Sustainability Assessment Framework for Infrastructure:

Application to Buildings

by

Jonghoon Kim

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved June 2016 by the
Graduate Supervisory Committee:

Samuel T. Ariaratnam, Chair
Oswald W. Chong
Wylie K. Bearup

ARIZONA STATE UNIVERSITY

August 2016

ABSTRACT

In the United States, buildings account for 20–40% of the total energy consumption based on their operation and maintenance, which consume nearly 80% of their energy during their lifecycle. In order to reduce building energy consumption and related problems (i.e. global warming, air pollution, and energy shortages), numerous building technology programs, codes, and standards have been developed such as net-zero energy buildings, Leadership in Energy and Environmental Design (LEED), and the American Society of Heating, Refrigerating, and Air-Conditioning Engineers 90.1. However, these programs, codes, and standards are typically utilized before or during the design and construction phases. Subsequently, it is difficult to track whether buildings could still reduce energy consumption post construction. This dissertation fills the gap in knowledge of analytical methods for building energy analysis studies for LEED buildings. It also focuses on the use of green space for reducing atmospheric temperature, which contributes the most to building energy consumption. The three primary objectives of this research are to: 1) find the relationship between building energy consumption, outside atmospheric temperature, and LEED Energy and Atmosphere credits (OEP); 2) examine the use of different green space layouts for reducing the atmospheric temperature of high-rise buildings; and 3) use data mining techniques (i.e. clustering, isolation, and anomaly detection) to identify data anomalies in the energy data set and evaluate LEED Energy and Atmosphere credits based on building energy patterns. The results found that buildings with lower OEP used the highest amount of energy. LEED OEP scores tended to increase the energy saving potential of buildings, thereby reducing the need for renovation and maintenance. The results also revealed that the shade and evaporation effects of green spaces around

buildings were more effective for lowering the daytime atmospheric temperature in the range of 2°C to 6.5°C. Additionally, abnormal energy consumption patterns were found in LEED buildings that used anomaly detection methodology analysis. Overall, LEED systems should be evaluated for energy performance to ensure that buildings continue to save energy after construction.

I dedicate this dissertation to

my lovely wife Kyungjoo and daughter Erin (Jiho)

&

Family

I am eternally grateful for your love, unwavering support, and continuing guidance.

Without all of you this would not have been possible.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Dr. Samuel Ariaratnam, for his excellent guidance, caring, patience, financially supporting and providing me for my research studies. You also gave me the opportunities to participate in the conferences, which afforded me the opportunity to have more connection and to become involved in the funded projects, which include Environmental Protection Agency (EPA), Southwest Gas (SWG), and Water Research Foundation (WRF). You also gave me the opportunities to teach your classes, Construction Project Management II and Trenchless Construction Methods, over the last 3.5 years.

I would like to thank Dr. Oswald Chong who exposed me to new research areas, which I had not been involved in, such as building energy performance analysis, data mining techniques, learning machine, big data applications, etc. You taught me methods of research, how to compose academic project proposals and patiently corrected my writing techniques. I would also like to thank Dr. Wylie Bearup for the guidance in my research endeavors for the past 3.5 years and helping me to develop my background in heavy and civil engineering fields. You also gave me a chance to teach your class, Building Construction Methods, Materials and Equipment. This allowed me the opportunity to gain invaluable teaching experience in academia.

I would like to thank my research group members, Hariharan Naganathan and Seungteak Lee from the School of Sustainable Engineering and the Built Environment at Arizona State University (ASU). Your advice, reviews, and recommendations have helped me complete my dissertation. I wish you all the best in your future endeavors and look forward to collaborating with you in the future.

I also would like to thank my friends, Jaemyung Lee, Dan Koo, Jinsung Cho, Jinyoung Hyun, and Sooyoung Moon, from Colorado State University, University of Louisville, Arizona State University and Korea Institute of Civil Engineering and Building Technology. You supported and encouraged me with your best wishes.

I would lastly like to thank my friends from the civil engineering industry, Curt Slagell, Jeff Callicott and Dave Goos with AZTEC Engineering and Robert Lyons, Andrew Baird, Chris Woolery and Jason Fenner with Kimly-Horn Engineering. Without your support and assistance, I would never have been exposed to good civil industry experiences, which included experiences in the roadway, drainage, utility disciplines for 6.5 years in the U.S.

Finally, I would like to thank my parents (Sungtaek and Sungmi Kim), my parents in law (Gunil and Yangmi Suh), American parents (Lee and Rita Melendez), sisters and brothers (Hyojung, Jiyeon, Hyuncheol, Manho, Youngbae, Yoon, Tom and Monica Taylor), nieces and nephews (Doyeon, Yoonsoo, Doyoon, Soontak, Dowon, Sungho and Yena).

Especially, I would like to thank to my wife, Kyungjoo and daughter, Jiho (Erin). You were always there cheering me up and stood by me through the good and/or bad times.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

1. INTRODUCTION

1.1 Research Background

1.1.1 Understating Leadership in Energy and Environmental Design Energy and Atmosphere Credits and Their Relationship with Building Energy Consumption

In the United States, buildings account for 20–40% of the total energy consumption (Pérez-Lombard et al., 2008). The operation and maintenance of buildings consume nearly 80% of this large amount of energy during their lifecycle (Cole & Kernan, 1996; Sartori & Hestnes, 2007). Approximately 70% of energy in the United States is generated by nonrenewable sources, e.g., coal and oil (U.S. Energy Information Administration, 2016). The negative effects of this type of energy consumption contribute to global warming, air pollution, and energy shortages. In order to reduce building energy consumption and related problems, sustainable development, especially sustainable construction, is being applauded by more official and unofficial organizations, including governments and environmental protection organizations.

What causes building energy consumption? According to previous studies (U.S. Green Building Council, 2015; U.S. Department of Energy, 2011), buildings accounted for 75% of all electricity generated and consumed in the United States due to heating, ventilating, and air conditioning (HVAC) systems. These systems are critical in the energy consumption of buildings in the United States. Previous studies found that another crucial factor, the atmospheric temperature, affected energy consumption and variations (Sailor, 2001).

The U.S. Green Building Council's Leadership in Energy and Environmental Design (LEED) was introduced in 1998, and it has become the dominant green building rating system globally. There is a common notion that LEED indicates energy efficiency, meaning LEED certification is often perceived as a mark of energy efficiency. To a lesser extent, few studies target the relevant energy efficiency standards that LEED adopts, e.g., the American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) 90.1. Adherence to ASHRAE 90.1 contributes to the energy efficiency of buildings. The Prescriptive and Performance Paths method is what LEED uses to model savings between baseline and design energy consumption (baseline meaning before energy efficient design is adopted and design meaning after energy efficient design is adopted). Energy efficiency requirements are further enumerated in LEED's Credit 3 for Energy and Atmosphere (EA). Though LEED-certified buildings are required to use ASHRAE 90.1 as de facto standards, jurisdictions that adopt ASHRAE 90.1 would be required to achieve similar energy efficiency levels even if they do not adopt LEED (as is the case in Arizona). The only difference is the adoption of LEED Credit 3, wherein buildings are required to achieve more than the 30% requirement.

In 2007, the U.S. Army issued Executive Order (EO) 13423 (President & Environmental, 2007) to call for the adoption of the Federal Leadership in High Performance and Sustainable Buildings. The executive order focused on reducing the lifecycle costs associated with environmental and energy attributes of federally owned building facilities by implementing the general guidelines of the Energy Policy Act of 2005. Its policies include improving energy efficiency and reducing greenhouse gas emissions. As a result of EO13423, Naval Facilities Engineering Command amended the policy in 2008 to require LEED Silver certification of all new military construction and major renovation projects in the U.S. Navy and Marine Corps building inventories.

Furthermore, the U.S. General Services Administration has upgraded its requirements to LEED Gold certification level on all new federal government buildings for a more sustainable future (Beatty, 2010).

In 2010, the previous study compared the energy consumption of U.S. Navy LEED-certified buildings and a commercial counterpart against EO13423's mandate to reduce building energy consumption (Menassa, Mangasarian, Asmar, Asce, & Kirar, 2012). Additionally, the study compared the LEED-certified buildings to the national average from the 2003 Commercial Building Energy Consumption Survey. The results of this research indicated that LEED certification alone could not guarantee the 30% savings for electricity called for by EO13423. Furthermore, the data showed that energy savings were not closely related to the number of points received in the Energy and Atmosphere (EA) category of the LEED certification process.

1.1.2 The Use of Green Space in Reducing Atmosphere Temperature

Urbanization and rapid development greatly increase the consumption of energy that emits greenhouse gasses. These gasses affect global climate and temperature (Kwok et al., 2016). Increasing global temperatures will result in a temperature increase in urban areas and abnormal climates, which lead to adverse effects on global climate (Santamouris, 2014). Research related to buildings and the urban heat island with external features—which include albedo, vegetation, and perforation rate—was affected by weather conditions, especially atmospheric temperature. The urban heat island effect exists because of the greater heat retention of buildings and manmade surfaces such as concrete and asphalt than that of vegetation (i.e., green spaces).

In 2013, previous research studies found that green space was an excellent way to decrease high atmospheric temperature in urban areas. It also generated oxygen to replace carbon dioxide. Furthermore, this study included the importance of green space in reducing heat island effects (McPherson, 1988; Wagner et al., 2013). Previous studies planned various landscapes around the buildings to control solar radiation and air infiltration as well as to provide shade and wind protection and thereby reduce energy consumption in the buildings (Sawka et al., 2013).

1.1.3 Applied Methodologies from past Research Studies

Previous research studies used various analytical methods to analyze the building energy data. Regression analysis and various nonlinear analysis methods were commonly applied to the study of building energy. Some of these examples include Cheng's (2015) nonlinear analysis and the Asadi et al. (2014) study on linear or multivariate regression analyses, which were applied to a large amount of energy consumption data. The Fourier et al. (2013) study's multiple linear regression was used to analyze the building's physical characteristics and energy performance.

Additionally, in other past research studies, nonlinear analysis methods were applied to external factors (e.g., atmospheric temperature and green space) and building energy consumption (Bessec, 2008; Henley & Peirson, 1997; Moral-Carcedp & Vicens-Otero, 2005; Santamouris et al., 2014). Several previous studies also addressed the building energy models and examined imaginary buildings as their case studies (Huang et al., 2009; Kalvelage et al., 2014; Kolokotronni et al., 2012; Wan et al., 2012; Wang, 2014; Yu et al., 2012).

Donovan et al. (2009) examined field surveys and electric bills to find out how to reduce summer energy electricity costs by approximately 5.2% using green space. Yang et al. (2010) found from field surveys that layout, density, and ratio of green coverage around buildings influenced urban heat effects in the residential buildings of Shanghai. They stated that shade and solar heat modified the urban heat island more than any other factors. Additionally, paved road space with shade was cooler than public lawn space without shade. Oliveira et al. (2011) asserted that even small urban green spaces could alleviate the urban heat island effect. They performed a case study of a 0.24 hectare neighborhood garden in Lisbon and found that the garden's highest temperature was 6.9 °C cooler than those of surrounding locations.

Overall previous research studies, which included building energy consumption performance, found that the relationship between environmental factors (e.g., atmosphere and green space) building energy consumption, and data analysis methods required the use of data mining techniques (e.g., K-means clustering, isolation, and isolation forest) to find better empirical results and overcome the limitation of data analysis methods. Therefore, this research study attempts to significantly address these gaps in knowledge by providing comprehensive case studies.

1.2 Research Objectives and Methods

The primary objectives of this dissertation are to: (a) find the relationship between building energy consumption, outside atmospheric temperature, and LEED EA credits, (b) examine the use of different green space layouts to reduce the atmospheric temperatures of high-rise buildings, and (c) use data mining techniques, including clustering and isolation, to identify data anomalies in the energy data set and evaluate LEED EA credits after construction based on building energy patterns.

The main focus of this dissertation is finding and analyzing building energy consumption data using various analytical methods such as data mining, which includes clustering, isolation, and anomaly detection. This dissertation is composed of three main phases, which are included in chapters 2, 3 and 4 and shown in Figure 1. The individual chapters will be expanded into journal format.

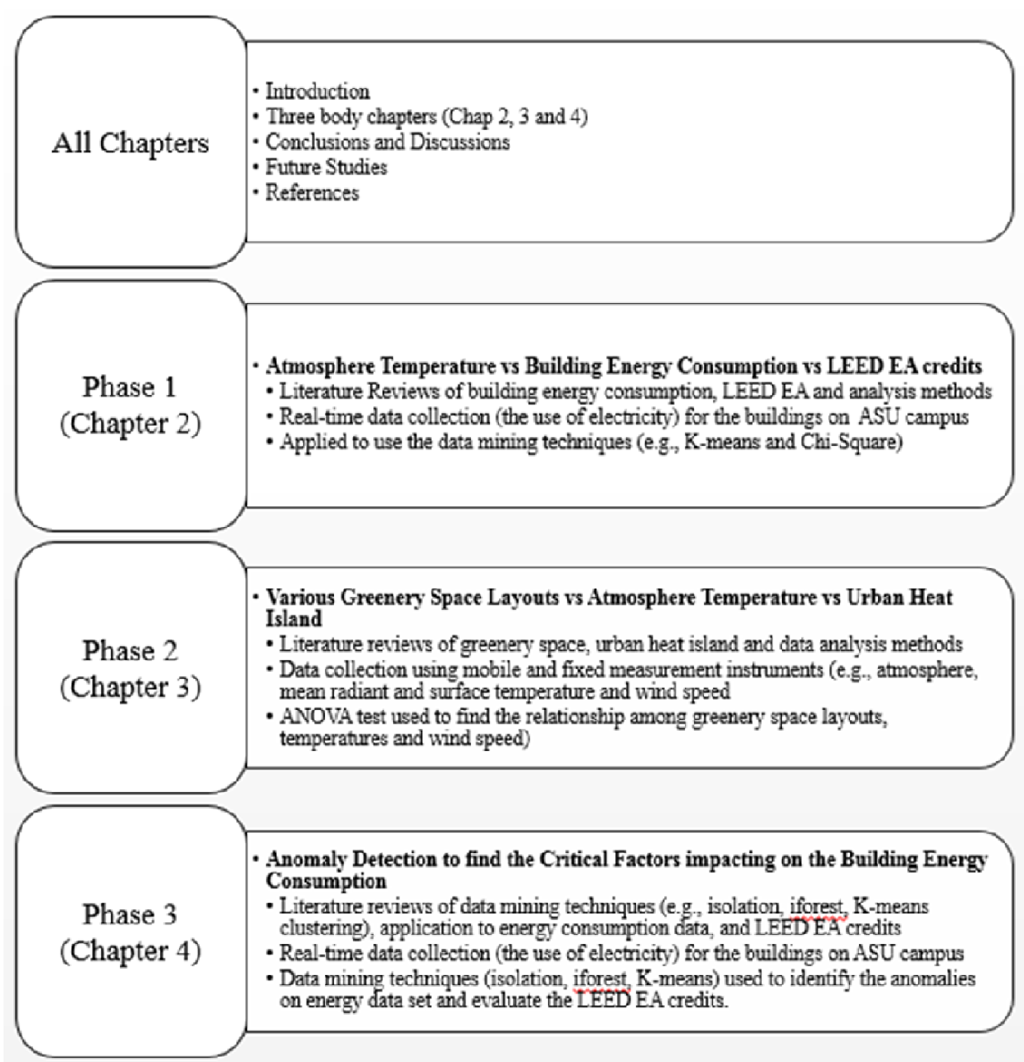


Figure 1. Research Study Flow Chart (Chapter 2, 3 and 4)

1.2.1 Chapter 2: Building Energy Consumption vs LEED EA Credits

Chapter 2 provides a detailed description of the effects of external factors on building energy efficiency designs and LEED EA credits. This study compared the relationship between the external factors and building energy consumption of LEED certified buildings at Arizona State University by establishing the relationships between the outside atmospheric temperature and the energy consumed in the building using real-time data. The study highlighted the fact that energy consumption data alone does not yield useful results, and a further pre-data process is needed to establish the cause and effect relationship. The findings of this study are being prepared in journal paper format for the *Journal of Engineering, Design and Technology.*

1.2.2 Chapter 3: The Use of Greenery Space Layouts in Reducing Air Temperature

Chapter 3 discusses how different green space layouts affect atmospheric air temperature around buildings, which contributes to the urban heat island effect. The aims are to develop an understanding of whether greenery arrangements affect the urban heat island. Field measurements were taken for three different layouts, namely greenery surrounding, in the center of, or distributed over a complex or building. The study shows that the layout of greenery can have a significant effect on urban heat islands given the same land area. This study is being prepared for the *Journal of Architectural Engineering.*

1.2.3 Chapter 4: Detect Anomalies Using Isolation Technique

Lastly, Chapter 4 presents the building energy consumption patterns of LEED buildings after the construction phase and finds energy consumption data abnormalities using clustering and isolation techniques. The purpose of this study is to examine how the energy consumption patterns affect LEED EA credits after construction. The results showed that LEED buildings with higher OEP points had stable energy patterns and consumed the least amount of energy. The findings of this study are being prepared for the *Journal of Energy Engineering.*

1.3 Dissertation Format

This dissertation is composed of three journal papers. Each of the three subsequent chapters represents an independent article. Therefore, each chapter will have its own abstract, introduction, objectives, methodology, discussion of results, and conclusions. The findings of chapters 2, 3, and 4 are being prepared in journal format for the *ASCE*.

Chapter 1 presents the basis of the current body of knowledge related to this research study, including the research background, problem statement, methodology, objectives, and scopes and format. Chapter 2 provides an understanding of the effects of external factors on building energy efficiency designs and LEED EA credits. Chapter 3 presents how different green space layouts affect the surrounding atmospheric air temperatures of high-rise buildings. Finally, Chapter 4 provides an understanding of the building energy consumption patterns of the post-construction phase and finds energy consumption data abnormalities by using clustering and isolation techniques during the data process. Chapter 5 includes the research summary and conclusions based on the case studies of Chapters 2 through 4 as well as the research limitations and contributions of the dissertation and future research studies. References and appendices are attached at the end of this dissertation.

2. UNDERSTANDING THE EFFECTS OF ENVIRONMENTAL FACTORS ON BUILDING ENERGY EFFICIENCY DESIGNS AND CREDITS

2.1 Abstract

Energy usage of buildings accounts for a large part of total energy usage in the U.S. The American Society of Heating, Refrigeration and Air-conditioning Engineers (AHSRAE) 90.1 standard has been used extensively to reduce energy consumption in buildings. ASHRAE 90.1 has been adopted by many states and organizations, including green building rating systems and codes, such as the LEED Energy and Atmosphere (EA) credits. It is often assumed that compliance with the LEED EA credit would be interpreted as improved energy efficiency. This study compared the relationship between the environmental factors and building energy consumption of three LEED certified buildings at the Arizona State University, by establishing the relationships of the outside atmospheric temperature and the energy consumed in the building using real-time data generated from different sources. The study shows that there is no linear dependency between the selected independent factors and energy use of the studied buildings. The study highlighted that energy consumption data alone does not yield useful results and further calibration of the dataset is needed by establishing the causation and effect relationships.

2.2 Introduction and Scope of Research

Buildings accounts for 20-40% of the total energy consumed (Pérez-Lombard et al., 2008) while building operations and maintenance consumed approximately 80% of such energy throughout their lifecycle (Cole & Kernan, 1996; Sartori & Hestnes, 2007) in the United States. As the energy is generated from non-renewable sources (e.g., fossil fuel and natural gas) mostly, it increases the risk of global warming and air pollution. Reducing energy consumption throughout a building's lifecycle will reduce the associated pollution.

The International Energy Conservation Code (IECC), and ANSI/ASHRAE/IESNA Standard 90.1 Energy Standard for Building are the two predominant building energy codes in the United States. These codes contain building energy design methods that would improve building energy efficiency by a certain proportion over the baseline building energy consumption.

The United States Green Building Council (USGBC) Leadership in Energy and Environmental Design (LEED) was introduced in 1998, and it has become the dominant green building rating system globally. There is a common notion that LEED equates to energy-efficiency. LEED-certification would often be perceived as achieving energy efficiency. To a lesser extent, few research targets the relevant energy efficiency standards that LEED adopts, i.e., ASHRAE 90.1. ASHRAE 90.1 is a standard what contributes to the energy efficiency of buildings. The "Prescriptive and Performance Paths" method is the method that LEED adopts to model savings between "baseline" and "design" energy consumption (baseline = before energy efficient design is adopted, and design = after energy efficient design is adopted). LEED's credit 3 for Energy and Atmosphere (EA) further enumerates energy efficiency requirements. While LEED-certified buildings are required to use ASHRAE 90.1 as de-facto standards, jurisdictions that adopt ASHRAE 90.1 would be required to achieve similar energy efficiency levels even if they do not adopt LEED (as in the case of Arizona). The only difference is the adoption of LEED Credit 3, where buildings are required to achieve beyond the 30% requirement.

A recent study on LEED buildings' energy consumption found that LEED-certified buildings consumed approximately 25-30% less energy than non-LEED certified buildings (Scofield et al., 2008). However, their research did not mention if ASHRAE 90.1 was a mandatory code in the studied region. Alternatively, an energy efficiency study by the American Physical Society (APS) noted that LEED buildings use more energy per square foot than the average for all existing commercial buildings (Richter et al., 2008). This study exhibited results that deviated from prior studies on LEED-certified buildings. As such, this paper attempts to establish a relationship, using real-time energy consumption data, between additional points from LEED credit 3, and actual energy consumption of three buildings on the ASU campus.

2.3 Research Objectives

This research attempts to understand the relationship between LEED credit 3 and energy efficiency of buildings, particularly when energy consumption is affected by outdoor air temperature and real-time data (by the hour). Establishing this relationship using real-time data is the unique approach adopted by this research. Real-time energy consumption data exhibits features that traditional energy consumption data does not include: 1) the ability to incorporate variability during different periods, and thus eliminate external elements on the analysis, and 2) changes to design and operational variables at different periods.

The three objectives of this research study include: 1) understanding the relationships between energy consumption and outdoor atmospheric temperature; 2) understanding the relationships between energy consumption and LEED energy credits; and 3) understanding the use of LEED energy credits as a tool to improve energy efficiency of buildings.

2.4 Literature Findings

Literature review included many studies on the relationships between energy use and factors driving energy use, especially on the use of statistical analyses for building energy consumption.

2.4.1 Building Energy Performance and Consumption

Diamond et al. (2006) studied the differences between modeled and actual energy performance of 21 LEED-certified buildings between December 2001 and August 2005 using 2003-2005 utility bills, by analyzing the energy consumed by these buildings. They found that 18 buildings were 26% more energy efficient than the baseline design case. However, the research also highlighted that 21 buildings were not conclusive, and there are other limitations like as-built versus as-designed discrepancies, and changing conditions such as occupancy, weather, use, etc.

Newsham et al.'s (2009) study on energy use data (from New Buildings Institute, and 100 LEED-certified commercial and institutional buildings) found that the LEED-certified buildings consumed between 18 and 39% less energy per floor area than non-LEED certified buildings. However, 28-35% of LEED buildings consumed more energy than comparable non-LEED certified buildings. The study also found an insignificant correlation between measured energy performance of LEED buildings, with the level of LEED certification and the number of LEED energy credits earned.

Juan et al. (2010) studied and suggested that LEED buildings are "healthier" and improve work productivity, and are potentially more energy efficient than non-LEED certified buildings. These buildings could potentially save 25-50% energy and are thus, more environmental friendly, as shown in previous studies.

Salmon et al.'s (2008) study on 121 LEED-certified buildings, using three different baselines (Energy Use Intensity – EUI, average Energy Star rating, and energy consumption model), found that the energy performance of the LEED-certified buildings vary significantly. The study found that the surveyed buildings' median EUI was 24% below the national average (Commercial Building Energy Consumption Survey - CBECS). Statistical analysis showed that the median energy performance of the LEED gold- and platinum-certified buildings were aligned with the energy consumption goals of Architecture 2030 (Wedding et al., 2008). The study also showed that adopting Energy Star ratings improved the energy performance of buildings regardless of their LEED certification status. Alternatively, the study also showed that the actual energy consumed at the occupancy stage of nearly half of the buildings was significantly higher than the computed design energy consumption. Similarly, Turner and Frankel's (2008) study showed that the average site energy intensity of the surveyed LEED buildings were 25% to 30% more energy efficient than the national average.

2.4.2 Building Envelopment

Cidell's (2009) study found that most LEED-certified buildings clustered in a few geographical locations. The research affirmed a previous study by Sinha (2008) that income level, educational level, and size of the service-sector correlated with the number of green buildings built in the region. This suggests that the green buildings might be more appealing to the more affluent regions.

Griffith et al.'s (2008) study showed that 62% of the sampled buildings and 47% of the sampled floor space could achieve net-zero energy use using current technologies and design practices. The energy-saving focused on building envelopes, lighting controls, plug and process load reduction, and energy efficient HVAC system, and their study showed a decrease in lifecycle energy consumption of 43% or more using ASHRAE 90.1-2004. Kneifel (2010) stated that ASHRAE 90.1 focuses on cost-effective solutions pertaining to the Lower Energy Case (LEC), and the study confirmed the statement in over half of the 192 sampled building. These results also showed how quickly energy efficiency measures could be applied in the context of design.

2.4.3 Analysis Methods

Regression analysis and various non-linear analysis methods were commonly applied to study building energy. Some of these examples include Cheng's (2015) non-linear analysis, and Asadi et al.'s (2014) linear or multivariate regression analyses, where they were applied on a large amount of energy consumption data. Focuier et al.'s (2013) multiple linear regression was used to analyze the building's physical characteristics and energy performance. Clustering analyses were used to analyze building energy performance and applied to both design and operational stages using design documents and weather data from the weather stations (Heidarinejad et a.l, 2014; Hsu, 2015; Petcharat, et al., 2012). Cluster analysis offers the potential to analyze and test energy consumption of buildings by energy use intensity categorization.

The above reviews showed that previous research comparing energy efficiency and consumption differences between LEED-certified and non-LEED-certified buildings did not focus on the application of relevant codes and standards, i.e. ASHRAE 90.1. LEED credits on "Energy and Atmosphere", more specifically Requirement 1 and Credit 3, are the only LEED requirements that target energy use and efficiency. This research aims at understanding if complying with ASHRAE 90.1 alone versus achieving more LEED credits in credit 3 would result in improving energy efficiency (rather than modeling LEED certification versus non-LEED certification.

2.5 Research Methodology

Real-time hourly energy usage data for three sample buildings on the ASU campus were selected for the study. Data from 2012 to 2014 on electricity consumption were collected. The energy consumption data collected are for the total energy consumed by the buildings by the hour. Only electricity data was collected as other sources of energy (like natural gas) were used only infrequently. The energy usage was adjusted by dividing the total electricity use by the corresponding building's floor area. The electricity usage data were consolidated into daily values to synchronize it with the average value of the outside atmospheric temperature. The resulting unit of the electricity usage is Kilowatt per square feet-day (Kw/Sqft-day). The real-time energy data was calibrated to minimize the effects of external elements (e.g. building maintenance schedule) using k-means clustering algorithm. K-means clustering is a method of vector quantization, originally derived from signal processing, that is popular for cluster analysis in data mining. *K*-means clustering aims to partition *n* observations into *k* clusters, in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. These results in a partitioning of the data space into Voronoi cells, and thus, would be used to isolate non-relevant factors.

The daily average outside atmospheric temperature was collected from the National Oceanic and Atmospheric Agency's (NOAA) climate data. The geographical location for the data is set at 33.4258N latitude and -111.922W longitude (city of Tempe), and at 33.4222N latitude and -111.8219W longitude (city of Mesa); both locations are at the Weatherup Center on the Arizona State University (ASU) campus. The distances between the weather station and buildings are insignificant to cause any changes in temperature or climate. As the relative humidity in the region is extremely low, the electricity use for heating and cooling is mainly used to deal with the temperature and not to extract moisture from the atmosphere. Thus, the energy consumed by the buildings could be lower than the national average.

2.6 Results: Findings and Analysis

The electricity used for the cooling and heating for each building was calibrated to address the outside air temperature.

2.6.1 Selection of LEED Buildings in Arizona State University, AZ

Three buildings managed by the Arizona State University were selected for the study. The buildings were selected as they had complete information and the ASU facilities management team had solid documentation of the buildings. The buildings included Bio Design Building B and Barrett Honors College building located in Tempe, AZ, and the Interdisciplinary Science and Technology Building 3 (ISTB 3) located in Mesa, AZ. As shown in Figure 2 and Table 1, these buildings are located in the same ASHRAE climate zone.

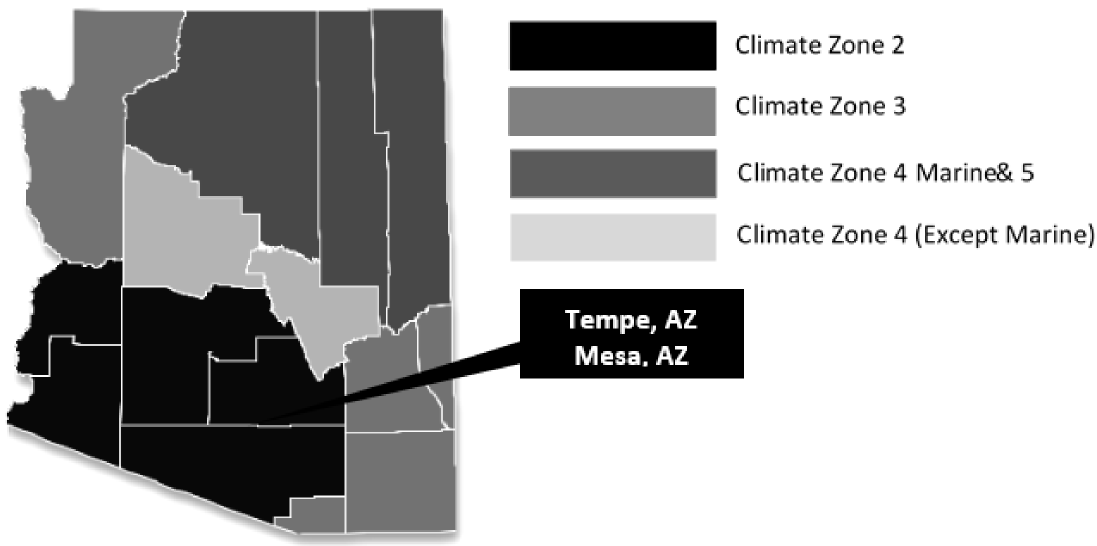


Figure 2. Climate Zone in Tempe and Mesa, Maricopa County (AZ) (Adopted from the 2014 International Energy Conservation Code)

Table 1. Climate Information Between Tempe and Mesa, AZ (Adopted from 2014 U.S. Climate Data)

Climate Information	**Tempe (9/1/2012-8/31/2014)**	**Mesa (9/1/2012-8/31/2014)**	**Observation Error**
Climate Zone	2	2	-
Sunny Days	296	296	-
Average July High (°F)	104.2	103.9	0.29%
Average January Low (°F)	33.8	34.1	0.88%
Elevation (ft)	1,192	1,273	-

The focus of this section is to understand how outdoor atmospheric temperature affects the energy consumption and performance (particularly electricity consumption) of LEED-certified buildings. The Optimize Energy Performance (OEP) factors (submitted during LEED certification process) were used to determine the LEED EA credits earned. Bio Design Building B earned 10 points, Barrett Honors College earned 4 points, and ISTB 3 earned 2 points. These points are listed with building information in Table 2.

Table 2. Green Building Facilities Information at Arizona State University, AZ

Building Name	Bio Design B	Barrett Honors College	ISTB 3
Facility Purpose	Office/Research	Class/Office	Office/Research
Facility Gross Area (SF)	179,559	89,298	47,276
Facility Net Area (SF)	80,940	42,663	21,304
LEED Certification	Platinum	Gold	Gold
Earned OEP in EA	10	4	2

LEED points for EA credit 3 represent the additional energy saving beyond the 30% specified on the basis of ASHRAE 90.1, and the computed values only represent the design values and not values at the operational phase. The energy saving is compared with the baseline design case (Turner & Frankel, 2008). The list of the selected buildings is shown in Table 2.

2.6.2 Statistical and Correlation Analyses

Figure 2 exhibits the scatter plots between the different aspects of atmospheric temperature and the electricity consumption in the three buildings. It is important to note that the ambient temperature in the Phoenix area is significantly higher during the summer, and lower during the winter, than the rest of the country (75°F during the summer, and 65°F during the winter). The heating load during winter is constantly less than 10% of the cooling load in the summer. The analyses concluded several important points:

1) The relationships between electricity consumption, atmospheric temperature, and earned OEP points differ in all the three buildings. The scatter plots for ISTB 3 and Barrett Hall seemed to suggest significant relationships between electricity consumption

and outdoor climate, but only a slight relationship was seen for Bio Design B. The impacts are significantly different, though.

2) Bio Design B seemed to be the most energy efficient building for the entire year, even though ISTB 3 seemed extremely efficient when its temperature was ideal (i.e., the heating and cooling systems were not running). These seem to suggest that changing temperatures for ISTB 3 leads to increasing electricity consumption, particularly heating and cooling loads.

3) The electricity consumption in Barrett College seemed to increase when outdoor temperature increased, and, unlike ISTB 3, the electricity consumption seemed to be less sensitive to the heating load.

4) While cooling and heating loads might have driven electricity consumption for ISTB 3 and Barrett College, the plots seem to suggest there could be other factors driving the electricity consumptions in both buildings. These factors include: a) air flow interference due to air filters, duct sizes, equipment efficiency, occupancy rates, building use, etc.; b) the buildings' design and operations (including energy system operation and design) were influenced by their use and functions, and the electricity consumption could be different – the electricity consumption for laboratory would be more consistent as occupancy rates often remain consistent throughout the day and evening, while the residential hall (Barrett College) would consume more energy in the evening and significantly lower energy during school breaks.

5) Despite the differences mentioned in point 4, the operational approaches the facilities management department adopt have more influence on the electricity

consumption. If the operators operate the buildings without regard to the occupancy, electricity consumption will become increasingly similar.

The bar charts in Figure 3 exhibit the average, minimum, and maximum electricity usages and their standard deviations. The bar charts clearly distinguish the electricity consumption patterns of the three buildings. The bar charts show that the Bio Design building is the most efficient among all three buildings, as its average electricity consumption is the lowest. ISTB 3 seems to be most efficient when the outdoor climate gets closer to ambient temperature. Bio Design building had the most consistent energy consumption pattern and it seems like external air temperature had the least effect on electricity use. The standard deviations of all three sets of data also showed that air temperature had the least impact on the Bio Design building, but the greatest on Barrett College. Further statistical analyses are presented below to provide more detail.

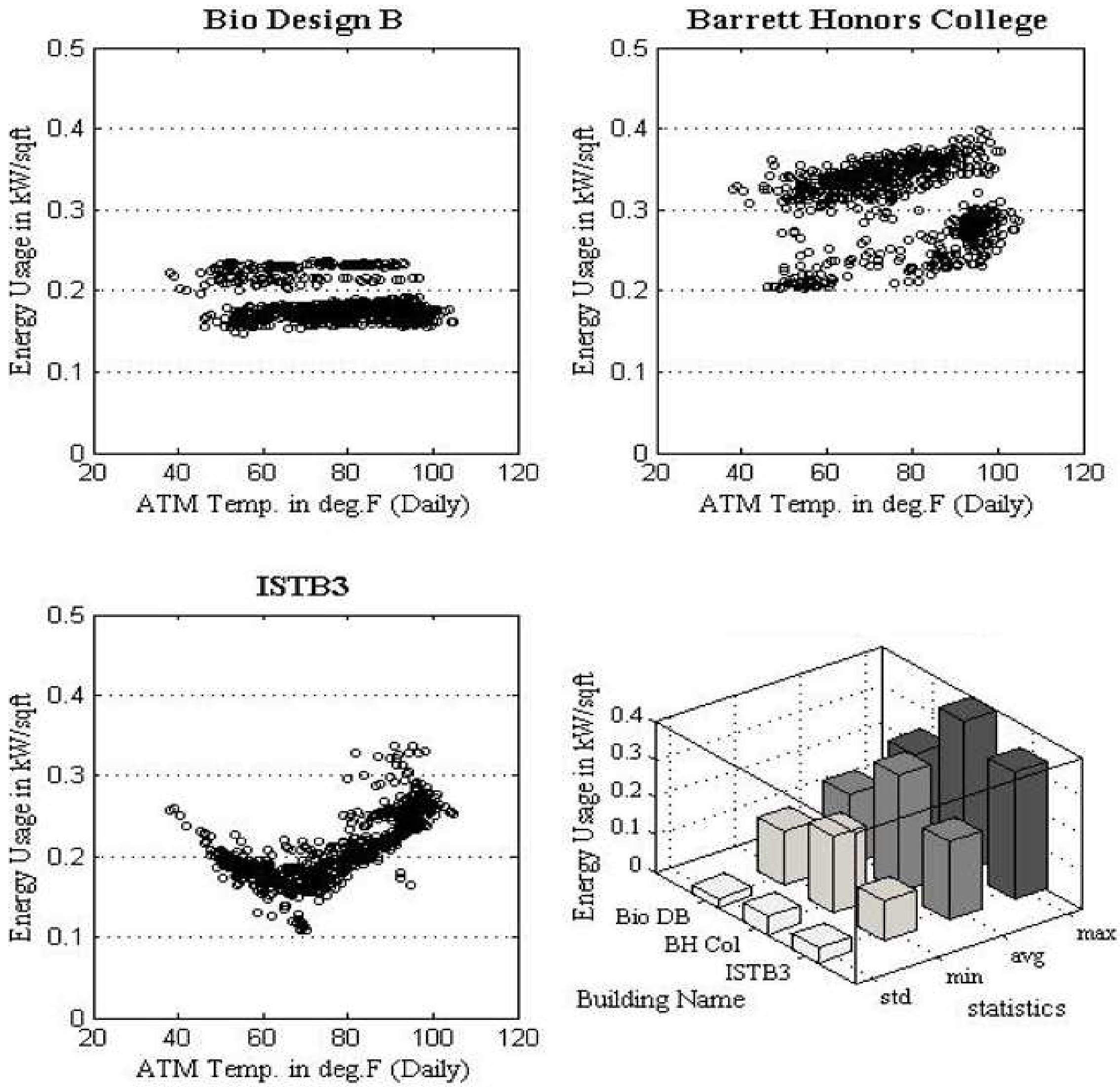


Figure 3. Scatter Plots of Energy Usages Versus Atmospheric Temperature in Daily Scale

2.6.3 Data Pre-Processing: Calibrating Electricity Consumption

Analyses in the previous section exhibit both significant and not so significant relationships between electricity consumption and outdoor air temperature. Further analyses are required to better understand the reasons behind the differences and plots.

Building Envelop: The "V-shape" scatter plot of the ISTB 3 building in Figure 2 suggested that the outdoor air temperature could heavily influence ISTB 3's electricity consumption. Thermal heat transfer through the envelope increases the demand to heat or cool indoor air, as lower or higher outdoor air temperature increases energy to cool or heat up outdoor air to adjust the ambient air temperature. A poorly sealed building would also cause an increase in energy use due to the gaps between outdoor and indoor air temperature. Thus, the relationship between energy use and outdoor air temperature could mean that one or both of the relationships are happening. Energy efficient building envelopes would decrease the significance of the relationships, since reducing heat transfer from the envelope would reduce the need to heat or cool indoor air. Good quality construction of building envelopes also reduces the significance of the relationship.

The significance of such a relationship on ISTB 3, and to a lesser extent on the Barrett College, does not mean that the Bio Design building has a better constructed and insulated building envelope. However, the low electricity consumption (average, minimum and maximum) of the Bio Design building suggests that the Bio Design building is an extremely energy efficient building, and this would explain the quality of its envelope. It is safe to assume that the envelope is well constructed. It is also safe to assume that the thermal heat coefficients of all three buildings are equivalent and were well constructed; thus, the differences in the energy performance would not be influenced by the building envelope.

What explains the differences? The explanation on the quality of building envelopes and their construction shows that other factors influence electricity use from changing outdoor air temperature. Preliminary observations show that minimum electricity use plots are scattered across the graph, and they seem to suggest that certain factors could be driving the relationship. A preliminary investigation suggested that the types of building occupants, hours of operations, and the types of equipment installed in the building could impact electricity consumption.

Treatment and Calibration of Data: The efficiency of equipment would have the most significant impact on the minimum electricity consumption of each building. Heating and cooling equipment would normally stop during the period when minimum energy use occurred. During these periods, other equipment may still be running. As a result, the data were calibrated accordingly, using various statistical methods such as K-means.

Therefore, the daily energy usages were adjusted by eliminating the minimum electricity use from the overall dataset. This filtering procedure would strengthen the effects of heating and cooling loads. Furthermore, preliminary analyses suggest that there are "unchecked" clusters among the data. The unchecked clusters are the result of mixing non-significant data with significant data, i.e., data influenced by outdoor air temperature versus those not influenced by outdoor air temperature. Bio Design B building and Barrett Honors College data could exhibit the presence of such clusters, and thus require further treatment to the data.

Data Clustering: Cluster analysis is the procedure that groups a set of objects in such a way that objects in the same group are more similar to each other than to those in the other groups (Anderberg, 2014). It is commonly used in exploratory data mining and statistical data analysis to eliminate unwanted influences or factors from a set of data. Cluster analysis is not a specific algorithm, and has several techniques. K-means is the technique used in this paper.

Clustering to eliminate data errors: The clustering of data suggests the need for further investigation of relationships. Unexpected clusters were found after the clustering exercise. The relationship between electricity use and outdoor air temperature is distorted if other non-related clusters are mixed with the related clusters. Two different groups clustered the data Bio Design B building - ‘Before calibration’ and ‘After calibration’. The clustering exercise also eliminated an error from the data. The error occurred from February 2, 2012 to 2013, when energy consumption data was inaccurately registered. The clustering exercise adjusted the inaccurately registered data. The trends shown on these two plots was similar, even if the maximum and minimum electricity use had an approximately 40% difference (from 0.1969 kW/SF to 0.1466 kW/SF). The scatter plots were calibrated and replots are exhibited in Figure 4.

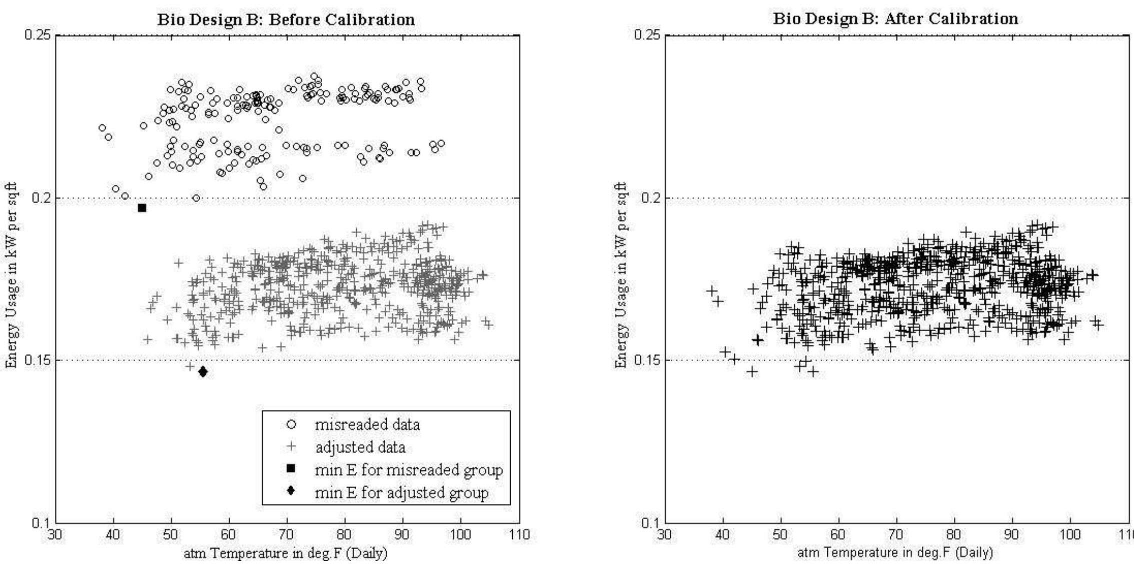


Figure 4. Scatter Plots of Energy Usages and Atmospheric Temperature: (“Before Calibration” and “After Calibration” of Building Temperature)

The clustering exercise was also conducted on the Barrett Honors College building’s electricity consumption data, and the data was separated into two groups. Figure 4 shows the scatter plots of electricity consumption and outdoor atmospheric temperature of the

Barrett Honors College. The upper-left figure is a scatter plot of energy usage of periods of vacations and semesters at ASU when electricity consumption is assumed to be lower.

Clustering of data that are influenced by the occupancy rate (Barrett Honors College): The scatters of the two clusters explained the effects on electricity consumption that is influenced by occupancy. High occupancy (during the semester) is correlated with high electricity use. The academic schedule of ASU is used as the factor for the clustering exercise. Mixing clusters of low occupancy with high occupancy affect the electricity use's relationships with outdoor air temperature and energy use intensity (per square foot). The clustering exercise thus separates the data into two clusters - one for high occupancy (during the semester), and one for low occupancy (during breaks). The analysis shows that occupancy is a factor that exhibits a significant impact on the cooling and heating loads. A fully occupied building exhibits a greater dependency on heating and cooling loads, and larger gaps between outdoor atmospheric temperature and indoor ambient temperature would increase heating and cooling loads.

The relationships between building occupancy and heating/cooling energy consumption is critical in the analysis of energy consumption in buildings. The separation shows the intimate relationship between occupancy and heating/cooling load that prior research has yet to establish. The upper left figure (titled Barrett Hall College occupancy) in Figure 5 shows the "vacation" group from the "semester" group (semester group represents data during the semester, and the vacation group represents data during breaks). The data points for the vacation group are denser as it gets closer to the ambient temperature, while the data points for the semester group are clustered away from the ambient temperature and clustered around the periods when the gaps of indoor and outdoor temperature were larger. This somehow suggests that electricity consumption is affected by both occupancy and outdoor air temperature, and there is a significant relationship between outdoor air temperature and occupancy. As a result, both clusters could be merged by subtraction of minimum energy usage points from both clusters.

Clustering of data that are influenced by occupancy rate (Bio Design building): However, this is not the case for the Bio Design B building. The data could not be linearly divided, and was found to be co-mingled. The two clusters were divided by the distance between the centroids within their clusters. The choice of initial centroids must consider the shape of the cluster. The two clusters are not radially shaped, but spread horizontally and slightly downward to the left. Before applying this algorithm, all of the dimensions were normalized in order to calculate the spatial distance of the vectors.

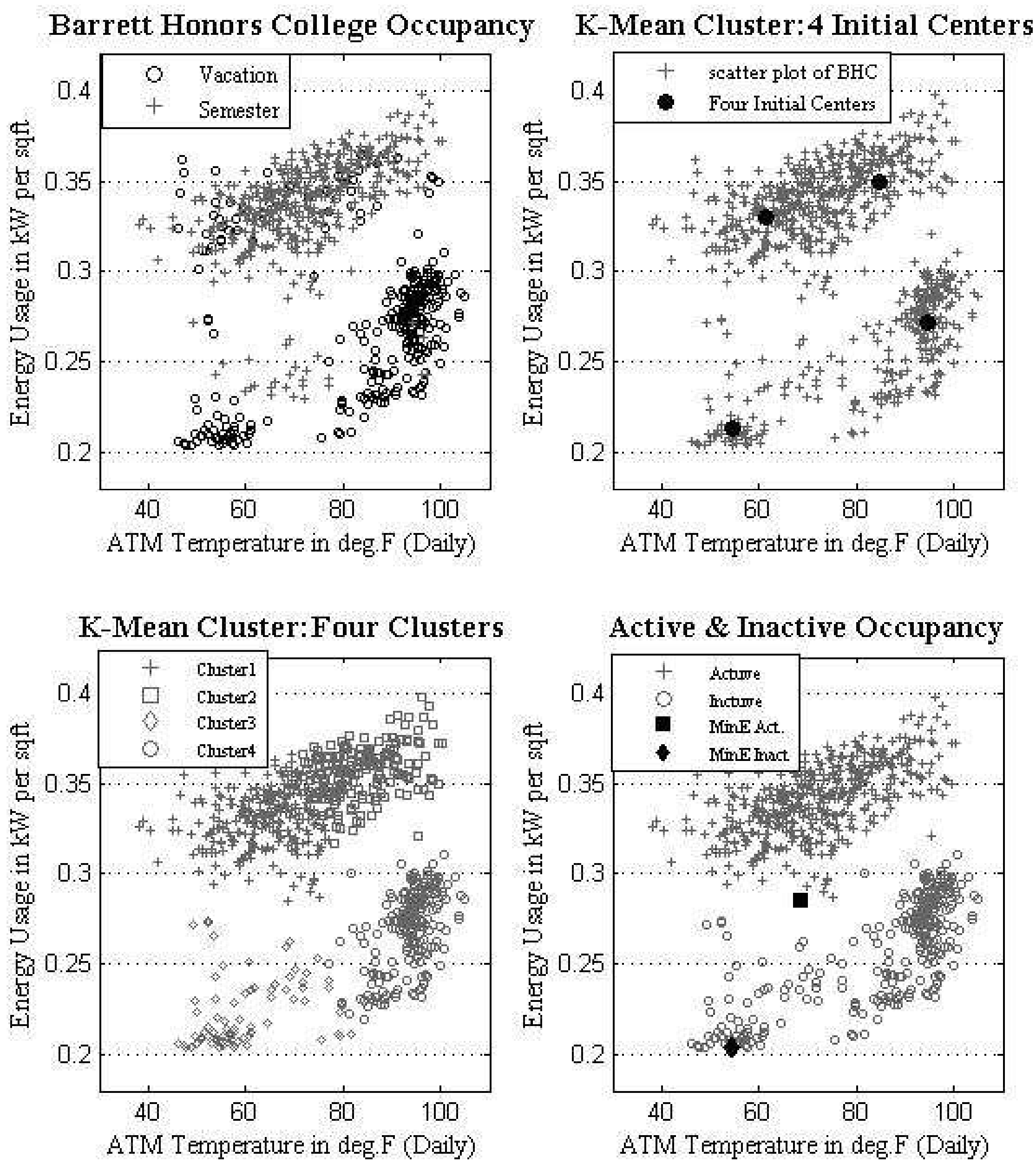


Figure 5. Scatter Plots of Energy Usages and Atmospheric Temperature:

(Vacation (Inactive Occupancy) and Semester (Active Vacancy))

The figure on the upper-right side is an initialization of the four cluster centroids. The two centroids (or points) in semester group belong to the semester groups and the lower two clusters belong to the vacation groups. The initial centroids were deployed along the centerlines of the two groups. Then, all points were assigned to the closest centroid. The recomputed centroids were used for the next iteration, and the iteration was stopped when the threshold of the distance between the previous centroid and newly calculated centroid became less than 0.001. The figure on the bottom left-hand side summarizes the final results of the four clusters. Finally, the two groups were created by adding two of the horizontal clusters to create active and inactive occupancy clusters, and similarly, the two separate minimum energy usage values of 0.2851 kW/SF and 0.2037 kW/SF for the active occupancy and the inactive occupancy were also chosen.

2.6.4 Calibrated Heating and Cooling versus Atmospheric Temperature

Building characteristics affecting relationships: The raw energy usage data has to be calibrated to address the corresponding minimum energy usages for a different scenario. The minimum energy usages were chosen for each cluster. Figure 6 shows the calibrated relationship of building energy usage compared to the outdoor atmospheric temperature. There are no visible separations for the scatter plots of all three buildings. In the bar chart in Figure 6, the ISTB3 building (recall that its OEP score is the lowest among the three studied buildings) had the most significant relationship between electricity consumption and outdoor atmospheric temperature.

On the other hand, the Bio Design B building (recall that it earned the most OEP score) consumed only a quarter of electricity compared to ISTB 3. The Bio Design B building exhibited a more stable electricity consumption pattern throughout the study period. Barrett Honors College displayed a more moderate relationship. This somehow suggests that there is an inverse relationship between electricity consumption and OEP scores (recall LEED credit 3) – Higher OEP score results in lower energy footprint and less significant relationships between energy consumption and outside air temperature.

This suggests that better building energy efficiency design could reduce the impact of external effects on energy consumption. This is, however, an inconclusive statement, but it is worth further investigation.

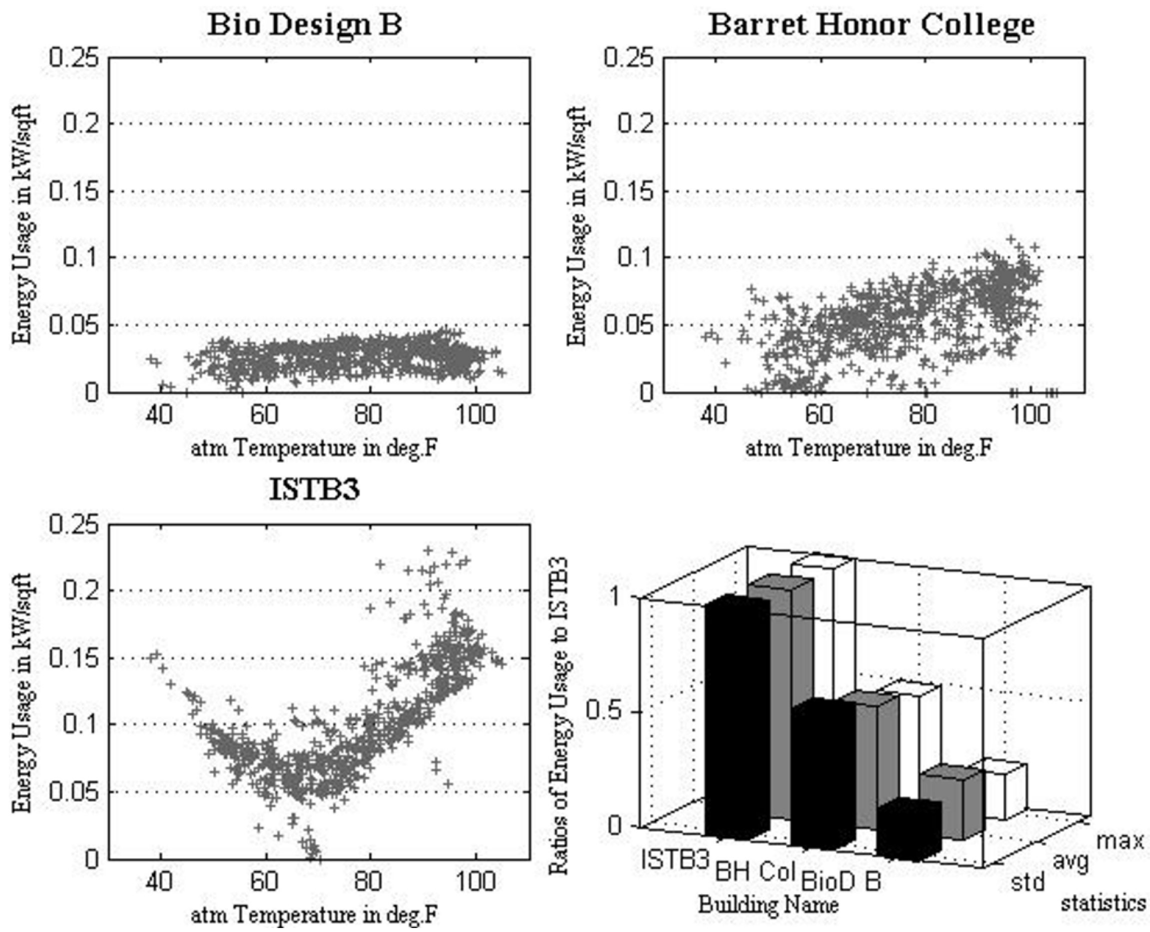


Figure 6. Calibrated Scatter Plots of Energy Usages by Corresponding Minimum Energy Usages Versus Atmospheric Temperature in Daily Scale

2.6.5 Chi-Square Analysis

Chi-square is used to further analyze the data. As discussed before, building performance is driven by a building's characteristics and its local climate condition. This analysis focuses on comparing buildings with similar conditions between the internal and external factors. The chi-square method is used to generate an understanding of the differences between LEED scores for credit 3, and its effects on building energy use. The chi-square technique is used as it could investigate the distributions of categorical variables (calibrated energy usage) and how they differ from the expected frequency (that is derived from three sampled buildings). The significance and differences between the expected frequencies and observed sample's frequencies in the categories were examined.

The difference between expected and observed frequencies of the categories was thoroughly examined. **The null hypothesis is described as "no significant difference in the energy efficiency between the three buildings."** The similarity between the frequencies of the categorization shown in Table 3 reveals that the energy efficiency of the studied buildings is very different (none of the buildings are similar to each other). The chi-square test aims to understand the implementation of the buildings' energy usage efficiency for different buildings; the chi-square statistic method was used to understand practical energy usages among three sampled buildings. A two-way table, which contains 728 daily energy use for two years, was used to categorize the magnitude of the energy usage by a 0.05kW/SF increment, and the results are shown in Table 3. Two-way Table of Energy Efficiency of LEED Buildings in ASU, AZ.

The chi-square analysis shows that daily energy use (through observation) of the Bio Design Building B can be categorized as the most energy efficient building among all three sampled buildings ((0-0.05 kW/square foot). On the contrary, a sizable proportion of the observations of the ISTB 3 building shows that it performs worse than the Bio Design B building (see Table 3). Barrett Honors College falls between both buildings.

Table 3. Two-way Table of Energy Efficiency of LEED Buildings in ASU, AZ

			Building Name			
Energy Usage in kW/SF			Bio Design B	Barrett Honors College	ISTB3	Total
More Efficient	Extreamly Better	(0.00-0.05)	728	304	41	1073
	Much Better	(0.05-0.10)	0	416	384	800
	Moderately Better	(0.10-0.15)	0	8	205	213
	Slightly Better	(0.15-0.20)	0	0	86	86
Less Efficient	About the Same	(0.20-0.25)	0	0	12	12
Total			728	728	728	2184

Figure 7 reveals that the buildings' expected counts for each category and the observed counts (of the data points among three LEED buildings) grouped in different external air temperature. The chi-square analysis shows that it is very common, from the general statistical point of view, to overcome the biases of building categorization due to the buildings' energy consumption pattern. In the data used for this research study, 2184 different observations of the energy usage for three buildings were recorded. The expected numbers of energy usage in each building category were significantly different from the occurrence of the energy efficiency categories within the research study. Therefore, the null hypothesis was rejected, implying that there was a significant difference in the energy efficiency between the three buildings and these energy efficiencies were not distributed proportionately to the occurrence of energy efficiency categories.

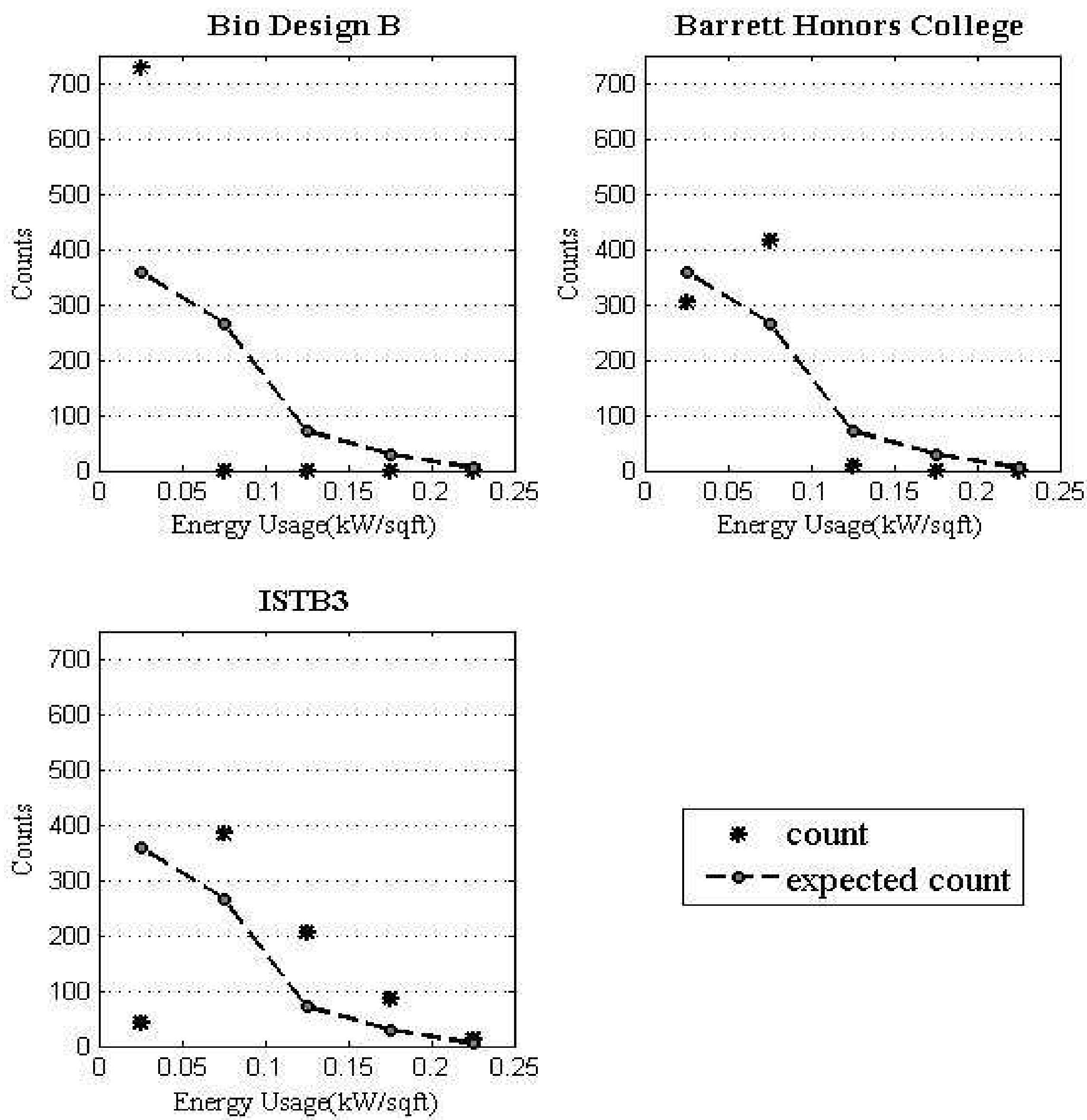


Figure 7. Actual Counts and Expected Counts in Chi-Square Test

2.7 Conclusions and Discussions

This study investigated the effects that both endogenous variables (e.g., LEED OEP) and exogenous variables (e.g., atmospheric temperature) have on the energy usage comparison of green buildings such as Bio Design B, Barrett Honors College and ISTB 3, on the ASU campus, AZ. The results of the study are concluded and summarized below:

1. Data Calibration and Adjustment: As shown in the results of this study, raw energy data was required to reconcile and harmonize to use for data analysis. In addition to enhance this raw data, calibration and adjustment of the data was needed to analyze accurately. The daily energy usages were adjusted by subtracting the minimum energy use. This approach allowed the research to filter electric uses from other equipment in the building except for heating or cooling.

2. Data Clustering (K-means): K-means technique was utilized to find and eliminate data errors. For the Bio Design B building-raw data, an energy usage data error occurred and recorded incorrectly for 6 months during the research study, from February 2012 to February 2014. Based on this raw data, there was an approximately 40% energy usage difference (from 0.1969 kW/SF to 0.1466 kW/SF) between 'Before calibration' and 'After calibration'.

3. Chi-Square Method: In order to verify if the results were reasonable, the chi-square technique was used to prove the findings. Observation counts on daily energy usage in the Bio Design B Building earned the highest OEP, scored 10, falling into the most efficient category. In contrast, the observation counts of ISTB 3 earned the lowest OEP, scored 2, therefore displaying less energy efficiency. As a result, the null hypothesis, which was defined as "no significant difference of the energy efficiency between the three buildings", was rejected.

4. Relationship between electricity consumption, atmospheric temperature and OEP points:

The three buildings earned different OEP points. Electricity consumption for ISTB 3 and Barrett Honors College, which earned the lowest and middle points, were significantly affected by atmospheric temperature; however, the Bio Design B building, which earned the highest OEP points, was less affected by the atmospheric temperature compared to the other two buildings. Also, the average electricity consumption for the Bio Design B building was the lowest. As a result, the building with the lowest OEP points (ISTB 3) used the highest amount of energy compared to the other two buildings with higher OEP. The study showed an inverse relationship between energy use and OEP points.

The LEED OEP scores tend to increase the energy saving potential of the building, leading to fewer needs for renovation and maintenance; however, it cannot be verified for a one-sided approach such as energy efficiency, as the analyses suggest. However, this study highlighted that calibrating energy data is a better approach to analyze energy use in buildings and that the relationships between LEED credits (EA) and energy efficiency are not as simple as assumed by previous research studies. Energy efficiency credits in green building standards and rating systems (e.g. LEED and IgCC) may not reduce energy use in reality.

2.8 Future Research

There are more complicated factors that influence energy use and these factors have to be integrated into design and engineering. Instead of focusing on the scores of these energy credits, such as on the LEED scoreboard, the research suggests that external factors could be more critical. Better energy efficiency can be achieved if these external factors are integrated with the energy efficiency credits in LEED. Based on the research results, here is a list of further research to validate results in the following areas:

1. Investigate the daily On-Peak and Off-Peak, and the monthly On-Peak and Off-Peak energy usages using the hourly temperature data and building occupancies.

2. Investigate how much energy is used for heating/cooling, lighting, and office equipment per building, and find out how to use the energy efficiently.
3. Quantify and weight LEED points based on greenhouse gas emissions from the building materials.
4. Analyze the relationships between energy usage On-Peak/Off-Peak regarding building energy usage per building function and LEED points.

3. URBAN GREENERY SPACE LAYOUTS AND URBAN HEAT ISLAND: CASE STUDY-ANALYSIS OF HIGH RISE APARTMENT COMPLEXES IN SOUTH KOREA

3.1 Abstract

Low Solar Reflectance Index (SRI) is the key to reducing urban heat island. The greenery space has a low SRI due to oxygen generation that further reduces environmental temperature. The focus of this research is to determine air temperature differences affected by different greenery space layouts. The aims are to develop an understanding of whether greenery arrangement has an effect on urban heat island. Field measurements were taken for three different layouts, namely greenery surrounding a complex/building, greenery in the center of a complex/building, and greenery distributed over a complex/building. The study results exhibit gaps between significant temperature differences when solar heat is present (i.e., in the daytime). The sites with greenery in the middle of a complex generated lower urban heat island effects than the other layouts. The study shows that layout of greenery space (and ultimately low SRI materials) can have a significant effect on urban heat island, given the same land area

3.2 Introduction

Urbanization and rapid development increase the consumption of energy that emits greenhouse gases in large quantity. These gases affect global climate and temperature (Pachauri et al., 2014). The effects of climate change can be devastating, as predicted by many scientists (Gornall et al., 2010). Increasing global temperatures will result in a temperature increase in urban areas and abnormal climates generating adverse effects on global climate (Santamouris, 2014).

3.2.1 The Relationship Between Greenery Space and Urban Heat Island (UHI)

Wagner et al. (2013) studied that greenery space is an excellent solution to overcome high temperature in urban areas and generate oxygen to replace carbon dioxide (through plant transpiration). Previous studies had shown the importance of greenery in reducing the effects of heat island effects (McPherson, 1988). Due to limited land space in many major cities, greenery is a luxury and takes away valuable space that could be utilized in other uses (like commercial and residential). As such, maximizing the use of greenery becomes critical; however, the understanding of greenery space and its effects on heat island on apartment buildings is still fairly limited.

3.2.2 Urban Heat Island Effects in Korea

The urban heat island effects result from urban development and greenery space reduction, which cause major climate change (McPherson, 1988). The Korea Meteorological Administration (2008) observed that the average temperature in Korea has increased by 1.5°C/ 100-year, which is higher than the average increase in global temperature (0.6±0.2°C/100-year). The urban heat island in Korea deviates considerably from an idealized, concentric heat island structure, mainly due to the location of the majority of commercial and industrial sectors and the local topography. Relatively warm regions extend in the east–west direction and relatively cold regions are located near the northern and southern mountains (Kim & Baik, 2005).

3.3 Research Hypothesis and Objectives

The research study attempts to develop an understanding of air temperature, relative humidity, and greenery space layout, focusing on a housing complex based on the following hypotheses: 1) greenery layout affects air temperature and relative humidity surrounding the complex, and 2) there are distinctions among greenery space layouts along with means to reduce air temperature in housing complexes due to the different layouts in housing complexes.

Regarding the hypotheses above, there are four research objectives, which are to: 1) demonstrate the sites where the air temperature is different, depending on greenery space layouts; 2) reduce air temperature within the buildings; 3) use greenery space layouts in reducing air temperature in different building layouts; and 4) investigate how greenery layout affects evaporation and shading.

3.4 Previous Research Studies

Research on the urban heat island effect in housing complexes have mostly focused on the effectiveness of greenery in reducing energy consumption and providing comfort to residences through mitigating that effect.

3.4.1 Understanding of the Concepts of Urban Heat Islands

Research related to high-rise housing complexes originates mainly from Asia. Giridharan et al. (2004; 2007) investigated the impact of the urban heat island on high-rise and high-density residential complexes in Hong Kong. The results for the studies showed that the urban heat island, with external factors such as albedo, vegetation, and perforation rate, was affected by weather conditions. Taib (2010) studied thermal comfort in high-rise housing via surveys and temperature measurement.

The UHI effect exists due to greater heat retention of buildings and man-made surfaces such as concrete and asphalt compared to the lesser heat retention and cooling properties of vegetation, which is more abundant in the countryside (Rosenthal et al., 2008). The urban heat island temperature effect can be measured in terms of the urban canopy layer, which refers to the space below the rooftops of buildings, and the mesoscale, which refers to regional temperature measurement (Voogt, 2002).

Wolf (2004) found that urban and suburban areas have hotter air and surface temperatures than rural surroundings. The hottest near-ground temperatures are found in areas with the least vegetation and the greatest urban development. The heat island effect has existed in New York City since the end of the 19th century. Monica Pena Sastre (Student of Urban Planning at Columbia University) found that a difference of at least 1.8ºF (1ºC) already existed at the beginning of the 20th century between the mean temperature in NYC and its surrounding rural areas, and this difference increased over the 20th century (Sastre, 2003). Annual analysis between 1900 – 1997 shows that mean temperatures in New York City based on the Central Park weather station were slightly higher than the surrounding region by approximately 2.2ºF (1.2ºC) to 5.4ºF (3.0ºC) (Rosenthal et al., 2008).

3.4.2 The Use of Greenery Space to save Energy Consumption

Huang (1987) simulated climate influences on housing with DOE-2.1, an energy simulation program created at the Lawrence Berkeley Laboratory to study cooling load, wind speed, shade evaporation and tree size. Guidelines have suggested that trees should be planted to control solar radiation, but should not obstruct the wind when planted near buildings; trees can block solar radiation in summer and reduce energy loss in winter (McPherson, 1988). The National Renewable Energy Laboratory in the United States (NREL, 1995) proposed innovative ways to reduce energy loss in housing complexes. Walker et al. (2009) suggested an energy-saving landscape that can reduce energy consumption about 25%, by providing shade in summer and blocking winds in winter.

The previous studies planned various landscapes around housing to control solar radiation in the summer and north winds in winter to reduce energy consumption in the housing units. However, these studies apply to low-rise housing units, and not for high-rise housing complexes in Asia.

3.4.3 Data Analysis Methods and the Use of Greenery Spaces

To investigate the heat island effect in Singapore, where high-rise buildings are common, the extracted thermal data from Google images was used and compared with land use (Jusuf et al., 2007). Simpson et al. (1996) proved with a climate simulation model that if trees were situated on the western side of the housing, air-conditioning energy could be reduced 10–50%. Simpson (2002) stated that the shade from greenery space affects energy consumption in housing. The results showed that the tree shape and size had a greater impact than their location in energy reduction.

Donovan et al. (2009) examined field surveys and electric bills, and the results suggested that greenery space at the west and south sides of buildings reduced summer energy costs by about 5.2%; however, the energy costs increased by approximately 1.5% on the north sides. Yang et al. (2010) suggested from field surveys that layout, density and ratio of green coverage around buildings influenced urban heat effects in a high-rise housing complex in Shanghai. They stated that shade and solar heat modify urban heat island more than any other factor; additionally, paved road space with shade was cooler than public lawn space without shade. Oliveira et al. (2011) asserted that even small urban greenery spaces could alleviate the heat island effect. They performed a case study of a 0.24-hectare neighborhood garden in Lisbon, and found that the garden's highest temperature was 6.9°C cooler than that of surrounding locations. Ewenz et al. (2012)

showed, through mobile measurements taken from a car, that even a small greenery space can mitigate the urban heat island of central business districts.

3.4.4 Approach to Reduce Urban Heat Island

Studies of urban heat island in Korean housing complexes mainly concern temperature change relative to floor area index and building-to-land ratio, evaluating indoor/outdoor thermal comfort, building shape, and finishing material. Whang et al. (2003) researched temperature changes associated with land use by measuring single-family homes and apartment complexes. They predicted temperature change by considering a discomfort index and land-use patterns. Hong et al. (2003) drew correlations between areas of greenery space, water permeability ratio, floor area index, building-to-land ratio, temperature using biotope survey data, and temperature derived from Landsat Thematic Mapper images in the southern Seoul residential area. The greenery space area and water permeability ratio were inversely proportional to the urban heat island, and floor area index and building-to-land ratio were directly proportional. Whang et al. (2008) measured thermal characteristics of surfaces, air currents, and solar radiation in summer. They suggested that the environment of high-rise housing complexes had a higher temperature and weaker wind speeds than other places, which can create an uncomfortable environment. They also stated that approaches are needed to improve thermal comfort in high-rise housing (Jusuf et al., 2007).

Previous research studies have indicated that urban heat islands were severe in highly populated areas where high-rise buildings areas are common and the lack of greenery space contributed to increasing atmospheric temperature. Prior research also stressed the importance of greenery spaces in alleviating the impacts of rising atmospheric temperature. Most research studies are related to urban heat islands in highly populated housing developments. The urban planning agencies set the rules for different development strategies, such as defining plot ratio, developmental use, floor area ratio, building façade and material types, design, and area of coverage. Developers are required to comply with these rules; however, there are no specific research studies focusing on greenery space layout and its effects on the surrounding atmospheric temperature and relative humidity around highly urbanized areas. The foci of this research are to: 1) Understand the relationships between different green space layouts and atmospheric temperature/relative humidity, and 2) Understand if the relationship could become a policy driver in using greenery space as a strategy to reduce heat island effect.

3.5 Research Limitations

From previous research studies, low-rise buildings were shown to have different cooling and heating loads depending on the location of trees (McPherson, 1988). However, this research study has limited research conditions due to numerous high-rise buildings within housing complexes in Korea. Therefore, this study aims to discover means by which to reduce air temperatures within buildings and how to use greenery space layouts in reducing air temperatures in housing complexes.

3.6 Research Methodologies

This research study was conducted during summer and focused on sunny days, between 12:00 and 16:00, when human activity and cooling loads are at their peak.

3.6.1 Selection of Greenery Space Layouts and Measurement Sites

To classify greenery space layouts in Korea, 70 apartment complexes in the Seoul metropolitan areas, which are called the City of Il-San, were selected as a case study.

Figure 8. Four Types of Greenery Space Layouts

shows four different layout types, which include greenery space around buildings, in the center of complexes, over the entire complex, and no greenery space, which were examined in this research study.

Green spaces in the center of the buildings
(Location: Il-San)

Green spaces in the entire buildings: center, around and near the buildings.
(Location: Il-San)

Green spaces around buildings
(Location: Il-San)

No Greenery Space
(Location: Il-San)

Figure 8. Four Types of Greenery Space Layouts

Mobile and fixed-measurement instruments were installed to collect data from the project sites. Several apartment complexes were selected based on similar building arrangement, shapes and space composition. These selected sites have different types of greenery spaces and were built in the same time period. Additionally, the selected apartment buildings have similar heights and floor area indexes. Apartment building exterior and pavement colors have the same composition and similar albedo as well. Finally, all sites are located within 500m from the selected buildings to avoid errors from variations in latitude and longitude.

Figure 9 shows the locations of the project sites.

Sites 1–3: the main group	Site 4: the comparison group
Site 1: Greenery space around the buildings	Site 2: Greenery space in the center of the buildings
Site 3: Greenery space over the entire buildings	Site 4: No greenery space

Figure 9. Locations of Field Survey Sites

Sites 1-3 are the main groups to compare to Site 4, which has no greenery space (public parking lot). This parking lot is located approximately within 400m from the Site 3, with a similar environmental geometry location as compared to the other three sites. Table 4 lists in detail the site conditions.

Table 4. Present Conditions of Sites 1 Through 4

Items	Site 1	Site 2	Site 3	Site 4
Site Name	5th Gangsun	8th Gangsun	9th Gangsun	Public Parking Lot
Greenery space layout	Greenery space around building	Greenery space in the center of complex	Greenery space on the entire complex	No Greenery Space
Site area (m^2)	11,945	10,930	7,389	3300
Green space area (m^2)	5,242	4,969	3,503	0
Green space ratio (%)	43.89	45.42	47.4	0
Built Time Period	Jan. 1994	Nov. 1993	May. 1993	Jan: 1994
Floors	18 Floors	18 Floors	19 Floors	0
Building Exterior Color	Ivory wall and red roof	Ivory wall and red roof	Ivory wall and red roof	N/A
Pavement of Road	Asphalt and block	Asphalt and block	Asphalt and block	Asphalt and block
Parking lot	Ground	Ground	Ground	Ground

3.6.2 Field Measurements: Mobile and Fixed Measuring Instruments

Three types of mobile instruments, which include Testo AG: 845, 425 and 454, were utilized to measure four elements, which include atmospheric temperature, surface temperature, mean radiant temperature (MRT), and wind speed. The surface temperatures were measured with the Testo 845 to compensate for the variation of temperature in the instrument. Based on the results, the measurement points were established and were collected using the outdoor air flow measurement equipment. Table 5 lists weather information during the field measurement.

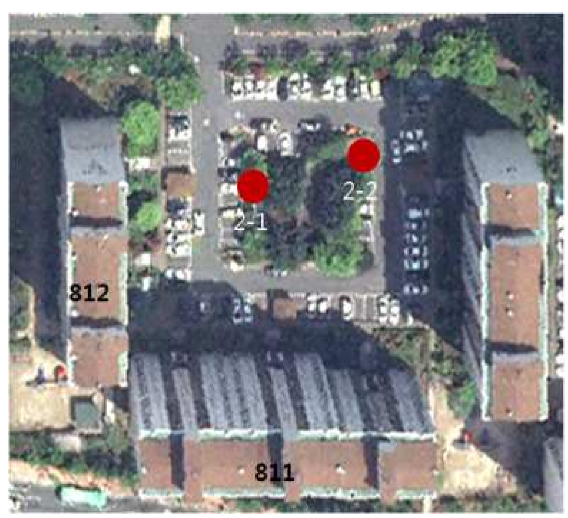
Table 5. Weather Information During the Field Measurement

Day	Weather	Wind Speed	Relative humidity	Visi bility (km)	* cloudiness points	Rainfall (mm)	Instrument	Time
20.Aug . 2010	Slightly covered skies Strong sun	<1m /s	68%	12	6	0	Testo 845	12:00 ~ 24:00
21.Aug . 2010	Clear Skies Very strong sun	2~3 m/s	62%	15	5	0	Testo 845	12:00 ~ 24:00
24.Aug . 2010	Very cloudy skies Rainy from 19:00	<1m /s	89%	15	6	0.5	Testo 845	12:00 ~ 19:00
4.Sep. 2010	Clear Skies Strong sun	1-2m/s	75.8%	17	5	0	Testo 435 Testo 845	12:00 ~02:00
7.Sep. 2010	Clear Skies Strong sun	2~3 m/s	56.5%	18	3	0	Testo 454 Testo 845	12:00 ~ 24:00

(*Cloud Data points: 0-Clear/ 10-Cloud)

To screen sites using precise measurements, four initial points were selected to measure the air temperatures and surface temperatures. This process was repeated three times in August with the Testo 845, every 2 hours beginning at 12:00, at each project site. Each point had the same albedo and was at a similar position, such as in front of an apartment's entrance or in the middle of a walkway. When the air temperatures were measured at a height of 1.5m, the surface temperatures had the same albedo. The collected data clearly shows a temperature difference between greenery space and no greenery space. To construct isotherms, inverse distance weighting was used within ArcGIS 9.3.

After the mobile field survey was conducted to measure the air temperatures and related humidity, two units of the automatic temperature-humidity measuring instruments were installed at each site. The data logger for the automatic temperature-humidity instruments measured and collected the air temperatures and relative humidity data every 3 minutes, from September 17 to 30, 2010. Figure 10 shows the site locations and two units of instruments in the aerial photos.



Site 1: greenery space around the buildings

Site 2: greenery space in the center of the complex

Site 3: greenery space over the entire complex

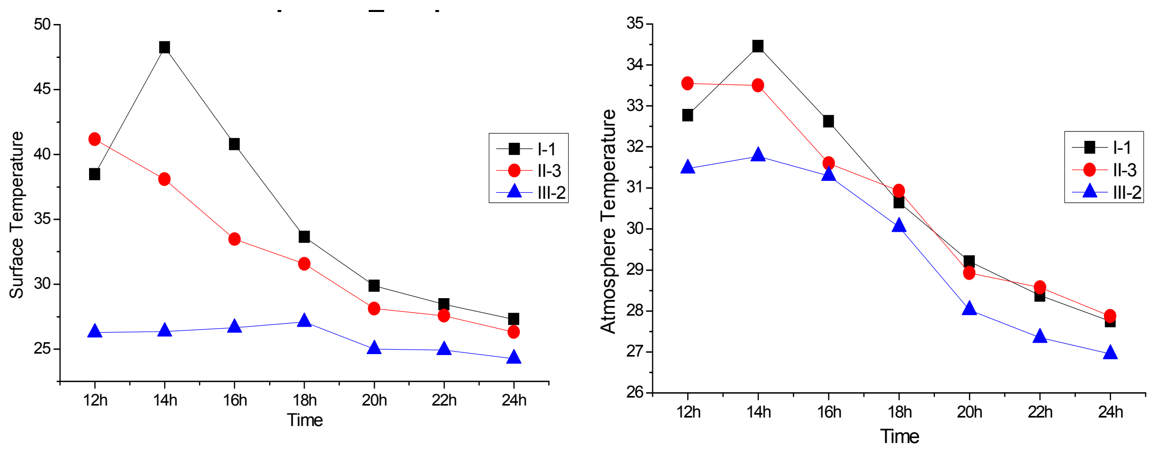
Site 4: Parking lots

Figure 10. Fixed-Measurement Points of Four Sites

3.7 Results: Temperature Reduction Effects by Greenery Space

3.7.1 Comparison Between Surface Temperature and Atmosphere Temperature

From the air temperatures measurement at Sites 1–3 layouts, Site 1 had the highest temperatures and Site 3 the lowest. Temperature variation was greatest at 14:00 for the surface temperatures and the air temperatures at the sites. After sunset, there was little difference in temperature variation. Figure 11 shows the comparison results between surface temperatures and air temperatures.



Site 1: Greenery space around the buildings

Site 2: Greenery space in the center of the buildings

Site 3: Greenery space over the entire buildings

Site 4: No greenery space

Figure 11. Comparison Results of Field Survey with Mobile Instrument (Left: Surface Temperature/ Right: Atmohsphere Temperatures)

Based on the results in Figure 11, U-shaped building spaces had no differences in wind speed, because the buildings blocked the wind. Site 3 was the lowest temperature and Site 1 had the highest air temperature. The air temperature at Site 1 through 3 dropped significantly after 22:00. However, Site 4 stored large amounts of heat during daytime due to the coverage by black asphalt. The asphalt at Site 4 did not have the cooling effects of greenery space. Between 12:00 and 16:00, the surface temperature at Sites 1 and 4 were higher than in Sites 2 and 3.

In daytime temperature, the mean radiant temperatures changed significantly; however, it had no significant difference after sunset. At Site 3, there was little difference between the air temperature and the mean radiant temperature as compared to other sites 1, 2 and 4. These sites had significant differences, from 5°C to 10°C. The measurement results of the air temperatures, surface temperatures, mean radiant temperatures and relative humidity at the sites are shown in Figure 12.

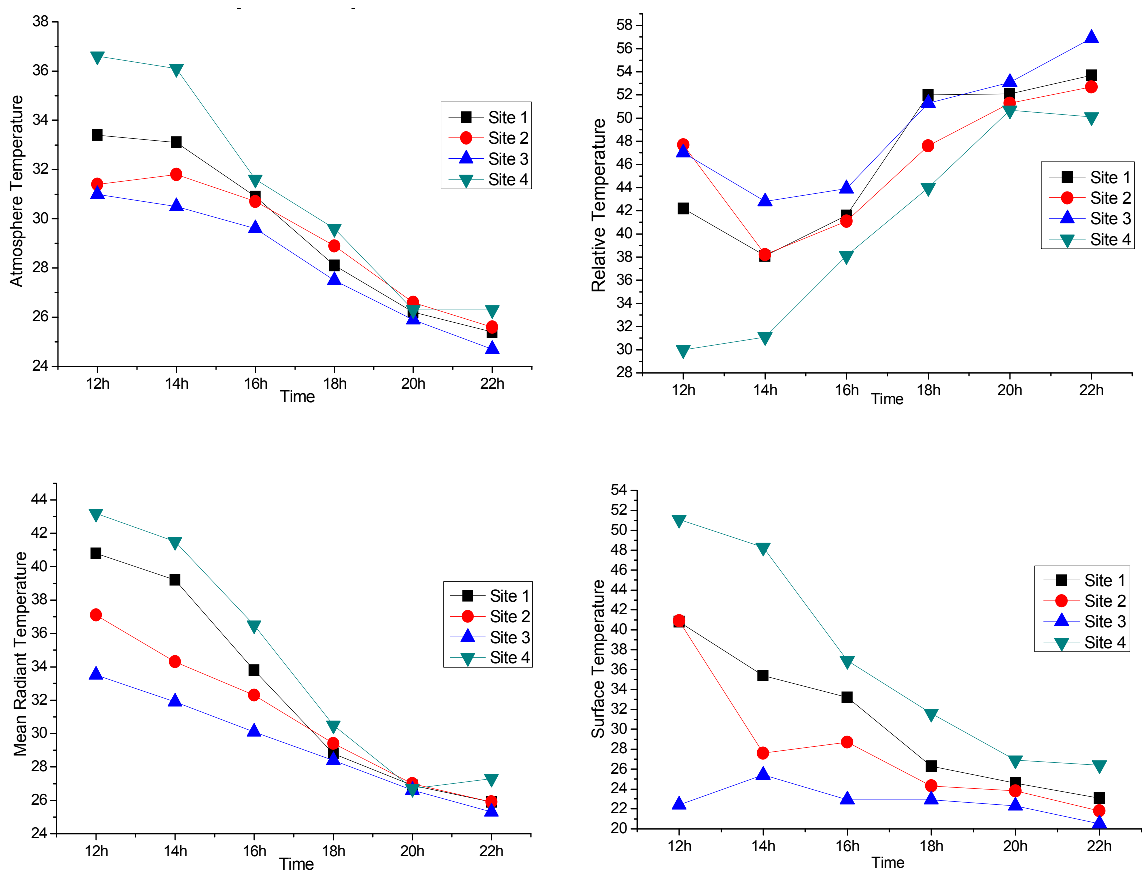


Figure 12. Measurement Results of Mobile Field Survey

The surface temperatures of Site 3 did not show a large change and had the tendency to remain constant. The portable measurement results of Sites 1–4 show that different greenery spaces layout had a spatial distribution in temperature during daytime. To examine if there is significance difference between the means of sites, the statistical method, which is the one-way analysis of variance (ANOVA), was used, and the results are shown in Table 6**.**

Table 6. Statistical Significance Test Results of Mobile Measurement

Time	Significant probability	Scheffe					
		1 & 2	1 & 3	1 & 4	2 & 3	2 & 4	3 & 4
14h/04/Sep	0.000	0.000	0.000	0.003	0.477	0.000	0.000
14h/07/Sep	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Based on the results in Table 3, there was a similarity between Sites 2 and 3. For further analysis to enhance experiment consistency and approach, it was necessary to perform continuous monitoring through fixed equipment for the same period to facilitate observation of temperature variation. To achieve reliable results for temperature reduction, the data logger was installed at Sites 1–4 to measure air temperatures and relative humidity. The meteorological data is shown in Table 7.

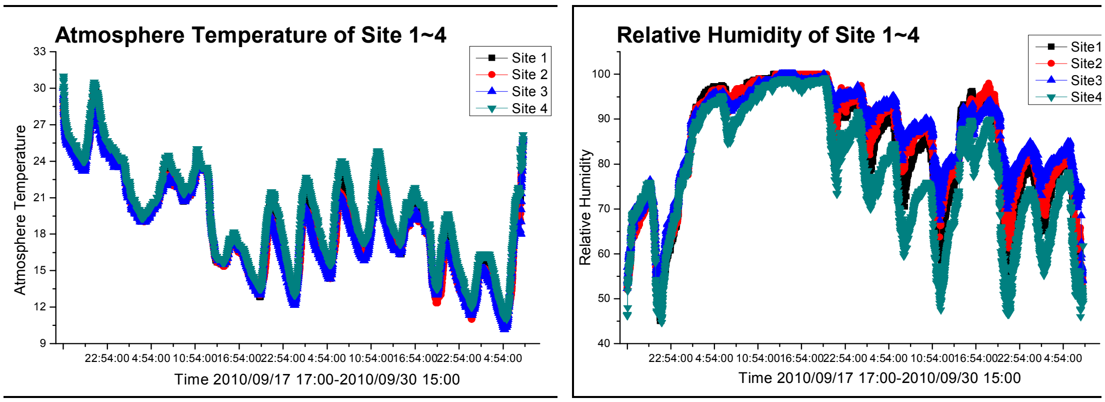
Table 7. Weather Information Data at the Project Sites

Time	**Average temperature (°C)**	**Maximum temperature (°C)**	**Minimum temperature (°C)**	**Average cloudiness points**	**Rainfall (mm)**
17.Sep.2010	22.3	28.8	18.0	8.0	0
18.Sep.2010	24.0	29.9	19.7	5.6	0
19.Sep.2010	20.5	22.4	18.9	10.0	35
20.Sep.2010	22.0	27.9	18.6	9.1	6.5
21.Sep.2010	20.4	22.6	15.9	10.0	45.5
22.Sep.2010	15.8	19.6	11.3	7.8	0
23.Sep.2010	15.3	24.3	10.0	0.8	0
24.Sep.2010	15.6	23.8	9.0	0.4	0
25.Sep.2010	17.0	24.7	11.7	2.0	0
26.Sep.2010	16.7	25.0	12.3	3.8	0
27.Sep.2010	17.1	24.1	11.9	6.5	0.5
28.Sep.2010	13.8	18.9	7.7	0.8	0
29.Sep.2010	10.7	16.5	5.5	5.8	0
30.Sep.2010	12.6	22.1	6.5	4.1	0

The temperature in September in the City of Il-san tended to drop gradually during the study period. The changes in air temperature and relative humidity during measurement period are shown in Site 1: Greenery space around building; Site 2: Greenery space in center of complex;

Site 3: Greenery space over entire complex; Site 4: Comparison group (No Greenery Space)

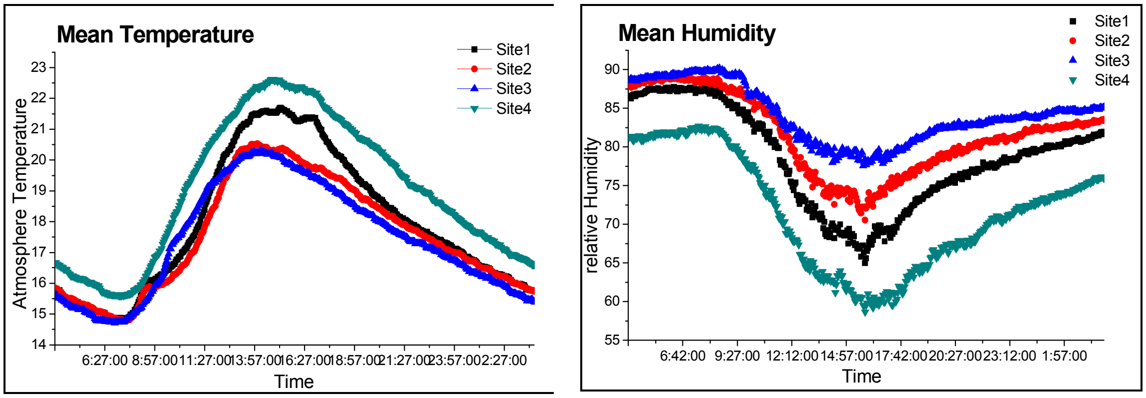
Figure 13.



Site 1: Greenery space around building; Site 2: Greenery space in center of complex;
Site 3: Greenery space over entire complex; Site 4: Comparison group (No Greenery Space)

Figure 13. Changes in Air Temperature and Relative Humidity During Study Period

To organize the data more precisely, the data were classified into 24-hour intervals. Analysis of daytime temperature results were based on sunny days. The 2-week site measurement results showed the same phenomena during the 2 weeks of research study period. To follow the temperature trend, the data were analyzed based on days 6, 7, 8, and 9 (excluding the highest day 1 and lowest days 11 and 12); 30-minute temporal data are shown in Figure 14.



Site 1: Greenery space around building; Site 2: Greenery space in center of complex;
Site 3: Greenery space over entire complex; Site 4: Comparison group (No Greenery Space)

Figure 14. Mean Temperature and Humidity During Measurement Period

Site 1, with greenery space around buildings, had higher temperatures than Sites 2 and 3; Site 2 was slightly warmer than Site 3. Site 4, the control, had a significantly warmer temperature distribution than Sites 1–3. During daytime, temperatures varied between Sites 1–3, but there were slight changes after sunset. Site 4, with no greenery space, had the warmest temperature distribution by far, and little decline after sunset. Average RH tended to follow the order Site 3 > Site 2 > Site 1 > Site 4, in contrast to that of air temperature, with Site 4 > Site 1 > Site 2 > Site 3. Based on the results of field measurements, the air temperature at Site 1 was higher and its RH lower than the other sites. The averages and standard deviations from fixed measurement show that the temperature of greenery space over the entire complex is similar to that in the center but lower than around the buildings, by 0.2–2°C, depending on the layout. To statistically verify each destination, data were subjected to the ANOVA test using SPSS in Table 8**.** This table shows differences in temperature change; however, the differences of the center and entire site layouts were proven neutral or slight. Therefore, if it is difficult to construct greenery space over an entire complex, a similar effect can be achieved by constructing it in its center. From measurements of the U-shaped complex, we discovered that temperature was affected by the layout, and had a daytime variation. Therefore, the lower the housing complex temperature, the more effective it is to establish greenery space in the center of a complex than to use the other layouts.

Table 8. Statistical Significance Test Results of Fixed Measurement

Levene	Significant probability	Scheffe					
		1 & 2	1 & 3	1 & 4	2 & 3	2 & 4	3 & 4
28.504	0.000	0.015	0.001	0.000	0.874	0.000	0.000

1=green space around building; 2=green space in center of complex; 3=green space over entire complex; 4=control group / $p<0.05$ or less

3.8 Finding and Analysis: Casual Analysis of Temperature Reduction by Greenery Space

3.8.1 Analysis by Evaporation Effect

The temperature differences from these experiments of greenery space layout were analyzed in terms of evaporation and shade effects. To determine the evaporation effect dependent on greenery space layout, isotherms were constructed using mobile measurement results and via fixed measurements over impervious surfaces. The isotherms were at 0.5°C intervals over impervious surfaces and depicted various temperatures around greenery spaces. The isotherms from mobile measurement results show that greenery space temperatures were 0.5–5°C lower than those of impervious surfaces. The cool air around greenery spaces moderates the hot air formed on those surfaces. The air temperatures for all project sites are shown in Figure 15.

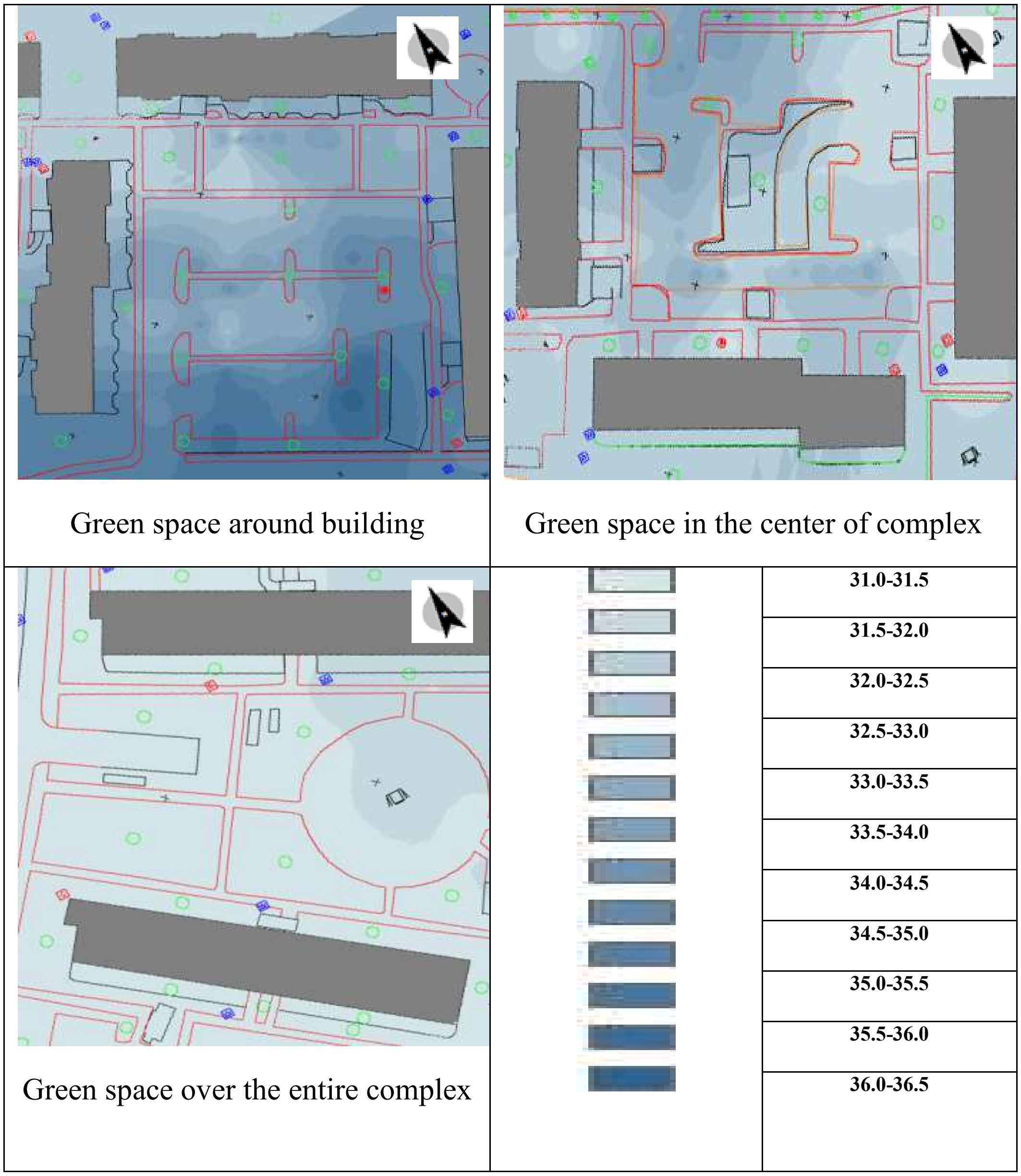


Figure 15. Air Temperatures for the Project Sites at 14:00

Through evaporation, greenery space should be cooler than areas with impervious surfaces, and this has been proven through several studies (NASA, 2010). The temperature of greenery space around buildings was lower than that of impervious surface space by 3–5°C. When there is cool air near greenery space, it does not affect the entire complex or persist near buildings. A small greenery space creates a little cool air over impervious surfaces. Such a small greenery space can cool a housing complex, but its size is not adequate to compensate the hot air produced within that complex. Consequentially, the greenery space around buildings was 3–5°C warmer than at the center of the complex.

Cool air from greenery space affects the temperature of impervious surface space; this space between greenery space and buildings had temperatures very similar to that of the greenery space. This result shows that given a certain size of greenery space, the temperature of impervious surface space can be cooled to a temperature similar to that of green space. Specifically, if green spaces of a certain size are situated like stepping stones, the temperature of impervious surface space between the greenery spaces can be reduced by their evaporation effects. According to the theory of landscape ecology, it is ideal if habitat patches are located contiguously, even if this cannot be done on a large scale. If it is a patch of sufficient size for larger animals, it should be preserved for wildlife habitat. However, if a large patch cannot be maintained for certain reasons, assembling small patches can provide suitable biotic habitats for smaller animals (like birds). This theory applies to greenery space evaporation effects under similar conditions, such that air temperature can be reduced about 3–5°C. The average temperature for greenery space in the entirety of the complex is lower by as much as 5°C relative to that for green space around buildings, and 3°C lower than for green space in the center of the complex. The isothermal analysis shows lower temperatures over the greenery space.

From the measurement results, the evaporation effects of greenery space of a certain size can mitigate the high temperature over impervious surfaces. Therefore, the inclusion of such greenery space within impervious surface spaces can reduce internal residence temperature over that of a continuous impervious surface space lacking green space. In other words, evaporation effects of green space impact impervious surface space when those green spaces are situated contiguously, like habitat patches. Within a residential complex where it is difficult to create large green spaces, distribution of certain sized green space between impervious surface spaces can reduce temperature, although this effect is weaker than with large-scale green spaces. However, the size of green space required to cool the hot air over impervious surfaces needs further study.

To determine evaporation effects depending on greenery space layout, we compared impervious surface space temperatures for the three different layouts. The results show that the greenery space around buildings maintains higher temperatures than other layouts. The difference in temperature persisted during both strong sunshine and after sunset. The air temperature over impervious surface space is shown in Figure 16.

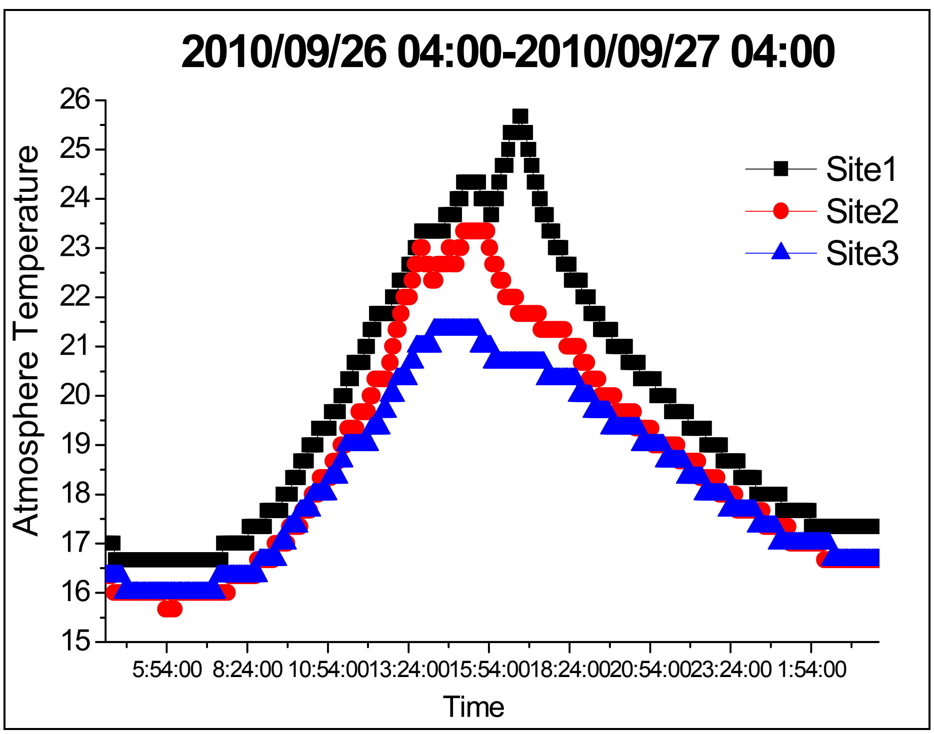


Figure 16. Air Temperature over Impervious Surface Space

3.8.2 Analysis by Shade Effect

The shade was created and modulated by buildings and greenery spaces during the daytime. With increased shade space in the living environment, there is more thermal comfort than in other situations at the same time. In the present experiment, the space with the most shade was Site 3, and Site 1 had the least shade. The numerous flat-type apartment complexes are shaped in either straight, U-shaped, or square-shaped types in Korea. With such configurations, depending on the movement of the sun, shade always forms around buildings, but spaces between them are constantly exposed to the sun, therefore heating the air. The schematic figure of hot air formation without green space is shown Figure 17.

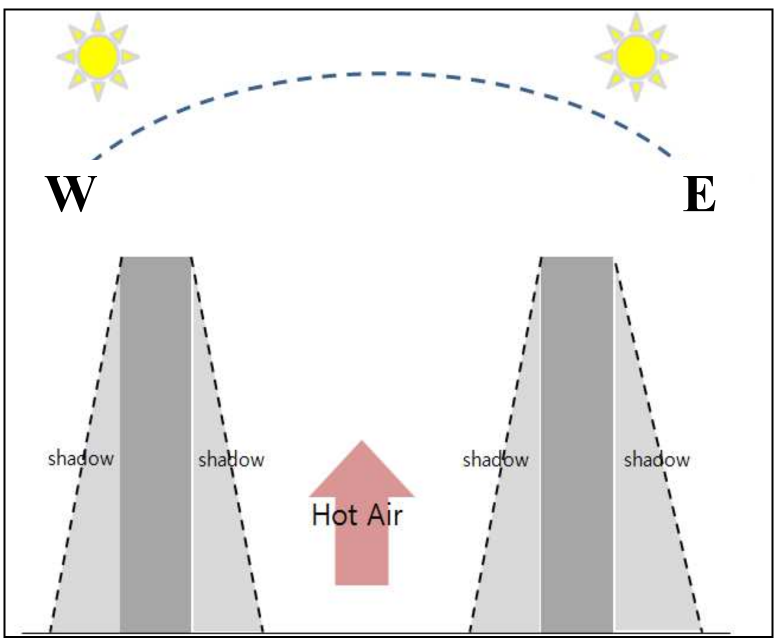


Figure 17. Schematic of Hot Air Formation Without Green Space

Based on the measurement results of greenery space layout, the major effect was analyzed between evaporation and shade of green space. These are the two major effects leading to the formation of the park cool island (PCI) effect. There are significant implications of the variance of PCI effects of green space within the layout. To control hot air in impervious surface space between buildings, we should use evaporation and shade effects on green spaces. If green spaces are sited around buildings, the two effects act near the buildings, and we cannot control hot air from the impervious surface spaces. The schematic figures of hot air information with green space are shown in Figure 18 and Figure 19.

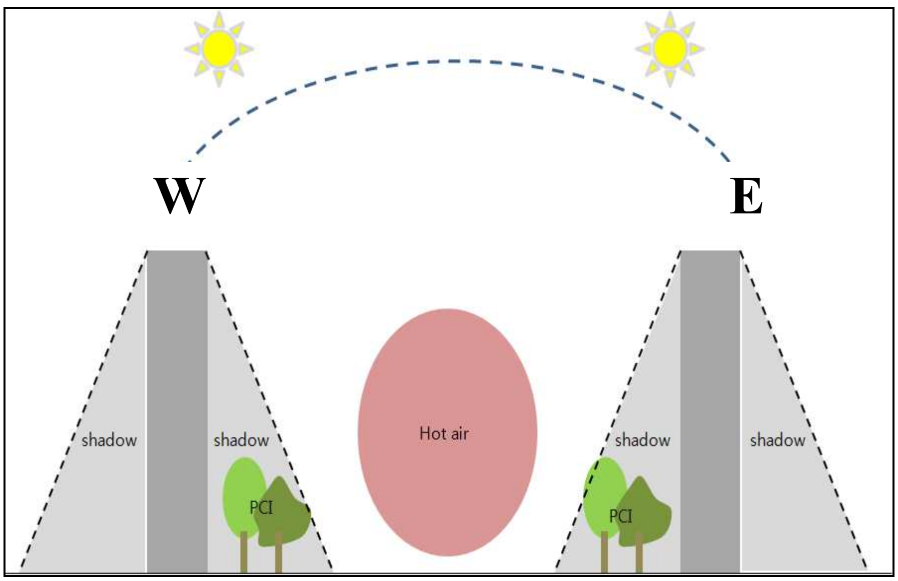


Figure 18. Schematic Figures of Hot Air Formation With Green Space (Greenery Space around Buildings)

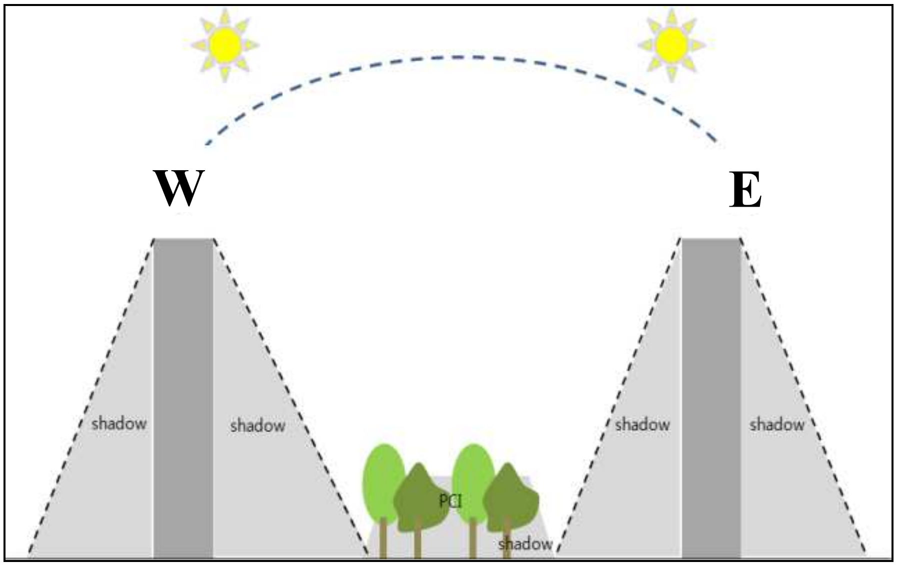


Figure 19. Schematic Figures of Hot Air Formation With Green Space (Greenery Space in Center of Complex)

However, if green space is placed in the center of a complex, shade from buildings and greenery space do not overlap, and we can moderate the hot air from impermeable surface spaces between buildings. This solution situates the PCI effect in the most vulnerable space, which does not benefit from the shade of tall buildings.

3.9 Conclusions and Discussions

To enhance temperature reduction within limited greenery spaces, the space was divided into three layout types and temperature changes observed. The air temperature and mean radiant temperature were lower than at other sites when green spaces were placed over the entire complex site, whereas, when greenery spaces were placed around the complex, it was the highest. In the mobile investigation, the air temperature was slightly higher for greenery space in the center of the complex relative to that over the entire complex. In the fixed investigation, the two complex types had similar air temperatures. The three types of complexes showed differences from 12:00 to 16:00, with a maximum at 14:00.

Two conclusions were found based on the results, which are as follows: 1) Impervious surface space interspersed between greenery spaces is cooled by the latter spaces even though they are not large enough. Small greenery spaces assembled in stepping stones have effects similar to that of a large greenery space. This is the PCI effect, and air cooled by green spaces spreads to surrounding areas. The PCI effect was studied on several occasions by Bongardt (2006), Chen Yu et al. (2006), and others. Through field survey and simulation, these studies demonstrated that green spaces had a PCI effect on themselves and surrounding impervious surface spaces. A field survey in the present experiment showed the same phenomenon. If green space is constructed around the impermeable pavement, the PCI effect, like patches in landscape ecology, will act over the pavement surfaces. Therefore, it is useful to reduce daytime temperature in cities and housing complexes; 2) To lower daytime temperature in high-rise building complexes with equal green space area, it is better to put the space in the center of the complexes rather than around buildings. If the buildings are of parallel shape, U-shape or square shape, hot air forms in the central parts of complexes, which are used mostly for parking lots. If green spaces are around buildings, shade and evaporation effects remain around these buildings and do not affect the central hot air, which elevates temperatures in the entire apartment complex. If the greenery spaces are constructed in the center of the complex instead of impermeable pavement, the shade and evaporation effects of the space moderate hot air and concentration of strong sunshine. This mitigates the urban heat island effect.

Where there is a shade effect, the air temperature difference is from 0.5°C to 2.5°C, and the surface temperature from 3°C to 8°C. These results are consistent with previous research (Onishi et al., 2010) and (Yang et al., 2010), which shows that impervious surface parking area with trees is more efficient than only a lawn plaza in reducing the heat island effect. From these results, it can be seen that green space shade reduces temperatures along with the evaporation effect. Although impervious pavement is necessary for a residential complex, green spaces should permit shading at 14:00 to reduce the temperature. Shading can also reduce the temperature at other times, so that green space should be created to continuously mitigate the heat island effect.

It is necessary to transform all available outdoor space into green space in housing complexes to realize temperature reduction, but this is difficult to accomplish. However, similar effects can be attained by placing green space in the center of complexes. If there are green spaces only around buildings, planting the complex parking lot with green space techniques can achieve effects similar to a green space in the middle of a complex.

Even if green space is positioned in the center of a complex, hot air can be created over impermeable roads. However, if green space is created around the impermeable pavement, the PCI effect, which is like patches in landscape ecology, will cool the hot air over the pavement. These green space layouts will reduce the cooling load during daytime and increase the thermal comfort of residents.

Oliveira et al. (2011) stated that only 0.24 ha of garden space can reduce temperatures by as much as 6°C over the surroundings. Three of the sites in the present study have a similar area (about 0.5 ha) and plant growth, but these can make a difference in temperature with a green space layout. With such layouts, we can mitigate urban heat island more effectively.

To investigate temperature changes versus greenery space layout in apartment complexes, all other variables were controlled. In the future, we must consider other variables that affect internal temperatures of residences to develop various methods of temperature reduction. By quantifying the effects of such variables on temperature reduction, we can obtain more accurate tools for mitigating urban heat islands. To reduce the heating effects of impervious surfaces, further study is needed to determine the minimum green space size that can reduce the temperature, since we did not address the effective size in this work. In the future, it is necessary to measure temperature reduction effects of varying amounts of green space.

4. THE USE OF CLUSRING AND ISOLATION FOREST TECHNIQUES IN REAL-TIME BUILDING ENERGY CONSUMPTION DATA: APPLICATION TO LEED BUILDINGS

4.1 Abstract

Buildings are the highest consumer of energy in the United States in many different sectors including transportation, industry, commercial, and residential buildings. To reduce building energy consumption, many different building technologies, programs, codes, and standards—such as Leadership in Energy and Environmental Design (LEED); Home Energy Rating System (HERS); and the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE)—have been developed based on building environmental performance assessment and energy simulation models. However, these programs, codes, and standards are utilized before or during the design and construction phases. For this reason, it is challenging to track whether buildings could still save energy after construction. The purpose of this study is to detect anomalies from the energy consumption dataset of LEED institutional buildings. The anomalies are identified using two different data mining techniques, clustering and isolation Forest. The paper demonstrates an integrated data mining approach that helps in evaluating LEED Energy and Atmosphere credits after construction.

4.2 Introduction and Research Scope

Based on the Commercial Building Energy Consumption Survey (CBECS) in 2016, there were 5.6 million commercial buildings in the United States in 2012, comprising 87 billion square feet of floor space (USEIA, 2016). This level indicates a 14% increase in the number of buildings and a 21% growth in floor space since 2003 (USEIA, 2013). A building's size, function, and geographic location are the key elements that affect the use of energy consumption.

In the United States, people spend 90% of their time indoors, working, living, shopping, and entertaining in buildings that consume much energy (Bose & Diette, 2016). Since most energy comes directly or indirectly from fossil fuels, buildings are responsible for large amounts of greenhouse gas (GHG) emissions, representing approximately 36% of the entire nation's annual energy consumption (UNEP, 2009). Building energy consumption and GHG generation have been increased and are projected by the EIA to increase another 30% by 2030 (Kwok et al., 2016).

The United States accounts for approximately 20% of world energy consumption. Buildings consume roughly half (49%) of energy consumed in the US, which is the same energy consumption from the combined sectors of transportation and industry. According to the U.S. Energy Information Agency (EIA), fossil fuels supply three-quarters (76%) of the energy consumed by the building sector. The use of fossil fuels to generate energy results in the production of carbon dioxide and other GHGs that scientists increasingly agree are driving climate change.

In order to achieve more than 60% energy consumption reduction for the building sector, new technologies, regulations, integrated building design, and other strategies will be required (Torcellini, 2006). However, these new technologies will not assure efficiency improvement by themselves. In order to save energy consumption, many different energy-saving programs, codes, and design standards (e.g., ASHRAE 90.1, LEED) have been tried, such as solar heating, passive cooling, natural ventilation flow, and use of daylight, for building sites and their surroundings; however, these standards are applied during the pre-construction phase to meet the requirements. Therefore, this study presented that LEED EA credits, which are representative as a design standard in this study, are still effective for energy saving after the construction phase. As such, this paper attempts to find energy consumption patterns using real-time energy consumption data using clustering analysis, to find anomalies using the isolation forest method and finally to examine the LEED EA credits after construction.

4.3 Research Objectives

This research study endeavors to understand the building energy consumption patterns for the post-construction phase and find energy consumption data anomalies using the clustering and isolation forest techniques during data processing. The research study has three primary objectives: (1) to identify building energy consumption patterns using clustering analysis, (2) to detect anomalies in the clustered dataset using the isolation forest method, and (3) to examine how these anomalies impact the LEED EA and OEP credits of the certified buildings.

4.4 Previous Research Studies

Over the decades, extensive research studies have been performed by various researchers on energy consumption, energy modeling, performance assessment, and LEED certifications. A few studies have evaluated the relationship between building energy consumption and LEED EA credits and the use of different statistical analysis to analyze building energy performance (Wu et al., 2016). The research identifies the gaps in prior studies and thus evaluates LEED EA credits using data mining methods to find the pattern of consumption and anomalies in those patterns.

4.4.1 Understanding LEED Rating Systems-EA Credits

The LEED standard is widely used for rating buildings' performance in the United States (Fowler et al., 2006). There are six categories of LEED credits, which include Location and Transportation, Sustainable Sites, Water Efficiency, Energy and Atmosphere, Materials and Resources, Indoor Environmental Quality, Innovation and Regional Priority and four different LEED certification levels, including certifield, silver, gold, and platinum that are based on the number of points awarded in Table 9.

Table 9. LEED Certification Comparison Between LEED NC v2.2 and LEED 2009 with EA Achievable Points (U.S. Green Building Council, 2009)

Certification Level	LEED-NC v2.2	LEED v2009
Certified	26-32	40-49
Silver	33-38	50-59
Gold	39-51	60-79
Platinum	52-69	80-110
EA Achievable Points	17	35

The LEED system has grown over the past years. Since LEED version 1.0 in 1998 and v2.0/2.2 in 2000, the certification system has been upgraded to nine rating system products, which include Homes, Neighborhood Development , Commercial Interiors, Core and Shell, New Construction, Existing Building, Schools, Retail, and Healthcare. In 2009, LEED v2009 or v3 was built based on the previous version of the rating system, and LEED v4, which launched in 2013, focused on increasing technical stringency from the past versions and developed new requirements for project types. The historical LEED Rating Systems (only focusing on EA credits) from v2.0 to v3.0 are shown in Table 10. The difference between the new rating systems and the old versions is that each point has been devalued and specified in LEED 2009. To achieve points for the EA category, EAc1

(Optimize Energy performance) has added nine more points and EAc2 (Renewable Energy) has added four more points.

Table 10. Historical LEED Rating Systems from v2.0 to v3.0 (EA Credits)

Description		Acquirable Points		
		v2.0/ v2.1	**v2.2**	**v3 or v 2009**
Prereq 1.	Fundamental Building Systems Commissioning	Required	Required	Required
Prereq 2.	Minimum Energy Performance	Required	Required	Required
Prereq 3.	CFC Reduction in HVAC&R Equipment	Required	Required	Required
EAc1	Optimize Energy Performance	1 to 10	1 to 10	1 to 19
EAc2	Renewable Energy	1 to 3	1 to 3	1 to 7
EAc3	Additional Commissioning	1	1	2
EAc4	Ozone Depletion	1	1	2
EAc5	Measurement & Verfication	1	1	3
EAc6	Green Power	1	1	2
Points		**17**	**17**	**35**

4.4.2 LEED EA Credits vs Building Energy Consumption

Building operation is responsible for about 30% of GHG emission and accounts for about 40% of primary energy consumption globally (Kwok et al., 2016). Furthermore, GHG emission from the building sector is expected to grow in the next decades as a result of rapid economic growth (Grubb et al., 1991). The U.S. Green Building Council (USGBC) and its LEED green building rating systems were designed (USGBC, 2007) to reduce the environmental and health impacts of buildings.

The LEED rating system is a framework that facilitates a streamlined implementation of sustainable construction principles (USGBC, 2003a). The three main types of benefits of sustainable constructions are environmental, economic, and health and community (Diamond et al., 2006). For environmental aspects in particular, benefits include improved air and water quality, reduced energy and water consumption, and reduced waste disposal (USGBC, 2003b, 2004). However, the LEED rating is tallied during the pre-construction phase to award points based on simultaneous construction and design development (USGBC, 2003b). In addition, the early design and pre-construction phases of a building are the most critical times to make decisions on its sustainability features (Azhar et al., 2010). Due to the LEED point systems, those designing LEED rating systems are more concerned with earning points than creating environmentally friendly buildings (Ding, 2008). No matter how unsustainable a building is, it can get LEED certification (Retzalff, 2009).

4.4.3 Anomaly Detections and Isolation Techniques

Anomaly detection techniques used to be classified into two main categories, anomaly detection and misuse detection. Anomaly detection means storing a user's behaviors in a database and comparing the user's current behavior with the data in the database. Also, if the deviation is large enough, this signifies an abnormality in the network (Chaturvedi et al., 2012; Lappas & Pelechrinis, 2007; Sun & Wang, 2009; Sushil et al., 2012). In contrast to misuse detection, anomaly detection utilizes the reverse approach. In other words, it defines normal system behavior and defines any other behavior as abnormal (Helman et al., 1992). The anomaly detection technique has been used in various fields to detect network intrusion or failure, such as credit card fraud (Van et al., 2015), auto insurance fraud (Nian et al., 2016), tax fraud (Bonchi et al., 1999), customer activity monitoring and profiling (Singh & Singh, 2015), malware/spyware detection (Aziz et al., 2015), data cleaning (Sapienza et al., 2015), and securities fraud (Barse et al., 2003). Usually anomaly detection techniques are applied for the same reasons: (a) to identify normality by calculating, (b) to determine a metric to calculate an observation, and (c) to observe metric measurements, including anomalies (Ian Davidson, 2007).

Existing models of anomaly detection constructs a standard profile and identify instances that do not confirm to the standard profile (Liu, Ting, & Zhou, 2008). Isolation-based anomaly detectors are a new kind of anomaly detector that does not rely on any density or distance measure (Liu et al., 2010). Several extensive studies have detected anomalies for both static and dynamic network topologies (Abe, Zadrozny, & Langford, 2006; Akoglu, Tong, & Koutra, 2015; Bhuyan, Bhattacharyya, & Kalita, 2014; Gogoi, Bhattacharyya, Borah, & Kalita, 2011). For automation purposes, isolation Forest (iForest) is probably the best approach for numeric attributes, since random forest cannot handle nominal attributes unless converted to numeric form (Carrasquilla, 2010). Also, Vengertsev et al. (2005) evaluated three types of anomalies to determine graph anomaly datasets and identified that iForest showed the best accuracy for global and local anomalies. In addition, iForest is a mass-based approach that employs the level of depth and gives a better scope of integration with other methods for better accuracy (Ting, 2009). Thus, with different datasets, isolation forest proves to be more accurate when compared with other anomaly detection methods, which serves as a motivation for implementing an isolation anomaly detection method in the energy consumption datasets of LEED buildings. The research involving data management, clustering of data, and anomaly detection using the isolation forest algorithm is explained in the following sections.

4.5 Research Methodologies: Data Management, Clustering and Isolation Frameworks

For this research study, two different data mining techniques, k-means clustering and isolation forest, were applied to analyze real-time building energy consumption data.

These techniques were selected for different purposes and to implement a unique approach to energy consumption analysis for LEED institutional buildings. The primary intention of this study was to identify LEED certified buildings with similar consumption patterns using k-means clustering and to detect anomalies in these patterns to evaluate the LEED EA credit points of the certified buildings. To detect anomalies, the modified isolation forest algorithm was proposed to build an ensemble of iTrees for the clustered dataset and then detect anomalies of the instances that had shorter paths on the iTrees stumps. Figure 20 represents the proposed research methodology to detect anomalies on the consumption pattern of LEED buildings, the clustered-isolation (CI) framework.

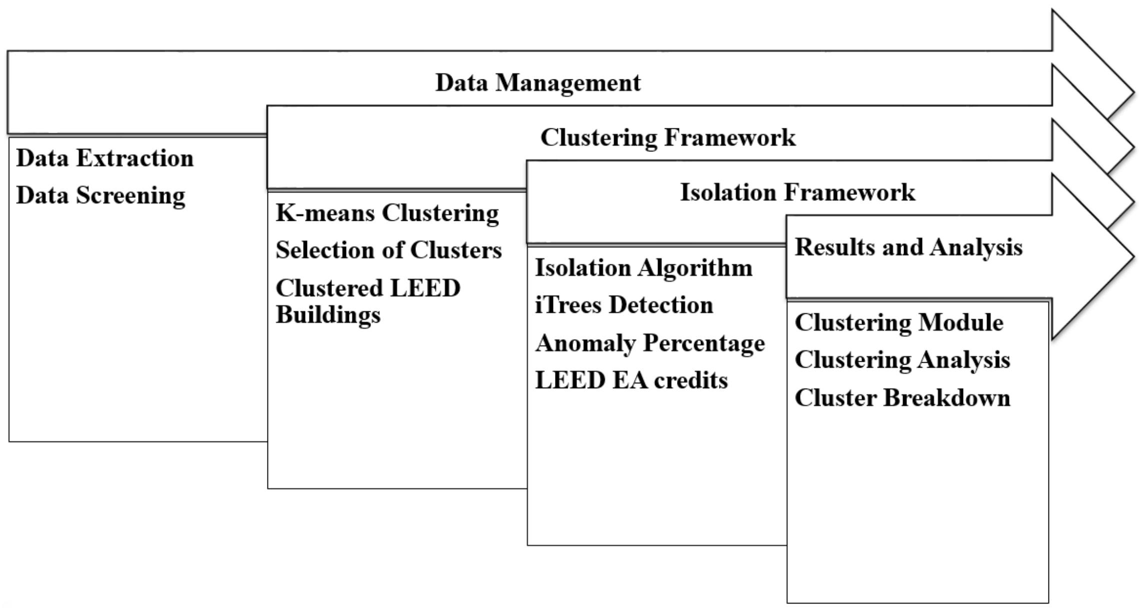


Figure 20. Research Methods Flowchart to Detect Anomalies Using CI Framework

4.5.1 Data Management

Data-gathering methods are often loosely controlled, resulting in out-of-range values, impossible data combinations, missing values, redundant information, and noisy and unreliable data. The process of data management involves two different steps, including data extraction and data screening. The primary data collected for the research include data from the energy information system (EIS) of an institution of nine LEED buildings, which are extensively used for different research, classes, and administrative purposes. They are yearly data of the daily totals from the nine buildings, which include building energy consumption (electricity).

4.5.1.1 Data Extraction

The second part of data collection involves data extraction from different reliable sources using the Python Beautiful Soup algorithm (Yih et al., 2006). The primary task of this data extraction was to understand the interpretations of the dataset. The output label needed to be clearly stated to help in correlating and analyzing the data features. This could be done using Fisher information, which provides a way of measuring the extent of how much one feature is dependent on another within the dataset. The dataset was analyzed for its ability to undergo dimensionality reduction, which helps to understand output visually. In this study, data extraction on external factors was performed to speed up the process of data collection. The factors include climate data such as temperature, humidity, and precipitation from the U.S. meteorological department. The algorithm and data extraction were learned at the machine learning repository at the University of California, Irvine, which has datasets of different meteorological data (UCI, 2015).

4.5.1.2 Data Screening

The dataset included 3,294 data points from nine different buildings with a small dimensionality, and there was a need to look for false positives in the data and omit them. Another Python script was written to check for these anomalies. Thus, this needs to be cleaned up or omitted to analyze certain models. Furthermore, the data screening process simplifies the search space a level further by consolidating valid samples.

4.5.2 Clustering Framework

Partitioning data into groups based on their consumption trends and pattern was essential to detect anomalies between different ranges of energy consumption. It helped in comparing with LEED EA credits with respective to the consumption range. The research involved yearly data on energy consumption of LEED certified buildings. According to Rodrigues et al. (2003), good clustering criteria include two parameters: compactness and separation. The k-means algorithm minimizes the mean square errors between each sample and its associated cluster center, where k refers to the pre-specified number of clusters (Rodrigues et al., 2003). The algorithms have the advantage of clear geometrical and statistical explanation and work conveniently with numerical attributes (Chicco, Napoli, & Piglione, 2006). K-means algorithms take the input parameter, *k*, and partition a set of *n* objects into k clusters so that the resulting intracluster similarity is high and at the same time the inter-cluster similarity is low (Han & Kamber, 2006). It has been shown that k-means algorithms perform better than another commonly used clustering algorithm, Kohonen Self-Organized Maps (SOM), on energy consumption (Rodrigues et al., 2003). Indeed, SOM performs better when the dimensionality is high (Jain, Murty, & Flynn, 1999), but K-means are better suited when dimensionality is relatively low (four clusters), which is the case for our application.

The research involved two steps: clustering using k-means clustering and anomaly detection using the isolation forest method. The primary focus of clustering analysis was to identify buildings with similar consumption patterns and group them together. The reason to do this was to connect with the isolation forest method. During isolation forest, the computer selects a subsample by itself. It was necessary to make sure that anomaly detection was performed on comprehensive similar consumption patterned data rather than random unpatterned datasets, which could impact the accuracy of the anomalies. Thus, clustering using k-means was used to cluster buildings with similar energy consumption together.

Since the research involved nine LEED buildings, it was possible to identify anomalies for small datasets using traditional methods. However, the goal of this research was to implement a novel technique for analyzing energy consumption data that could handle both large and small datasets.

4.5.3 Isolation Framework

The proposed CI framework detects anomalies from the clustered buildings to learn the pattern of anomalies for LEED buildings with similar consumption patterns. For this study, isolation forest is used as the anomaly detection method because it is faster and more reliable than other outlier detection methods. Hodge and Austin (2004) studied three different types of anomaly detection—unsupervised clustering, supervised classification, and semi-supervised recognition—and identified semi-supervised detection to be the most effective method with the greatest accuracy. Ensemble-based minimum margin active learning is a simple, novel method for detecting anomalies using unsupervised learning (Hodge & Austin, 2004; Yamanishi, Takeuchi, Williams, & Milne, 2004). To detect anomalies and to select high-confidence unlabeled factors, a new and novel isolation forest algorithm was adopted that was faster and had greater accuracy than Oak Ridge Cyber Analytics (ORCA) and random forest (Liu et al., 2008). The most important assumptions of the isolation forest algorithm are that the anomalies are a minority and the attribute values are different from each other (Liu et al., 2008). The isolation forest algorithm is best suited to high dimensionality (Liu et al., 2008), where the presence of irrelevant attributes (unlabeled data) is high (the case in big data), and in situations where the training set requires no anomalies, which is an important requirement for the CI framework.

The isolation forest has three stages: training, testing, and evaluation. The method builds an iTree for the consumption dataset, and then normal consumption patterns are clustered at the top end of the tree, whereas the anomalies stay at the roots. The advantage of iTrees is that it can provide results of high dimensionality and efficiency with few subsampling data. It requires minimal time and memory to run the program and select the best among the irrelevant attributes.

4.6 Results and Analysis

The section has three different subsections including the clustering module, where some clusters and clustering of buildings based on energy consumption are explained using k-means algorithms. The second part includes the isolation forest module, where each cluster is processed by training and testing using the isolation forest algorithm. The final section includes the LEED EA credits of the buildings and their respective anomaly points, which provide us an overview of the impact of anomalous points on LEED EA credits, particularly on "Optimizing Energy Performance."

4.6.1 Clustering Module

As discussed in the previous section, clustering is performed using a k-means clustering that is more suitable (for this research) and reliable. The segmentation of millions of data points and the dataset is done after cleaning the data. Figure 21 explains the number of clusters selected using the consumption data.

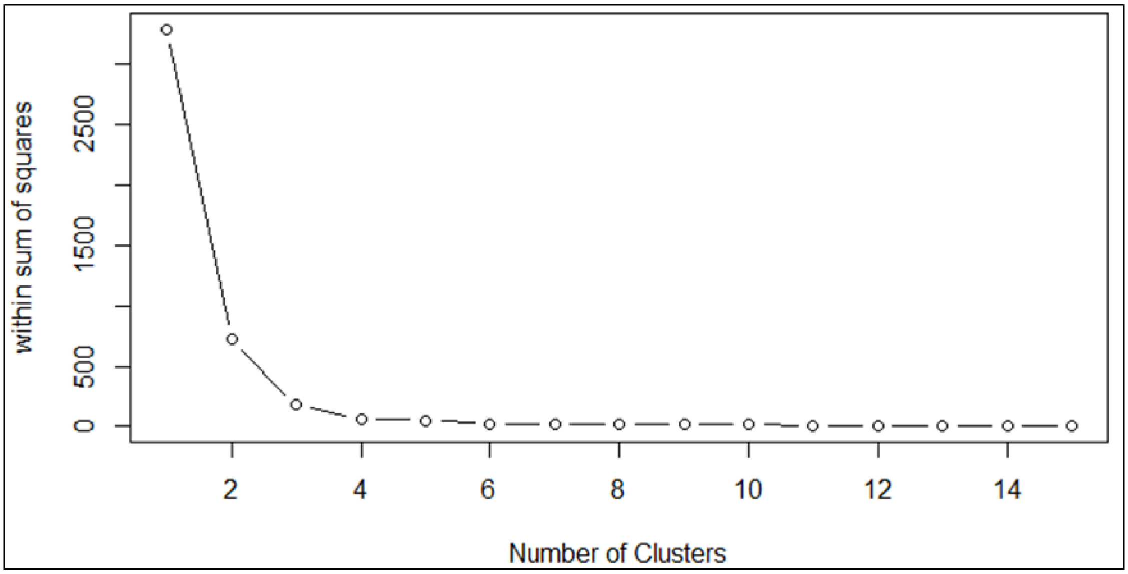


Figure 21. Selection of Clusters

The first step is to identify the points for all samples in a spatial domain. Once the point is fixed in the spatial domain, the centroid is plotted for all clusters. Third, the points nearest to the centroids are identified, and the centroids are recalculated and shifted. This step gives the weighted averages of all points, and finally, iteration is continued until saturation. Cluster analysis is a bottom-up approach because statistical analysis is involved. The selection of clusters is done using the clustering algorithm. From Figure 21, it is noted that there is a steep curve from 1 until 4 and then the graph is more saturated from 4 to 14 clusters. This helps to identify that the number of clusters for the datatype can be effectively four.

4.6.2 Clustering Analysis

The k-means algorithm was used to cluster buildings into four different clusters. Each cluster had different sets of buildings, and some of them overlapped with other clusters.

Figure 22 illustrates the four different clusters in the dimensionality space. Using the k-means clustering algorithms, data were plotted to visualize the clusters of buildings, as shown in Figure 22. The different colors in the scatter plots represent different clusters connected with the discriminant coordinates on the X- and Y-axes, respectively. From the figure, it is evident the buildings are equally segmented and thus there are not many changes in the number of data points.

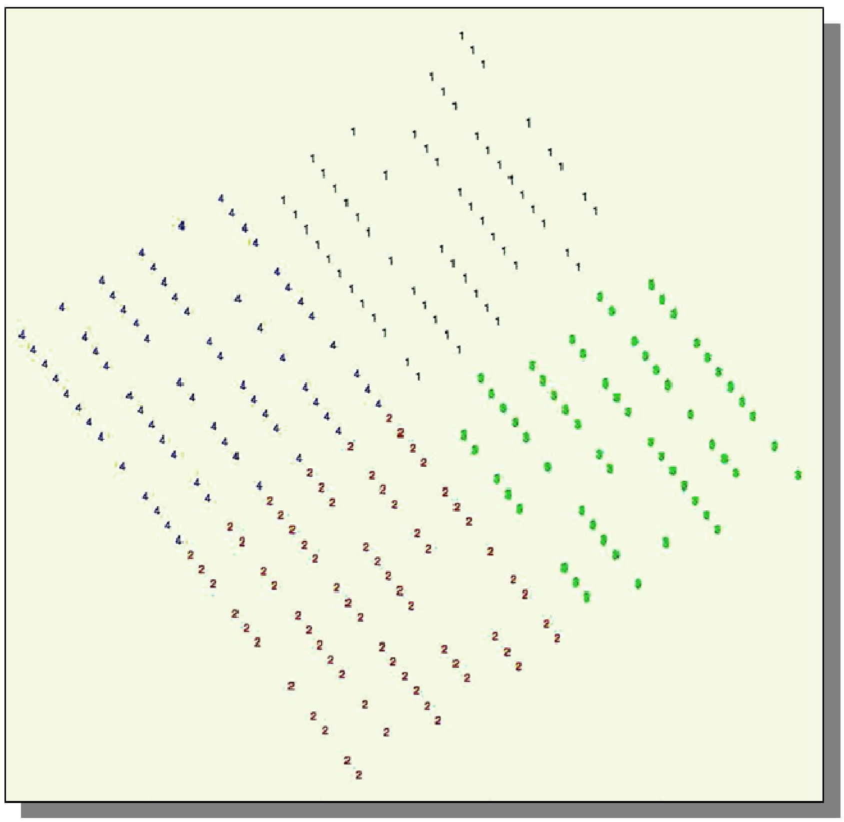


Figure 22. K-means Clusters (1-4)

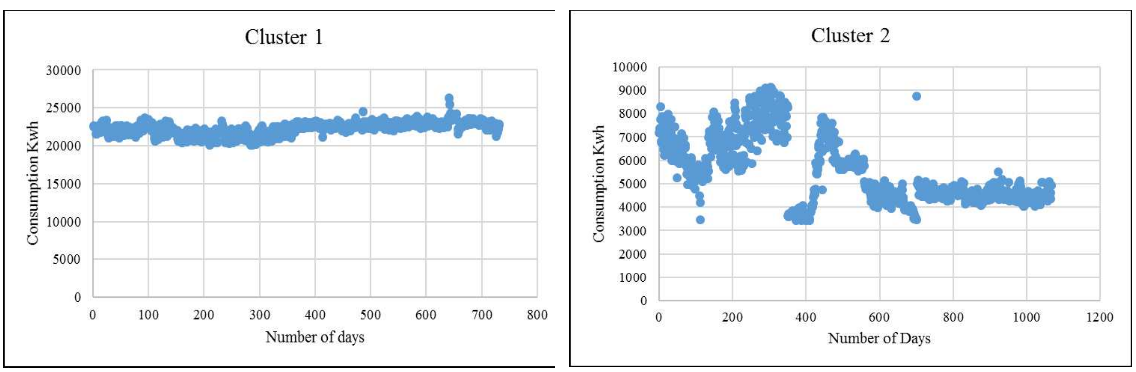
Cluster 1 has two buildings and 731 data points. Similarly, Cluster 2 has 1,066 data points with three buildings, Cluster 3 has 747 data points with three buildings, and Cluster 4 has 746 points with three buildings. Thus, the clusters share almost close to equal numbers of data points. Table 11 shows the different clusters and their respective buildings and numbers of data points.

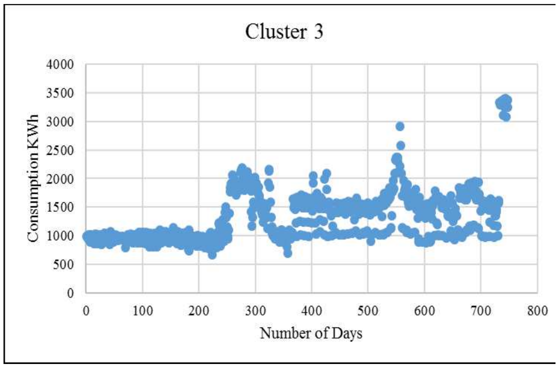
Table 11. Cluster and Respective Buildings and Data Points

Clusters	Building Number	Number of Data points (Days of Energy Usage)
1	3,8	731
2	1,5,9	1066
3	2,4,5	747
4	1,6,7	746

4.6.3 Cluster Breakdown

After clustering, each cluster was tested by the scatter plots to understand if the clusters were well grouped together. Figure 23 shows four different clusters and the buildings scattered on each cluster. It is clear that Cluster 1 has a group of energy consumption patterns that are plotted on different spatial dimensions. Based on the scatter plots, the outliers were visualized efficiently in a few clusters from the figure below.





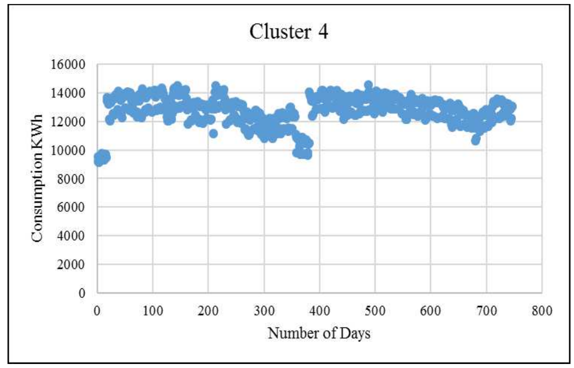


Figure 23. Breakdown Scatterplots of Different Clusters

Even though Cluster 1 looked like a group all together, it was necessary to see a deeper analysis of this dataset to find energy consumption patterns. It is obvious from Clusters 2, 3, and 4 that there were outliers. The primary goal was to check whether these outliers were anomalies regarding consumption to determine the impact on the LEED EA credits.

4.6.4 Isolation Framework

After clustering, the next step was to identify the anomalies using the isolation forest algorithm. Isolation forest was identified as an efficient anomaly detection method due to its faster process and better accuracy than the random forest and local outlier factor (LOC) methods (Liu et al., 2010). Each cluster of data was also utilized to identify anomalies using the iForest algorithm. Figure 24 shows a flowchart to explain the process of the isolation forest algorithm.

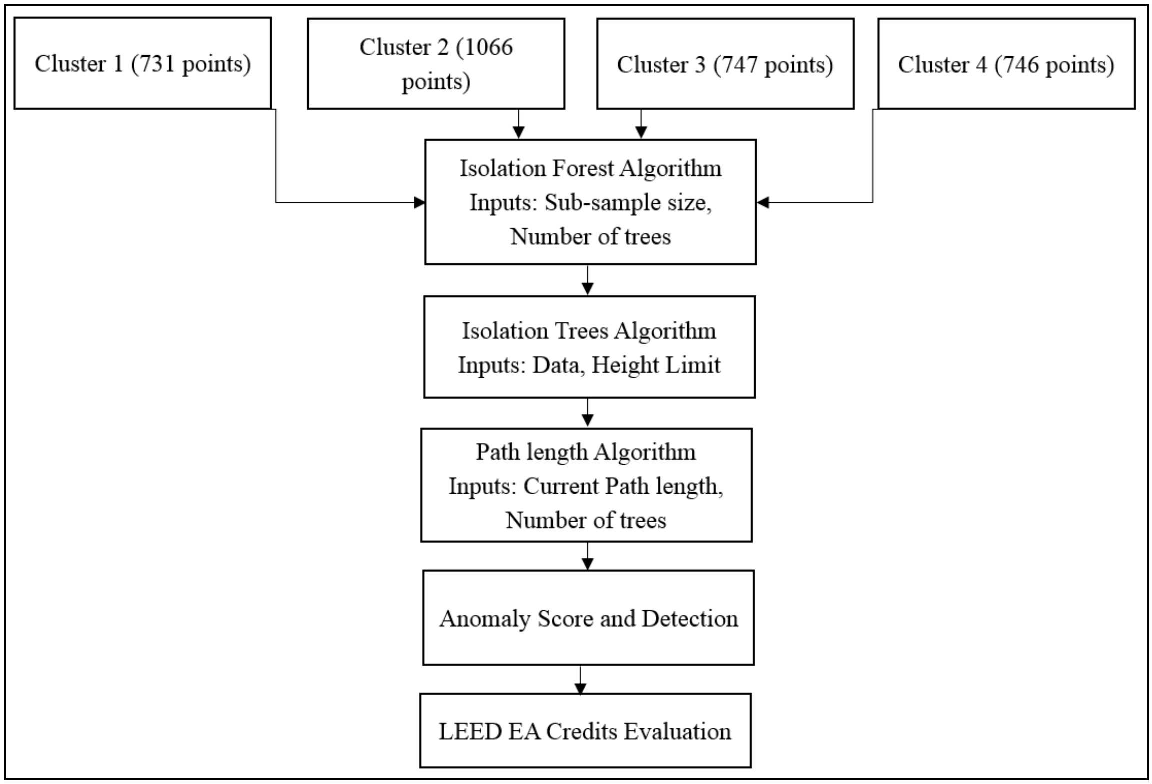


Figure 24. Flow Chart: Process of Isolation Forest Algorithm

4.6.5 Isolation Forest Validation

The isolation framework involves three different algorithms, including the iForest algorithm, in which the sample size and number of trees are the inputs. The second algorithm is the iTrees algorithm, in which the tree height and height limit are given as inputs. The final algorithm is the path length algorithm, which provides the longest and shortest path lengths of the anomalies along with the anomaly points. Finally, the anomaly score is determined for each anomaly point and is important to understand the anomalous behavior of energy consumption. The following subsections explain the anomalies of each cluster, the buildings connected with these anomalies, and LEED EA credit evaluation.

4.6.5.1 Cluster 1 Anomaly Detection

As discussed earlier, Cluster 1 contained two different LEED buildings of similar consumption pattern together and had about 730 points. *Figure 23* shows that Cluster 1 did not have many anomalies. The iForest algorithm was performed using the R programming language, and the outputs were stored as csv files for further processing. Anomaly detection using the iForest algorithm involves two different processes. The first stage is training, in which the machine learns the pattern of consumption of similar data points. In this study, the machine was trained with 60% of the data points and tested with 40% of the data points. The advantage with Isolation Forest is that it requires very few subsamples to identify anomalies and thus requires less processing time than other detection methods.

The first step in running the iForest algorithm was to identify the subsample size. Empirically, the sample size was determined to be 2^8 or 256, within which the algorithm could provide enough details to perform anomaly detection over a wide range of data points (Liu et al., 2012). With the available dataset, 256 sample sizes were determined for each cluster, which contributed close to 30% of the data points of each cluster. The second step was to finalize the number of trees or the iterations to which the data must be processed. Based on the recommendation from (Liu et al., 2008), the path lengths usually converge well before 100 iterations, and thus 100 iterations used to converge the number of trees for this study. The iforest algorithm was processed with the subsample ($n = 256$ and number of trees $t = 100$). The other important inputs were the height limits of the trees beyond which the trees did not converge or branch out. For this type of data to converge, the height limit is determined to be 10 for all four clusters. So the path length numbers were equal to or lesser than the height limit.

After all the required inputs were provided, the first steps of the iForest algorithm and iTree algorithm were processed to identify the anomalies from the 100 iterations and 256 samples. This is a new approach to anomaly detection from energy consumption data. This study was ensured that each cluster was treated in a similar way and regularly tested the algorithm for any discrepancies. Figure 25 shows the possible anomalous range and the specific anomalies in that range of data points. From Cluster 1, 57 anomalies identified as well as their path lengths and anomaly scores.

Cluster 1 has two buildings grouped under them that includes Building 3 and Building 8 out of which Building 3 had only six anomalous points and Building 8 had the rest of the 51 anomalous points.

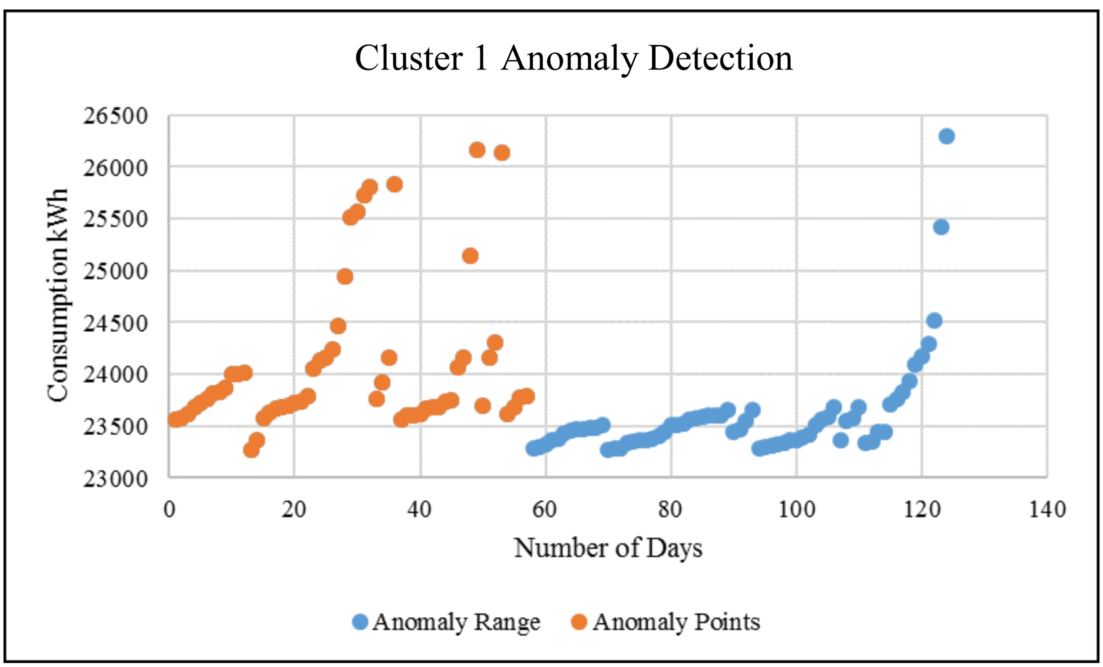


Figure 25. Cluster 1: Anomaly Detection

4.6.5.2 Path Length and Anomaly Score

According to the isolation algorithm, the anomalous points are close to the root of the tree and the normal points further away, and a deeper tree has more convergence. The anomaly score ranges from 0.5 to 1, and when the value tends to be closer to 1, it indicates the higher strength of the anomalous point. Also, data points with short path length numbers are closer to the root of the tree, which means the data points are anomalous points for the given dataset.

From Cluster 1, the average path length was 2.83 (path length starts from the 1st node as 1), which indicates that the anomalous points in this cluster made sense and were accurate. Also, the average anomaly score of the cluster anomalies was 0.83, which again proves that the anomalous points had a higher degree of accuracy. Thus, the inference from this cluster is that Building 8 had more anomalies, which will be discussed under the LEED credit evaluation section.

4.6.5.3 Cluster 2 Anomaly Detection

From Figure 4, it is clear that Cluster 2 has anomalies from a total of three buildings it possesses. The isolation forest and isolation tree algorithms were repeated with the subsample size of 256 from about 1,066 points, and the number of trees was kept as 100 iterations. Figure 26 shows the anomaly ranges of the cluster and the anomaly points on the consumption pattern.

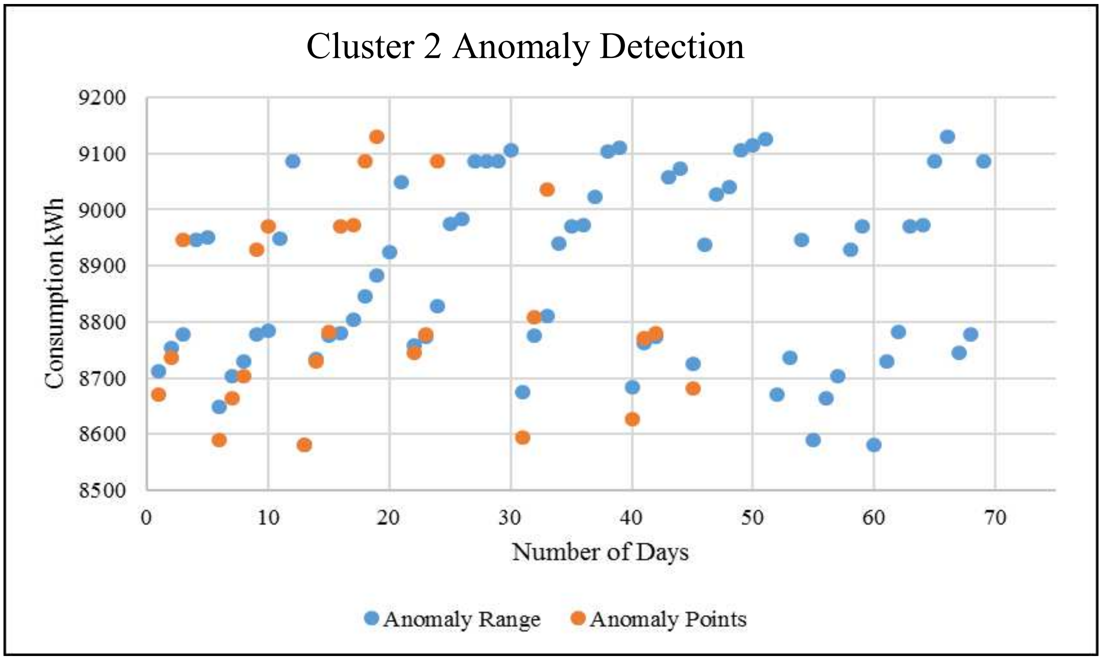


Figure 26. Cluster 2: Anomaly Detection

The buildings connected in Cluster 2 are Building 1, Building 5, and Building 9. The anomalies were all in Building 1, and Building 5 and 9 had no anomalies in this cluster.

4.6.5.4 Path Length and Anomaly Score

The average path length of the cluster anomalies was 4.8. This indicates that the anomalies were identified in this cluster after several converging iterations. Also, the average anomaly score of this building was 0.79, which was closer to 1, explaining the strength of the anomaly points.

4.6.5.5 Cluster 3 Anomaly detection

Cluster 3 contained three buildings: Building 2, Building 4, and Building 5. All of these buildings shared similar consumption patterns, and an isolation algorithm was utilized to identify the anomalies from 748 data points. Figure 27 shows the anomaly range and anomaly points based on data points from Cluster 3.

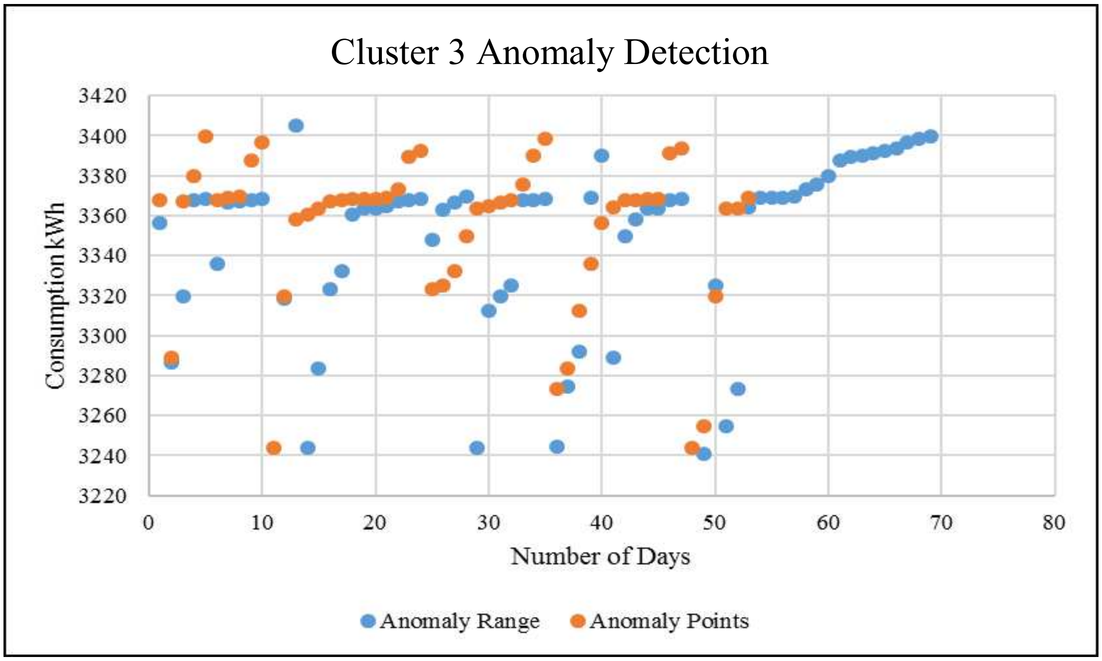


Figure 27. Cluster 3: Anomaly Detection

Of the total of 53 anomaly points, Building 5 had the most among all three buildings with 40 points, and Building 4 had 13 points. This indicates that Building 5 must have an impact on their buildings' LEED points.

4.6.5.6 Path Length and Anomaly score

The average path length of the cluster was 4.3, and the average anomaly score stayed at 0.75, which was on the higher side in terms of accuracy. Thus, the points at Building 5 need more investigation to optimize the energy performance.

4.6.5.7 Cluster 4 Anomaly Detection

The final cluster, Cluster 4, included Buildings 6 and 7 and some consumption patterns of Building 1. The cluster had a total of 747 points, out of which 256 random samples were given as input for the iForest algorithm. With the same module, 48 points of anomalies were identified in Cluster 4. Figure 28 shows the anomalous points over the range of anomalous data points.

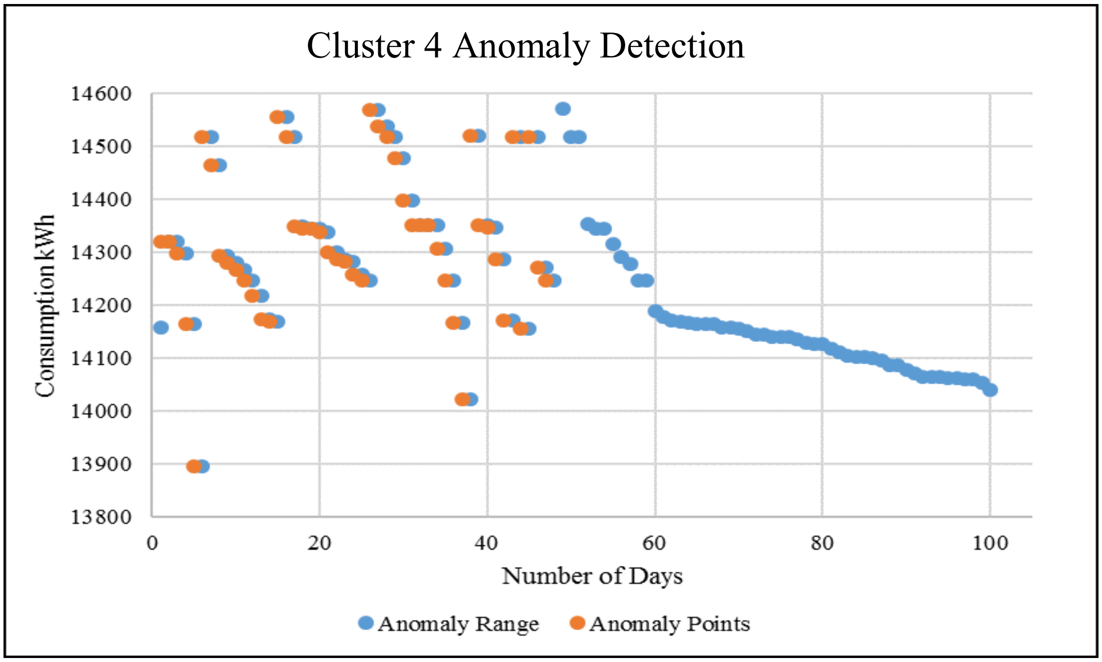


Figure 28. Cluster 4: Anomaly Detection

From the analysis, 30 anomalous points were in Building 6, and 18 points of anomalies were from Building 7. The one building in this cluster repeated from Cluster 2 was Building 1. However, any anomalies could not be found from the consumption pattern and hence, no cumulative anomaly calculation is required in this paper.

4.6.5.8 Path Length and Anomaly Score

The average path length and the average anomaly scores were 3.3 and 0.74, indicating that the degree of accuracy was good in this cluster as well. The motive to identify the path length and anomaly score was to make sure the anomaly points of the similar consumption pattern (clusters) were accurate and had no discrepancies. Thus, anomaly detection was performed using the isolation forest algorithm and isolation tree algorithm, and the results are discussed. To run the algorithms, R programming was used during the entire process with Matlab plotting syntax.

4.6.5.9 LEED Credit Evaluation

Using iForest and iTrees, the anomalies were determined from each cluster and its respective buildings. Except Building 2, all of the buildings' anomalies ranged from as low as six points to the maximum of 53 points. The table below shows all of the buildings and their LEED E/A credit scores on their subcredits. From the analysis, it is evident that Building 1 and Building 8 had the highest numbers of anomalous points with 51, followed by Building 5 with 40 points. All of these buildings are LEED gold certified and have optimized energy performance points close to 10. Table 12 below shows the buildings and their percentages of anomalous points based on their LEED credit scores.

Table 12. LEED Buildings and Anomalies

Building Number	Anomalous Data Points	Total Data Points	Percentage of Anomalies	LEED EA Score	OEP Score	Certification
1	51	365	14%	13	10	Gold
2	0	365	0%	33	7	Platinum
3	6	365	1.6%	15	10	Gold
4	13	365	3.5%	3	2	Silver
5	40	365	11%	3	2	Gold
6	30	365	8.2%	15	10	Platinum
7	18	365	5%	6	4	Gold
8	51	365	14%	7	5	Gold
9	0	365	0%	5	3	Silver

From the table, it can be inferred that the percentage of anomalies in each building ranged from 1.6% to 14%. Building 1 and 8 had similar anomaly percentages in relation to their LEED EA scores, which were 13 and 7. They are the gold certified buildings with the highest anomaly percentages when compared to the silver certified Building 4 (3.5% anomalies). Also, both the buildings scored 10 (out of 10) on the Optimized Energy Performance (OEP) credit at the time of pre-construction, which is contradictory to the results achieved by the analysis. This indicates that the trend of anomalies in these buildings has changed the optimized energy performance, and there is a requirement to re-examine the energy credits of these buildings in the post-construction phase. There is a similar case with Building 5, with 11% anomalies, but the buildings' EA and OEP credits are lower (3 and 2), making it clear regarding energy points.

The other important building is Building 6, which has 8% anomalies and is LEED platinum certified with 15 points of total EA credits and a full 10 points of OEP credits. This is again an issue when compared with Building 2, which is LEED platinum certified and has no anomalies. Also, Building 2 has 33 EA points (out of 35 in the new version), which proves the efficiency of this building.

The other inference from the table is that Building 3 is relatively efficient with fewer anomalies, which means that the energy consumption of the Building 3 does not have a lot of abnormal behaviors and has good points on both EA and OEP credits (15 and 10) respectively.

4.7 Conclusions and Discussions

Research and models have been developed to determine real-time anomalies, and alarm systems have been implemented to notify users of them. Data mining techniques of clustering and isolation were utilized as methods in the field of energy consumption analysis to quicken and improve the accuracy of anomaly detection. LEED credits are assigned to the buildings based on their energy and other credit performances immediately pre-construction phase. As discussed earlier, anomalies are abnormal behaviors of building energy consumption during different times of the year. The research involves the integration of two data mining techniques, clustering and isolation forest. Clustering was used to identify buildings with similar consumption patterns and to group them together. The isolation forest algorithm was used to identify the anomalies from these clusters and to connect these anomalies with their respective buildings. Later, this was examined with LEED EA and OEP credits of the buildings to understand the impact of the anomalies on LEED credits.

This study was done to detect the abnormalities in the buildings, which in turn could affect the energy performance of the buildings. With variations in energy performance, the credit scores can largely vary, which makes LEED certification more questionable post construction. The data used in this research were preprocessed and were of high quality, indicating that the anomalies were none other than abnormal behaviors in the energy consumption of the buildings. Thus, the research recommends regular inspection of energy performance to improve and disapprove certification for LEED buildings.

This research study was limited to nine LEED buildings due to the lack of reliable data. A future study needs to integrate the isolation forest technique as an important process of energy analysis of all buildings through which anomalies will be detected and treated, thus helping the analyst to know more about the energy performance of the building. The findings from this research make it evident that irrespective of the type of LEED certification, buildings must be evaluated for their ENERGY performance at regular intervals, and take necessary steps to maintain their credit scores.

5. RESEARCH CONCLUSIONS AND DISCUSSIONS

5.1 Summary of Results and Contributions

The main aims of this dissertation are to (a) find the relationship between building energy consumption, outside atmospheric temperature, and LEED EA credits, (b) examine the use of different green space layouts to reduce the atmospheric temperatures of high-rise buildings, and (c) use data mining techniques such as clustering, isolation, and anomaly detection to identify data anomalies of building energy consumption and examine LEED EA and OEP credits to understand the impact of the anomalies on LEED credits.

The analysis presented in Chapter 2 contributed in four ways. Based on the previous research studies, it was hard to handle the real-time building energy data and find the relationship between the use of building energy and LEED EA credits. First, data calibration and adjustment were required to clean, complement, and analyze the raw energy data. In this manner, the data analysis could improve and get better results. Second, a data mining technique (K-means clustering) was applied to examine and eliminate the data errors in an energy data set. Third, the chi-square method was utilized to verify the results of data mining and determine whether they were reasonable. Lastly, the LEED building with the lowest OEP points consumed the highest amount of energy. Therefore, this study showed that there was a relationship between building energy consumption and LEED OEP points. This study also showed that the results of the data mining technique matched those of the previous studies, which applied different methods (e.g., regression and non-regression). However, this study highlighted that calibrating energy data was a better approach to analyzing energy consumption in buildings and that the relationships between LEED OEP points and energy efficiency are not as simple as previous research studies assumed.

The analysis presented in Chapter 3 presents a method to reduce atmosphere temperature using different green space layouts. For this study, four green space layouts were selected to observe atmospheric temperature changes. Based on the results, two conclusions were found. First, the impervious surface space interspersed between green spaces was cooler even if these spaces were not large enough due to the park cool island effect and the cooling of air by green spaces. Based on the results, when green space was constructed around the impermeable pavement, the park cool island effect would act over the pavement surfaces. Thus, it was a very effective way to reduce daytime atmospheric temperature in building areas. Second, this study showed that green spaces around buildings were more effective for lowering daytime atmospheric temperatures by approximately 2 °C to 6.5 °C due to the higher shade and evaporation effects. In addition, these spaces did not affect the central hot air, which elevated temperatures in the entire building area. Thus, these spaces mitigated the urban heat island effect.

In Chapter 4, the data mining techniques of clustering and isolation were utilized as methods in the field of energy consumption analysis to quicken and improve the accuracy of anomaly detection. Buildings are assigned LEED credits based on their energy and other credit performances immediately following the preconstruction phase. The goal of this study was to detect the abnormalities in the buildings, which in turn could affect their energy performance. The credit scores can vary widely with variations in energy performance, which makes post-construction LEED certification more questionable. The data used in this research were preprocessed and high quality, indicating that the anomalies were none other than abnormal behaviors in the energy consumption of the buildings. Thus, the research recommends regular inspection of energy performance to improve and disapprove certification for LEED buildings.

5.2 Limitations of the Study and Future Research

This dissertation was composed of three topics, and each chapter provided different case studies with various factors (e.g., LEED EA Credits and OEP score, atmospheric temperature, real-time building energy consumption data, and green space) and analytical methods (e.g., K-means clustering, chi-square, isolation forest, and anomaly detection). The studies in each chapter had limited research conditions and future research studies for the following chapters.

In Chapter 2, this study investigated the effects that both endogenous variables (e.g., LEED OEP) and exogenous variables (e.g., atmospheric temperature) have on the energy usage comparison of green buildings. However, The LEED OEP scores tended to increase the energy saving potentials of the buildings, which reduced the need for renovation and maintenance. However, the analyses suggest that this cannot be verified for a one-sided approach such as energy efficiency. However, this study highlighted that calibrating energy data is a better approach to analyzing energy use in buildings and that the relationships between LEED EA credits and energy efficiency are not as simple as previous research studies assumed. Energy efficiency credits in green building standards and rating systems (e.g., LEED and International Green Construction Code) may not reduce energy use in reality.

In Chapter 3, the research condition was limited due to the different cases between the previous and current studies. The previous studies focused on low-rise buildings. Therefore, these buildings had different cooling and heating loads based on the green space layouts. However, the current study focused on high-rise buildings. Therefore, this study aims to discover means of reducing air temperatures within buildings and using green space layouts to reduce air temperatures in housing complexes. For the future research study, other variables that affect the internal temperatures of residences need to be considered for the development of various methods of temperature reduction. Accurate tools are needed to mitigate urban heat islands by quantifying the effects of such variables on temperature reduction. Further study is needed to reduce the heating effects of impervious surfaces and determine the minimum green space size that can reduce the temperature, since minimum green space in this work did not address effective size.

In Chapter 4, the research study was limited to nine LEED buildings due to the lack of reliable data. A future study needs to integrate the isolation forest technique as an important process of energy analysis for all buildings in which anomalies will be detected and treated. This would help the analyst know more about the energy performance of the building. The findings from this research make it evident that irrespective of the type of LEED certification, buildings must be evaluated for their energy performance at regular intervals and take necessary steps to maintain their LEED EA credit scores.

REFERENCES

Abe, N., Zadrozny, B., & Langford, J. (2006). Outlier detection by active learning. *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '06*, 504. http://doi.org/10.1145/1150402.1150459

Akbar, S, Khan, M. N. A., (2014). Critical Analysis of Density-based Spatial Clustering of Applications with Noise (DBSCAN) Techniques. *International Journal of Database Theory & Application*, Vol. 7 Issue 5, p17-27. 11p. doi: 10.14257/ijdta.2014.7.5.02.

Akoglu, L., Tong, H., & Koutra, D. (2015). Graph based anomaly detection and description: A survey. *Data Mining and Knowledge Discovery*, *29*(3), 626–688. http://doi.org/10.1007/s10618-014-0365-y

Anderberg, M. R. (2014). Cluster Analysis for Applications: Probability and Mathematical Statistics: A Series of Monographs and Textbooks (Vol. 19). Academic press.

Asadi, S., Amiri, S. S., & Mottahedi, M. (2014). On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design. Energy and Buildings, 85, 246-255.

Azhar, S., Brown, J., & Sattineni, A. (2010, June). A case study of building performance analyses using building information modeling. In Proceedings of the 27th international symposium on automation and robotics in construction (ISARC-27), Bratislava, Slovakia (pp. 25-27).

Aziz, A., Uyeno, H., Manni, J., Amin, M., & Staniford, S. (2015). U.S. Patent No. 9,106,694. Washington, DC: U.S. Patent and Trademark Office

Barse, E., Kvarnstrom, H. & Jonsson, E. (2003). Synthesizing Test Data for Fraud Detection Systems. Proc. of the 19th Annual Computer Security Applications Conference, 384-395.

Beatty, M. (2010, October 28). "GSA Moves to LEED Gold for All New Federal Buildings and Major Renovations". Retrieved April 13, 2016, from http://www.gsa.gov/portal/content/197325

Bessec, M., & Fouquau, J. (2008). The non-linear link between electricity consumption and temperature in Europe: a threshold panel approach. Energy Economics, 30(5), 2705-2721.

Bhuyan, M. H., Bhattacharyya, D. K., & Kalita, J. K. (2014). Network Anomaly Detection: Methods, Systems and Tools. IEEE Communications Surveys & Tutorials, 16(1), 303–336. http://doi.org/10.1109/SURV.2013.052213.00046

Bonchi, F., Giannotti, F., Mainetto, G., Pedreschi, D. (1999). A Classification-based Methodology for Planning Auditing Strategies in Fraud Detection. Proc. of SIGKDD99, 175-184.

Singh, P., & Singh, M. (2015). Fraud Detection by Monitoring Customer Behavior and Activities. International Journal of Computer Applications, 111(11).

Bongardt, B(2006), Stadtklimatologische Bedeutung Kleiner Parkanlagen - dargestellt am Beispiel des Dortmunder West Park (Urban climatic importance of small gardens on the example of Dortmund West Park), Doctor degree thesis.

Bose, S., & Diette, G. B. (2016). Health Disparities Related to Environmental Air Quality. In Health Disparities in Respiratory Medicine (pp. 41-58). Springer International Publishing.

Building Energy codes 101: An Introduction 5, U.S. Department of Energy (2010). www.ashrae.org/file%20library/docLib/Public/20100301_std901_codes_101.pdf

Carrasquilla, U. (2010). Benchmarking Algorithms for Detecting Anomalies in Large Datasets. Review Literature And Arts Of The Americas, 1–16.

Chaturvedi, S. K., Richariya, V., & Tiwari, N. (2012). Anomaly detection in network using data mining techniques. International Journal of Emerging Technology and Advanced Engineering, ISSN, 2250-2459.

Chen Yu, Wong Nyuk Hien, Zain, M.F.M. and Surod, M. (2006), Thermal benefits of city parks, Energy and Buildings Vol.38, 105-120.

Cheng, J. C., & Ma, L. J. (2015). A non-linear case-based reasoning approach for retrieval of similar cases and selection of target credits in LEED projects. Building and Environment, 93, 349-361.

Chicco, G., Napoli, R., & Piglione, F. (2006). Comparisons among clustering techniques for electricity customer classification. *IEEE Transactions on Power Systems*, *21*(2), 933–940. http://doi.org/10.1109/TPWRS.2006.873122

Cidell, J. (2009). Building Green: The Emerging Geography of LEED-Certified Buildings and Professionals. *The Professional Geographer*, *61*(2), 200–215. doi:10.1080/00330120902735932

Cole, R. J., & Kernan, P. C. (1996). Life-cycle energy use in office buildings. *Building and Environment*, *31*(4), 307–317. doi:10.1016/0360-1323(96)00017-0

Diamond, R., Opitz, M., Hicks, T., Vonneida, B., and Herrera, S. (2006). "Evaluating the site energy performance of the first generation of LEED-certified commercial buildings," published in the Proceedings of the 2006 Summer Study on Energy Efficiency in Buildings, American Council for an Energy Efficient Economy, Washington DC, August. LBNL-59853.

Diamond, R. (2011). Evaluating the Energy Performance of the First Generation of LEED-Certified Commercial Buildings.

Donovan, D.H. and Butry, D.T. (2009), The value of shade : Estimating the effect of urban trees on summertime electricity use, Energy and Buildings 41, 662-668.

Ewenz C., Bennett J., Jent C., Vinodkuma, Guan H., Clay R. (2012) Urban Heat Island Mapping in Adelaide, The 13th National Street Tree Symposium, ISBN 978-0-9805572-1-3, 1-8.

Foucquier, A., Robert, S., Suard, F., Stéphan, L., & Jay, A. (2013). State of the art in building modelling and energy performances prediction: A review. Renewable and Sustainable Energy Reviews, 23, 272-288.

Fowler, K. M., & Rauch, E. M. (2006). Sustainable building rating systems summary (No. PNNL-15858). Pacific Northwest National Laboratory (PNNL), Richland, WA (US).

Giridharan, R., Ganesa, S. and Lau, S.S.Y. (2004), Daytime urban heat island effect in high-rise and high-density residential developments in Hong Kong, Energy and Buildings Vol.36, 525-534.

Giridharan, R., Lau, S.S.Y., Ganesa, S. and Givoni, B. (2007), Urban design factors influencing heat island intensity in high-rise high-density environments of Hong Kong, Building and Environment Vol.42, 3669-3684.

Gogoi, P., Bhattacharyya, D. K., Borah, B., & Kalita, J. K. (2011). A survey of outlier detection methods in network anomaly identification. *Computer Journal*, *54*(4), 570–588. http://doi.org/10.1093/comjnl/bxr026

Gornall, J., Betts, R., Burke, E., Clark, R., Camp, J., Willett, K., & Wiltshire, A. (2010). Implications of climate change for agricultural productivity in the early twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences, 365(1554), 2973-2989.

Griffith, B., Long, N., Torcellini, P., Judkoff, R., Crawley, D., & Ryan, J. (2008). *Methodology for modeling building energy performance across the commercial sector*. National Renewable Energy Laboratory. Technical Report NREL/TP-550–41956.

Grubb, M., Rayner, S., Tanabe, A., Russell, J., Ledic, M., Mathur, A., & Brackley, P. (1991). Energy policies and the greenhouse effect: A study of national differences. Energy policy, 19(10), 911-917.

Han, J., & Kamber, M. (2006). Data Mining: Concepts and Techniques. *Annals of Physics*, *54*, 770. http://doi.org/10.5860/CHOICE.49-3305

Heidarinejad, M., Dahlhausen, M., McMahon, S., Pyke, C., & Srebric, J. (2014). Cluster analysis of simulated energy use for LEED certified US office buildings. Energy and Buildings, 85, 86-97.

Helman, P., Liepins, G., & Richards, W. (1992). Foundations of intrusion detection [computer security]. In Computer Security Foundations Workshop V, 1992. Proceedings. (pp. 114-120). IEEE.

Henley, A., & Peirson, J. (1997). Non-Linearities in Electricity Demand and Temperature: Parametric Versus Non-Parametric Methods. Oxford Bulletin of Economics and Statistics, 59(1), 149-162.

Hodge, V. J., & Austin, J. (2004). A survey of outlier detection methodologies. *Artificial Intelligence Review*. [Online]. Available at http://doi.org/10.1023/B:AIRE.0000045502.10941.a9 [Cited April 25, 2016]

Hong, S.H. (2003), Analysis of Temperature profiles by Green Area of Apartments in Gangnam, Seoul, a master's thesis in University of Seoul.

Hsu, D. (2015). Comparison of integrated clustering methods for accurate and stable prediction of building energy consumption data. Applied Energy, 160, 153-163.

Huang, Y.J, Akbari, H., Taha, H., Rosenfeld, A.H. (1987), The potential of vegetation in reducing summer cooling loads in residential building, Journal of Climate and Applied Meteorology Vol.26, 1103-1116.

Huang, Y, Miller, N., and Schlegel, N. (2009). Effects of global climate changes on building energy consumption and its implications on building energy codes and policy in California. California Energy Commission, Berkeley.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM Computing Surveys*. http://doi.org/10.1145/331499.331504

Juan, Y. K., Gao, P., & Wang, J. (2010). A hybrid decision support system for sustainable office building renovation and energy performance improvement. Energy and buildings, 42(3), 290-297.

Jusuf, S.K, Wong, N.H, Hagen, E, Anggoro, R and Hong, Y. (2007), The influence of land use on the urban heat island in Singapore, Habitat International, Vol.31, Issue 2. 232-242.

Kalvelage, K., Passe, U., Rabideau, S., & Takle, E. S. (2014). Changing climate: The effects on energy demand and human comfort. Energy and Buildings, 76, 373-380.

Kim, Y. H., & Baik, J. J. (2005). Spatial and temporal structure of the urban heat island in Seoul. Journal of Applied Meteorology, 44(5), 591-605.

Kneifel, J. (2010). Life-cycle carbon and cost analysis of energy efficiency measures in new commercial buildings. Energy and Buildings, 42(3), 333-340.

Kolokotroni, M., Ren, X., Davies, M., & Mavrogianni, A. (2012). London's urban heat island: Impact on current and future energy consumption in office buildings. Energy and buildings, 47, 302-311.

Korea Meteorological Agency. (2008). Understanding of Climate Change and Scenario Application.

Korean Institute of Construction and Technology (2008), A study about technology on developing Eco-Housing estate that countermeasures Urban Climate Changes.

Kwok, K. Y. G., Kim, J., Chong, W. K., & Ariaratnam, S. T. (2016). Structuring a Comprehensive Carbon-Emission Framework for the Whole Lifecycle of Building, Operation, and Construction. Journal of Architectural Engineering, 04016006

Lappas, T., & Pelechrinis, K. (2007). Data mining techniques for (network) intrusion detection systems. Department of Computer Science and Engineering UC Riverside, Riverside CA, 92521.

Lee, Howard B, MacQueen, James B. (1980). A K-Means cluster analysis computer program with cross-tabulations and next-nearest-neighbor analysis. *Educational and Psychological Measurement*, Vol 40(1), spr 1980. pp 133-138.

Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2010). Can Isolation-Based Anomaly Detectors Handle Arbitrary Multi-modal Patterns in Data?

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2012). Isolation-based anomaly detection. ACM Transactions on Knowledge Discovery from Data (TKDD), 6(1), 3

Mcpherson, E.G., Herrington, L.P. and Heisler, G.M. (1988), Impacts of vegetation on residential heating and cooling, Energy and Buildings Vol.12, 41-51.

Moral-Carcedo, J., & Vicens-Otero, J. (2005). Modelling the non-linear response of Spanish electricity demand to temperature variations. Energy economics, 27(3), 477-494.

NASA (2010), New Clues about the Urban Heat Island Effect at ACCUWEATHER.com.

Newsham, G. R., Mancini, S., & Birt, B. J. (2009). Do LEED-certified buildings save energy? Yes, but…. *Energy and Buildings*, *41*(8), 897–905. doi:10.1016/j.enbuild.2009.03.014

Nian, K., Zhang, H., Tayal, A., Coleman, T., & Li, Y. (2016). Auto insurance fraud detection using unsupervised spectral ranking for anomaly. The Journal of Finance and Data Science.

Oliveria, S., Andrade, H. and Vaz, T. (2011), The cooling effect of green saces as a contribution to the mitigation of urban heat: a case study in Lisbon. Building and Environment, 46 : 2186-2194.

Onishi, A. et al(2010), Evaluating the potential for urban heat island mitigation bu greening parking lots, Urban Forestry & Urban Greening 9, 323-332.

Pachauri, R. K., Allen, M. R., Barros, V. R., Broome, J., Cramer, W., Christ, R., ... & van Vuuren, D. (2014). Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.

Pérez-Lombard, L., Ortiz, J., & Pout, C. (2008). A review on buildings energy consumption information. *Energy and Buildings*, *40*(3), 394–398. doi:10.1016/j.enbuild.2007.03.007

Petcharat, S., Chungpaibulpatana, S., & Rakkwamsuk, P. (2012). Assessment of potential energy saving using cluster analysis: A case study of lighting systems in buildings. Energy and Buildings, 52, 145-152.

Richter, B., D. Goldston, G. Crabtree, L. Glicksman, D. Goldstein, D. Greene, D. Kammen, M. Levine, M. Lubell, M. Savitz, D. sperling, F. Schlachter, J. Scofield, and J. Dawson. (2008). Energy Future: Think Efficiency. http://www.aps.org/energyefficiencyreport/index.cfm. College Park, MD: American Physical Soceity.

Rodrigues, F., Duarte, J., Figueiredo, V., Vale, Z., & Cordeiro, M. (2003). A Comparative Analysis of Clustering Algorithms Applied to Load Profiling. In *Machine Learning and Data Mining in Pattern Recognition* (pp. 73–85). http://doi.org/10.1007/3-540-45065-3_7

Rosenthal, J. K., Crauderueff, R., & Carte, M. (2008). Urban Heat Island Mitigation Can Improve New York City's Environment: Research on the Impacts of Mitigation Strategies on the Urban Environment.

Sailor, D. J. (2001). Relating residential and commercial sector electricity loads to climate—evaluating state level sensitivities and vulnerabilities. Energy, 26(7), 645-657.

Salmon, W., Turner, C., Analyst, S., & Owens, B. (2008). Energy Performance of LEED ® for New Construction Buildings.

Santamouris, M. (2014). Cooling the cities–a review of reflective and green roof mitigation technologies to fight heat island and improve comfort in urban environments. Solar Energy, 103, 682-703.

Santamouris, M., Cartalis, C., Synnefa, A., & Kolokotsa, D. (2015). On the impact of urban heat island and global warming on the power demand and electricity consumption of buildings—a review. Energy and Buildings, 98, 119-124.

Sapienza, A., Panisson, A., Wu, J., Gauvin, L., & Cattuto, C. (2015). Anomaly Detection in Temporal Graph Data: An Iterative Tensor Decomposition and Masking Approach. Proceedings of AALTD. Portugal, 117.

Sastre, M. (2003). The history of the urban heat island effect in New York City. Unpublished master's thesis, Columbia University Graduate School of Architecture, Planning and Preservation, Avery Library, New York.

Sartori, I., & Hestnes, a. G. (2007). Energy use in the life cycle of conventional and low-energy buildings: A review article. *Energy and Buildings*, *39*(3), 249–257. doi:10.1016/j.enbuild.2006.07.001

Sawka, M., Millward, A. A., Mckay, J., & Sarkovich, M. (2013). Growing summer energy conservation through residential tree planting. Landscape and Urban Planning, 113, 1-9.

Scofield, J. H. (2008). A Re-examination of the NBI LEED Building Energy Consumption Study, 764–777.

Scofield, J. H. (2009). Do LEED-certified buildings save energy? Not really…. *Energy and Buildings*, *41*(12), 1386–1390. doi:10.1016/j.enbuild.2009.08.006

Shashua-Ba, L. and Hoffaman, M.E. (2000), Vegetation as a climatic component in the design of an urban street : An empirical model for preditiong the cooling effect of urban green area with trees, Energy and Buildings Vol.31, 221-235.

Simpson, J.R. and Mcpherson, E.G. (1996), Potential of tree shade for reducing residential energy use in California, Journal of Arboriculture 22(1), 10-18.

Simpson, J.R. (2002), Improved estimates of tree-shade effects on residential energy usse, Energy and Buildings 34, 1067-1076.

Sinha, V. (2009). Critics say new green rating hurts affordable housing. *EWashington Business Journal 23* May. http://www.bizjournals.com/washington/stories/2008/05/26/story2.htm

Sun, S., & Wang, Y. (2009, March). A weighted support vector clustering algorithm and its application in network intrusion detection. In Education Technology and Computer Science, 2009. ETCS'09. First International Workshop on (Vol. 1, pp. 352-355). IEEE.

Taib, N. (2010), An Assessment of Thermal Comfort and Users' Perceptions of Landscape Gardens in a High-rise Office Building, Journal of Sustainable Development Vol.3, No.4, 153-164.

The Ministry of Environment in Korea (2008), Analysis of Climate Change in Korea and East Asia Area and Study of the Atmosphere and Ocean Effects.

The National Renewable Energy Laboratory, 1995, Landscaping for Energy Efficiency, Department of Energy, United State of America.

Ting, K. M. (2009). Title : Adaptive Anomaly Detection using Isolation Forest, (December).

Torcellini, P., Pless, S., Deru, M., & Crawley, D. (2006). Zero energy buildings: a critical look at the definition. National Renewable Energy Laboratory and Department of Energy, US.

Turner C. and M. Frankel (2008). Energy Performance of LEED for New Construction Buildings- Final Report. *New Buildings Institute*, White Salmon, VA

UCI. (2015). *UCI Machine learning Repository*. . [Online]. Available at http://archive.ics.uci.edu/ml [Cited April 25, 2016].

UNEP (2009). United States Green Building Council, Buildings and climate change: a summary for decision-makers. United Nations Environmental Programme, Sustainable Buildings and Climate Initiative, Paris, 1-62.

U.S. Energy Information Administration - EIA - Independent Statistics and Analysis. (2016).
[Online]. Available at
http://www.eia.gov/electricity/monthly/epm_table_grapher.cfm?t=epmt_1_01
[Cited May 31, 2016].

USGBC (2003a). United States Green Building Council, Leadership in energy and environmental design. [Online]. Available at http://www.usgbc.org/leed/leed_main.asp [Cited April 27, 2016].

USGBC (2003b). United States Green Building Council, Why build green? [Online] Available at http://www.usgbc.org/AboutUs/whybuildgreen.asp [Cited April 27, 2016].

USGBC (2004). United States Green Building Council, An introduction to the United States Green Building Council and the LEED Green Building Rating System [Online] Available at https://www.usgbc.org/Docs/Resources/usgbc_intro.ppt#27 [Cited April 27, 2016].

U.S. Green Building Council. (2007). LEED Initiatives in Government by Type [Online]. Available at https://www.usgbc.org/ShowFile.aspx? DocumentID=1741 [Cited April 27, 2016].

U.S. Green Building Council (2015), "Green Building Facts," http://www.usgbc.org/articles/green-building-facts.

U.S. Department of Energy (2011), "Buildings Energy Data Book," U.S. Department of Energy,
Washington DC.

U.S. Energy Information Administration (2013). Independent Statistics and Analysis. [Online] Available at http://www.eia.gov/consumption/commercial/reports/2012/buildstock [Cited April 25. 2016]

U.S. Energy Information Administration (2016). Independent Statistics and Analysis. . [Online] Available at https://www.eia.gov/consumption/commercial/reports/2012/energyusage [Cited April 27. 2016]

Van Vlasselaer, V., Bravo, C., Caelen, O., Eliassi-Rad, T., Akoglu, L., Snoeck, M., & Baesens, B. (2015). APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions. Decision Support Systems, 75, 38-48.

Vengertsev, D., Thakkar, H., & Science, C. (2005). Anomaly Detection in Graph : Unsupervised Learning , Graph-based Features and Deep Architecture.

Voogt, J. A. (2002). Urban heat island. In T. Munn (Series Editor) & I. Douglas (Vol. Ed.) Encyclopedia of global environmental change: Vol 3. Causes and consequences of global environmental change (pp. 660-666). Chichester, UK, John Wiley & Sons.

Wagner, I., Krauze, K., Zalewski, M., (2013). Blue aspects of green infrastructure. Sustainable Development Applications, 4, pp. 144–155.

Walker. L.R. and Newman, S.E. (2009), Landscaping for Energy Conservation, Colorado State University.

Wan, K. K., Li, D. H., Pan, W., & Lam, J. C. (2012). Impact of climate change on building energy use in different climate zones and mitigation and adaptation implications. Applied Energy, 97, 274-282.

Wang, H., & Chen, Q. (2014). Impact of climate change heating and cooling energy use in buildings in the United States. Energy and Buildings, 82, 428-436.

Wedding, G. Christopher, and Douglas Crawford-Brown (2008). "Improving the link between the LEED green building label and a building's energy-related environmental metrics." Journal of Green Building 3.2 (2008): 85-105.

Whang, H.K., Lim, J.Y., Song, D.S., Kim, T.Y. (2008), Characteristics Analysis of Thermal Environment in Apartment Complexes in Summer, Korea Institute of Architectural Sustainable Environment and Building System, Autumn Symposium Vol.10, 138-142.

Whang, J.W. and Kim, S.J. (2003), A Comparative Analysis on the Urban Heat Island by Residential Types, Journal of Korea Planners Association, Vol.38 No. 7, 235-244.

Wolf, K. L. (2004). Trees, parking and green law: Strategies for sustainability. Stone Mountain, GA: Georgia Forestry Commission, Urban and Community Forestry.

Wu, P., Mao, C., Wang, J., Song, Y., & Wang, X. (2016). A decade review of the credits obtained by LEED v2. 2 certified green building projects. Building and Environment, 102, 167-178.

Yamanishi, K., Takeuchi, J. I., Williams, G., & Milne, P. (2004). On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms. Data Mining and Knowledge Discovery, 8(3), 275–300. http://doi.org/10.1023/B:DAMI.0000023676.72185.7c

Yang, F, Stephen, S.Y. Lau and Feng Qian (2010), Summertime heat island intensities in three high-rise housing quarters in inner-city Shanghai China: Building layout, density and greenery, Building and Environment 45, 115-134.

Yu, F. W., Chan, K. T., & Sit, R. K. (2012). Climatic influence on the design and operation of chiller systems serving office buildings in a subtropical climate. Energy and Buildings, 55, 500-507.

Yih, W. T., Goodman, J., & Carvalho, V. R. (2006, May). Finding advertising keywords on web pages. In Proceedings of the 15th international conference on World Wide Web (pp. 213-222). ACM.

APPENDIX A

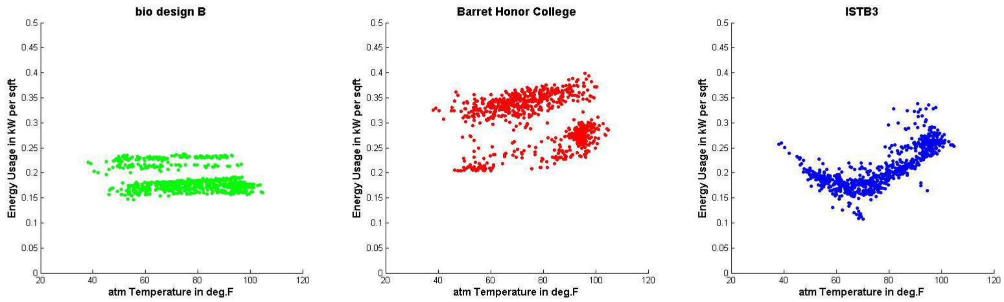
CHAPTER 2: SAMPLES OF PLOTS BASED ON BUILDING ENERGY USAGE DATA

A.1. Weather Data Sample

Time Info	Low Temp (F*)	Mean Temp (F*)	High Temp (F*)	Precip. (in)	Mean Wind Speed (mph)
9/1/2012	84	95.2	105.1	0	6.33
9/2/2012	88	96.6	107.1	0	5.41
9/3/2012	88	95.4	107.1	0	5.87
9/4/2012	80.1	93.1	104	0	9.55
9/5/2012	81	90.5	100.9	0	10.93
9/6/2012	82.9	93	102	0	8.17
9/7/2012	72	83.5	102	0	7.6
9/8/2012	73	84	93.9	0.51	6.33
9/9/2012	79	86.8	93.9	0	11.85
9/10/2012	75.9	84.9	98.1	0.03	5.87
9/11/2012	72	80.1	98.1	0.03	6.21
9/12/2012	73	81	93.9	0.02	2.99
9/13/2012	75	79	100	0	5.18
9/14/2012	79	88.5	100	0	11.16
9/15/2012	78.1	85.8	98.1	0	8.98
9/16/2012	73	86.1	100	0	4.26
9/17/2012	73.9	87.5	100	0	5.06
9/18/2012	77	89.2	102	0	4.49
9/19/2012	78.1	91	105.1	0	4.49
9/20/2012	79	91.2	105.1	0	4.26
9/21/2012	78.1	91.1	105.1	0	5.18
9/22/2012	78.1	91.4	106	0	3.8
9/23/2012	79	92.2	106	0	4.83
9/24/2012	79	89.9	102.9	0	5.06
9/25/2012	75.9	88.9	100.9	0	7.83
9/26/2012	73.9	85.6	99	0	6.44
9/27/2012	73.9	85.4	97	0	4.6
9/28/2012	75	85.6	98.1	0	4.49
9/29/2012	75	86.6	98.1	0	5.52
9/30/2012	75	87.6	102.9	0	4.14
. .					
8/29/2014	82.9	96.3	108	0	5.87
8/30/2014	82.9	96.1	111	0	6.44

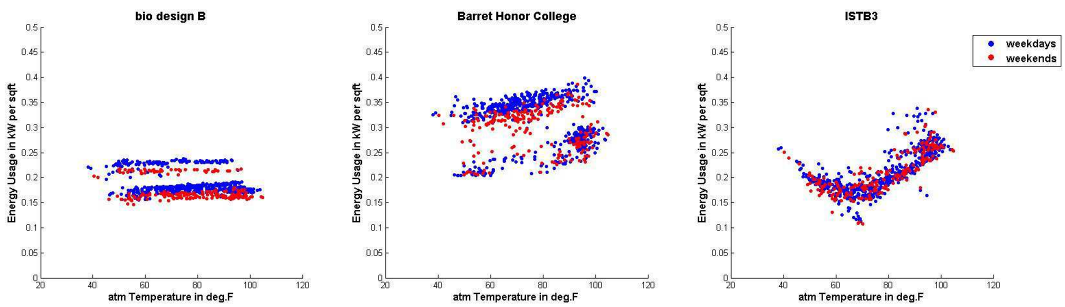
8/31/2014	82.9	97.3	111	0	6.79

A.2. Temperature vs Energy Usage for three buildings (Raw Data)

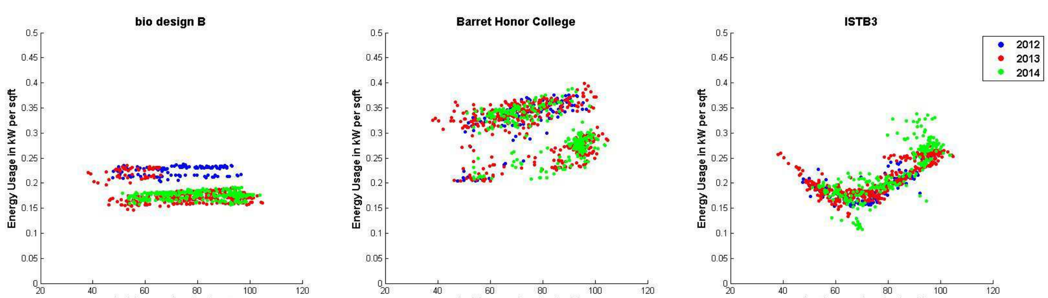


A.3. Temperature vs Energy Usage for three buildings (Separated by weekdays and weekends)

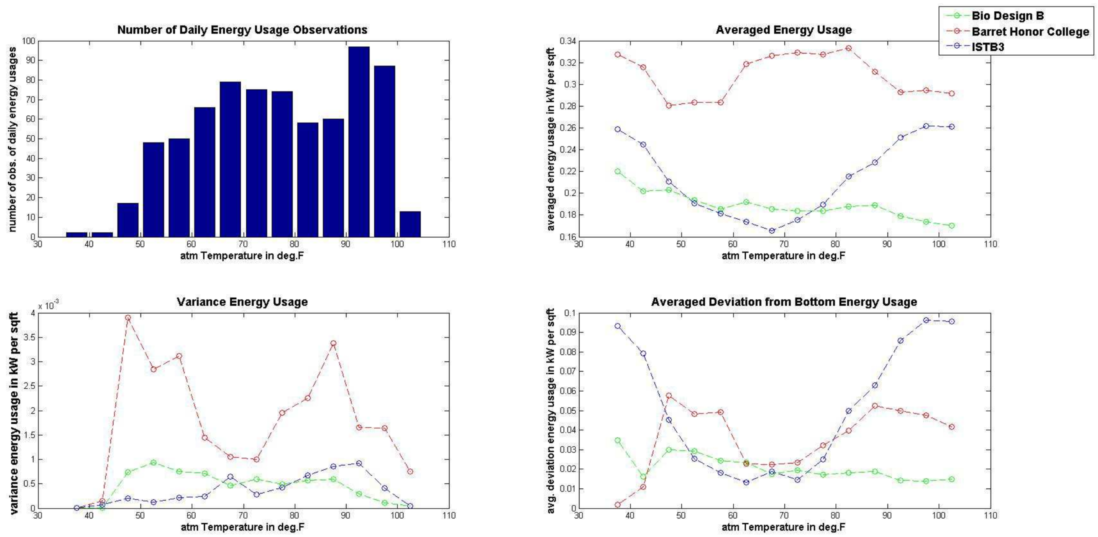
.



A.4. Temperature vs Energy Usage for three buildings (Separated by year-2012 to 2014)



A.5. Deviation of Temperature vs Energy Usage



APPENDIX B

CHAPTER 3: SAMPLE OF WEATHER DATA

B.1. Weather Hourly Data