Guiding building owners towards energy conservation measures

Sean Kuusinen (EBIO)

April 9, 2014

Honors Thesis Committee Members
Dr. Carol Wessman (Advisor, EBIO)
Dr. Gregor Henze (CEAE)
Dr. Barbara Demmig-Adams (EBIO)

for my parents
&
grandpa jack

Acknowledgments

A huge thank you is owed to Ecosystems for providing full and unrestricted access to their data—without it this project would have certainly been infeasible. Thank you to Henry Mueller who provided the original idea and the subsequent guidance through the world of building systems and retrofits. My thesis advisor, Carol Wessman, has been instrumental in keeping me focused and on track as well as providing feedback throughout the writing process—thank you Carol. I am greatly appreciative to my committee members Gregor Henze and Barbara Demmig-Adams for contributing their time, resources, and feedback. Finally, a big thank you goes to the Undergraduate Research Opportunity Program for providing funding.

Abstract

Energy conservation measures have been identified as the most cost effective way to reduce carbon emissions. However, a lack of available information regarding energy conservation prevents building owners from investing in energy efficiency. This research provides the groundwork for supplying building owners with a simple model to guide retrofit decisions in their buildings. Using readily available characteristics of buildings that have received a lighting retrofit, the achieved reduction in energy consumption was analyzed using a classification and regression tree. This statistical method determines which building attributes are most related to reduction in energy use. The results of this process show that simple building attributes, including square footage, business type, and vintage, are only responsible for a small portion of the variation in energy-use reduction following energy conservation measures. However, the classification and regression tree and the random forest methods provide insight into how building attributes can be used to explain energy reduction following a lighting retrofit. With more attributes, these simple visual tools may show how energy efficiency analysts can communicate potential savings to building owners, reducing costs to owners and carbon emissions in the atmosphere.

Table of Contents

Introduction

As global greenhouse gas emissions increase and climate change accelerates, humanity has approached a crossroads in terms of how it chooses to use energy (Anderson 2013). The 4.86 million commercial buildings in the United States, totaling 7.16 billion square feet of floor space, offer a tremendous opportunity to reduce carbon emissions (EIA 2003). However, practical and financial barriers often prevent building owners from investing in energy conservation measures (Prindle and Fontane 2009).

Current methods for evaluating energy conservation measures (ECMs) in buildings are not easily scalable, meaning not efficient or practical when applied to a large data set or large numbers of buildings (Yeonsook 2011). A new model for ECMs has been introduced, but the extensive modeling and statistical knowledge required presents another type of barrier that is likely to discourage building owners from pursuing energy conservation (Yeonsook 2011). Additionally, complex energy efficiency modeling tools have been shown to be overall less accurate than simple models (Earth Advantage Institute 2009). A simple tool that relies on easily-attained building characteristics may allow building owners to understand the potential savings that result from ECMs. This would remove some of the major barriers to energy efficiency, leading to increased investment and reduced carbon emissions. This research seeks to understand if simple building characteristics can be used to evaluate the potential savings resulting from ECMs, specifically:

Question: Can simple building attributes be used to classify the energy use reduction as the result of energy conservation measures, and can this be used to create a simple visual model to aid decision makers?

A classification and regression tree (CART) approach is used to explain the variation in percent reduction in energy use by splitting the data into homogeneous groups and using combinations of explanatory variables, in this case building attributes. This approach identifies those building characteristics that can best explain energy savings. This model can then be used to evaluate potential savings in buildings based solely on their attributes.

Background

Due to moral, political, environmental, and economic implications, anthropogenic, or human-caused, climate change is likely to be the foremost issue of the 21st Century. Earth will likely become uninhabitable for humans and other large life forms if the total carbon emitted into the atmosphere reaches a critical point, somewhere within the range of 5,000 to 10,000 gigatons (Hansen et al. 2013). There is at least three times that much carbon available to humans in the form of unburned fossil fuels (Hansen et al. 2013). In order to avoid a catastrophic increase in temperature, the rate at which carbon is released into the atmosphere will need to be drastically slowed (Beinhocker et al. 2008).

Commercial buildings represent one of the largest opportunities to reduce carbon emissions. In the U.S. alone, over two billion tons of CO_2 are emitted from commercial buildings annually, which accounts for nearly 20% of the country's emissions (2010 Buildings Energy Data Book). An analysis performed on 643 commercial buildings showed a median energy savings of 16% after implementing energy conservation measures (ECMs) (Mills 2004). Energy consumption in buildings is expected to rise 1.7% per year until 2025, while total floor area will increase 1-2% annually (Ryan 2004). Accounting for external costs such as those resulting from increased ocean acidification and decreased biodiversity, energy efficiency investments yield more energy per dollar than investments in fossil fuels (Lovins 1997). However, policies and investments that reduce carbon emissions are unlikely if lawmakers, business owners, and voters continue to view such actions as barriers to economic growth (Prindle and Fontane 2009).

Market barriers to energy efficiency

In order to grow the economy while reducing emissions, the U.S.'s *gross domestic product (GDP) per ton of CO_2 emissions*, or carbon intensity, needs to increase by ten times in the next forty years (Beinhocker et al. 2008; U.S. Energy-Related Carbon Dioxide Emissions 2012). Attaining a tenfold increase in efficiency has been recognized as viable by companies as big as General Electric (Hawken et al. 1999). Many of the technologies that will enable a radical increase in efficiency already exist—light emitting diodes (LEDs) use about one-fifth the energy of conventional incandescent lightbulbs (Hawken et al. 1999).

However, the American Council for an Energy Efficient Economy determined that the lack of information regarding efficiency opportunities is the largest reason that these projects are overlooked by business owners (Prindle and Fontane 2009). Business owners often forego investing in energy conservation measures (ECMs) for three reasons. First, most businesses have pressing needs requiring capital investment (Prindle and Fontane 2009). Second, given the unknown return of an energy upgrade, building owners often view ECMs as a risky investment (Yeonsook 2011). For example, a business owner may opt for the guaranteed return of adding another staff member to the sales team over a retrofit with an unknown return on investment. Third, for a business owner looking to reduce costs, the rate of return for most energy conservation projects is too small to justify the investment (Schendler 2012).

Smaller building owners face additional barriers. Over 80% of commercial buildings in the United States are less than 50,000 square feet (Prindle and Fontane 2009). Due to financial constraints businesses of this size are unlikely to hire an engineer or consultant to evaluate energy savings because the cost of hiring an engineer is often greater than the savings that result from the ECMs (Prindle and Fontane 2009). It is estimated that by investing in ECMs, building owners in the United States could save a total of $30 billion by 2030 and keep billions of tons of CO_2 out of the atmosphere (Mills 2009). This is according to an analysis performed on 643 commercial buildings which showed an annual median 16% energy savings per building (Mills 2004, 2009).

Peripheral benefits of energy efficiency investments

In addition to cost savings, energy conservation helps companies to connect with consumers and employees who increasingly put a premium on environmental and social responsibility (*Engagement 2.0* 2010). At the turn of the 21st Century, only a dozen companies had reported their corporate social responsibility (CSR) efforts using the international standard for reporting provided by the Global Reporting Initiative (GRI). In 2012, a little over a decade later, over 3,500 companies reported their CSR efforts using the GRI standards (GRI Report 2013). Cumulatively, nearly 6,000 companies have submitted reports at some point in the last twelve years (GRI Report 2013).

There are certainly more companies that emphasize CSR. Some companies such as JBS, the world's largest meat processor, report to the Carbon Disclosure Project, a reporting format that emphasizes climate change (JBS CDP Report 2013). Other companies such as Dole, the world's largest banana producer, have chosen to publish independent yet comprehensive CSR reports (Luske 2012). While it is impossible to know exactly how many companies have CSR campaigns, a study published in 2010 suggests that 81% of companies have at least some CSR information on their websites (CSR Trends 2012). Often, these campaigns address carbon emissions reduction either directly by setting a reduction target or indirectly by pledging to reduce emissions when possible (CSR Trends 2012). These numbers illustrate the fact that corporate social responsibility has become essential to allow businesses big and small to remain competitive in a world that increasingly demands responsibility.

Political support for Energy Conservation Measures

The business case continues to drive investments in energy conservation. In addition, during the last decade politicians and governments have increasingly shown support for energy conservation measures (ECMs). In 2007, U.S. President George W. Bush signed the Energy Independence and Security Act. Among the many provisions in the act was the requirement that all new Federal Buildings are to be fossil-fuel free by 2030 (U.S. House Bill H.R. 6 2007). Further legislation required the phasing-out of relatively inefficient light fixtures. In 2011, U.S. President Barack Obama initiated the “Better Buildings Initiative” to reduce energy consumption by 20% in all commercial buildings by 2020 (White House papers 2011). Cities like Chicago have targeted as much as 50% reduction in commercial building energy use (City of Chicago Climate Action 2011). Internationally, there is support for ECMs as well. The Intergovernmental Panel on Climate Change showed that energy efficiency is the most cost effective way to reduce carbon emissions (IPCC 2013). The political support for energy conservation measures reinforces the environmental, economic, and consumer pressures to increase energy efficiency.

Current methods for evaluating energy efficiency

Currently, many energy efficiency analysts employ a “best-guess” method in smaller buildings to determine opportunities for energy efficiency. Energy efficiency experts routinely assume that there are opportunities for efficiency improvements as technology has gotten better and buildings are often outdated. While this is quite effective for lighting, it does not reveal less conspicuous opportunities in HVAC systems, windows and insulation that are impossible to evaluate at-a-glance.

In bigger buildings, energy use is benchmarked or recorded at a regular, baseline level and then compared to similar buildings across a portfolio (Yeonsook 2011). Candidate ECMs are then evaluated using transient simulation models to compare the relative benefits of a set of ECMs (Yeonsook 2011). This strategy is not scalable, or not efficient or practical when applied to a large data set or large numbers of buildings, and does not include a risk analysis (Booth 2012). Additionally, there is evidence that some complex models have a lower accuracy than simple models (Earth Advantage Institute 2009).

Yoensook 2011 introduced a new methodology for ECM decision making that supports large-scale retrofit decisions. This model defines energy flows within a building with a small set of parameters, and then calibrates and quantifies uncertain parameters in the model, a likelihood function then measures how closely parameter values match observations, which can support a risk-based decision making process (Yeonsook 2011). However, the extensive modeling and statistical knowledge required for this approach presents another type of barrier that is likely to discourage building owners from pursuing ECMs. A new methodology would ideally allow building owners to understand the amount potential energy reduction in their building based only on a few simple building attributes.

Methods

Two models were developed for the present analysis. The first model was created using the means of whole-building energy consumption for buildings in the Mountain West range, which includes Arizona, Colorado, Idaho, Montana, Wyoming, New Mexico and Nevada (EIA 2003).

The second model used empirical energy consumption data from buildings that received energy conservation measures, specifically lighting retrofits. The latter model used several building attributes to find how the building attributes influence the percent reduction in energy use following a retrofit.

Initial Model – Whole building energy consumption by square footage

The initial model correlated expected energy use intensity, expressed as kbtus/sqft/year, similar to the measure of automobile energy efficiency, to the size of the building for different building types. This showed the expected energy consumption for each building type based on its size. This model was based on median energy use intensity for buildings throughout the Rocky Mountains based on data from websites such as eSource, the Department of Energy, and the Environmental Protection Agency. Energy use intensity was used to determine kilowatt hours consumed per month per square foot. Then, using data from multiple studies that found percentage of energy used for cooling, lighting, and miscellaneous loads in each building type, baseline energy use was calculated (U.S. Climate Action Report 2010 & Hojjaht and Michaels 2004). Assuming that most buildings in the Rocky Mountain zone use electricity for cooling and gas for heating, the amount of electricity used in each season was extrapolated into these categories: hot (Jun-Aug), cool (Dec-Feb), and swing (Mar-May & Sep-Nov). These data were then visualized to allow building owners to understand the potential for energy use reduction based on how their building's energy performance compares to that of similar buildings.

Second Model – Reduction in energy use following a retrofit

The second model used a different dataset from the previous model. Empirical data were collected from sixteen buildings around the Front Range area in Colorado that received lighting retrofits. A Classification And Regression Tree (CART) was used to analyze whether the variation in the reduction of energy use after a lighting retrofit can be explained with the attributes of the building. In other words, are the simple attributes of a building correlated to the amount of energy saved following a lighting retrofit? If so, building owners could use this model to understand the energy efficiency reduction potential in their buildings.

In order to determine which building characteristics influence the reduction in energy use following a retrofit, the CART method was used. The resulting model categorizes buildings based on how their attributes affect energy reduction following a retrofit. After the tree is built using the data, the tree is simplified which produces a streamlined model without sacrificing accuracy (Reddy 2010).

Data

The buildings used in the analysis conformed to the following parameters:

1. Located in The Front Range of Colorado (Denver, Boulder and associated suburbs)
2. Commercial building
3. Received a lighting retrofit within a 2 year timeframe
4. Energy consumption data from utility provider available *before* and *after* the retrofit

Climate affects electricity load because it is used to cool buildings. Fortunately, climate is generally consistent throughout the Front Range, with little variation in annual average temperature and precipitation (Figure 1).

Figure 1: Average monthly temperatures for Boulder (left) and Denver (right). Source: http://www.weather.com/weather/wxclimatology/monthly/graph/USCO0105

The lighting retrofits performed in the various buildings in the sample varied. In most cases, compact fluorescent lighting (CFL) was updated with more efficient technology. In some cases, light emitting diodes (LEDs) replaced either incandescent lighting or CFLs.

Monthly whole-building energy consumption (WBEC) date for each building were obtained directly from the utility provider, usually Xcel Energy or Platte River Power Authority, for the two years preceding the retrofit and the year following.

Additionally, the following data were acquired for each building:

1. Business type occupying the building (i.e. office, retail, worship)
2. Square footage
3. Vintage, or year built

Generally, these data are available publicly on the county assessor's website. When the building square footage was ambiguous, which was the case for five of the buildings, it was confirmed either using documentation from an analyst who had visited the building or by measuring the size of the building using satellite images (Google Earth Pro). Percent decrease in energy usage following lighting retrofit was calculated for each building using whole building energy consumption data.

All buildings included here were used for commercial purposes. The buildings were classified based on use (office, retail, and house of worship) since the type of equipment and amount of energy used in a building often depends on the type of business occupying the space. Additionally, the hours of operation varied based on the business type. Retail businesses often stay open later during the peak season in December while offices are closed. Classifying buildings into business types served to control for this variation.

Classification and Regression Tree

The classification and regression tree (CART) method was used to determine how building attributes affect percent reduction in energy use. In the CART process buildings were split, or classified, into distinct groups with other like buildings based on the attributes that best split the data set into homogenous groups. Then, each group was assigned a percent reduction that best fits it. The model is then simplified and the output was a decision, similar to a flow chart that shows percent reduction expected based on the attributes of a building. Finally, the CART was tested on a new dataset to see how accurately it classified data. This resulted in a rate of error, or a misclassification rate.

In the present analysis, two CART models were created. The first CART was built with all sixteen data points, which means this model has no misclassification rate. The second CART was built with eleven data points and then cross-validated with the remaining five data points in order to find a misclassification rate.

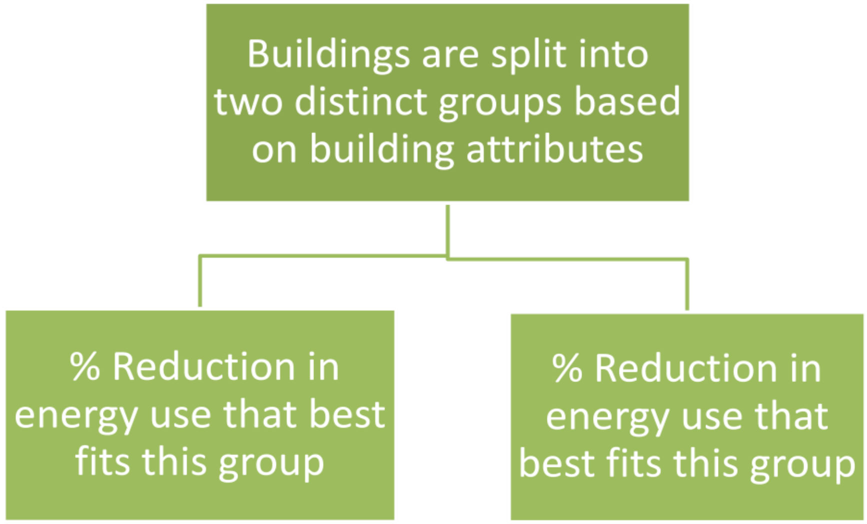

Figure 2: A conceptual model of a classification and regression tree. The process finds the building attribute that best splits buildings into distinct groups. Then, each group is assigned a response variable that best describes the group.

A CART explains the variation of a response variable, in this case the percent reduction in energy use, by splitting the data into homogeneous groups using combinations of explanatory variables, here the building's attributes (Breiman et al. 1984). This is a simple way of visualizing a multiple linear regression, as CART creates a tree that is a visual representation of the classification and regression procedure (Ostendorp 2011). This occurs in two steps. First, the tree must be built through a process that classifies the data based on its values and the sum of squares between classifications (Fridely 2010). The second step, pruning, serves to reduce the complexity of the tree, leaving a model that is the most effective while still being simple (Reddy 2010). The CART algorithm executes a comprehensive search to build a parsimonious tree, or one with the fewest number of branches possible (Reddy 2011). CART models are also non-parametric, meaning they do not rely on assumptions about the data's distribution (Reddy 2011).

The first step of the process uses part of the data as a training dataset to build the tree by testing the explanatory variables on the response variable (Reddy 2011). Each node constitutes a binary decision, such as square footage above and below a certain size, and the tree cascades as such to the end nodes (Ostendorp et al. 2011). The end nodes represent the tree's response to a set of inputs (Ostendorp et al. 2011). At each node, the data are grouped based on how an explanatory variable, in this case the building attributes, best splits the data into homogenous groups (Ostendorp 2011). An ideal split separates the dataset into classes based on homogeneity with regard to one variable, and this process continues until all end nodes contain points of uniform class (Ostendorp 2011).

Once the tree is built, each end group on the tree has a typical response value based on how the explanatory variables split (Reddy 2010). This provides an idea of which attributes most affect the amount of energy reduction following a retrofit. However, often this means that each node only contains one data point, a circumstance called "over-fitting" (Ostendorp 2011). In this case, the tree must be "back pruned," or simplified to reduce the number of splits leaving only the ones that most affect the response variable (Ostendorp 2011). Finally, the CART must be tested, or cross-validated, with a dataset not included in the original model to determine the misclassification rate, or how often the model incorrectly classifies buildings. In this research, two CARTs were created. The first uses all sixteen data points to maximize the sample space, leaving no cross validation data set. In the second CART, eleven data points were used to build the tree, and then the remaining five were used for cross validation.

Back Pruning and Cross Validation

The second step of the CART process reduces the number of splits and leaves only the most influential explanatory variables (Ostendorp 2011). Generally, the tree is back-pruned to the first node that is within one standard of error of the minimum R_α value, which balances the misclassification error rate with a cost penalty for complexity (Breiman et al. 1984). This finds the most parsimonious, or simplest tree, for a minimum R_α value (Ostendorp 2011). In other words, the remaining tree balances the best configuration with a minimized misclassification rate, and also gives realistic misclassification rates of the final tree (Reddy 2011). Often, back pruning can reduce the number of splits by up to 80% (Ostendorp 2011). Since so few buildings were included in the model, back pruning was not necessary, and when applied to the CART resulted in only a single node.

Cross-validation refers to the process of finding how accurately the model classifies new data points. By using a subset of data not included while building the tree, a misclassification rate, $R\varphi$, can be found. This serves to find how well the model can classify completely new data points. Cross-validation was performed on the second CART. The entire CART process is detailed in Appendix A.

Random Forest

The random forest partitioning approach uses a large set of classification trees to find which explanatory variables most affect the response variable (Shih 2011). Random forests randomly test many different sub samples of predictors (i.e. building attributes) at each split on the tree (Strobl et al. 2009). Because of this, random forests can more accurately determine how a predictor affects the response (Strobl et al. 2009). In addition, this approach ranks the explanatory variables in order of how strongly they affect the response (Shih 2011). The results of a random forest have been show to more accurately classify data than a single classification tree and is particularly useful for small datasets (Strobl et al. 2008).

In order to find the effect of the explanatory variables on the response, the random forest algorithm randomly shuffles the values of the response and explanatory variables, thereby breaking any correlations inherent in the data set (Shih 2011). The difference in model accuracy before and after randomization is and averaged over all trees, telling us how important the predictor is for the outcome (Strobl et al. 2009). It is calculated by permuting, or changing, the combination and magnitude of each explanatory variable and applying it to each split on each tree (Shih 2011). Then, the difference between residual sum of squares before and after each split is calculated, which is then summed over all splits of all trees for that variable and normalized by the standard deviation (Shih 2011). This final number is node impurity, or the measure of a variable's influence on the response variable (Shih 2011).

In this research, the random forest method was executed once with 500 CARTs created in the single run. From these CARTs node impurity, or how much each explanatory variable affects the response, was calculated.

Results

Initial Model – Whole building energy consumption by square footage

The initial model was created using data from public sources. Table 1 shows all the publicly available Energy Use Intensity (EUI) data that were available online. This model demonstrates the potential for energy conservation measures (ECMs) in commercial buildings (Figure 3).

Table 1: Energy Use Intensity (kbtu/sq.ft/year) throughout the Rocky Mountains for different building types, Q indicates that data was not available (EIA, eSource, EPA).

Total West 2003	83
Education	78
Food Sales	Q
Food Service	244
Health Care	180
Lodging	104
Mercantile and Service (Retail)	85
Office	72
Public Assembly	91
Public Order and Safety	Q
Religious Worship	28
Warehouse and Storage	39
Other	Q
Vacant	Q

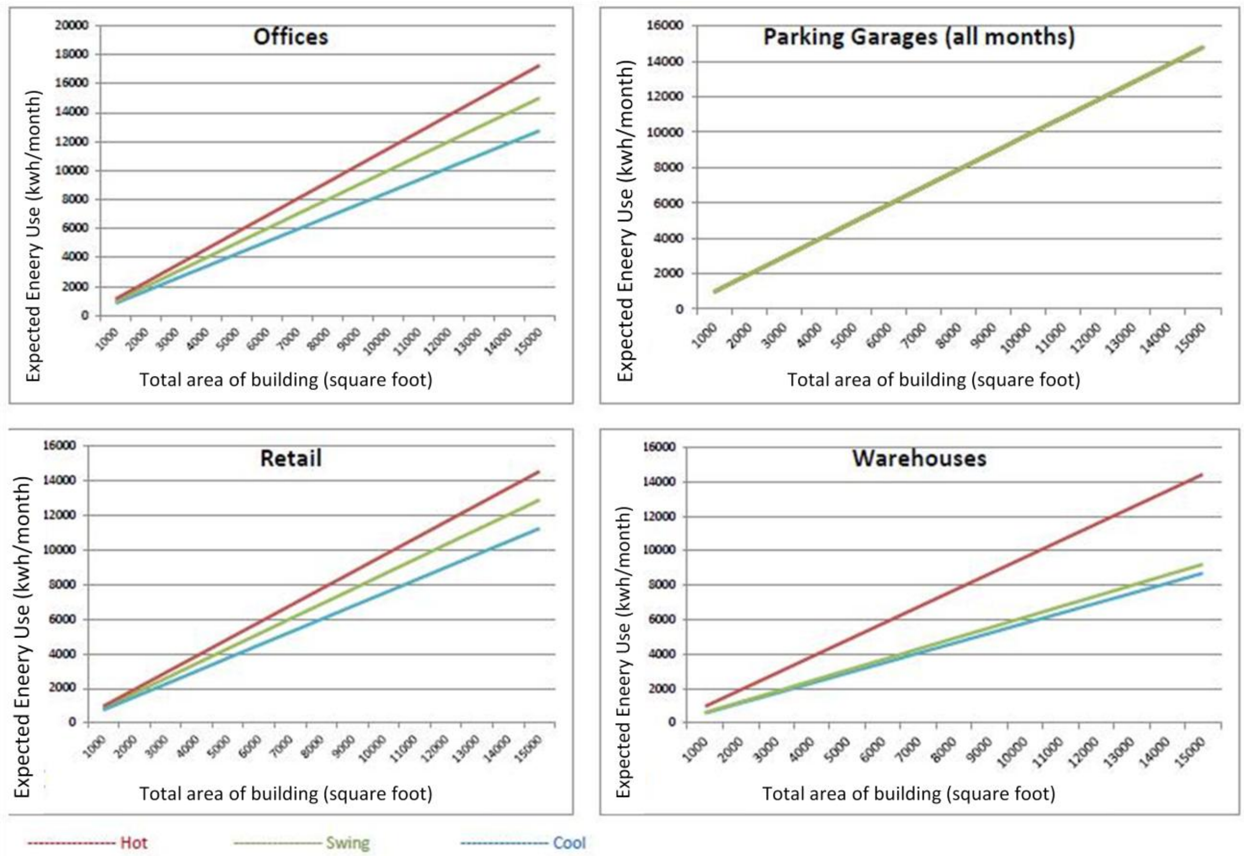

Figure 3: The initial model based on data collected from online sources (DOE, EPA, eSource). The graphs project the expected mean energy use (in kWh) per square foot for (A) offices, (B) parking garages, (C) retail, and (D) warehouses in the Rocky Mountains. Each color represents a different season—blue: cool (Dec-Feb), green: swing (Mar-May & Sept-Nov), and red: hot (Jun-Aug). This allows a building in the Rocky Mountain region to be compared to others, showing the amount of potential savings that may result from ECMs.

Second Model – Classification and Regression Tree

The second model, based on empirical data, should present building owners with the opportunities for energy savings from a different perspective. The model uses both qualitative and quantitative building attributes and shows their correlation with the percent reduction in energy use following a lighting retrofit. There was a high level of variation in the percent reduction of whole-building energy use. Further analyses, such as the coefficient of determination (R^2) were done to begin to pinpoint which building parameters may explain the variation. The attributes of the sixteen buildings included in the model are summarized in Table 2.

Table 2: Summary statistics of the 16 buildings included in the model. Square footage includes any basement or garage space in the building. Vintage refers to the year in which construction was completed. Percent reduction in whole building energy consumption (WBEC) is the difference in the average whole building electricity energy consumption before and after the lighting retrofit for the same months of the year (i.e. March through November both before and after the retrofit). The p-value is presented for the percent reduction in WBEC—all differences were significant except in two cases, which are in bold. The bottom of the table provides a complete summary of each variable including minimum, maximum, range, mean, median, and standard deviation.

Use type	Sq footage	Vintage	% Reduction of WBEC	p-value of % reduction of WBEC
Office	2777	1910	37.95	.014
Office	6708	1972	18.05	.008
Office	36942	1981	29	.0001
Office	1727	1977	40.8	.0002
Office	7063	1961	17.3	.012
Office	6519	1977	23.3	.014
Retail	1275	1948	17.6	.0005
Retail	2165	1976	15.5	.0004
Retail	502	1981	47.9	3.68e-6
Retail	10000	1964	14.7	.008
Retail	938	1971	18.5	**.064**
Worship	10192	1957	-3	**.40**
Worship	14079	1963	12.2	.0002
Worship	12816	1947	10.25	.017
Retail	13121	1986	7.9	.011
Office	2320	1951	25	.021
–				
Mean	8072	1964	20.81	
Median	6614	1968	17.83	
Std Dev	9025	19	13.01	
Min	502	1910	-3.00	
Max	36942	1986	47.90	
Range	36440	76	50.90	

The initial tree was built using all sixteen buildings to reduce the effect of outliers on the tree, or in other words using the maximum sample space possible (Figure 4). This tree identifies square footage as the most influential variable, with three different classes of size. Cross-validation was not performed on the first tree, as the entire dataset was used to create it, and hence a misclassification rate cannot be calculated. In other words, there is no confidence rate for this tree.

In the second tree, two-thirds of the data were used to build the tree and the remaining third was use to determine the misclassification rate. The latter tree has a low misclassification rate ($R\varphi = .05$), meaning only 5% of the validation dataset was misclassified (Figure 5). The latter tree supports the first tree since both have square footage as the only influential variable and show percent reduction. Neither of the trees responded to the back pruning algorithm, likely because the dataset was too small to create a circumstance of over fitting, or having one data point at each end node.

The random forest algorithm shows that the explanatory variables, including square footage, vintage, and building type, explain 14.1% of the variation in the percent reduction in energy use. The random forest algorithm shows that square footage is the most influential variable for on the latter 14.1% of variance in percent reduction, followed by vintage and then building use type (Figure 6).

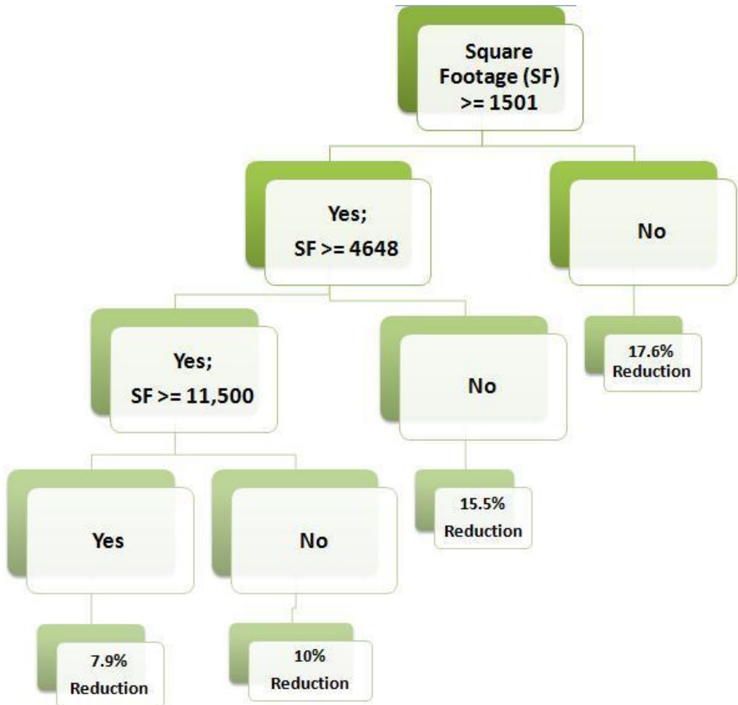

Figure 4: Classification and regression tree using all available data points. Back pruning was not viable as it removed all nodes other than the root. The misclassification (Rφ) rate was not discoverable because the entire data set was used to build the tree, thus cross validation was not possible.

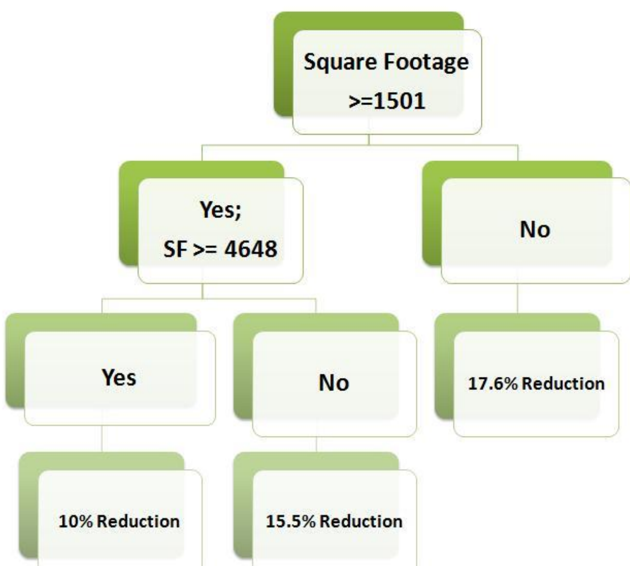

Figure 5: A classification tree built using eleven data points, then cross-validated with the remaining five points. The misclassification rate (Rφ) is 0 on the right most and middle end nodes, and Rφ = .05 on the left end node. This tree has similar classifications and response variables to the first tree that used the entire sample space.

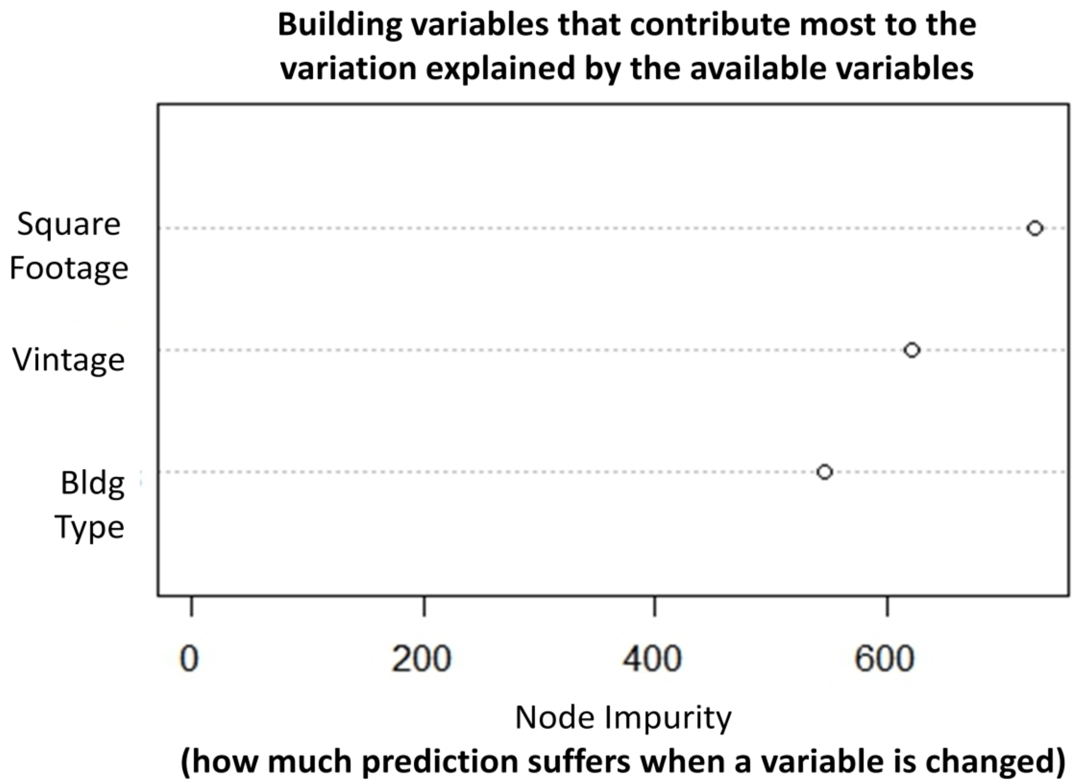

Figure 6: Use of the Random Forest algorithm to reveal building attributes that most affect the response variable. In this case, square footage has the highest influence, followed by vintage then building type. Node impurity measures how much each explanatory variable affects the response variable by finding how the predictive power of the model changes as the variable changes. It is calculated by permuting, or changing, the combination and magnitude of each explanatory variable and applying it to each split on each tree. Then, the difference between residual sum of squares before and after each split is calculated. This is summed over all splits for that variable over all trees to get node impurity.

Discussion

The initial model presented here aims to help building owners to see how their building compares to others, and therefore the amount of energy and money that could be saved as the result of ECMs. The present study shows that the simple building attributes, including square footage, vintage, and business type, only explained 14.1% of the variation in percent reduction in energy-use following a retrofit. Square footage is the building attribute most responsible for variation in percent reduction energy use that can be explained by the variables available here.

The CART and random forest models serve to showcase these simple models which can be used to help building owners understand the potential energy savings in their buildings. With more building attributes than presented here, these methods may be useful to convey potential savings following a retrofit.

The classification and regression tree model, which has a relatively low misclassification rate (Rφ), shows that out of the available variables, square footage is the building attribute that can be used to most confidently explain percent reduction in energy use following a retrofit. This is confirmed by the random forest procedure, in which square footage was found as the most influential variable. However, the amount of variation in the energy use reduction explained by the building attributes available here is low, only 14.1%. This is consistent with other correlative models that only use building characteristics to explain the variation in energy use, which explain 11.9 to 14.9% of variation (Guerra et al. 2009). In one study, 71% of the unexplained variation was due to occupancy behavior, a metric not included in the present model (Sonderegger 1978). The simple models reported here thus cannot be used with confidence to show business owners exactly how much they will save with a retrofit, but further research accounting for additional characteristics and behaviors may be able to do so.

The CART and Random Forest processes allow a visual representation of how building attributes may affect reduction in energy consumption following a retrofit. These methods may be used to simplify the visualization of potential energy savings in commercial buildings. Simple models have the potential to allow business owners to easily understand the potential energy savings in their buildings using easily measured attributes of their buildings. Moreover, a recent study found that simpler energy models tend to be more accurate than complex ones (Earth Advantage Institute 2009). A comparison of error rates for simple versus complex energy models shows that the simplest model has the lowest error rate (Earth Advantage Institute 2009). By focusing on the most influential drivers of energy use in buildings, simple models can capture more of the variation in energy reduction follow a retrofit. This highlights the importance of focusing on the most influential drivers, such as occupancy. The methods presented in this study may be used to find those most influential variables. Using variables such as envelope type, window to wall ratio, number of occupants, occupancy behavior and so forth may lead to a better understanding of which variables most affect energy use.

More energy saved in smaller buildings

The CART and the random forest algorithms show that square footage has the most influence on percent reduction in energy use following a retrofit (14.1%) of the available variables. In addition, building size in this data set exhibits an inverse relationship with percent energy reduction following a retrofit—smaller buildings show a larger percent reduction in energy use. This suggests that smaller buildings somehow respond to the reduction in lighting more strongly. In general, lighting retrofits reduce the cooling load because the newer technologies burn at a lower temperature than older ones. Perhaps the reduction in cooling load is higher per square foot in smaller buildings because these buildings are more sensitive to internal temperature changes. In some instances, smaller buildings have been shown to have a higher Energy Use Intensity (EUI), or energy consumed per square foot (Chung et al. 2006). This might suggest that smaller buildings have more potential to reduce energy use per square foot than larger buildings, which should be viewed as good news for small buildings since it may boost the smaller return on investment (ROI) that typifies these projects. On the other hand, this is not an argument against retrofitting bigger buildings, as these large lighting retrofits benefit from leveraging the fixed costs of the retrofit by replacing many lights.

In contrast, other data show that building size usually has a linear relationship with thermal capacity, or the amount of heat stored within a building (Antonopoulos and Koronaki 1998). Further research on the correlation between building size and percent energy reduction would resolve this discrepancy—either the inverse relationship seen in the present study is a new finding or an anomaly in the data due to small sample size.

Systems Approach to Energy Conservation Measures

A systems approach to energy conservation occurs when lucrative efficiency projects are bundled those that have a slower payback, which leads to significant emissions reductions and rates of return that satisfy the shrewdest financial officers (Schendler 2012). Of the energy conservation measures regularly addressed, including HVAC systems, windows, insulation, and equipment, lighting retrofits are the simplest and provide the quickest payback, which makes them the "lowest hanging fruit" (Hawken et al. 1999; Schendler 2012). Unfortunately, picking only the lowest fruit leaves the less lucrative but still carbon-reducing efficiency projects untouched; a circumstance termed "cream skimming" (Schendler 2012). The reason for cream skimming includes many of the economic barriers discussed above, such as capital restraints and high return on investment (ROI) thresholds (Howarth and Andersson 1992). The systems approach is a potential solution to cream skimming. In order to achieve a higher ROI, high return projects, such as a lucrative lighting retrofit, can be bundled with those that have a lower return, such as a new economizer for an HVAC system. The CART and random forest methods may be applied to these scenarios to find which combination of retrofits and building attributes maximizes the return on investment following a retrofit.

Vintage and building type

Additionally, the vintage of building construction appears to have an impact on the reduction in energy use. This is primarily due to the effect of age on the materials and the advances in the efficiency of building materials and building technologies (Guerra et al. 2009). For example, advances in insulation technologies have resulted in energy savings of as much as 30% in some cases (Taylor and Lucas 2010). Ever-changing building codes also affect the energy efficiency of a building (Laustsen 2008). Over time, stricter building codes have led to higher standards of energy efficiency, which helps explain why vintage is an important variable (Laustsen 2008).

Business type occupying the building had the smallest influence, which is surprising considering that the type of lighting retrofit often depends on building type. Office buildings usually have compact fluorescent (CFL) tubular fixtures, while retail stores often feature track lighting with incandescent bulbs. It is common for offices to upgrade to newer CFL technology while retail stores upgrade to more efficient LEDs. Additionally, occupancy tends to vary between business type, and occupancy is an influential driver of energy consumption (Chung et al. 2006). Despite these trends, business type was not measured as affecting percent reduction in energy use following a retrofit.

Climate change and building temperature

By 2050, Colorado temperatures are expected to rise significantly, bringing temperatures typical of the Kansas border to the Front Range (Ray et al. 2008). This affects whole building energy consumption, as warmer temperatures will lead to increased use of air conditioning (Sailor and Pavlova 2003). Generally, an increase in external temperature has a linear correlation with internal temperature, with variations depending on building specifications (Coley and Kershaw 2010). Models predict that with each degree Celsius of warming of outside air, there will be a 2-4% increase in electricity demand, and hence lighting retrofits will be especially important to reduce the cooling load on buildings (Sailor et al. 2003). A standard incandescent light reaches temperatures around 170°C, almost four times hotter than an LED (Crawford 2014). Replacing incandescent and fluorescent lighting in buildings throughout Colorado will significantly reduce the cooling load due to lighting, helping to offset increasing outdoor temperatures. It will also help offset the increased cost of cooling in commercial buildings.

Limitations and Further Research

Further work with this type of model will benefit from a few important factors, as there are clear limitations of this research. First, this model suffered from a small sample size of sixteen buildings. Additionally, there are many building attributes that were ignored to keep the model as simple as possible. Future studies may address other building attributes, including but not limited to: building materials, building envelope, window-to-wall ratio, surface area-to-volume ratio, aspect, microclimate, heating ventilation and air conditioning unit specifications, occupancy behavior, lighting retrofit treatment, window quality, insulation type, r-factor of building materials, surrounding vegetation, wind exposure, damper operation, and mixed mode window operation, among others. Building attributes subtly affect energy use (Guerra et al. 2009). The classification and regression tree approach along with the random forest approach can be used to find which attributes explain the variation of reduction in energy use following a retrofit.

Additionally, the present model includes only three years of utility data for each building, in most cases only one year after the retrofit. In order to control for yearly variations (i.e. one particularly hot summer), future models would benefit from more years of data.

Conclusions

Energy conservation measures mitigate carbon emissions while allowing economies to operate business as usual in a changing climate. This will be an integral part of the effort to slow climate change, however the market barriers that prevent building owners from investing in energy conservation measures need to be circumvented. The CART and Random Forest methods were used to show that square footage, vintage and building type explain 14.1% of variation in percent energy reduction following a retrofit. Using the same statistical processes with more building attributes will allow analysts to determine which variables most affect percent reduction in energy use following a retrofit, which will lead to better models that more confidently communicate potential savings to owners. Providing easily accessible and understandable tools to building owners will help them do so, ultimately enabling commerce to contribute to the reduction of global carbon emissions.

Appendix A: R Code

```
#Part 1: Set up

#Load the RPART package
library(rpart)

#Begin by importing the data
data = read.csv(file.choose())

#Divide data into a training and cross-validation set
train = data[1:12,]  #Training set
valid = data[13:16,] #Cross validation set

#Part 2: Building a CART

#Set seed
set.seed(2)

#View data
head(data)
xtabs ( ~ reduc, data=data)
colnames(data)
View(data)

#Develop a CART of this data
default = rpart.control(minbucket=3, cp=0.05, xval=3) #Set up some default options so we get as
deep a tree as possible
tree = rpart(reduc ~ sf + age + class, data = train , method = "class", control = default)

#Plot tree results
plot(tree,main="CART",branch=0.75,margin=0.1)
text(tree,cex=0.9,use.n=TRUE)

#This shows how large a tree is needed
plotcp(tree)

#Now make some predictions with this tree on the validation set
preds = predict(tree,newdata=valid,type="class")
preds = as.numeric(preds) - 1

#Find misclassification rate
errors = as.numeric(preds != valid)
Rt = sum(errors)/length(errors)
```

```
errors10 = as.numeric(preds[valid==0] == 1)
errors01 = as.numeric(preds[valid==1] == 0)
R10 = sum(errors10)/length(errors)
R01 = sum(errors01)/length(errors)

#Show the misclassification rates
print(Rt)
print(R10)
print(R01)

#Part 3: Random Forest

library(randomForest)

# Run the random forest the algorithm
fit = randomForest( reduc ~ sf + age + class, data = data, mtry=1, ntree=500)

# Show results
print(fit)

# Textual representation the importance of each variable in classifying reduction
importance(fit)

# Visual representation of importance of each variable in classifying reduction
  varImpPlot(fit, main="Most Important Variables")
```

References

Anderson, Kevin. "Climate Change: Going beyond Dangerous." Interview. *Slideshare.net*. Tyndall Centre, University of Manchester. (July 2011).

Antonopoulos, K.A., and E. Koronaki. "Apparent and Effective Thermal Capacitance of Buildings." *Energy* 23.3 : 183-92 . (1998).

Beinhocker, Eric, Jeremey Oppenheim, Ben Irons, Makreeta Lahti, Diana Farrell, Scott Nyquist, Jaana Remes, Tomas Naucler, and Per-Anders Enkvist. "The Carbon Productivity Challenge: Curbing Climate Change and Sustaining Economic Growth."*Mckinsey Global Institute* (2008).

Booth, A.t., R. Choudhary, and D.j. Spiegelhalter. "Handling Uncertainty in Housing Stock Models." Building and Environment 48 (2012): 35-47

Buildings Energy Data Book 2010. Department of Energy & Energy Information Agency 2010. Web. 19 Nov. (2013).

Chung, William, Y. V. Hui, and Y. Miu Lam. "Benchmarking the Energy Efficiency of Commercial Buildings." *Applied Energy* 83: 1-14 . (2006).

City of Chicago Climate Action. Chicago Climate Action Plan. (2010). Available from http://www.chicagoclimateaction.org/e

Climate Change 2013. The Physical Basis. Rep. 5th ed. Stockholm, Sweden: Intergovernmental Panel on Climate Change. (2013).

Coley, David, and Tristan Kershaw. "Changes in Internal Temperatures within the Built Environment as a Response to a Changing Climate." *Building and Environment* 45.1: 89-93. (2010).

Colorado Climate Report 2008. Ray, Andrea J,' Joseph J. Barsugli, Kristen B. Averyt. A Report for the Colorado Conservation Board. (2008).

Crawford, Trent. LED vs CFL vs Incandescent Light Bulbs. Sewell Development Corporation. (2014).

CSR Trends 2012. Rep. Craib Design & Communications and PriceWaterHouseCoopers LLP, (2010).

Earth Advantage Institute and Conservation Services Group. *Energy Performance Score Report: Findings & Recommendations Report*. Issue brief. Vol. 1. : Energy Trust of Oregon. (2009).

EIA. 2003 Commercial Buildings Energy Consumption Survey. U.S. Energy Information Administration, Washington, DC. (2003). Available from http://www.eia.gov/emeu/cbecs/contents.html

Engagement 2.0. Rep. Aon Hewitt. (2010). Available from http://www.aon.com/attachments/engagement_2%200_oct2010.pdf

Greening, Lorna A., David L. Greene, and Carmen Difiglio. "Energy Efficiency and Consumption — the Rebound Effect — a Survey." *Energy Policy* 28.6-7 (2000): 389-401.

GRI Report List 1999-2013. Rep. Global Reporting Initiative, 18 Sept. (2013).

Santin, Olivia Guerra, Laure Itard, and Henk Visscher. "The Effect of Occupancy and Building Characteristics on Energy Use for Space and Water Heating in Dutch Residential Stock." *Energy and Buildings* 41.11: 1223-232. (2009).

Hansen, James, Makiko Sato, Gary Russell, and Pushker Kharecha. "Climate Sensitivity, Sea Level, and Atmospheric CO2." *NASA Goddard Institute for Space Studies*. (2013).

Hawken, Paul, Amory B. Lovins, and L. Hunter Lovins. *Natural Capitalism: Creating the next Industrial Revolution*. Boston: Brown and Little. (1999).

Heo, Yeonsook. "Bayesian Calibration of Building Energy Models for Energy Retrofit Decision-Making Under Uncertainty." *Georgia Institute of Technology*. (2011).

Hojjati, Behjat, and Joellee Michaels. *U.S. Commercial Buildings Energy Intensity Using Site Energy by Census Region and Principal Building Activity 1992-2003*. U.S. Energy Information Agency. (Dec. 2004).

Howarth, Richard B., and Bo Andersson. "Market Barriers to Energy Efficiency." *Energy Economics* 15.4 (1993): 262-72.

Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990 - 2011. Rep. Washington, DC: U.S. Environmental Protection Agency. (2013).

JBS S/A Climate Change Investor Report. Carbon Disclosure Project 2013. Web. 25 Sept. 2013.

Johnson, Laurie T., Starla Yeh, and Chris Hope. "The Social Cost of Carbon: Implications for Modernizing Our Electricity System." *Journal of Environmental Studies and Sciences* (2013).

Kondratenko, I., Clarke, J. A., Johnstone, C. M. and Lever, M. *Using Simulation to Formulate Domestic Sector Upgrading Strategies for Scotland.* Energy and Buildings Magazine , Elsevier Publishers. pp. 759-770. (2004).

Laustsen, Jens. *Energy Efficiency Requirements in Building Codes, Energy Efficiency Policies for New Buildings*. Paris: International Energy Agency. (2008).

Lovins, Amory B., and L. Hunter Lovins. "Climate: Making Sense and Making Money." *Rocky Mountain Institute* (1997). *RMI Library*. Web. 16 Sept. 2013.

Leo Breiman. Classification and regression trees. Group. (1984).

Luske, Boki. *Comprehensive Carbon Footprint Assessment Dole Bananas*. Rep. Soil and More International. (Mar. 2012).

May-Ostendorp, Peter, Gregor P. Henze, Charles D. Corbin, Balaji Rajagopalan, and Clemens Flesmann. "Model-predictive Control of Mixed-mode Buildings with Rule Extraction." *Building and Environment* 46 (2011): 428-37.

Mills, E. “Building commissioning: a golden opportunity for reducing energy costs and greenhouse gas emissions.” *LBNL Report*. Lawrence Berkeley National Laboratory. (2009).

Mills, E., Friedman, H., Powell, T., Claridge, D., Haasl, T., and Piette, M. A. “The cost-effectiveness of commercial-buildings commissioning: a meta-analysis of energy and non-energy impacts in existing buildings and new construction in the United States.” *LBNL Report-56637*. Lawrence Berkeley National Laboratory. (2004).

Prindle, William, and Andre De Fontane. "A Survey of Corporate Energy Efficiency Strategies." *American Council for an Energy Efficient Economy* (2009): 5-77--89.

Reddy, T. Agami. *Applied Data Analysis and Modeling for Energy Engineers and Scientists*. New York: Springer (2011).

Roth, K. W., Westphalen, D., Feng, M. Y., Llana, P., and Quartararo, L. “Energy impact of commercial building controls and performance diagnostics: market characterization, energy impact of building faults and energy savings potential.” *Report prepared for U.S. Department of Energy*. TIAX LCC, Cambridge, MA. (2005).

Ryan, J. D., and Nicholls, A. “Commercial building R&D program multi-year planning: opportunities and challenges.” *Proceedings of the ACEEE Summer Study on Energy Efficiency in Buildings*. American Council for an Energy-Efficiency Economy, Washington, DC. (2004).

Rysanek, Adam, and Ruchi Choudhary. "Multi-Period Optimization of Building Refurbishment Decisions: Assessing Options and Risk Under Economic Uncertainty." *Proceedings of Buildings Simulation* (2011).

Sailor, D. J., and A. A. Pavlova. "Air Conditioning Market Saturation and Long-term Response of Residential Cooling Energy Demand to Climate Change." *Energy* 28.9: : 941-51. (2003).

Satchwell, A., Goldman, C., Larsen, P., Gilligan, D., and Singer, T. "A survey of the U. S. ESCO industry: market growth and development from 2008 to 2011." *Technical Report*, Lawrence Berkeley National Laboratory. (2010).

Schendler, Auden. *Getting Green Done: Hard Truths from the Front Lines of the Sustainability Revolution*. New York: PublicAffairs, (2009).

Schendler, Auden. "Priming the Pump for Emissions Reduction." *Journal of Industrial Ecology* 10.4 (2006): 8-11.

Schendler, Auden. "Rotten Fruit: Why 'Picking Low-Hanging Fruit' Hurts Efficiency And How To Fix The Problem." *EDC Magazine* (5 Nov. 2012).

Shih, Stephanie. "Random Forests for Classification Trees and Categorical Dependent Variables: An Informal Quick Start R Guide." Stanford University & University of California, Berkeley (2011).

Sonderegger, RC. "Movers and stayers: the resident's contribution to variation across houses in energy consumption for space heating." *Energy and Building* 1: 313-324. (1977–1978).

Strobl, Carolin; Anne-Laure Boulesteix; Thomas Kneib; Thomas Augustin; and Achim Zeileis. Conditional variable importance for random forests. *BMC Bioinformatics*. 9: 307. (2008).

Strobl, Carolin; James Malley; and Gerhard Tutz. An Introduction to Recursive Partitioning: Rational, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests. *Psychological Methods*. 14(4): 323-348. (2009).

Taylor, ZT; RG Lucas. "An estimate of Residential Energy Savings from IECC Change Proposals Recommended for Approval at ICC's Fall, 2009 Initial Actions Hearing." PNNL 19367. (2010).

"Tools, Research Services, and Advice for Utilities and Corporations." *Tools, Research Services, and Advice for Utilities and Corporations | E Source*. N.d.

U.S. Climate Action Report 2010. Rep. 5th ed. Washington, DC: United States Department of State (2010).

U.S.A. Environmental Protection Agency. Energy Star. *Energy Star Building Upgrade Manual*. U.S. EPA. (Oct. 2007).

U.S. Energy-Related Carbon Dioxide Emissions 2012. Energy Information Agency. (Oct. 2013).

U.S. House. 110th Congress, 1st Session. *H.R.6, Energy Independence and Security Act.* Washington: Government Printing Office. (2007).

White House "President Obama's plan to win the future by making American businesses more energy efficient through the "Better Buildings Initiative". (2011). Available from http://www.whitehouse.gov/the-press-office/2011/02/03/president-obama-s-plan-win-future-making-american-businesses-more-energy