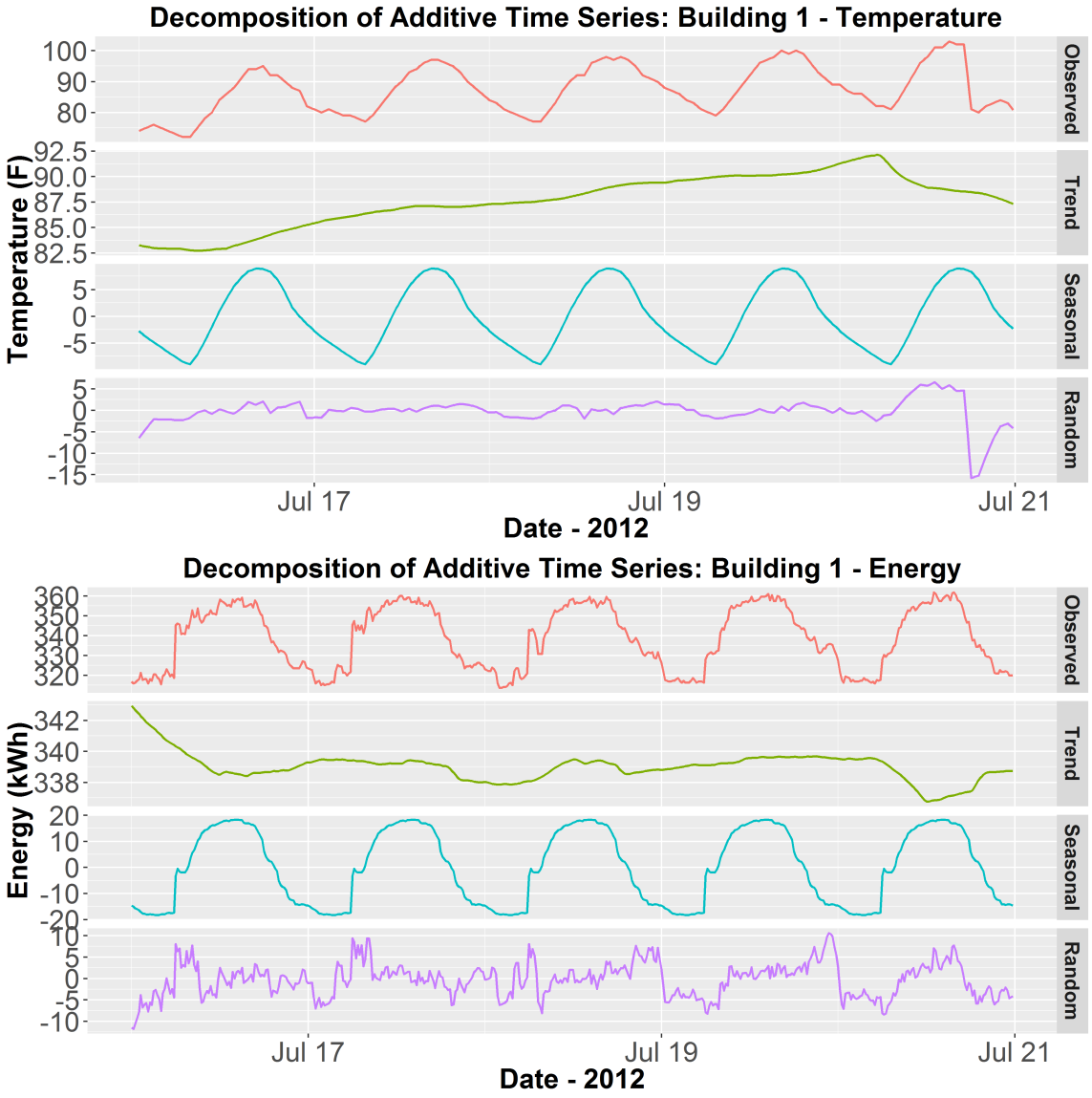

Figure 4.1. *Time Series Decomposition of Temperature, top, and Energy, bottom, for Building 1*

various components: observed, trend, seasonal, and random. The additive nature of classical decomposition is seen as the observed data can be replicated by summing the trend, seasonal, and random components. Beginning with the temperature data, it is commonly known that temperature follows a short term periodicity of one day. The seasonal component captures this typical temperature fluctuation throughout each day

and repeats this exact seasonal component fives times over. The trend component captures the overall magnitude and progression of temperature, while the random component displays any deviations in the observed data from the trend component and seasonal component data in aggregate. Moving to the electricity data, the exact same breakdown applies. The trend component represents the magnitude of the consumption, 340kWh, and graphically the trend component appears to move substantially, but has less than a 2% variability over the week interval. The random component exhibits many jagged features as it encompasses events unexplained by the daily periodicity, most likely unscheduled human interaction with plug load, such as lab experiments or non-HVAC equipment. Finally, the seasonal component is seen repeated 5 times over, again due to the periodicity, but it is this repeated seasonal component that provides the most interesting insights; the typical daily operation of a building.

Figure 4.2 shows one period, or one day, of the seasonal component of electricity data in Figure 4.1. This average electricity consumption can also be described as a building's typical daily operation, or daily operation signature. Figure 4.2 shows the daily building operation signature of summer weekdays gathered from over two years of data for Building 1, with electricity consumption along the y-axis and the hour of the day along the x-axis. The features of the curve indicate various usage characteristics and operational tendencies. As shown, there is an apparent spike in usage at approximately 6 a.m., indicating a scheduled HVAC event by the building management system - in alignment with the "turn on" time from Chapter 3, followed by a sharp decrease, showing the tendency for the HVAC units to overshoot demand and drop in usage - a characteristic of the building unidentified by the simpler derivative analysis. The building then undergoes a gradual increase in use from growing occupancy (plug load and

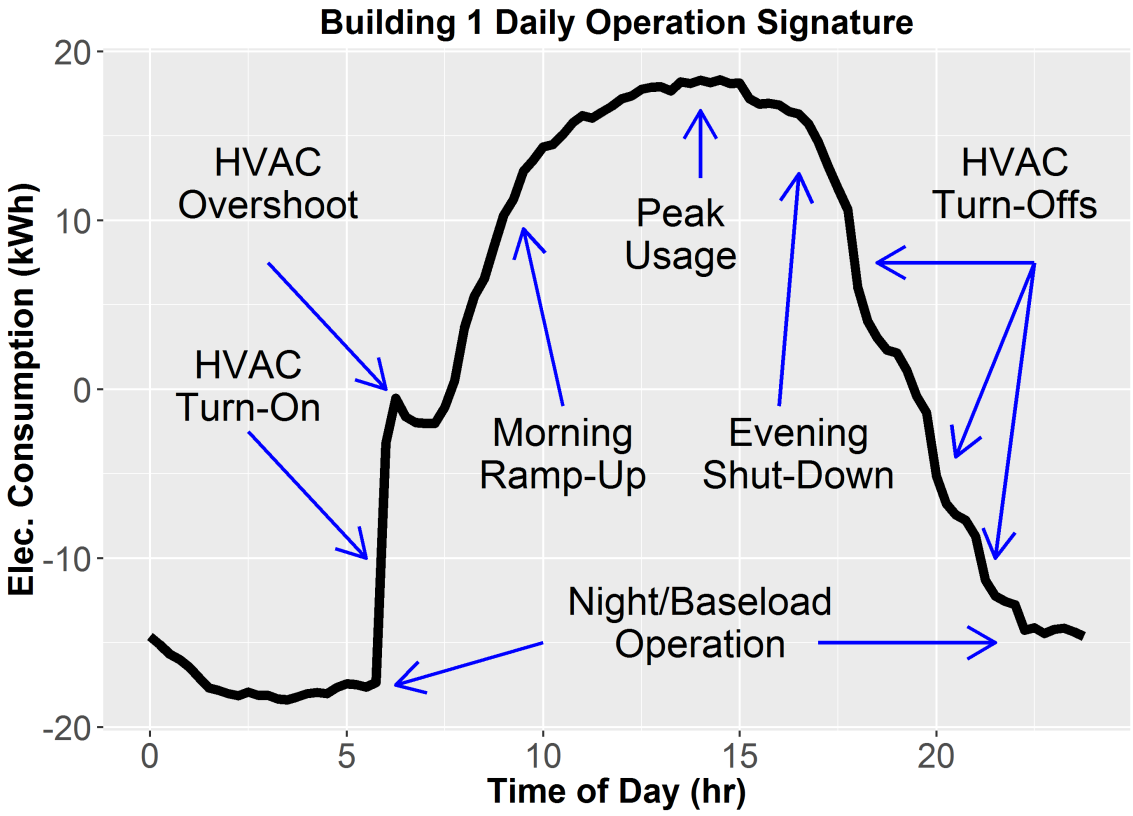

Figure 4.2. *Building 1 Daily Operation with Operational Characteristics Identified. Determined from two years of data considering only July/August weekdays.*

HVAC) peaking at about 1 p.m., and then a gradual fall throughout the rest of the day as occupancy decreases. Finally, three more significant drops occur at 5 p.m., 7 p.m., and 8 p.m. indicating more scheduled HVAC events, or "turn off" times.

4.2.1 Time Series Decomposition with Subsetting Criteria

Building operation depends on a variety of different predictors and associated responses. For example, the day of the week and the weather contribute to patterns and correlations that provide insight to building operation and characteristics. The method of subsetting allows for the division of the data into various groups based on defined criteria (e.g. day

of week) and provides an opportunity to compare various characteristics among the criteria. Figure 4.3 presents the energy consumption daily operation signature of Building 1 in kWh versus time of day after using a time series decomposition of all two years of data and subsetting criteria for day of the week. While applying subsetting techniques the continuity of the energy data is compromised. For example, in a day of the week analysis all Mondays do not represent a continuous data set. To account for this, each seasonal component is shifted so that time zero, midnight, starts at 0kWh allowing for accurate comparisons between the data. The plot shows the obvious difference in the weekend and weekday lines, as well as the differences between the weekdays, particularly Monday and Friday. The discrepancy on Fridays can be attributed to lower load requirements in this office/lab building due to reduced occupancy. Monday's increased consumption is the result of increased HVAC requirements following a weekend of reduced HVAC use and a lower weekend internal set point temperature.

Figure 4.4 expands this analysis to all six buildings to allow for further comparison and insight. Building 2 shows highest consumption on Monday, lowest weekday consumption on Friday and lowest weekend consumption on Sunday. There is also one distinct "turn on" period, 4:45-5:30a.m. and two distinct "turn off" periods at 5:45p.m. and 11p.m., furthering validating the derivative analysis findings. Most interesting however, is the weekend operation signature compared to the weekday operation signature. Building 1 displays various "turn on" and "turn off" events during the weekdays, but not the weekends. Building 2 however, does show similar HVAC events, although at a different magnitude compared to the weekday, indicating peculiar weekend behavior. This finding also indicates the size of a unit and its overall impact on the electricity consumption, showing a 30kWh load over 15-minute intervals corresponding to a 120kW

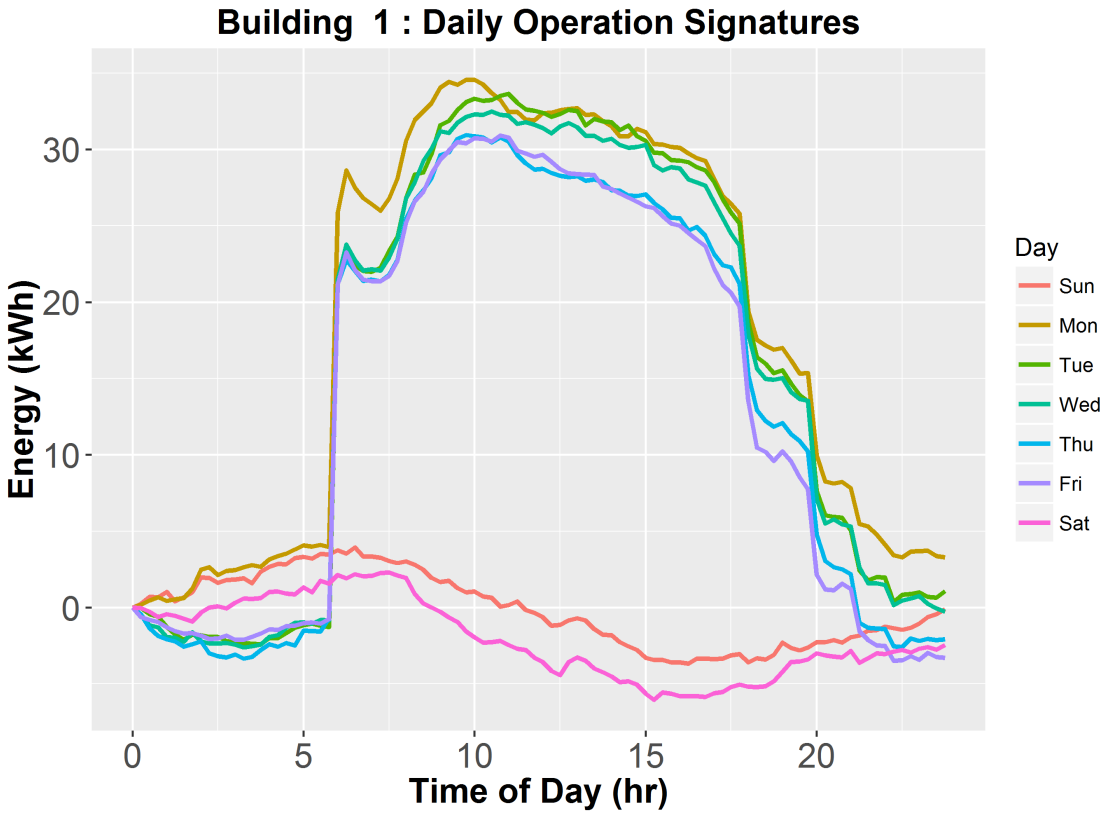

Figure 4.3. *Building 1 Daily Operation Signatures for Various Days of the Week*

unit. Considering the constant operation observed in the curve from 4:45 a.m. until 10:45 p.m. this piece of equipment consumes approximately 2160 kWh a day - a calculation that requires time series analyses to confirm electricity consumption succeeding a "turn on" event and preceding a "turn off" event. Derivative analysis provides identification measures, where time series allows for consumption quantification along with other subtler characteristics. Through the elimination of this HVAC event on the weekends, the building manager would save approximately 225 MWh, or about $16,000 annually (considering a $0.07/kWh rate). Furthermore, rescheduling all of these events, weekday and weekend, to typical occupied hours, 7a.m to 7 p.m. would result in another 258 MWh, or $18,100 annually.

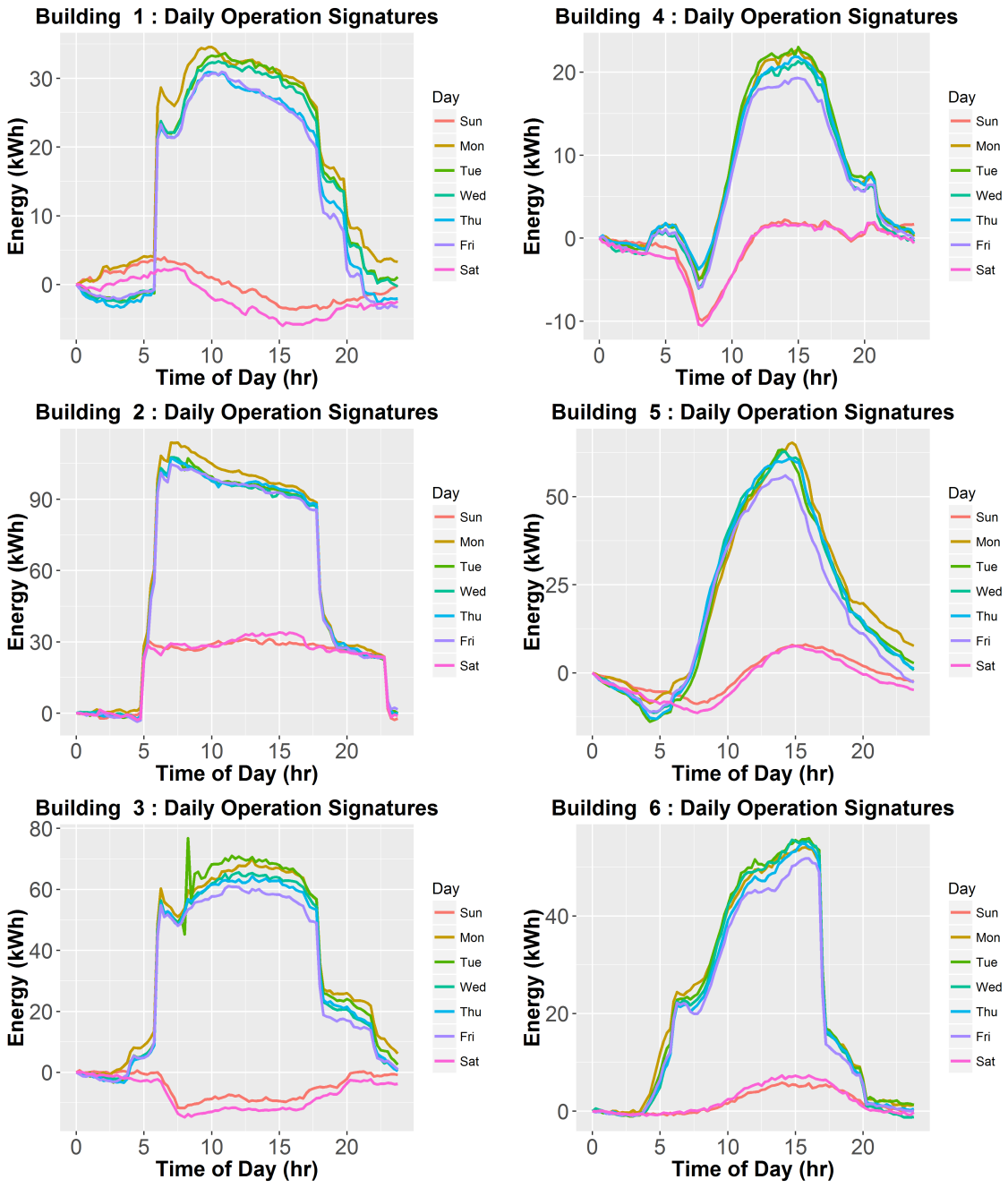

Figure 4.4. *Daily Operation Signatures for 6 Buildings Identifying Operational Characterisitcs for Each Day of the Week*

Building 3 shows similar characteristics as Building 1 and 2, including scheduled HVAC events and lower weekend/weekday usage. However, the most interesting characteristic in this building is the consumption spike at 8 a.m. for a brief period on Tuesdays

only, an uncharacteristic building consumption feature is observed and may require the attention from a building manager. The event does not affect consumption significantly, but could pose a demand issue considering the sharp increase and decrease of the event. Building 4 begins an analysis of the San Jose buildings, which operate in a milder climate and consist of office space only, compared to Richardson's lab/office combined buildings. Building 4 shows a much smoother operation signature, with small scheduled HVAC events at 4:00 a.m. and 9:15 p.m., along with dips at 7:30 a.m. and 7 p.m. These dips were later discovered to be the result of daylight sensing exterior lighting, turning on at sunset and off at sunrise. Specifically, exterior lighting events occur at differing times throughout the year and hinders the seasonal component, but can be identified and addressed using the derivative method and is described in depth in Chapter 5. Building 5 displays an extremely smooth operation signature throughout each day of the week, indicating efficient building operation, although a high usage into the late hours of the day shows a lack of unoccupied HVAC set point changes in the evening. The introduction of such set point changes would allow HVAC units to shut down in the evening, saving substantial amounts of electricity. Lastly, the daily operation signature of Building 6 shows an operation unlike any of the other building's in shape, but also has HVAC events from 4:30-6:00 a.m., 5 p.m., and 8 p.m. In summary, all of the building operation signatures subsetted by day of the week indicate various characteristics in each building and allow for quick, insightful, and quantifiable analysis.

The building data can be further subsetted into heating and cooling seasons, heating season corresponding to the months of December-February, while cooling season

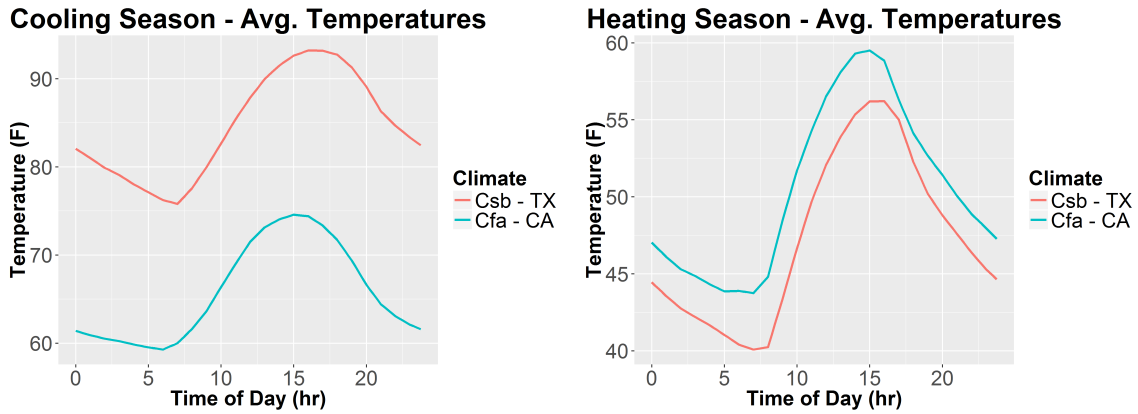

Figure 4.5. *Cooling and Heating Season Climate Temperature Differences - Richardson, TX and San Jose, CA*

includes June-August. Further included in the subsetting criteria are temperature constraints to ensure outliers are omitted, such as unusually warm heating days. The criteria is set using a full day (96 data point) mean temperature and the correct temperature ranges are computed separately for each climate. Figure 4.5 shows the large discrepancy between the two climates and the need for individualized temperature constraints. Richardson shows extremely high cooling season temperatures, as well as colder heating season temperatures, while San Jose displays a mild climate and a much smaller difference between cooling and heating season temperatures. Considering these values Richardson cooling season is defined as having a day mean temperature above 70F (21.1C) and heating season below 65F (18.3C), while San Jose cooling season is above 60F (15.6C) and heating season below 60F.

Through subsetting heating and summer seasons for Richardson buildings and computing the day of the week classical time series decomposition results in Figure 4.6 and reveals unique insights into cooling/heating season operation signatures. First, all heating and cooling season day operation signatures show drastically different shapes, cooling season having smooth and rounded features and heating season exhibiting largely

square features. Peaks in the cooling season months for all buildings occur at approximately 1:00-3:00pm, while heating season peaks occur at 6 a.m. These difference in times of peak usage with different temperature operation signatures strongly support the events are due to HVAC. This deduction was verified with the building manager. Cooling season operation signatures more gradually approach these peaks, while heating season shows a sudden jump in energy usage followed by a constant linear decline in usage throughout the rest of the day. All buildings also show an earlier turn on for Fridays in heating season operation signatures only, indicating a difference in the building's schedule for Fridays. Building 2, unlike Building 1 and 3, shows a much more constant usage in cooling season months, and also demonstrates a variable load during weekend operation signatures, indicating continued HVAC usage in weekend operation signatures. Overall each building shows a distinct difference in magnitude between cooling and heating season operation signatures. This difference indicates the relative energy requirements of the heating versus cooling systems in the building, and identifies the building as having electric heating sources due to the larger heating season magnitudes of usage. Additionally and as expected, during cooling season operation signatures, more energy is consumed during warmer times of the day (i.e. mid-afternoon), and similarly during the heating season, less HVAC is used during these mid-afternoon hours, also due in part to occupancy induced thermal load. Even more, this supports the peak usage time of 6 a.m. for heating season, the coldest time of the day, and 8 a.m., where morning arrivals lead to higher infiltration rates. Whereas the cooling season peaks are throughout the interval from 1 p.m. to 3 p.m., the warmest parts of the day. Interestingly, the HVAC also turns off in a staged manner in Building 1 during heating season months

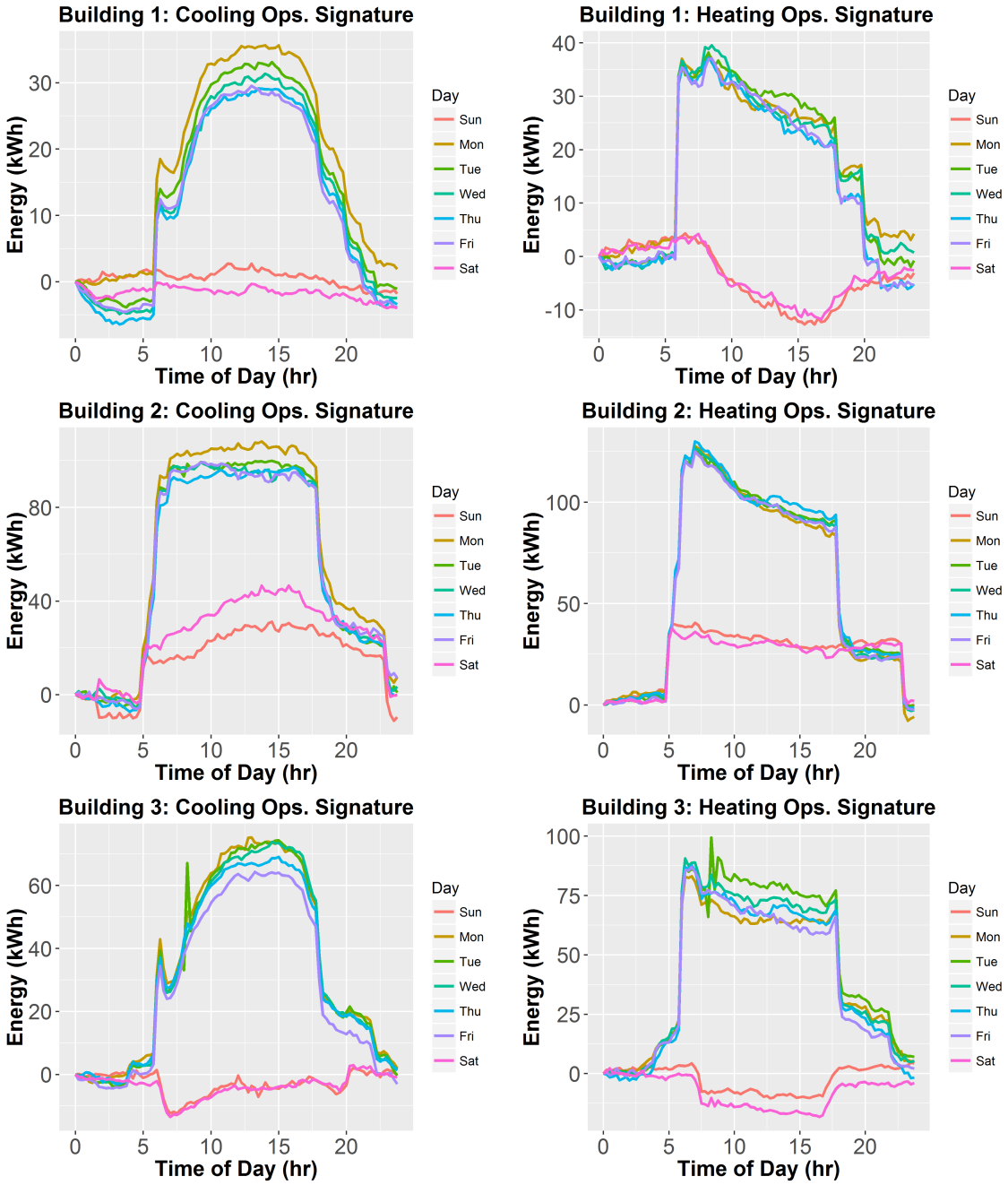

Figure 4.6. *Cooling and Heating Season Daily Operation Signatures for Richardson, Texas Buildings*

as indicated by the staged ramp down of energy consumption. This, too, was verified by the building manager.

Figure 4.7 continues the cooling/heating season daily operation signature analysis for Buildings 4-6. Building 4 shows a slight difference in shape, having less blocky features in heating season operation signatures, yet also more defined turn on and turn off HVAC events. Building 5 shows very similar operation signatures throughout, except for an evening turn on, necessary for colder temperatures heading into night hours. Building 6 displays a much smoother heating season operation signatures than cooling season, ramping constantly through the morning in the heating season. The building also shows a differing late turn off time from cooling to heating seasons, turning off at 8 p.m. in the cooling season and 7 p.m. in the heating season. Overall each building shows a much higher magnitude of consumption during the cooling season months, indicating non-electric heating and electric cooling systems, which was verified by the building manager.

To ensure that the previously described insights are significant, the uncertainty of each building time series data sets is assessed. In a classical time series analysis the random component of the decomposition captures the error in the analysis. For each decomposition, the uncertainty, μ, is computed by: $\mu = \sigma/\sqrt{(n)}$, where σ is the standard deviation of the random component and n is the number of days in the analysis. Table 4.1 reports the uncertainty for each building classical time series decomposition. In most cases the uncertainty is less than 2 kWh giving confidence that the insights of each signature are not the result of noise in the data. Those buildings with higher uncertainties also show relatively higher magnitudes in operation, therefore the relative uncertainty is still significantly low. These analyses are computed using two years of data and result in n values of approximately 200 days after subsetting. If the analyses

were computed on larger datasets, such as 4 years of data, uncertainties would be further minimized and confidence of building operation signature insights increased.

Building	Heating +/- kWh	Cooling +/- kWh
Building 1	0.61	1.46
Building 2	2.93	1.53
Building 3	1.17	1.76
Building 4	1.15	0.65
Building 5	1.56	1.73
Building 6	0.65	0.45

Table 4.1. Uncertainty Quantification for Each Building Time Series Heating/Cooling Operation Analysis

The heating/cooling season time series analysis is summarized in Figure 4.8 on all six buildings. Due to varying magnitudes of consumption, between office buildings and lab spaces, each building operation signature is shifted and normalized to values between 0-1. Additionally, each building is normalized by its largest operation signature in either heating or cooling season conditions, allowing for quick heating/cooling season comparisons. Each individual consumption data set provide multiple insights and could be studied at length, but a few comparisons are particularly striking. The first comes in the magnitude of the operation signatures, as described earlier, in cooling/heating season conditions. Buildings 1, 2, and 3 (Richardson buildings) show large consumption in heating season conditions, while Buildings 4, 5, and 6 (San Jose) consume most in cooling season operation signatures. Again, this indicates electric heating among all Richardson buildings, while natural gas heating and electric cooling among San Jose buildings. There are also apparent differences in scheduling of the Richardson buildings, particularly in the evening where distinct time differences in HVAC turn offs are seen. Finally, overall shape differences among buildings show differences in occupancy

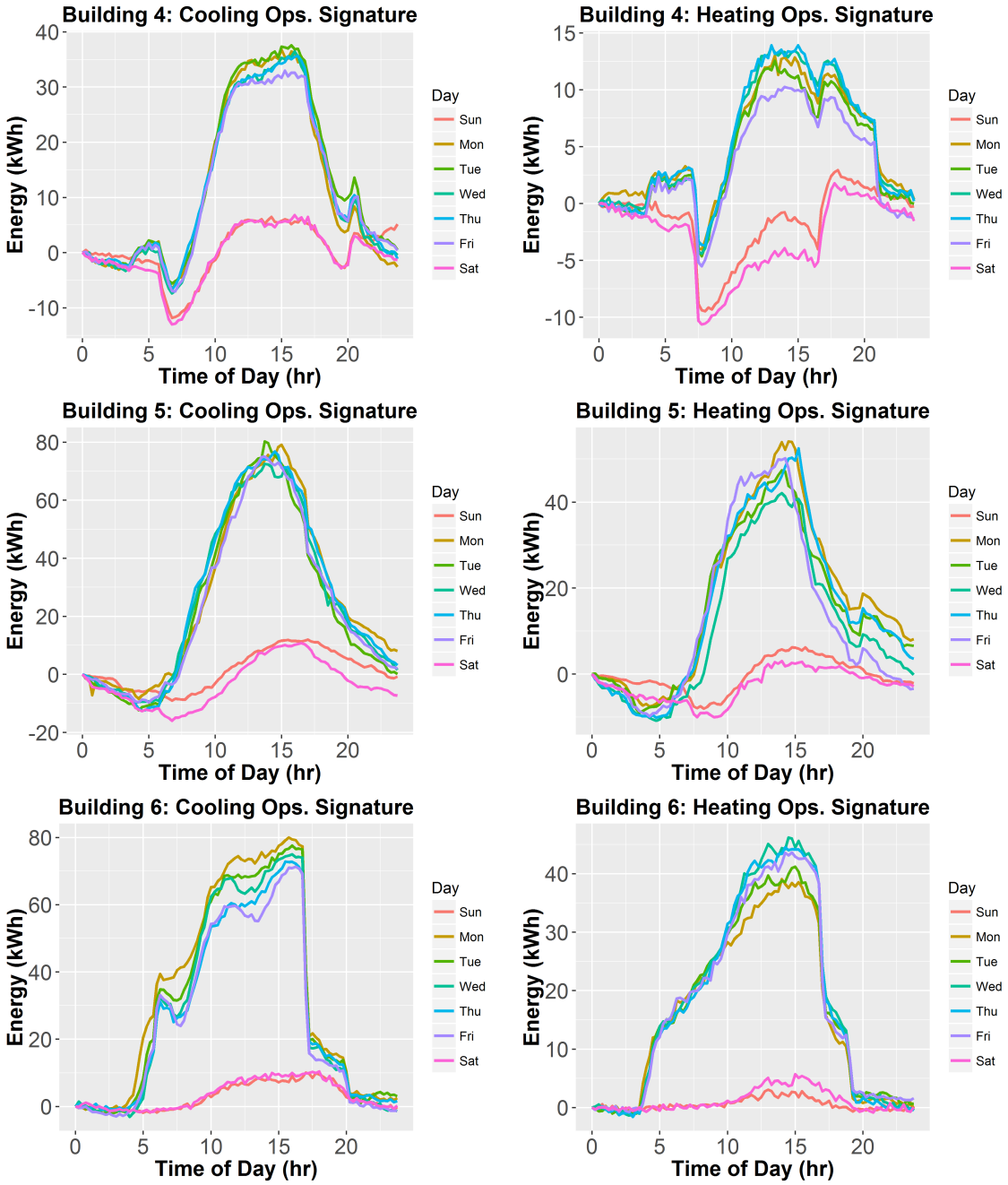

Figure 4.7. *Cooling and Heating Season Daily Operation Signatures for San Jose, California Buildings*

and size as in Building 5 compared to Building 4, as well as between buildings of differing uses, such as the blocky operation signatures of lab buildings (Richardson) and smooth occupancy driven operation signatures of office buildings (San Jose).

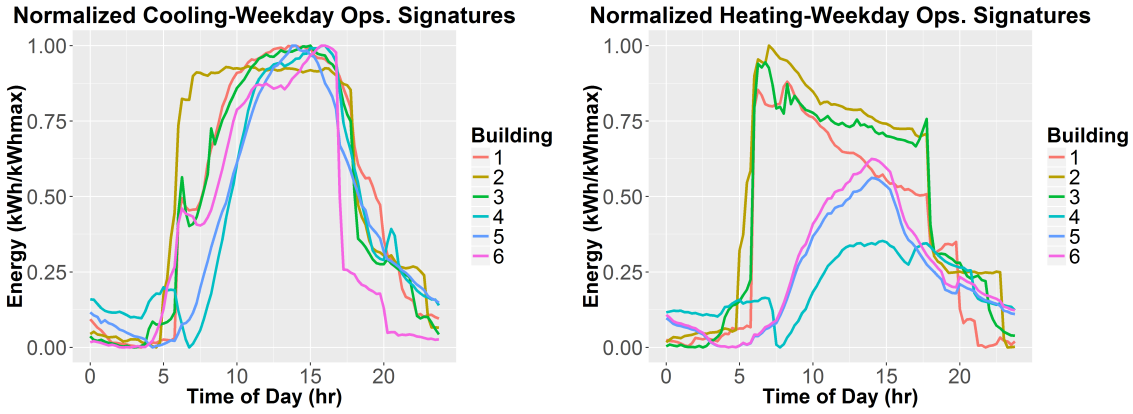

Figure 4.8. *Daily Operation Signatures of All 6 Buildings in Cooling and Heating Conditions Normalized to each Building's Peak Season Operation*

Solar irradiance is known to impact the thermal loads on buildings[12,58] and by additionally subsetting mean daily solar irradiance into sunny and cloudy days, the impact of irradiance can be determined. In cold heating season temperatures, building's heating loads are lessened during high solar irradiance days compared to low solar irradiance conditions. A cold heating season day is defined as having a mean temperature less than 48F (8.9C), while a sunny day is defined with a mean solar irradiance above 210 W/m^2 and a cloudy day as less than 70 W/m^2. The mean solar irradiance is calculated considering all 15-minute data points, including night values (0 W/m^2). Both the temperature and solar irradiance constraints were identified to allow for at least 10 days to pass each criteria set, create a large distinction between sunny and cloudy days, and enable an accurate time series decomposition. Figure 4.9 shows Building 1 daily operation signatures with varying daily solar irradiance on the left ($\mu = \pm 1.84 kWh$ for sunny, $\mu = \pm 1.73 kWh$ for cloudy) and the associated average solar irradiance (W/m^2) vs. time of day. The building operation signature is computed using the seasonal component plus the mean of the trend values to account for magnitude differences. Mean trend values are $407 \pm 17 kWh$ and $416 \pm 35 kWh$ for sunny and cloudy days respectively.

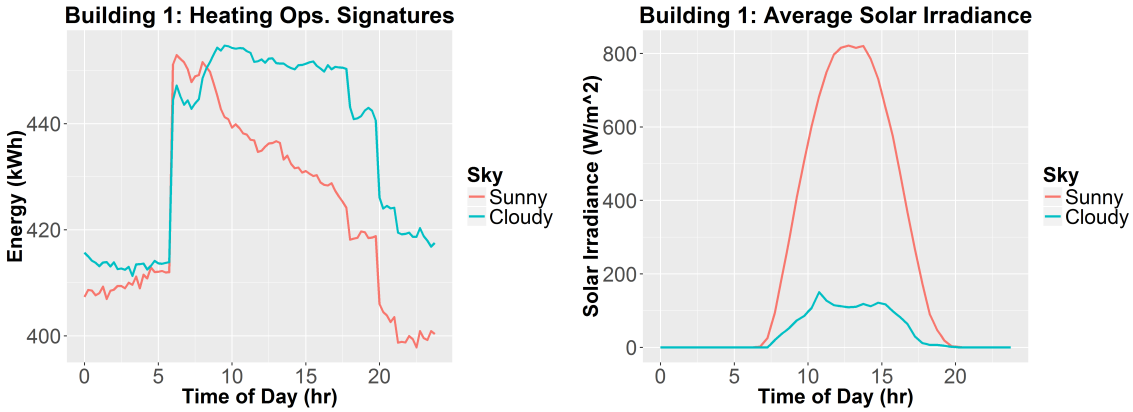

Figure 4.9. *Left: Operation Signatures of Cold Heating Season Days in Cloudy and Sunny Weather for Building 1 ($\mu = \pm 1.84 kWh$ for Sunny, $\mu = \pm 1.73 kWh$ for Cloudy) Right: Average Solar Irradiance for Sunny and Cloudy Days*

The figure shows that during cloudy days, much more overall energy is required to maintain the set point, while the sunny days tail off in consumption throughout the day as the solar irradiance assists in contributing heating load. Computing this difference between the signatures show cloudy days consume 1.2 MWh ($84) more than equally cold sunny days. Also, the lower morning usage for cloudy days and higher morning load for sunny days shows the effect of night cloud cover maintaining higher temperatures as compared to clear nights. Overall, these observations give insights into a building's solar-thermal load, which may point to opportunities for energy savings by using window shading accordingly during sunny days.

5 Disaggregation

One of the toughest challenges of analyzing building electricity consumption data is the inherent mixing of several types of consumption components in the data[45]. Commercial building electricity data consists of many electricity signals, such as baseload, HVAC, plug load, and exterior lighting; where baseload is the minimum electricity consumption required for a building (e.g. 24-hour lighting and security/monitoring systems), HVAC includes heating, ventilation, and air conditioning electricity consumption, plug load is the electricity consumed by occupant activities (e.g. computers, refrigerators, copiers, televisions, interior lighting, etc.), and exterior lighting includes lighting used for parking lot or security lighting at night. Singularly each electricity consumption component can provide a multitide of insights, but arriving at this data in a disaggregated form has conventionally required the employment of special sensors or load monitoring systems.[45,46]. Using EDIFES, without the need for additional sensoring or metering, disaggregation of the total electricity consumption into its various components can be achieved. This section describes the progress made in disaggregating the building electricity data, including the disaggregation of exterior lighting, HVAC, and occupancy based plug loads in buildings.

5.1 Exterior Lighting Disaggregation

In some cases, exterior lighting loads are included in building electricity consumption data. Exterior lighting may include security, parking lot, or display lighting systems which are only necessary in dusk, dawn, and nighttime hours. These lighting systems also tend to be significant load contributors due to vast surface area of coverage needed. Two buildings, 3 and 4, contain exterior lighting. Figure 5.1 depicts a sinusoidal line from the derivative analysis on electricity consumption produced by system_finder() throughout the two year span of Building 4. Looking more closely, a similar sinusoidal trend is demonstrated by plotting the sunrise and sunset times of each day. This was determined through the EDIFES sun_tagging() function which uses the function sunrise.set from the package *StreamMetabolism*[59], and allows for cross referencing of the turn on/off events with the times associated with the sun rising and setting. Using a similar derivative method the EDIFES function ext_light_finder() detects whether a building demonstrates exterior lighting. Just as in the derivative analysis, standard deviations are identified, but only on an hour interval centered at the sunrise or sunset time of each day. For example, if the sunrise occurred at 6:30 a.m. the function will search for a "turn off" event from 6-7 a.m. Each time a "turn off/on" event occurs during a sunrise/set interval, the occurrence is tallied and the overall occurrences are divided by the total number of days in the data set to determine the frequency. When the finder sees over a 75% frequency from both the sunrise and sunset occurrences the building is determined to possess exterior lights. Figure 5.2 shows the energy consumption of Building 4, and highlights the lighting events in the building of which are identified by ext_light_finder().

Once identified as having exterior lighting, the data is passed on to the EDIFES function light_value(). The function analyzes all of the occurrences and aims to determine a

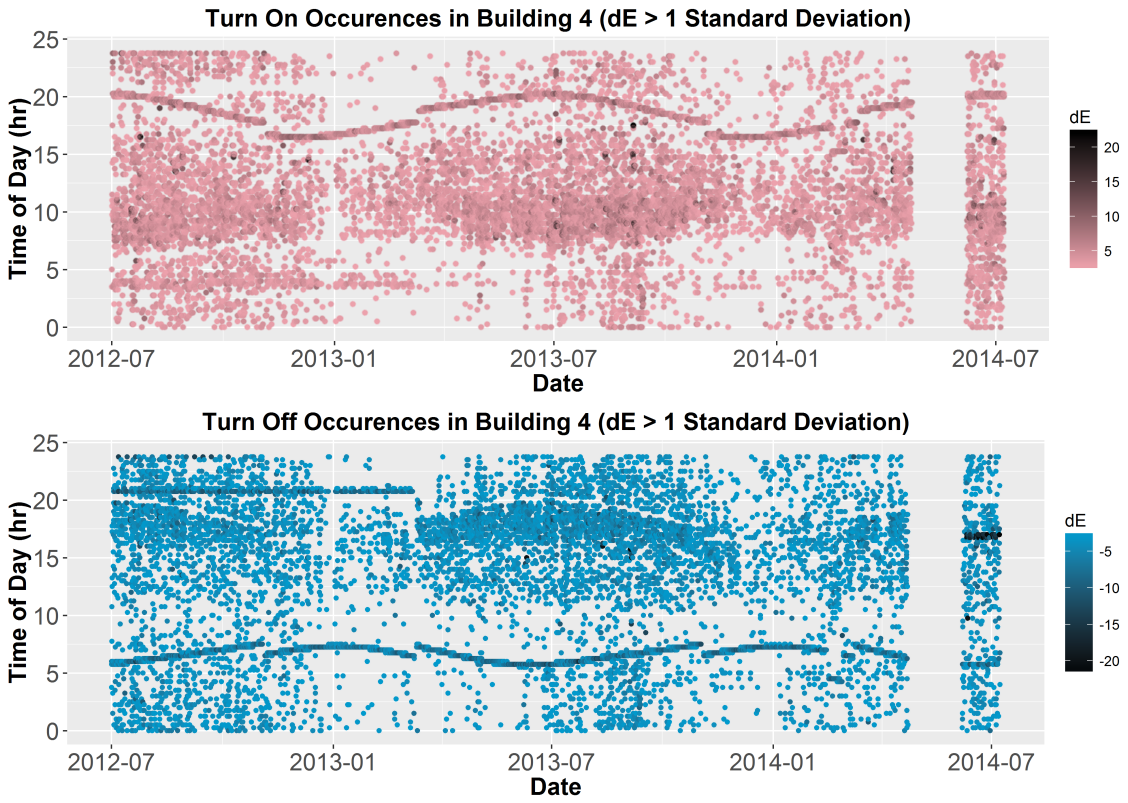

Figure 5.1. *Building 4 Turn On and Off Events, Standard Deviation* ±1

distinct value for the exterior lights. This determination is surprisingly complex for multiple reasons: 1) the lights turn on and off opposite of the current electricity consumption trend (i.e. the morning consumption of a building is ramping up in use, but as the sun rises the lights turn off, a decrease in use, mixing the two results) and 2) consumption is time integrated and the "turn on" in the middle of one interval will be observed over two intervals (i.e. if 60kW lights turn on at minute five of the 15 minute interval they will be seen in two consecutive energy consumption data points as an increase of 5 kWh in the first, and another 10kWh in the second - this is "carry over"). Therefore, considering the combination of these two properties, a significant and varying portion of the light load is masked in the data - meaning only perfectly timed, with no other

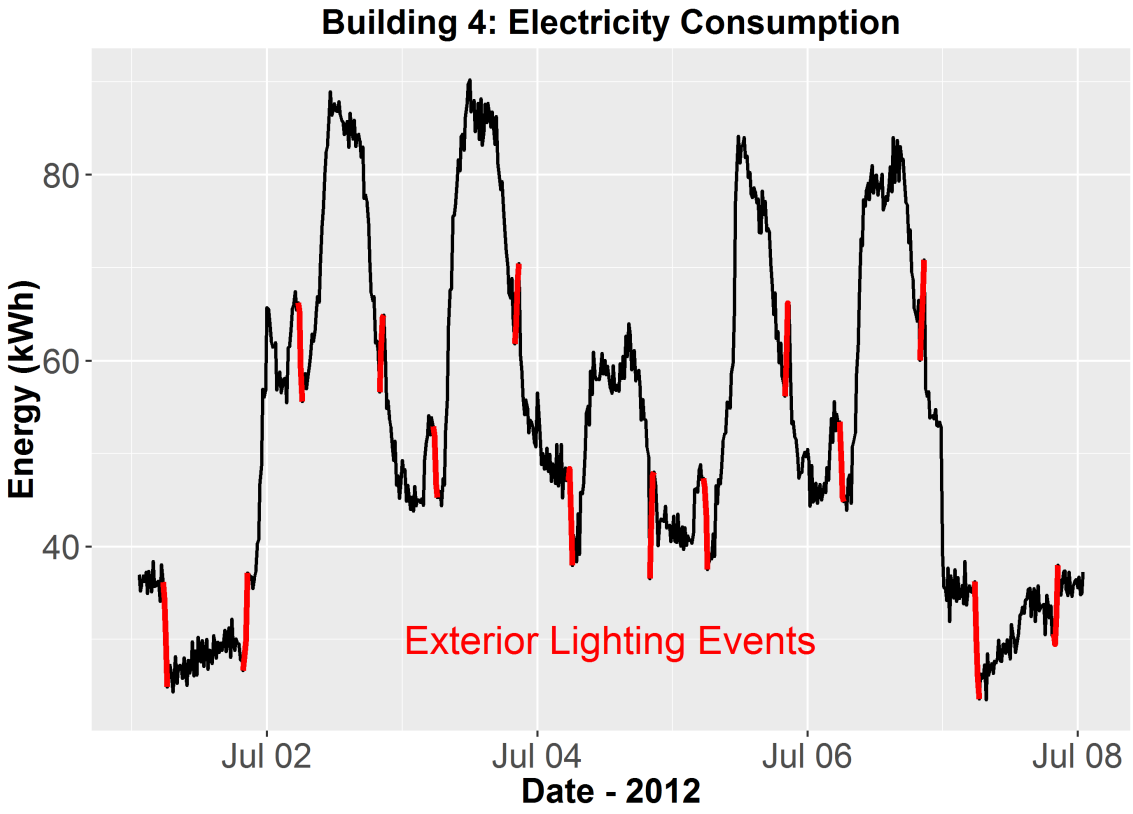

Figure 5.2. *Building 4 Total Electricity Consumption with Exterior Lighting Events Identified*

loads turning on or off in the interval, lighting events could be read in full in one time step. This then requires that the lighting event occurs over at least a range of two intervals. However, after much analysis it was determined that all lights *do not* turn on/off at the same time. They respond to independent lighting sensors, further opening the time interval to expect observing lighting events and also allowing a larger window for loads of other equipment to mask the lighting events.

Considering the multiple issues with determining the value of the lights, a statistical approach was taken. The values of these lighting events are aggregated from 30 minutes before and after the sunrise/set time to account for differing turn on/off times of the light banks, and also to account for "carry over" in the time integration of consumption.

Additionally, due to the opposite direction of energy consumption, larger changes in energy observed are actually those of lighting events that occurred at intervals when the least amount of other loads interfered. Therefore, the large values are deemed closer to the actual exterior lighting value than other observations. The top quartile of the observations were determined to qualify as the larger values for analysis and the median value of the quartile is reported as the value for the lights. These constraints were created after numerous disaggregation attempts, where the top quartile median was found to minimize the error in disaggregation. For Building 4, the value is 11kWh per 15 minute interval, which corresponds to a 44kW load.

Once the lighting value is determined, the last function ext_lighting_disag(), takes in the value and disaggregates the lights from the total electricity consumption data. The same issues persist in the disaggregation as did in the determination of the value of the lights, and to account for them, a weighting approach was taken to disaggregate. Looking 30 minutes before and after the sunrise or sunset, the function identifies all positive/negative energy changes (depending on the nature of the event, turn on/off corresponds to sunset/rise respectively) and sums the values, this sum is the weight denominator. Then the value, 11kWh for example, is distributed among the positive energy change events (for lighting turn ons). This is done with each event's energy change, divided by the sum of the weight, and multiplied by the value. For example, a sunset, "turn on" event occurs, over the interval 30 minute prior and 30 minute after, three positive "turn on" energy changes were found, 3, 2, and 5 kWh, the sum of which is 10kWh. At the event where 3kWh was observed, the lighting value $11 * 3/10$ is calculated, the next step would add the $2/10$ weight, followed by the $5/10$ weight, finally coming to the total value of the lights, 11kWh. This process is automatically computed and Figure 5.3 shows the

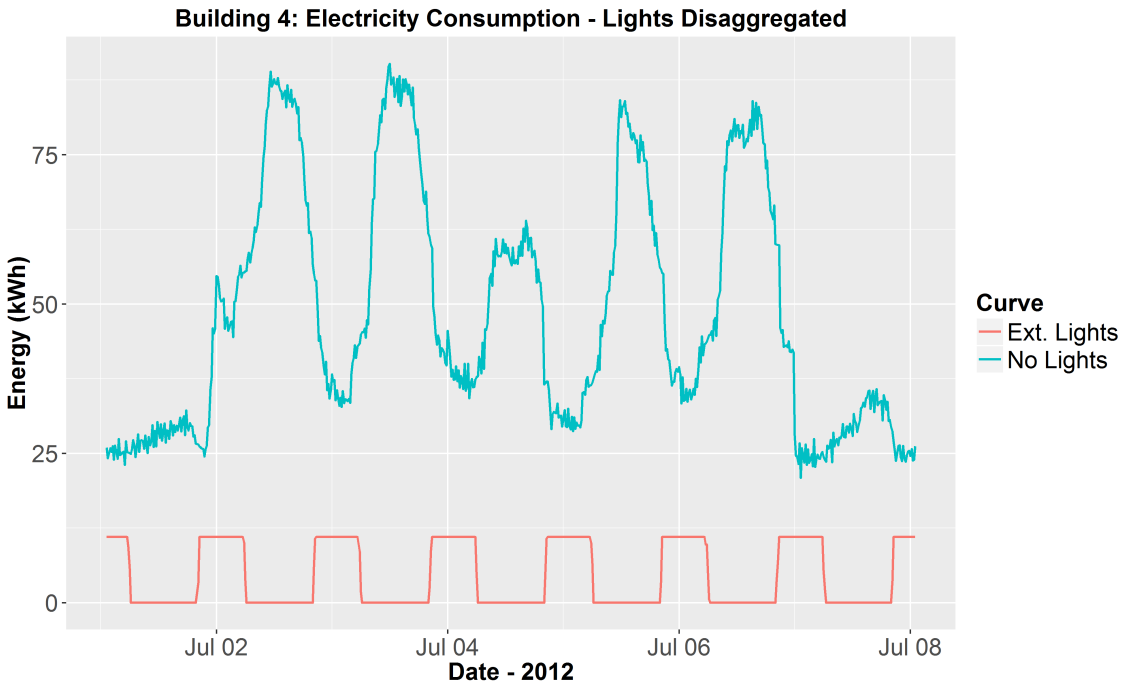

Figure 5.3. *Building 4 Electricity Consumption with Exterior Lighting Disaggregated.*

automated lighting disaggregation. It is worth noting that at the points of disaggregation there appear no artifacts (bad data) from removing the lights and ext_light_finder() no longer identifies the lights in the data, validating this method.

Now that the light values are known, the exact electricity consumption is determined for an entire year, as well as the cost. At a load of 44kW the lights consume approximately 193 MWh per year or $38,500 at $0.20/kWh in San Jose, CA. Provided the lighting type of the building, a quick calculation can be made to assess a lighting retrofit. In this case, if the lights are currently halogen (25 Lumens/Watt) and an LED (45 Lumens/Watt) retrofit were considered, the building owner would save approximately 86 MWh and $17,100 per year if lumens were kept constant, or if HID lighting were considered (120 Lumens/Watt) 153 MWh and $30,500 per year in savings. This also allows for quick return on investment calculations, such as, a 3 year return on investment would require

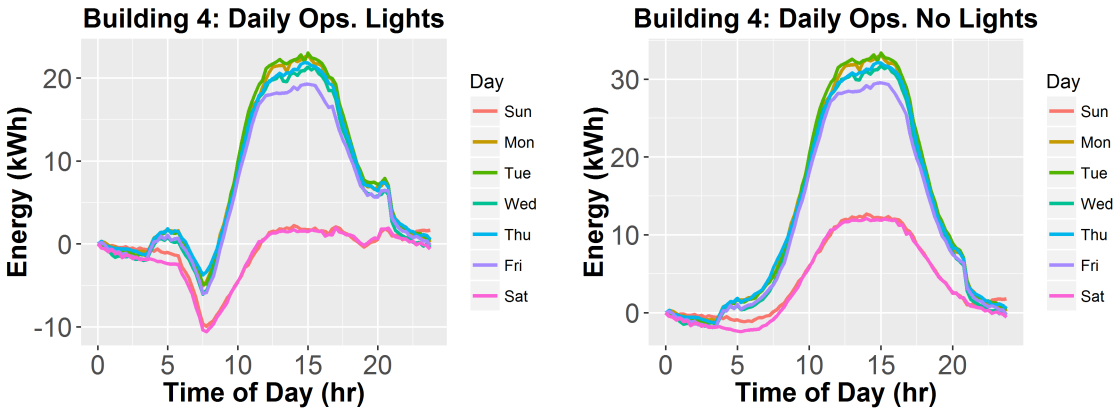

Figure 5.4. *Building 4 Day of the Week Time Series Figure Before Lighting Disaggregation on the Left and After on the Right.*

the retrofit to be less than $51,300 or $91,500 for halogen and HID respectively. Finally, this disaggregation step is incredibly important as the variability in time of the lights severely hurts system identification and other analysis that look for scheduled patterns in the data. Figure 5.4 shows the significant difference between the day of the week time series analysis before and after lighting disaggregation. The operation of the building after the disaggregation now exhibits the true HVAC and occupancy tendencies of the building, and allow for new analyses to be applied.

5.2 Heating Ventilation and Air Conditioning Disaggregation

HVAC is the single largest component of a building's energy consumption, and consequently the largest opportunity for energy and cost savings[1]. However, HVAC is also very difficult to accurately disaggregate from the largely unpredictable occupancy loads of a building, and each building's response to weather characteristics typically vary drastically. Considering these hurdles, weather data and observed scheduling remain the best variables to attempt disaggregation when submetering countless equipment loads are

not an option. In this section an approach is presented to model HVAC usage vs. temperature, while accounting for random occupancy/plug load behaviors.

As shown earlier in the "Early Data Analysis" chapter, temperature and energy do not correlate well when directly compared (current interval temperature vs. electricity). They do, however, continue to increase as various constraints are applied to the data sets, one of those being mean daily electricity consumption against mean daily temperature, but this analysis also has its own flaws. It uses mean daily values that associated the temperatures in the evening with the electricity consumption earlier that morning, a non-physical comparison. There are also other factors in a building, such as scheduling and occupancy, and both are a function of the "time of day" and not temperature. These can be difficult issues to deal with as HVAC, the most correlated and only significant causal relationship to temperature, changes set points throughout a day and into the weekends. This leads to a discontinuous HVAC energy consumption, which obviously does not correlate well to a continuous temperature data set. Finally, buildings do respond to temperature, but due to large thermal masses, they respond to the temperatures that occurred many hours before. Therefore, it was found that the aggregate of weather characteristics define a building's HVAC operation at far higher correlations than compared to a singular lag.

The two issues for disaggregating the HVAC component from the total electricity consumption can be summarized as such: 1) each time interval (each day includes 96) presents its own set point value and occupant tendencies and 2) each time interval responds to its own distinct accumulation of weather characteristics. The second may be described as an accumulated lag of the building's response to weather. To compute

these lags, EDIFES uses the function lag_stat() which determines the mean value temperatures for a specified lag. The function computes the "lags" for all values 1-96. For example, a lag of 1 will report the mean of the current and last temperature, while 96 will report the mean temperature of the current and past 96 (a full day) data points. Again, this is done for each time interval considering each of the 96 lag possibilities.

The analysis then computes how well each lag, with each time step, is able to linearly model energy versus temperature. This is done by taking one time interval at a specific lag and examining the distribution of mean temperatures computed. In over 700 days of data, a significant amount of similar temperatures occur and allow for the grouping of those similar events. The next EDIFES function in the analysis is hvac_explicit(), (named for determining the explicit temperature contribution for energy compared to taking the mean daily temperatures and mean daily energy consumption, an implicit solution) which groups the data into similar temperature ranges of at least 30 days, satisfying the central limit theorem of statistical significance[57], and minimizing noise in the data. For example, the 30 hottest temperatures observed (from the lag at a particular time interval) are grouped together, then the next 30 hottest, and so on, until all days have been accounted for and grouped, resulting in approximately 20 groupings. The mean temperature of the group is recorded as well as the mean energy of the group. Due to the number of groupings the range of temperatures in each group from the mean value is less than 1 degree Celsius (except in some cases at the very hottest and coldest groupings where outliers may exist) - demonstrating the similarity in temperature among each group.

These mean temperatures and energy values can now be modeled through a linear regression. Figure 5.5 shows the values of Building 3 at midnight and at 12pm, using a

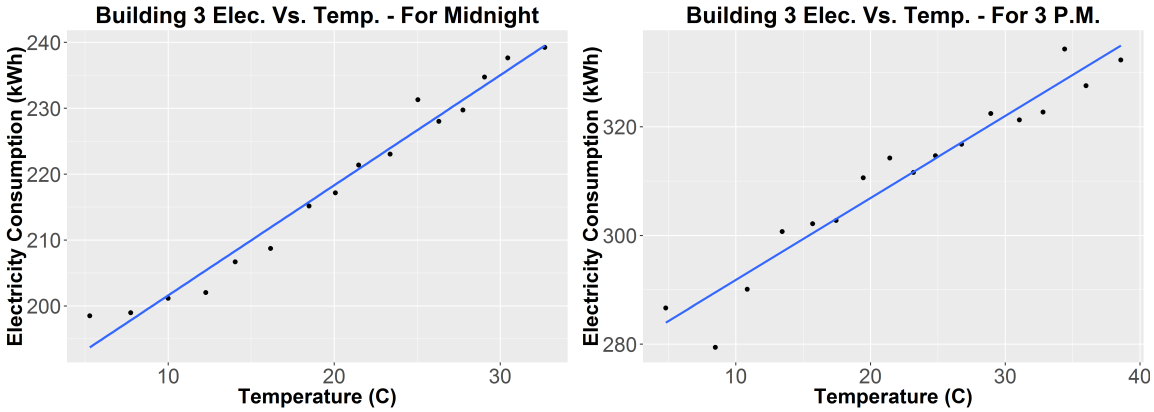

Figure 5.5. *Grouping Plots at Midnight and 3 p.m. for Building 3. The blue line displays the linear regression best fit.*

lag of 4 hours. Both plots in the figure show a very linear progression of the data, having R-Squared values above 0.9. This result is far better than any other analysis using simple correlations (the R-Squared value is the correlation squared in a linear model, therefore the correlation is much higher than 0.9 and is considered a significantly strong relationship[55]).This high correlation is achieved because each time interval is considered independently, therefore ensuring HVAC characteristics such as set points are the same at the same time of day. Conducting the analysis across the 30 days allows for a normal distribution that minimizes the effects of random occupancy loads. The function hvac_explicit() computes a linear regression for each data set and reports the adjusted R-Squared value, or fit, of the model.

EDIFES then uses the function thermal_lags() to accumulate all of the R-Squared values associated with each and every time interval and lag combination. The output of the function is then a 96 x 96 matrix (96 intervals and 96 lags) holding all of the R-Squared values for each lag at each time interval. Various approaches can then be taken to determine the most appropriate lag for a building. One solution is to choose one singular lag that results in the highest sum of its 96 R-squared values. Second is to choose

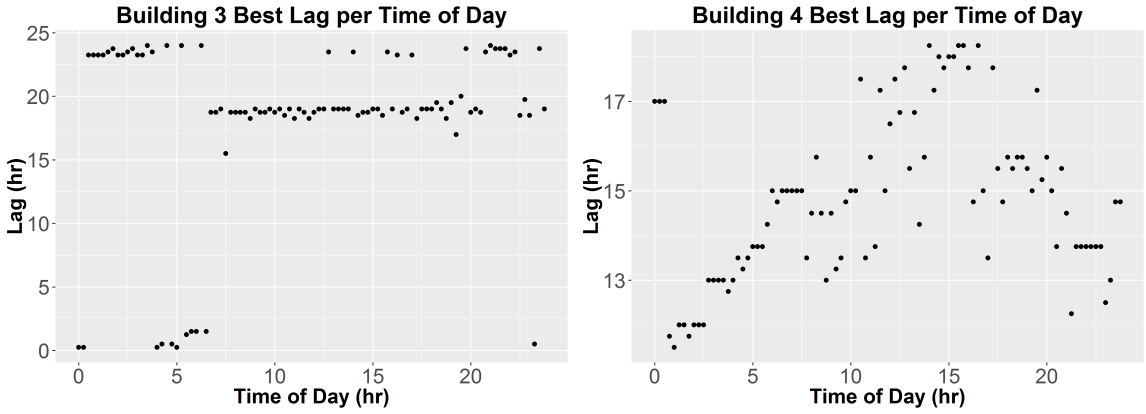

Figure 5.6. *Building 3 & 4 Highest R-Squared Lags for each Time Interval of the Day*

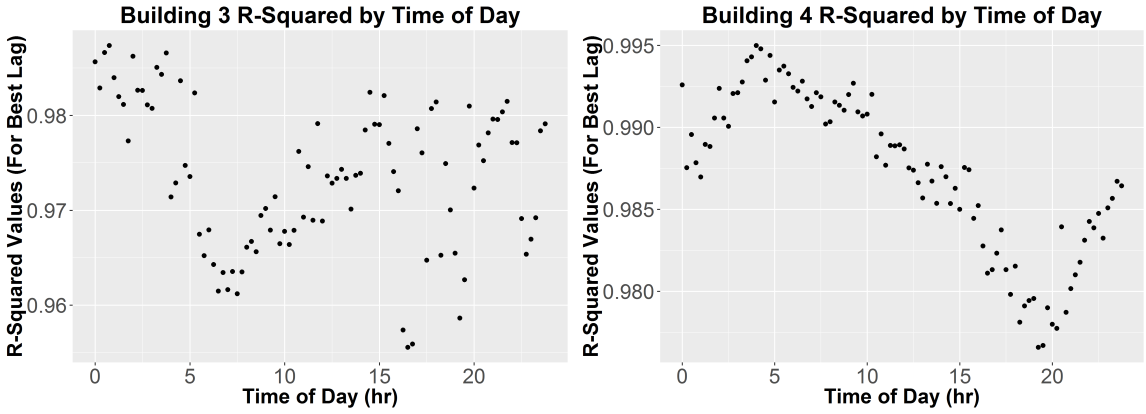

Figure 5.7. *Building 3 & 4 R-Squared Values for the Associated Best Lags of each Time Interval.*

the lag value (mode, median, or mean of these values) that is found to be the highest correlated value for each time and apply it over all time intervals. Third, and the option that is chosen in this analysis, is to use the best lag for each time interval separately. Figure 5.6 and Figure 5.7 show the best lag value for each time of the day for Building 3 and 4 respectively.

Of particular interest are the lags values in Figure 5.6 for Building 3. During night or unoccupied times the building has a lag of about 23 hours, but during occupied times the lag changes to about 18 hours. Further, the lag drops to a less than 2 hours for certain

times, particularly from 4-6 a.m. - the time at which HVAC systems turn on in Building 3. Due to the turning on of HVAC systems, largely air handlers, the current ambient temperature is much more important and correlated to energy as compared to any other time in the day. Air handlers primary purpose (and largest corresponding electricity load) is to heat and cool air and thus, the HVAC consumption is at its maximum during the "turn on" times. This therefore results in high correlations between electricity usage and current temperature at this particular time of the day. Then after about 7/8 a.m. the building has established near steady state conditions once again and cools/heats at a lower rate that is more indicative of the overall thermal load on the building, accounting for the past 18 hours. The same phenomenon occurs during the unoccupied night operation, although the difference of 5 hours of thermal lag is a result of the fact that a set point change (occupied to unoccupied) occurs in the evening. Building 4, on the other hand does not directly show this, but does show a more defined overall lag, as all lags lie between 11-19 hours. Figure 5.7, nonetheless, shows an interesting trend for Building 4, which uncovers temperature set point change times at approximately 4 a.m. and 9 p.m. due to the shifts in the R-Squared data. The set point changes are further discussed later in this chapter.

EDIFES can now account for the effective thermal lags in the buildings, and using function best_lags() determines the best lags for each time interval and passes this knowledge to hvac_explicit_2() which recalculates the models for each time interval, only using the best lag for that interval. The output of this function includes 96 linear regression models, one for each time of the day. Each linear model includes an intercept

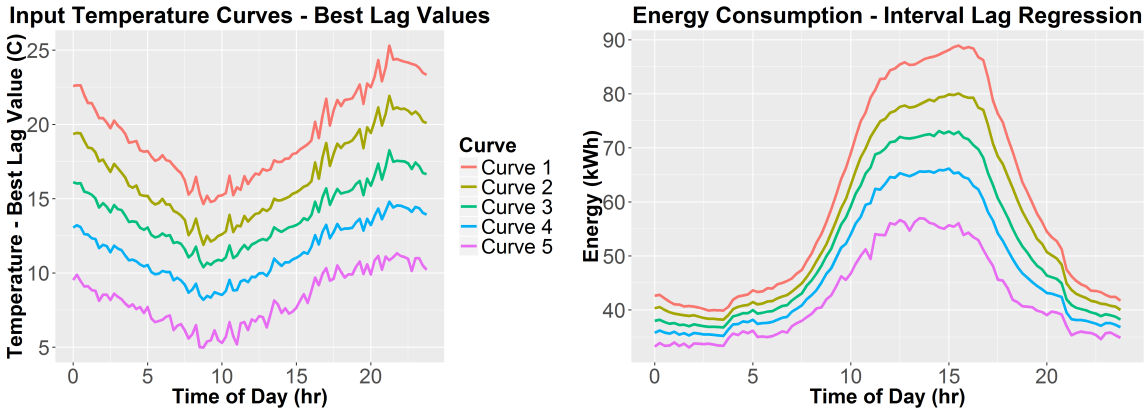

Figure 5.8. *Building 4 HVAC Lag Regression Energy Consumption as an output of Best Lag Temperature Data*

and slope for the regression, with the slope holding information on the relationship between energy and temperature in kWh/C. The slope, however, is different for each interval and for conciseness, Table 5.1 displays the mean slope, or HVAC energy response to temperature, for each building. Building 1 displays its negative relationship with temperature, with -53.24 kWh/C consumed per 15-minute interval, while Building 3 shows the largest association with a 60.21 kWh/C response. Unsurprisingly, the magnitude of each building's slope trend with relative buildings size, with the exception of Building 6. Building 6 is the fourth largest building, but has the smallest response in electricity to temperature.

	kWh Per Degree C
Building 1	-53.24
Building 2	44.84
Building 3	60.21
Building 4	23.38
Building 5	33.54
Building 6	17.18

Table 5.1. Mean Slope of Lag Regressions, kWh Per Degree C

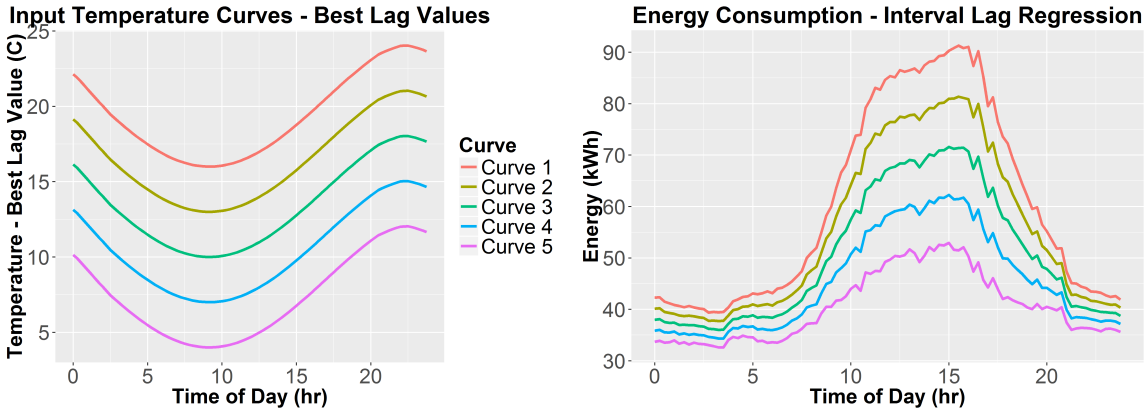

Figure 5.9. *Building 4 HVAC Lag Regression Energy Consumption as an output of Smooth Temperature Data*

Finally, the function hvac_explicit_disag() can then take in the models and the corresponding temperature lag values for any given day and output the associated HVAC electricity consumption. Figure 5.8 displays the output HVAC electricity consumption for various temperature data sets, of which are not continuous as each value is associated with a differing lag. Figure 5.9 shows the output if smooth temperature data sets were applied to the model, consequently the energy output becomes jagged due to the incorrect temperature lag values. These electricity consumption curves, however, are not just the HVAC, they do still include typical occupancy plug load. One approach for disaggregating the plug and baseload is to take the lowest energy consuming temperature condition (either hottest or coldest values, in this case coldest for Building 4) from the HVAC electricity consumption to result in the variable HVAC component. Unfortunately, this method results in unsatisfactory results for a substantial amount of days, but does work for a few days in the data set. The issue arises from situational occurrences of HVAC and plug load. For example, some nights HVAC units will turn on at atypical

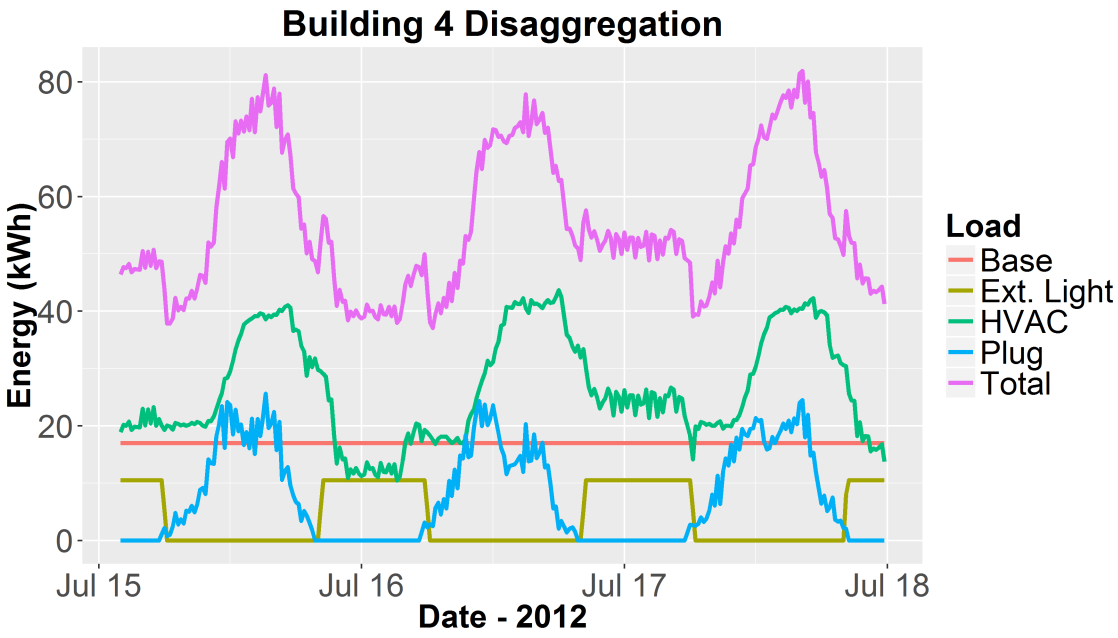

Figure 5.10. *Building 4 Disaggregation Using HVAC lag, exterior lighting, and HVAC event disaggregation functions*

times. Therefore, more work must be done to adjust for the atypical situational occurrences demonstrated by the HVAC data. Disaggregation for Building 4 is shown in Figure 5.10.

Figure 5.10 is developed using the HVAC lag, exterior lighting, and HVAC event disaggregation functions. The exterior lighting function is used to disaggregate the lights followed by the HVAC functions. However, the second HVAC function, HVAC_event(), is currently not scalable to entire data sets and other buildings. Further research into refining the algorithm promises to expand its use to entire data sets of various types and allow for automated disaggregation of building energy. Despite the limited applicability, results from disaggregation indicate the methods used create reasonable findings. In particular is the plug load electricity consumption. The plug load is determined from taking the total energy consumption and removing the HVAC, exterior lighting, and

baseload components, therefore the plug load is entirely derived from the other functions. The plug load exhibits a doubled peaked behavior, rising in the morning, then dipping around noon, followed by a second peak before decreasing through the afternoon to evening hours. This plug load operation is identical to what is seen widely in literature sources for occupancy levels in office buildings[60–62]. Where occupancy drops around noon as employees leave the building or pause their current work - therefore decreasing the plug load. Following lunch at noontime, the plug load rises again as employees resume work in the afternoon, succeeded by a final decrease as employees leave for the rest of the day. In conclusion it is believed that the HVAC lag and event approach, along with exterior lighting disaggregation provides promising results for future investigation.

5.3 Set Point Identification

In commercial buildings, most HVAC units are programmed with varying internal set points for occupied and unoccupied times to minimize energy consumption and maximize employee/customer comfort. Considering the knowledge gained on HVAC in the previous section, the internal set point changes can be determined. The HVAC lag models provide the change in energy consumption as a function of temperature, and with knowledge of the energy difference at a set point change, the set point temperature difference was found.

Two functions have been created to detect the set point change simply given the data and models developed by HVAC lag. EDIFES function set_point_finder() applies a time series analysis on the entire dataset and uses the seasonal component to determine if

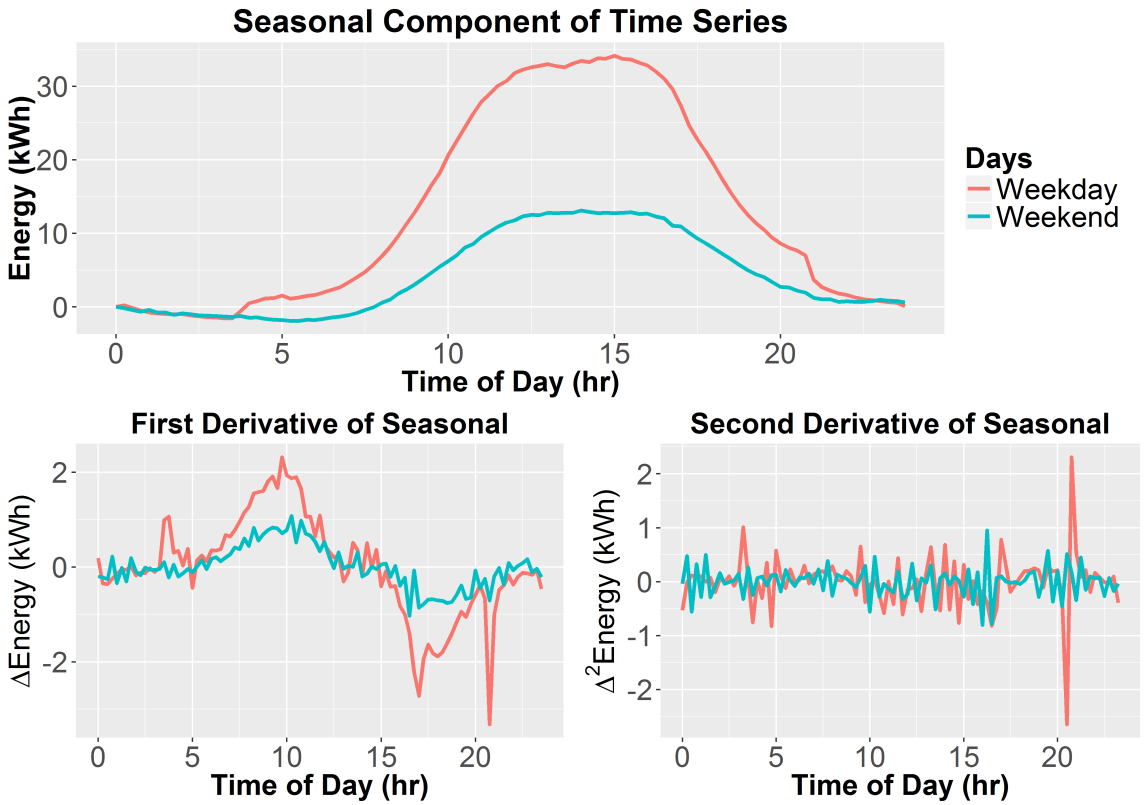

Figure 5.11. *Set Point Methodology, Building 4 Seasonal Time Series and First and Second Derivative Plots showing spikes indicating set point events*

a set point change exists. By analyzing the seasonal component, overall characteristics can be found, such as a daily set point change. Figure 5.11 shows the seasonal component and a jump in usage can be seen in the weekday consumption early in the morning and late in the evening. To identify this event the second derivative of energy is taken to find sudden spikes in the time series energy consumption. The function then takes the mean and standard deviation of the second derivative and pin points the significant occurrences, with one last set of criteria - a morning set point will occur before noon and will result in a positive change (i.e. disregard negative spikes) and the evening set point will occur after noon and will result in a negative change (i.e. disregard positive spikes).

Using this analysis set_point_finder() determines all of the set point change times during a day and reports the following logic: "Yes/No - Weekday Set Point Change", "Yes/No - Multiple Weekday Set Point Change", "Yes/No - Weekend Set Point Change", "Yes/No - Multiple Weekend Set Point Change". Table 5.2 shows the output from the function identifying all set point information and identifying 4 a.m. and 9:15 p.m. as set point changes. Following set_point_finder(), the function set_point_value() takes in the overall data, information from the previous set point function, and the models from HVAC lag regression to determine the internal set point change. The set point change for Building 4 was reported to be 3.6F, or 4F as confirmed by the building manager.

Set Point Finder Results
Yes - Weekday Set Point Change
No - Multiple Weekday Set Point Change
No - Weekend Set Point Change
No - Different Weekend Set Point Change
4
21.25

Table 5.2. Set Point Finder Output - Identified Weekday Set Point Change at 4 a.m. and 9:15 p.m. for Building 4

6 Equipment Identification

Disaggregation aims to break up the main load components from the overall electricity curve, and in many cases go further into determining what specific pieces of equipment make up each component. Equipment identification is the next step beyond disaggregation and allows for the pin pointing of specific equipment for analysis to determine equipment performance over time. However, based on the fact that equipment is a subcomponent of a disaggregated curve of the total consumption, extracting information on equipment can be quite complex. This section explores a method for extracting values of equipment within the building electricity consumption data.

6.1 Method of Equipment Finding

Building electricity consumption is time integrated, as described in detail in Chapter 2, Early Data Analytics, resulting in mixed consumption of various equipment. One approach to finding equipment is to simply analyze the raw derivative values. Tabulating the data for unique derivatives, only 194 unique derivative values are found (307 if negative and positive pairs are not considered as the same) - considering the 70,000+ observations this number is relatively small. Figure 6.1 shows the 194 derivatives and the amount of times they occur in the data set.

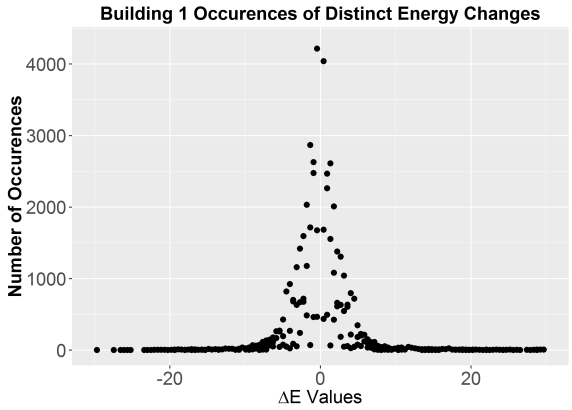

Figure 6.1. *Building 1 Occurrences of Distinct Changes in Electricity*

Although, 194 values are significantly small, more constraints are required to determine real equipment values rather than mixed values - as is the case in most of the derivatives. To find real equipment values one can leverage the knowledge that a piece of equipment turning on/off will have an impact on the current time interval and also on the following interval. In the rare cases where no other equipment turns on/off the information found in the second interval can be used to determine the value of the piece of equipment. If a piece of equipment turns on during an interval it results in a partial energy change (in that interval) for the load of the equipment, while the second interval change is the full value of the equipment over 15-minutes. If no other equipment turns on during that 30-minute time frame, the second interval can be directly converted to a power value for the equipment, and - more importantly - the values of this equipment can be used to back calculate when the equipment turned on in the first interval. For example, 10kWh is observed in the first interval and 15kWh in the second interval. The second interval can then be calculated to be a piece of equipment the size of 60kW, and the first interval can be back calculated to show the piece of equipment turned on at

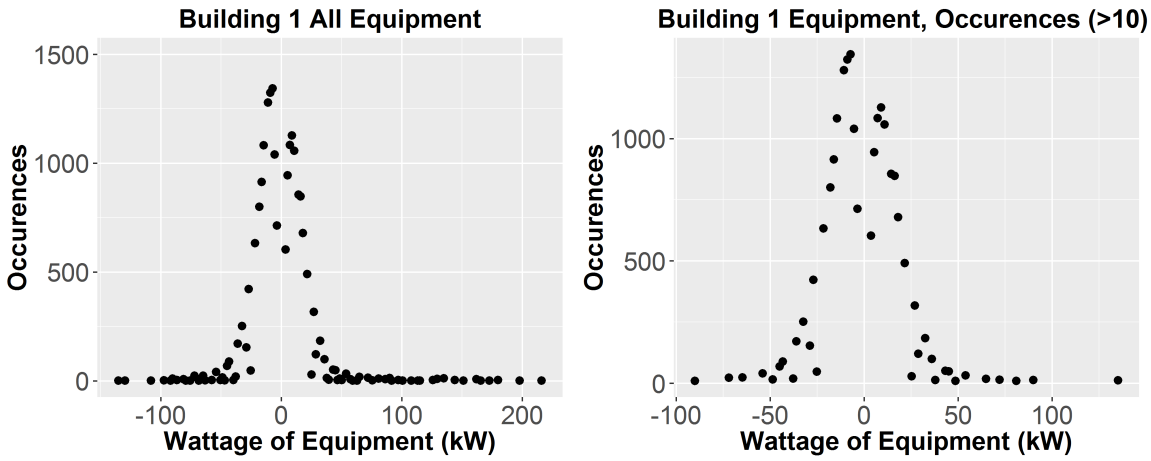

Figure 6.2. *Building 1 Occurrences of Detected Equipment by "Turn On/Off" Wattages*

minute five of the first interval. This time is useful as it provides another means of determining if a piece of equipment is real. If the equipment did not turn on at an expected interval of the electricity meter, then it does not exist. An expected interval of the meter is one which occurs at the frequency at which the meter operates. In this analysis only equipment which occurred at a half second interval (in 15 minutes this is 1,800 possible turn on/off intervals) are considered real occurrences. Therefore, this analysis requires two criteria: 1) two consecutive positive or negative energy changes occur and 2) a "real" interval turn on time must be observed, where a real turn on interval is one that occurs at a half second interval.

6.2 Identification of Equipment

Performing this analysis on Building 1 and omitting equipment that did not turn on/off at the expected intervals gives the Figure 6.2. The first plot shows all of the determined pieces of equipment - a total of 51 distinct values of equipment. The second narrows the criteria to equipment that were observed at least 10 times - lowering the total to

24 distinct values. Tables 6.1 and 6.2 show the values of the equipment found and the number of occurrences they were observed.

Equipment Wattage (1)	Occurences (1)	Equipment Wattage (2)	Occurences (2)
9000	1127	36000	99
7200	1083	43200	51
10800	1057	45000	48
5400	944	54000	32
14400	855	25200	28
16200	847	64800	18
18000	679	72000	14
3600	603	90000	13
21600	491	37800	13
27000	317	135000	12
32400	184	81000	10
28800	121	48600	10

Table 6.1. Building 1 'Turn On' Equipment by Wattage and Occurrences

Equipment Wattage (1)	Occurences (1)	Equipment Wattage (2)	Occurences (2)
-7200	1344	-36000	171
-9000	1323	-28800	154
-10800	1279	-43200	89
-14400	1082	-45000	68
-5400	1039	-25200	47
-16200	914	-54000	40
-18000	800	-64800	24
-3600	713	-72000	23
-21600	632	-37800	19
-27000	422	-48600	16
-32400	251	-90000	10

Table 6.2. Building 1 'Turn Off' Equipment by Wattage and Occurrences

The meter resolution for Building 1 was found to be 1.8kW, but the values (except for at the low wattage end) do not jump intervals by 1.8kW, but by larger increments. In fact, one multiple stands out among the rest, 9kW. Equipment values are seen from 9kW up to 135kW, ascending in 9kW intervals, indicating that the building may have multiple

pieces of the same equipment at a value of 9kW or that a piece of equipment can be operated at different consumption levels (e.g. a variable speed drive). Figure 6.3 displays the 9kW hour multiple values (from 9kW - 81kW) by their hour of occurrence vs. the date they occurred, as well as indicating the magnitude through a color scale. Looking at Figure 6.3, lines across the building's "turn on" times from the derivative analysis are present showing that the pieces of equipment are most likely related to the HVAC system. Knowing that the 9kW pieces of equipment are related to the HVAC system now allows for further analysis into the non-"turn on/off" time occurrences and may help in further disaggregation of Building 1 HVAC. Finally, this analysis was constrained due to the meter resolution of 1.8kW. For buildings that exhibit higher resolutions, additional and smaller, equipment may be identified.

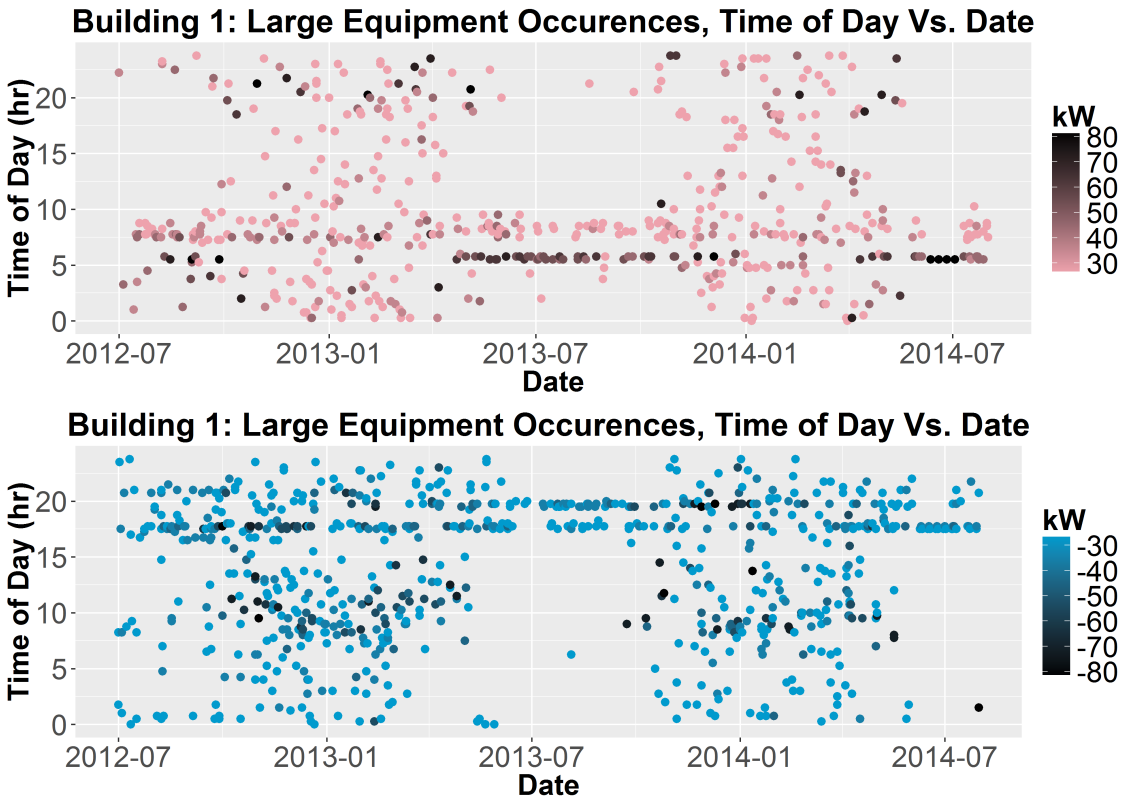

Figure 6.3. *Building 1 Occurrences of Equipment which corresponds to a Multiple of 9kW (27kW - 81kW). Top shows the "Turn On" Times of Equipment and the Bottom shows the "Turn Off" Times of the Equipment.*

7 Discussion

EDIFES is capable of cleaning, processing, and computing analytics on buildings to identify building markers and characteristics, as well as make various energy recommendations. This work focused on the contributions to EDIFES in Early Data Analysis, Derivative Analysis, Classical Time Series Decomposition, Disaggregation, and Equipment Identification. In each section, functions and characteristics were identified to build the EDIFES building marker library.

7.1 Early Data Analysis

Early analysis displayed various issues with the data, such as missing data, mean shifts, and merging three data sets. Functions were developed to automatically handle the various issues and produce clean data primed for analysis. Through analyzing the data with various correlation plots and constraints, it was determined that simple correlations between building electricity and temperature do not identify strong statistical relationships. Those relationships investigated included current interval electricity vs. temperature (-0.44), electricity vs. lagged temperature (-0.5), and mean of day electricity consumption vs. mean of day temperature (-0.56). One of the problems in finding a strong correlation was a result of differing HVAC systems in buildings. EDIFES is now able to

classify the buildings by their HVAC system types as described in Table 2.9, as well as determine the temperature change point in their operation. These findings helped justify approaches to various analyses following. Another marker was identified in the building's meter resolution, determined for all buildings in Table 2.3, which now provides information on the size of equipment that can be disaggregated.

7.2 Derivative Analysis for System Identification

The derivative analysis provides a quantitative method for extracting information on building systems, identifying their relative size and more importantly their scheduling. Derivative analysis is able to identify both stationary systems, such as "turn on/off" times of HVAC systems, but also non-stationary systems such as exterior lighting - a difference between time series decomposition which can only detect and quantify stationary systems. These systems, both their schedules and values, signify building markers which are noted for each building through the function system_finder(). All building systems can be found in Appendix B and C.

7.3 Time Series Decomposition for Building Daily Operation Signatures

Classical time series decomposition of 15-minute building electricity consumption data provides a visual and quantitative understanding on how buildings operate in specific

conditions. Similar to a derivative analysis, time series decomposition identifies building scheduling showing "turn on and off" times in the seasonal component. Additionally, classical time series decomposition is able to characterize the operation of the building before, after, and between those events, which is necessary for quantifying the effect of each. This analysis also provides a means for "searching" the data with various constraints to determine building tendencies through different days of the week, cooling and heating season operations, and to examine how solar irradiance impacts a building. Further questions and constraints may be added to the data for analysis, although statistical significance must be maintained: days > 30 as per the central limit theory[57], and 3-4 years (or more) of data may be required to ask questions of increased specificity. Both an advantage and disadvantage of this analysis is its ability to minimize random behavior. In most cases, the incorporation of over 30 days of data per criteria minimizes random occupancy behavior, but it fails to identify non-stationary systems such as exterior lighting for which a derivative analysis may be necessary.

The seasonal component specifically is, by itself, a marker of the building, as it could be described as the building daily operation signature. The building daily operation signature holds much information on the typical operation of a building, such as the "turn on/off" times that are also found in the derivative analysis and more importantly the behavior of the consumption between the "turn on/off" events. Using the electricity values from the time series analysis, a scheduling recommendation and associated savings can be computed for building systems. In most cases presented in this thesis, the buildings exhibited tendencies to "turn on/off" too early or late. Considering a 7 a.m. - 7 p.m. reschedule for each building the following savings are summarized in Table 7.1. Totaling to 722 MWh and $60,200 per year in savings.

	Energy Savings (MWh) Per Year	Monetary Savings ($) Per Year
Building 1	40	2828
Building 2	483	33810
Building 3	125	8757
Building 4	27	5400
Building 5		
Building 6	47	9400

Table 7.1. Energy and Monetary Savings for all Six Buildings if Rescheduled Major System to a 7a.m. - 7p.m. schedule. Note: Building 4 System was found after Exterior Light Disaggregation.

Constraints such as day of the week, and weather values can be prescribed to the data to determine operation signatures under other conditions. Particularly, the analysis was able to identify graphically the difference in operation between weekdays and weekends, as well as between individual weekdays. The differences between weekdays can assess the occupancy levels in a building for a certain day, or the knowledge of weekend operation signatures can inform a manager to adjust usage parameters during unoccupied times. Further constraints on the weather characteristics were applied to investigate the difference between cooling and heating season operation. Cooling and heating season operation signatures were not only different in magnitude for most of the buildings, but also completely different in shape. Buildings peaked their usage at different times, depending on weather conditions and may have staged HVAC units differently as well. Finally, irradiance criteria was added to the analysis to assess the impact of thermal radiation on the electricity consumption of a building in heating seasons. It was found that cold and cloudy days consumed 1.2 MWh ($84) more than cold sunny days, showing a significant impact. Further analysis is required using irradiance time series to robustly quantify the role of irradiance in building electricity consumption. In conclusion, classical time series decomposition provides an additional tool for understanding

and discovering new operational tendencies and building markers for which to build specific algorithms for characterization.

7.4 Disaggregation

Disaggregation methods and functions were developed to specifically disaggregate exterior lighting and HVAC. Using the derivative analysis, EDIFES was able to detect exterior lights, determine the load associated with them, and then disaggregate them from the total electricity consumption curve. From the disaggregation the total electricity consumption was calculated, and a halogen to HID retrofit in Building 4 could see massive savings in energy and money. Table 7.2 shows these values for Building 4, as well as Building 3 which was also detected to possess exterior lights - these values total 412 MWh and $48,600 of savings per year.

	Energy Savings (MWh) Per Year	Monetary Savings ($) Per Year
Building 3	259	18130
Building 4	153	30514

Table 7.2. Energy and Monetary Savings for Buildings with Exterior Lights Considering a Halogen to HID Retrofit.

Following lighting disaggregation, it was possible to address HVAC disaggregation. Knowing that HVAC electricity consumption is caused by weather conditions, further analysis was conducted to determine a relationship in weather "lags". Where building electricity values were associated with the mean temperature of a past time frame. Using this knowledge, optimal lags were determined for each building, and for each individual

time interval (96 per day at 15-minute intervals). With the lags calculated, linear regression models were developed for each time interval with electricity as a function of temperature. This method led to a model which can describe the electricity consumption of a building as a function of exterior temperature. Table 5.1 displays the mean slope of electricity consumption for each of the six buildings. These values are extremely important as they provide a base for which to continue HVAC disaggregation and also lead to further questions to be answered. One such question is the set point change a building undergoes from unoccupied to occupied settings. Building 4 was analyzed for an HVAC set point change and determined they occurred at 4 a.m. and 9:15 p.m. and corresponded to a change of 4 Fahrenheit. Further analysis can leverage the findings from this work and create a dynamic HVAC disaggregation capability.

7.5 Equipment Identification

Just as difficult as building energy disaggregation (i.e. equipment is a subcomponent of a disaggregated consumption data set), is the identification of equipment from 15-minute interval data. Leveraging an understanding of time integrated consumption data, 70,000+ observations, and that some of the observations would uniquely correspond to time frames which only one piece of "turned on/off" - equipment was identified. Building 1 was analyzed in detail and using a specific set of criteria "24 pieces of equipment" were found. Through further inquiry it was determined that many of these pieces of equipment were multiples of 9kW and also occurred largely at the HVAC "turn on/off" times, showing that the system is comprised of 9kW units. This knowledge is

crucial to HVAC disaggregation and further work will investigate the occurrences outside of the "turn on/off" times to determine what factors prompt HVAC systems to turn on or off.

8 Conclusion

Energy Diagnostics Investigator for Efficiency Savings (EDIFES) has been developed for scalable data analytics to conduct virtual energy audits on commercial buildings. Built as a software package, EDIFES leverages data analytics applied to building electricity data and readily available weather data to determine building markers, characteristics, and operational tendencies. These analytics constitute derivative analysis methods, time series decomposition, disaggregation, and equipment finding. Through these analyses building systems are identified, including Heating Ventilation and Air Conditioning (HVAC), lighting, and plug load or other equipment, with characteristics such as load and system scheduling. Once identified, EDIFES conducts preliminary virtual energy audits to diagnose efficiency issues, determines the impact (i.e. return-on-investment or payback) of potential retrofit actions (e.g. rescheduling HVAC to occupied hours or conducting a lighting retrofit), and then can be used for measurement and verification (M&V) or continuous commissioning. The six buildings presented, from two different climate zones, were analyzed in full through EDIFES and resulted in over a GWh/year of energy recommendations associated with over $100,000 in yearly cost savings. Additionally, preliminary disaggregation of load components in a building, crucial for diagnostics and prognostics, was explored with an approach to quantify HVAC systems as

a function of temperature. Each of the six buildings were also characterized with each of EDIFES 40+ functions and these traits are continually documented as more buildings undertake analysis in the hope of creating the Building Energy Genome. This analysis requires simply a building's utility 15-minute interval electricity data, approximate location, and squarefootage to enable full capabilities.

9 Future Work

The current state of EDIFES provides a base software package from which to build further analytics and much more sophisticated processes. The first area of further work is in disaggregation - the corner stone of building energy efficiency analysis issues. Disaggregation allows analysis to dig deeper into specific components, allowing algorithms to focus on particular characteristics. The most difficult, and most important component of disaggregation is HVAC. This work was investigated as an approach for quantifying the effect of temperature on building electricity consumption, but has yet to be fully used in disaggregation. Irradiance must also be included in the HVAC disaggregation model. Perhaps in a similar manner as temperature - adding a second variable to the regression, as its significant impact was shown through time series analysis. Once disaggregation of the HVAC component is finished it will uncover the plug load in a building leading to a whole new set of plug load analysis questions and functions.

The formal creation of the Building Energy Genome must also be developed to ensure EDIFES can be a malleable software tool which learns from each building analyzed. Currently, various characteristics are observed and noted. In future versions of EDIFES these characteristics will be tied to other particular traits in the building - these traits

may point to certain specific analyses resulting in much more sophisticated and detailed virtual energy audits. EDIFES may also benefit from machine learning algorithms to dynamically assess buildings and uncover new characteristics which may never have been noticed otherwise.

Considering these advancements, EDIFES must be implemented entirely in a high performance computing cluster with large data base capabilities. As EDIFES expands, so too does computation time - currently a full analysis on two years of data requires 20+ hours of computation time on 3GB of memory using a 2 GHz dual core processor. A high performance computation can cut this time to within minutes, necessary for marketability. Further, data bases for building energy must be adequate to hold thousands of buildings and allow access to the entirety of the data sets for analysis. As EDIFES adds further capabilities, each one must be tested throughout all buildings to ensure its applicability. Finally, EDIFES requires a consumer friendly user interface to allow for simple understanding of EDIFES data requirements and access to virtual energy outputs. With data analytics as an emerging field, the capabilities described above are entirely plausible and will create a software that has the potential to save massive sums of energy and money.

Appendix A

Preparation of this document

This document was prepared using pdfLaTeX. The (free) programs and versions implemented are as follows (links to the current versions are included):

- LaTeX implementation: **MiKTeX 2.9**

  http://www.miktex.org/

- TeX-oriented editing environments: **TeXStudio**

  http://www.texstudio.org/

- Bibliographical Database: **BibTeX**

  http://www.bibtex.org/

  and **Zotero**

  https://www.zotero.org/

Appendix B

Appendix Figures

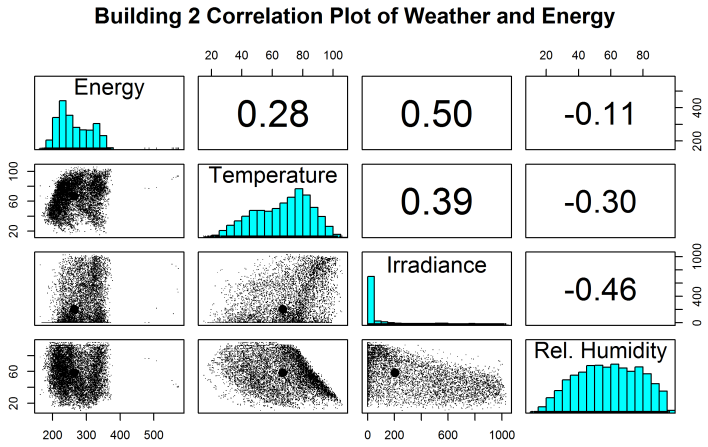

Figure B.1. *Correlation Plot of Weather Data and Energy for Building 2*

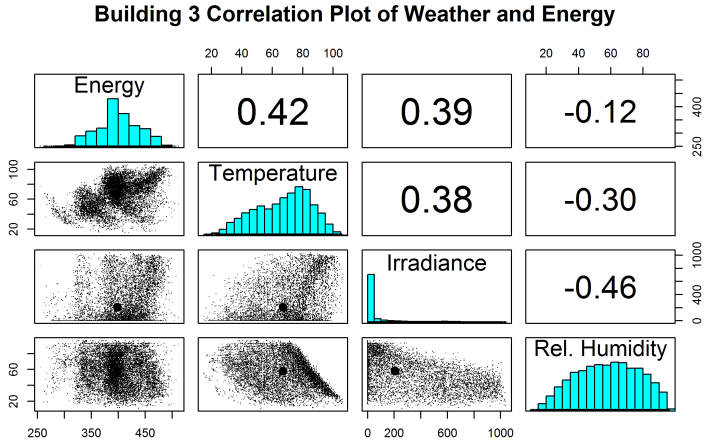

Figure B.2. *Correlation Plot of Weather Data and Energy for Building 3*

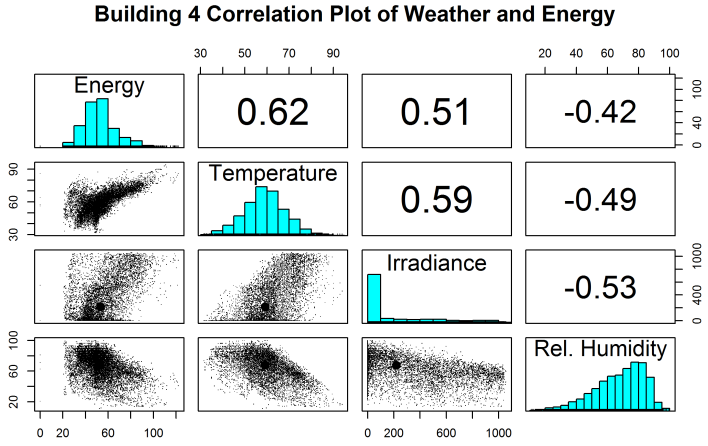

Figure B.3. *Correlation Plot of Weather Data and Energy for Building 4*

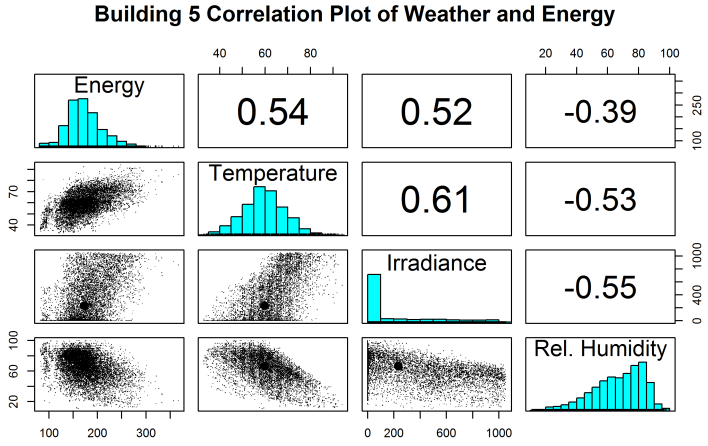

Figure B.4. *Correlation Plot of Weather Data and Energy for Building 5*

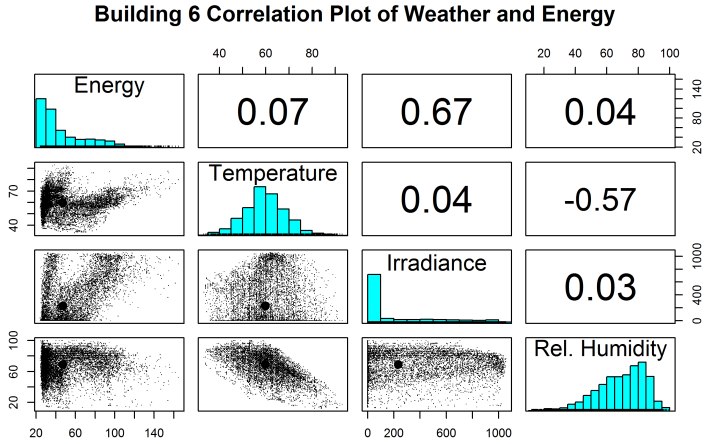

Figure B.5. *Correlation Plot of Weather Data and Energy for Building 6*

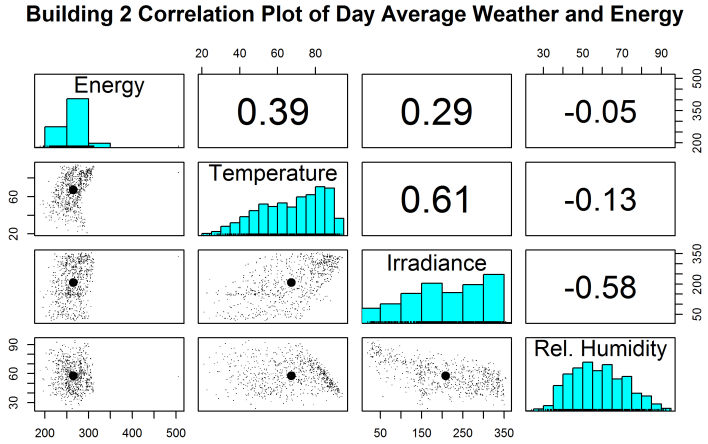

Figure B.6. *Correlation Plot of Average Weather and Energy Data for Building 2*

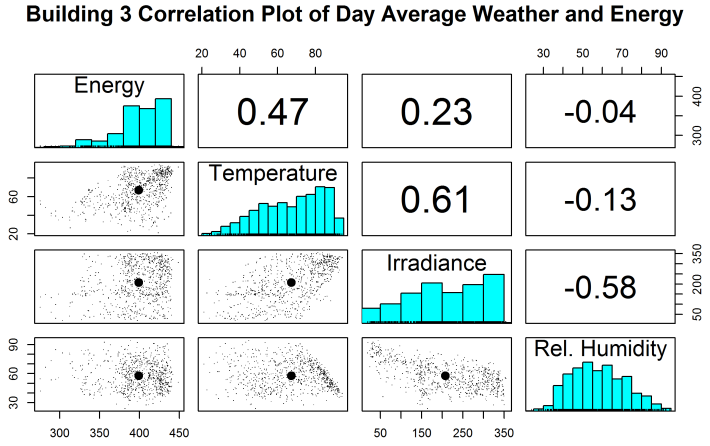

Figure B.7. *Correlation Plot of Average Weather and Energy Data for Building 3*

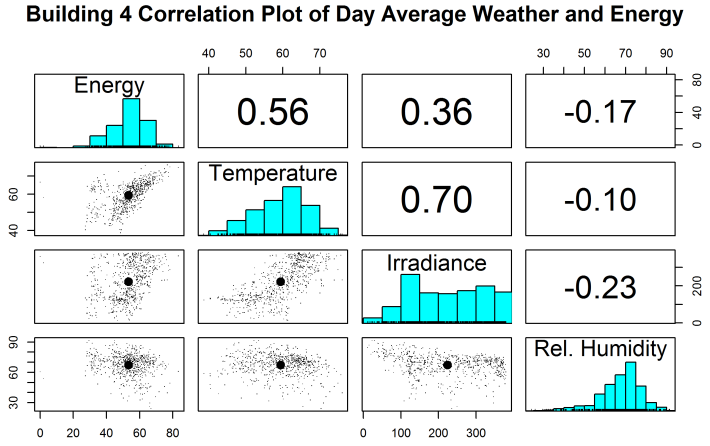

Figure B.8. *Correlation Plot of Average Weather and Energy Data for Building 4*

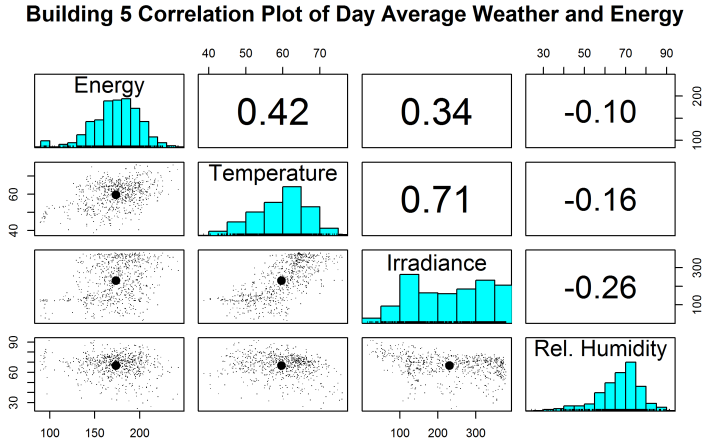

Figure B.9. *Correlation Plot of Average Weather and Energy Data for Building 5*

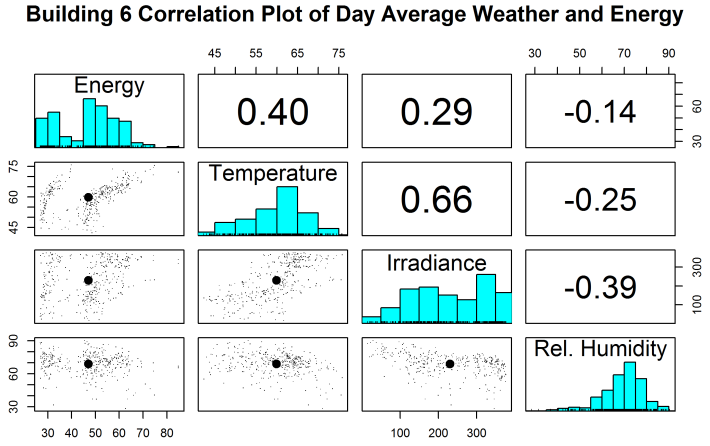

Figure B.10. *Correlation Plot of Average Weather and Energy Data for Building 6*

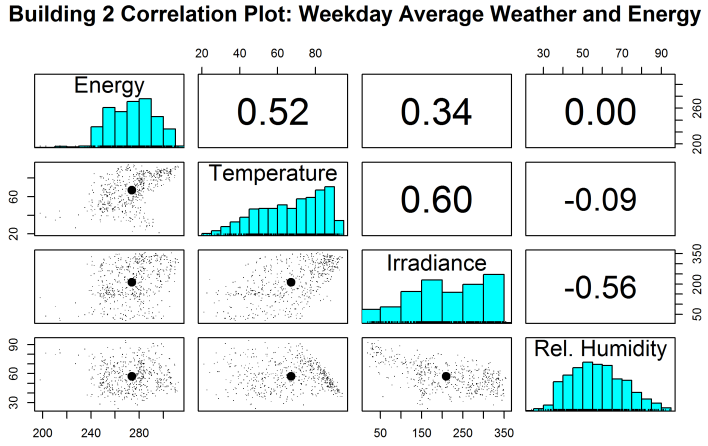

Figure B.11. *Correlation Plot of Average Weather and Energy Data of Weekdays for Building 2*

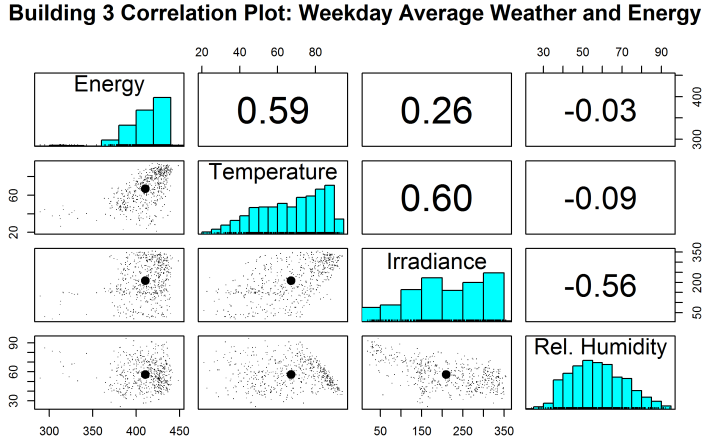

Figure B.12. *Correlation Plot of Average Weather and Energy Data of Weekdays for Building 3*

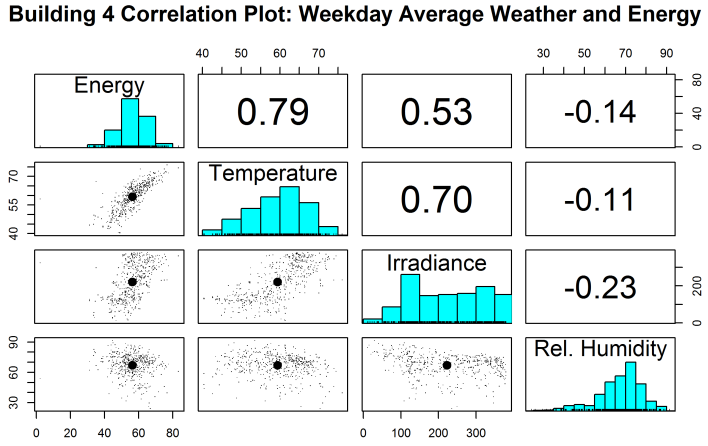

Figure B.13. *Correlation Plot of Average Weather and Energy Data of Weekdays for Building 4*

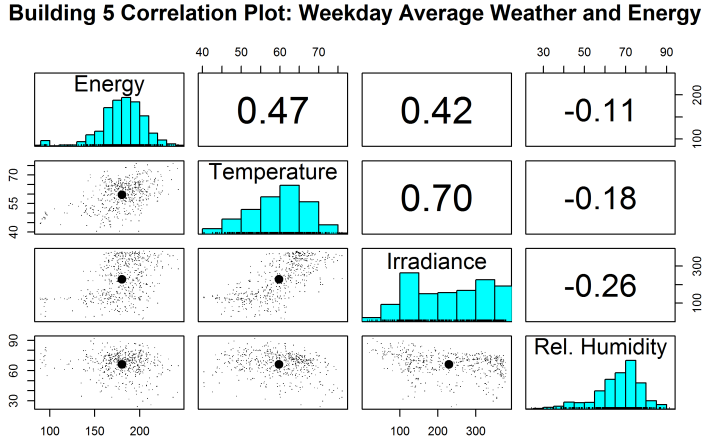

Figure B.14. *Correlation Plot of Average Weather and Energy Data of Weekdays for Building 5*

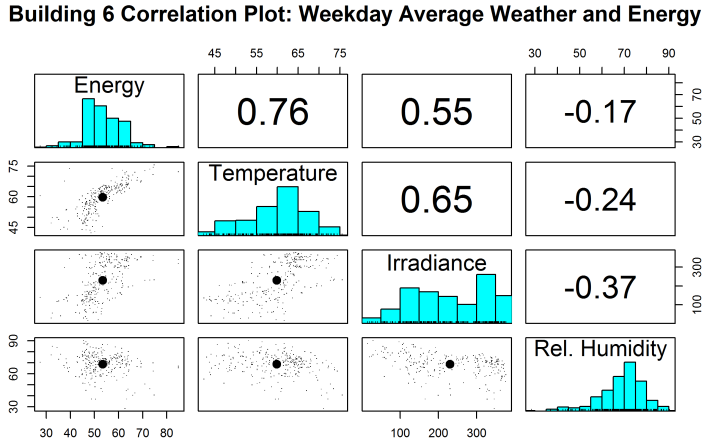

Figure B.15. *Correlation Plot of Average Weather and Energy Data of Weekdays for Building 6*

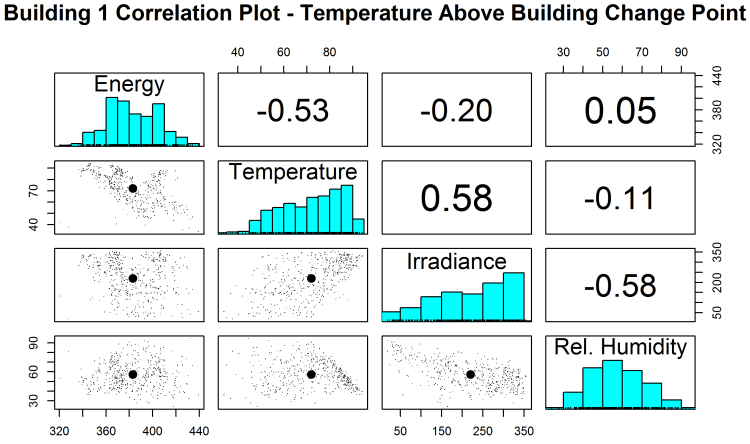

Figure B.16. *Correlation Plots of Weather and Energy Data Above the Change Point, Building 1*

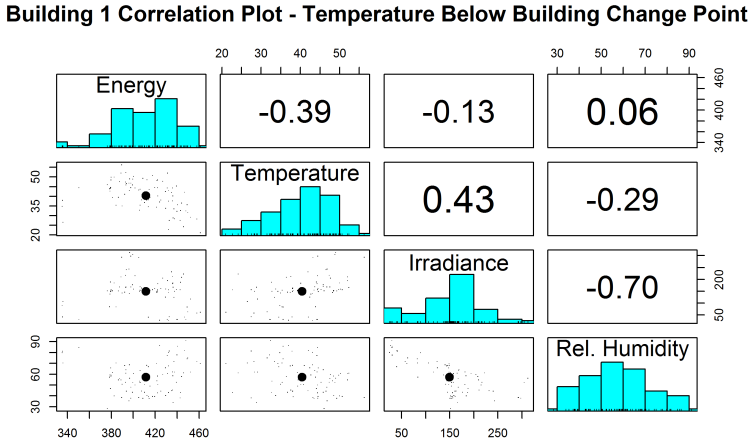

Figure B.17. *Correlation Plots of Weather and Energy Data Below the Change Point, Building 1*

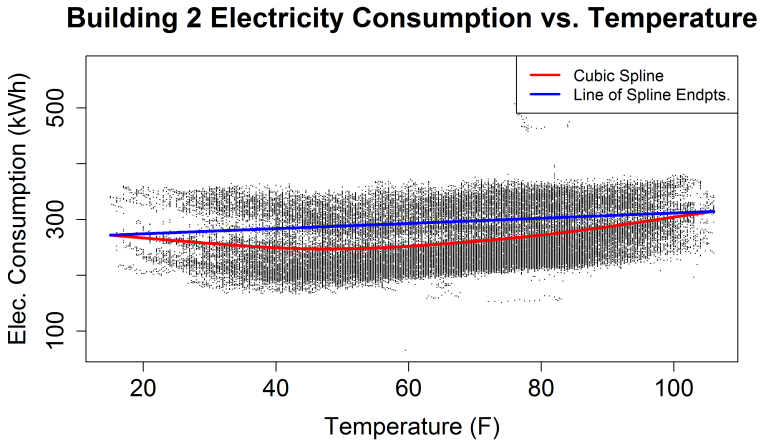

Figure B.18. *Energy vs. Temperature with a fitted spline and line through endpoints, Building 2*

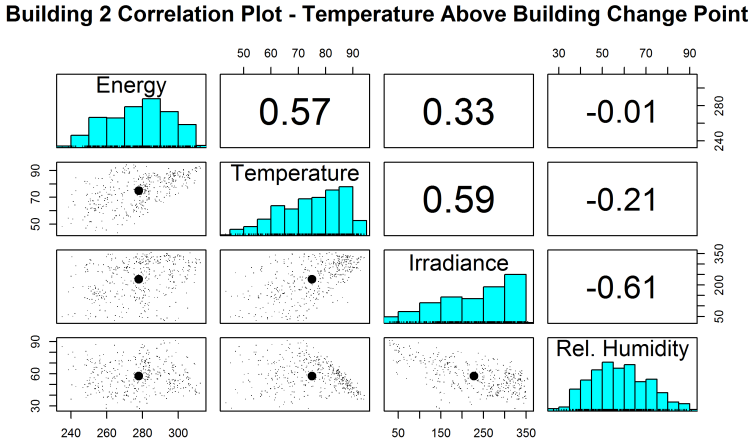

Figure B.19. *Correlation Plots of Weather and Energy Data Above the Change Point, Building 2*

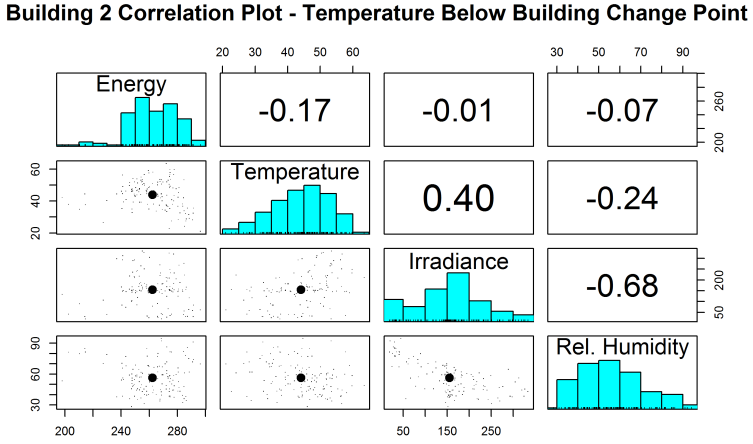

Figure B.20. *Correlation Plots of Weather and Energy Data Below the Change Point, Building 2*

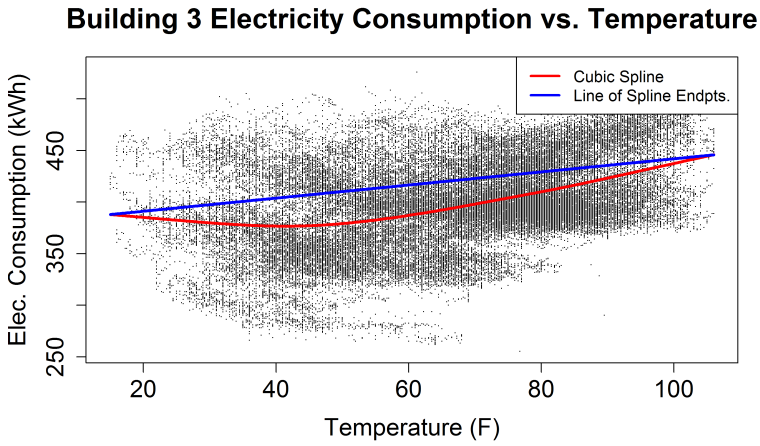

Figure B.21. *Energy vs. Temperature with a fitted spline and line through endpoints, Building 3*

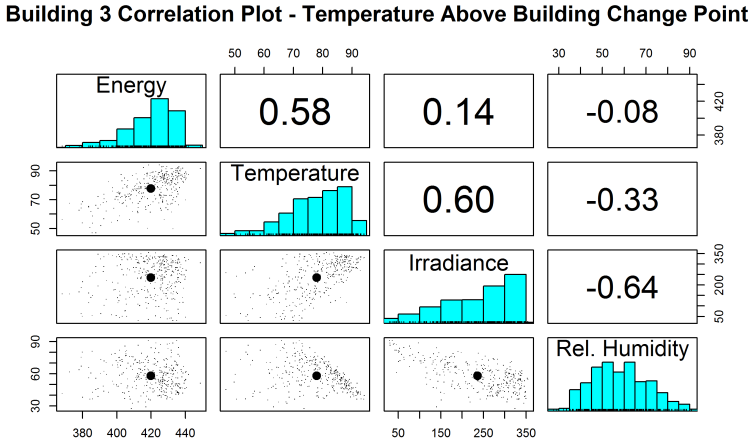

Figure B.22. *Correlation Plots of Weather and Energy Data Above the Change Point, Building 3*

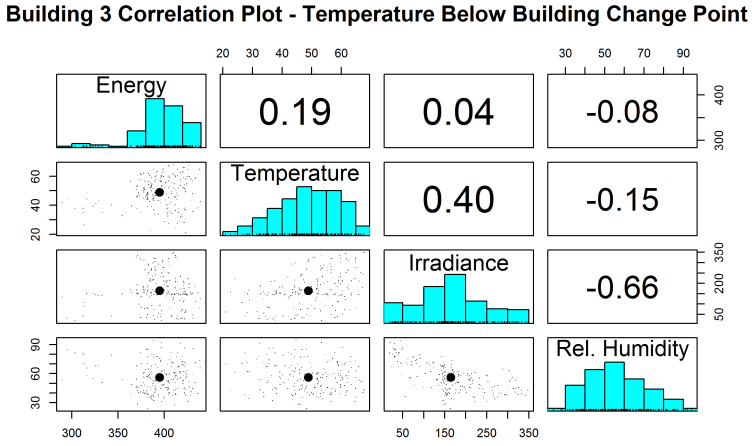

Figure B.23. *Correlation Plots of Weather and Energy Data Below the Change Point, Building 3*

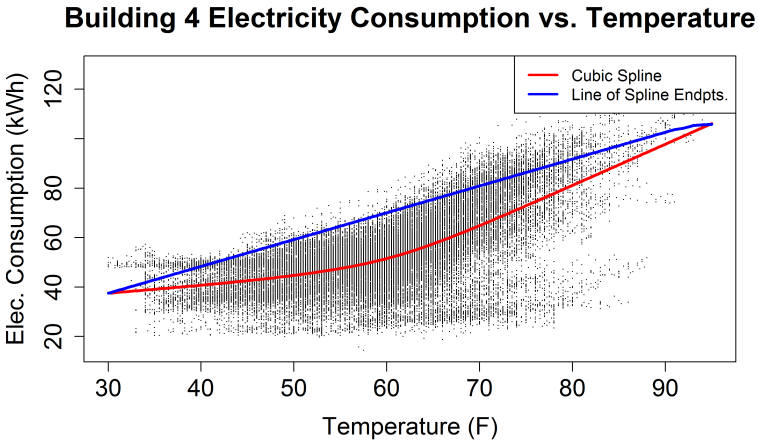

Figure B.24. *Energy vs. Temperature with a fitted spline and line through endpoints, Building 4*

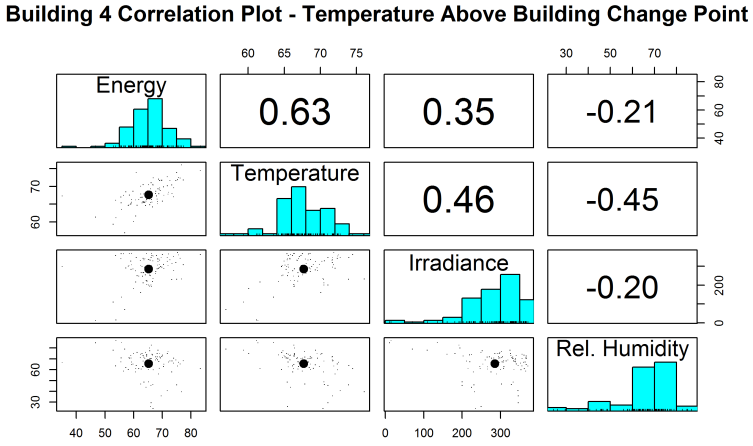

Figure B.25. *Correlation Plots of Weather and Energy Data Above the Change Point, Building 4*

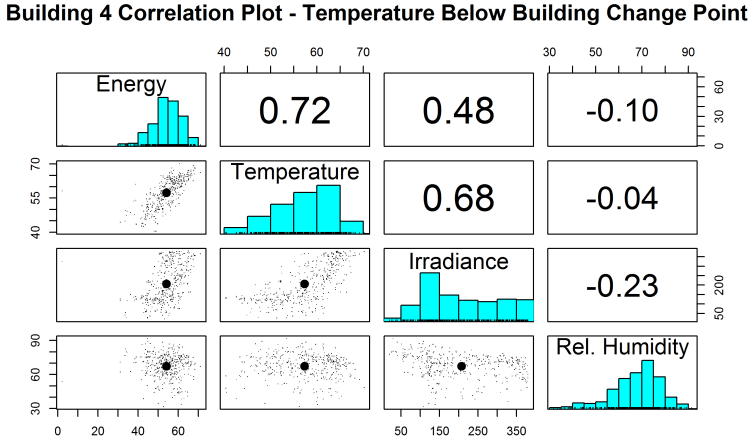

Figure B.26. *Correlation Plots of Weather and Energy Data Below the Change Point, Building 4*

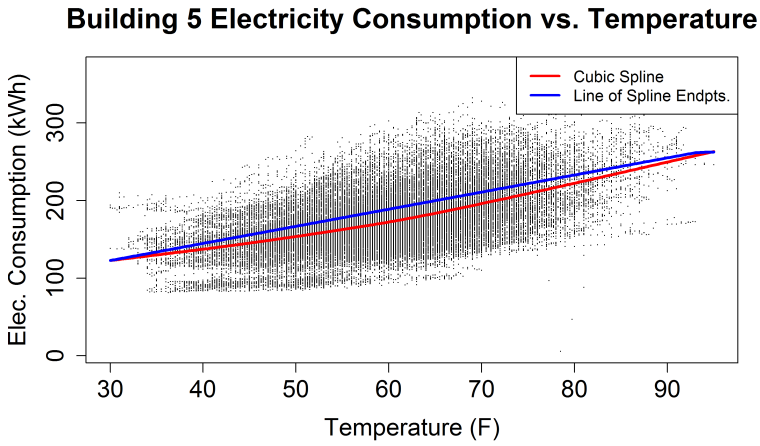

Figure B.27. *Energy vs. Temperature with a fitted spline and line through endpoints, Building 5*

Figure B.28. *Correlation Plots of Weather and Energy Data Above the Change Point, Building 5*

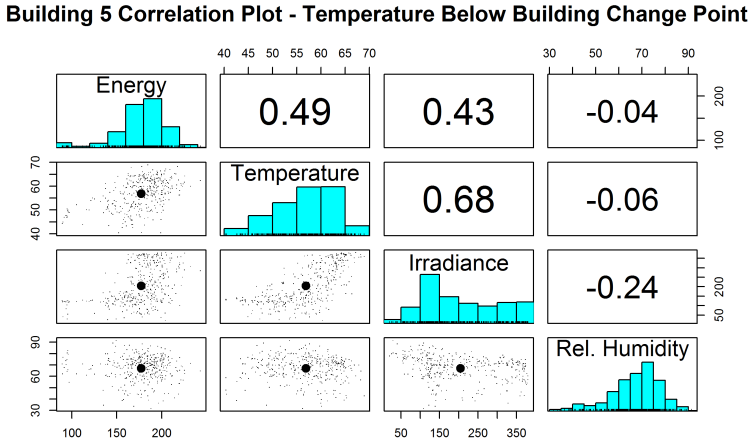

Figure B.29. *Correlation Plots of Weather and Energy Data Below the Change Point, Building 5*

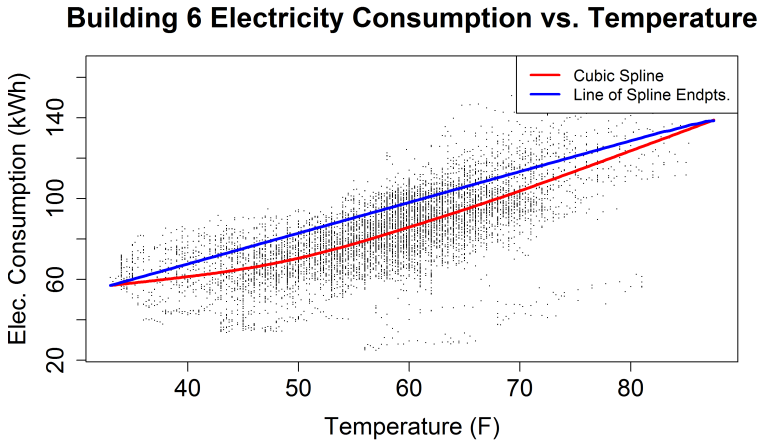

Figure B.30. *Energy vs. Temperature with a fitted spline and line through endpoints, Building 6*

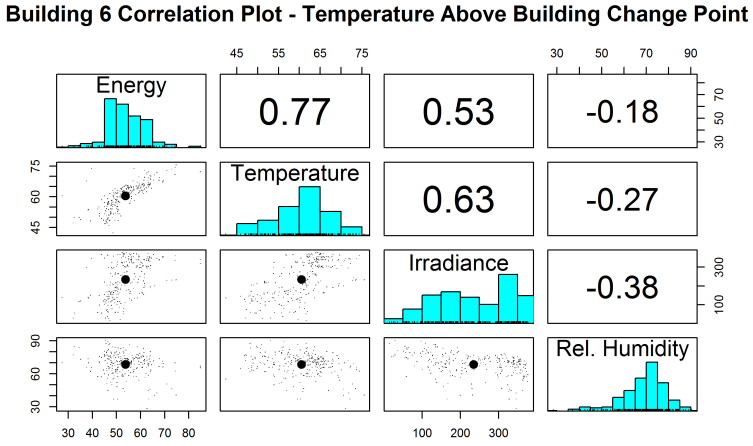

Figure B.31. *Correlation Plots of Weather and Energy Data Above the Change Point, Building 6*

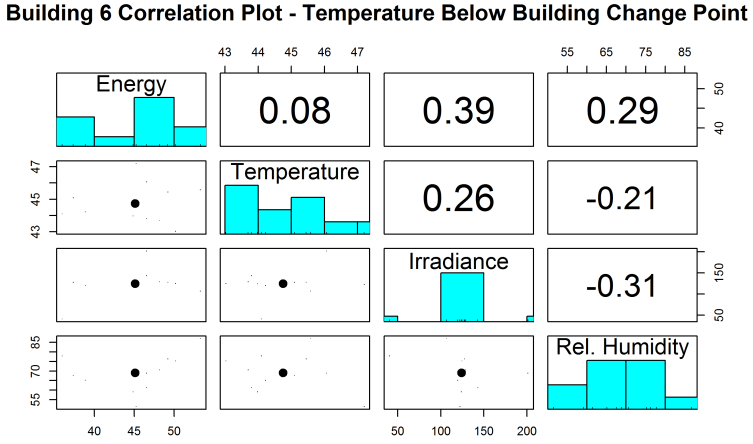

Figure B.32. *Correlation Plots of Weather and Energy Data Below the Change Point, Building 6*

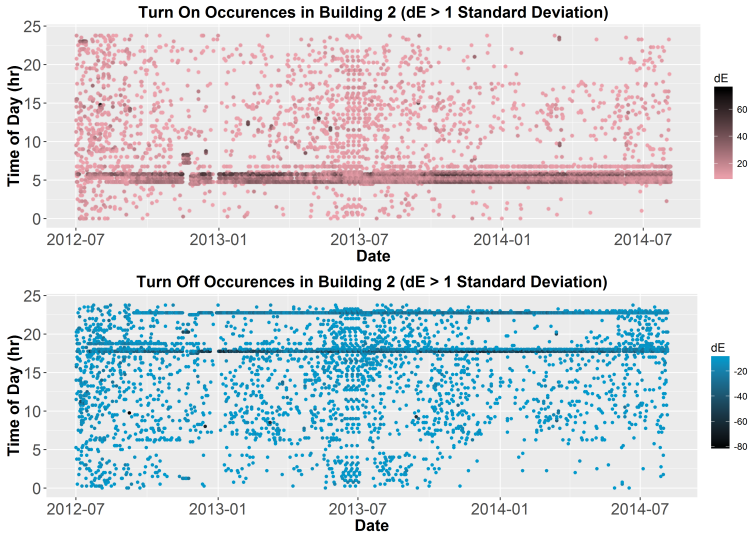

Figure B.33. *Building 2 Turn On and Off Events, Standard Deviation +1/-1*

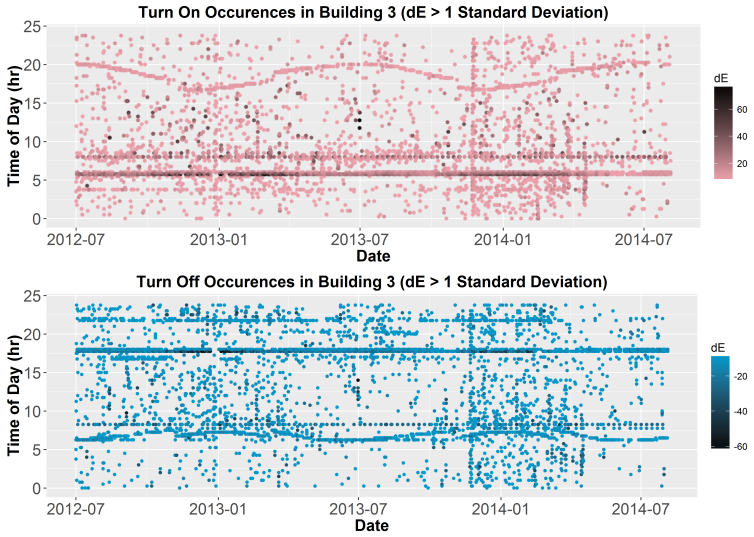

Figure B.34. *Building 3 Turn On and Off Events, Standard Deviation +1/-1*

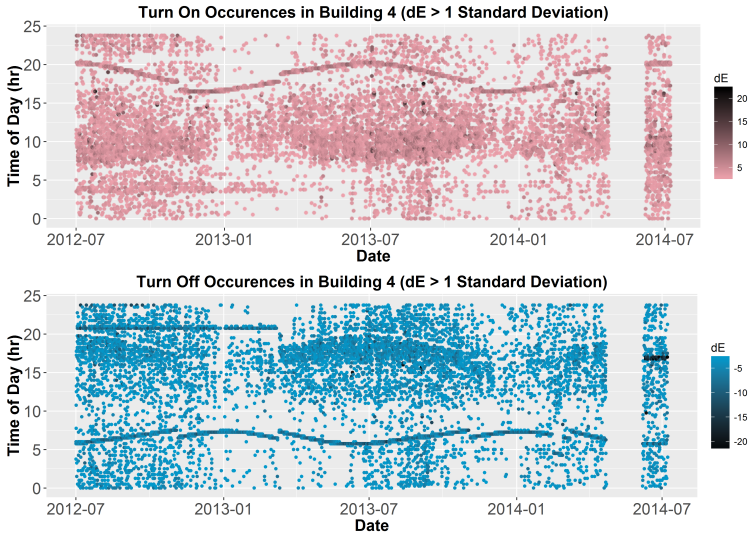

Figure B.35. *Building 4 Turn On and Off Events, Standard Deviation +1/-1*

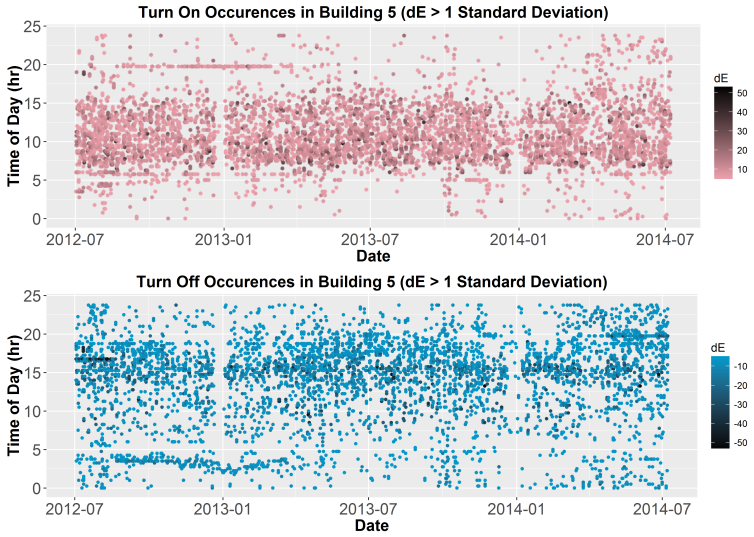

Figure B.36. *Building 5 Turn On and Off Events, Standard Deviation +1/-1*

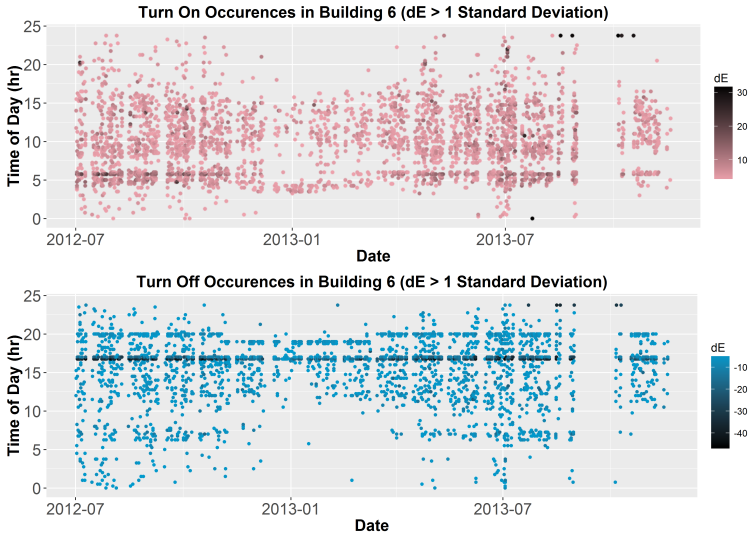

Figure B.37. *Building 6 Turn On and Off Events, Standard Deviation +1/-1*

Appendix C

Appendix Tables

1 Derivative Analysis

	Occurences	Time	Average dE (kWh)	Frequency	Energy (W)
1	690.00	00:04:45	27.84	0.91	19209.60
2	524.00	00:05:45	39.26	0.69	20572.20
3	490.00	00:05:15	21.12	0.64	10346.85
4	297.00	00:06:00	14.06	0.39	4175.55
5	260.00	00:06:45	11.78	0.34	3063.15
6	219.00	00:05:00	11.68	0.29	2557.80
7	206.00	00:05:30	13.70	0.27	2822.40

Table C.1. Turn On Events for Building 2

	Occurences	Time	Average dE (kWh)	Frequency	Energy (W)
1	657.00	00:22:45	-22.28	0.86	-14639.40
2	538.00	00:17:45	-36.83	0.71	-19812.15
3	372.00	00:18:00	-14.14	0.49	-5261.85

Table C.2. Turn Off Events for Building 2

	Occurences	Time	Average dE (kWh)	Frequency	Energy (W)
1	535.00	00:05:45	36.29	0.70	19415.70
2	410.00	00:06:00	13.14	0.54	5387.40
3	218.00	00:08:00	22.43	0.29	4888.80

Table C.3. Turn On Events for Building 3

	Occurences	Time	Average dE (kWh)	Frequency	Energy (W)
1	518.00	00:17:45	-22.12	0.68	-11460.15
2	393.00	00:18:00	-11.16	0.52	-4384.80
3	207.00	00:21:45	-10.43	0.27	-2159.10
4	198.00	00:06:15	-11.51	0.26	-2278.35
5	160.00	00:08:15	-19.04	0.21	-3046.05

Table C.4. Turn Off Events for Building 3

	Occurences	Time	Average dE (kWh)	Frequency	Energy (W)
1	225.00	00:09:45	4.85	0.31	1090.29
2	213.00	00:10:15	4.71	0.29	1004.03
3	212.00	00:09:15	4.81	0.29	1020.27

Table C.5. Turn On Events for Building 4

	Occurences	Time	Average dE (kWh)	Frequency	Energy (W)
1	235.00	00:20:45	-7.10	0.32	-1668.61
2	214.00	00:17:00	-5.69	0.29	-1218.35
3	208.00	00:07:15	-6.65	0.28	-1384.03

Table C.6. Turn Off Events for Building 4

	Occurences	Time	Average dE (kWh)	Frequency	Energy (W)
1	173.00	00:08:30	11.87	0.24	2053.41
2	154.00	00:09:00	11.33	0.21	1744.08
3	154.00	00:09:30	10.54	0.21	1622.61
4	151.00	00:08:45	11.60	0.21	1750.86
5	149.00	00:08:15	12.42	0.20	1851.27

Table C.7. Turn On Events for Building 5

	Occurences	Time	Average dE (kWh)	Frequency	Energy (W)
1	173.00	00:15:30	-12.57	0.24	-2175.03
2	165.00	00:18:45	-8.08	0.23	-1333.41
3	153.00	00:15:15	-13.77	0.21	-2107.41

Table C.8. Turn Off Events for Building 5

	Occurences	Time	Average dE (kWh)	Frequency	Energy (W)
1	179.00	00:05:45	10.89	0.43	1949.22

Table C.9. Turn On Events for Building 6

	Occurences	Time	Average dE (kWh)	Frequency	Energy (W)
1	271.00	00:16:45	-27.46	0.65	-7442.68
2	243.00	00:17:00	-13.76	0.58	-3344.04
3	173.00	00:20:00	-7.09	0.41	-1226.38

Table C.10. Turn Off Events for Building 6

ïż£

Complete References

[1] US Senate Briefing. International energy outlook 2013. US Energy Information Administration, 2013. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.369.6789&rep=rep1&type=pdf, doi:10.1.1.369.6789.

[2] David Austin. Addressing market barriers to energy efficiency in buildings. Congressional Budget Office, 2012. http://cbo.gov/sites/default/files/cbofiles/attachments/AddressingMarketBarriersToEnergyEfficiencyInBuildings_WorkingPaper_2012-10.pdf.

[3] Financing small commercial building energy performance upgrades: Challenges and opportunities. National Institute of Building Sciences, Council on Finance, Insurance and Real Estate, 2014. https://c.ymcdn.com/sites/www.nibs.org/resource/resmgr/CC/CFIRE_CommBldgFinance-Final.pdf.

[4] Better buildings challenge. United States Department of Energy, Building Technologies Office. URL: http://energy.gov/eere/efficiency/buildings.

[5] Niccolo Aste and Claudio Del Pero. Energy retrofit of commercial buildings: case study and applied methodology. Energy Efficiency, 6(2):407–423, 2013. doi:10.1007/s12053-012-9168-4.

[6] Bo Shen, Lynn Price, and Hongyou Lu. Energy audit practices in china: National and local experiences and issues. Energy Policy, 46:346–358, 2012. doi:10.1016/j.enpol.2012.03.069.

[7] PE David Gerlach PhD. A case study of multiple energy audits of the same building: Conclusions and recommendations. ASHRAE Transactions, 120:GG1, 2014. http://search.proquest.com/openview/b599a9420d997ee048f655f3a12e44b5/1?pq-origsite=gscholar.

[8] Surapong Chirarattananon and Juntakan Taweekun. A technical review of energy conservation programs for commercial and government buildings in thailand. Energy Conversion and Management, 44(5):743–762, 2003. URL: http://www.sciencedirect.com/science/article/pii/S0196890402000821, doi:10.1016/S0196-8904(02)00082-1.

[9] J. Harris, J. Anderson, and W. Shafron. Investment in energy efficiency: a survey of australian firms. Energy Policy, 28(12):867–876, 2000. doi:10.1016/S0301-4215(00)00075-6.

[10] K. J. Lomas, H. Eppel, C. J. Martin, and D. P. Bloomfield. Empirical validation of building energy simulation programs. Energy and Buildings, 26(3):253–275, 1997. doi:10.1016/S0378-7788(97)00007-8.

[11] Drury B. Crawley, Jon W. Hand, Michael Kummert, and Brent T. Griffith. Contrasting the capabilities of building energy performance simulation programs. Building and Environment, 43(4):661–673, 2008. URL: http://www.sciencedirect.com/science/article/pii/S0360132306003234, doi:10.1016/j.buildenv.2006.10.027.

[12] P. G. Loutzenhiser, H. Manz, C. Felsmann, P. A. Strachan, T. Frank, and G. M. Maxwell. Empirical validation of models to compute solar irradiance on inclined surfaces for building energy simulation. Solar Energy, 81(2):254–267, 2007. doi:10.1016/j.solener.2006.03.009.

[13] M. Mirsadeghi, D. Costola, B. Blocken, and J. L. M. Hensen. Review of external convective heat transfer coefficient models in building energy simulation programs: Implementation and uncertainty. Applied Thermal Engineering, 56(1):134–151, 2013. doi:10.1016/j.app1thermaleng.2013.03.003.

[14] Joshua D. Rhodes, William H. Gorman, Charles R. Upshaw, and Michael E. Webber. Using BEopt (EnergyPlus) with energy audits and surveys to predict actual residential energy usage. Energy and Buildings, 86:808–816, 2015. URL: http://www.sciencedirect.com/science/article/pii/S0378778814009220, doi:10.1016/j.enbuild.2014.10.076.

[15] D. B. Crawley, L. K. Lawrie, F. C. Winkelmann, W. F. Buhl, Y. J. Huang, C. O. Pedersen, R. K. Strand, R. J. Liesen, D. E. Fisher, M. J. Witte, and J. Glazer. EnergyPlus: creating a new-generation building energy simulation program. Energy and Buildings, 33(4):319–331, 2001. doi:10.1016/S0378-7788(00)00114-6.

[16] T. Agami Reddy, Itzhak Maor, and Chanin Panjapornpon. Calibrating detailed building energy simulation programs with measured data - part 1: General methodology (RP-1051). HVAC&R Research, 13(2):221–241, 2007. doi:10.1080/10789669.2007.10390952.

[17] Nelson Fumo. A review on the basics of building energy estimation. Renewable & Sustainable Energy Reviews, 31:53–60, 2014. doi:10.1016/j.rser.2013.11.040.

[18] Nan Li, Zheng Yang, Burcin Becerik-Gerber, Chao Tang, and Nanlin Chen. Why is the reliability of building simulation limited as a tool for evaluating energy conservation measures? Applied Energy, 159:196–205, 2015. doi:10.1016/j.apenergy.2015.09.001.

[19] Hai-xiang Zhao and Frederic Magoules. A review on the prediction of building energy consumption. Renewable & Sustainable Energy Reviews, 16(6):3586–3592, 2012. doi:10.1016/j.rser.2012.02.049.

[20] Margaret F. Fels. PRISM: An introduction. Energy and Buildings, 9(1):5–18, 1986. URL: http://www.sciencedirect.com/science/article/pii/0378778886900034, doi:10.1016/0378-7788(86)90003-4.

[21] J. K. Kissock and C.W. Eger. Understanding industrial energy use through sliding regression analysis, 2007. URL: https://www.udayton.edu/engineering/centers/industrial_assessment/resources/docs/pdf/UnderstandingIndEnergySlidingRegAnal_ACEEE_2007.pdf.

[22] International Performance Measurement & Verification Protocol: Concepts and Options for Determining Energy and Water Savings, volume 1. 2002. URL: http://www.nrel.gov/docs/fy02osti/31505.pdf.

[23] J. S. Haberl. ASHRAE's guideline 14-2002 for measurement of energy and demand savings: How to determine what was really saved by the retrofit. In Proceedings of the Fifth International Conference for Enhanced Building Operations. URL: http://esl.tamu.edu/docs/terp/2005/esl-ic-05-10-50.pdf.

[24] M. F. Fels, K. Kissock, M. A. Marean, and C. Reynolds. Advancing the art of prism analysis. Home Energy, 12(4), 1995. URL: http://www.osti.gov/scitech/biblio/460392.

[25] ENERGY STAR performance ratings technical methodology. URL: https://www.energystar.gov/ia/business/evaluate_performance/supermarket_tech_desc.pdf.

[26] J. K. Kissock, J. S. Haberl, and D. E. Claridge. Inverse modeling toolkit: Numerical algorithms for best-fit variable-base degree day and change point models. 2003. URL: http://oaktrust.library.tamu.edu/handle/1969.1/153708.

[27] J. K. Kissock, T. A. Reddy, and D. E. Claridge. Ambient-temperature regression analysis for estimating retrofit savings in commercial buildings. Journal of Solar Energy Engineering, 120(3):168–176, 1998. URL: http://dx.doi.org/10.1115/1.2888066, doi:10.1115/1.2888066.

[28] F. Sever, K. Kissock, D. Brown, and S. Mulqueen. Estimating industrial building energy savings using inverse simulation. ASHRAE Transactions, 117, 2011. URL: http://academic.udayton.edu/kissock/http/Publications/ASHRAE2011-86073_Final.pdf.

[29] ARPA-e OPEN 2015 project selections. U.S. Department of Energy, 2015-11-23. URL: http://arpa-e.energy.gov/sites/default/files/documents/files/OPEN_2015_Project_Descriptions.pdf.

[30] Clayton Miller, Zoltan Nagy, and Arno Schlueter. Automated daily pattern filtering of measured building performance data. Automation in Construction, 49:1–17, 2015. doi:10.1016/j.autcon.2014.09.004.

[31] Imran Khan, Alfonso Capozzoli, Stefano Paolo Corgnati, and Tania Cerquitelli. Fault detection analysis of building energy consumption using data mining techniques. Energy Procedia, 42:557–566, 2013. URL: http://www.sciencedirect.com/science/article/pii/S1876610213017591, doi:doi:10.1016/j.egypro.2013.11.057.

[32] Jui-Sheng Chou and Abdi Suryadinata Telaga. Real-time detection of anomalous power consumption. Renewable & Sustainable Energy Reviews, 33:400–411, 2014-05. doi:10.1016/j.rser.2014.01.088.

[33] Halldor Janetzko, Florian Stoffel, Sebastian Mittelstaedt, and Daniel A. Keim. Anomaly detection for visual analytics of power consumption data. Computers & Graphics-Uk, 38:27–37, 2014. doi:10.1016/j.cag.2013.10.006.

[34] Tianzhen Hong, Mary Ann Piette, Yixing Chen, Sang Hoon Lee, Sarah C. Taylor-Lange, Rongpeng Zhang, Kaiyu Sun, and Phillip Price. Commercial building energy saver: An energy retrofit analysis toolkit. Applied Energy, 159:298–309, 2015. doi:10.1016/j.apenergy.2015.09.002.

[35] O. T. Masoso and L. J. Grobler. The dark side of occupants' behaviour on building energy use. Energy and Buildings, 42(2):173–177, 2010. doi:10.1016/j.enbuild.2009.08.009.

[36] Richard Barras and D Ferguson. A spectral analysis of building cycles in britain. Environment and Planning A, 17(10):1369–1391, 1985.

[37] Lambert H Koopmans. The spectral analysis of time series. Academic press, 1995.

[38] Andrew V. Metcalfe and Paul S.P. Cowpertwait. Introductory Time Series with R. Springer New York, 2009. URL: http://link.springer.com/10.1007/978-0-387-88698-5.

[39] S. Beveridge and Cr Nelson. A new approach to decomposition of economic time-series into permanent and transitory components with particular attention to measurement of the business-cycle. Journal of Monetary Economics, 7(2):151–174, 1981. doi:10.1016/0304-3932(81)90040-4.

[40] Cs Bretherton, C. Smith, and Jm Wallace. An intercomparison of methods for finding coupled patterns in climate data. Journal of Climate, 5(6):541–560, 1992. doi:10.1175/1520-0442(1992)005<0541:AIOMFF>2.0.CO;2.

[41] Bw Ang. Decomposition methodology in industrial energy demand analysis. Energy, 20(11):1081–1095, 1995-11. doi:10.1016/0360-5442(95)00068-R.

[42] A. M. M. Masih and R. Masih. Energy consumption, real income and temporal causality: Results from a multi-country study based on cointegration and error-correction modelling techniques. Energy Economics, 18(3):165–183, 1996. doi:10.1016/0140-9883(96)00009-6.

[43] Luca Ghelardoni, Alessandro Ghio, and Davide Anguita. Energy load forecasting using empirical mode decomposition and support vector regression. Ieee Transactions on Smart Grid, 4(1):549–556, 2013. WOS:000325485600059. doi:10.1109/TSG.2012.2235089.

[44] Akin Tascikaraoglu and Borhan M. Sanandaji. Short-term residential electric load forecasting: A compressive spatio-temporal approach. Energy and Buildings, 111:380–392, 2016. WOS:000369191100034. doi:10.1016/j.enbuild.2015.11.068.

[45] K. Carrie Armel, Abhay Gupta, Gireesh Shrimali, and Adrian Albert. Is disaggregation the holy grail of energy efficiency? the case of electricity. Energy Policy, 52:213–234, 2013. URL: http://www.sciencedirect.com/science/article/pii/S0301421512007446, doi:10.1016/j.enpol.2012.08.062.

[46] M. L. Marceau and R. Zmeureanu. Nonintrusive load disaggregation computer program to estimate the energy consumption of major end uses in residential buildings. Energy Conversion and Management, 41(13):1389–1403, 2000. URL: http://www.sciencedirect.com/science/article/pii/S0196890499001739, doi:10.1016/S0196-8904(99)00173-9.

[47] J. Zico Kolter and Matthew J. Johnson. REDD: A public data set for energy disaggregation research. In Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA, volume 25, pages 59–62. Citeseer, 2011. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.454.5796&rep=rep1&type=pdf, doi:10.1.1.454.5796.

[48] R. Dong, L. J. Ratliff, H. Ohlsson, and S. S. Sastry. Energy disaggregation via adaptive filtering. In 2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton), pages 173–180, 2013. doi:10.1109/Allerton.2013.6736521.

[49] R. Dong, L. Ratliff, H. Ohlsson, and S. S. Sastry. A dynamical systems approach to energy disaggregation. In 52nd IEEE Conference on Decision and Control, pages 6335–6340, 2013-12. doi:10.1109/CDC.2013.6760891.

[50] J. Zico Kolter and Tommi Jaakkola. Approximate inference in additive factorial hmms with application to energy disaggregation. In International conference on artificial intelligence and statistics, pages 1472–1482, 2012. URL: http://machinelearning.wustl.edu/mlpapers/paper_files/AISTATS2012_KolterJ12.pdf.

[51] Franz Rubel and Markus Kottek. Comments on:"the thermal zones of the earth" by Wladimir Koppen (1884). Meteorologische Zeitschrift, 20(3):361–365, 2011. URL: http://www.ingentaconnect.com/content/schweiz/mz/2011/00000020/00000003/art00010.

[52] Climatic data online - application router, 2016-01-12. URL: http://www7.ncdc.noaa.gov/CDO/cdopoemain.cmd?datasetabbv=DS3505&countryabbv=&georegionabbv=&resolution=40.

[53] Owen Vallis, Jordan Hochenbaum, and Arun Kejariwal. A novel technique for long-term anomaly detection in the cloud. In 6th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 14), Philadelphia, PA, June 2014. USENIX Association. URL: https://www.usenix.org/conference/hotcloud14/workshop-program/presentation/vallis.

[54] William Revelle. psych: Procedures for psychological, psychometric, and personality research. Northwestern University, Evanston, Illinois, page 165, 2014.

[55] Richard Taylor. Interpretation of the correlation coefficient: A basic review. Journal of Diagnostic Medical Sonography, 6(1):35–39, 1990-01-01. URL: http://jdm.sagepub.com/content/6/1/35, doi:10.1177/875647939000600106.

[56] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2013. ISBN 3-900051-07-0. URL: http://www.R-project.org/.

[57] P. Hall and C. C. Heyde. Martingale Limit Theory and Its Application. Academic Press, 2014. URL: http://www.stat.yale.edu/~mjk56/MartingaleLimitTheoryAndItsApplication.pdf.

[58] H. Manz, P. Loutzenhiser, T. Frank, P. A. Strachan, R. Bundi, and G. Maxwell. Series of experiments for empirical validation of solar gain modeling in building energy simulation codes - experimental setup, test cell characterization, specifications and

uncertainty analysis. Building and Environment, 41(12):1784–1797, 2006. doi:10.1016/j.buildenv.2005.07.020.

[59] Stephen Sefick. Stream Metabolism-A package for calculating single station metabolism from diurnal Oxygen curves, 2015. R package version 1.1.1. by Stephen Sefick Jr. URL: http://CRAN.R-project.org/package=StreamMetabolism.

[60] Jie Zhao, Bertrand Lasternas, Khee Poh Lam, Ray Yun, and Vivian Loftness. Occupant behavior and schedule modeling for building energy simulation through office appliance power consumption data mining. Energy and Buildings, 82:341–355, 2014. URL: http://www.sciencedirect.com/science/article/pii/S0378778814005714, doi:10.1016/j.enbuild.2014.07.033.

[61] Xiaohang Feng, Da Yan, and Tianzhen Hong. Simulation of occupancy in buildings. Energy and Buildings, 87:348–359, 2015-01-01. doi:10.1016/j.enbuild.2014.11.067.

[62] Chuang Wang, Da Yan, and Yi Jiang. A novel approach for building occupancy simulation. Building Simulation, 4(2):149–167, 2011. URL: http://link.springer.com/article/10.1007/s12273-011-0044-5, doi:10.1007/s12273-011-0044-5.

View publication stats