

Utility Scale Building Energy Modeling and Climate Impacts

A Dissertation Presented for the
Doctor of Philosophy
Degree

The University of Tennessee, Knoxville

Brett Bass

May 2021

© by Brett Bass, 2021
All Rights Reserved.

Abstract

Energy consumption is steadily increasing year over year in the United States (US). Climate change and anthropogenically forced shifts in weather have a significant impact on energy use as well as the resilience of the built environment and the electric grid. With buildings accounting for about 40% of total energy use in the US, building energy modeling (BEM) at a large scale is critical. This work advances that effort in a number of ways. First, current BEM approaches, their ability to scale to large geographical areas, and global climate models are reviewed. Next, a methodology for large-scale BEM is illustrated, displaying its capability to create a digital twin of a utility service area consisting of more than 178,000 electrical meters in and around Chattanooga, Tennessee. This urban BEM (UBEM) framework is unique in its ability to scale beyond localized tax assessor data, which can be a limiting factor in the size of UBEM analyses. A partnership was formed with a Chattanooga electrical utility to use real 15-minute electricity data to assign building parameters and empirically validate the models. Several analyses were performed on the buildings in the service area, including simulating several building technologies and climate change resilience. After the utility-scale analysis, the scope was broadened to the entire US. A method was created by which climate models can be used to project building energy use for all commercial buildings in the US through 2100 using a floor-area scaling technique. US building energy climate research to this point has either been localized to individual building types in specific regions of the country or has evaluated energy use across the US as a whole. With simulated error rates of less than 4% compared to commercial building energy survey data, this bottom-up method can be used to effectively forecast building energy related to climate change. The utility scale UBEM framework was also expanded to model every building in the US individually. A modeling effort of this size has never been done on an individual building basis (more

than 125 million buildings). The methodology can show that US nation-scale analyses can be accomplished using high-performance computing (HPC) resources and can be used as a baseline for UBEM researchers in the future while the models can be used for simulation-informed analysis across the country.

Table of Contents

1	Introduction	1
1.1	Motivation	1
1.2	Research Contribution	1
1.3	Building Energy Modeling	3
1.3.1	Physical Building Energy Modeling	3
1.3.2	Data-Based Building Energy Modeling	5
1.4	Urban Building Energy Modeling	5
1.4.1	Bottom-up Urban Building Energy Modeling	5
1.4.2	Top-down Urban Building Energy Modeling	9
1.5	Building Energy Modeling Uncertainty	10
1.6	Urban Building Energy Modeling Limitations	13
1.7	Climate Modeling	14
2	Methods	16
2.1	Virtual Electric Power Board (Chattanooga)	16
2.1.1	Data Aggregation	17
2.2	Building Type Classification	18
2.2.1	Missing Values	22
2.2.2	Euclidean Distance	22
2.2.3	Dynamic Time Warping	23
2.2.4	Machine Learning	23
2.3	Software	25
2.4	Model Evaluation	28

2.4.1	3D Visualization	28
2.5	Energy Conservation Measures	28
2.5.1	Energy Efficiency	29
2.5.2	Demand Side Management	29
2.5.3	Emissions	29
2.6	Climate Impacts	31
2.6.1	Virtual EPB Climate Impacts	33
2.6.2	US Commercial Buildings Climate Impacts	35
2.7	Model America	38
3	Results and Discussion	42
3.1	Virtual EPB	42
3.1.1	Building Type Assignment	42
3.1.2	Energy Conservation Measures	48
3.1.3	3D Visualization	58
3.1.4	Virtual EPB Climate Impacts	63
3.2	US Commercial Buildings Climate Impacts	67
3.2.1	Commercial Floorspace Extrapolation	69
3.2.2	Weather files	69
3.2.3	Constant Floorspace	71
3.2.4	Traditional Urban Growth	76
3.3	Model America	78
4	Conclusion	84
	Bibliography	88
	Appendices	95
	Vita	102

List of Tables

1.1	Research Novelty and Contributions - Improvements to the AutoBEM framework are covered in Section 2.1. Energy, cost, and emissions savings for a utility are explored in Section 2.5. Climate impacts on a utility as well as US commercial building stock are described in Section 2.6. The methodology by which every US building was modeling can be found in Section 2.7.	4
1.2	Pros and Cons of UBEM approaches - Bottom-up UBEM has been used more recently due to decreases in data aggregation, storage, and computation costs while still allowing for individual building insights. This is especially true when the focus is buildings alone. Top-down models typically involve other variables (e.g., econometric, transportation, energy generation) in their analysis when the scope is beyond buildings alone.	11
1.3	Uncertainty Estimation Method Features - Various BEM uncertainty quantification methods, common examples and their features are shown. . . .	11
2.1	Machine Learning Windows and Statistics - Time windows and statistics were chosen to retain as much data as possible in during relevant times of the year. Weekends could be a significant differentiator between two building types. . . .	26
2.2	Machine Learning Grid Search Hyperparameters - Hyperparameter values used for grid search are shown with optimal hyperparameter values from cross-validation in bold. For more on these metrics, see [13], [17], [44], [50].	26

2.3	Machine Learning Cross Validation Comparison - Cross validation metrics for k-nearest neighbors (KNN), extreme gradient boosting (xgbTree), and random forest (RF). RF had superior mean and median classification accuracy and κ , which compares observed accuracy to expected accuracy, taking into account the chance of randomly classifying correctly.	27
2.4	Energy Efficiency ECMs - EPB Energy Efficiency ECMs focus on reducing the annual electricity end-use. Overall energy use is not necessarily decreased for the ECMs that are switching to natural gas. Electricity use is also not decreased for the PV ECM, though the PV generation was used to offset end-use and considered electricity savings.	30
2.5	Demand Management ECMs - EPB Demand Side ECMs may actually increase annual electricity use though the peak hour for a utility is typically more important from a cost and emissions standpoint.	30
2.6	Chattanooga Emissions Rates - EPB Emissions rates for the EPB service area [62]. For emissions analysis, only annual energy savings were used to calculate emissions reduction; peak demand emissions reduction was not considered.	30
2.7	IPCC Climate Model Parameters - IPCC climate model characteristics used to create Future Meteorological Year (FMY) weather files.	34
2.8	IPCC Climate Model Variables - Name and units of meteorological variables in IPCC data.	34
2.9	US Commercial Prototype Building and Climate Zone Parameters. - 1,410 commercial prototype building models were used from all US climate zones, resulting in 16,920 simulations.	36
2.10	Building Type Assignment Methodology For US Buildings were categorized by their physical characteristics determined from the prototype for each building type. Buildings were further categorized within each physical bin based on prototype construction floorarea weights.	41

3.1	Building Type Classification Quantitative Results - Each of the three methods have similar CVRMSE values. The random forest has an edge primarily because the maximum CVRMSE for that method is much lower than the other two methods, bringing the mean down. American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) considers <15% monthly and <30% CVRMSE as "investment grade" ([5])	44
3.2	Building Type Classification Qualitative Results - While the Euclidean Distance classifier had the highest accuracy, it was primarily due to the number of residential predictions coupled with the amount of residential buildings in the sample.	46
3.3	Building Type Classification Sensitivity and Specificity - Random Forest was best at differentiating commercial vs residential buildings while Euclidean Distance over-predicted residential.	46
3.4	Results of Classifications Greater than the Mean Confidence For Each Method - The mean and median CVMRSE improves for the Euclidean and DTW methods but worsens for the RF compared to Table 3.1	49
3.5	Results of Classifications Greater than the first quarile Confidence For Each Method - Filtering the data to even higher confidnce level, the mean and median CVMRSE improves for the Euclidean and DTW methods but worsens for the RF compared to Table 3.4.	49
3.6	Simulated vs Acutal US Commercial Building Energy USe - Error values for comparison of simulated commercial building energy use in the US with actual commercial building energy use reported from CBECS 2012 [61].	70
3.7	Model America HPC Parameters - 80% of THETA was used for the majority of the runs with just over 1 million buildings per run on average. .	82
3.8	Model America HPC Post-Processing Statistics - The runs were untarred and zipped in parallel with the largest states being split up so as to reduce the maximum time shown below.	82

1	US Commercial Prototype Building Multipliers - Multiplier Values used for scaling US commercial building energy to represent the total building stock.	96
2	US Commercial Floorspace per Climate Zone - Commercial floorspace percentages per climate zone display the degree to which each zone influences total US energy.	97

List of Figures

1.1	Physical BEM Workflow - Physical BEM includes a set of inputs describing the physics of the buildings and simulates the building outputting specified parameters of interest[27]. The amount of data input to develop a representative model can be one of the difficulties of physical BEM. EnergyPlus is a commonly used BEM software [57].	6
1.2	Data-based BEM Workflow - A data-based BEM approach can input similar data to a physical BEM approach, though not all data is required. Data-based models are not built on a physics simulation engine and therefore do not require specific simulation data. The algorithm and data pre-processing can also be adjusted on a case by case basis. In this way, data-based models may be much more flexible than physical models [24] but as they do not use a simulation engine, they tend to be brittle and fail to generalize to data from new buildings.	6
1.3	Earth's Climate Interactions - Earth's energy balance between atmosphere, land surface, ocean, and sea ice is key to long-term climate prediction. Each component has representative equations for each grid point which are changing over time. [52].	15
1.4	Representative Concentration Pathways - The Intergovernmental Panel on Climate Change defines Representative Concentration Pathways scenarios that range from 1.5°C to 4.9°C by 2100 [53]. This could have significant impacts on buildings, cities, and utilities.	15

2.1	Prototype Building Model Renderings - These Department of Energy (DOE) prototypes represent 70% of total US Commercial floorarea. Six vintages were used for these prototypes which collectively cover any year a building could be built. [28].	19
2.2	Real vs Prototype Simulation Data - Example comparison of real building energy use intensity (EUI) to prototype building simulation output EUI for two different prototypes. When evaluating methods of building type assignment, the real building is compared to all 97 building type and Standard combinations.	19
2.3	Energy Use By Building Type - Building type has a significant impact on both annual electricity use and natural gas use. Building type has a more significant impact on electricity use than natural gas use across building types for climate zone 4A.	21
2.4	Energy Use By Standard - Building standard impacts annual natural gas use more than electricity use in building energy models in climate zone 4A. Older buildings use more energy than newer buildings.	21
2.5	Euclidean Distance vs DTW Distance - Euclidean Distance versus dynamic time warping distance [51]. Sine curve is compared to cosine as an example for which Euclidean Distance does not adequately compare two time series due to phase shift.	24
2.6	Weather Variable Impacts on Building Energy Simulation - Dry bulb temperature, on average, tends to dominate changes in building energy use as shown here for DOE's Medium Office reference model. The next most important meteorological variables tend to be relative humidity, direct normal incident radiation, direct horizontal incident radiation, and wind speed. . . .	32

3.1	Building Type Classification Confidence Distributions - The distance-based classification methods have clear drop-off and reduction with the majority of distances being smaller and few large distances. The RF shape of the RF classifier probability is less helpful for confidence purposes as it lacks a clear cutoff point.	47
3.2	Smart Thermostat Monthly Demand Savings - Smart thermostats with utility-signaled 2.2 °C of building pre-conditioning 2 h before peak hour has the potential to save an average of 27% of a building's demand, but this sample of 101,082 buildings varies from 0–93% by individual building and time of year.	51
3.3	Smart Thermostat Monthly Demand Savings By Vintage and Building Type - Potential demand reduction achievable with 2.2 °C pre-conditioning by building vintage and building type varies from 0 to 93% of peak demand.	52
3.4	Gas Water Heater Monthly Demand Savings - Changing water heaters from electricity to natural gas can relieve peak electric demand. This study found that a building's demand could be reduced as much as 80% but averages approximately 5% of building peak.	54
3.5	Gas HVAC Monthly Demand Savings - Distributions of potential demand reduction for a building, broken down by month and building vintage, achievable with swapping HVAC equipment from electricity to natural gas or dual-fuel equipment with a utility signal to reduce electricity use during peak hour.	54
3.6	Annual Electricity Savings For Six ECMs - Distribution of potential annual electricity savings for two fuel-switching technologies (water heater and HVAC), and four traditional energy efficiency measures.	56
3.7	EPB PV Generation - The total PV potential in the EPB service area relates to the length of the day, with maximum PV generation in July when days are longest and minimum PV generation in December when days are shortest.	56

3.8	CO₂ Emissions Savings for Four ECMs - Distribution across buildings in the utility's service area of operational emissions savings for four traditional energy efficiency measures.	57
3.9	Total Cost Savings By Total Electricity and Demand - Combined utility-scale energy and demand annual retail-rate electricity cost savings. . .	57
3.10	Total Demand and Electricity Savings By Building Type - Medium office buildings, the most prevalent commercial building type, account for the majority of total, utility-scale, demand-related cost savings. Smart thermostats have very little or negative annual electricity savings. Residential and medium office commercial buildings, due to number of buildings, dominate utility-scale potential electricity cost savings.	59
3.11	Average Demand and Electricity Savings By Building Type - Average demand-related electricity cost savings by building type can indicate cost-conscious opportunities for short-term wins in demand management. Average energy-related electricity cost savings by building type are on average 4x higher than demand-related savings and typically offer the best cost savings for building owners.	60
3.12	Total Demand and Electricity Savings By Vintage - Average demand-related electricity cost savings by building vintage can indicate cost-conscious opportunities for short-term wins in demand management. Average energy-related electricity cost savings potential by building vintage indicates significant savings opportunities. By combining building type and vintage for energy and demand, an initial estimate of cost savings can be used to determine if purchase and installation of these technologies may be cost-effective.	61
3.13	Cesium EPB 3D Visualizations - Geospatial visualization of buildings provides insights not readily available from visualizations such as box-and-whisker plots [16]. The different scales and granularity that can be attained using these visualizations are useful in utility-scale BEM. (Virtual EPB Visualization)	62

3.14	Dry Bulb Temperature for Each RCP and Year - Average dry bulb temperature increases for RCP 8.5 from 2030 to 2100 and decreases for RCP 2.6 over the same time period.	64
3.15	Total EPB Projected Energy Use - Simulated energy use (GWh) decreases for the EPB service area for the three highest emission RCP scenarios (4.6, 6, 8.5) while it decreases for the low emission scenario (RCP 2.6).	64
3.16	Single Building Residential Heating Energy Use - Heating Energy use (kWh) for a representative EPB IECC building decreases significantly for high emission scenarios and increases for low emission scenarios.	66
3.17	Single Building Residential Cooling Energy Use - Cooling Energy use (kWh) for a representative EPB IECC building increase for all RCPs. . . .	66
3.18	Single Building Residential Monthly Demand - High mitigation scenarios result a greater winter demand as average temperatures stay flat or decrease. High emission scenarios result in a greater summer demand as average temperatures increase.	68
3.19	US Commercial Floorspace Trend - Linear fit to commercial floorspace historical data. A linear fit is a realistic approximation of commercial floorspace in the US.	70
3.20	Average US Weather Variables for Various RCPs - These weather variables shown here are averages of all of the climate zones, unweighted by the number of buildings in each zone. This provides a different angle than the building floorspace weighted energy simulation results.	72
3.21	US Total Commercial Building Energy Use for RCPs - The total energy in the US decreases significantly for the highest emission scenario and increases for the lowest emission scenario. The two middle emission scenarios also decrease significantly.	74
3.22	US Total Commercial Building Electricity Use for RCPs - Electricity use in the US (mostly used for cooling) with constant building area increases in all scenarios by 2100.	75

3.23	US Total Commercial Building Natural Gas Use for RCPs - Natural Gas use in the US (mostly used for heating) with constant building area decreases in the three highest emission scenarios but increases for the mitigation scenario.	77
3.24	US Total Commercial Building Energy Use for RCPs and Increasing Commercial Floorspace - If commercial building area continues to increase at the same rate (assuming no changes in building technology), US commercial building energy will increase for all scenarios with the relative increase proportional to the constant floorspace results.	79
3.25	Buildings per state and county - The three largest counties from California, Texas, and Illinois were removed for better visualization of the bottom figure. The largest number of buildings are in states such as California, Texas, and Florida.	81
3.26	Model America HPC Statistics Correlations - The percent of successful runs (Tarred-up Buildings (%)) has a negative correlation with wall time. . .	83
1	US Total Commercial Building Energy Use for RCPs By Climate Zone - Total energy by by climate zone changes are related to the temperature of each zone. Colder zones total energy decreases as average global temperatures increase while warmer zones total energy increase along with the global temperature. The impact of the zones with the most commercial buildings can be seen with the order of magnitude of zones 2A, 3A, 4A, and 5A. . . .	98
2	US Total Commercial Building Electricity Use for RCPs By Climate Zone - Total electricity results by climate zone are also affected by the temperature of each zone. Warmer climates see huge increases in electricity under low mitigation scenarios. The electricity decrease for most of the zones with the most commercial floorspace (3A, 4A, 5A) is what leads to the overall electricity decrease in the US for higher emission scenarios.	99

3	US Total Commercial Building Natural Gas Use for RCPs By Climate Zone - Total natural gas results by climate zone are affected by the temperature of each zone. All climates see decreases in natural gas under low mitigation scenarios. The increase in temperature under these scenarios leads to less heating energy use.	99
4	US Total Commercial Building Electricity Use for RCPs and Increasing Commercial Floorspace - The total electricity use with increasing commercial floorspace increased by at least 55% for each scenario from 2030 to 2100.	100
5	US Total Commercial Building Natural Gas Use for RCPs and Increasing Commercial Floorspace - The total natural gas use with increasing commercial floorspace increased for all scenarios but the increase was much less than the electricity change.	101

Chapter 1

Introduction

1.1 Motivation

Energy consumption is steadily increasing year over year in the United States (US). Increasing population, climate change, and an abundance of new technologies have led to this end-use energy inflation. This increase in energy end use goes hand in hand with surging emissions as energy producers continue to meet energy demands. It is critical to understand and model this energy consumption in order to predict and mitigate any adverse outcomes climate change may have on humanity in the coming years. This modeling task is difficult, as energy consumption in the US involves a complex web of geographically and economically interconnected energy consumers across the country. One of the largest components of this web is buildings, which use about 40% of US energy [19]. Modeling buildings provides a significant opportunity to transform the built environment by examining new technologies, optimizing building and grid efficiency, and testing the resiliency and reactions of buildings under various scenarios. The results of these analyses will inform decision-making and policy in the years to come.

1.2 Research Contribution

As urban building energy modeling (UBEM) is a relatively nascent field [43], it contains many novel research areas to explore. While advances in high-performance computing (HPC)

have allowed for the modeling of large numbers of buildings yet most analyses do not take advantage of this computing power and model tens to hundreds of buildings [3] while relying on city-based resources such as tax assessors data or city Geographic Information System (GIS) databases. Another issue with many of these analyses is that they lack the actual high-resolution data necessary to assess the quality of and calibrate their building energy models. An approach called “Automatic Building Energy Modeling (AutoBEM)”, has been created that uses real 15-minute electricity data from an electrical distribution utility, allowing for the modeling of parts of eight counties representing more than 178,000 premises in Chattanooga. This approach is unique in that it allows for scalability beyond a single data source.

Significant improvements were made to the AutoBEM framework that substantially reduced the pre-adjusted error rates (adjustments to simulation results can be made to eliminate simulation bias and provide more representative simulations). These improvements include tuning building prototypes, building prototype and vintage assignment strategies, improved footprint matching, and handling of multiple electrical meters per building, among others. Utility-scale analyses were undertaken on these improved building models, including building retrofits, demand reduction techniques, cost and emissions savings opportunities, renewable energy potential, microgrid analyses, and climate impacts. These analyses are useful for utilities, as they allow them to estimate the efficacy of new building technologies or the impact of climate change on their grid. As one of the primary novelties of this modeling framework is its scalability, expanding beyond Chattanooga was an important goal. An effort was made to expand the the scope beyond 178,000 buildings to every building in the US using HPC resources (more than 125 million buildings). This work will lay the groundwork for future mega-scale analyses while demonstrating the capability for HPC to such a large number of buildings and allowing for utilization of simulation-informed analysis across the US.

This work also makes research contributions beyond the scope of a single utility. Previous US building energy climate research has focused on single climate zones, individual building types, or forecasting total energy use based on past data. The geographic scale, climate heterogeneity, and granularity of this work make it novel and valuable to the research community.

These contributions provide methodologies and results at a scale not previously seen in building energy modeling (BEM), illustrating how advances in computing technology can benefit this field greatly. It will become increasingly important to model vast geographic areas to predict outcomes and find optimal strategies to reduce energy use, emissions, and costs while making BEM ubiquitous in urban planning. A summary of the topic's novelty and the contributions of this work is shown in Table [1.1](#).

1.3 Building Energy Modeling

BEM is an arduous task requiring a multitude of skill sets, as there are countless physical phenomena occurring in a building at any given time and very few buildings are exactly the same or have similar energy profiles. These interactions can be captured using physical and data models.

1.3.1 Physical Building Energy Modeling

Physical BEM takes inputs describing a building - including geometry, construction materials, lighting, HVAC, refrigeration, water heating, control strategies, and occupancy schedules, among others - and combines them with weather data to calculate thermal loads, system responses to those loads, and energy use, along with several other related metrics. These calculations are performed using physics equations from many fields, including fluid dynamics, heat transfer, electrical engineering, construction technology, lighting, and others. The complexity increases in buildings (as in many other systems) as the interactions between various physical phenomena must be taken into account [\[39\]](#). A primary benefit of using a physics-based model is that simulation results will generally be reasonable, as they are based on the laws of physics. However, supplying the model with adequate inputs to capture the exact physics of a building can be nearly impossible since a building energy model typically averages about 3,000 inputs. Another benefit of physical modeling is the ability to test specific changes to a building, such as a different technology or improved efficiency, to observe the reaction in terms of the building's energy consumption. EnergyPlus is a commonly used BEM software and is the simulation engine used to model building energy

Table 1.1: Research Novelty and Contributions - Improvements to the AutoBEM framework are covered in Section 2.1. Energy, cost, and emissions savings for a utility are explored in Section 2.5. Climate impacts on a utility as well as US commercial building stock are described in Section 2.6. The methodology by which every US building was modeling can be found in Section 2.7.

Topic	Novelty	Contribution
Improvements to AutoBEM framework	AutoBEM is a fully integrated bottom-up UBEEM framework by which any buildings in a region in the US can be modeled.	Tuning building prototypes
		Improving building footprint selection
		Handling of multiple electrical meters per building
		Real electricity data
		Building type and vintage assignment comparison
		AutoBEMGen/AutoSIM [1][49] upgrades/improvements
Energy, Cost, Emissions Savings and Climate Impacts for Utility	Simulation of many energy conservation measure for 178k+ buildings. No ECM or climate analyses have been done at this scale using a bottom-up methodology.	More than 2 million simulations on HPC resources
		Demand management analyses for various ECM
		Common energy efficiency measures evaluated
		Models simulated using climate model weather data
Future Meteorological Year Simulations for all Commercial Buildings in US	Current building energy climate forecasts in the US are geographically limited, limited to few building types or utilize a top-down approach.	1,440 building prototypes from all US climate zones used
		16,920 simulations using climate model weather data
		Prototype simulation energy scaling technique developed and implemented
Model America	No bottom-up UBEEM analysis has ever been done at this scale. (> 125M buildings)	Data aggregation/cleaning
		Building property classification
		HPC Generation/Simulation
		Big Data Movement

use in this work. The specific inputs driving the engine as well as potential outputs can be found in the input/output reference [58]. A typical framework for physical BEM is shown in Figure 1.1.

1.3.2 Data-Based Building Energy Modeling

Data-based building energy models make use of various sensors throughout a building, storing data from any relevant building systems. These data are aggregated and split into training and testing sets. A model is chosen that is best suited for the data and is trained using the training set. The model's accuracy is determined using the testing set. If the model is of sufficient quality, it can be used to predict specific independent variables. Data-based models differ from physics-based models in that data-based models utilize an independent variable (here, typically building energy consumption), while physics models only use it for calibration. This makes data-based models a good choice for forecasting energy use or demand in buildings, as past occurrences are often a good indicator of future outcomes. However, data-based models can be more unpredictable than physical models because they usually cannot extrapolate, leading to the possibility of erroneous predictions. This unpredictable behavior has been handled in several ways, including incorporating physical equations into algorithms. A typical framework for data-based BEM is shown in Figure 1.2.

1.4 Urban Building Energy Modeling

While BEM focuses on modeling a single building, UBEM analyzes tens to millions of buildings. UBEM can be either bottom-up or top-down. A comparison of the different UBEM methods is shown in Table 1.2.

1.4.1 Bottom-up Urban Building Energy Modeling

Bottom-up UBEM uses the previously described physical and data-based models but at a much larger scale. This is accomplished in one of two ways: Each building can be individually

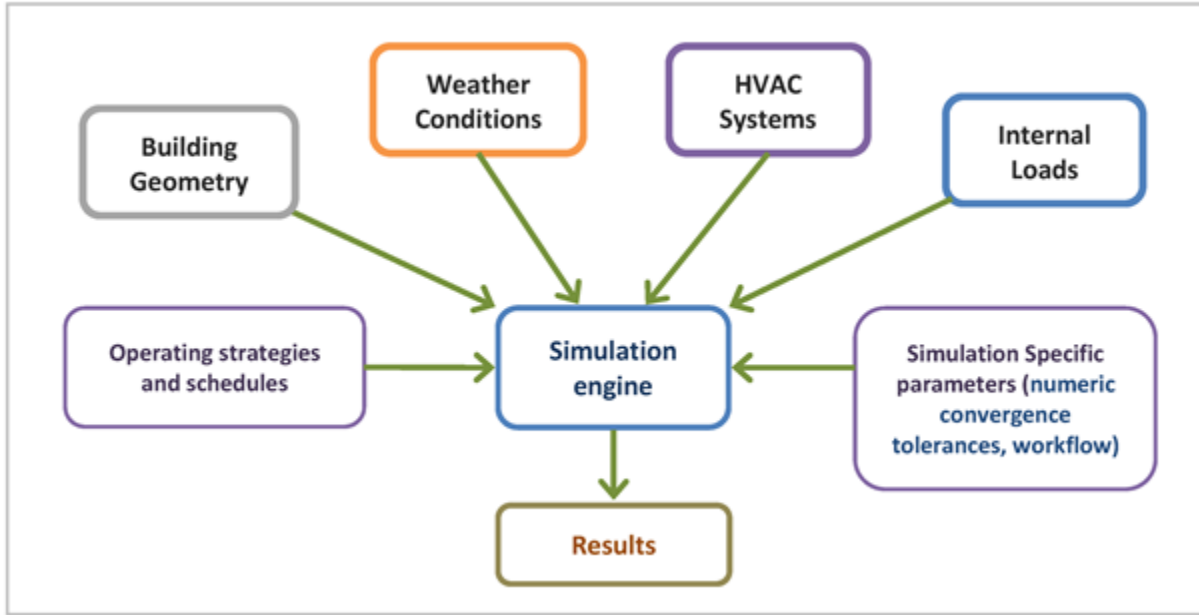


Figure 1.1: Physical BEM Workflow - Physical BEM includes a set of inputs describing the physics of the buildings and simulates the building outputting specified parameters of interest[27]. The amount of data input to develop a representative model can be one of the difficulties of physical BEM. EnergyPlus is a commonly used BEM software [57].

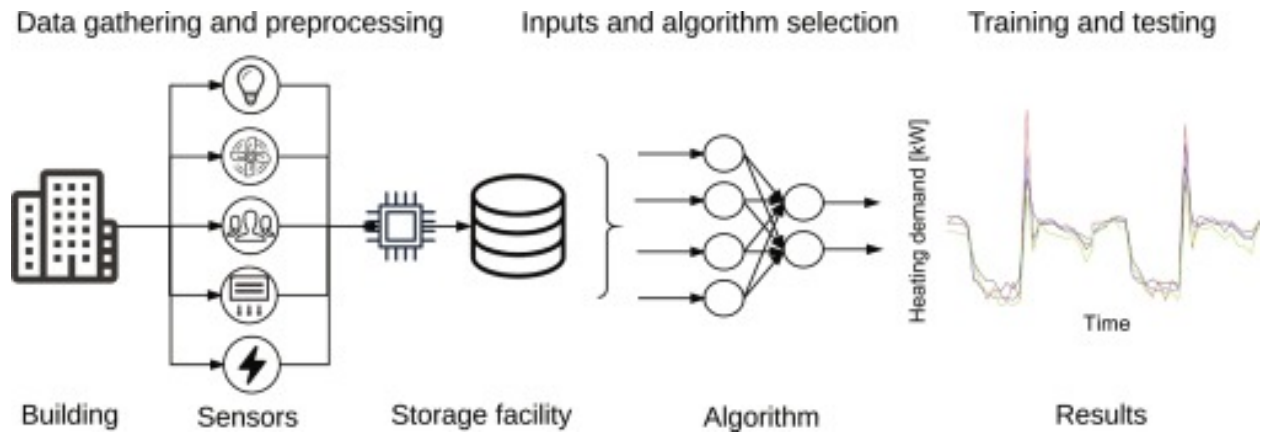


Figure 1.2: Data-based BEM Workflow - A data-based BEM approach can input similar data to a physical BEM approach, though not all data is required. Data-based models are not built on a physics simulation engine and therefore do not require specific simulation data. The algorithm and data pre-processing can also be adjusted on a case by case basis. In this way, data-based models may be much more flexible than physical models [24] but as they do not use a simulation engine, they tend to be brittle and fail to generalize to data from new buildings.

modeled, or a set of representative buildings can be modeled and the results scaled up to represent all of the buildings. Each method entails certain trade-offs.

Bottom-up physical BEM can be extremely computationally expensive at a large scale, as each model must be simulated. This computational challenge has become easier in recent years due to advancements in HPC. Physical building energy models require the most data, as detailed information about the building's geometry, height, number of floors, construction materials, window-to-wall ratios, HVAC type, and occupancy schedules are required to accurately represent the building. For physical BEM, actual energy data are not required but are helpful for accuracy validation and model calibration. These data can be acquired from various public or private sources but can be challenging to obtain, nonexistent, or unreliable. For this reason, simplifications often must be made. These frequently include the use of building archetypes that describe much of the building stock. The parameters of these building archetypes can be applied to buildings on an individual level. The methodology for collecting data used to create building energy models is one of the primary distinguishing factors of UDEM approaches, as having more data that describe a building often leads to more accurate results. One issue that arises when physically modeling a large number of buildings is that these models can lack detail, since zoning and other building parameters must be automated. On the other hand, individually modeling each building allows the model to capture the building's unique geometry and systems as well as the interactions between buildings. Higher resolution weather data can also be used, if necessary.

Various UDEM analyses have been performed in the field in different geographic areas and using different methodologies. A study of 332 residential buildings in Kuwait City introduced a Bayesian calibration method for archetype assignment with improved error rates compared to deterministic approaches [14]. A similar Bayesian approach was applied on 2,663 buildings in Cambridge, Massachusetts, comparing annual and monthly calibration approaches [54]. A smaller analysis of 22 urban buildings in California evaluated a data-driven urban energy simulation method that aimed to capture the inter-building dynamics of dense urban areas. The results indicated that the framework could adequately predict energy consumption at various time intervals and partially capture inter-building dynamics [38]. A study in Boston modeled 83,541 buildings to outline a workflow for a large number of

buildings. This analysis was not calibrated but was roughly crosschecked using US national consumption data. Metered energy data was the main inhibitor of using these models to guide energy policy [15]. While the use of actual energy use was the main obstacle of that analysis, it is an important element of the usefulness of the simulation outputs in this work. SUNtool [45] and CITYSIM [46] are examples of full-scale streamlined frameworks that use urban data to create building models of every building in a database and capture interactions between buildings and microclimates.

These bottom-up physical methods are most similar to the AutoBEM framework explored in Section 2.1. However, there are a number of differentiating factors. First, the number of buildings analyzed ($> 178,000$ for Virtual EPB and $> 125,000,000$ for Model America) is much greater than in any of the above-mentioned research, with most current UBE methods covering fewer than 2,000 buildings. Another significant differentiating factor of this analysis compared to other previous research is the use of 15-minute smart meter electricity data for simulation bias adjustment and empirical validation. A major issue with various UBE methods is their reliance on aggregating building properties from datasets such as tax assessor data, which can limit the scale and geographical reach of the modeling effort. AutoBEM does not rely on these data and is therefore not limited by their availability [36]. The assignment of building properties is further explored in Section 2.2.

Physically modeling buildings that represent a majority of the building stock and scaling them to the full number of buildings drastically lightens the computational load and data collection effort - that is, modeling only the archetypes and scaling the results rather than applying building archetype parameters to many unique building geometries. This can allow for more detailed representative models, as fewer models need to be created, though the unique building geometries and interactions between buildings are omitted. Accruing the correct multipliers to scale the representative buildings up to the full sample can also be difficult or inaccurate, depending on the location being modeled.

This representative building simulation with scaling methodology is used for US commercial building climate projections explored in Section 2.6.2. It is useful for this scenario because it enables one to estimate energy use of millions of buildings with orders of magnitude fewer simulations.

Some work has also been performed in data-based modeling of representative buildings. One method of doing so involves clustering buildings in order to create models of these representative clusters, scaling by the amount of buildings belonging to that cluster.

Bottom-up data-based BEM of individual buildings can be done by aggregating distinct building model outputs. The data necessary for this method could be as little as historical energy consumption data, though other building and environmental parameters are usually necessary for substantive predictions.

An example of a bottom-up data-based approach was done in Cambridge, Massachusetts on 6,499 buildings [31]. They used tax-assessors data as well as geographical survey information to create a feature vector and use parametric and non-parametric learning methods to predict building energy use. They were able to explain about 75% of energy consumption variance in these buildings, meaning the model captures about 75% of the spread of the energy consumption.

ResStock [64] and ComStock [26] are two additional tools that are currently being developed by the National Renewable Energy Laboratory which can be used with aggregated energy data in a hybrid approach to estimate end-use load profiles for every location in the US [41]. This methodology is not fully bottom-up and thereby loses some of the granularity of modeling each building individually.

1.4.2 Top-down Urban Building Energy Modeling

Top-down UBEM considers an entire group of buildings rather than modeling them individually. Data are collected and an algorithm is trained to predict the energy usage of the group of buildings. This requires far fewer initial data, as individual building data are not required. Econometric variables such as income, gross domestic product, fuel prices, and climate data are often used because they are available at the same scale, which also expands the scope beyond buildings alone. The main drawback of the top-down approach is that individual building energy data are not attainable. For this reason, most recent analyses have used bottom-up UBEM, as individual building analyses are essential for decision-makers.

There are examples of analyses using a top-down UBEM approach. IMACLIM is a French top-down computable general equilibrium model that uses demographics, labor, productivity

and international energy prices to predict many outputs, including energy use and carbon output in France [23]. An energy-economic model was developed in Japan focusing on a rural, residential Japanese region. This model focuses on the structures of energy supply and demand in the region and takes into account both the technological and economic aspects of energy conversion to assess CO₂ emission reduction possibilities [4].

1.5 Building Energy Modeling Uncertainty

Uncertainty in BEM is an important concept, as the number of factors influencing energy use in buildings is by nature uncertain [55]. There are two types of uncertainties at play in BEM: epistemic and aleatoric. Epistemic uncertainty stems from a lack of knowledge. An example of epistemic uncertainty in BEM is the power density of equipment within a building, which cannot be easily represented without measuring each piece of equipment's exact usage patterns. Aleatoric uncertainty refers to the intrinsic randomness in a system. An example of aleatoric uncertainty in BEM is occupant behavior, which can be estimated but involves some random events. Other sources of uncertainty in BEM relate to the weather, building envelope, and HVAC system.

Uncertainty quantification is currently handled using one of two types of analysis: forward or inverse. Forward uncertainty analysis makes use of known input uncertainty and propagates it through the model to determine output uncertainties. Inverse uncertainty analysis quantifies unknown input variables using measured data. The two methods are linked, as sampling-based inverse uncertainty requires many forward uncertainty propagation simulations, while the results of inverse uncertainty analysis can be used to estimate the efficacy of building technologies.

Forward uncertainty analysis can be divided into two subcategories: probabilistic and non-probabilistic [34]. Probabilistic methods are typically used when data are plentiful, while non-probabilistic approaches are used when data are sparse. Most current studies rely on probabilistic propagation for uncertainty quantification. Probabilistic uncertainty quantification is further subdivided into sampling and non-sampling methods. Sampling methods treat the model as deterministic and run it many times with different samples,

Table 1.2: Pros and Cons of UBE approaches - Bottom-up UBE has been used more recently due to decreases in data aggregation, storage, and computation costs while still allowing for individual building insights. This is especially true when the focus is buildings alone. Top-down models typically involve other variables (e.g., econometric, transportation, energy generation) in their analysis when the scope is beyond buildings alone.

	Pros	Cons
Top-Down	Less detailed individual building data unnecessary	No individual building results
	Can use time-based modeling techniques	No representation of end-uses
	Can be easier/faster to model large area	Depend on past energy use
Bottom-Up	Detailed building data not required	No ability to model different technologies in building
Data-Based	Can use time-based modeling techniques	Depend on past energy use
Bottom-Up	Ability to model different technologies in building	Requires most detailed data
Physical	End-use energy consumption detail	Computationally expensive
		Many assumptions typically required

Table 1.3: Uncertainty Estimation Method Features - Various BEM uncertainty quantification methods, common examples and their features are shown.

Direction	Function/Features	Sub-type	Example/Features
Forward	Propogates uncertainty through building energy models	Probabilistic	Monte Carlo
			Flexibility with different data types/probability distributions Computational cost
	Intuitiveness/Ease of use	Non-Probabilistic	Min-Max Less computational cost than sampling Requires domain knowledge Results can be less reliable than sampling
Inverse	Infers inputs based through building energy models based on measured data	Frequentist	Maximum Likelihood Estimation
			Infer single input point Confidence intervals available
	Requires measured data	Bayesian	Bayes Theorem Infer input probability distribution Requires domain knowledge/data to create prior distributions

while non-sampling methods rely on model perturbation. Sampling methods are widely used in current BEM research, while non-sampling methods have not been widely adopted. Both techniques have disadvantages; sampling methods are computationally expensive, while non-sampling methods require extensive changes to existing modeling systems.

Monte Carlo sampling is a sampling-based probabilistic method widely used due to its intuitiveness, ease of implementation, and reliability, as well as its flexibility in dealing with different data types and probability distributions [30]. It uses sampled input probability distributions to estimate output probability distributions. The main disadvantage of Monte Carlo sampling is its computational cost, though this can be mitigated with efficient sampling methods or surrogate modeling.

A non-probabilistic method used in forward uncertainty analysis is the use of a minimum and maximum input interval to estimate the minimum and maximum output intervals. This is a straightforward method, but the input interval may be difficult to determine and the results can be unreliable compared to probabilistic methods.

Inverse uncertainty analysis is conducted using either a frequentist or Bayesian technique. Frequentist methods rely on measured data to infer single input parameter values and their deviations. A commonly used frequentist implementation is maximum likelihood estimation, which compares measured data to model predictions to find input parameters that maximize the associated function [40]. Confidence intervals for the input point are inherently available when using linear models but numerical optimization techniques are required for non-linear problems. Bayesian techniques use expert knowledge or available data sources to initialize input parameter distributions (priors) with measured data and Bayes' Theorem (a way to update predicted probabilities or distributions with new information) to update priors and create posterior distributions [12]. The incorporation of prior knowledge about the system may lead to improved input distributions. A comparison of these BEM uncertainty estimation techniques is shown in Table 1.3.

These BEM uncertainty quantification techniques are critical for modeling individual buildings, as the analysis is focused and detailed input data are available, but they can be less useful or impractical when modeling large areas. The number of buildings involved in large UBEM invokes the law of large numbers, which states that, as sample size increases,

the sample mean approaches the true mean. For this reason, UBEM often uses prototype buildings, intended to represent a cross-section of common building types and cover 75% of built commercial construction [56]. As the number of buildings of a given type within a sample increases, the mean of those buildings will approach the representative prototype building as the impact of outlier buildings is diminished.

Uncertainty is also very difficult to quantify for large numbers of buildings, especially using common probabilistic sampling techniques. This is because, in UBEM, most of the input parameters are typically assumed or assigned based on prototype buildings or other coarse data, making input distributions difficult to create. The other difficulty of using probabilistic sampling techniques in UBEM is the computational challenge involved. Simulating large spatial areas is already computationally demanding, and sampling these input parameters and simulating these buildings multiple times could be impractical depending on the number of buildings in the study.

For these reasons, uncertainty quantification was not directly applied in this analysis or included in the AutoBEM framework. Probabilistic methods would have been extremely computationally expensive for even a small number of samples per building, quickly bringing the number of simulations into the millions. While inverse uncertainty analysis could have been used, the lack of granularity of individual building input data would have made these results less valuable. The invocation of the law of large numbers with more than 178,000 buildings and use of prototype building energy models mitigated the need for uncertainty quantification. The use of bias adjustments to simulated data and empirical validation using measured data also confirms that a deterministic approach is sufficient.

1.6 Urban Building Energy Modeling Limitations

Though BEM has been an active field for many years, UBEM has only recently come to the forefront, with advances in computing allowing for the modeling of more and more buildings. Because it is a new field, there are some limitations related to current UBEM strategies, one of which is related to individual building simulation efficacy. While aggregated building simulation results are useful for large numbers of buildings, simulation quality will

likely decrease as the number of buildings in a sample approaches one, if not adjusted to measured energy data.. The irregularity of building energy use becomes an issue, as any particular building may not be accurately represented by a prototype building. This prototype building limitation could be mitigated in future work through the use of sensing technology or crowdsourcing to obtain individual building properties such as occupancy schedule, construction materials, and window-to-wall ratios.

1.7 Climate Modeling

It is important to understand how the built environment will be affected by climate change. The field of climate modeling has benefited from advances in HPC, which have improved the spatial and temporal resolution of predictions. Global Climate Models (GCMs) are complex mathematical representations of the major climate system components (atmosphere, land surface, ocean, and sea ice) and their interactions [22]. These interactions are shown in Figure 1.3. Climate models produce simulations of past data to be compared with observed data to instill confidence in future predictions. In recent years, more complex simulations have been run in an effort to obtain better representations and a higher-fidelity picture of what the future may hold. These simulation results provide data to inform decision-making and aid in preparation for the future in many different fields. Climate model data are used to estimate building energy use in Section 2.6.

As it is uncertain how emissions will trend in the coming years, the Intergovernmental Panel on Climate Change (IPCC) created the representative concentration pathways (RCPs). These climate scenarios were created to standardize the work of many climate researchers across the globe. The RCPs contain a common definition of emissions values through the year 2100 for each pathway. The names of the four RCPs define the level of radiative forcing (W/m^2) expected in 2100. Radiative forcing measures the influence a variable plays in altering the balance of incoming and outgoing energy in Earth's atmosphere. The different pathways represent projections varying from a decline in radiative forcing to a steady rise [63]. These RCPs are used in this work to quantify uncertainty in future weather by illustrating how building energy will be effected by each pathway.

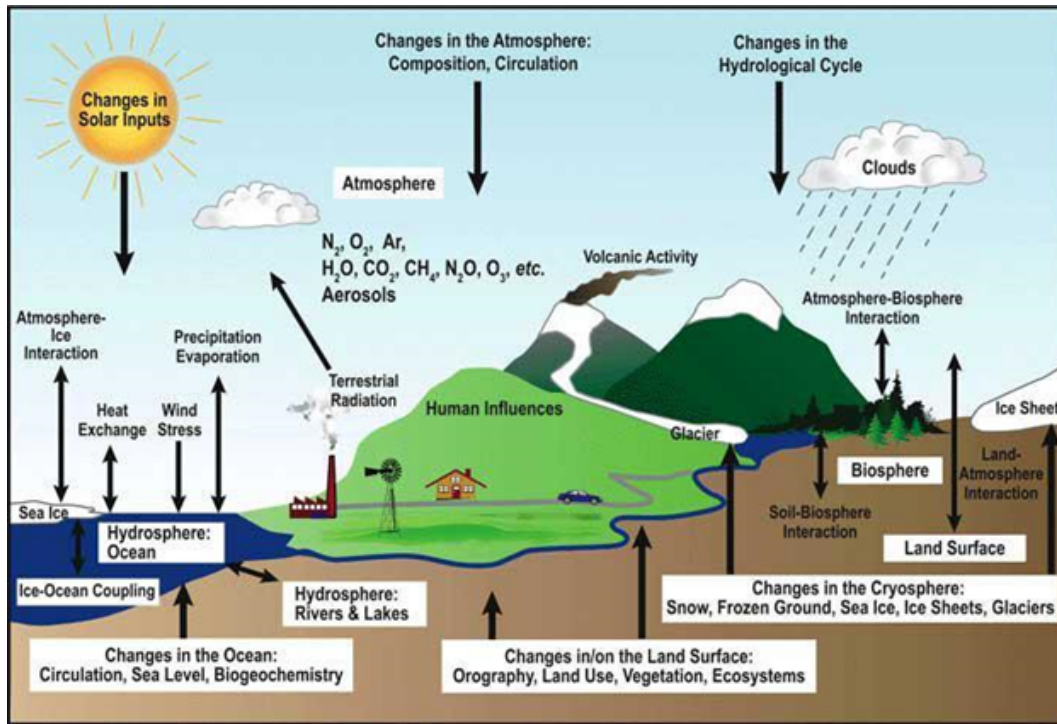


Figure 1.3: Earth's Climate Interactions - Earth's energy balance between atmosphere, land surface, ocean, and sea ice is key to long-term climate prediction. Each component has representative equations for each grid point which are changing over time. [52].

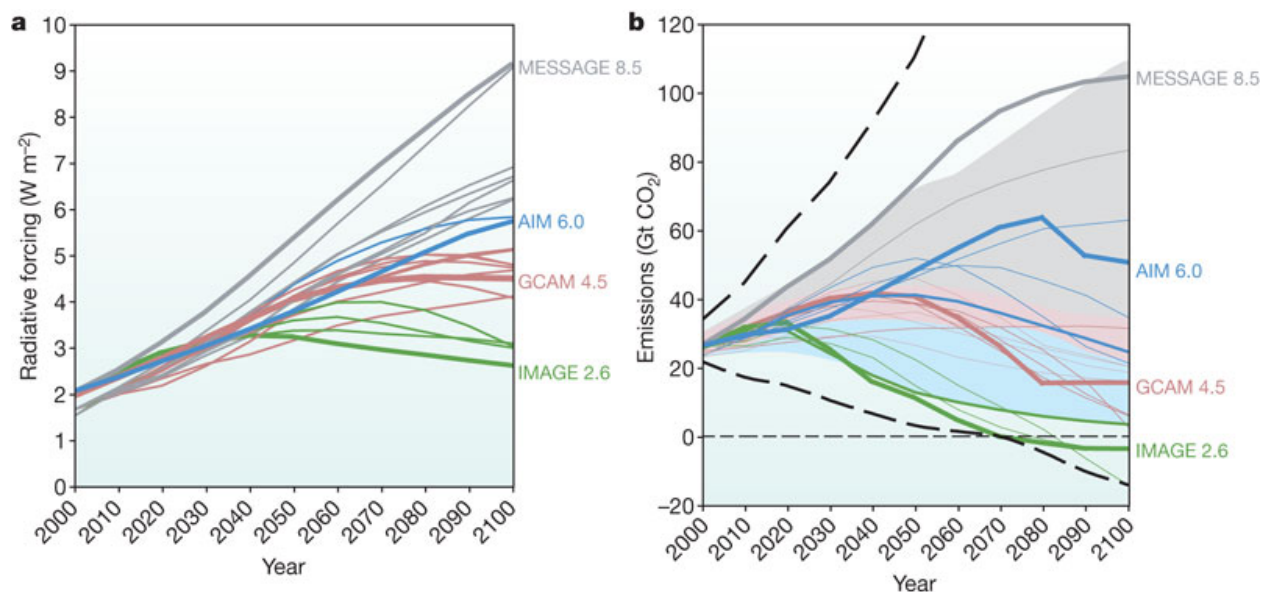


Figure 1.4: Representative Concentration Pathways - The Intergovernmental Panel on Climate Change defines Representative Concentration Pathways scenarios that range from 1.5°C to 4.9°C by 2100 [53]. This could have significant impacts on buildings, cities, and utilities.

Chapter 2

Methods

2.1 Virtual Electric Power Board (Chattanooga)

Our group at Oak Ridge National Laboratory (ORNL) has created a framework by which it is possible to model each building in a utility service area. A partnership was formed with the Electric Power Board (EPB) of Chattanooga, Tennessee, to model their service area, consisting of about 178,000 building electrical meters. They shared 15-minute real electricity data for the year 2015. Their geographical area covers parts of eight counties in southeast Tennessee and northwest Georgia. Creating a digital twin of a utility service area is valuable because it allows the utility to simulate countless scenarios to optimize its buildings and grid. The utility may want to estimate the electricity, cost, and emissions savings from a retrofit technology applied to their service area or forecast how their grid will be affected by climate change projections. There are many more examples of possible analyses and modifications that make utility-scale BEM worthwhile. These digital twin building technology and climate applications are explored in Sections [2.5](#) and [2.6.1](#) respectively.

To create a digital twin of the EPB service area, a UBEM method had to be chosen. Bottom-up physical BEM offers several advantages over data-based models, including retrofit modeling, which is particularly important to a utility. The primary disadvantages of physical modeling are related to the quantity of data necessary and computational challenges. This analysis used HPC - including some of the world's fastest supercomputers (TITAN, THETA) - to minimize computational difficulties while allowing for flexibility in the analysis by

modeling all buildings individually. The data sources and algorithms are collectively called “Automatic Building Energy Modeling ([AutoBEM](#))” [37].

Significant improvements have been made to the framework for which a utility is modeled. These improvements range from building archetype tuning to improved footprint matching and duplicate-premise energy reduction strategies. One of the largest sources of improvement was the refinement of the commonly used residential prototype, with the upgraded model providing a more representative residential building simulation.

2.1.1 Data Aggregation

The first step in developing building energy models is collecting data. Most UBEM approaches use tax assessor data or a city or county data source to create their building energy models. While this is a valid approach, it is not scalable beyond a single city or county without the aggregation of many data sources. Even if the data can be gathered appropriately, there are often gaps if some counties share less or different data than others. A method for developing building energy models in a scalable fashion without tax assessor data was needed. The EPB service area, consisting of eight different counties across parts of two states, did not have applicable tax assessor data, making it a viable area to evaluate these scalable techniques.

The first data that must be aggregated describe the building's physical makeup in terms of the building's 2D footprint and building height. Originally, building footprints were gathered using ORNL image recognition. New building footprints taken from Microsoft's freely available building footprint data, consisting of over 125 million footprints in all 50 states, were matched to electrical meters in the EPB service area. These data were created from semantic segmentation using a deep neural network architecture (ResNet340) to recognize building pixels and polygonization to convert pixel blobs into polygons. The pixel error rate on their evaluation set was 1.15% with a precision and recall of 94.5% [33]. Building heights were found using LiDAR for the state of Tennessee as well as a small region in the state of Georgia. These two features together describe the physical shape of the building. The building geometries were then associated with an electrical meter by finding the distance between each meter's GPS coordinates and the detected buildings.

An issue that arose when matching electrical meters to detected buildings was buildings with multiple electrical meters. Modeling each of these premises as the full building would result in a significant overestimation of building energy use. Of the approximately 178,000 meters in the EPB service area, there were only 115,504 distinct buildings. For this reason, only distinct buildings were modeled, with simulation results post-processed by dividing the electrical use for a multi-meter building by the number of shared meters for that building. This multiple meter issue is exacerbated in very large buildings, where it becomes critical to downscale the results from the whole building.

Once the building geometries have been established and joined to their electrical meters, establishing parameters relating to how the building functions is important for each model. Prototype building type and vintage are used to assign many of these parameters. Prototype building models are a set of models consisting of 16 building types that represent much of the built environment, with current models representing about 75% of commercial buildings in the US [56]. The vintage of the building is determined by the year it was built and can affect various building parameters, including efficiencies. Building type and vintage prototypes were assigned to each individual building by comparing the real 15-minute electricity use to the simulated electricity use for the 97 prototype buildings and vintage combinations for this climate region (ASHRAE-169-2006-4A). The matched building type and vintage features (HVAC system, insulation, occupancy, floor-to-ceiling height, window-to-wall ratio, etc.) were then assigned to the building geometry. Several methods were evaluated for classifying building type and vintage. 3D renderings of the building prototypes are shown in Figure 2.1. These methods have collectively been described previously [36], [37].

2.2 Building Type Classification

The impact of building type on simulation results is shown in Figure 2.3 for climate zone 4A (Chattanooga). Each building type and vintage combination was simulated using the same geometry and simulation parameters to determine the impact of building type and vintage on simulation outputs. The difference in energy use across building types is significant. Electricity use is especially impacted by building type, with the simulated annual electricity

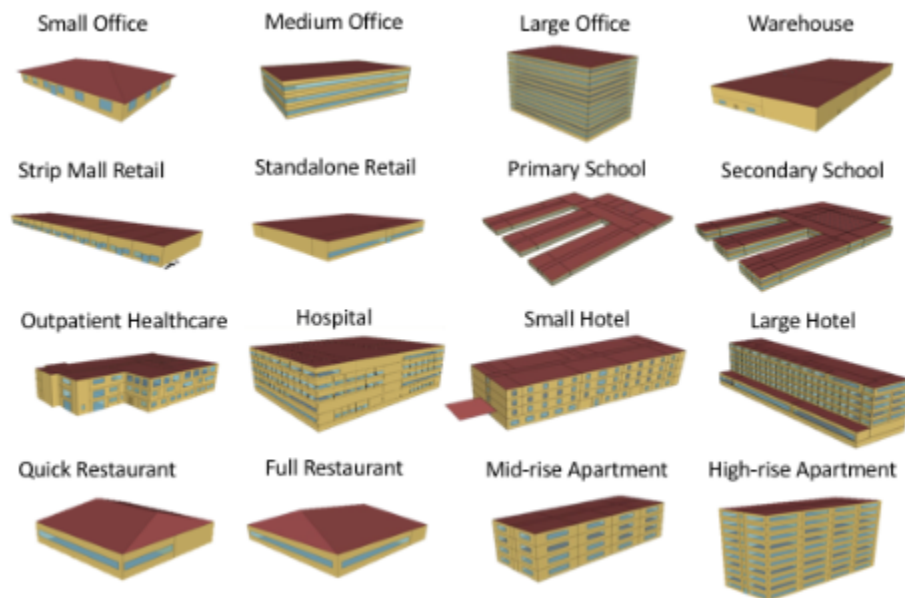


Figure 2.1: Prototype Building Model Renderings - These Department of Energy (DOE) prototypes represent 70% of total US Commercial floorarea. Six vintages were used for these prototypes which collectively cover any year a building could be built. [28].

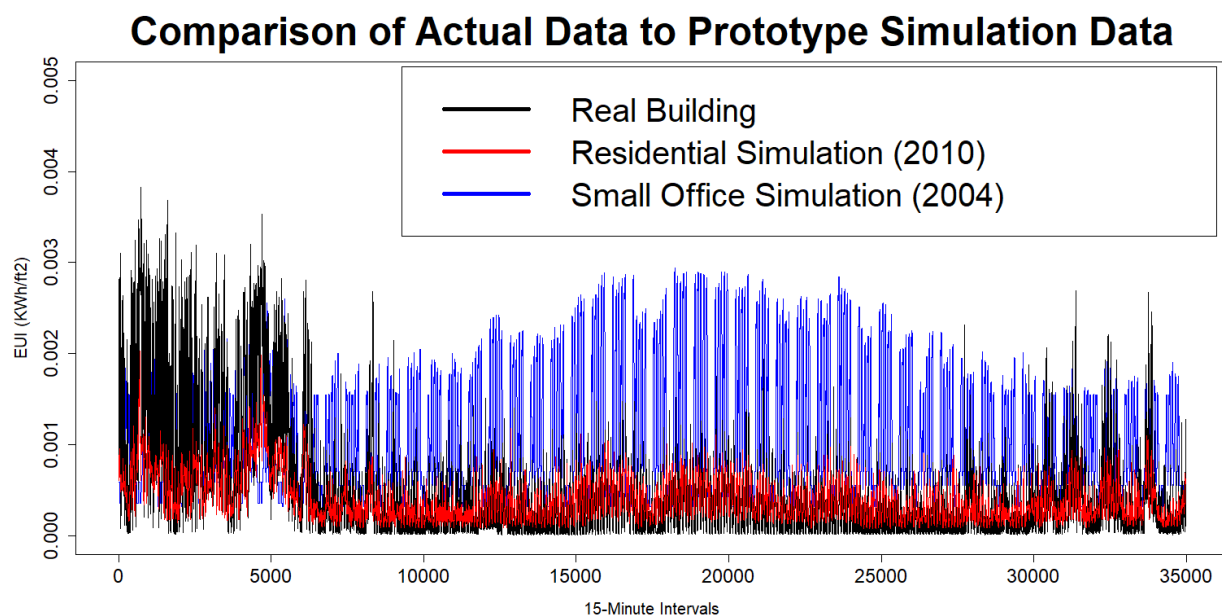


Figure 2.2: Real vs Prototype Simulation Data - Example comparison of real building energy use intensity (EUI) to prototype building simulation output EUI for two different prototypes. When evaluating methods of building type assignment, the real building is compared to all 97 building type and Standard combinations.

output of less than 500 GJ for building types like warehouses and close to 4,000 GJ for hospitals with the same geometry. Natural gas use is also affected but to a lesser extent, while several building types use no natural gas.

The building standard also has an effect on the simulation outputs, though its impact is less than building type. The impact of building standard on simulation results is shown in Figure 2.4 for climate zone 4A (Chattanooga). The impact of building standard on simulated energy use is intuitive: Older buildings use more energy. This is likely due to improvements in efficiency and other building technologies. Buildings built before 1980 use the most energy, while the 2013 which is the most recent standard used for the analysis, uses the least. The discrepancy between the old and new standards is more extreme for natural gas use compared to electricity.

As building type classification is one of the most important parameters in BEM, it was important to ensure that the methods used were optimal. A sample of 100 premises was randomly chosen to evaluate the various building type classification techniques. Building energy use intensity (EUI)(typically kWh/ft²) was used to prevent building size from heavily influencing the classifications. A sample of 100 buildings allows the pros and cons of each method to be determined, including simulation quality and building type classification accuracy. To evaluate the methods of building type assignment, each real building's actual EUI was compared with every prototype building simulation output EUI (simulated for the same year the real data were obtained) to obtain the prototype most similar to the real building. An example of a real building compared to two sample prototype buildings is shown in Figure 2.2.

In many cases where utility data are available, some post-processing may be done to adjust the building classifications based on the utility rate structures, though this was not considered for this analysis. Even as post-processing may better classify the actual building type, the correct building type classification may not necessarily lead to higher-quality building energy simulations, as an individual building may perform differently compared to the average building for a particular building type. For example, an individual office's energy use may more closely resemble that of a school prototype than the office prototype. The end goal of the analysis should be considered when determining whether data should be

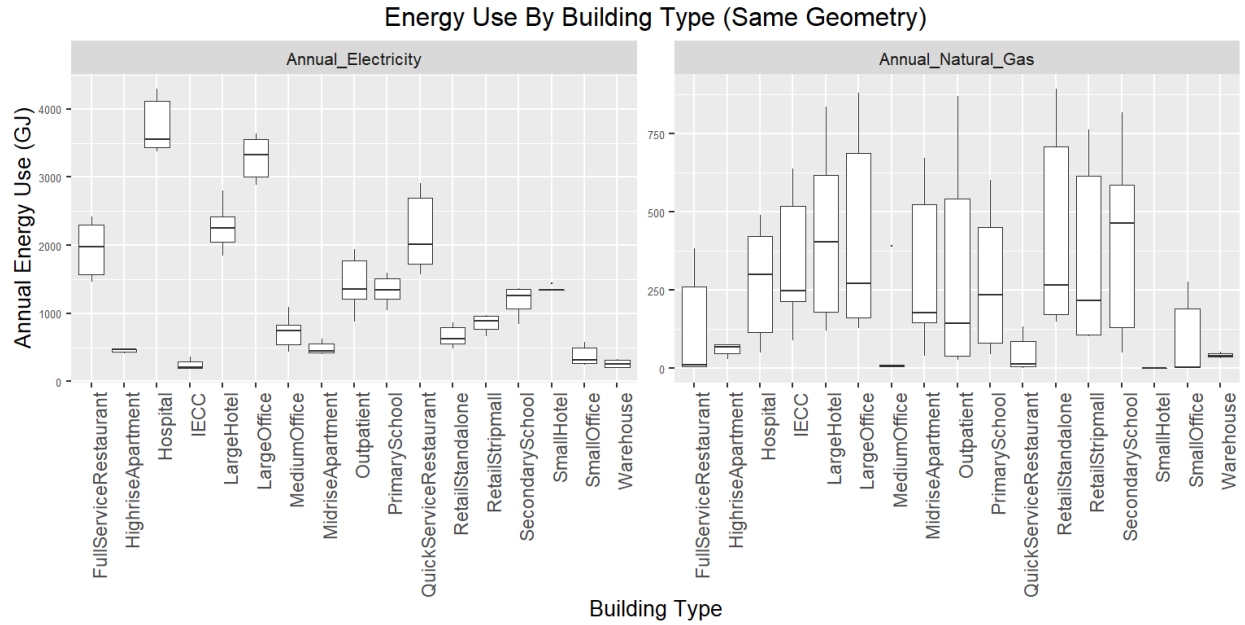


Figure 2.3: Energy Use By Building Type - Building type has a significant impact on both annual electricity use and natural gas use. Building type has a more significant impact on electricity use than natural gas use across building types for climate zone 4A.

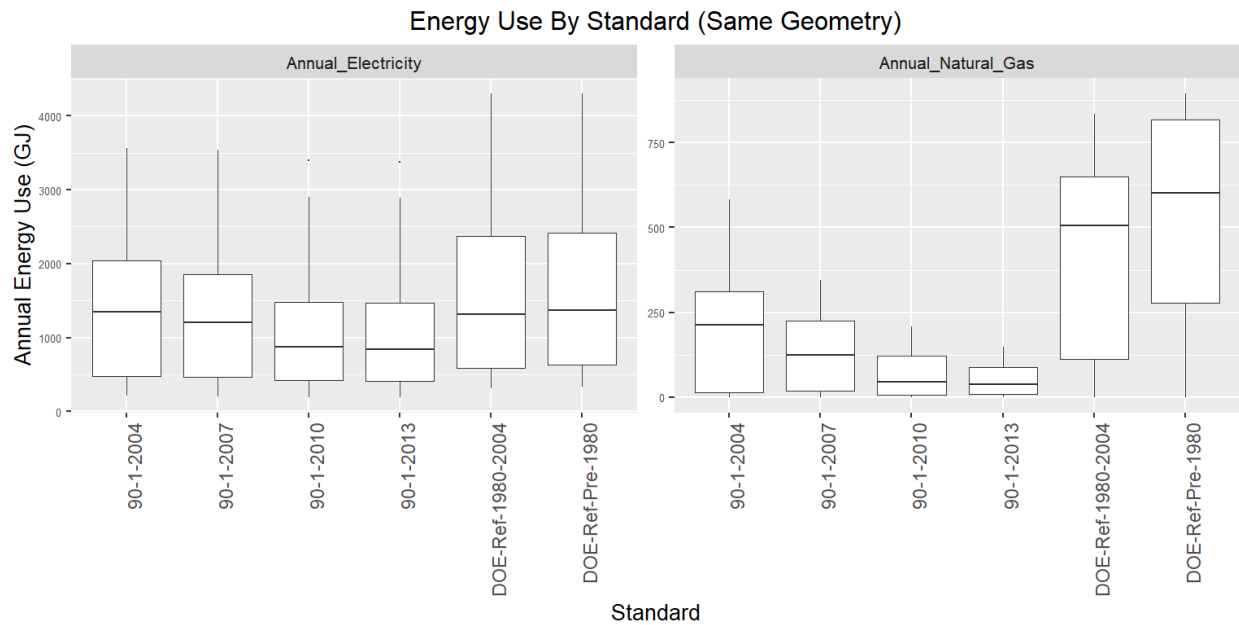


Figure 2.4: Energy Use By Standard - Building standard impacts annual natural gas use more than electricity use in building energy models in climate zone 4A. Older buildings use more energy than newer buildings.

post-processed. Post-processing the building type classification based on sensed or known data may provide the correct building classification but may also increase simulation error in some cases.

The building type classifications were evaluated in two ways: error rates of simulated electricity to actual electricity usage and assigned building type compared to actual building type. Actual building type was found manually by searching each of the 100 buildings. Building electricity data were adjusted using a single annual adjustment factor. The runtime of each method was also considered, as it plays an increasingly important role as the number of buildings grows. Building type classification methodologies based on smart meter data are directly related to my published work [10].

2.2.1 Missing Values

As with any real data, there are often gaps and missing values. To assign building type and vintage by comparing actual 15-minute electricity use to 15-minute prototype building simulation outputs, these missing data must be handled in some way. Various imputation strategies were employed to handle missing data, including omission, autoregressive integrated moving average (ARIMA), and univariate dynamic time warping (DTW). The consequences of using each method were compared.

2.2.2 Euclidean Distance

The first and most straightforward method of comparing the real 15-minute EUI data to the prototype 15-minute EUI data was measuring the Euclidean Distance between the EPB sample and each of the 97 prototype simulation combinations. For the Euclidean Distance comparison, imputation strategies were ignored, as missing points could be omitted with each point being compared directly to the corresponding point from each time series. This resulted in a comparison between however many points were in the EPB sample and the same number of points from each prototype. The prototype and standard combination with the smallest distance between each observation was chosen as the label for that observation.

2.2.3 Dynamic Time Warping

The next method of comparing the EPB data to the prototype combinations was DTW. DTW is a commonly used measure of the similarity between two time series. DTW finds the optimal global alignment between two time series, accounting for temporal distortions. The algorithm optimally maps one time series onto another and, similar to Euclidean Distance, compares each point in one time series to every other point and returns the warped distance. In this way, even if the time series are not exactly in phase, their points are compared and the warping distance remains small. This method may be a good fit for electricity data, where the same patterns may occur at different points throughout the year. As one would expect, this vast comparison is very computationally expensive (quadratic time and space complexity), and many modifications have been made in an attempt to expedite this process. For this analysis, an approximation called FastDTW was used ([48]). A comparison of Euclidean Distance to DTW distance is shown in Figure 2.5. DTW warps to another section of the time series and maps similar queries together, which may result in a better match. DTW cannot be used directly on time series with missing data. The data had to be either omitted or imputed. For this analysis, the missing data were imputed for comparison using the small and large gap strategy previously described.

2.2.4 Machine Learning

The final method used to label building type and standard was using a machine learning (ML) classifier. The ML classifier was constructed in a different way than the previous two methods. Rather than directly comparing the data to the prototypes, this method extracted time and statistics-based features from the data, with the prototypes considered as the labels. The Caret package in R was used to build, tune, and compare these models [32]. Several time-based statistics were extracted from the time series (shown in Table 2.1 below).

These time windows were chosen because they summarize critical structures of the time series. For example, one would expect the EUI of a large office on the weekend to be different than normal and completely different when compared to other building types. Weekly windows were originally used but were removed because they resulted in lower

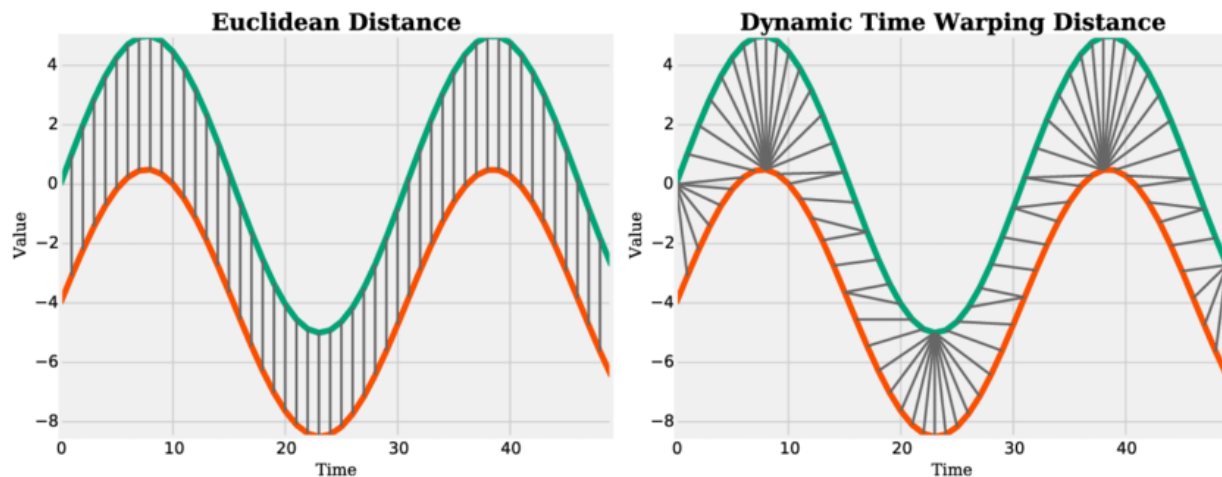


Figure 2.5: Euclidean Distance vs DTW Distance - Euclidean Distance versus dynamic time warping distance [51]. Sine curve is compared to cosine as an example for which Euclidean Distance does not adequately compare two time series due to phase shift.

cross-validation metrics. Three different models were evaluated with a hyperparameter grid search being used for each to compare optimal models for each. These models were k-nearest neighbor (KNN), random forest (RF), and extreme gradient boosting (xgbTree). KNN is a classifier that assesses the distance between a test vector and all training vectors with the label being the vector at which distance is minimized [44]. xgbTree and RF are both decision tree classifiers that recursively partition data based on feature values for which each partition serves as a test for a feature of test data [42]. Boosting (xgbTree) relies on shallow trees for which error is minimized by minimizing bias, while RF uses fully grown decision trees and minimizes error by minimizing variance [13], [17].

As there were 97 different classes with one observation per class, cross-validation could not be done with the raw labels. Instead, the labels were changed to building type only (removing vintage), thereby incorporating at least 3 labeled observations (6 for most) into the training data set which allowed for classic cross-validation to get a rough estimate of what the hyperparameters should be to split the building types. The random forest ultimately had the highest classification accuracy and was the final model used to create the building energy models for the EPB samples. The hyperparameter grid values are shown in Table 2.2 while the cross-validation results are shown in Table 2.3.

2.3 Software

With all building features aggregated, the models can be created using OpenStudio, an open source analysis platform that facilitates integrated whole-building energy analysis [59]. The model is then simulated using EnergyPlus, US Dept. of Energy’s \$100M flagship whole building energy simulation program that can model energy consumption, heating, cooling, ventilation, lighting, plug and process loads, and water use in buildings [57]. This is done at scale using AutoBEMGen [1]; a Python software developed to take a set of building features in a comma separated value format and create an OpenStudio model and EnergyPlus input file. This multi-threaded framework allows many building energy models to be generated much faster than they could be built from scratch. AutoSIM [49] is then used to simulate these EnergyPlus building energy models on HPC resources, distributing the models to

Table 2.1: Machine Learning Windows and Statistics - Time windows and statistics were chosen to retain as much data as possible in during relevant times of the year. Weekends could be a significant differentiator between two building types.

Time Window	Statistics
Monthly	Maximum
Yearly	Mean
Weekends	Median
	Standard Deviation
	Minimum

Table 2.2: Machine Learning Grid Search Hyperparameters - Hyperparameter values used for grid search are shown with optimal hyperparameter values from cross-validation in bold. For more on these metrics, see [13], [17], [44], [50].

Method	Hyperparameter	Value
KNN	Kernel	Rectangular Gaussian Triangular Epanechnikov
	Kmax	30, 40, 50 , 60
RF	Trees	500
	Mtry	2, 125 , 390
	Min Node Size	1
	Split Rule	Gini Extra Trees
xgbTree	N rounds	50, 100 , 150
	Max Depth	1-3
	Eta	0.3 -0.5
	Gamma	0
	Col Sample By Tree	0.6 , 0.8
	Min Child Weight	1
	Subsample	0.5 , 0.75, 1

Table 2.3: Machine Learning Cross Validation Comparison - Cross validation metrics for k-nearest neighbors (KNN), extreme gradient boosting (xgbTree), and random forest (RF). RF had superior mean and median classification accuracy and κ , which compares observed accuracy to expected accuracy, taking into account the chance of randomly classifying correctly.

Method	Median Acc	Mean Acc	Median κ	Mean κ
KNN	78.4%	80.1%	77.1%	78.8%
RF	84.3%	82.2%	83.3%	81.1%
xgbTree	80.3%	81.0%	79.0%	79.7%

each node appropriately and efficiently (based on previous runtime). These software were improved and streamlined in a number of ways, including expanding the framework to scale beyond Chattanooga and updating all code to support the latest versions of OpenStudio and EnergyPlus to introduce new capabilities.

2.4 Model Evaluation

Models were evaluated by comparing simulated 15-minute simulation electricity use to the actual 15-minute use [21]. Normalized mean bias error (NMBE) and coefficient of variation of root mean square error (CVRMSE) were used to evaluate the data. As the real data were somewhat unreliable, several adjustments were made to calculate the accuracy metrics. For some analyses, bias adjustments were made to provide more reliable results for the EPB service area.

2.4.1 3D Visualization

Visualizing the buildings in the EPB service area in a 3D, geospatial manner displays a useful context and perspective to gain valuable insights about the data. Filters and colors can be used to identify groups of building types, buildings with high energy use, buildings with high savings potential, and so on.

The visualization was created by spatially joining the building data (metadata, savings) to a shapefile of the EPB service area around Chattanooga. FME Workbench was then used to convert the shapefile into Cesium 3D Tiles [47], an open specification for streaming massive heterogeneous geospatial datasets[16]. The visualization at the time of the publication can be found online ([Virtual EPB Visualization](#)).

2.5 Energy Conservation Measures

By creating an error-informed digital twin of all buildings in the service area, any technology or policy can be assessed to determine building-specific and utility-scale information regarding the energy, demand, emissions, and financial impact. Several energy conservation

measures (ECMs) were applied, as well as measures that impact demand and other elements related to building performance. These ECMs modified the building energy models created using OpenStudio to change various parameters of the building. Since every ECM were implemented on every single building, the full service area could be analyzed as well as certain regions or districts within the territory. These ECMs are divided into two categories: energy efficiency measures and demand management measures. These ECMs and analyses were previously published [6], [9].

2.5.1 Energy Efficiency

Energy efficiency ECMs focus on lowering energy consumption within a building. These ECMs are mostly efficiency improvements and technology retrofits. They are useful to both a utility and the energy consumer by simply reducing the amount of energy used. The energy efficiency ECMs are shown in Table 2.4

2.5.2 Demand Side Management

Peak demand for a utility is the hour of a month at which energy consumption is at its maximum. The peak is especially important to both utilities and customers, as it is when the utility must use its least efficient and most costly generation assets. Demand-side management ECMs may slightly reduce annual energy usage but focus on the peak hour (even if they increase energy usage in the hours prior to the peak hour). The demand-shaving ECMs are shown in Table 2.5

2.5.3 Emissions

The emissions for the EPB service area are determined directly from the annual energy savings derived from a particular ECM. The Environmental Protection Agency (EPA)'s Emissions and Generation Resource Integrated Database (eGRID) was used to determine various emissions per energy saved [62]. The pollution types and emission rates are shown in Table 2.6.

Table 2.4: Energy Efficiency ECMs - EPB Energy Efficiency ECMs focus on reducing the annual electricity end-use. Overall energy use is not necessarily decreased for the ECMs that are switching to natural gas. Electricity use is also not decreased for the PV ECM, though the PV generation was used to offset end-use and considered electricity savings.

ECM Type	ECM Information
Gas HVAC	Convert electric HVAC system to natural gas system
Gas Water Heater	Convert electric water heater system to natural gas system
Lighting	Reduce lighting power density to 0.85 W/sf
Infiltration	Reduce infiltration by 25%
Insulation	Increase roof insulation to IECC 2012
HVAC Efficiency	Improve HVAC COP to IECC 2012 standard
PV	Maximum possible installed PV per building

Table 2.5: Demand Management ECMs - EPB Demand Side ECMs may actually increase annual electricity use though the peak hour for a utility is typically more important from a cost and emissions standpoint.

ECM Type	ECM Information
Smart Thermostat (4)	Pre-heat/pre-cool by 4°F for peak hour +4 hours
Smart Thermostat (8)	Pre-heat/pre-cool by 8°F for peak hour +8 hours
Smart Water Heater	Turn off heating coil for peak hour

Table 2.6: Chattanooga Emissions Rates - EPB Emissions rates for the EPB service area [62]. For emissions analysis, only annual energy savings were used to calculate emissions reduction; peak demand emissions reduction was not considered.

Pollutant Type	Emission rate (lb/MWh)
NO _x	0.513
SO ₂	0.803
CO ₂	992.271
CH ₄	0.074
N ₂ O	0.015

2.6 Climate Impacts

Similar to ECMs, various scenarios were tested once error-informed urban-scale building energy models had been created. A weather file was necessary for each model simulation, as it initializes many of the physical phenomena occurring during a simulation. While these weather files are typically simulated using any meteorological year (AMY) or typical meteorological year (TMY) files, the buildings can also be simulated using climate model projections to create future meteorological year (FMY) files.

FMY building energy simulation is important, as there are a number of concerns regarding how possible climate trajectories may impact many of the systems on which humans rely. Climate change impacts not only a nation's building codes but also its critical infrastructure, such as the electric grid. Generation utilities must be able to handle higher cooling loads and electrical distributors in order to plan infrastructure deployment (e.g., feeders, substations, transformers) to accommodate weather-induced shifts in building energy loads within their service territories [2].

Weather can significantly impact non-base HVAC loads within a building. The weather files for meteorological years acquired from different providers can impact annual energy use $\pm 7\%$ and monthly heating/cooling loads by $\pm 40\%$ in different building types [11]. However, simply describing it as weather and interchanging a single file masks the complexity of the underlying meteorological variables. While they are dependent upon the primary HVAC system and other variables, changes in dry bulb and/or wet bulb (for a hydronic system) tend to dominate the impacts of HVAC energy use. Figure 2.6 quantifies this impact using DOE's Medium Office Reference Building [18].

This study used data from the Coupled Model Intercomparison Project 5 (CMIP5), where each RCP scenario was projected using an ensemble of climate models. Using an ensemble allows for different initialization parameters and enables variability to be quantified in simulations. While some CMIP5 climate models may share a common lineage - and therefore common biases - the ensemble technique provides a more thorough solution than using any individual model [20]. Variability and uncertainty are beyond the scope of this study. Ensemble names (e.g., r1i1p1, r2i1p1) indicate that ensemble members differ only in

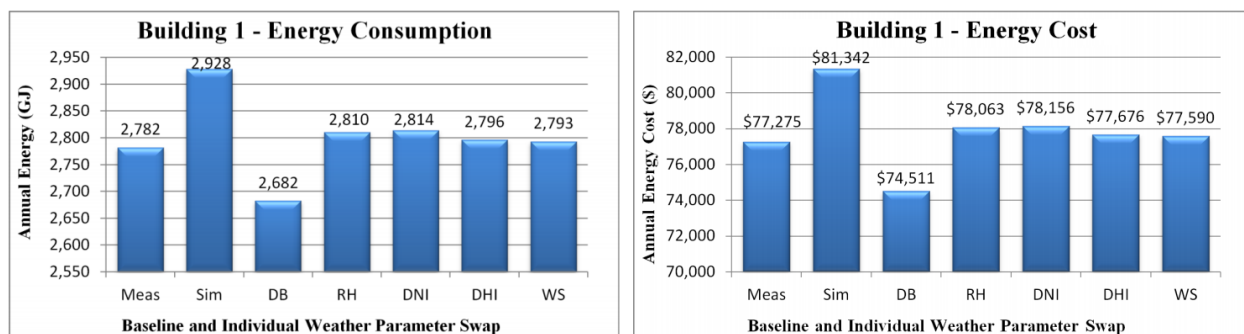


Figure 2.6: Weather Variable Impacts on Building Energy Simulation - Dry bulb temperature, on average, tends to dominate changes in building energy use as shown here for DOE's Medium Office reference model. The next most important meteorological variables tend to be relative humidity, direct normal incident radiation, direct horizontal incident radiation, and wind speed.

their initial conditions (i.e., the model physics are the same for all five ensemble members, but the members were initialized from different initial conditions outside of the control simulation). Hence, the differences between ensemble members represent internal variability. While only one ensemble (r1i1p1) was used in this analysis, the data and methods presented in this paper can be extended to multiple ensembles.

For both of the following analyses, FMY files needed to be created for simulation. The years 2030, 2045, and 2100 were chosen as future years. Assuming the IPCC scenarios as the basis for our analysis, the climate model output data (Table 2.7) were morphed into a format that EnergyPlus could use to simulate buildings. The most impactful meteorological variables (Figure 2.6) were identified for each of the RCPs and future years. The data for each scenario were downloaded in netCDF format and were morphed into the format required by EnergyPlus (epw). Next steps included selecting the area of the Earth to be used for analysis, unit conversions, and downscaling three-hour data to hourly data. This downscaling was performed linearly with the awareness that this simple method was unlikely to precisely represent the variability of some meteorological variables (e.g., solar radiation). EnergyPlus was used to simulate building energy models.

2.6.1 Virtual EPB Climate Impacts

For the EPB service area, only one climate zone had to be used. The climate model grid point closest to Chattanooga was chosen. Combining the four RCP scenarios and the three years resulted in 12 FMY files. To reduce the computational burden, one building of each type was chosen to represent a portion of the service area's buildings, with the table providing the multiplier for utility-scale impacts. A representative sample of each building type in the EPB service area was chosen based on the median area. An exact building geometry from the EPB utility was used. This sample of buildings was then simulated using each FMY file, scaled up by the number of each building type in the service area to obtain a representation of the entire area.

Importantly, since the utility partner provides only electricity (i.e., not natural gas), all building types were assumed to have both electric heating and electric cooling for their HVAC systems. This has the effect of defining an optimistic value for maximum technical adoption

Table 2.7: IPCC Climate Model Parameters - IPCC climate model characteristics used to create Future Meteorological Year (FMY) weather files.

Project	CMIP5
Model	MIR-CGCM3
Modeler	Meteorological Research Institute
Experiment	2.6, 4.5, 6, 8.5
Time Frequency	3 hours
Modeling Realm	atmosphere
Ensemble	r1i1p1
Version	20120119

Table 2.8: IPCC Climate Model Variables - Name and units of meteorological variables in IPCC data.

Variable Long Name	Variable Short Name	Unit
Near-Surface Air Temperature	tas	K
Surface Downwelling Shortwave Radiation	rsds	W/m ²
Surface Diffuse Downwelling Shortwave Radiation	rsdsdiff	W/m ²
Surface Air Pressure	ps	Pa
Near-Surface Specific Humidity	huss	1

potential of energy and demand savings on the utility's electrical distribution network. This climate research related to the EPB service area was previously published [7].

2.6.2 US Commercial Buildings Climate Impacts

The FMY analysis for all commercial buildings in the US was conducted in a similar but more expansive fashion. In this analysis, FMY files were created for all 15 US climate zones. The climate model grid points for this analysis were chosen by proximity to representative cities for each climate zone. The same variables and years as in the FMY EPB study were used. This resulted in 180 FMY files, derived from every combination of 15 climate zones, three years, and four RCPs. A total of 1,410 commercial prototype and reference buildings were derived from valid combinations of 15 climate zones, 16 building types, and approximately six vintages (Table 2.9). These prototype buildings were taken directly from the US Energy Codes Program [56] and are not geometries from actual buildings. As they are not buildings from an urban scale, they may be more somewhat more representative of an actual building of that type, as there is more detail in the model (in areas such as zoning and other building parameters). An annual simulation of each building was executed using EnergyPlus with the building's standard TMY file as well as the 180 FMY files.

Values for the construction area of each prototype building and climate zone for 2003-2007, compiled by Pacific Northwest National Laboratory (PNNL) [29], were used to scale from individual buildings to the national level. This scaling method makes two major assumptions: first, that the commercial prototype building construction area from 2003-2007 is representative of the current commercial building stock distribution in the US, and second, that there are equal numbers of each prototype vintage in the building stock distribution today. While one might assume that there would be more buildings of an older vintage, the distribution of buildings may be expected to shift toward newer buildings as the projections used in this study shift farther into the future. For this reason, a uniform distribution was chosen. The authors are hopeful that future research will circumvent these limiting assumptions by modernizing building type multipliers and vintage distributions from novel sources of urban- to nation-scale building data.

Table 2.9: US Commercial Prototype Building and Climate Zone Parameters. - 1,410 commercial prototype building models were used from all US climate zones, resulting in 16,920 simulations.

Building Types	Vintages	Climate Zones	Representative Cities	Years	RCP Pathways
Full Service Restaurant	Pre-1980	1A	Miami, FL	2012	2.6
High-rise Apartment	1980-2004	2A	Houston, TX	2030	4.5
Hospital	2004	2B	Phoenix, AZ	2045	6
Large Hotel	2007	3A	Atlanta, GA	2100	8.5
Large Office	2010	3B	Las Vegas, NV		
Medium Office	2013	3C	San Fransicso, CA		
Mid-rise Apartment		4A	Baltimore, MD		
Outpatient		4B	Albuquerque, NM		
Primary School		4C	Seattle, WA		
Quick Service Restaurant		5A	Chicago, IL		
Retail Standalone		5B	Denver, CO		
Retail Stripmall		6A	Minneapolis, MN		
Secondary School		6B	Helena, MT		
Small Hotel		7	Duluth, MN		
Small Office		8	Fairbanks, AK		
Warehouse					

Equations 1, 2, and 3 illustrate how multipliers are calculated and used to scale up building simulation energy use from individual buildings to a nation-scale building simulation energy use for the entire US for a given year.

$$M_A = \frac{\sum A_{c03-07}}{A_{T_{year}}} \quad (2.1)$$

$$N_{c03-07} = \frac{A_{c03-07}}{A_p} \quad (2.2)$$

$$E_{T_{year}} = (N_{c03-07} * E_b) * M_A \quad (2.3)$$

- A_{c03-07} - Commercial prototype building area built between 2003-2007
- $A_{T_{year}}$ - Total commercial building area for given year
- M_A - Commercial floorspace multiplier
- A_p - Individual prototype building area
- N_{c03-07} - Number of commercial buildings built between 2003-2007
- E_b - Simulated energy use for individual prototype buildings
- $E_{T_{year}}$ - Total energy for prototype buildings for entire US for given year

The Commercial Buildings Energy Consumption Survey (CBECS) provides commercial building data for different years, with 2012 the most recent. This survey is used to obtain the total commercial floorspace ($A_{T_{year}}$) in the US. CBECS contains both commercial electricity and gas use, which can be compared to simulated energy use from TMY weather data to provide some empirical validation of simulation accuracy [60].

The scaling of energy use simulations for each scenario was evaluated using two additional methods. First, the number of buildings and floorspace were held constant through future years (2030, 2045, 2100) to enable a clearer comparison of increasing energy use from a stable baseline due solely to climate change. Second, a more realistic estimate of energy

use is created by taking into account urban growth and land use changes. This was done by extrapolating historical data for CBECS' number of commercial buildings and floorspace to future years using a linear fit. This simplification introduces error and uncertainty, but enables a more representative picture of future growth for commercial energy use in the US compared to a scenario in which there is no increase in buildings. This work is submitted and in review [8].

2.7 Model America

The scalable data sources and algorithms (described in Section 2.1) were created for modeling all buildings for a utility. However, in order to scale this approach nationwide, an effort to create a model of every U.S. building was identified as a 5-year goal for the Building Energy Modeling team at ORNL under the concept name "Model America".

Creating and simulating building energy models for every building in the US is valuable for many organizations. Considering this is the by far the largest bottom-up UBEM analysis ever performed (more than 125 million buildings considered in all 50 states), the data aggregation methods and HPC strategies can be used as a baseline for future analyses of this scale. The models and simulation results themselves are useful in many different ways as they can be used for simulation-informed analysis throughout the US. An example of this is the integration of the building energy models into urban systems modeling efforts in various regions around the country. Several other use cases for the expansive set of building energy models are shown in the list below.

- Utilities
 - The utility scale analyses described in Section 2.1.
 - * Demand Management
 - * Energy Efficiency
 - * Grid Resilience
 - * Climate Impacts
 - * Cost Savings

- * Rate Structures

- Home/Building Owners

- Non-metropolitan areas may be able to utilize building energy modeling as they are rarely the focus UDEM.
- Automated personal financing can be achieved as models can estimate savings and payback periods of building technologies

- * Leading to increased building technology adoption and advances in building efficiency across the country

- Companies

- Building technology companies may focus their sales and marketing efforts on high impact buildings by estimating the efficacy of their technology in every building model for a region.

Being that the number of buildings for this analysis is so large, expansive data aggregation and compute resources were required. The 2D Footprints for every building were gathered from Microsoft's dataset [33]. A partnership with Google was formed to provide some of the critical spatial and temporal data for the buildings. The building heights were provided by Google assigned to each building type using 30 meter interpolated to 5 meter spatial height resolution. These footprint and the height provide the physical shape of each building. Inevitably, missing data issues arise with this number of buildings. For this case, missing height data was interpolated using the median of the nearest 20 buildings. An estimate of the year built for each building was provided by Google using a pixel adjustment technique on image data back to 1985. This estimate was used to be classify buildings into the prototype vintage bins (previously referenced in Table 2.9). Missing year built building data was assigned according to the distribution of commercial building year constructed from CBECS [60]. The climate zone of the building determines many of the parameters of a building model (insulation, etc.). The climate zone for each building was assigned based on the geographical location of the building.

The building prototype classification is based on a building’s physical characteristics as well building archetype construction weights. The physical characteristics used were building height and building 2D area with boundary conditions for the bins derived from the archetype’s physical traits [18]. These bins are shown in Table 2.10. Within each bin, the buildings were assigned according to commercial construction weights from the year 2003 to 2007 [29]. It was assumed that these construction weights are representative of the current built environment. As these only represent commercial buildings, residential buildings were factored in using a overall residential percentage of 95.88% residential to 4.12% commercial.

With all of these aggregated features for each of the more than 125 million buildings, the models can be generated and simulated. This is accomplished using THETA, a supercomputer at Argonne National Laboratory. The HPC process on THETA is described below.

1. Pre-processing of the building data determines how to effectively distribute the tasks to THETA.
2. The building data and AutoBEMGen software [1] are sent to each node of THETA used for a given run using AutoSIM [49].
3. The buildings are generated in OpenStudio [59] and simulated in EnergyPlus [57] on each node.
4. The models and simulation results were aggregated and copied to permanent storage.
5. Models are reformatted from runs to state directories in order to facilitate staging future analysis.
6. Models are zipped by county for public sharing.

Building models separated by county could then be publicly shared and available for download (This work will be submitted for publication shortly after the submission of this document [35])

Table 2.10: Building Type Assignment Methodology For US Buildings were categorized by their physical characteristics determined from the prototype for each building type. Buildings were further categorized within each physical bin based on prototype construction floorarea weights.

Physical Bin Parameters	Prototype Buildings
< 40ft Height	FullServiceRestaurant
< 6,000ft ² (2D) Area	QuickServiceRestaurant
	SmallOffice
	Residential
< 40ft Height	RetailStandalone
> 6,000ft ² (2D) Area	RetailStripmall
< 15,000ft ² (2D) Area	Outpatient
< 40ft Height	PrimarySchool
> 15,000ft ² (2D) Area	SecondarySchool
	Warehouse
> 40ft Height	MediumOffice
< 80ft Height	SmallHotel
< 18,000ft ² (2D) Area	MidriseApartment
> 40ft Height	Hospital
< 80ft Height	LargeHotel
> 18,000ft ² (2D) Area	
> 80ft Height	HighriseApartment
	LargeOffice

Chapter 3

Results and Discussion

3.1 Virtual EPB

3.1.1 Building Type Assignment

For this analysis, three different datasets were created with building type and standard being classified from one of each of the methods. The models were then simulated using EnergyPlus and compared to the actual data to obtain accuracy metrics. The methods were compared in two different ways; error rates of simulated electricity to actual electricity usage and assigned building type compared to actual building type. The actual building type was found using a manual search. Building electricity data was adjusted using a single, annual adjustment factor. Both of these metrics may be valuable depending on the goal of an analysis.

Runtime was also considered and would become an increasingly important factor as these results are scaled beyond 100 buildings. The Euclidean Distance calculation was the fastest with a runtime 25x faster than dynamic time warping which must compare all sets of points though this time can be reduced with tuning. The random forest was took only a few seconds to train as it only needed to be trained on 97 samples.

Quantitative Summary

Coefficient of Variation of Root-Mean Squared Error (CVRMSE) is a quantitative metric used for building energy modeling that measures uncertainty in the model compared to real

data. CVRMSE is calculated by normalizing the Root-Mean Squared Error (RMSE) by the average value of the dependent variable. By nature, RMSE can range between 0 and infinity and since the errors are squared before they are averaged, high weights are given to larger errors. This results in larger errors than other metrics such as Mean Absolute Error (MAE). The equation for CVRMSE is shown in equation 3.1 for which Y_i represents measured data, $\bar{Y}$ represents simulated data, and N represents the number of data points for an annual 15-minute simulation (35,040).

$$CVRMSE = \frac{1}{\bar{Y}} \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y})^2}{N}} \quad (3.1)$$

CVRMSE is often computed for building energy models on monthly or hourly data for a year, whereas these numbers are computed for 15-minute data over a year. Also, missing data for this calculation was addressed by omitting "NA" values and the aforementioned imputation strategies. The performance metrics based on comparison of the resulting BEM and measured data shown in Table 3.1.

Qualitative Summary

Qualitative results are sorted into three categories; direct accuracy, general accuracy, and commercial accuracy. Direct accuracy was determined by comparing the exact prediction with the actual building type. This was difficult to classify in some instances as a church or a car dealership could not be directly classified into building prototypes. General accuracy corrects this issue slightly by classifying actual buildings into their closest representative prototype building as well as combining categories such as small, medium and large office into a general "office" label. The final category is commercial accuracy which is simply residential (detached) or commercial (other). The accuracy values are shown in Table 3.2.

While accuracy is a reasonable metric, it can be a bit misleading given the dataset was comprised of about 80% residential detached houses. The ability of these classifiers to differentiate between residential buildings and commercial buildings is important as it has a large impact on building properties and energy use. Though this is a multi-class problem with 17 different buildings types, an estimate of the binary commercial classification quality

Table 3.1: Building Type Classification Quantitative Results - Each of the three methods have similar CVRMSE values. The random forest has an edge primarily because the maximum CVRMSE for that method is much lower than the other two methods, bringing the mean down. American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) considers <15% monthly and <30% CVRMSE as "investment grade" ([5])

Method	Min	Median	Mean	Max
RF	.7%	38.6%	44.1%	206%
Euc	.5%	38.5%	48.6%	545%
DTW	.5%	38.7%	49.1%	560%

can be obtained by simplifying the actual building type and the prediction to commercial or residential. As the simplification to commercial vs residential is done in a post-processing step, a single decision threshold is used.

Sensitivity (true positive rate - method predicts commercial and building is commercial) and specificity (true negative rate - method predicts residential and building is residential) are useful metrics for binary classification exercises as they highlight class imbalance issues. The sensitivity and specificity values are shown in Table 3.3. While the Euclidean classifier had the best direct accuracy of the three methods (Table 3.2), it's over-prediction of residential in a highly residential dataset is problematic as it shows the classifier does not have the ability to separate these important building distinctions. In contrast, the lower direct accuracy of the Random Forest (Table 3.2) may be caused by certain real commercial buildings in the dataset that behave more closely to a different commercial prototype, likely making these predictions more representative than a residential classification.

It should be noted that no post-processing was done for any of these methods. If 15-minute electricity is available, often billing rates may be available and may be used to change a building classification with awareness that assigning the correct building type may not lead to a more representative energy simulation.

Confidence

All of these methods utilize some sort of distance metric from Euclidean Distance, to warping distance (DTW), and class probability (RF). These distances can be viewed as a pseudo-confidence factor with the premises with the smallest distance or highest similarity to the prototypes being the labels of the highest confidence. This confidence level allows the utility to determine to what degree they trust certain labels as well as if some buildings need to be labeled in another way if the distance is out of the normal range for that method. For visualization, the distances for each method were scaled between 0 and 1 and are shown in Figure 3.1.

For both the Euclidean Distance and DTW distance methods, the majority of prototype-actual building matches has a small distance that increases for several observations. The observations with the smallest distance would be the predictions with the highest confidence.

Table 3.2: Building Type Classification Qualitative Results - While the Euclidean Distance classifier had the highest accuracy, it was primarily due to the number of residential predictions coupled with the amount of residential buildings in the sample.

Method	Direct	General	Commercial
RF	62%	63%	78%
Euc	80%	80%	81%
DTW	71%	71%	77%

Table 3.3: Building Type Classification Sensitivity and Specificity - Random Forest was best at differentiating commercial vs residential buildings while Euclidean Distance over-predicted residential.

Method	Commercial Sensitivity	Commercial Specificity
RF	78.9%	78.3%
Euc	0.05%	100%
DTW	36.8%	87.8%

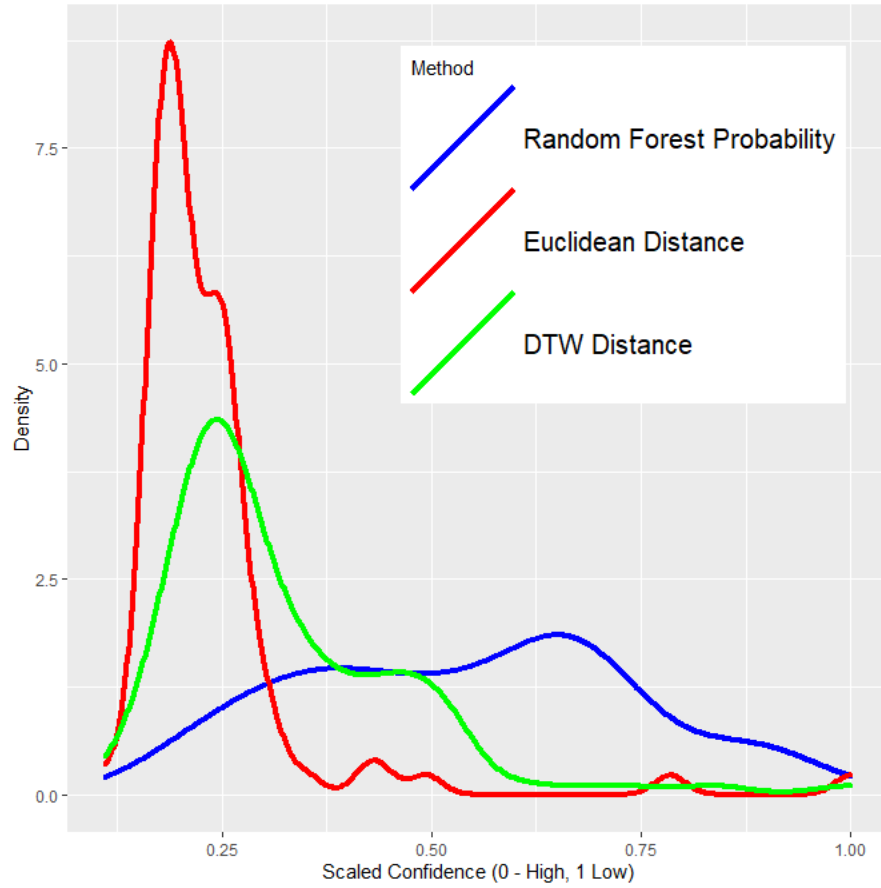


Figure 3.1: Building Type Classification Confidence Distributions - The distance-based classification methods have clear drop-off and reduction with the majority of distances being smaller and few large distances. The RF shape of the RF classifier probability is less helpful for confidence purposes as it lacks a clear cutoff point.

The shapes of these two methods is different than the random forest because of the way they are calculated. For this dataset, the maximum random forest probability was 37.4%. The reason the percent was this low is because of the similarity between prototype building vintages within building type bins which decrease the maximum probability per class. For example, the EUI signature of a 2010 small office may be similar to a 2007 small office; leading to a split in probability voting. A confidence could still be used for this method, but one would have to consider the top probability classes to ensure a high-confidence prediction.

If one was to filter the dataset to only predictions below the mean, and below the first quartile distance for each method, the error rates would be expected to decrease. These scenarios are shown in Tables 3.4 and 3.5. The increased filtering does limit the number of buildings in each; taking the average to 45 buildings for the mean filter and 20 buildings for the quartile filter.

These results are generally as expected with a decrease in CVRMSE for the Euclidean Distance and DTW methods but the CVRMSE increases for the random forest method. This leads one to believe that the random forest probability may not be as effective as measuring confidence as the distance methods or building type probabilities across vintages need to be included.

3.1.2 Energy Conservation Measures

Urban-scale building energy modeling can offer increasingly compelling capabilities for assessing demand-savings opportunities or evaluation of new business models for utilities, independent energy savings estimates for building owners, or actionable roadmaps for cities' sustainability plans. In this study, these capabilities are summarized for the number of buildings that could benefit from the technologies discussed. For each measure, there were a number of buildings in which the measure resulted in an increase in energy or demand. These buildings were omitted from the measure-aggregated results reported. Aggregated, simulated electricity and demand savings as well as derivatives of these savings are shown in Figures in the following sections.

Table 3.4: Results of Classifications Greater than the Mean Confidence For Each Method - The mean and median CVMRSE improves for the Euclidean and DTW methods but worsens for the RF compared to Table 3.1

Method	Min	Median	Mean	Max	Acc
RF	18.7%	45.8%	51.8%	138%	71%
Euc	5.5%	35.2%	37.5%	78.4%	97%
DTW	20.2%	35.2%	41.3%	206%	87%

Table 3.5: Results of Classifications Greater than the first quarile Confidence For Each Method - Filtering the data to even higher confidnce level, the mean and median CVMRSE improves for the Euclidean and DTW methods but worsens for the RF compared to Table 3.4.

Method	Min	Median	Mean	Max	Acc
RF	18.7%	48.1%	57.2%	138%	59%
Euc	5.5%	29.9%	31.6%	66.8%	94%
DTW	20.2%	29.9%	32.4%	55.4%	100%

Demand Savings

Smart thermostats, for pre-conditioning buildings as thermal batteries prior to peak hour demand, is among the traditional demand management technologies that may be cost effective for utility programs to implement. While utility-managed thermostats can lower energy, demand, emissions, and costs for the utility, these are often passed on, either directly or indirectly, to the program participant or rate payer. Additionally, technologies such as smart thermostat settings that could affect comfort always allow setpoint override so as not to participate at that time. Figure 3.2 shows the wide range of potential demand savings for over 100,000 buildings under a maximum technical adoption scenarios, which the percentage reported is the percent of kW during the building's peak that could be reduced through pre-conditioning.

Figure 3.3 demonstrates the power of building-specific, bottom-up aggregation to allow visual analytics of demand savings by vintage or building type. In these figures, demand reduction is instead a percentage of kW reduction for the entire service territory. As an example, March and November with vintages of 90.1-2010 and 90.1-2013 are relatively high due to the lower energy demand during those shoulder months combined with the fact that those are the two most popular vintages of buildings in the utility's service territory.

Smart thermostat savings opportunities grouped by building type explain the aggregate results even more thoroughly. March and November have significantly more savings opportunities in the medium office of which there are many in the EPB service area. The other building type with significant demand savings potential is the retail stripmall. Switching a building's water heater from electric to natural gas or other fuel type is another strategy that could reduce peak electric demand. This study indicates the greatest savings opportunity in the utility's service territory for such a strategy in April and October. These savings are shown in Figure 3.4.

HVAC systems can also be changed from electricity to natural gas for heating. In fact, there are dual-fuel systems that offer additional resilience with the potential for a utility-controlled signal to request the equipment dynamically swap between electricity or natural gas for heating. Such a demand management strategy will be most effective in the winter

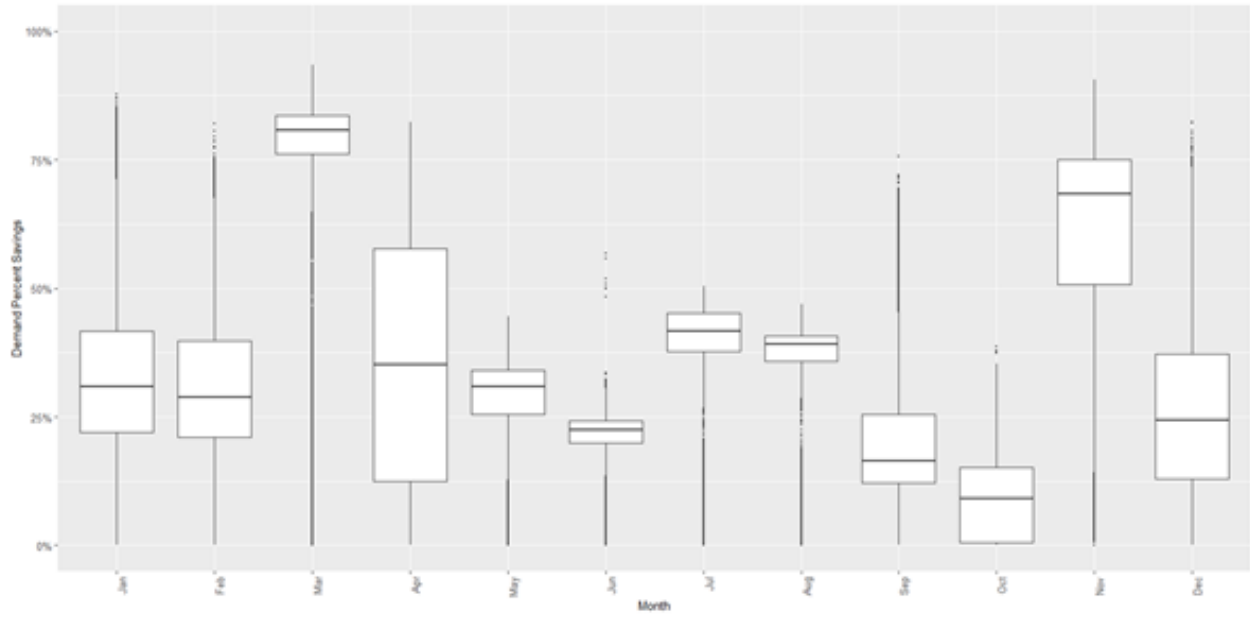


Figure 3.2: Smart Thermostat Monthly Demand Savings - Smart thermostats with utility-signaled 2.2 °C of building pre-conditioning 2 h before peak hour has the potential to save an average of 27% of a building's demand, but this sample of 101,082 buildings varies from 0–93% by individual building and time of year.

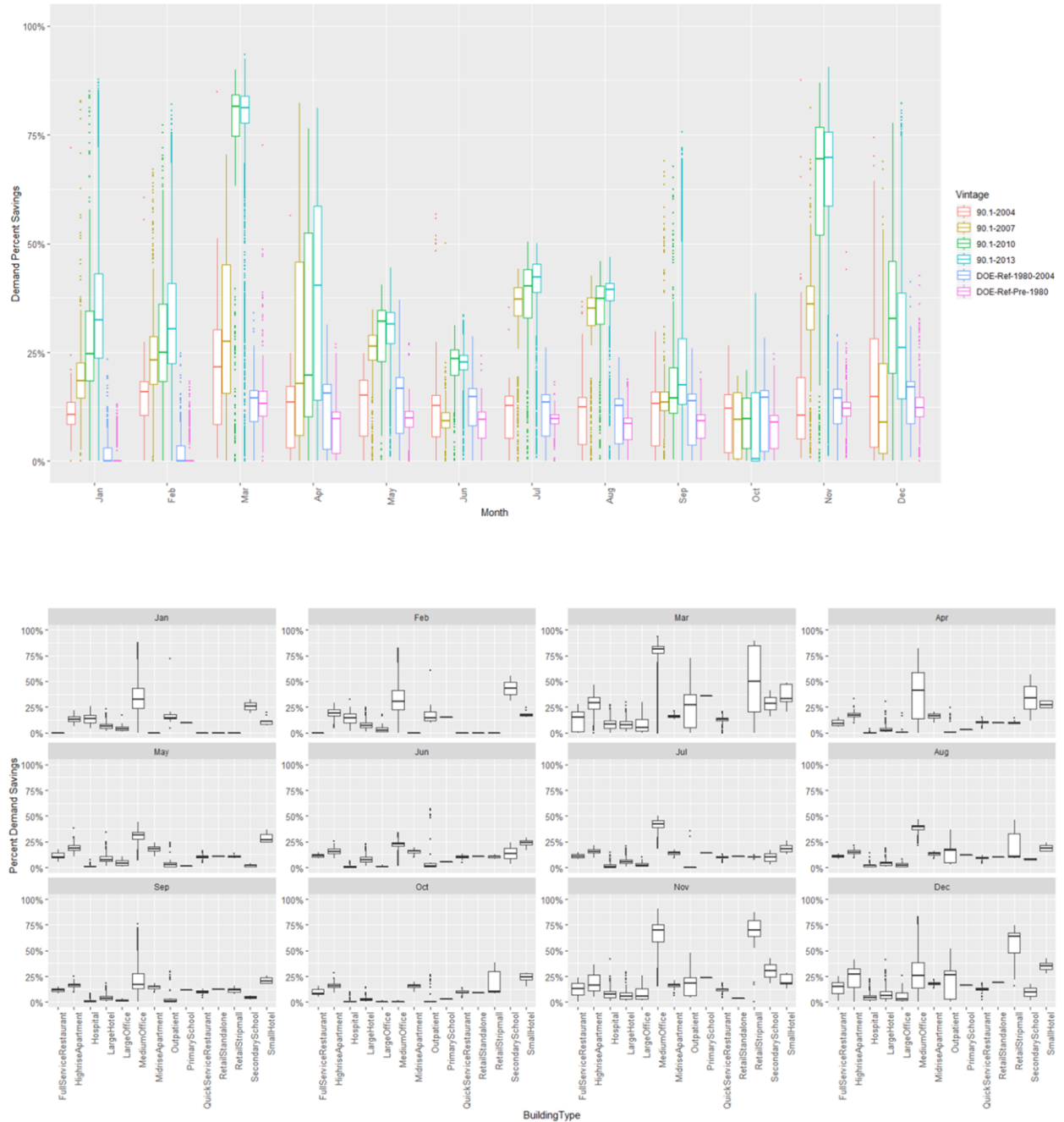


Figure 3.3: Smart Thermostat Monthly Demand Savings By Vintage and Building Type - Potential demand reduction achievable with 2.2 °C pre-conditioning by building vintage and building type varies from 0 to 93% of peak demand.

months. Results, (Figure 3.5), indicate that older buildings have the greatest average savings potential throughout the winter and even have some savings opportunity in the summer. Newer buildings have close to zero savings opportunity in the summer and much lower savings opportunities in the winter months with greater savings potential from older commercial vintages (90.1-2004). While not shown here, the study found Quickservice restaurants to be the only building type that consistently showed potential demand savings from this strategy during summer months.

Energy Savings

While demand management remains one of the biggest opportunities for today's utilities, traditional energy efficiency remains one of the most cost-effective opportunities for long-term savings for building owners. In the utility's territory, if the system-wide annual bill was shared equally among all buildings and cost USD 5 per year, USD 1 would be for demand and USD 4 for energy. The technologies shown are among the most common and cost-effective building energy efficiency measures considered. Annual electricity savings by vintage are shown in Figure 3.6. It should be noted that swapping the water heater or HVAC (two columns on the left) results in higher electricity energy savings by shifting related costs to other fuel types. This study shows that the oldest buildings have the most electricity savings potential. While not shown here, preliminary results from the study indicate that annual electricity savings are greatest for newer vintages when considering only commercial buildings.

From a total energy and cost perspective, the traditional energy efficiency measures of a more efficient HVAC (typically at end-of-life), reducing infiltration (sealing leaks between the indoors and outdoors), adding insulation (further reducing conductive heat transfer), and swapping to more efficient lighting technologies, often are at the top of most building efficiency discussions.

The maximum adoptable PV potential for the EPB service area is shown in Figure 3.7. This represents the PV generation if every roof in the EPB service area is 70% covered with PV with a cell efficiency of 15% and an inverter efficiency of 98%. This generated electricity can be viewed as savings with buildings that generate more than the building used resulting

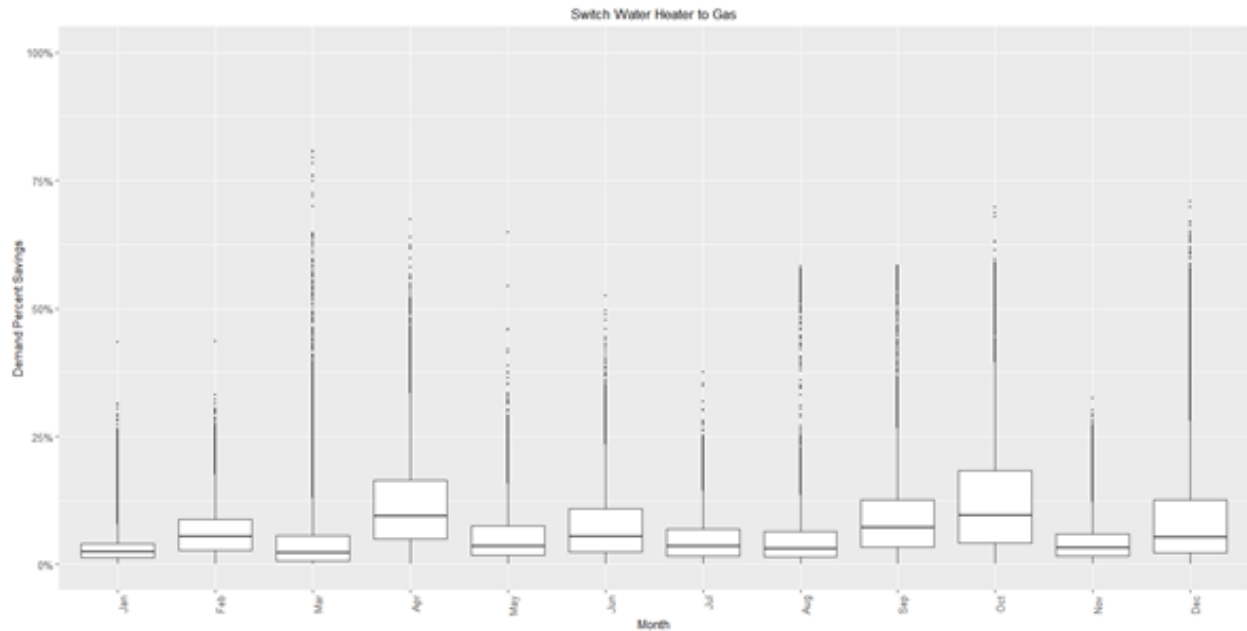


Figure 3.4: Gas Water Heater Monthly Demand Savings - Changing water heaters from electricity to natural gas can relieve peak electric demand. This study found that a building's demand could be reduced as much as 80% but averages approximately 5% of building peak.

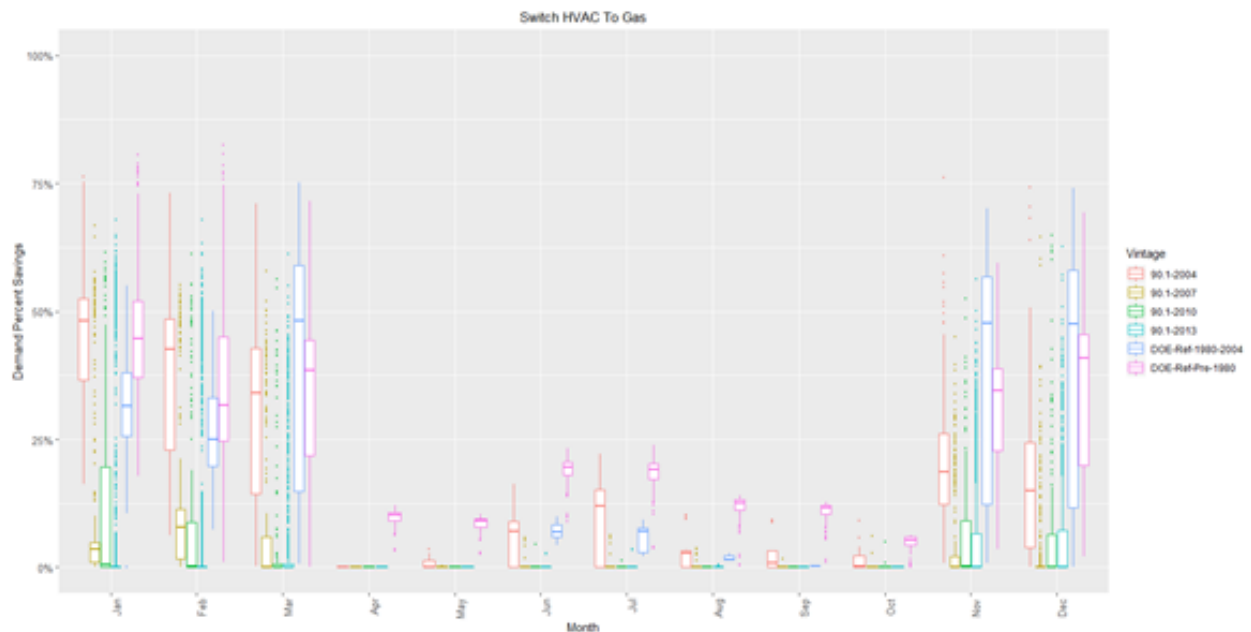


Figure 3.5: Gas HVAC Monthly Demand Savings - Distributions of potential demand reduction for a building, broken down by month and building vintage, achievable with swapping HVAC equipment from electricity to natural gas or dual-fuel equipment with a utility signal to reduce electricity use during peak hour.

in negative values (turning the meter backward). PV installation results in demand savings for some months when the peak hour is during daylight, but does not contribute to demand reduction when the peak hour is before sunrise or after sunset; typically in the winter months.

Emissions Savings

Long-term environmental impact of the building stock may be better considered via greenhouse gas emissions required for the creation and operation of a building. Many activities relevant to urban-scale energy modeling are in service of individual cities defining sustainability plans with activities for curbing emissions of buildings and vehicles. In this study, annual emissions savings were calculated directly from annual electricity savings using EPA's eGRID [62] and thus identical except for lbs/MWh scaling factor and resulting emissions unit. In reality, peak demand electricity generation often has higher emissions than typical generation so peak demand shaving would result in higher values of emissions savings. This is not accounted for in this analysis. Changing from electric to gas ECMs is also not shown in the emissions figure, as the total savings from reducing electricity savings would not be realized as gas would entail emissions as well. As all emission plots are similar in shape with a different scale, only CO₂ emissions are shown in Figure 3.8.

Cost Savings

While demand, energy, and emissions vary considerably in relative importance by stakeholder, almost all stakeholders consider potential cost savings. In these results, retail-rate electricity cost savings were calculated using the US national average for 2015 of USD 0.1041/kWh and USD 10.5/kW. Since residential buildings largely do not elect a demand-sensitive (e.g., time-of-use) pricing structure, these results show limited cost savings for demand response compared to annual electricity savings. The utility-wide total potential retail cost savings for both electricity and demand of each technology is shown in Figure 3.9. It should be noted that not all savings could be realized for the switch from electricity to gas as the additional cost of gas, and concomitant potential loss of revenue for the utility, is not considered here.

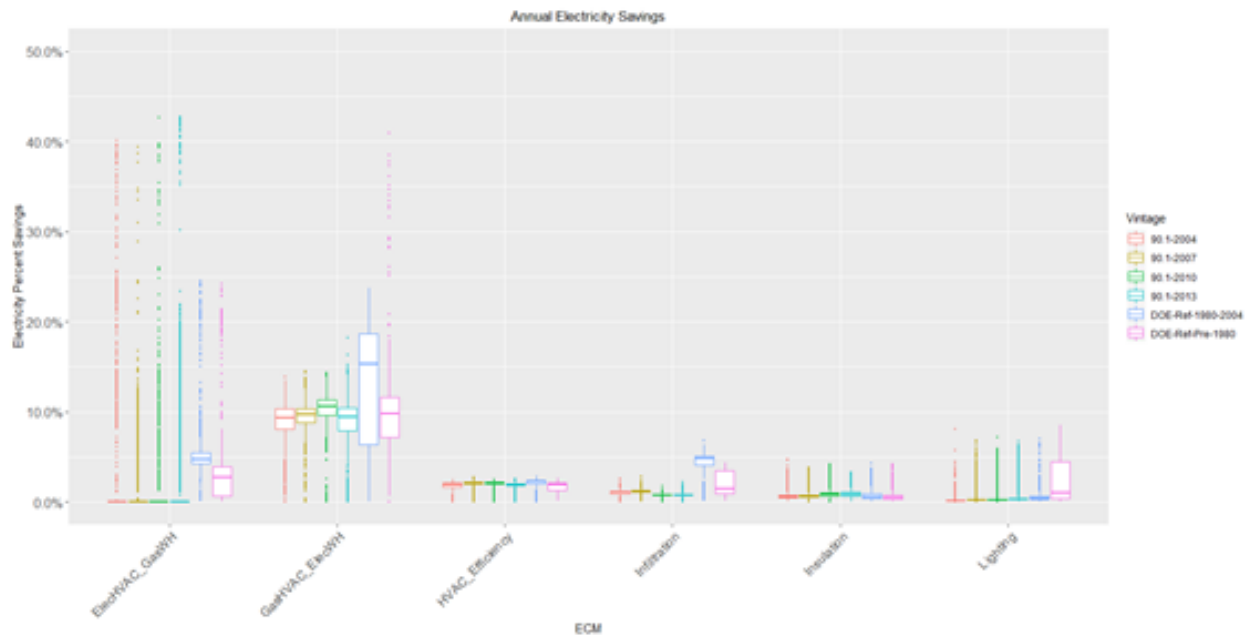


Figure 3.6: Annual Electricity Savings For Six ECMs - Distribution of potential annual electricity savings for two fuel-switching technologies (water heater and HVAC), and four traditional energy efficiency measures.

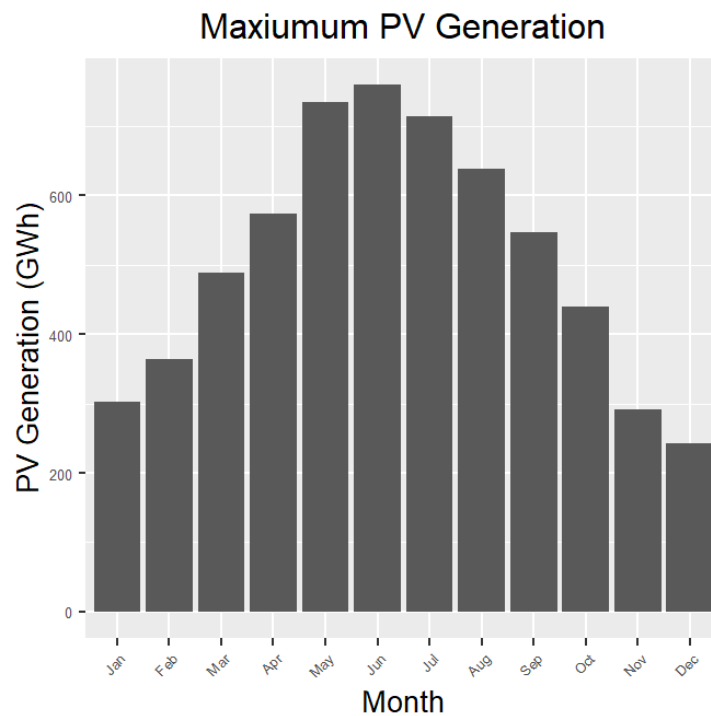


Figure 3.7: EPB PV Generation - The total PV potential in the EPB service area relates to the length of the day, with maximum PV generation in July when days are longest and minimum PV generation in December when days are shortest.

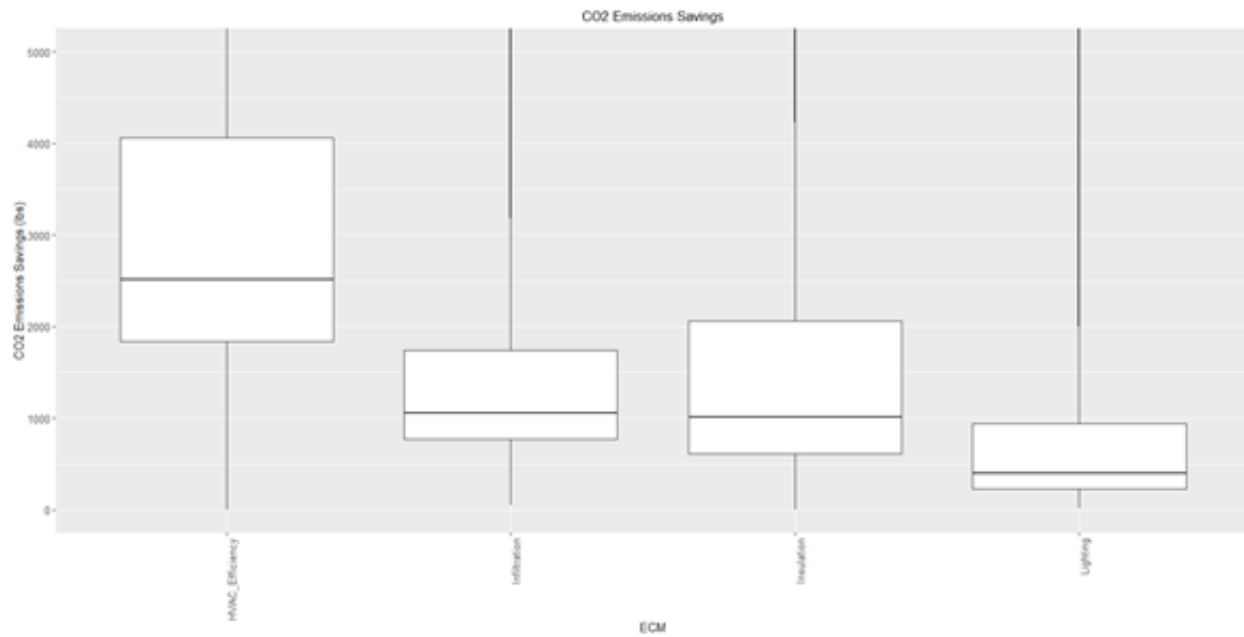


Figure 3.8: CO₂ Emissions Savings for Four ECMs - Distribution across buildings in the utility's service area of operational emissions savings for four traditional energy efficiency measures.

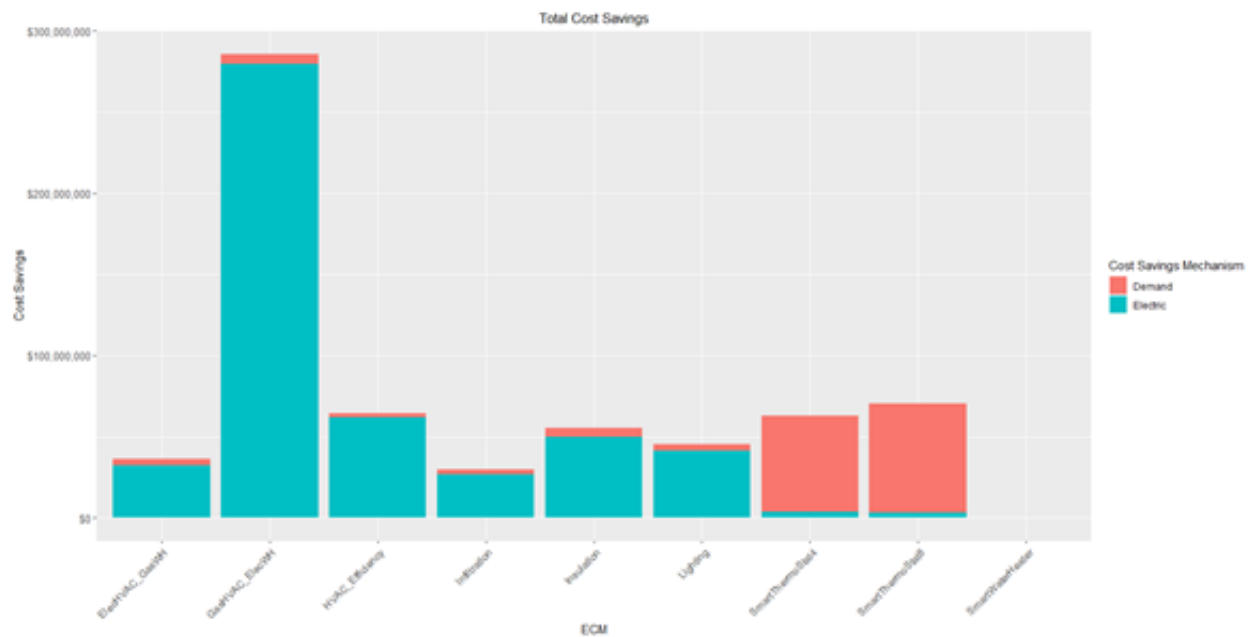


Figure 3.9: Total Cost Savings By Total Electricity and Demand - Combined utility-scale energy and demand annual retail-rate electricity cost savings.

While Figures 3.9 and 3.10 constitute business intelligence that can lead to long-term program formulation and utility activities, the domination of individual building types at utility-scale can obfuscate the short-term, per-building cost effectiveness for deployment; Figure 3.11 normalizes the demand and energy potential cost savings, respectively, while normalizing for number of buildings to achieve the savings that could be seen at an individual building. In these results, it may be interesting to note reversal of trends from the totals. For example, switching to a gas water heater is actually a more effective cost reduction than switching to gas HVAC, while the totals indicated the opposite. High-rise apartments appear to be excellent candidates for demand-related cost savings for several measures. When considering energy-related cost savings, switching to a gas water heater has a significant amount of savings in large hotel buildings as well as hospitals. Lighting efficiency is estimated to have the greatest average potential savings opportunity for quick-service restaurants, outpatients, and full-service restaurants. However, since there are particular building types that lack a significant number of actual buildings in the utility's service area, these averages may be biased.

In addition to demand- and energy-related cost savings, many older buildings may be reaching end-of-life for existing equipment or ready for a retrofit to modernize/upgrade the building, further increasing the timeliness and likelihood of deployment for energy efficient building technologies. Figure 3.12 shows similar average per-building potential cost savings for demand and energy, respectively, broken down by vintage.

Generally, and perhaps counter-intuitively, older buildings are estimated to have lower demand savings potential for most technologies other than gas HVAC swapout for DOE-Ref-Pre-1980. In contrast to previous results, there are some cost savings for the smart water heater ECM, the DOE-Ref-1980-2004 vintage in particular had savings.

3.1.3 3D Visualization

The 3D visualization is shown in Figure 3.13. An initial view of the downtown area of the city of Chattanooga provides a good starting point for with varied building heights providing depth to the image. Buildings can be colored by various parameters including EUI, building

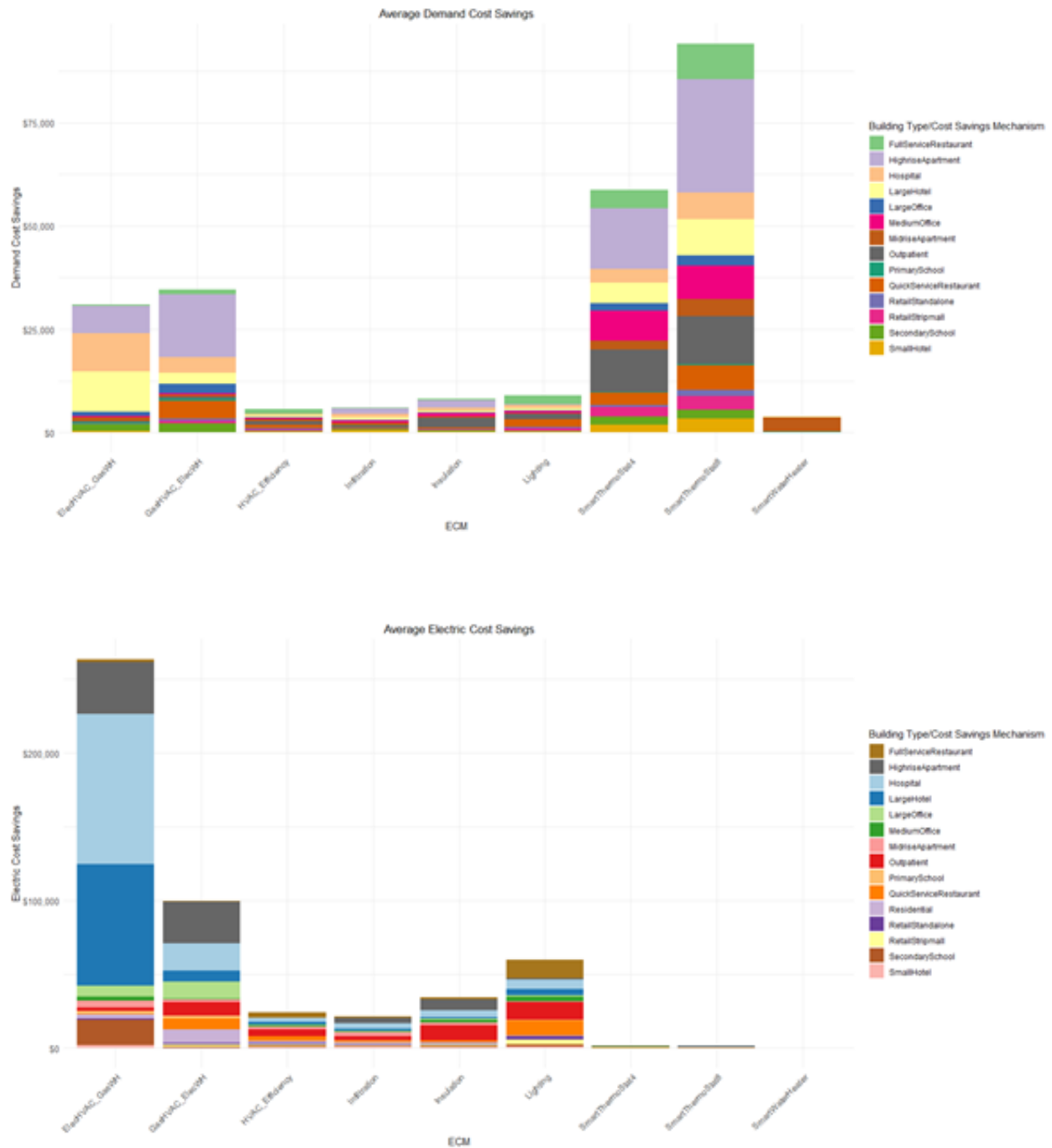


Figure 3.11: Average Demand and Electricity Savings By Building Type - Average demand-related electricity cost savings by building type can indicate cost-conscious opportunities for short-term wins in demand management. Average energy-related electricity cost savings by building type are on average 4x higher than demand-related savings and typically offer the best cost savings for building owners.

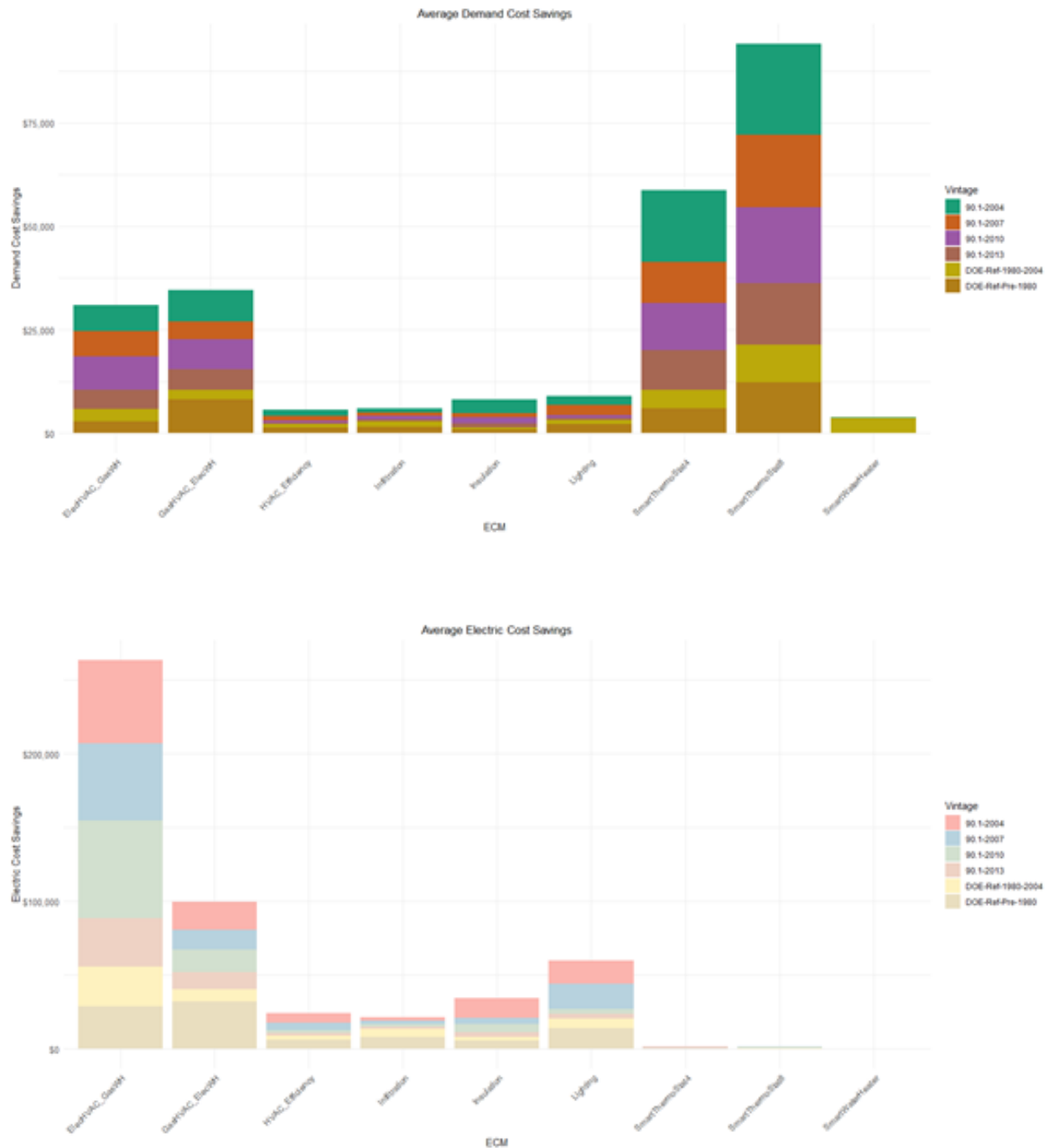


Figure 3.12: Total Demand and Electricity Savings By Vintage - Average demand-related electricity cost savings by building vintage can indicate cost-conscious opportunities for short-term wins in demand management. Average energy-related electricity cost savings potential by building vintage indicates significant savings opportunities. By combining building type and vintage for energy and demand, an initial estimate of cost savings can be used to determine if purchase and installation of these technologies may be cost-effective.

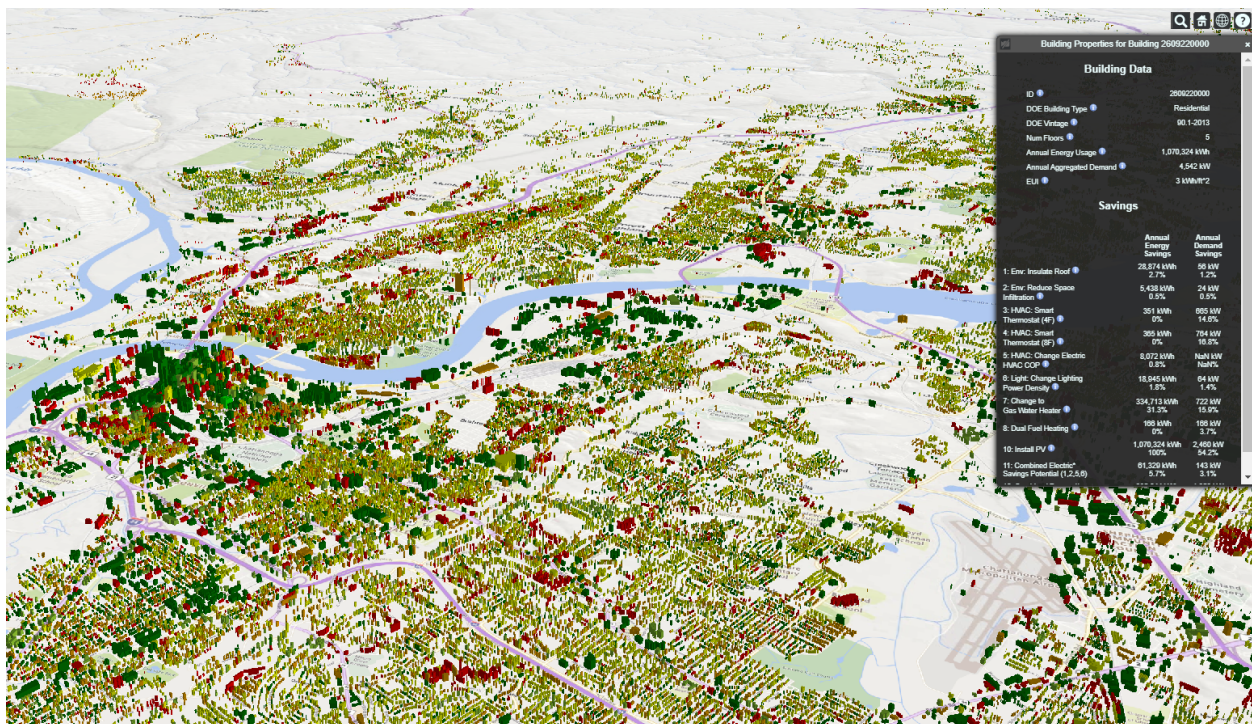
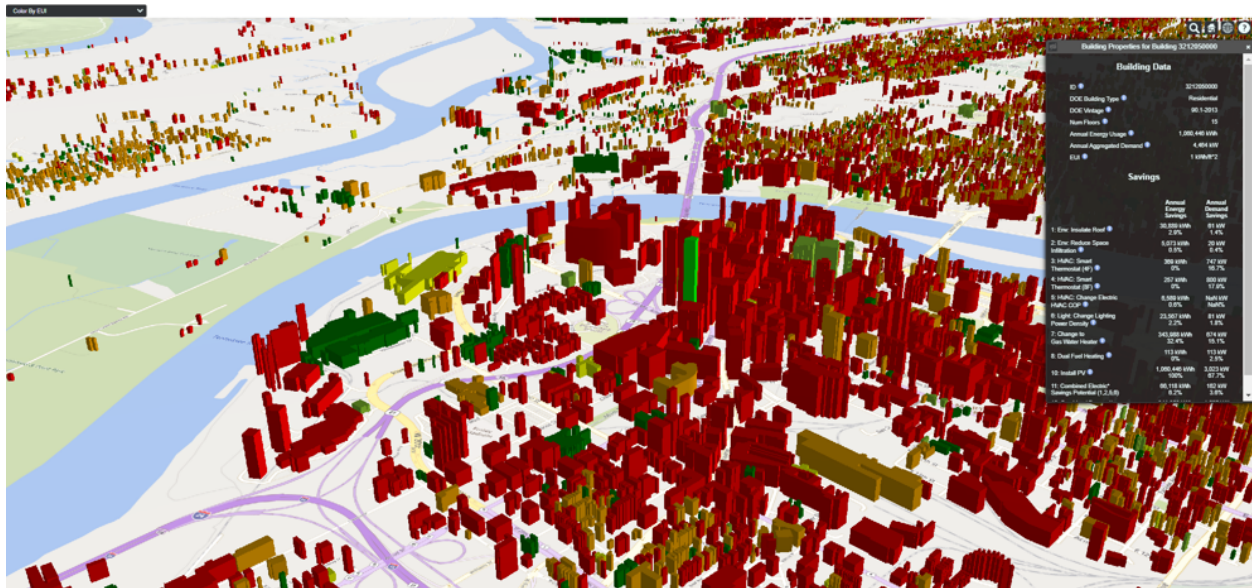


Figure 3.13: Cesium EPB 3D Visualizations - Geospatial visualization of buildings provides insights not readily available from visualizations such as box-and-whisker plots [16]. The different scales and granularity that can be attained using these visualizations are useful in utility-scale BEM. ([Virtual EPB Visualization](#))

type, vintage, savings potential, and more. Clicking on an individual building provides a drop-down table with building characteristics as well as building energy simulation data.

3.1.4 Virtual EPB Climate Impacts

Weather files

An initial investigation into the weather files themselves allows for an intuitive understanding underlying the building simulation results. The first observation that is apparent is the difference between average dry bulb temperature between 2015, TMY, and the climate scenarios. The baseline files temperature and pressure is significantly higher than the RCP scenarios. This difference is likely explained by how coarse the grid points are for this batch of climate models. Rather than choosing an exact location, the grid point with the location that was closest to the coordinates of Chattanooga was selected. For this reason, total energy values are shown as the different RCPs and years may still be effectively compared. It also seems that 2015 was a significantly warmer year than the TMY for Chattanooga.

Comparing only the scenarios across the years in (figure 3.14), the temperature values make sense and the trend is apparent. For the scenario with the greatest mitigation (RCP 2.6), the dry bulb temperature remains relatively constant while the dew point decreases slightly on average by 2100. For the highest radiative forcing scenario (RCP 8.5), the average temperatures increase significantly. There is a greater standard deviation as years go farther into the future for most cases other than for the RCP 8.5 scenario in which the 2030 year has the greatest standard deviation. This case is interesting as the temperatures are also significantly lower than the other 2030 scenarios which is unexpected given the increase in radiative forcing.

Energy Use

The simulated energy use results are shown for the EPB area in (Figure 3.15) in GWh. While TMY (typical) and 2015 (actual) data are provided for completeness, discontinuities between these and future years (IPCC) should be disregarded; only 2030, 2045, and 2100 are from the same location and model source for direct comparison. It is important to note

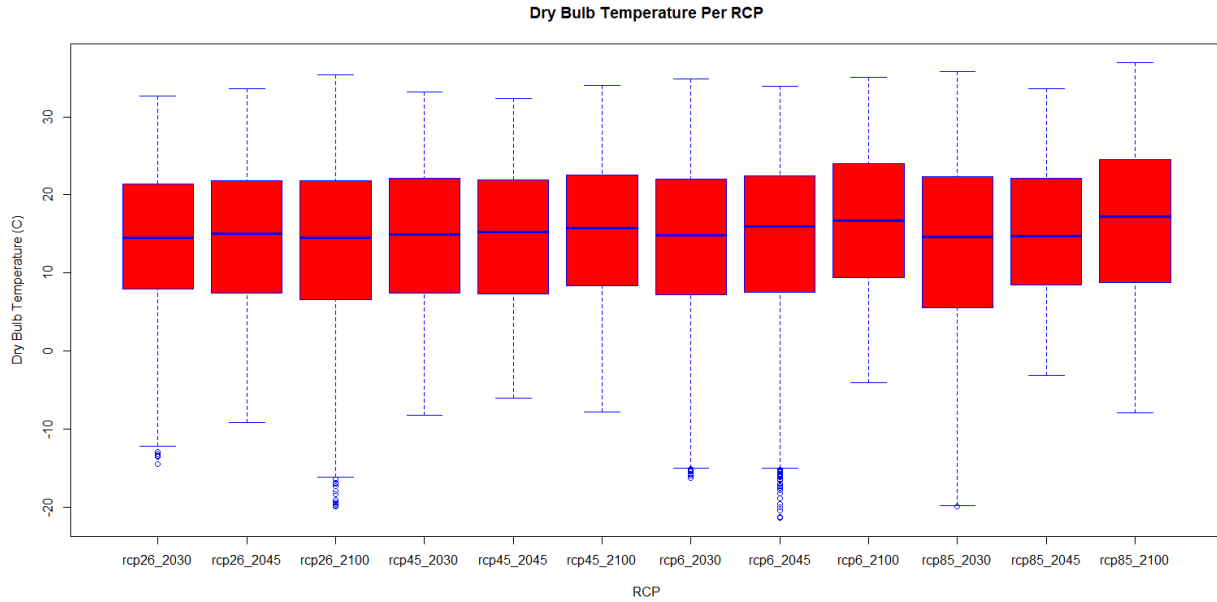


Figure 3.14: Dry Bulb Temperature for Each RCP and Year - Average dry bulb temperature increases for RCP 8.5 from 2030 to 2100 and decreases for RCP 2.6 over the same time period.

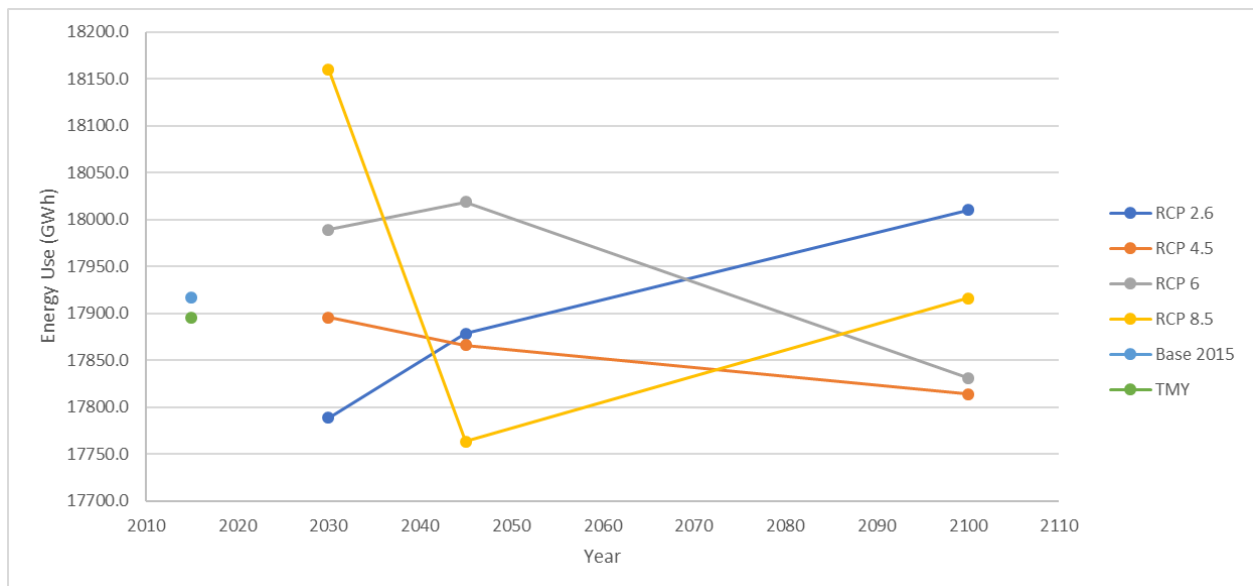


Figure 3.15: Total EPB Projected Energy Use - Simulated energy use (GWh) decreases for the EPB service area for the three highest emission RCP scenarios (4.6, 6, 8.5) while it decreases for the low emission scenario (RCP 2.6).

that only static multipliers are used on existing building simulations with different weather files; the current study does not consider sprawl of the built environment or land use changes during those time periods.

The trend seems somewhat clear other than RCP 8.5, 2045. The highest mitigation scenario (RCP 2.5) increase in energy from 2030 to 2100 while the other three scenarios decrease in total energy with a change proportional to their scale of emission escalation.

Taking a closer look at cooling and heating energy for the building type that dominates the EPB service area (IECC - Residential) makes the trend more clear. The heating energy and cooling energy for a representative residential buildings are shown in (Figures 3.16, 3.17).

It is apparent that the heating energy makes a greater percent of total energy use than cooling energy for the residential buildings which explains the increase in energy use for RCP 2.6 with the dry bulb temperature decreasing. This also explains the decrease in total EPB service area energy for the other three scenarios without nearly as much emission meditation. In addition to the scale of heating to cooling energy; the change in heating energy is greater than the change in cooling energy across the scenarios. For RCP 8.5, from 2030 to 2100, the heating energy decreases by about 1,800 kWh while the cooling energy only increases by about 550 kWh.

These total energy use results are significant. This states that with increasing temperatures for RCP scenarios, total energy use will actually decrease in Chattanooga. It should be noted that this could be impacted by the assumption that all buildings in the EPB area use electricity for heating while in reality, this would not be exactly the case.

Demand

Utilities are very sensitive to pricing and peak generation hours for each calendar month. This can often constitute 25% of a non-residential energy bill and is the worst-case scenario that utilities have to build or purchase power for to supply without blackouts or brownouts. As such, many utilities and organizations are interested in how to best adapt their infrastructure to be resilient against challenges from climate change. The high and low mitigation scenarios (2.6, 8.5) for years 2030 and 2100 are shown in (Figure 3.18).

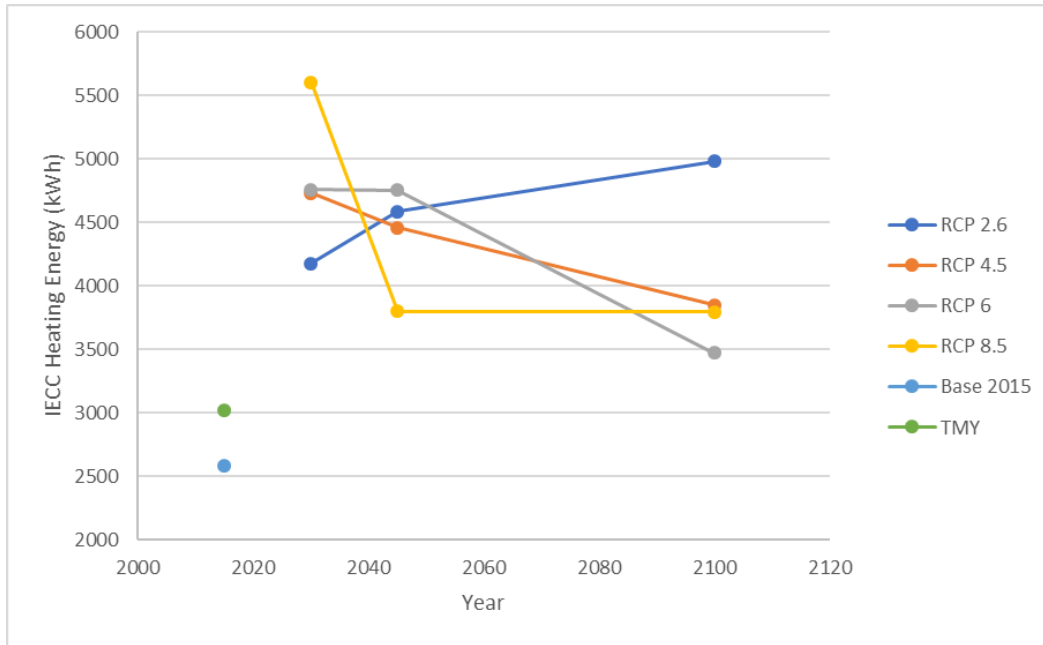


Figure 3.16: Single Building Residential Heating Energy Use - Heating Energy use (kWh) for a representative EPB IECC building decreases significantly for high emission scenarios and increases for low emission scenarios.

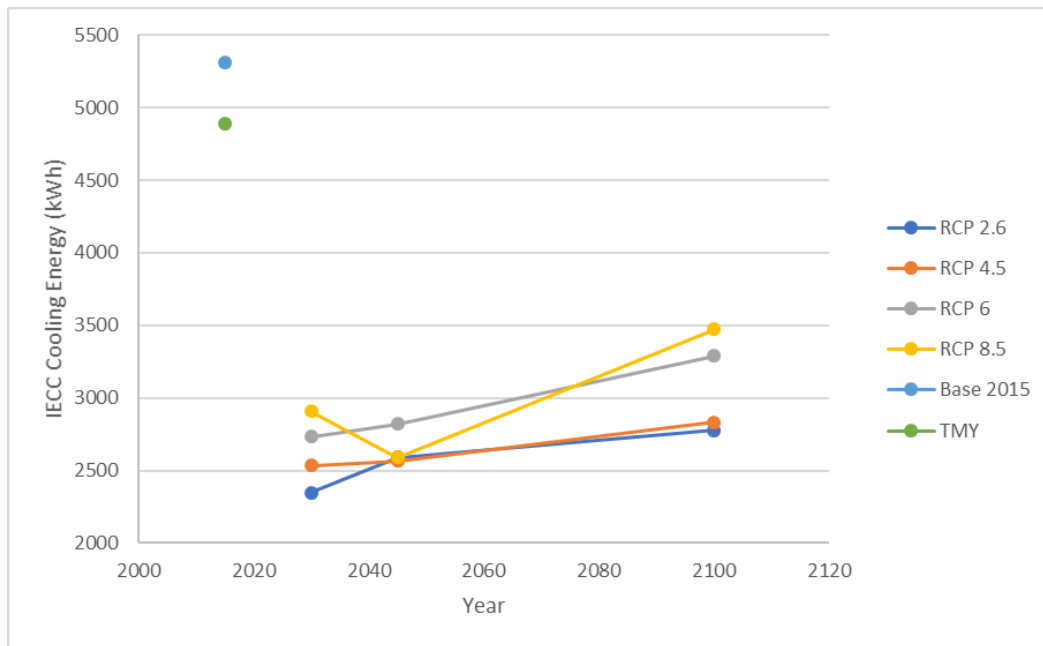


Figure 3.17: Single Building Residential Cooling Energy Use - Cooling Energy use (kWh) for a representative EPB IECC building increase for all RCPs.

The demand results show some interesting trends. For example, for RCP 2.6, the demand results for the winter are mixed over time with demand decreasing significantly during January from 2030 to 2100 but increasing in February and December. One would expect the demand to increase from 2030 to 2100 in the winter in the high mitigation scenario as temperatures decreased. It could be that the peak demand hour could have changed as climate model forecasts move farther into the future. Also for RCP 2.6, summer cooling demand is lower for the peak hours of June and July. For the high emission scenario, heating demand decreases significantly in January and February (as expected), but slightly increases during December which could be a similar situation as the low mitigation scenario for the winter months in which the peak hour may adjust based on the different climate scenario.

It is interesting to note that the transitional months of the spring and fall remain relatively constant across the extreme years and RCPs with more minor changes in the summer months than the winter months as well. This observation makes sense with the total energy use as the change in cooling energy over different RCPs and years was smaller than the change in heating over the same scenarios.

3.2 US Commercial Buildings Climate Impacts

The impact climate has on US commercial building energy use was analyzed in a similar fashion to the EPB analysis but on a larger scale including many different climate zones. In this scenario, real individual building data for every building in the scope of the study was unavailable. For this reason, and to ensure the models and simulations were of sufficient quality, they are simulated with TMY files and compared to CBECS building energy data for 2012. As electricity and natural gas account for 93% of total commercial building energy use, only these values as well as their sum were compared [61]. A comparison of simulated results compared to the CBECS values is shown in Figure 3.6. With all simulated values having less than 4% error, the models, simulations, and scaling values were deemed to be of sufficient quality prior to extrapolation.

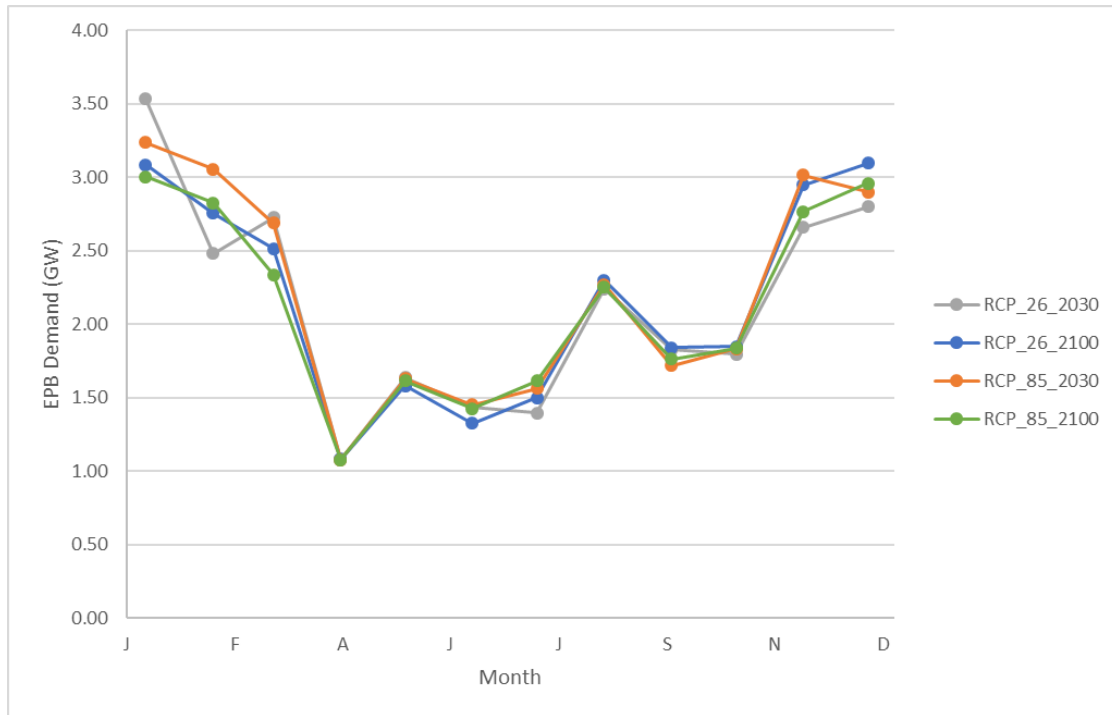


Figure 3.18: Single Building Residential Monthly Demand - High mitigation scenarios result a greater winter demand as average temperatures stay flat or decrease. High emission scenarios result in a greater summer demand as average temperatures increase.

3.2.1 Commercial Floorspace Extrapolation

Future weather results are compared in two ways. First, total commercial building square footage remained constant throughout the US from 2012 to 2030, 2045, and 2100 so that the direct impact of weather may be better observed. The next was to increase this area as years progress based on past trends. While this method introduces a source of error with any fit, it may provide a better depiction of what energy use will look like in the future as more commercial buildings are constructed. The linear fit (Figure 3.19) is a decent approximation of the data with an R^2 value of 0.89, and could be used to evaluate future weather energy impacts to commercial buildings with increasing floorspace.

The extrapolation methods described below are intended to be informative and useful for future studies, but include many of the typical limitations of such studies. Current building energy use and performance projection studies for efficient building technologies are largely based on prototype buildings. Despite buildings becoming larger in the U.S., the footprint and conditioned area of the prototype buildings have not changed over time. Equipment placed in these buildings is becoming more efficient over time, but this is not reflected unless the technology is a building code requirement. Many urban buildings are evolving toward mixed-use patterns that are not considered in canonical building equipment schedules. Systemic modifications to occupancy and building use due to pandemics is not reflected. Specifically, this study projects building growth but without accounting for urban sprawl or concomitant microclimate variation. Building use and mixed-use building operation is not considered. While the study includes projections out to 2100, improved efficiency and changes in occupancy over such time spans are not considered. In summary, this work does not speculate or project the building size, shape, occupancy, use, or equipment efficiency.

3.2.2 Weather files

Investigation into the weather files allows a more complete understanding regarding the underlying causes of change in simulated energy use. The trend of important meteorological variables over time for each RCP is shown in Figure 3.20.

Table 3.6: Simulated vs Actual US Commercial Building Energy Use - Error values for comparison of simulated commercial building energy use in the US with actual commercial building energy use reported from CBECS 2012 [61].

	Energy (GJ)	Electricity (GJ)	Gas (GJ)
Actual (CBECS)	6,825,633,248	4,474,800,000	2,350,833,248
Simulated	7,084,270,519	4,654,391,371	2,429,879,148
% Error	3.8%	4.0%	3.4%

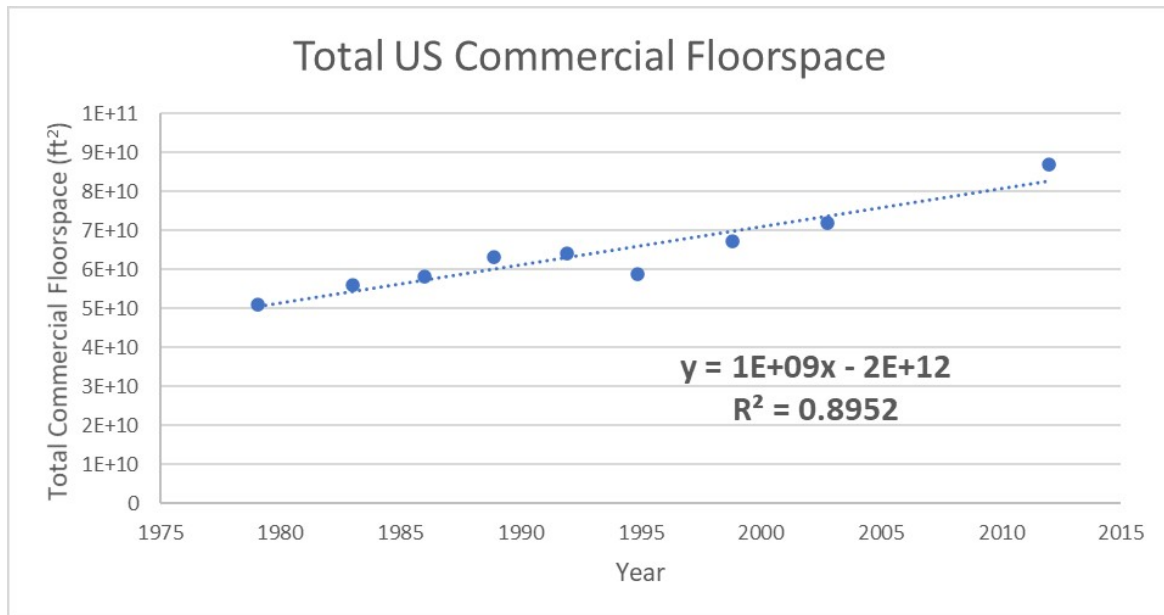


Figure 3.19: US Commercial Floorspace Trend - Linear fit to commercial floorspace historical data. A linear fit is a realistic approximation of commercial floorspace in the US.

Since dry bulb is the most impactful meteorological variable, its trend from 2030 to 2100 tends to dominate energy use fluctuations. The dry bulb temperature is directly related to the radiative forcing of each scenario with the average temperature for RCP 2.6 staying relatively constant while RCP 8.5 increasing the most significantly by 3.5°C-4.5°C. The dew point trend is similar to dry bulb temperature with the exception that RCP 2.6 increases in 2045 before decreasing in 2100. The pressure in all scenarios decreases except for RCP 2.6, with the largest decreases in RCP 4.5 and RCP 6. Pressure does not follow the radiative forcing trend exhibited by temperature variables.

While the path from 2030 to 2100 may be different under each scenario, the difference from 2030 to 2100 is the least for RCP 2.6 (decreasing slightly), with larger changes for scenarios with stronger radiative forcing as RCP 8.5. The national direct normal radiation is inversely proportional to the amount of radiative forcing, with direct normal radiation increasing significantly for RCP 2.6 and slightly for RCP 4.5 while RCP 6 decreases slightly and RCP 8.5 decreases significantly.

3.2.3 Constant Floorspace

Total Energy

The commercial square footage for each climate zone in 2012 is included in Table 1 of the Appendix. This shows which climate zones most impact national energy use with climate zone commercial percentages summarized in Table 2. Moist climate zones (A) tend to have the most commercial floorspace, accounting for about 76.35% of total US commercial floorspace whereas dry zones (B) make up 18.49% and marine zones (C) constitute 4.59%.

For distributions of temperature, warmer climate zones (1-3) make up 48.15% of total commercial floorspace in the US while colder climate zones (4-8) make up 51.85%. These summaries help intuit the following electricity and gas results where increasing average temperatures decrease gas usage (generally heating) and increase electricity usage (generally cooling). ASHRAE typically uses base temperatures (i.e. 22°C minus internal heat gains) of 18.3°C for heating and 10°C for cooling. As the average temperature approaches a building's base temperature, the total energy use will decrease.

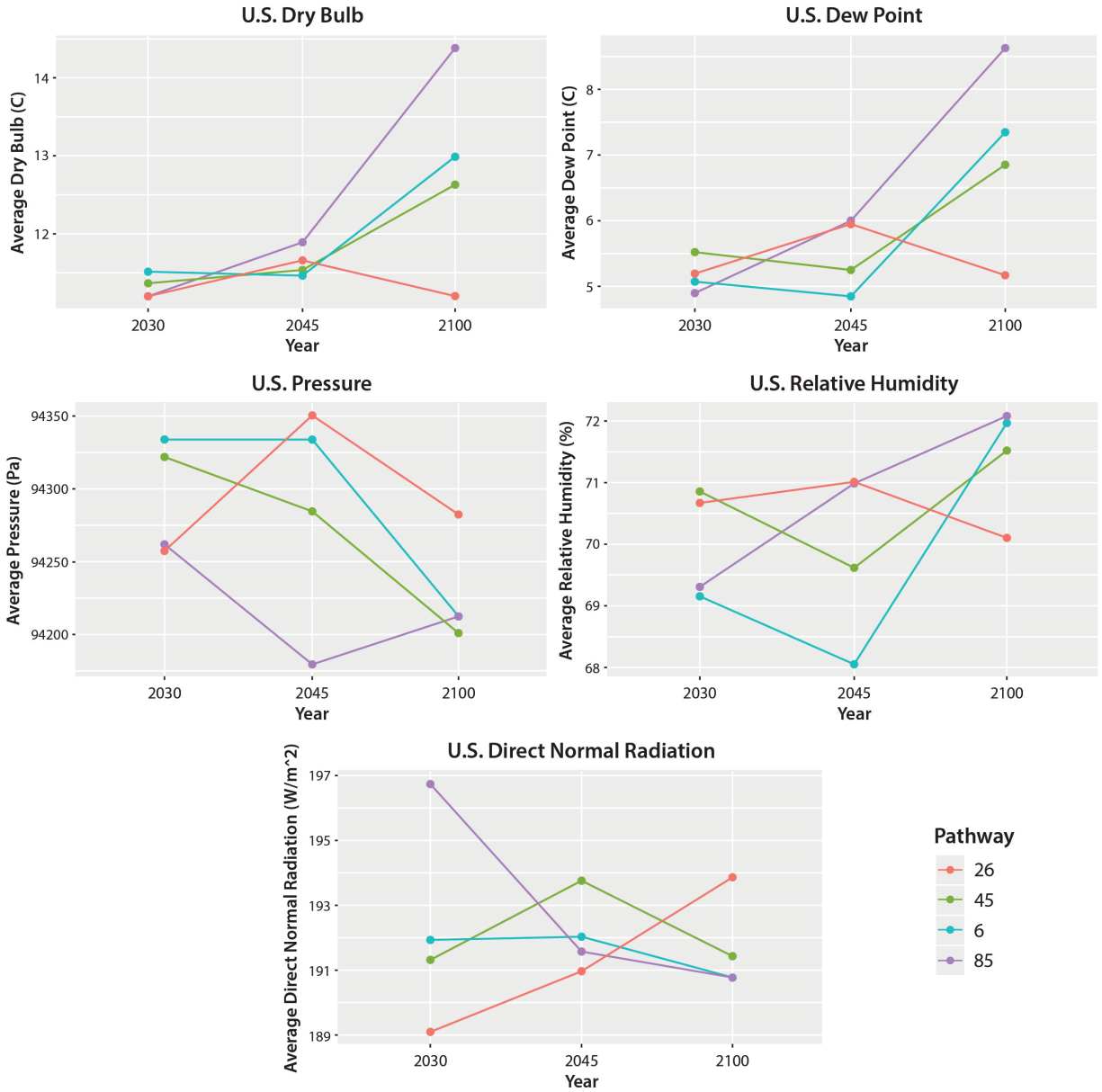


Figure 3.20: Average US Weather Variables for Various RCPs - These weather variables shown here are averages of all of the climate zones, unweighted by the number of buildings in each zone. This provides a different angle than the building floorspace weighted energy simulation results.

Total commercial energy use across the US, including Alaska and Hawaii, is represented in Figure 3.21 where total floorspace is assumed to remain constant from 2012 to 2100 to highlight energy use changes due solely to climate change. The RCP energy trends show RCP 2.6 increasing by 2.3% while RCP 8.5 decreases by 7.5%. RCP 4.5 and 6 both decrease from 2030 to 2100 by 4.1% and 3.7%, respectively.

Differences in 2030 are worth noting since many variables often, but not always, change more dramatically by 2100. Due to the combination of factors that affect building energy use, while the ending point for some of the higher radiative forcing scenarios is similar, there is a larger decrease in total energy for RCP 8.5. This result is intuitive since the United States, as a nation, resides in heating-dominated (i.e. cold) climate zones with annual heating outpacing cooling by a factor of approximately five. National average of 54% of residential energy use for space conditioning and 32% of commercial energy use for space conditioning involves 9% of that energy going to space cooling and 45% of that energy going to space heating. While climate change may ultimately require adaptations of the built environment to withstand the effects (e.g. more significant extremes) of climate change, from a building energy use perspective alone, climate change is a net energy saver for the United States [25].

We include climate zone specific trends in Figure 1 of the Appendix. These total energy trends quantify energy increases for warm climate zones as more extreme radiative forcing scenarios require more cooling whereas cold climate zones require less heating energy use. The opposite is true for lower radiative forcing scenarios across climate zones.

Electricity

Analysis of simulated electricity indicates how cooling and other electrical loads will be affected by future weather, and better informs trends seen in total energy usage. Increasing average temperatures generally result in increased electricity use, as quantified by simulation results shown in Figure 3.22. The scale by which the electricity use changes is proportional to the change in radiative forcing. RCP 2.6 and 6 increase in total electrical use (1.3% and 2.8% increases from 2030 to 2100), while RCP 4.5 and 8.5 decreases in 2045 prior to larger increases in 2100 resulting in total increases of 1.0% and 3.5% respectively. Due to a lack of cooling needs, cold weather climate zones such as 6, 7, and 8 is much less and even decreases

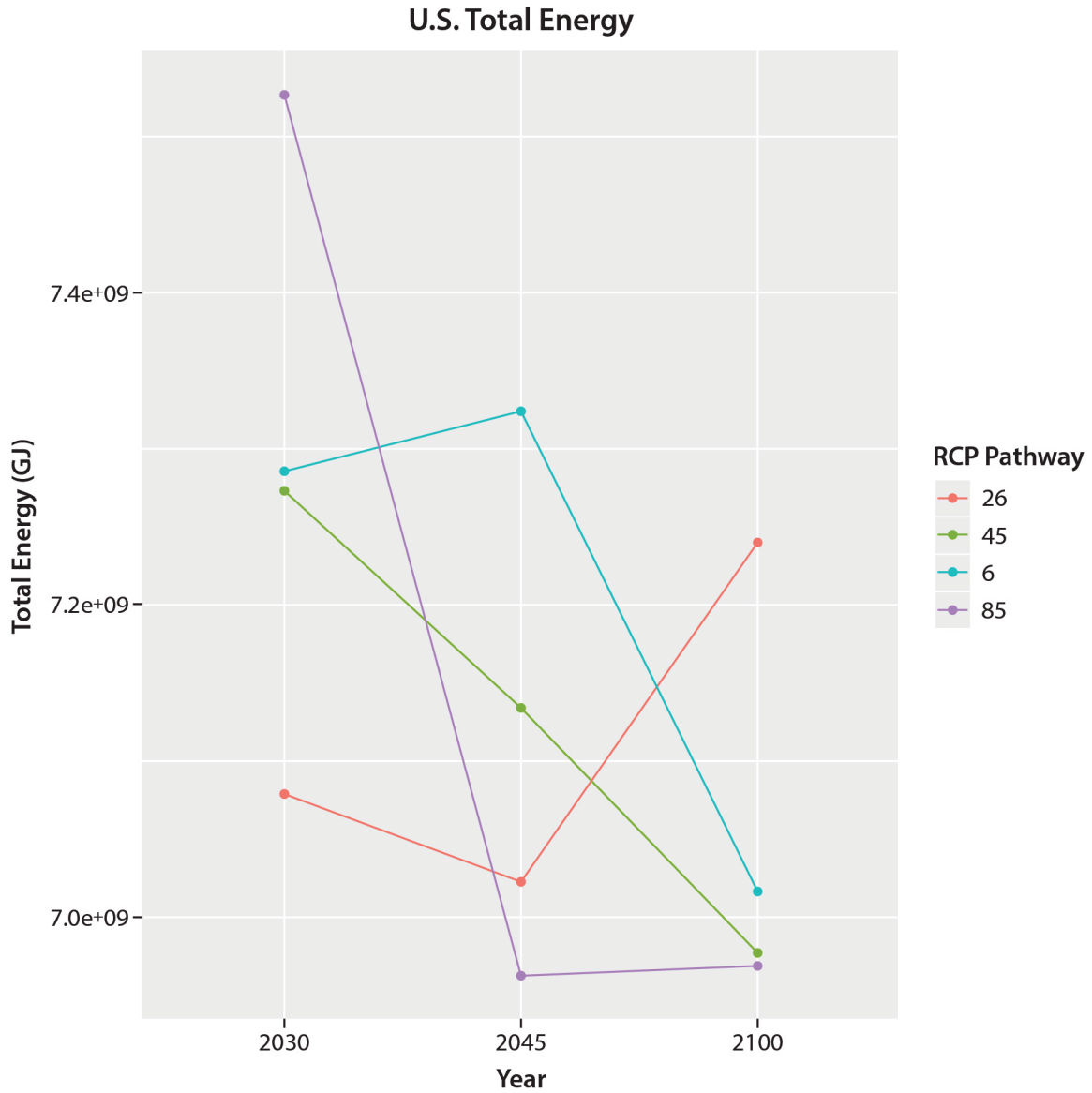


Figure 3.21: US Total Commercial Building Energy Use for RCPs - The total energy in the US decreases significantly for the highest emission scenario and increases for the lowest emission scenario. The two middle emission scenarios also decrease significantly.

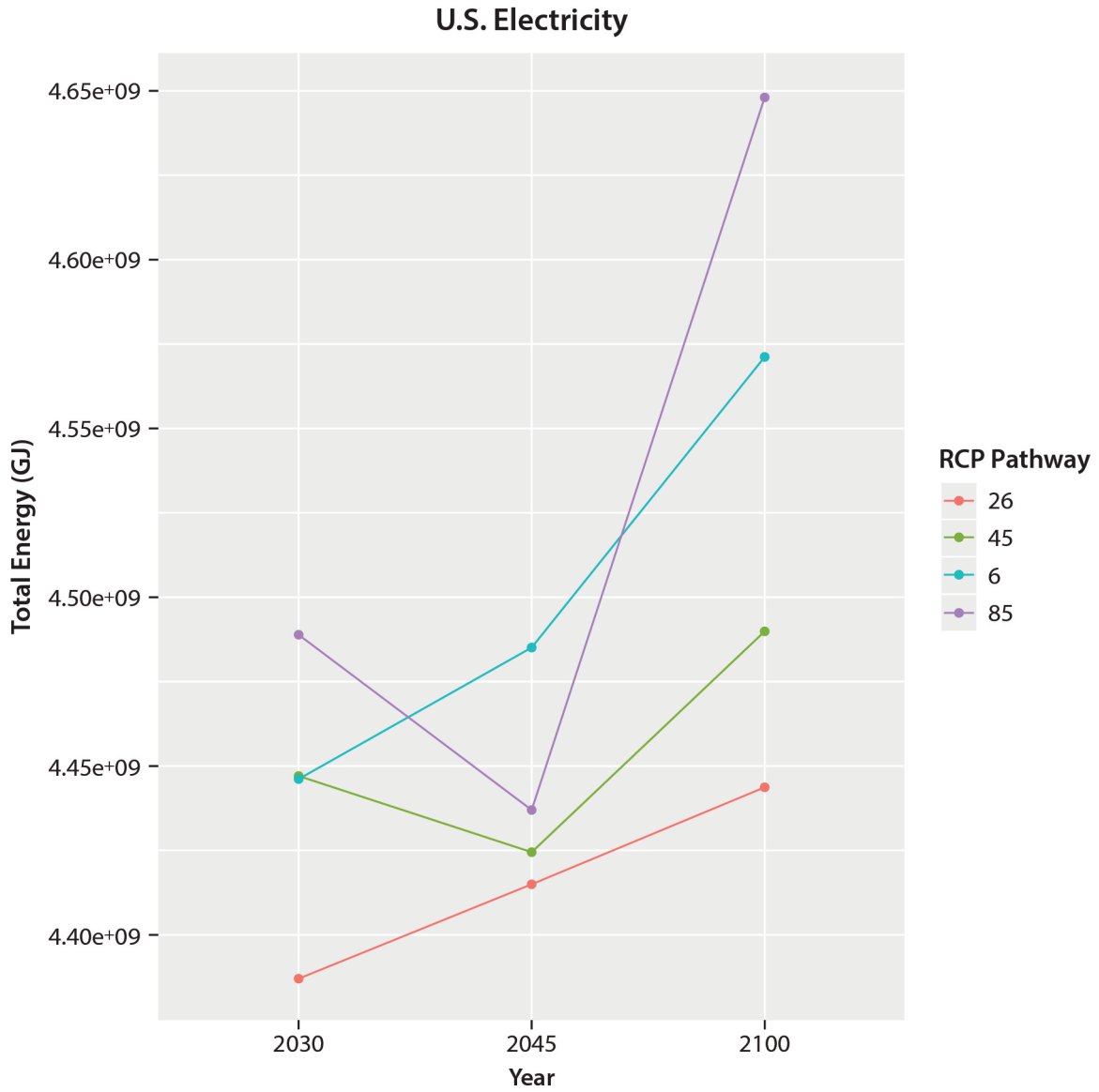


Figure 3.22: US Total Commercial Building Electricity Use for RCPs - Electricity use in the US (mostly used for cooling) with constant building area increases in all scenarios by 2100.

for some RCP pathways. Changes in electrical use are quantified by climate zone in Figure 2 of the Appendix.

Natural Gas

The use of heating fuel sources, primarily natural gas, would be expected to have opposite trends than electricity as quantified in Figure 3.23. For the most extreme radiative forcing and greatest temperature increase pathway (8.5), the natural gas use decreases. Conversely, natural gas use increases under RCP 2.6. Electricity use and natural gas use are both anticipated to decrease from 2030 to 2045 for RCP 8.5 before cancelling out to a flat transition from 2045 to 2100. Changes for RCP 2.6, 4.5 and 6 are relatively small from 2030 to 2045 compared to the larger changes occurring between 2045 and 2100. Percent changes for the four RCPs range from an increase in gas usage of 3.9% from 2030 to 2100 for RCP 2.6 to decreases in gas usage of 12.0%, 13.9%, and 23.6% over the same period for RCPs 4.5, 6, and 8.5, respectively.

Since electricity and natural gas are inversely correlated, it is important to contrast the amount the total natural gas use decreases compared to the amount the total electricity increases for the different RCP pathways to determine the net energy gains or losses. While the total electricity used is much more than gas due to electricity use for base loads, the change in gas from 2030 to 2100 is much greater as reflected in the total energy use of Figure 3.21. The simulated gas usage is broken down by climate zone as shown in Figure 3 in the Appendix.

3.2.4 Traditional Urban Growth

While the previous results focus on relative change to electricity, natural gas, and total energy under climate change alone (constant commercial floorspace from 2030 to 2100), this section uses CBECS total commercial floorspace historical data from the last 40 years to predict more realistic (absolute) energy use in the future.

Total simulated energy used by commercial buildings throughout the country is shown in Figure 3.24. In calendar year 2019, approximately 125 million buildings, residential

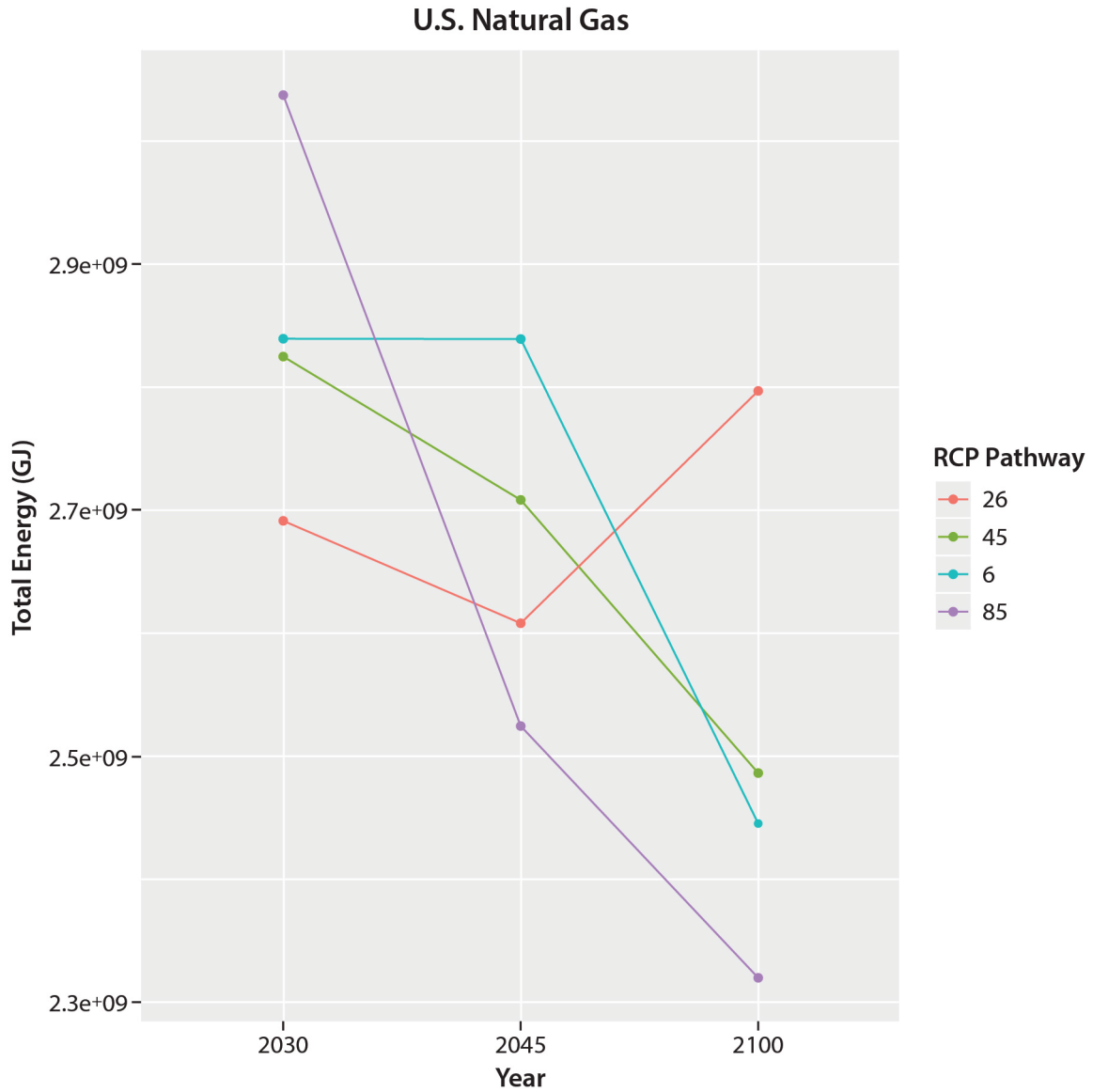


Figure 3.23: US Total Commercial Building Natural Gas Use for RCPs - Natural Gas use in the US (mostly used for heating) with constant building area decreases in the three highest emission scenarios but increases for the mitigation scenario.

and commercial combined, consumed approximately \$412 billion USD in energy. As the number of buildings increases, the total energy use will likely increase by approximately 65% if not sufficiently eroded by building energy efficiency efforts. The trends reported here, while dominated by building growth, is similar to previous results. Energy use increase is the smallest for the highest radiative forcing scenario (55.3% increase) and the largest for the lowest radiative forcing scenario (71.6% increase); mirroring the results from the constant floorspace analysis. National electricity use is similar to the total energy usage except the rate of change is quicker (Appendix Figure 4) and some natural gas increase (Appendix Figure 5) but not as significant as with total energy.

3.3 Model America

The Model America concept of generating and simulation a building energy model of more than 125 million US buildings was a significant challenge in terms of building-specific data retrieval, algorithms, compute for generation and simulation, and big data management. A total of 122,714,640 building energy models were created. This resulted in 3.2TB of OpenStudio models and 2.0TB of EnergyPlus models separated by state and county. Summaries of the number of buildings per state and county are shown in Figure 3.25.

While the primary result of this analysis is the freely available building energy models, which can be used in many different use cases (described in Section 2.7), there are some interesting statistics regarding the data movement, building generation/simulation, and data sorting/post-processing. A summary of the run parameters and their statistical range is shown in Table 3.7. The wall time in Table 3.7 represents the total time to accomplish this workflow. The majority of the runs were completed using 80% of THETA for 2 hours, and averaged over 1.1 million buildings. These values were optimized over time as different run parameters were evaluated.

Out of the 125,714,640 building sensed in the US, 122,930,327 were aggregated and made available (97.8%). The entire modeling process was accomplished through 141 successful generation/simulation runs with an average building retrieval rate (successful generation/simulation/transfer) from the supercomputer of 77% across all runs, though this

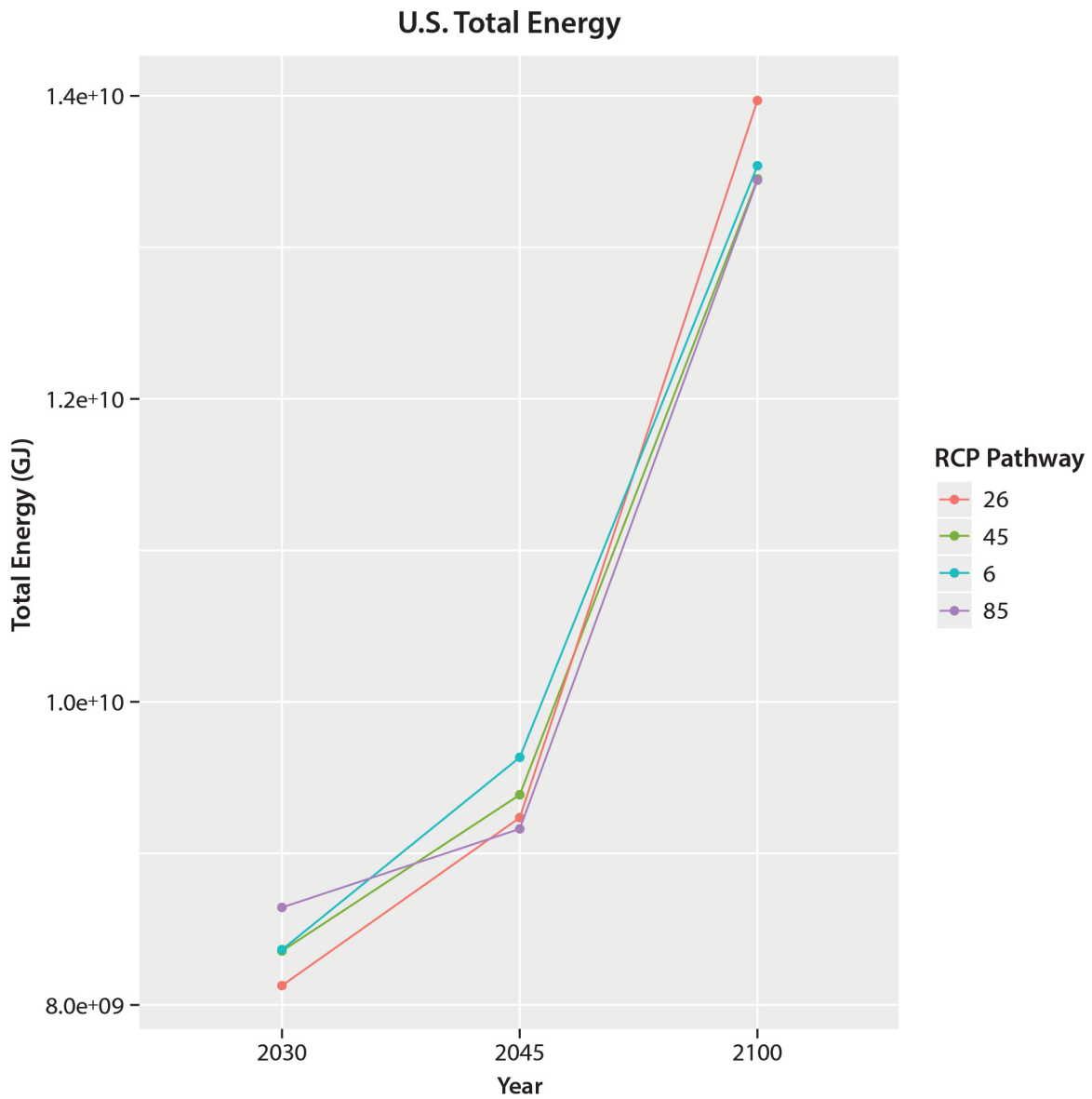


Figure 3.24: US Total Commercial Building Energy Use for RCPs and Increasing Commercial Floorspace - If commercial building area continues to increase at the same rate (assuming no changes in building technology), US commercial building energy will increase for all scenarios with the relative increase proportional to the constant floorspace results.

rate increased to 94% for the final 30% of runs as HPC parameters were optimized. One of the main causes of this issue is the variance in generation and simulation times across different building geometries and building types. These long generation and simulation times for certain buildings are problematic as they get cut off as the data is churned through during a runs. This specifically occurs as wall time approaches time requested if the data has not been churned through completely.

Post-processing of the data to make it readily available was also a significant undertaking. The data was untarred from each run by state and zipped by state into county sized zip files. These processes were run in parallel, drastically reducing the time necessary for the processing. The states with the largest number of buildings were split for increased speed and to prevent the THETA's lustre file system from being overwhelmed with files. A summary of the times for these processes is shown in Table [3.8](#).

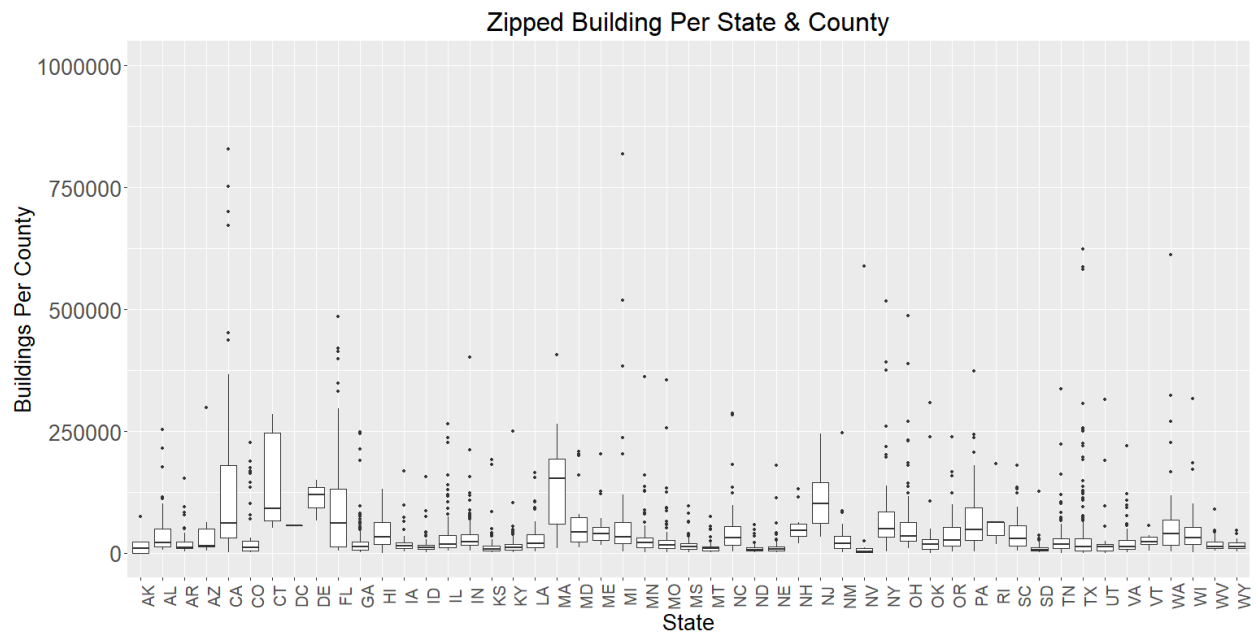
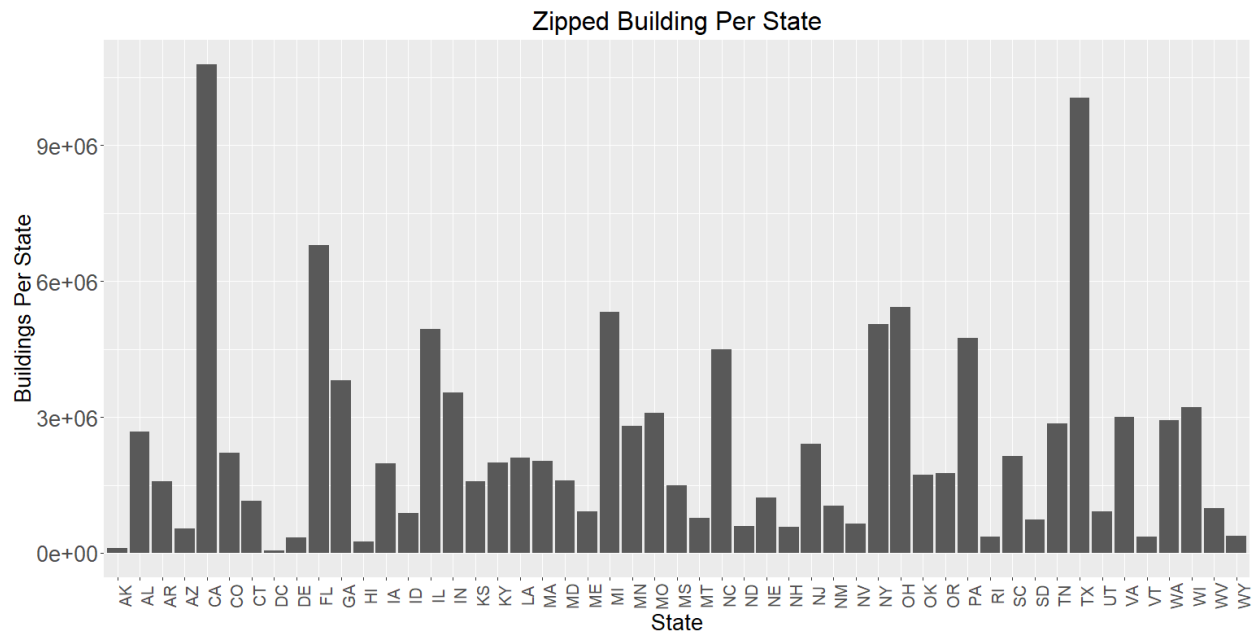


Figure 3.25: Buildings per state and county - The three largest counties from California, Texas, and Illinois were removed for better visualization of the bottom figure. The largest number of buildings are in states such as California, Texas, and Florida.

Table 3.7: Model America HPC Parameters - 80% of THETA was used for the majority of the runs with just over 1 million buildings per run on average.

	Requested time (min)	Nodes	Churns	Cores	% of Theta	Node Hours
Min	30	128	1	8,192	3%	58
Mean	119	2,816	6	180,249	64%	5053
Median	120	3,514	8	224,896	80%	5220
Max	180	3,514	8	224,896	80%	8397

	Total Buildings	Tarred-up buildings (%)	Bldgs/sec	core minutes/Bldg	Pre-processing Time (min)	Wall Time (min)
Min	16,384	1%	1	4	0.07	4
Mean	1,131,912	77%	153	51	23	106
Median	1,116,160	92%	98	22	20	122
Max	1,799,168	100%	491	1496	123	143

Table 3.8: Model America HPC Post-Processing Statistics - The runs were untarred and zipped in parallel with the largest states being split up so as to reduce the maximum time shown below.

	Time to Untar Buildings by State (Days)	Time to Zip each State (Hours)
Min	1.23	0.4
Mean	3.74	26.0
Median	2.60	21.1
Max	12.08	78.9

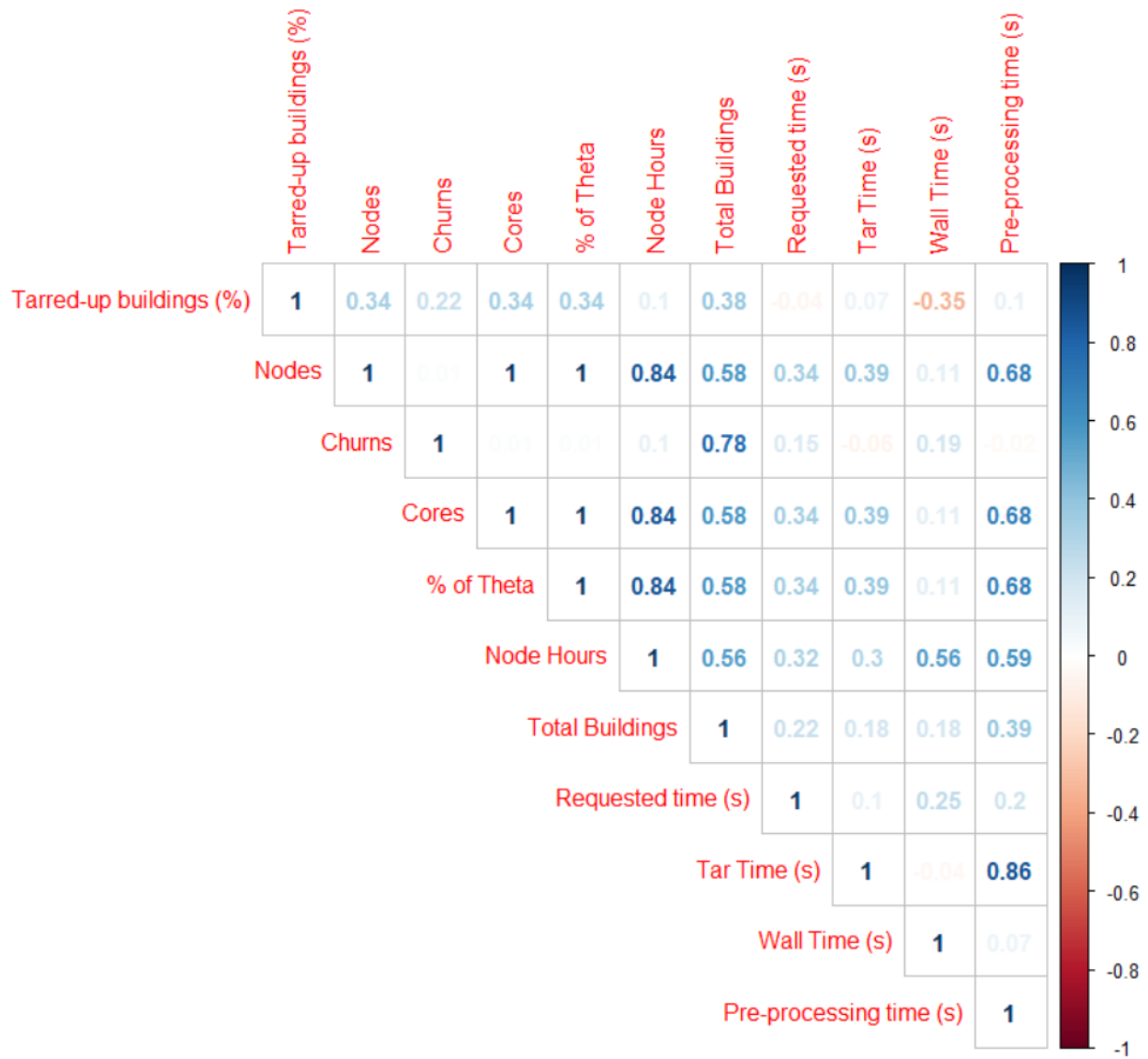


Figure 3.26: Model America HPC Statistics Correlations - The percent of successful runs (Tarred-up Buildings (%)) has a negative correlation with wall time.

Chapter 4

Conclusion

This research aimed to identify and optimize methods by which to generate, simulate, and analyze building energy from a utility-scale to the entire US. It focused primarily on ideal methodologies for this scale of BEM, analysis of energy and demand savings opportunities for a utility through various building technologies, and the impact of climate change on a utility and buildings across the US. Large-scale BEM efforts such as these maximize the impact of the results, serving to motivate decision makers to enact policies or implement technology that results in real and efficient change.

Assignment of building type and vintage is currently an outstanding challenge in the emerging area of urban-scale energy modeling. This study leveraged 15-minute, whole-building electricity use and building energy models for 100 buildings to assess data preparation as well as building type and vintage assignment methods. Omission of large gaps ($>75\%$ of data), Auto Regressive Integrated Moving Average (ARIMA) for filling small gaps (<1 Week), and Univariate Dynamic Time Warping (DTW) for filling large gaps (>1 Week) was found to be effective in this study. Three building type classification methods were compared involving Euclidean Distance, DTW, and machine learning with random forest and time-based statistics. Euclidean Distance was the fastest and had the best overall classification accuracy, whereas the random forest performed better for commercial buildings. The run-time of DTW would be a significant hindrance as the number of building increased. For each method, a pseudo-confidence was obtained via the similarity distance (Euclidean,

DTW) or the class probability (RF). The distance metrics of were significantly more useful for confidence as filtering by high confidence predictions resulted in better quantitative metrics.

For the Virtual EPB analysis, over 178,000 OpenStudio and EnergyPlus models were generated, over 2 million annual simulations were performed on high performance computing results, and baseline models were empirically validated against 15-minute electrical consumption of each building. This study elucidates methods and showcases results for statistical distributions of potential energy and demand savings of 8 building technologies under a maximum technical adoption scenario.

Energy efficient building measures were implemented in each building include energy-efficient lighting, space-sealing, roof/attic insulation, improved HVAC efficiency, and rooftop photovoltaics (PV). Lighting and infiltration savings were notably impactful in older buildings while maximum rooftop PV potential for the utility's service area was estimated at up to 750 GWh in a single month.

Annual demand savings, defined as the sum of energy use during the peak hour of each calendar month, were shown for four measures including two scenarios for smart thermostat building space pre-conditioning and smart water heaters. The smart thermostat with an 4.4°C offset resulted in the greatest demand reduction. The 2.2°C offset simulated an average of 7% lower annual demand offset but would likely result in fewer customer overrides than the liberal 4.4°C approach.

Cost and emissions savings were evaluated for each of the building technologies with emissions being directly proportional to annual energy savings. The 2015 US national average of USD 0.1041/kWh for electricity savings and USD 10.5/kW were used to measure cost savings. The smart thermostat cost savings (dominated by demand savings) were second only to switching the HVAC from electric to natural gas, which in reality would have some of the cost savings offset by the fuel use.

The EPB climate analysis described methods for translating the Intergovernmental Panel on Climate Change (IPCC) Representative Concentration Pathways (RCP) scenario data for years 2030, 2045, and 2100 into hourly Future Meteorological Year (FMY) weather files for Chatanooga, TN. The results showed that dry-bulb temperature is the most influential variable in affecting simulated building results. It was shown to decrease over time for RCP

2.6 while increasing over time for RCPs 4.5, 6, and 8.5. This led to an increase in total energy from 2030 to 2100 for RCP 2.6 and a decrease in total energy from 2030 to 2100 for RCPs 4.5, 6, and 8.5. This was shown to be caused by the larger proportion of total energy used by heating electricity vs cooling electricity as well as the larger decrease in decrease in heating energy vs decrease in cooling energy for the EPB area.

The monthly demand profile mostly followed these same patterns as the total energy use. Demand decreases for low mitigation scenarios and increases for high mitigation scenarios in the winter. There are some exceptions likely due to and adjusted demand peak hour as forecasts go farther into the future. spring, summer, and fall months were mostly unchanged comparatively across RCPs and into the future as cooling energy was impacted less by the climate scenarios.

A similar methodology was applied to all commercial buildings in the US. Again, IPCC RCP scenario data for the years 2030, 2045, and 2100 was translated into hourly FMY weather files, but all US climate zones were considered for this analysis. DOE prototype building energy models were simulated with the EnergyPlus simulation engine to assess climate change impacts to commercial building energy use for the United States. Simulation results were scaled up using building construction multipliers to the total commercial floorspace in the US in 2012. TMY-simulated values were compared to actual simulation values from CBECS 2012 with total energy, electricity, and gas scaled simulation results being within 4% of actual values.

Relative to a 2012 building stock baseline, RCP 2.6 results in a 2.3% increase in total U.S. commercial energy use whereas RCP 8.5 results in a 7.5% decrease in energy use. This is due in large part to the nation's five-fold difference between energy use for space heating over space cooling; RCP 2.6 is likely to result in a 3.9% increase in gas (heating) whereas RCP 8.5 would result in a 23.6% decrease in gas use from 2030 to 2100.

Trends from the last 40 years were used to extrapolate future commercial construction for a more realistic depiction of anticipated commercial building energy use in the future. With increasing commercial floorspace, total building energy, electricity, and gas all increased with the amount of electricity used increasing much more than the amount of gas (reflecting the results seen when floorspace remained constant). While climate change may ultimately

require adaptations of the built environment to withstand the effects of climate change, from a building energy use perspective alone, climate change is a net energy saver for the United States.

The Model America effort was conducted on a scale different than any UBEM analysis to date. The methods and parameters used may lay the groundwork for future analyses of this scale. In addition, the models can be used for simulation-informed analyses across the country including rural areas where this data may have not been available. Demand management, energy efficiency, grid resilience, and climate impacts may all be evaluated for buildings across the country. The models can also be used for automated personal financing of building technology investment, increasing the adoption of these technologies and building energy efficiency.

Bibliography

- [1] Adams, M. B. (2018). Autobemgen, version 00, u.s. copyright txu 2-141-227. [4](#), [25](#), [40](#)
- [2] Allen, M. R., Fernandez, S. J., Fu, J. S., and Olama, M. M. (2016). Impacts of climate change on sub-regional electricity demand and distribution in the southern united states. *Nature Energy*, 1(8):1–9. [31](#)
- [3] Ang, Y. Q., Berzolla, Z. M., and Reinhart, C. F. (2020). From concept to application: A review of use cases in urban building energy modeling. *Applied Energy*, 279:115738. [2](#)
- [4] Ashina, S. and Nakata, T. (2008). Energy-efficiency strategy for co2 emissions in a residential sector in japan. *Applied Energy*, 85(2):101 – 114. [10](#)
- [5] ASHRAE (2014). Guideline 14-2014: Measurement of energy, demand, and water savings. [ix](#), [44](#)
- [6] Bass, B. and New, J. (2019). Potential demand reduction from buildings in a simulated utility. [29](#)
- [7] Bass, B. and New, J. (2020). Future meteorological year weather data from ipcc scenarios. [35](#)
- [8] Bass, B. and New, J. (2021). Future meteorological year weather files and commercial building energy impacts from ipcc scenarios. [38](#)
- [9] Bass, B., New, J., and Copeland, W. (2020). Potential energy, demand, emissions, and cost savings distributions for buildings in a utility’s service area. [29](#)
- [10] Bass, B., New, J., Ezell, E., Garrison, E., Copeland, W., and Im, P. (2021). Utility-scale building type assignment using smart meter data. [22](#)
- [11] Bhandari, M. S., Shrestha, S. S., and New, J. R. (2012). Evaluation of weather data for building energy simulations. 49:109–118. [31](#)
- [12] Booth, A., Choudhary, R., and Spiegelhalter, D. (2012). Handling uncertainty in housing stock models. *Building and Environment*, 48:35 – 47. [12](#)
- [13] Breiman, L. (2001). Random forests. *Mach. Learn.*, 45(1):5–32. [vii](#), [25](#), [26](#)

- [14] Cerezo, C., Sokol, J., AlKhaled, S., Reinhart, C., Al-Mumin, A., and Hajiah, A. (2017). Comparison of four building archetype characterization methods in urban building energy modeling (ubem): A residential case study in kuwait city. *Energy and Buildings*, 154:321 – 334. [7](#)
- [15] Cerezo Davila, C., Reinhart, C. F., and Bemis, J. L. (2016). Modeling boston: A workflow for the efficient generation and maintenance of urban building energy models from existing geospatial datasets. *Energy*, 117:237 – 250. [8](#)
- [16] Cesium (2020). Cesium. [xiv](#), [28](#), [62](#)
- [17] Chen, T. and Guestrin, C. (2016). Xgboost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. [vii](#), [25](#), [26](#)
- [18] Deru, M., Field, K., Studer, D., Benne, K., Griffith, B., Torcellini, P., Liu, B., Halverson, M., Winiarski, D., and Rosenberg, M. (2011). *US Department of Energy commercial reference building models of the national building stock*. [31](#), [40](#)
- [19] Energy Information Administration (2018). How much energy is consumed in u.s. residential and commercial buildings? [1](#)
- [20] Flato, G., J., Marotzke, B., Abiodun, P., Braconnot, S., Chou, W., Collins, P., Cox, F., Driouech, S., Emori, V., Eyring, C. Forest, P., Gleckler, E., Guilyardi, C., Jakob, V., Kattsov, C., Reason, and Rummukainen, M. (2013). Climate change 2013: The physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change. [31](#)
- [21] Garrison, E., New, J., and Adams, M. (2019). Accuracy of a crude approach to urban multi-scale building energy models compared to 15-min electricity use. [28](#)
- [22] Geophysical Fluid Dynamics Laboratory (2020). Climate modeling. [14](#)
- [23] Gherzi, F. (2018). The imaclim-p model version 3.4. [10](#)
- [24] Goy, S., Tardioli, G., Monsalvete, P., and de Uribarri, A. (2017). *Building energy demand modeling: from individual buildings to urban scale*. [xi](#), [6](#)

- [25] Hadley, S. W., Erickson, D. J., Hernandez, J. L., Broniak, C. T., and Blasing, T. (2006). Responses of energy use to climate change: A climate modeling study. *Geophysical research letters*, 33(17). [73](#)
- [26] Hale, E., Horsey, H., Johnson, B., Muratori, M., and Wilson, Eric, e. a. (2018). The demand-side grid (dsgrid) model documentation. [9](#)
- [27] Hyunwoo, L. and Zhiqiang, Z. (2017). Review on stochastic modeling methods for building stock energy prediction. *Building Simulation*. [xi](#), [6](#)
- [28] Irwin, R., Chan, j., and Frisque, A. (2016). Energy performance for ashrae 90.1 baselines for a variety of canadian climates and building types. [xii](#), [19](#)
- [29] Jarnagin, R. and Bandyopadhyay, G. (2010). Weighting factors for the commercial building prototypes used in the development of ansi/ashrae/iensa standard 90.1-2010. [35](#), [40](#)
- [30] Kim, Y.-J., Ahn, K.-U., and Park, C.-S. (2014). Decision making of hvac system using bayesian markov chain monte carlo method. *Energy and Buildings*, 72:112 – 121. [12](#)
- [31] Kolter, Z. and Ferreira, J. (2011). A large-scale study on predicting and contextualizing building energy usage. [9](#)
- [32] Kuhn, M. (2019). Caret. [23](#)
- [33] Microsoft (2018). Microsoft building footprints. [17](#), [39](#)
- [34] Moens, D. and Vandepitte, D. (2005). A survey of non-probabilistic uncertainty treatment in finite element analysis. *Computer Methods in Applied Mechanics and Engineering*, 194(12):1527 – 1555. Special Issue on Computational Methods in Stochastic Mechanics and Reliability Analysis. [10](#)
- [35] New, J., Adams, M., Bass, B., Clinton, N., Berres, A., Garrison, E., Guo, T., and Dumas, M. (2021). Model america: a crude building energy model of u.s. buildings. [40](#)
- [36] New, J., Garrison, E., Adams, M., Bass, B., and Guo, T. (2020). Scaling beyond tax assessor data. [8](#), [18](#)

- [37] New, J. R., Adams, M., Im, P., Yang, H., Hambrick, J., Copeland, W., and Bruce, Lilianand Ingraham, J. A. (2018). Automatic building energy model creation (autobem) for urban-scale energy modeling and assessment of value propositions for electric utilities. [17](#), [18](#)
- [38] Nutkiewicz, A., Yang, Z., and Jain, R. K. (2018). Data-driven urban energy simulation (due-s): A framework for integrating engineering simulation and machine learning methods in a multi-scale urban energy modeling workflow. *Applied Energy*, 225:1176 – 1189. [7](#)
- [39] Office of Energy Efficiency and Renewable Energy (2012). About building energy modeling. [3](#)
- [40] Omlin, M. and Reichert, P. (1999). A comparison of techniques for the estimation of model prediction uncertainty. *Ecological Modelling*, 115(1):45 – 59. [12](#)
- [41] Present, E., Wilson, E., Parker, A., and Frick, N. (2019). End-use load profiles for the u.s. building stock. [9](#)
- [42] Quinlan, J. R. (1986). Induction of decision trees. *Mach. Learn.*, 1(1):81–106. [25](#)
- [43] Reinhart, C. F. and Cerezo Davila, C. (2016). Urban building energy modeling – a review of a nascent field. *Building and Environment*, 97:196 – 202. [1](#)
- [44] Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press. [vii](#), [25](#), [26](#)
- [45] Robinson, D., Campbell, N., Gaiser, W., Kabel, K., Le-Mouel, A., Morel, N., Page, J., Stankovic, S., and Stone, A. (2007). Suntool – a new modelling paradigm for simulating and optimising urban sustainability. *Solar Energy*, 81(9):1196 – 1211. CISBAT 2005. [8](#)
- [46] Robinson, D., Haldi, F., Kämpf, J., Leroux, P., Rasheed, A., and Wilke, U. (2009). Citysim: Comprehensive micro-simulation of resource flows for sustainable urban planning. [8](#)
- [47] Safe Software (2020). Fme workbench. [28](#)

- [48] Salvador, S. and Chan, P. (2004). Fastdtw: Toward accurate dynamic time warping in linear time and space. [23](#)
- [49] Sanyal, J., New, J., and Adams, M. B. (2017). Autosim: World’s fastest building simulator - energyplus workflows on high performance computing resources, version 00, u.s. copyright txu 2-141-960. [4](#), [25](#), [40](#)
- [50] Schliep, K., Hechenbichler, K., and Lizée, A. (2016). Package ‘knn’. [vii](#), [26](#)
- [51] Schödl, P. (2015). Scalable time series similarity search for data analytics. [xii](#), [24](#)
- [52] Science Learning Hub – Pokapū Akoranga Pūtaiao (2017). Climate models. [xi](#), [15](#)
- [53] Silverman, D. (2012). Emissions and radiative forcing models for ipcc 5 in 2013. [xi](#), [15](#)
- [54] Sokol, J., Cerezo Davila, C., and Reinhart, C. F. (2017). Validation of a bayesian-based method for defining residential archetypes in urban building energy models. *Energy and Buildings*, 134:11 – 24. [7](#)
- [55] Tian, W., Heo, Y., de Wilde, P., Li, Z., Yan, D., Park, C. S., Feng, X., and Augenbroe, G. (2018). A review of uncertainty analysis in building energy assessment. *Renewable and Sustainable Energy Reviews*, 93:285 – 301. [10](#)
- [56] US Department of Energy (2019). Commercial prototype building models. [13](#), [18](#), [35](#)
- [57] US Department of Energy (2020). Energyplus. [xi](#), [6](#), [25](#), [40](#)
- [58] U.S. Department of Energy (2020). Input output reference. [5](#)
- [59] US Department of Energy (2020). Openstudio. [25](#), [40](#)
- [60] US Energy Information Administration (2015a). A look at the u.s. commercial building stock: Results from eia’s 2012 commercial buildings energy consumption survey (cbees). [37](#), [39](#)
- [61] US Energy Information Administration (2015b). Use of energy explained: Energy use in commercial buildings. [ix](#), [67](#), [70](#)

- [62] US Environmental Protection Agency (2018). egrid summary tables 2018. [viii](#), [29](#), [30](#), [55](#)
- [63] Wayne, G. (2013). The beginner's guide to representative concentration pathways. [14](#)
- [64] Wilson, E. (2017). Resstock â€” targeting energy and cost savings for u.s. homes. [9](#)

Appendices

Table 1: US Commercial Prototype Building Multipliers - Multiplier Values used for scaling US commercial building energy to represent the total building stock.

ASHRAE	1A	2B	2A	3B	3C	3A	4B	4C	4A	5B	5A	6B	6A	7	8	Total
largeofficedetailed	6733	4019	21549	18833	7718	29439		10170	74886	8006	29259		8808	714		220134
mediumofficedetailed	8564	19319	53760	47312	8987	50661	2409	12956	78714	22628	70100	2304	19744	2182	452	400091
smallofficedetailed	5553	19132	70387	31411	5142	63693	3131	8106	61903	21289	60858	1999	15956	2139	310	371009
retailstandalone	14839	33515	146887	82723	12648	157835	7882	28338	168410	52421	226899	6033	62737	7181	900	1009246
retailstripmall	9094	16772	65543	41385	6838	67525	1490	7061	66706	13316	67664	1057	10099	454	89	375093
primaryschool	4213	10870	61754	29519	3149	62485	1990	6195	59208	14795	60881	2479	11120	1541	219	330418
secondaryschool	10600	15188	100737	54194	7244	125250	4185	16049	133185	28960	150955	5687	27474	4984	817	685508
hospital	2669	6336	31675	18032	2584	30950	1473	7005	40719	14448	53727	1567	14631	2230	87	228131
outpatient	2464	8877	37541	18181	4050	38451	1538	11967	54105	14407	70020	2202	22655	2578	136	289171
fullservicerestaurant	597	1625	6985	3140	381	7361	385	676	8412	2054	9465	265	2052	242	9	43650
quickservicerestaurant	517	1343	6061	4178	457	6735	352	947	5887	1697	8501	205	1678	236	15	38809
largehotel	7193	8264	41074	52454	6986	42021	2449	8106	63408	13212	60813	3817	14994	2534	237	327562
smallhotel	675	1972	19033	7523	1439	17713	1344	2597	20810	5913	24117	2038	7052	1356	257	113837
warehouse	23107	38402	171392	152055	10174	196261	4488	28813	161816	45541	236838	3241	30823	2853	146	1105951
highriseapartment	100657	5046	100023	49002	11456	43130		23678	165791	7632	76963	1062	8245	557		593241
midriseapartment	16998	6178	72363	57037	17233	54583	1451	24543	112070	21067	74219	3724	20730	2148		484343

Table 2: US Commercial Floorspace per Climate Zone - Commercial floorspace percentages per climate zone display the degree to which each zone influences total US energy.

Climate Zone	Percent of Commercial Floorspace
1A	3.24%
2A	15.22%
2B	2.98%
3A	15.03%
3B	10.08%
3C	1.61%
4A	19.29%
4B	0.52%
4C	2.98%
5A	19.37%
5B	4.34%
6A	4.21%
6B	0.57%
7	0.51%
8	0.06%

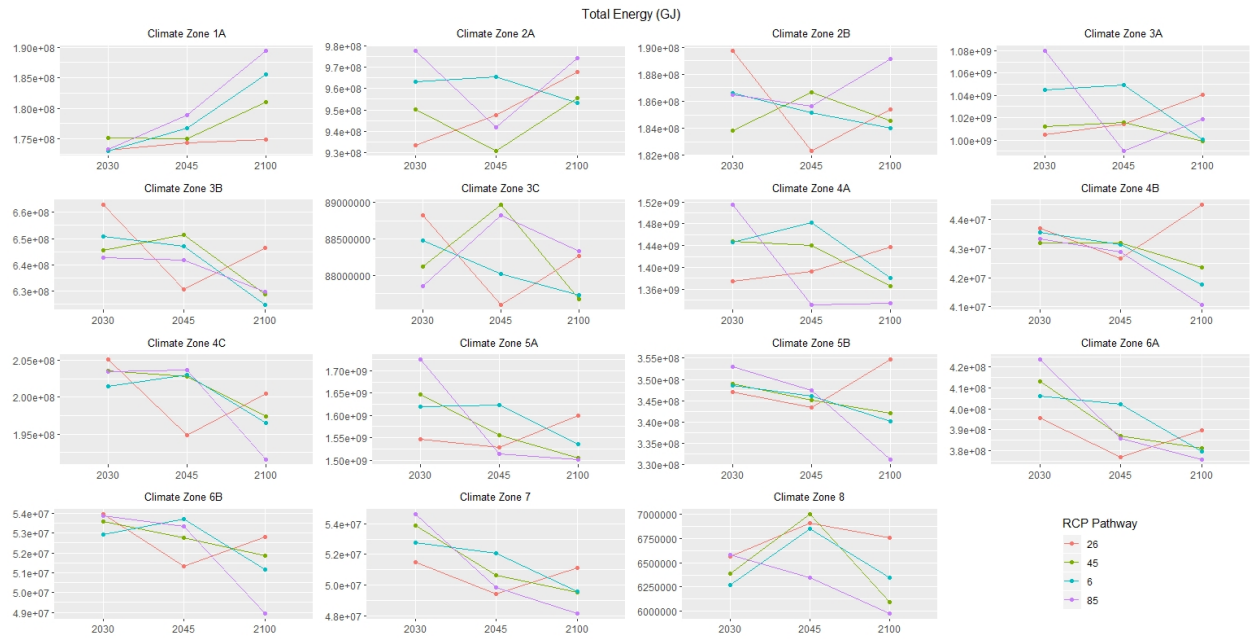


Figure 1: US Total Commercial Building Energy Use for RCPs By Climate Zone - Total energy by by climate zone changes are related to the temperature of each zone. Colder zones total energy decreases as average global temperatures increase while warmer zones total energy increase along with the global temperature. The impact of the zones with the most commercial buildings can be seen with the order of magnitude of zones 2A, 3A, 4A, and 5A.

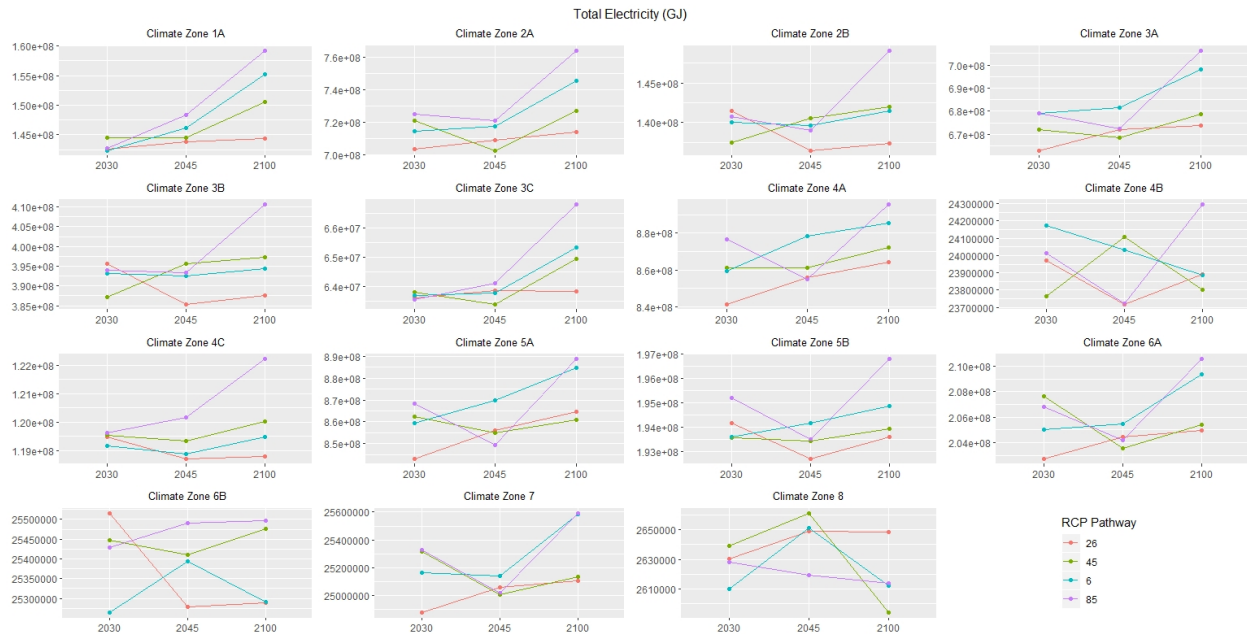


Figure 2: US Total Commercial Building Electricity Use for RCPs By Climate Zone - Total electricity results by climate zone are also affected by the temperature of each zone. Warmer climates see huge increases in electricity under low mitigation scenarios. The electricity decrease for most of the zones with the most commercial floorspace (3A, 4A, 5A) is what leads to the overall electricity decrease in the US for higher emission scenarios.

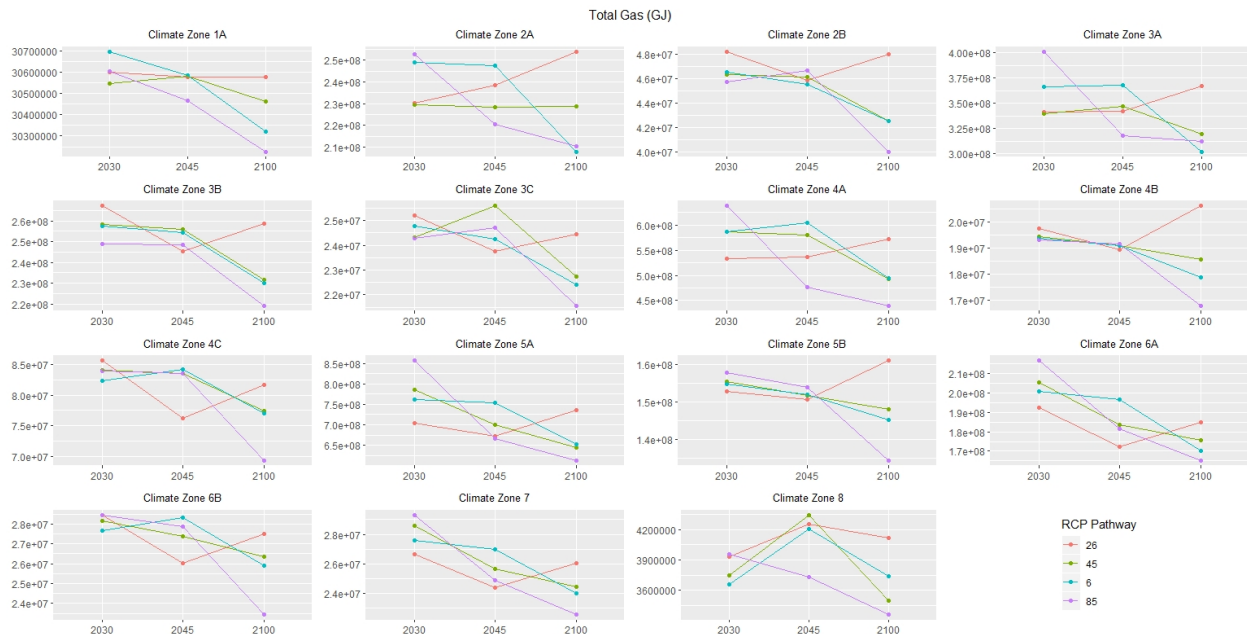


Figure 3: US Total Commercial Building Natural Gas Use for RCPs By Climate Zone - Total natural gas results by climate zone are affected by the temperature of each zone. All climates see decreases in natural gas under low mitigation scenarios. The increase in temperature under these scenarios leads to less heating energy use.

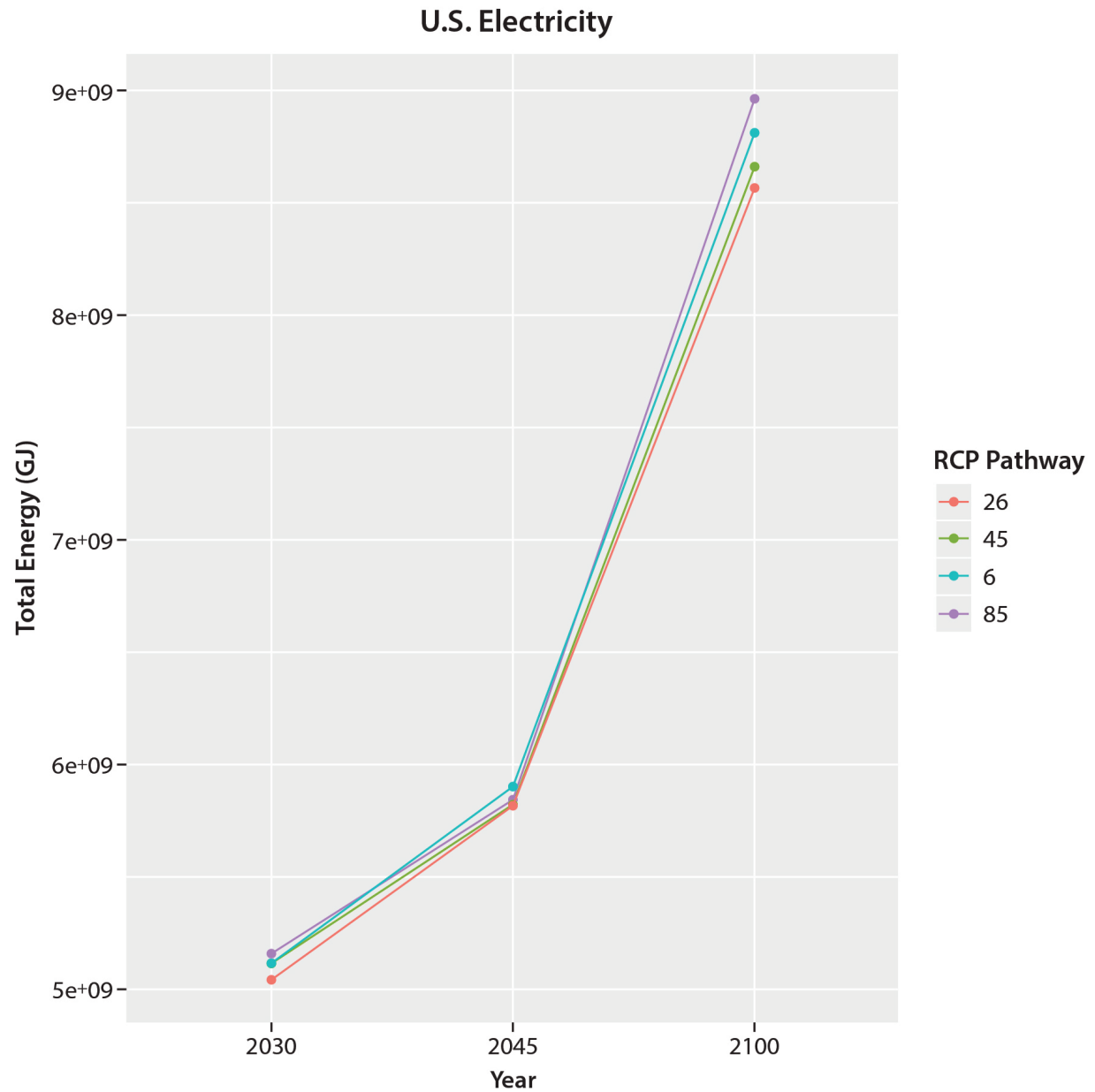


Figure 4: US Total Commercial Building Electricity Use for RCPs and Increasing Commercial Floorspace - The total electricity use with increasing commercial floorspace increased by at least 55% for each scenario from 2030 to 2100.

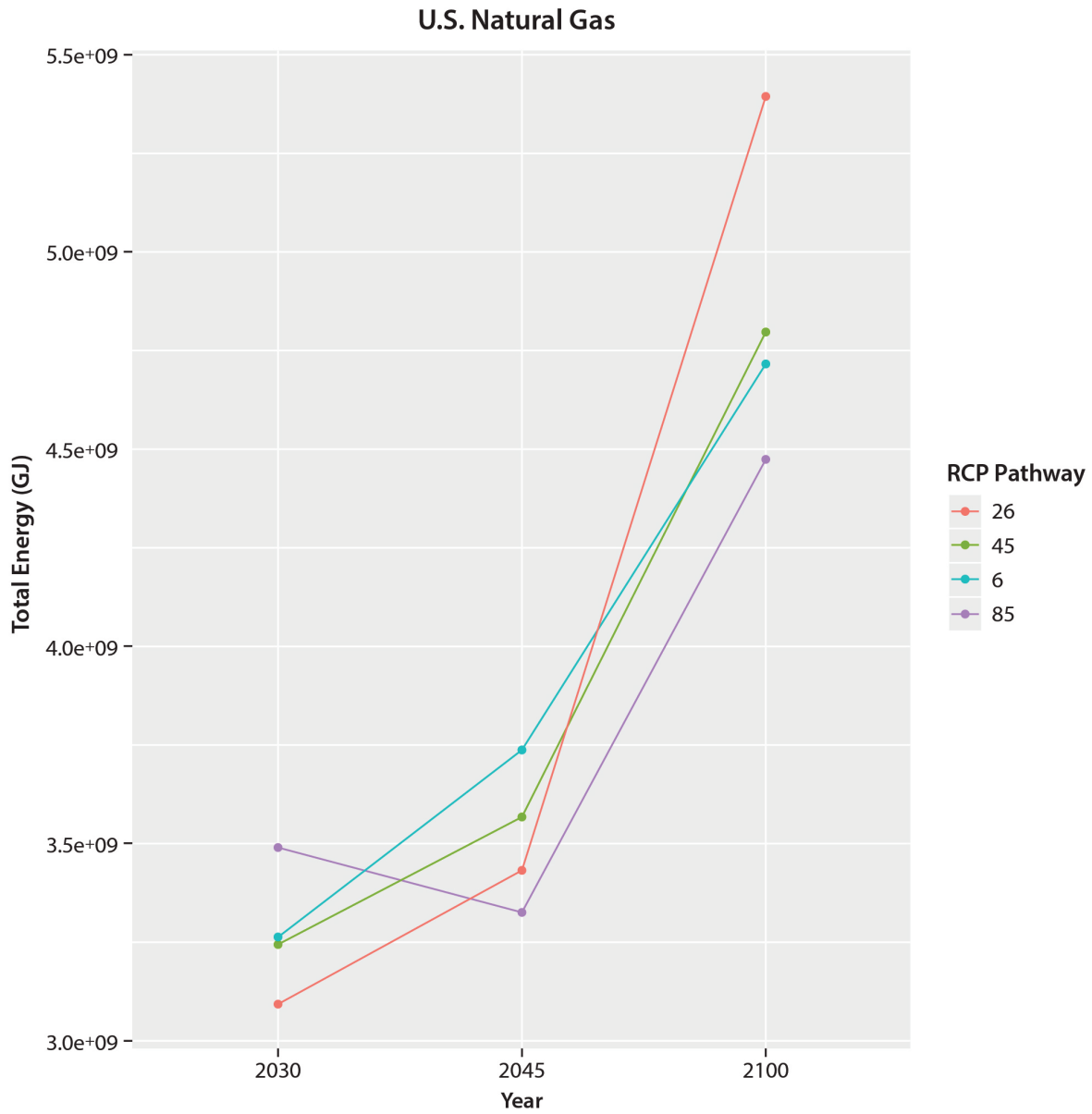


Figure 5: US Total Commercial Building Natural Gas Use for RCPs and Increasing Commercial Floorspace - The total natural gas use with increasing commercial floorspace increased for all scenarios but the increase was much less than the electricity change.

Vita

Brett Bass was born near Cleveland, Ohio in 1994 and raised in Avon Lake, Ohio. Before attending the University of Tennessee, Knoxville, he attended the University of Dayton where he earned a Bachelor of Arts in Mechanical Engineering in 2016 and a Master of Science in Renewable and Clean Energy Resources. in 2017, While at Dayton he received The class of 1902 Award of Excellence for Outstanding Mechanical Engineering Achievement and was on Dean's list all semesters while playing division 1 football. Brett's focus shifted from engineering to data while working for Emerson Climate Technologies in 2017, on projects using data to predict and prevent failure in refrigerated cargo containers. He chose to attend the University of Tennessee, Knoxville to pursue a Doctor of Philosophy in Data Science and Engineering. His research interests while at the University of Tennessee include urban-scale building energy modeling and climate modeling.