ABSTRACT

Title of dissertation: DETERMINING MEASUREMENT REQUIREMENTS FOR WHOLE BUILDING ENERGY MODEL CALIBRATION

Matthew Dahlhausen
Doctor of Philosophy Mechanical Engineering, 2020

Dissertation directed by: Professor Jelena Srebric, Ph.D.
Department of Mechanical Engineering

Energy retrofits of existing buildings reduce grid requirements for new generation and reduce greenhouse gas emissions. However, it is difficult to estimate energy savings, both at the individual building and entire building stock level, because building energy models are poorly calibrated to actual building performance. This uncertainty has made it difficult to prioritize research and development and incentive programs for building technologies at the utility, state, and federal level. This research seeks to make it easier to generate building energy models for existing buildings, and to calibrate buildings at the stock level, to create accurate commercial building load forecasts. Once calibrated, these building models can be used as seeds to other building energy model calibration approaches and to help utility, state, and federal actors to identify promising energy saving technologies in commercial buildings. This research details the economics of a building energy retrofit at a singular building; contributes significantly to the development of ComStock, a model of the commercial building stock in the U.S.; identifies important parameters for calibrating ComStock; and calibrates ComStock for an example utility region of Fort Collins, CO against individual commercial building interval data. A study of retrofit costs finds that measure cost and model uncertainty are the most significant sources of variation in retrofit financial performance, followed financing cost. A wide range of greenhouse gas pricing scenarios show they have little impact on the financial performance of whole building retrofits. A sensitivity analysis of ComStock model inputs across an exhaustive range of models identifies 19 parameters that explain 80% of energy use and 25 parameters that explain 90% of energy use. Building floor area alone explains 41% of energy use. Finally, a comparison of ComStock to Fort Collins, CO interval meter data shows a 6.92% normalized mean bias error and a 16.55% coefficient of variation of root mean square error based on normalized annual energy per floor area. Improvements in meter classification and ComStock model variability will further improve model fit and provide an accurate means of modeling the commercial building stock.

DETERMINING MEASUREMENT REQUIREMENTS FOR WHOLE BUILDING ENERGY MODEL CALIBRATION

by

Matthew Dahlhausen

Dissertation proposal submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2020

Advisory Committee:
Professor Jelena Srebric, Ph.D., Chair
Professor Reinhard Radermacher, Ph.D.
Research Professor Yunho Hwang, Ph.D.
Professor Bao Yang, Ph.D.
Professor Donald Milton, MD, DrPH, Dean's Representative

Dedication

To my parents.

Acknowledgments

No dissertation is a monument of one. Many people helped me along the journey, and some deserve special thanks.

Dr. Srebric, my advisor, has encouraged me since I started as a masters student in her lab at Penn State in 2012. She has witnessed and guided me through my whole career in building energy modeling so far. I appreciate her for her patience in allowing this work to come together over time and helping me develop my research and modeling skills.

My lab mates, especially Mohammad Heidarinejad, Daniel Dalgo, Yang-Seon Kim, and Nick Mattise have given an enormous amount of their time to work on projects together and provide feedback to my research, journal articles, and presentations.

My co-workers at Integral Group, especially Stefan Gracik and Stet Sanborn, taught me a lot about using building energy modeling to design net zero energy buildings. Much of what I learned about modeling HVAC systems at Integral Group is embedded in openstudio-standards and ComStock. Integral Group also supported me financially to continue my dissertation.

NREL, my current employer, deserves enormous thanks. They have funded the last few years of dissertation work. I could not have completed this dissertation without the building energy modeling tools EnergyPlus and OpenStudio that NREL develops, and I'm honored to be in the role of working on the code base directly, particularly managing openstudio-standards. My NREL coworkers, especially Andrew

Parker, reviewed a draft, and Anthony Fontanini, Ry Horsey, Lixi Liu, Rajendra Adhikari, Rawad El Kontar, Janghyun Kim, Chris CaraDonna, Amy LeBar, and Marley Praprost, were critical in getting ComStock running, analyzing enormous quantities of end use data, and allowing me space to work this dissertation around my NREL projects. Special thanks also goes to the U.S. Department of Energy which funded the End Use Load Profiles project, which this work contributed to.

Three individuals are largely responsible for my career as an engineer focusing on energy efficiency. My high school physics teach, Mr. Robert Shurtz, remains the best teacher I've had and gave me a strong foundation in physics and math. In college, Professor Charles Sullivan inspired me to pursue a career in building energy efficiency and approached learning with a care and depth that made mastery of complex concepts easy. Ivar Frislid taught me everything about blower door tests and residential energy use and gave me an invaluable understanding how buildings actually work away from models and design drawings.

Lastly, I need to thank friends and family for their support including Becky Miller and Mikhaila Clements, my mom Elizabeth Mease, my father Michael Dahlhausen and his partner Felice Dahlhausen, Norma Smith, a second mother to me, my brother Tom Dahlhausen, and my sister Katie Dahlhausen who completed a dissertation concurrently.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACH	Air Changes per Hour
AHU	Air Handling Unit
AMI	Advanced Metering Infrastructure
ASHRAE	American Society of Heating, Refrigeration, and Air Conditioning Engineers
BC	Bayesian Calibration
BEM	Building Energy Modeling
BMS	Building Management System
CBECS	Commercial Buildings Energy Consumption Survey
CDD	Cooling Degree Day
CVRMSE	Coefficient of Variation of the Root Mean Square Error
DOAS	Dedicated Outdoor Air System
DX	Direct Expansion
ESCO	Energy Service Company
EEM	Energy Efficiency Measure
EFLH	Equivalent Full Load Hours
EPD	Equipment Power Density
EUI	Energy Use Intensity
EVO	Efficiency Valuation Organization
HDD	Heating Degree Day
HVAC	Heating, Ventilation, and Air Conditioning
IAQ	Indoor Air Quality
IPMVP	International Performance Measurement and Verification Protocol
LBNL	Lawrence Berkeley National Laboratory
LEED	Leadership in Energy and Environmental Design
LHS	Latin Hypercube Sampling
LPD	Lighting Power Density
LRD	Load Research Data
MCMC	Markov Chain Monte Carlo
NREL	National Renewable Energy Laboratory
NMBE	Normalized Mean Bias Error
OA	Outdoor Air
PNNL	Pacific Northwest National Laboratory
QOI	Quantity of Interest
RTU	Roof Top Unit
SA	Sensitivity Analysis
SWH	Service Water Heating
VAV	Variable Air Volume
VFD	Variable Frequency Drive

Chapter 1: Introduction

1.1 Energy Use in Commercial Buildings

Energy savings in existing buildings is an area for sizable, cost-effective energy use reduction and greenhouse gas mitigation for economic and environmental benefits [1,2]. Buildings are capital intensive and long-lasting, with a median lifetime of 70 years [3]. A building is often renovated over the course of its lifespan to incorporate new uses, tenants, and bring the building up to code. 86% of all construction costs go toward renovations of existing buildings rather than new construction [4], yet renovation rates are only 2.2% of the annual building stock, with an 11% average energy savings [5]. The expense and scale of this infrastructure means that large scale shifts in energy use and predominant fuel sources take decades. Without increased investment and focus on energy retrofit of existing buildings, as much as 80% of 2005 thermal energy consumption may remain past 2050 [6]. Along with decarbonization of the electric grid, the building renovation rate will need to increase several times over, and average energy savings per project improve to >50% energy savings, to meet greenhouse gas emission reduction targets and Architecture 2030 goals [5]. Few projects in the U.S. have achieved this level of energy savings, with one recent study identifying only 50 such projects, known as deep or advanced energy retrofits [7, 8]. There is a need to understand why these projects are uncommon, what factors influence savings, and develop analysis tools to make these projects more feasible.

1.2 Energy Retrofits

The paucity of deep energy retrofit projects can be attributed to a host of factors. Chief among these is the codependent influence of capital constraints and investment uncertainty. Lack of access to capital, insufficient payback, and energy savings uncertainty are the top barriers to making energy retrofits more prevalent [9, 10]. Most projects are funded with limited internal capital, sometimes with assistance from grants, rebates, and other incentives. These projects have tended toward specific lighting, controls, and Heating, Ventilating, and Air Conditioning (HVAC) equipment measures with reliable savings, as it can be very expensive to go through an extensive energy audit to identify further measures and may not significantly reduce the energy savings uncertainty. When many Energy Efficiency Measures (EEMs) are done together, savings interaction will increase the overall savings uncertainty. EEMs are often selected by simple-payback which results in good financial payback on a per-measure basis, but this ranking does not consider how measure integration can unlock greater energy savings by lowering heating and cooling loads to enable alternative mechanical systems. Lower heating and cooling loads enable downsizing of central mechanical equipment, which allows the significant equipment replacement cost savings to be used to pay for other measures.

This means that choosing measures with optimal payback individually may not yield the optimal retrofit decision overall. The uncertainty in savings for larger retrofit projects targeting deeper savings has meant greater expense in measurement and selecting EEMs to be able to verify savings. The larger the project scope and options considered, the greater the expense in data collection. It takes an enormous amount of effort to develop a fully calibrated energy model of a building that can be used to accurately determine measure savings. In most buildings, the data is not available. But even in buildings with newer building management systems (BMS) that trend hundreds of points, the sensors go out of calibration, there may be multiple systems to collect data in different formats, and the effort to clean, process, and make this data usable for actionable decisions requires an expense that is well beyond the budget available. Faced with this challenge, an energy auditor must triage data collection and may omit EEMs simply because it is too difficult to estimate savings. Overall, uncertainty and capital budgets make energy retrofits an economic problem, not just a physical one.

1.3 Stock Models

The challenge of energy retrofits and low energy design extends to the entire commercial building stock. Utilities, states, and governments are under increasing pressure to decarbonize the electric grid through a mix of energy efficiency and renewable energy generation. Knowing which technologies to support for further research and development and which technologies to scale up through market incentive programs is critical to this goal. From the individual building up to the commercial building stock, there is need for representative, calibrated building energy models to use a starting point for individual building energy model calibration and to best estimate retrofit technologies for large scale development and deployment.

1.4 Dissertation Structure

This chapter, (Chapter 1), frames the issue of energy use in buildings and the difficulties of energy retrofits. Chapter 2 presents a literature review of energy retrofit analysis methods, building energy modeling (BEM), and calibration approaches. Chapter 3 presents the research hypothesis and objectives of the dissertation. Chapter 4 presents methodology for analyzing energy retrofits under uncertainty, and a methodology for quickly creating exploratory reduced-order building energy models that will later be used in the calibration process. The chapter also includes a description of the proposed calibration methodology for targeting data collection and details how this will be tested. Chapter 5 details ComStock, a commercial building stock modeling tool. Chapter 6 studies the feature importance in ComStock to develop priorities to investigate further for calibration. Chapter 7 presents an example calibration of ComStock for Fort Collins, CO. Lastly, Chapter 8 summarizes the work and contributions made to the field of building energy modeling and explores how this work can be used in future analysis.

Chapter 2: Literature Review

This chapter presents an overview of methodologies for selecting EEMs, how energy savings are calculated, the basics of building energy modeling, and the approaches to calibrating building energy models.

2.1 Energy Retrofits

Currently, in industry, it is common for energy efficiency measures to be selected based on the preferences of the energy auditor or facility manager. Figure 2.1 shows how this expert process can produce very different recommendations for a given building, with measure costs and payback estimates varying by a factor of two [11]. In cities where audits of commercial buildings are required, audits have a small impact on reducing energy use, with primary savings limitations being audit quality, limited access to capital, and uncertainty in energy savings projections [12].

The energy retrofit literature has sought to improve energy audit and recommendation repeatability and reliability by establishing methodologies for choosing EEMs in energy retrofits [13]. However, only a subset of the literature considers the integrative aspects of measure selection, and even fewer consider the uncertainty involved in both endogenous factors to the building (baseline parameter un-

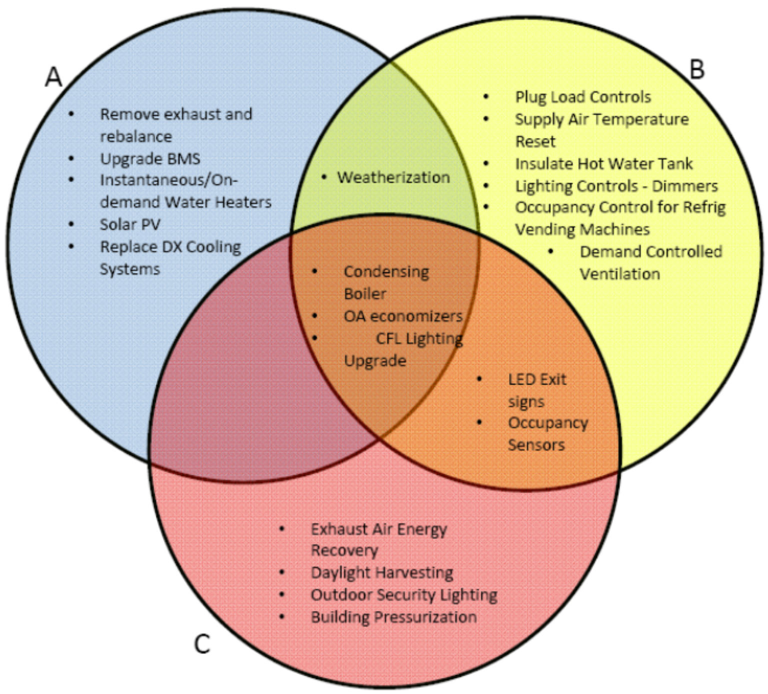

Figure 2.1: This graphic presents EEM recommendations for a office building in Philadelphia, PA, from three different energy auditors [11].

certainty, measure performance, and building use) and exogenous factors (energy price, weather, energy policy, and greenhouse gas policy variability). These factors are listed in Table 2.1.

Table 2.1: Endogenous and Exogenous Factors Influencing Retrofit Savings

Endogenous Factors	**Exogenous Factors**
Factors internal to the building and building operation	*Factors external to the building*
Baseline model parameter uncertainty Measure performance Building use and operational changes	Weather Measure cost Energy price Government regulations and policy Greenhouse gas policy

One recent study demonstrated how the ideal measure package changes with uncertainty of technology performance, capital costs, energy prices, carbon tariffs, and grid decarbonization [14]. This study also evaluated the range of outcomes under three decision criteria: maximum weighted average of options, maximum under the most pessimistic scenario, and the smallest regret to minimize difference in expected outcome. This approach captures the interactions among retrofit measures, and found that technology performance, capital costs, and energy prices caused the most significant difference in financial performance. This study implemented all measures at once, which is not always feasible, depending on available capital and desire to wait until end-of-life to replace equipment. Another study demonstrated a process to implement measures in a package ordered depending on capital availability [15]. This approach reduced financial risk exposure of a large retrofit project by staying within an internal budget, but it did not consider what measure package would result in the optimal savings. In both approaches, measure integration and packaging are important. Interestingly, extending the project timeline incorporates major equipment replacements, that are already embedded in capital plans, into a comprehensive retrofit package. This allows targeted load reductions to precede equipment replacement, which can reduce the equipment cost for replacement equipment, with the potential cost of forgoing possible energy cost savings. This trade-off between replacing equipment before the expected end of its service life and forgoing possible energy savings from implementing measures sooner is an important consideration in energy retrofits and deep energy retrofits. This trade-off has not been explored in depth to see how it may influence measure selection and constitutes the first research objective.

2.2 Energy Efficiency Evaluation

After EEMs are selected, there is a separate approach in the industry to determine actual energy savings, known as Measurement and Verification. MEasurement and verification standards are governed by the International Performance Measurement and Verification Protocol (IPMVP) [16] and established in ASHRAE Guideline 14 [17]. There are several methods for determining energy savings for an energy retrofit project, detailed in Table 2.2.

The retrofit isolation approaches are most common for small, one-off, easy to measure EEMs, like installing VFDs on pumps and fans, or replacing older T8 or T12 fluorescents with T5 or LED lamps. However, for larger projects that involve

Table 2.2: IPMVP Energy Savings Estimation Methods [16]

IPMVP Option	Savings Calculation	Application
Retrofit Isolation	Isolated measurement of key parameters influencing EEM	Single, simple EEMs that are easily metered before and after implementation. Does not include measure interaction.
Whole Facility	Regression model of whole building utility data	Major renovation targeting many building systems. Unable to estimate savings a priori.
Calibrated Simulation	EEMs implemented in an energy model calibrated to whole building utility data	Major renovation targeting many building systems; changes expected in building use.

multiple, interactive EEMs, it is not possible to isolate and properly attribute energy savings to individual EEMs. To capture savings for larger retrofit projects where interactive effects are substantial, there are two different approaches. A whole-building statistical *black-box* model approach, where building utility data is used with other predictors to construct a generalized regression model to predict energy use as if the baseline building kept operating, and a *physical model* approach where the building physics are represented and EEMs and overall savings are modeled explicitly [18,19]. These two methods are explained in further detail in the following sections.

2.2.1 Measurement and Verification Model Calibration Standards

Whole building energy models are simplified representations of all the complex parameters and interactions that determine how much energy a building uses, including weather, occupants, and innate characteristics of building systems. In measurement and verification, the whole-building model is matched to energy data, creating a *baseline* building model, which is the building before EEMs are implemented. This model is then used to predict energy use in the period after retrofit, using real weather and occupancy from the post-retrofit period. This prediction is compared against actual utility data to determine energy savings.

To make an accurate savings estimate, the baseline model must be a good predictor of baseline building energy use. The measurement and verification standards community has established Normalized Mean Bias Error (NMBE) and Coefficient of Variation of the Root Mean Square Error (CVRMSE) as the two metrics to establish model calibration, shown in to Eqns.2.1 and 2.2 [17].

$$CVRMSE = 100\frac{\sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{(n-p)}}}{\bar{y}} \tag{2.1}$$

$$NMBE = 100\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)}{(n-p)\bar{y}} \tag{2.2}$$

where y_i is the utility data for time i, $\bar{y}$ is the mean of all utility data, $\hat{y}$ is the simulated energy use for timestep i, n is the number of datapoints (12 for monthly and 8760 for hourly annual calibration), and p is the number of parameters in the baseline model, taken to be 1 for model calibration. The discrepancy between simulated energy use and utility data must have a CVRMSE and NBME lower than the standards specified in Table 2.3.

NBME ensures that the overall energy use in the model is matched to the real building, so as not to grossly over or under predict aggregate savings. CVRMSE

Table 2.3: Minimum calibration criteria specified by standards organizations [17]

	Monthly Criteria (%)		Hourly Criteria (%)	
Standard or Guideline	NMBE	CVRMSE	NMBE	CVRMSE
ASHRAE Guideline 14-2002 [17]	5	15	10	30
IPMVP [16]	-	-	5	20

ensures that the model does not over- or under-predict seasonal dependent uses such as HVAC energy in monthly data, and occupant or lighting related energy use in hourly data. Recently, there has been some discussion about the suitability of these metrics to establish model calibration, because while they ensure good agreement with utility data (low output-side error), this agreement can be reached along with large discrepancies between specific input values and the real values in a building (high input-side error) [20]. This discrepancy is primarily a concern for physical models, and is a focus of this dissertation.

2.2.2 Black-Box Models for Predicting Energy Savings

Black-box models are the most common choice for measurement and verification, as they are relatively simple compared to physical models, and most just need utility data and weather data to create a good prediction. Most common are regression models, where utility data is modeled with a generalized linear model with terms for outdoor air temperature, day-of-week, hour-of-day, month, and other such variables [21]. These models meet calibration standards for most buildings, with 75% or more buildings meeting calibration criteria, with an exception of CVRMSE for a mean-week model that predicts energy use solely as an average of time-of-week [21]. Ongoing research seeks to compare and establish wide ranging testing for measurement and verification models [22] to improve model accuracy and reduce the largest errors. While useful for measurement and verification, black-box models cannot *predict* energy savings as can a physical model, and so cannot be used to evaluate EEM options in the first place. If the model relies on data such as occupancy, which may change substantially in the post-retrofit period, or insufficient baseline energy data exists to generate a regression model, a physical model is necessary.

2.2.3 Physical Building Energy Models for Measurement and Verification

Building Energy Modeling (BEM) was developed in response to the Arab oil embargo in the 1970s to calculate thermal load and energy requirements for buildings [23]. BEM tools have since matured and are used in many applications in the buildings industry to make early design decisions, perform load calculations, optimize controls, demonstrate code compliance, and evaluate energy efficiency programs. Tools range from simple spreadsheets to integrated solvers, such as EnergyPlus, that couple thermal processes in building envelopes with HVAC system models [24]. A detailed comparison of building energy simulation software is available in [24]. This dissertation uses EnergyPlus [25] and OpenStudio [26], an Application Programming Interface (API) for EnergyPlus, as they are are open-source tools under active development, and the only such open-source tools at present with

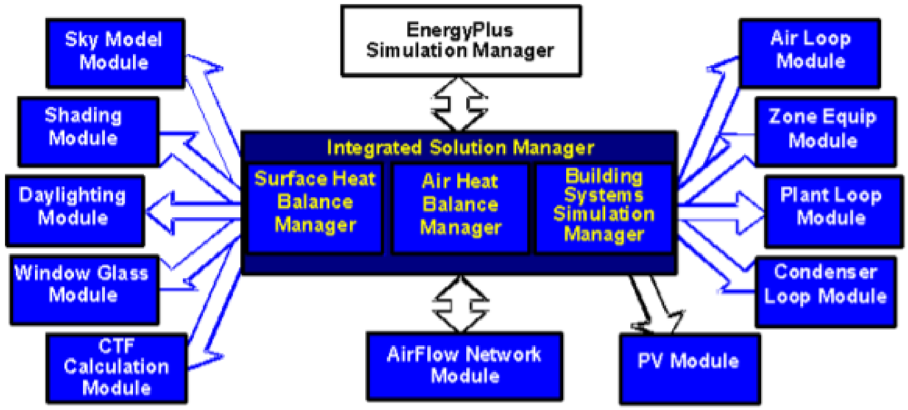

Figure 2.2: Integrated solvers of the EnergyPlus solution manager [25]

substantial analysis and cloud computing support.

EnergyPlus is an integrated ODE solver for calculating air and surface temperatures in a zone from convective, conductive, and radiative exchange, illustrated in Figure 2.2. The core of the solver assumed a zonal mixed air model which iteratively calculates a zone's air temperature based on internal loads, surface exchange, and air mixing, as shown in the governing Eqn. 2.3.

$$C_z \frac{dT_z}{dt} = \sum_{i=1}^{N_{sl}} \dot{Q} + \sum_{i=1}^{N_{surfaces}} h_i A_i (T_{si} - T_z) + \sum_{i=1}^{N_{zones}} \dot{m}_i C_p (T_{zi} - T_z) + \dot{m}_{inf} C_p (T_\infty - T_z) + \dot{Q}_{sys} \tag{2.3}$$

where:

$C_z \frac{dT_z}{dt}$ = energy stored in zone air

$C_z = \rho_{air} C_p C_T$

ρ_{air} = zone air density

C_p = zone air specific heat

C_T = sensible heat capacity multiplier

T_z = zone air temperature

$\sum_{i=1}^{N_{sl}} \dot{Q}$ = sum of zone convective internal loads

$\sum_{i=1}^{N_{surfaces}} h_i A_i (T_{si} - T_z)$ = convective heat transfer from zone surfaces

$\sum_{i=1}^{N_{zones}} \dot{m}_i C_p (T_{zi} - T_z)$ = heat transfer from interzone air mixing

$\dot{m}_{inf} C_p (T_\infty - T_z)$ = heat transfer due to infiltration of outside air

T_∞ = outdoor air temperature

$\dot{Q}_{sys} = \dot{m}_{sys} C_p (T_{sup} - T_z)$ = system energy provided to the zone

$\dot{m}_{sys}$ = air system mass flow rate

T_{sup} = supply air temperature

The air system assumes perfect operation, i.e. the terminal units if present are controlled ideally, and the plant systems can meet load up to their capacity if needed. Plant systems are approximated with performance curves for all types of equipment, including chillers, cooling towers, boilers, fans, pumps, DX heating and cooling coils, etc. There are hundreds of parameters that go into specifying a full EnergyPlus model. Most parameters are singular values. However, internal loads are modeled with a peak value multiplied by a fractional schedule, and most HVAC equipment performance is specified by quadratic, cubic, or biquadratic curves with specific curve coefficients. Schedule values and curve coefficients are difficult to specify as singular variables in calibration, and thus tend to be excluded from calibration workflows.

Achieving model calibration is comparatively much more difficult for physical building energy models than for black-box regression models. This is because of the hundreds of available parameters to tune and match to the baseline building. Most importantly, it is possible to achieve calibration criteria without good agreement between model parameters and real building parameters, as the quantity and similarity of some input parameters can lead to model over-fitting [20, 27]. Section 2.3 goes into greater detail on model calibration and presents current solutions to the over-fitting problem.

2.2.4 Summary of Energy Savings Measurement and Verification

Energy savings estimation and measurement and verification for energy retrofits includes a range of analysis methods. Of the analysis methods, only building energy modeling - representation of the physical model of the building explicitly - can be used to predict energy savings from a group of EEMs that are highly interactive. BEM is used in this dissertation to explore EEM selection strategies, develop reduced-order models of buildings to prioritize energy auditing scope, and estimate savings uncertainty from implementing a retrofit measure package. Many BEM models together are used to construct a model of the national building stock.

2.3 Building Energy Model Calibration

2.3.1 Overview of Approaches

In order for the building energy model to be used for the retrofit savings prediction, the model must be calibrated to accurately represent the baseline building conditions. Calibration involves matching the input model parameters as close to reality as possible, from on-site audits, standards, and industry benchmarks, so that the model accurately predicts the energy use in the baseline period. There have been many varied approaches to building energy model calibration [18,28,29]. When done well, the building energy model becomes a good representation of the building and can even be used for model-based commissioning to identify operational issues in HVAC systems or other faults [30,31]. However, it is rare for a model at first pass to accurately predict building performance. Model errors arise from four distinct reasons: specification uncertainty [18,32,33], parameter uncertainty [18,34], numerical uncertainty [18,32], and scenario uncertainty [18,35], detailed in Table 2.4.

Table 2.4: Causes of Building Energy Model Uncertainty

Uncertainty Source	Description	Typical Influence on Model
Specification Uncertainty	Model misspecification comes from simplifying physical representations of the building, such as combining zones, omitting or simplifying pieces of mechanical equipment, and setting up controls incorrectly	large; (>100% errors possible)
Parameter Uncertainty	many physical properties or efficiencies in buildings are hard to know exactly. Some, like infiltration, are very difficult to measure and have wide variations over similar building types. This also includes sensor and measurement uncertainty of energy and BMS data	significant source of error; large (>100% errors possible) when multiple parameters are far from true values. Sensor error is small, typically 3 to 5% [32]
Numerical Uncertainty	Numerical uncertainty is caused by modeling continuous thermal processes in discrete timesteps. If timesteps exceed typical response rates, or control logic in the physical system, this can lead to improper interpolations and miss things like equipment cycling. This is controlled with smaller timesteps and more accurate control logic	small; errors typically <1% when timesteps are <10 min, though can be larger if equipment short-cycling is present.
Scenario Uncertainty	Variables like future weather and occupant behavior are inherently uncertain. These can be analyzed in different future scenarios, but not appreciably reduced	medium; errors can be significant for behavioral influences [35], and weather can vary considerably year by year.

As there are many sources of uncertainty, model calibration is typically done by a building science expert manually adding or removing model components and adjusting parameter values to improve calibration to meet the criteria given in Table 2.3 [28,36–38]. Modelers focus on values that are highly influential in the simulation, with common parameters varying by building type and location [20,27,30,37,39–41]. This process is very time consuming, so many have proposed automated methods to reduce parameter uncertainty [19, 28].

Automated methods typically approach the problem by first running a sensitivity analysis to determine influential model parameters [42], and then once sensitive parameters are determined, using optimization algorithm to find the parameter values that best reduce NMBE and CVRMSE [43, 44]. This process of using sensitivity analysis followed by optimization is detailed in ASHRAE RP-1501 [45].

There are, however, several complications with this approach. Most importantly there is model over-fitting, also known as the identifiability problem, where in a model with multiple parameters (hundreds in building energy models) many combinations can produce the same outcome, especially if the output space is small (e.g., monthly utility data) [18]. This is especially true with building occupants, lighting, and plug load parameters that have many degrees-of-freedom in fractional schedules and load values [35]. As there are many combinations of input parameters than can achieve good performance on calibration criteria, it is easy for parameters to get stuck at local minimums [32] or at parameter constraints that ultimately create a bias error when evaluating EEMs [18, 46, 47]. This can be partly mitigated, but not avoided, by carefully setting parameter constraints [18, 42].

Model over-fitting is largely a fault of using monthly utility data for calibration. Monthly data is used because utilities, namely electricity and gas, are billed on a monthly cycle. While 15 minute interval data is now common, gas data is still kept in monthly increments for most buildings. This aggregation to monthly data is a significant numerical uncertainty problem that makes it impossible to differentiate the influence of schedules versus parameter values. It does not provide sufficient resolution to understand the main drivers for electric energy end-uses [48] which can be important in selecting and estimating savings from energy efficiency measures [49]. Using monthly data in calibration can miss major discrepancies in model specification, such as fan curves and equipment capacities, such that values are tuned to reach calibration performance, even though the model is poorly specified [27]. One study showed an *average* hourly bias of 48% HVAC energy use from using monthly data in calibration [31], meaning many times the model over- or under-predicted HVAC energy use by 48%. Calibration with monthly data can mask or introduce faulty parameter values that may be critical to EEM analysis.

For retrofit analysis, it is insufficient to just show a good match to **output-side** utility data. A good calibration will also show good model correlation to **input-side** data, meaning the parameter values important for retrofit analysis are in good agreement with real values. One study found that calibration to just monthly data with ASHRAE Guideline 14 [17] results in a less than a 40% correlation to real input parameters [36].

The issues with deterministic approaches and inability to reliably produce input-side parameters have inspired the use of stochastic models that can better capture input parameter uncertainty, and therefore energy savings uncertainty, from an energy retrofit. In particular, Bayesian calibration of models is a means of developing probability distributions for model parameters [50].

2.3.2 Bayesian Calibration

Bayesian calibration for building energy models is based on the work of Kennedy and O'Hagan [50], and detailed in several articles and theses [32, 51–54]. It is based on Bayes theorem, given in equation 2.4, which states that the probability of parameters θ occurring given observed data y is proportional to the prior probability distributions of parameters $p(\theta)$ times the likelihood function $p(y|\theta)$:

$$p(\theta|y) \propto p(\theta)p(y|\theta) \tag{2.4}$$

This is expressed in mathematical form by equation 2.5 [48, 55]:

$$y(x) = \eta(x, \theta) + \delta(x) + \epsilon(x) \tag{2.5}$$

where $y(x)$ are energy measurement observations, $\eta(x, \theta)$ is a representation of the energy simulation at known conditions x and calibration parameters θ, $\delta(x)$ is a model inadequacy parameter meant to represent the discrepancy between the model and the real building, and $\epsilon(x)$ is an error term for random measurement error. $\delta(x)$ is modeled as a Gaussian process, which interpolates between model points and gives a probabilistic representation of the energy model. Measurement error $\epsilon(x)$ is assumed to be mean zero and independently distributed. Detailed representation is given in [53,55].

The overall steps of the Bayesian calibration process are as follows:

1. **Step 0)** Gather a list of prior uncertainty estimates for model parameters [54, 56]. These are typical ranges of values for common model parameters and their distributions. Typical distributions are either uniform or triangular. Prior uncertainty values are difficult to measure, and come from standards or existing literature [32–34, 57–59].

2. **Step 1)** Create an initial model from building data for use in calibration. The detail and starting accuracy of the initial model can influence the calibration results if the model contains significant misspecification errors, especially for optimization-based deterministic methods [32, 36]. This will place bounds on uncertain parameters.

3. **Step 2a)** Run a parameter screening with a sensitivity analysis technique such as the Morris method [32, 46, 55, 56, 60]. A discussion of sensitivity analysis methods is available in Hopfe [57] and Menberg [40]. The Morris method is decent at producing reliable parameter rankings quickly, though it does not capture all complex parameter interactions that may be present in a model. The goal of the parameter screening is to reduce the number of parameters that will be included in the calibration. Non-influential parameters will not change much in the process, but will add substantial simulation time.

4. **Step 2b)** Optionally use optimization to estimate starting parameter values

and ranges for uncertain parameters.

5. **Step 3)** Run an energy model [32, 56, 61] with Latin Hypercube Sampling (LHS) to generate a large set of values that cover the parameter ranges.

6. **Step 4)** Combine input, output, and observation data in a Gaussian Process Emulator (GPE) that gives a probabilistic representation of the model [32,62]. There are some comparisons of which GPE to use [63,64], and this is an ongoing area of research.

7. **Step 5)** Implement Bayes Theorem on a subset of output data with a Markov Chain Monte Carlo (MCMC) method and iterate until a convergence criteria is achieved to generate the posterior probability distribution. Chong [52] compares No-U-Turn-Sampler (NUTS), Random-Walk Metropolis, and Gibbs sampling, and recommends using NUTS as the MCMC algorithm.

The Bayesian Calibration workflow is represented graphically in Figure 2.3.

Bayesian calibration provides a significant improvement in both output and input side error over a deterministic (optimization) approach [32, 65]. A calibrated Bayesian model with just as-built drawings and utility data can perform comparably to an expert-tuned model with more detailed submeter and equipment specification data [55]. The output of a probability distribution for parameters facilitates easy calculation of energy savings uncertainty for retrofit measures [46].

Bayesian calibration shows promise, yet so far most implementations have calibrated against only monthly utility data and investigated few (generally ≤ 4) parameters [46,54,56,60]. As discussed previously, since monthly data masks dynamic

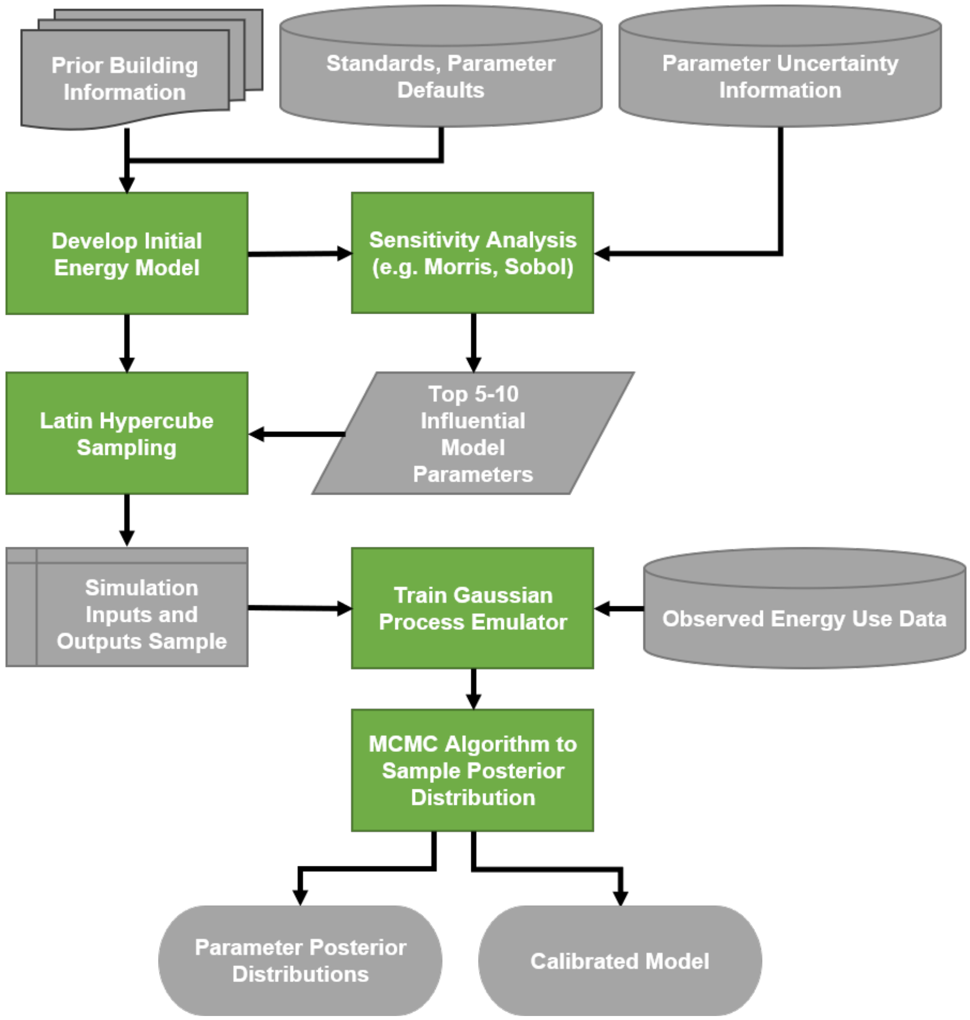

Figure 2.3: This graphic presents the typical Bayesian calibration workflow.

changes, it limits the ability to restrict posterior distribution for schedule-dependent parameters [32]. Chong [52] extended Bayesian calibration using hourly data by taking a subset of hourly data instead of a full year, and by using the NUTS algorithm to speed up the MCMC process. There exists ample opportunity to explore Bayesian calibration with greater numbers of parameters, hourly data, additional output comparison data, and data targeted to specific energy efficiency measures. However, Bayesian calibration is currently limited because of lack of knowledge of input parameters into the model.

2.3.3 Representative Seed Models for Calibration

While Bayesian calibration handles uncertainty nicely, it requires knowing which parameters a particular model is most sensitive to and reasonable prior distributions for those parameters. Without good assumptions for input parameters, Bayesian calibration performs poorly [52]. There exists a need to identify important model parameters and establish uncertainty bounds [58]. Representative models can be derived from bottom-up models of a building stock that cover a wide range of building types and characteristics. So for, such stock models focus on the residential sector [66], [67] or have limited geography and a limited set of input parameters [68]. Recent work covered in this dissertation develops ComStock [69], a model of the U.S. commercial building stock. ComStock is, to the authors knowledge, the first attempt at a bottom-up model of the U.S. commercial building stock.

2.3.4 Conclusion of Calibration

This section presented the current approaches used in building energy model calibration and explained the problems in many current approaches that rely on expert modelers or optimization methods. Bayesian calibration was introduced as a way of retaining uncertainty through the energy model calibration process to better predict and give an uncertainty range for retrofit savings. Bayesian calibration requires knowledge of input parameters and their distributions, which is a significant limitiation to development of current Bayesian calibration research.

Table 2.5: Summaries of selected calibration literature. More can be found in the literature reviews by Reddy [28], Coakley [18], and Fumo [19]. Parameters included if few considered.

Paper	Building Type	Location	Method	Metered Data	Calibration Parameters
Bertagnolio [70]	office	Belgium	manual refinement from sensitivity analysis	monthly, hourly electricity and gas	45 parameters (including schedule)
Chaudhary [36]	office (synthetic)	unspecified	*Autotune* genetic optimization	monthly electricity	35 singular parameters
Chong 1 [62]	mixed use (cooling only)	Singapore	Bayesian calibration	hourly chiller electricity	2 parameters - chiller capacity, chiller rated COP
Chong 2 [62]	office (cooling system only)	Pennsylvania	Bayesian calibration	hourly cooling electricity	5 parameters - (2x) chiller capacity, (2x) chiller rated COP, cooling tower capacity
Cipriano [71]	office	Spain	sensitivity analysis and LHS	hourly indoor temperatures	53 singular parameters
Hale [37]	office	Florida	Manual calibration	monthly electricity	8 parameters - ground temperature, cooling setpoint, lighting and equipment schedules, equipment power density, fan size, fan static pressure, supply air temperature
Harmer [30]	mixed use	Boulder, Colorado	sensitivity analysis and LHS	hourly electricity and steam	20 singular parameters
Heo [56]	office	Cambridge, UK	Bayesian calibration	monthly gas	19 parameters screened - 4 for BC window opening estimate, window discharge coefficient, indoor heating thermostat setpoint, infiltration rate
Heo [55]	office (synthetic)	Chicago, Illinois	Bayesian calibration	monthly total energy use	61 parameters screened - 4 for BC outdoor air flow rate, infiltration rate, heating system efficiency, equipment power density
Kim [72]	office (synthetic)	unspecified	Bayesian calibration	monthly heating and cooling	18 parameters screened
Kim [32]	office	South Korea	Bayesian calibration	10 min heat extraction of air handler and daily gas	107 parameters screened - 4 for BC window U-value, window SHGC, infiltration rate, chiller COP
Li [48]	office/lab	Atlanta, Georgia	Bayesian calibration with meta-model	daily chilled water and peak demand	6 parameters - cooling setpoint, (2x) occupancy density, (3x) outdoor air flow rates
Macumber [44]	office	Florida	genetic optimization	monthly electricity	5 parameters - ground temperature, cooling setpoint, lighting power density, equipment power density, fan static pressure
Muehleisen [60]	gatehouse	Golden, Colorado	Bayesian calibration	monthly electricity and gas	4 parameters - infiltration, equipment power density, lighting power density, occupant density
O'Neill [73]	office	Great Lakes, Illinois	sensitivity analysis and optimization	monthly and hourly total and plug electricity	10 parameters
Raftery [31]	office	Kildare, Ireland	manual calibration	monthly and hourly electricity	unspecified
Tian [61]	Retail	Tianjin, China	sensitivity analysis and Bayesian calibration	monthly electricity and gas	4 parameters - window u-value, window SHGC, occupant density, equipment power density
Westphal [74]	office	Santa Catarina, Brazil	sensitivity analysis and manual calibration	monthly and hourly electricity	unspecified

Table 2.5 summarizes a subset of the recent literature on using automated methods for building energy model calibration, including sensitivity analysis and Bayesian calibration methods. Areas that are under-explored in the literature are Bayesian calibration with hourly data (and calibrating with hourly data generally), including what submeter data would be helpful in model calibration and retrofit savings estimation [36, 75]. Collecting extensive submeter data entails significant expense and expertise, even in buildings with modern BMS systems, and is impractical for a many buildings with limited staff and budget. It is not clear how valuable this data is to determining energy savings from energy efficiency measures, that is, specifically, it is not clear whether the expense in measurement and model calibration will often exceed or stay below the corresponding value of uncertainty reduction the measurement provides. Raftery [27] suggests a hierarchy of model data, showing the importance of measured, logged data and how measured energy end uses (and therefore savings estimates from related measures) can change dramatically as calibration improves. This implies that some measured submeter data is vital to reducing uncertainty in input-side error and making savings estimates. Bertagnolio [70] similarly finds the need to collect meter data to reduce input side error. As energy savings uncertainty is a significant limitation to the uptake of energy retrofits, there is a need to prioritize building auditing, submetering, and sensor collection in a way that will improve model calibration and reduce energy savings uncertainty. Lower savings uncertainty will hopefully lead to more actionable energy savings projects and less risky financial investment. Building stocks models calibrated with end use data can help prioritize model input parameters for specific buildings and develop seed models for calibration.

Chapter 3: Research Hypothesis and Objectives

The objective of this research is to improve the estimation of savings from energy efficiency measures by determining parameter importance in stock energy models and establishing a foundation for calibration. This chapter states the research hypothesis, given in Table 3.1 and the research objectives, given in Table 3.2, that describe the actions taken to achieve these objectives.

Table 3.1: Research Hypothesis

Research Hypothesis	Reliable energy retrofit decisions are possible with new energy modeling approaches based on stock energy data from actual buildings.

Table 3.2: Research Objectives

1. Develop a methodology to consider interaction effects of energy efficiency measures, accounting for the influence of (1) capital cost constraints, (2) uncertainty associated with measure costs, and (3) future energy and carbon tax escalation on the retrofit decision making.
2. Create a process to develop initial building energy models that compromise stock models for sensitivity analysis and calibration.
3. Establish calibration metrics and identify important model parameters for stock model calibration.
4. Demonstrate and report stock model calibration metrics for a region.

Chapter 4: Energy Retrofit Methodology

The first section focuses on establishing a methodology to consider energy savings of energy retrofits under exogenous cost factors. The second section presents a brief overview of software to generate reduced-order models; simplified building energy models to begin the calibration process. The last section details the sensitivity and calibration analysis process.

4.1 Energy Savings Under Uncertainty

This section details a methodology established for evaluating the impacts of (1) capital cost constraints, (2) uncertainty associated with measure costs, and (3) future energy and carbon tax escalation on the retrofit decision making. The methodology was demonstrated for a case study of an actual building with sub-metered energy data including interval data for different end-uses deployed to calibrate a baseline building energy model. The calibrated model enabled considerations of different retrofit scenarios to include intrinsic and extrinsic uncertainties associated with the decision making for a building retrofit. Furthermore, this study also developed software for interoperability with OpenStudio [26] based on R scripts [76], allowing deployment of the methodology to retrofit decision making for other buildings. A case study for an office building in Philadelphia, PA, demonstrated the significance of the difference in capital availability on the optimal choice of retrofit measures.

4.1.1 Retrofit Path Methodology

Energy retrofit measure selection is dependent on capital availability, financial criteria, and uncertainties in energy savings and energy costs. Including measure interactions and savings uncertainties is necessary to properly account for a measure's impact on overall building performance. This measure integration and packaging increases the number of options to consider, and requires energy simulations to handle the complexity of measure interactions. Installing measures longitudinally based on a fixed capital budget adds further complexity, as the order in which measures are installed becomes significant. Load reduction measures allow equipment downsizing, and there is a performance difference for differently-sized systems with the same energy efficiency measures. This consideration greatly increases the number of energy simulations. Figure 4.1 shows the analysis process to generate all possible retrofit path-options, including the downsizing difference, under capital constraints to calculate their impact on the optimal retrofit measure package. As indicated in the figure, the established methodology comprises five steps including (1) Develop a calibrated energy model, (2) Select Energy Efficiency Measures (EEMs), (3) Generate unique simulations for measure combinations, (4) Run building energy simulations, (5) Analyze retrofit path options for different

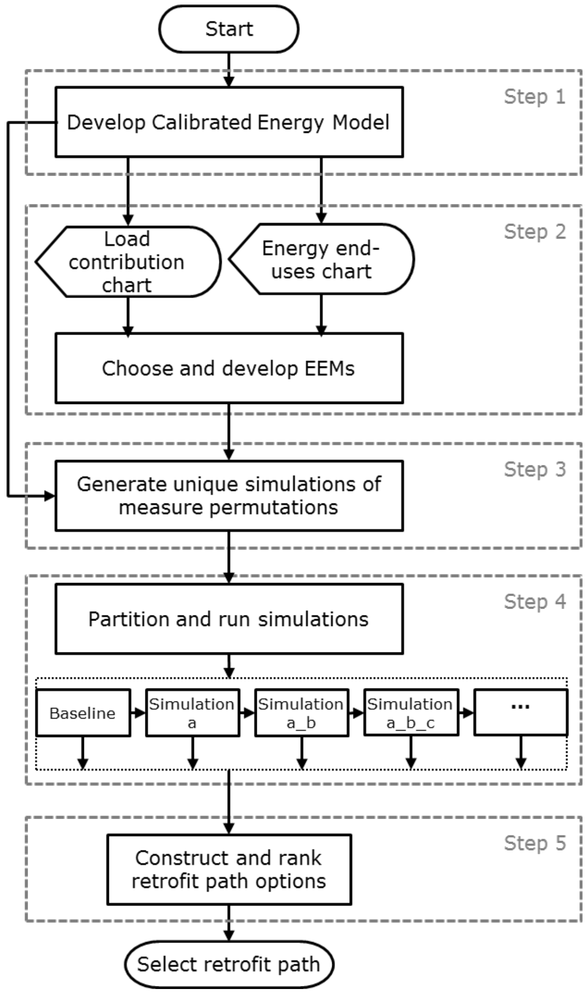

Figure 4.1: The flowchart of established methodology for retrofit decision making [49].

financial scenarios, and identify optimal options. The code to demonstrate this methodology for the case study presented in this paper is available on Github at (`https://github.com/CITY-at-UMD/retrofitLCC`).

4.1.1.1 (Step 1) Develop a calibrated energy model

The first step in the evaluation of different EEMs using building energy simulation tools requires developing a calibrated baseline building energy model. The calibrated baseline energy model needs to meet the requirements of the American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) Guideline 14 2002 [17]. Most calibrations of building energy models use monthly electricity and gas consumption from the utility bills due to their ubiquitous availability [77]. However, if sub-metered interval data for a building are available, a more accurate calibration method uses the sub-metered interval data to calibrate the building energy model with the 15 minutes sub-metered building energy data [41]. This study uses 15 minutes sub-metered energy end-uses interval data for the calibration of the baseline building energy model.

4.1.1.2 (Step 2) Select energy efficiency measures

The selection of EEMs depends on the building principal functionality, age, size, and financial constraints. Use of the building energy simulations allows identification of energy end-use breakdown, key load contributions, and measures that will most likely reduce peak building loads. These measures may not be financially justifiable on their own, but may be acceptable after including the savings associated with equipment replacement. They include measures for building enclosure, solar control, plug load / lighting control, and HVAC equipment control or replacements. Evaluating multiple measures can be time-consuming if it involves manually creating an energy model for each measure and combination of measures. The OpenStudio Parametric Analysis Tool (PAT) [37] implements measures from the Building Component Library (BCL) [78], and partially automates this process. This study develops new building component library measures and implements them with scripts using the OpenStudio API [26,79], an open-source object-oriented software for manipulating models in EnergyPlus [25].

4.1.1.3 (Step 3) Generate unique simulations for measure permutations

Energy efficiency measures can be combined together and installed in different orders to create unique retrofit paths. To distinguish between retrofit paths, each measure is given a letter “a”, “b”, “c”, etc. A string of letters identifies a unique retrofit path. Particularly, this study looks for the benefit that comes from installing measures that reduce building load prior to replacing the central heating or cooling equipment, which may be downsized depending on the new loads. These HVAC measures are dependent on other measures, whereas building lighting or occupancy sensors are independent of other measures. For example, if HVAC equipment measure “c” is dependent on measures “a” and “b”, which are independent, then the measure combination a_b_f will be identical to b_a_f, but not a_f_b. To simulate sequence a_f_b, the process is to (1) simulate the model and auto-size the equipment capacity for a_f, (2) read the HVAC equipment capacity for “f” from the output and hard-size that value in the energy model, and then (3) implement measure “b” and run the energy simulation to get the result for a_f_b. Without this process, if the model is auto-sized and several measures are implemented that reduce load, the energy simulation may run the simulation with a lower equipment capacity than what is present in the building, introducing a small error for the predicted performance.

Once all permutations are generated, a script removes redundant simulations, e.g. a_b_f and b_a_f. Then, another script takes the base model, adds each measure to it in succession, and saves it as a run script for the energy simulation software.

4.1.1.4 (Step 4) Run building energy simulations

This study uses OpenStudio and R scripts to implement and run the energy simulations automatically for the selected measures. The developed scripts distinguish between path dependent and path independent simulations to facilitate the process of running multiple simulations in parallel. The 2566 unique energy simulations in this case study are partitioned by the first measure for seven virtual machines on a central server, reducing simulation time by about three-quarters compared to running on a single machine. The energy simulation outputs are collected into a data file on the virtual machines, and then combined together to create a database of all simulation results.

4.1.1.5 (Step 5) Analyze retrofit path options for different financial scenarios

The life-cycle cost analysis is deployed to each unique energy simulation combined together in sequence to create a retrofit path for a given measure order and financial scenario. This study considers the impact of three financial variables: (1) Measure cost, (2) Future greenhouse gas price, and (3) Capital availability. It is possible to include many more financial variables in the analysis, as there is no need to rerun the energy simulations and the computational cost of evaluating other financial scenarios is much cheaper than generating unique energy simulations. However, any measure performance variation or calibration sensitivity,would require additional measures or choosing a subset of the energy simulations to re-simulate based on most promising retrofit implementation paths. Therefore, it is important to carefully choose the measures and calibrate the model as specified in Steps 1 and 2. The different measure combinations and financial scenarios in this case study generate a half million retrofit path options. These are filtered by financial scenario to generate a ranked list of optimal-path options for a given scenario.

4.1.2 Case Study

This methodology is demonstrated with a case study. The case study is a commercial office building at the Philadelphia Navy Yard, shown in Figure 4.2. The building was originally built as a barracks, and underwent major renovation in 1999

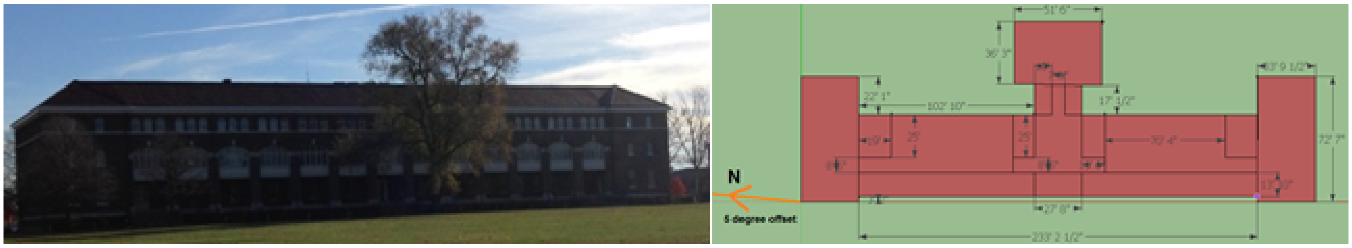

Figure 4.2: Office building at the Philadelphia Navy Yard: (a) Front side of building and (b) Floor plan in the energy model [49].

to become an office building. The building is approximately 75,000 ft^2 (6,968 m^2), of which approximately 60,000 ft^2 (5,574 m^2) is conditioned space, and approximately 40,000 ft^2 (3,716 m^2) of that is office space spread over 3 stories and a conditioned basement. The Energy Utilization Index (EUI) in this study is referenced to the conditioned floor area only. The building exterior wall is 1.5ft (0.46m) thick brick and has a window-to-wall ratio of 17%. Three VAV units with DX cooling serve the building. A gas boiler serves heating coils at each AHU and provides reheat for terminal boxes in each thermal zone. A gas hot water heaters provide service hot water.

4.1.2.1 Energy Model Calibration

The building energy model uses OpenStudio [26], an interface-type of middleware for EnergyPlus [25]. The model simplifications include an assumption of an identical floor plan on each story, which is nearly the case in the building. The fenestration is modeled with a set window-to-wall ratio on each exterior wall, rather than modeling each window individually, to improve simulation speed with little accuracy loss for load calculations [80]. Mechanical equipment and lighting specifications are detailed from design drawings. Plug loads are modeled with a set area density for office, conference, and lobby areas from sub-meter data, and equipment schedules were then adjusted to match the measured hourly plug load profiles [41, 81]. Air infiltration is assumed to be a uniform, constant 0.2 ACHnat across the exterior enclosure [82]. The model uses Actual Meteorological Year (AMY) weather data from the Philadelphia International Airport located a few miles from the building site. A detailed summary of building instrumentation and calibration are available in the literature [83, 84].

The model is calibrated to 10-month hourly sub-metered energy data for heating, cooling, service hot water, fan, lighting, total building electric, and total building gas energy use. Plug load and miscellaneous electric use, including water systems pumps and elevators use, is assumed equal to the total building electrical energy use less all other metered electrical loads – cooling, fans, and lighting. January 2012 data are not available, as sub-meter data was not installed until late January. Furthermore, HVAC sub-meter data in December 2012 are not comparable, as the building underwent a major controls upgrade. The lack of data for these periods increases the uncertainty in heating energy use for model calibration, as nearly a fourth of annual heating degree days occurred in January. The calibration disregards anomalous service hot water use data in April and May, when water use spiked, coinciding with a construction period on the second floor. Table 4.1 shows coefficient of variation of the root square mean error (CVRSME) and normalized mean bias error (NMBE) calibration statistics for each end use following ASHRAE Guideline

14 [17]. The final model calibration adjusted solely building temperature setpoints to 71.5±1.5°F (21.9±0.6°C) for heating and 72.5±1.5°F (22.5±0.8°C) for cooling, to match observed variation in the deadband range for VAV temperature control. This calibration method based on end-uses was possible with the availability of sub-meter data. Typically, most of the existing case studies calibrate energy models with monthly energy use by fuel type due to the lack of sub-meter data.

Table 4.1: Building 101 Energy End Use Calibration Metrics [49]

	CVRMSE	NMBE	Months
Calibration Target	15%	5%	All
All Electricity	6.1%	-0.6%	All
Plug Loads	5.9%	-2.1%	All
Lighting	4.9%	2.6%	All
Fans	9.3%	-1.2%	Omit December
Cooling	12.6%	3.6%	Omit December
All Gas	12.0%	2.3%	Omit December
Heating	12.6%	1.7%	Omit December
Service Hot Water	9.5%	2.3%	Omit April and May

4.1.2.2 Select Energy Efficiency Measures

The baseline building energy end-uses and the component contribution to peak heating and cooling load are helpful indicators to determine promising EEMs. Figure 4.3 shows the contribution of each energy end use to total building energy use. Heating energy use, and then internal equipment and lighting energy use, dominate the energy use of the building. Measures targeting these systems can save a

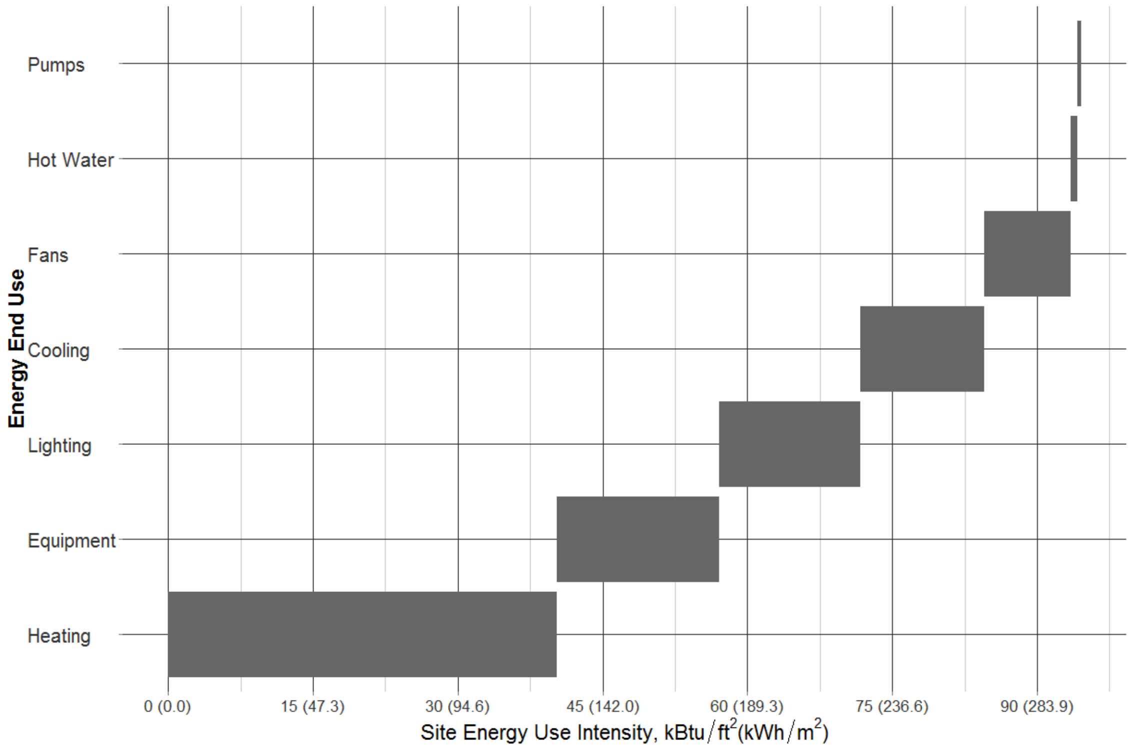

Figure 4.3: Breakout of building energy by end use, which helps to target EEMs that can reduce the largest energy end uses in the building. EEMs targeting heating, lighting, and equipment energy use will save the most energy.

substantial amount of energy.

Figure 4.4 shows that infiltration and conduction through exterior walls and windows are the main contributors to peak heating loads, and are offset partially by lighting and internal equipment. Solar heat gain, interior lighting, and equipment are the main contributors to the peak cooling load. EEMs that reduce these peak load contributions enable HVAC equipment downsizing upon replacement.

This study considers seven energy efficiency measures, shown in Table 4.2. Several of these measures, including measure "a", "b", and "c", were commonly

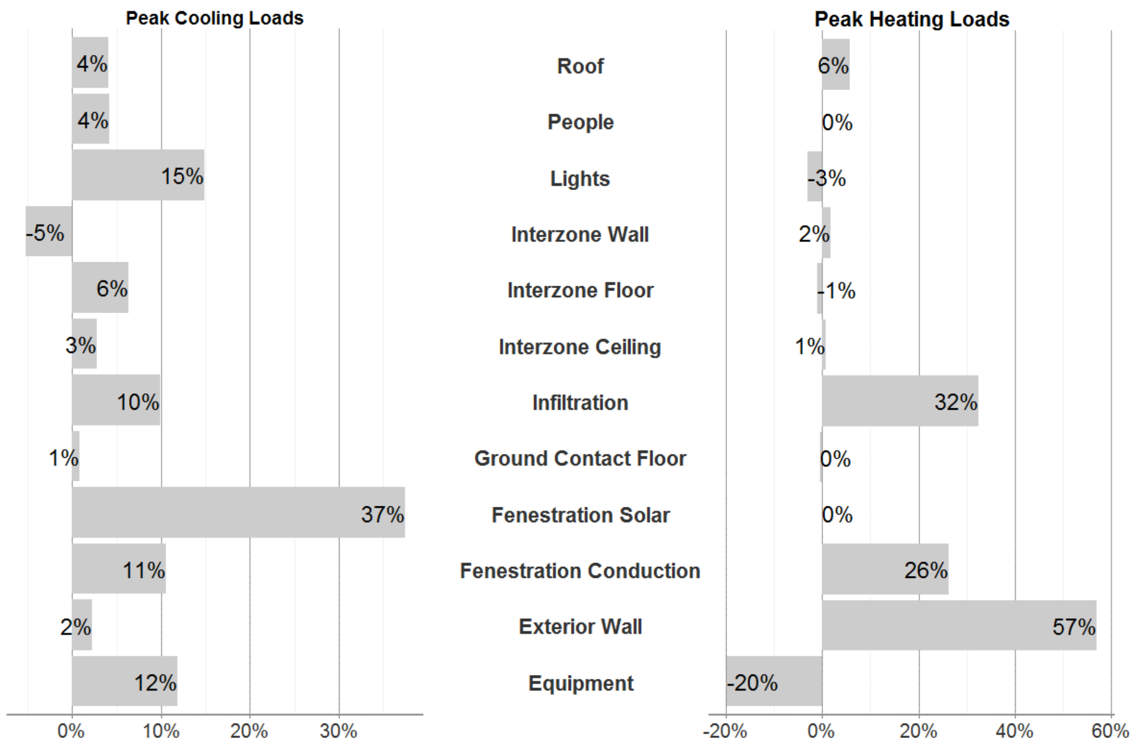

Figure 4.4: Percent contributions of component loads to thermal zones' peak heating and cooling. EEMs that reduce fenestration solar gain will greatly reduce required cooling equipment capacity, and EEMs that lower infiltration and reduce conduction losses through windows and walls will reduce required heating equipment capacity.

Table 4.2: Energy Efficiency Measure Descriptions [49]

Letter	Energy Efficiency Measure	Description	Source
a	Wall Insulation	Add an exterior insulation and finish system, with 4 in (0.1 m) EPS board, R-16, reduce infiltration by 30%	[85]
b	Light Power Density Reduction	Reduce conference and office lighting power density from 1.15 to 0.9 W/ft^2 (12.38 to 9.69 W/m^2)	[86]
c	Occupancy Sensors	Reduce lighting fraction from 0.2 to 0.05 during unoccupied hours on weekdays, and 0.15 to 0.05 on weekends	[86]
d	Infiltration Reduction	Reduce outdoor air infiltration by 15%	Engineering Judgment
e	Window Film	Reduce Solar Heat Gain Coefficient (SHGC) from 0.764 to 0.38	[8]
f	Condensing Boiler	Replace boiler with a 90% efficient condensing boiler, autosize capacity and flow rates for loop, lower supply temperature to 140°F (60°C)	[87]
g	Condensing Unit	Replace condensing units with auto-sized unit with high speed Energy Efficiency Ratio (EER) 11.5 and low speed EER 16.2	[88]

Table 4.3: Energy efficiency measure costs for measures in this study [49].

Energy Efficiency Measure	Cost / Unit	Cost	Yr-1 Savings	Simple Payback (yrs)	Source
Wall Insulation	$4.78/ft^2$ ($51.45/m^2$) wall area	$927,930	$6,301	147	RSMeans,4in.(0.1m) EPS insulation, Commercial renovation Exterior Insulation and Finish System, 25% mark-up for Multiple Stories
Light Power Density Reduction	$4.78/ft^2$ ($51.45/m^2$)	$202,886	$6,323	32	RSMeans,Fluorescent high-bay 4 lamp fixture, 1W/ft^2 (10.76 W/m^2), 59FC, 4 fixtures per 1000 ft^2 (92.9 m^2)
Occupancy Sensors	$1.06/ft^2$ ($11.41/m^2$)	$44,991	$2,384	19	RSMeans,5 fixtures per 1000 ft^2 (92.9 m^2), including occupancy and time switching
Infiltration Reduction	$150,000	$150,000	$1,749	86	Engineering judgment
Window Film	$18.93/ft^2$ ($203.76/m^2$) glazing	$182,311	$4,259	43	RSMeans,Solar Films on Glass average of min/max value
Condensing Boiler	$20,706 + $13.82/MBH ($20,706 + $4.05/kW)	$42,176	$3,960	11	RSMeans,commercial gas boilers
Condensing Unit Replacement	$7,909 + $766/ton ($7,909 + $2693.91/kW)	$116,631	$4,864	24	RSMeans,packaged air-cooled refrigerant compressor and condenser

recommended by energy audits for the cases study building [89], and other measures were included to reduce peak heating and cooling loads. Cost assumptions come from RSMeans [90] using standard union rates in Philadelphia and are summarized in Table 4.3. Energy efficiency measures were implemented as BCL measure scripts modifying the OpenStudio model of the building.

4.1.2.3 Scenario Parameters

Life-cycle costs for different retrofit scenarios are compared to life-cycle cost for a baseline scenario. The baseline scenario assumes that the expected lifespan of the outdoor air-cooled condensing units is 20 years, meaning a replacement in 5 years, and that the expected lifespan of the boiler is 25 years, meaning a replacement in 10 years [91]. The resulting energy costs, capital costs, and greenhouse gas emissions costs over the building lifetime are combined into a cash flow that is then discounted to calculate the Net Present Value (NPV) of the scenario. The scenarios assume a lifetime of 20 years with a real discount rate of 3%. Natural gas and electricity prices are adjusted according to the NIST energy price escalation rates for census region 1 [92]. In addition, four greenhouse gas emissions prices are considered: no cost for emissions, and the default, low, and high greenhouse gas price scenarios from NIST. Lastly, measure costs for each measure are considered at full price and half-price, reflecting the possibility of measure cost reductions and significant additional efficiency incentives that are not accounted for in the greenhouse gas price. Each scenario for ordering retrofit measures is considered under five capital availability scenarios: \$1.00/ft^2 (\$10.76/m^2), \$2.00/ft^2 (\$21.53/m^2), \$3.00/ft^2 (\$32.29/m^2), \$5.00/ft^2 (\$53.82/m^2) and \$100.00/ft^2 (\$1076.39/m^2), reflecting different annual capital allotments available to fund energy efficiency measures. The \$100.00/ft^2 (\$1076.39/m^2) scenario is an intentionally high value that practically imposes no financial limitation to implementing measures, meaning all measures for a given retrofit scenario are able to be implemented in the first year. In the other scenarios, the capital limitation causes a delay in implementation for an energy efficiency measures.

4.1.3 Results

The NPV of the baseline case where the equipment is replaced ranges from -\$35.58/ft^2 (-\$382.98/m^2) to -\$42.63/ft^2 (-\$458.87/m^2), depending on the greenhouse gas price scenario and the cost modifier for the measures. This includes the cost of replacing the central HVAC equipment and the energy costs over the project lifetime. The equipment will need to be replaced at the end of its service life, so the relative financial performance of a retrofit path is measured in reference to this baseline with equipment replacements and no other measures. For example, in the default NIST GHG price scenario, and measures at full costs meaning a cost modifier of 1.0, the net present value of the baseline case is -\$38.50/ft^2 (-\$414.41/m^2). If a retrofit path were to have a net present value of -\$39.50/ft^2 (-\$425.17/m^2), it would mean that it is \$1.00/ft^2 more expensive than the baseline case.

Figure 4.5 shows the comprehensive range of all possible retrofit paths relative to this baseline case for different financial scenarios, including the cost modifier, greenhouse gas price, and capital availability. Retrofit paths are shown based on the average annual site energy use of the building over the 20 year financial lifetime, and the net present value compared to the baseline case for the same financial scenario. The majority of the retrofit paths show negative net present value relative to baseline, with the greatest differences between retrofit paths determined by whether they include the wall insulation measure, which greatly reduces the net present value of retrofit paths that include it. Another major source of variation is the cost modifier that adjusts the measure costs. Figure 4.5(a) shows the retrofit paths for measure costs at full price (cost modifier of 1.0, gray) and measure costs at half price (cost modifier of 0.5, black). Each cost modifier scenario shows two distinct clusters of retrofit measures; the retrofit paths that include the wall insulation measure comprise the cluster with lower net present value. The box (b) in Figure 4.5(a) is the domain in Figure 4.5(b), expanded to show the influence of the greenhouse gas price scenario. For most retrofit paths, there is a difference of less than \$0.75/ft^2 (\$8.07/m^2) between the high and no greenhouse gas price scenarios for most retrofit paths. Within a given greenhouse gas price scenario, there is a further difference in retrofit path financial performance depending on the capital availability. In general, the capital availability yields a more significant difference than does the choice of greenhouse gas price scenario.

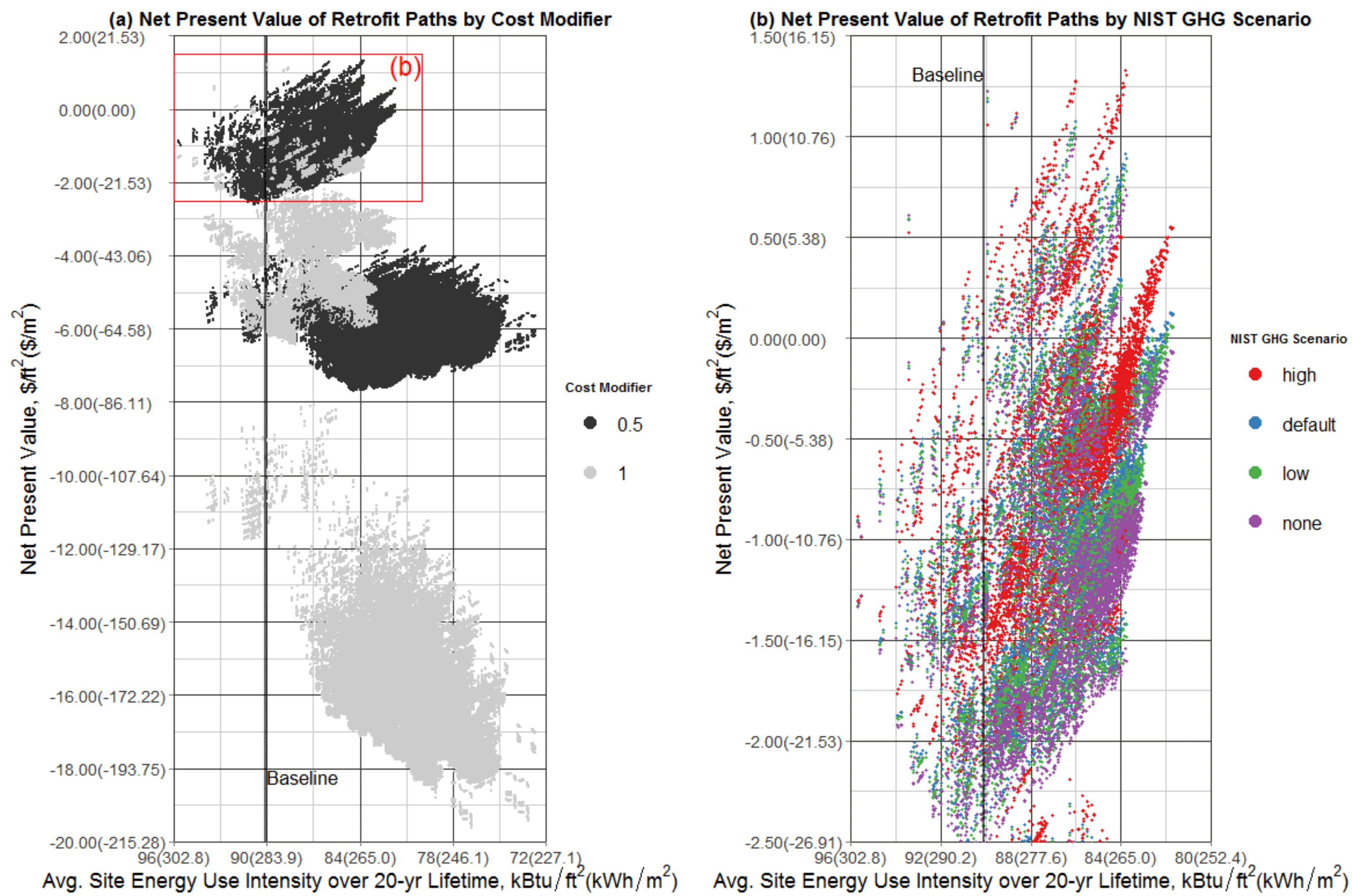

Figure 4.5: (a) Net present value of retrofit paths relative to the net present value of the baseline case. The retrofit paths are colored by whether they are part of the financial scenario when measure costs are at full price (cost modifier of 1.0, black), or at half price (cost modifier of 0.5, gray). (b) Net present value of retrofit paths relative to the net present value of the baseline case. The retrofit paths are colored by their NIST GHG price scenario [49].

Figure 4.6 shows the optimal paths relative to the baseline case for a range of financial scenarios. Optimal paths are those that have the most positive net present value relative to the baseline case, and include equipment replacements before the end of the service life of the equipment. The optimal paths include a combination of the measures reducing lighting intensity (measure "b"), occupancy sensors (measure "c"), replacing the boiler (measure "f"), and replacing condensing units (measure "g"). In the scenario with measure costs at full price, only permutations of path g_f_c, implementing the equipment replacements and then the occupancy sensors, show a positive net present value. Furthermore, when measure costs are at half price to model the hypothetical case where efficiency measures are much cheaper than they are at present, the g_f_c path remains the optimal option for all but the highest greenhouse gas price scenario. In the highest greenhouse gas price scenario, the optimal path includes reducing the lighting intensity before the other three measures. In Figure 4.6, each retrofit path is presented by a line, with the points showing a specific financial scenario. The shape of the point indicates the NIST GHG Scenario, and the color of the point indicates the capital availability. For each retrofit path, reducing the capital availability reduces the net present value of that option, and increase the average annual site energy use of the building over the 20 year financial lifetime. This makes intuitive sense: as the amount of available annual capital increases, measures are able to be implemented sooner, which means a longer time over which energy cost savings can accrue. In this case study, not having enough money to implement the optimal path at once greatly reduces the achievable financial benefits from that retrofit option.

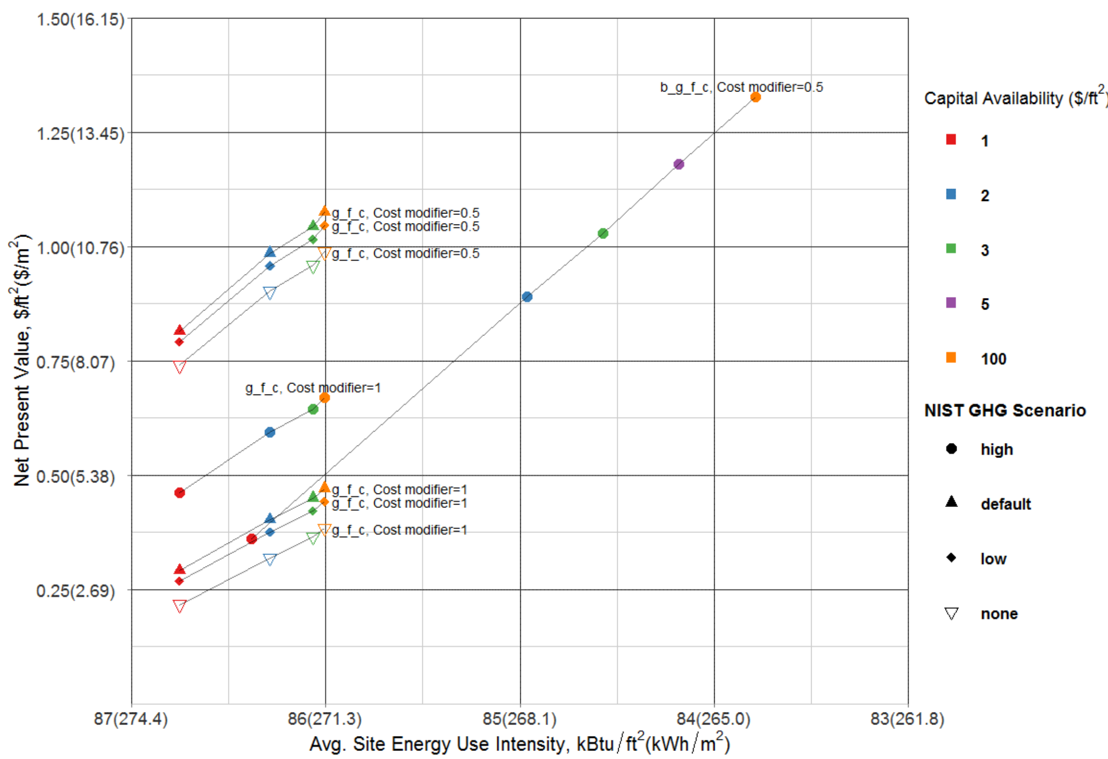

Figure 4.6: Optimal path options depending on the financial scenario [49].

Another consideration is how the financial performance depends on the order of measure instillation. Figure 4.7 shows the influence of changing the measure order for the b_g_f_c path for all financial scenarios. The maximum difference is the difference between the optimal ordering of measures and the worst ordering of measures within a given financial scenario. For this measure path, the largest difference between the optimal and worst ordering occurs in the lowest capital availability scenario. In this scenario, the size of the difference is comparable in magnitude to the net present value of the retrofit path. This means that when capital is not available to install measures, choosing which measures to install first can be as important as choosing which measures to install. The influence of measure order is less important in higher capital availability scenarios. In the case of no capital restrictions, the $100.00/ft^2 ($1076.39/m^2) scenario, the maximum difference in measure ordering is around $0.05/ft^2 ($0.54/m^2), which is much smaller than the variation in financial scenarios shown in Figure 4.5 and 4.6. This finding suggests that for this case study, the importance of installing measures with the optimal financial return is much more important than making sure they are ordered correctly to get the downsizing benefit, and that difference in financial performance under capital restriction is mostly explained by not implementing the measures with the optimal energy savings sooner. Overall, Figure 4.7 shows that for the optimal retrofit path, the difference in net present value between the optimal and the worst ordering of measures depends on the amount of capital available to implement energy efficiency measures. Furthermore, Figure 4.8 shows that for retrofit paths with a positive net present value relative to the baseline case, the maximum relative greenhouse gas

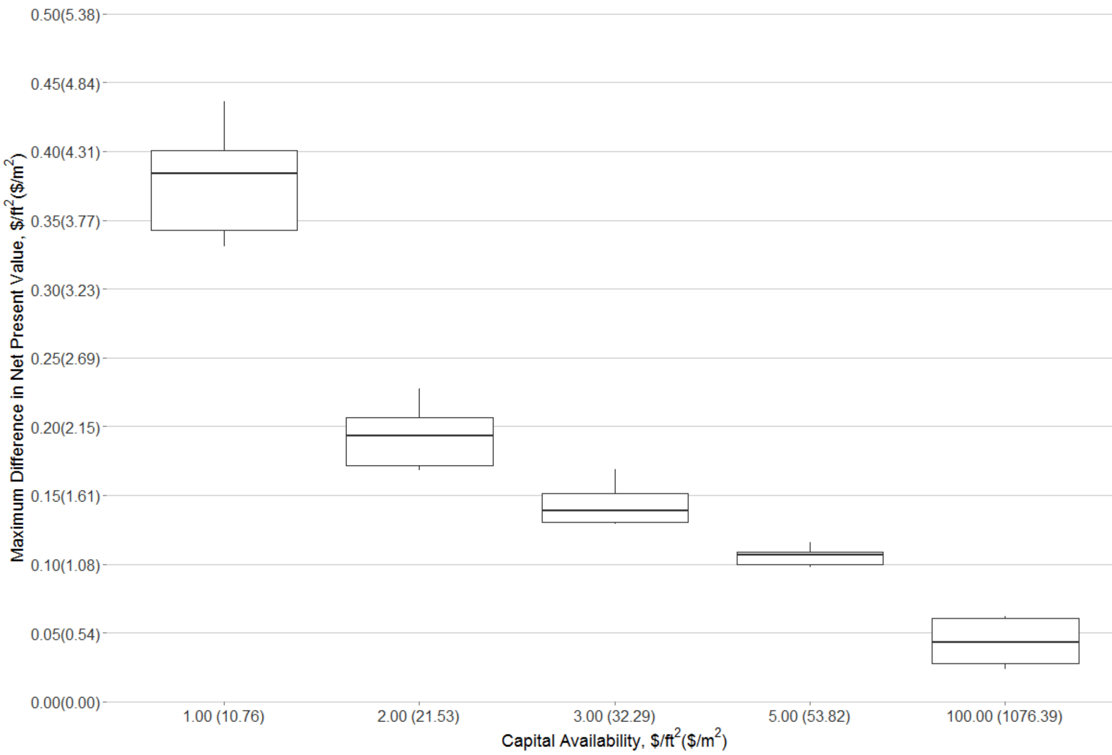

Figure 4.7: Net present values with changing capital availability resulting in different implementation order from the optimal retrofit path, worst retrofit path, which changes depending on the financial scenario [49].

emissions reductions possible over 20 years is 85% of the emissions of the baseline case. With measures at full price, fewer paths are available, and only a 5% emission reduction is possible.

4.1.4 Discussion

The aim of selecting the considered energy efficiency measures in this study is to reduce overall energy use and reduce peak demand served by the central heating and cooling equipment. However, none of the measures has a simple payback under the typical 7 year requirement of institutional investors, shown in Table 4.3.

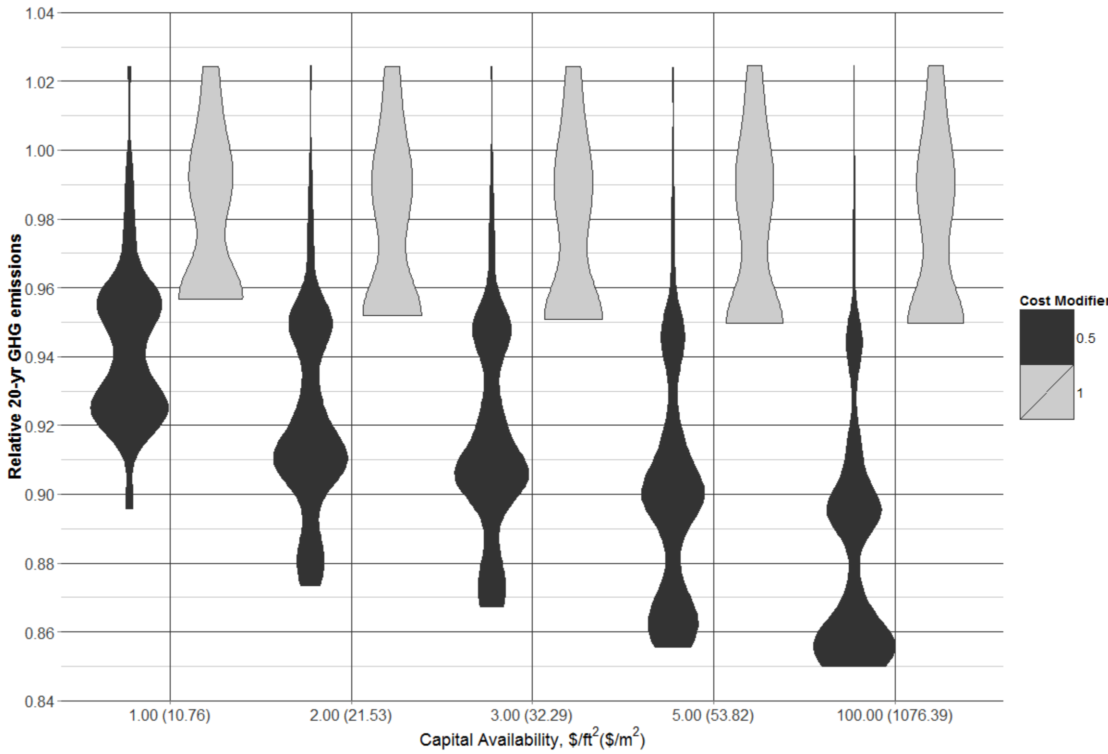

Figure 4.8: Violin plot showing the relative 20-year greenhouse gas emissions dependent on the capital availability and measure costs for retrofit paths with positive net present value relative to the net present value for the baseline case. The x-axis shows the capital availability per year available for retrofit, and the y-axis shows relative ghg emissions as a fraction of baseline emissions. Results from all measure packages with positive NPV are included in the violin density plots, and form clusters as more measures are able to be adopted when their costs are cheaper as shown by the different cost multipliers. Even at unlimited capital availability, there is only a 5% reduction in 20-yr ghg emissions when measures are at full price. Half cost measures allows more to be adopted, scaling to a 15% reduction in 20-yr ghg emissions with unlimited capital availability [49].

A life-cycle cost approach opens up further options, especially assuming an existing planned replacement cost of the HVAC systems. The load reduction benefits are minimal, and other factors dominate the financial performance of the retrofit options.

4.1.4.1 Minimal Impact of Load Reduction Benefits

While the selected measures reduce peak load considerably, this is reflected only in the replacement costs for the HVAC equipment and does not extend to significantly reduced energy use, and therefore cost, over the year. Furthermore, some measures have counteracting effects that negate the load reduction savings. For example, the window film measure reduces cooling loads, and thus annual electricity use by 8%, but this is offset by an increase in heating requirements for the building, raising annual gas use by 13%. The net result is a 1% increase in annual site energy use, but a 4% decrease in annual source energy use, a 3% decrease in annual energy cost, and 5% decrease in greenhouse gas emissions. The wall insulation measure behaves similarly. When wall insulation is added, the annual natural gas use reduces by 35% and the annual electricity use increases by 7%, for a net 11% energy savings, but is neutral for annual energy cost and greenhouse gas emissions. The reason for the electric use increase is that there was a modest cooling effect in the shoulder seasons for the building with lower insulation that offset the cooling requirements. This countering effect could be mitigated by reducing plug load consumption, or including natural ventilation or other free-cooling option. In addition, the natural gas rate per unit energy is cheaper than the electricity rate in this study, so electricity use is more significant for marginal energy cost savings.

For all measures, the improvement in financial performance from the demand reduction is of secondary importance to the aggregate savings of an energy efficiency measures, and often smaller than the marginal increase in heating that some of the cooling measures provided or vice versa. This does not negate demand reduction as a consideration for choosing energy efficiency measures, but suggests that this impact is only significant in cases where downsizing opens up further technology options to meet building loads instead of simply replacing equipment with a more efficient model, or if there are other significant demand response financial incentives that were not considered here.

4.1.4.2 Important Parameters for Financial Performance

For a given measure selection, the most significant determination in retrofit path financial performance comes from variations in measure cost. This matches a similar case study that considered uncertainty in measure performance and financial scenarios, which found measure cost and energy price to be the most significant determinants of measure package performance [14]. The case study presented here found less energy and greenhouse gas emissions savings opportunity, with only 10% emissions mitigated over the 20 year lifetime compared to the baseline that includes equipment replacement, and only 14% emissions mitigated compared to the baseline scenario that does not include equipment replacement. Part of the explanation for this difference is the case study present in this paper did not consider a micro Combined Heat and Power (CHP) system, which would offset the emissions from the relatively coal-intensive energy supply where the case study is located. The savings estimates in this case study are less than what was found in prior energy audits [89, 93], with discrepancies from the difference in weather assumptions and increased occupancy in the building after the audits. The later study considered more EEMs, and found a retrofit package including daylight dimming, upgraded lighting, and weatherization reduces site energy use by 23% and source energy use by 24%, with simple payback of 7.6 years assuming no incentives. The difference in recommendations is attributed to the different types and technical performance of EEMs considered, and the inclusion of a daylight dimming measure.

Uncertainty associated with measure costs and future measure costs are a significant part of financial performance uncertainty within this study and between studies. For example, newer LED technology could replace the fluorescent tube lighting common to most commercial buildings. This study assumed measures at full price and half price to represent potential reduced measure costs from cheaper technology or market scaling. As measure prices and financial scenarios are likely to change frequently, this study provides the developed code so others can test different financial scenarios and measures available on Github at (`https://github.com/CITY-at-UMD/retrofitLCC`).

Lowering the capital availability reduces the financial benefit from implementing measures, as the energy cost savings accrue over a smaller portion of the project lifetime, and later implementation is more heavily discounted. For a given retrofit path, the impact of the capital availability was as significant as the difference in greenhouse gas pricing scenarios, meaning that limited capital for energy retrofit projects imposed a similar barrier to achieving a given energy savings or emission reduction level. This is significant, because it suggests that funding retrofits through an annual budget or a revolving loan fund whereby the accrued energy cost savings is used to fund further retrofits may limit the extent of energy savings and emission reductions.

For a given selection of measures, measure order matters more with limited capital availability. However, the difference in load reduction is small compared to the difference in energy cost savings from implementing more cost-saving measures sooner. This effect is smaller than the uncertainty estimated for measure performance in the existing literature [14], suggesting that measure analysis should include measure uncertainty and interaction, but the impact of marginal load reduction is not as necessary for consideration. Furthermore, this study benefited from sub-meter calibration to help determine load reduction opportunity. This level of meter detail is rarely available in most buildings undergoing a retrofit, and imposes significant risk from the uncertainty of meeting buildings loads with demand reduction, given that building energy models can be prone to over-specification and mis-characterization of the source of buildings loads.

4.1.5 Conclusion

Building energy retrofit projects often use a single-measure ranking based on simple payback to analyze the financial performance of retrofit options. Such a ranking does not include the potential financial savings from load reduction measures that also reduce the cost of replacement heating and cooling equipment. This study considered the life-cycle cost of several Energy Efficiency Measures (EEMs) over an exhaustive list of measure combinations based on the building cooling and heating loads as well as financial scenarios with different capital availability. The selected EEMs are (1) wall insulation, (2) lighting power density reduction, (3) occupancy sensors, (4) infiltration reduction, (5) window film, (6) condensing boiler, (7) condensing unit replacement. The scenarios include consideration of NIST greenhouse gas pricing projections, full and half-price measure costs, and capital availability ranging from \$1.00/ft^2-yr, a minimum value for the capital constrains, to \$100.00/ft^2-yr, a representative of no capital constraint. In the most pessimistic scenario where measure costs are full-price, a capital constraint of \$1.00/ft^2-yr (\$10.76/m^2-yr), and no greenhouse gas emissions price, the net present value is \$0.22/ft^2 (\$2.37/m^2). For the most optimistic scenario, where the measure costs are half-price, no capital constraint, and the highest greenhouse gas emissions price, the net present value is \$1.33/ft^2 (\$14.32/m^2).

Measure costs were the most significant source of variation in financial performance, followed by the capital availability and greenhouse gas pricing scenario. The difference in measure ordering, and the importance of load reduction were relatively insignificant compared to the importance of the financial scenario, and are smaller than typical uncertainty measure performance and model calibration. Therefore, until model calibration of building loads and uncertainty of measure performance are more reliable, it is not useful to consider marginal capital cost savings on equipment from demand reduction or difference in measure installation order. Larger energy savings targets, for this building in excess of around 40% site energy savings or under 60 kBtu/ft^2-yr (189.3 kWh/m^2-yr) for this climate, will require more substantial building component and heating and cooling equipment changes than the marginal improvements considered in this case study, and will likely entail behavioral change and internal equipment load reductions as well. While low energy savings potential depends on the particular building and range of measures considered, this methodology suggests that the nonlinear dependence of energy and greenhouse gas savings potential on capital availability, and the relative lack of significance of the greenhouse gas price on financial performance implies that measures cost reduction and increasing capital availability are key concerns for emissions reductions. Therefore, increasing investment in energy retrofits is key to reducing greenhouse gas emissions.

Chapter 5: Reduced-Order Models and ComStock

This section details the process used to generate reduced-order models that are the starting point for calibration efforts.

5.1 Introduction to Reduced-Order Models

Reduced-order models are simplified building energy models that condense the number of zones, space types, and HVAC systems in a model while still retaining the underlying building energy model physics and HVAC system operation [94]. The intent of these models is to trade-off some model accuracy due to model misspecification with much simpler model creation [95]. The U.S. DOE prototype models of the national building stock are reduced-order models that are widely used to analyze retrofit savings of EEMs and develop energy policy [96], and such synthetic models are regularly used to create synthetic data for which to test against for comparative purposes [36, 55, 61, 72]. Reduced-order models need to be able to approximate building surfaces areas, exterior exposure, ventilation rates, and percentages of spaces types to be representative of a building. Heidarinejad [97] has shown the influence of building shape and extended building simulation software to generate several basic shapes that together can capture most building geometries.

Raftery [27] proposed the idea of zone typing that suggested thermal zones can be grouped without loss of model fidelity if they share a similar space type, exterior adjacency, and conditioning method. This means that for most buildings, core and perimeter models with care taken to match shape, space type percentages, and HVAC system layout can serve as an accurate representation of a building without the effort needed to model each HVAC thermal zone exactly.

5.1.1 Creating Reduced-Order Models from High Level Input Data

Generating a reduced-order model for a specific building requires an energy audit or knowledge of the buildings characteristics. Several characteristics are easy to determine, such as building use type, size, age, and HVAC system type. Others require a walk-through survey, such as lighting power density and construction type, and some require metering or analysis of BAS system data, such as supply temperatures and plug load schedules. Unfortunately the importance of building characteristics on performance does not correlate nicely with ease of determination. Plug load densities and part load efficiencies of HVAC equipment are difficult to determine but are highly important parameters [41]. For such parameters, it is prudent to use representative data from studies of similar buildings as a starting point for calibration, as the cost of data collection may be prohibitive.

The U.S. DOE prototype models [96] are a commonly used source for building parameter assumptions. These models are based on ASHRAE Standards, especially ASHRAE Standard 90.1 [86] for performance characteristics. The prototype models

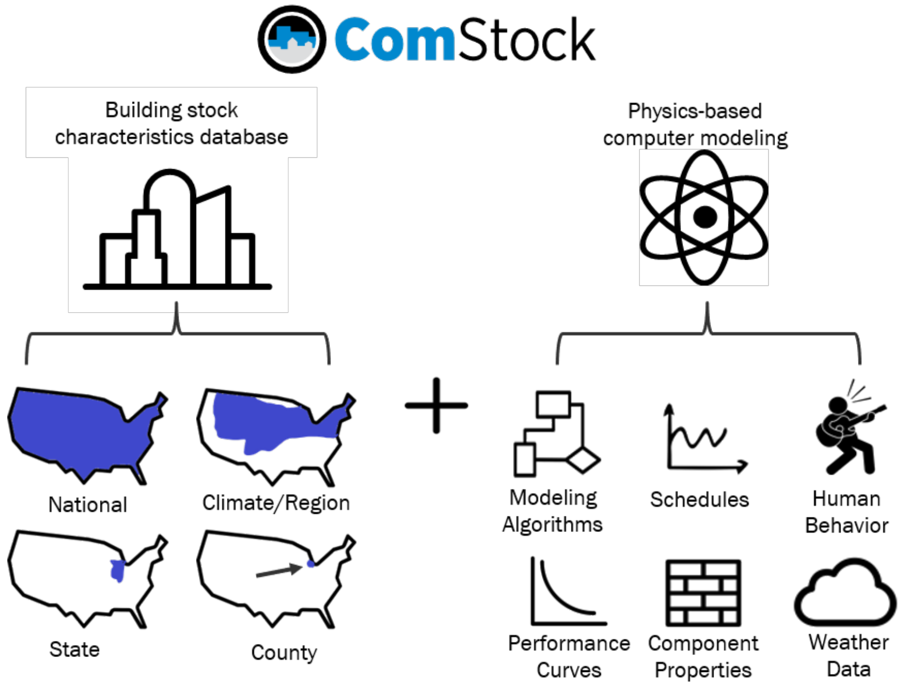

Figure 5.1: ComStock combines distributions and building characteristics of the U.S. building stock with building energy modeling.

are maintained by PNNL with performance characteristics and model generation methods available through openstudio-standards [26], a library for OpenStudio to generate the prototype building models and perform ASHRAE 90.1 baseline model generation for code compliance and LEED reporting.

5.2 ComStock

While containing useful building parameter assumptions, the DOE prototype models are limited in their scope, not representing the wide range of size, shape, HVAC systems, and other characteristics present in the building stock. For this reason, NREL developed stock modeling tools ResStock [67] and ComStock [69] for

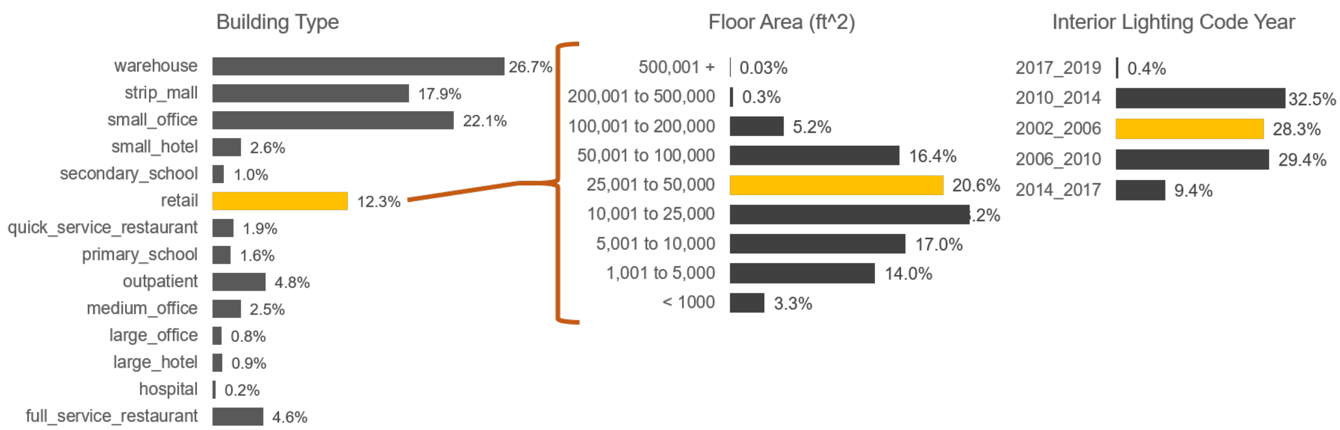

Figure 5.2: Example parameter probability distributions for ComStock. Floor area is dependent on building type, while year of lighting code replacement is independent based on a retrofit probability that increases with lighting system age. A large sampling from this distribution of dependent and independent parameters generates a representative model set of the building stock that can then be weighted to capture total energy use.

the residential and commercial building stock respectively to better represent the variability present in buildings, as shown in Figure 5.1. Instead of modeling one prototype building, many buildings are modeled with characteristics sampled from high level distribution data available from CBECS [98], CoStar™, and a Department of Homeland Security Database. Figure 5.2 gives an example distribution sampling of two parameters for a retail building. Building parameters can be independent or dependent on other parameters, for example building size being dependent on building type. Lighting power density is independent of other parameters and based on ASHRAE Standard 90.1 values matched to the code year when the lighting system was last renovated. A full list of input parameters in the macro sampling is given in Table 5.1. Input parameters then feed into the 'Create Typical Building

From Model' OpenStudio measure using methods from the openstudio-standards library [26]. A full ComStock run comprises 350,000 building energy models. These are run on Eagle, a supercomputer at NREL, taking between 5-30 minutes per simulation. The full timeseries enduse data from a ComStock run is over 600 GB and is uploaded to a Amazon Web Services S3 bucket and transfered to an Amazon Web Services Athena database for queries.

The actual count of buildings is aggregated from CoStar$^{\mathrm{TM}}$, a real estate company that maintains a database of all leasable commercial buildings in the U.S.. The buildings in CoStar$^{\mathrm{TM}}$ are classified by size and building type and used to get accurate counts and floor area of buildings in a given region. The aggregate of these building models creates a stock model and can be used to construct prior parameter distributions for a given building when only high level characteristics are known.

Geometry for ComStock uses an abstracted dual-bar method that mimics the building's exposure and surface area to volume ratio. The dual bar method uses two separated rectangles placed perpendicular to each other. Given an aspect ratio, perimeter ratio, and orientation, the width and height of the bars are adjusted to have the same the same surface exposure in each orientation as a similar building of a single shape. Figure 5.3 gives an example. Each building type is composed of a percentage of different space types, with default ratios from the DOE prototype models [96], and are allocated to core and perimeter zones. Some space types such as lobby areas or retail may be fixed to the first floor, and large single-height spaces such as gymnasiums can be separated out from the dual bars entirely and modeled as a separate rectangle.

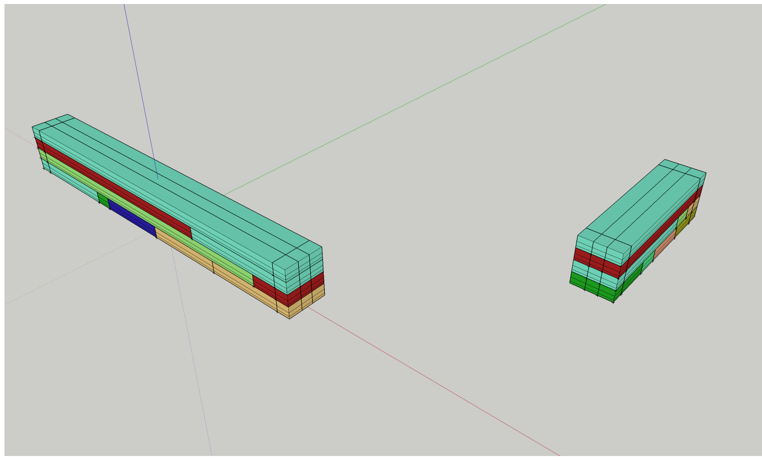

Figure 5.3: Example dual bar created for a building. This is a 4-story building with an aspect ratio of 2.0 and perimeter multiplier of 2.0. The different colors represent different space types within the building.

While some building characteristics are easily available in national survey data [98] such as building type, size, location, and HVAC systems, others require more detailed survey data such as lighting [99]. Some data, such as plug load schedules, are not readily available for all building types and need to be carefully inferred from calibration of ComStock to utility data.

Table 5.1: Macro level parameters used to specify a particular building energy model in ComStock.

Parameter	Description
Climate Zone, County ID, State ID	Location parameters used to determine the building location for weather file selection and aggregation to the county, state, and utility territory level
Building Type	(15) different building type categories representing diverse use cases
Year Built, Built Code	The year of construction used to determine which one of (8) ASHRAE 90.1 code sets to use for building properties and the starting year for retrofit frequency determination
HVAC System Type	(67) unique varieties of HVAC systems, including RTUs, centralized VAV systems, zonal systems, and small residential systems
Floor Area	The floor area of the building, binned into (10) different size bins
Number of Stories	Number of above-ground floors, with buildings between 15-25 stories and over 25 stories grouped as separate bins
Building Shape	(11) unique building shapes
Aspect Ratio	The building aspect ratio, the width versus length, from 1 to 6
Rotation	Degrees of rotation from North (0 degrees)
Service Water Heating Fuel	Gas or Electric
Weekday/Weekend Start Times and Duration	The occupancy start time and duration for both weekdays and weekends, used for creating schedule variability
Building System Code Year	Separate code year to use for a given building systems (envelope, HVAC, SWH, interior and exterior lighting, interior equipment) depending on a retrofit frequency for that system

Chapter 6: Parameter Importance for Calibration

6.1 Parameter Importance Analysis for Commercial Building Stock Modeling

Parameter importance analysis is used to determine highly influential parameters to adjust during calibration. This is especially important in commercial building models where hundreds of variables across different building types determine how the building uses energy. Parameter importance is closely linked with sensitivity analysis. Most sensitivity analysis methods generate a sample space and then evaluate a function against that sample space. In our situation, the function (a building energy model) is computationally expensive. Even a small set of 30 parameters with 2 values each means 2^{30} simulations for a full factorial analysis for just one building type, which is computationally prohibitive. Methods such as Latin hypercube sampling greatly reduce the number of function evaluations, sampling the parameter space uniformly while accounting for variation. The 350,000 simulations in ComStock are a sample of the millions commercial buildings that exist. Given the computational expense, this study determines feature importance by developing a regression model based on the ComStock run outputs and determining parameter importance by their weight in the regression model.

Most building energy modeling sensitivity studies preceding calibration use total annual energy use as the output variable [18] [28] . However, for purposes of hourly or HVAC calibration, this is insufficient. In residential buildings, using the regression parameters of temperature-dependent change-point models better identifies important heating or cooling variables than use of total annual use alone [100]. Commercial buildings, having greater variability and being much more likely to be driven by internal loads, are less explainable with change-point models, requiring a different set of output variables to determine hourly and HVAC energy influences. These output variables are known as quantities of interest (QOIs).

6.1.1 Quantities of Interest (QOIs)

Quantities of Interest (QOIs) are numerical properties of a set of comparison data used to determine the model fit. These are similar to the concept of shape factors [21] [101] used to condense high dimensional time series data, such as daily hourly load profiles, down into a few parameters to represent the data. Quantities of interest are based on a comparison of annual hourly electric load data, typically 8760 hours. To condense this down into QOIs, the data is split into seasons (heating/winter, cooling/summer, and shoulder) and base and peak magnitudes.

6.1.2 Detailed Quantities of Interest

- **Annual Energy Use, kWh (1)** Average annual electric consumption for the whole year. Sum all 8760 hrs.

- **Average Daily Base Magnitude By Season, kW (3)** Average minimum daily magnitude (kW). For each season, create an average daily load profile from all days in that season. The base magnitude is the minimum value that occurs during the average day.

- **Average Daily Peak Magnitude By Season, kW (3)** Average maximum daily magnitude (kW). For each season, create an average daily load profile from all days in that season. The peak magnitude is the maximum value that occurs during the average day.

- **Average Daily Peak Timing By Season, hr (3)** Average maximum daily timing (hr). For each season, create an average daily load profile from all days in that season. The peak timing is the hour when the maximum value occurs during the average day.

- **Top 10 Daily Peak Magnitude By Season, kW (2)** Top 10 daily magnitude (kW). For heating and cooling seasons only, create an average daily load profile from days with the 10 highest peaks in that season. The peak magnitude is the maximum value that occurs during the average top 10 day.

- **Top 10 Daily Peak Timing By Season, hr (2)** Top 10 daily timing (hr).

For heating and cooling seasons only, create an average daily load profile from days with the 10 highest peaks in that season. The peak timing is the hour when the maximum value occurs during the average top 10 day.

These quantities of interest can be calculated at three different levels:

- **Stock Total** The QOI is calculated using the entire aggregate data from all buildings in the stock model and the LRD data. This represents how closely the stock model matches the LRD data at a system level.
- **Individual Building** The QOIs are calculated for an individual building. This allows buildings and building types to be compared against each other, say comparing the peak magnitude and timing distributions for retail buildings vs office buildings.
- **Individual Building in Relation to the Aggregate** This calculates the QOIs at the building level *in relation* to the aggregate building model data. For example, the average daily peak magnitude QOI represents the magnitude of building energy use that coincides with the system aggregate peak magnitude. In this case, the timing QOIs are the *difference* between the individual building peak and the aggregate system peak.

Additionally, energy use (kWh) and rate (kW) QOIs can be normalized by floor area. As building size is the overwhelmingly most significant parameter in determining building energy use across the stock, normalizing by floor area helps to determine which buildings contribute disproportionately to QOIs per area.

6.1.3 QOI Seasonal Determination

Seasonal determination helps to separately calibrate building operation when the building is likely to be under different HVAC operational modes. These seasons will vary by location and also by building type and building characteristics. For example, a natatorium is likely to have heating year round, and a large office building with a data center will have cooling year round. There are three considerations when determining season classification:

- **The number of seasons** This is perhaps the easiest to stipulate, as buildings can either be in heating mode, cooling mode, or in mixed-neutral mode. This translates nicely to three seasons: summer-cooling, winter-heating, and shoulder-mixed-neutral.

- **The variable used to determine seasons** E.g. Outdoor drybulb temperature, outdoor sol-air temperature, percent of time building is a heating or cooling regime, etc. are all viable parameters. Some parameters like outdoor air temperature emphasize seasonality and time-of-year, whereas others like actual building heating or cooling load try to capture the building regime directly. There is a trade-off between accurately capturing the building mode and keeping a broad enough determination that is generalizable across all buildings. For this reason, outdoor air temperature is the chosen parameter.

- **Time resolution of seasons** Seasons can be determined at the monthly level down to the daily level, based on average outdoor air temperature. Greater

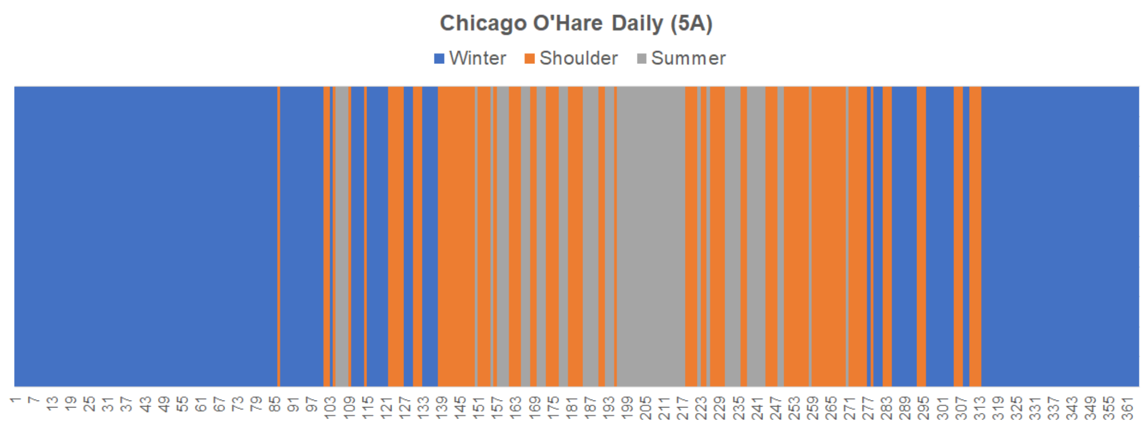

Figure 6.1: This graphic presents shows the seasonal determination for Chicago, IL when seasons are determined by daily outdoor average dry-bulb temperature.

resolution will better capture the building mode, but at a greater risk that all buildings in a utility territory do not follow the same season.

Based on the seasonal determination considerations, seasons here are determined by monthly average outdoor dry bulb temperature. Daily average temperatures <55°F are winter-heating, >=55°F and <=70°F are shoulder, and >70°F are summer-cooling. Considering the seasonal time resolution, Figures 6.1 and 6.2 compare using daily vs. monthly average outdoor drybulb temperatures. Monthly seasons maintain the same season across the same grid territory with limited loss in resolution.

Most utility regions serve a geographic area contained within one climate zone. However, to determine feature importance across the entire stock, it is helpful to break out feature importance by climate zone. As climate zones can span a large area (climate zone 4A for instance includes both Wichita, KS and New York, NY,

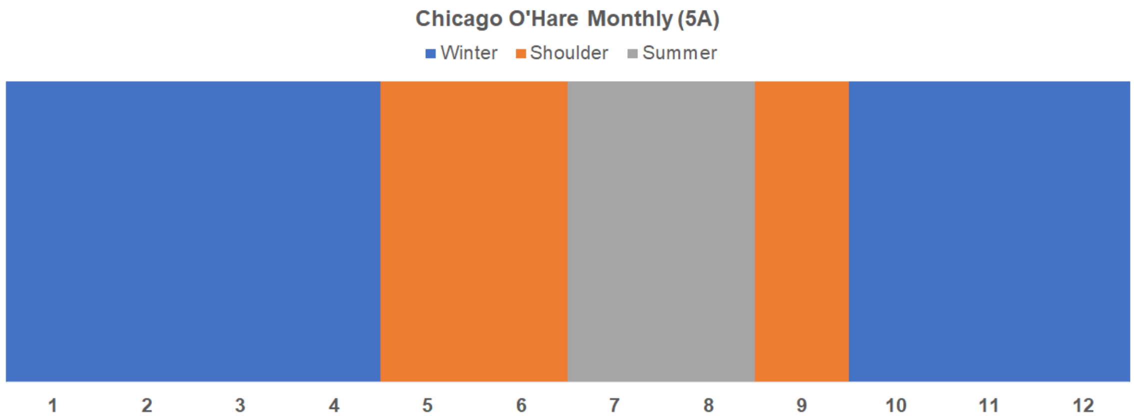

Figure 6.2: This graphic presents shows the seasonal determination for Chicago, IL when seasons are determined by monthly outdoor average drybulb temperature.

it is necessary to choose which months represent each season for that climate zone. Figure 6.3 describes this method. All weather files for the region are processed to determine the season for each month. Then, like seasons profiles are grouped together by building count and the modal season profile is selected to represent the climate zone.

climate zone	city	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	# bldgs
1A	Fort Lauderdale, FL	s	s	c	c	c	c	c	c	c	c	c	c	2145
1A	Key West, FL	s	s	c	c	c	c	c	c	c	c	c	c	118
1A	Miami, FL	s	s	c	c	c	c	c	c	c	c	c	c	2824
1A	Honolulu, HI	c	c	c	c	c	c	c	c	c	c	c	c	704
1A	Kahului	c	c	c	c	c	c	c	c	c	c	c	c	290

c = cooling/summer
s = shoulder
h = heating/winter

group by like seasons, and select the modal season to represent the climate zone

climate zone	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	# bldgs
1A	s	s	c	c	c	c	c	c	c	c	c	c	5087
1A	c	c	c	c	c	c	c	c	c	c	c	c	994

Figure 6.3: A visual description of the method for determining the season to use for each month for a given climate zone.

6.1.4 Parameters for Sensitivity Analysis

Building energy use is determined by a wide range of building characteristics, with hundreds of parameters in commercial building energy models [25]. Only a subset of these are particularly impactful, but this varies by building type, location, and other high-level characteristics. Several studies have put forward parameters to use for commercial building sensitivity analysis [33] [41] [70]. The parameters used for ComStock sensitivity analysis are drawn from these studies, as well as additional parameters to include a wider range of end uses and building types. Figures 6.1, 6.2, 6.3, 6.4, and 6.5 detail the parameters used for the ComStock sensitivity analysis. Some parameters are easy to capture and have limited variability, e.g. gas furnace nominal efficiency, while others are trickier to capture in a single variable, e.g. plug load schedule. Qualities of building parameters are explained in more detail below.

Categorical vs. Continuous Parameters. Some characteristics such as building service water heating fuel are categorical variables, and some are continuous, such as roof U-value. Sensitivity analysis works best with continuous variables. There are several methods to convert categorical variables to continuous. For this sensitivity analysis, categorical variables with few options, such as building type, are converted to onehot variables. Each building type becomes a separate column with 1 if the building is of that type and 0 if otherwise. Categorical variables with many options, such as HVAC system archetype, are converted to continuous numerical values by averaging the total annual energy use of all buildings with that characteristic. This includes the variable in the output, showing which system archetypes are more energy intensive than others. The presence of certain high energy use HVAC systems being highly correlated with certain building types is controlled for by including floor area and building type variables in the parameter set.

Implicit vs. Explicit Modeling Parameters. Building characteristics can be captured in several different ways in an energy model. Building wall construction is a good example. At the level of thermal balance calculation, specific heat, material density, thermal resistance, and absorptance at different spectra are the primary determinants of how the wall behaves. These parameters are rarely explicitly set by modeling practitioners however. Instead, modelers will choose a wall assembly which will implicitly define the set of thermal characteristics of the wall from either measured data or detailed thermal modeling analysis. The ComStock sensitivity study uses direct, continuous parameters where possible, which may be set explicitly (building aspect ratio) or implicitly (Wall U-value determined from wall construction type and vintage).

Schedule Parameters. Schedule parameters are highly influential in determining building energy use [41] but are very difficult to quantify as they are composed of values for each hour or sub-hourly increment and may be different by season or day of week. Schedules can be reduced down to a set of shape factors [21] [101] to describe the schedule. This study condenses schedules down to equivalent annual full load hours (EFLH), which calculates the number of hours a load would need to run at full load to match the aggregate integration of schedule value over time for the full year.

Parameter Interaction. Buildings are non-linear systems, which means some variables are highly co-dependent and require careful treatment in sensitivity analysis to capture their impact. Window solar heat gain and window-to-wall ratio are examples: window solar heat gain matters much more in building with a high window-to-wall ratio. Accounting for parameter interaction is handled by the choice of analysis method, which is explained in the next section.

Table 6.1: Parameters relating to high-level building characteristics.

Parameter	Description
Building Type	The building primary use and occupancy type
Location	The building location, represented by climate zone.
Size	Rentable floor area of the building.
Vintage	The age of the building, used in inferring the age of building systems, components, and efficiencies

Table 6.2: Sensitivity parameters relating to the building envelope.

Parameter	Description
Aspect Ratio	The ratio of length to width of the building, with length referring to the East-West axis of the building
Internal mass area ratio to floor area	The amount of internal thermal mass in the form of furniture and equipment, normalized by floor area
Number of floors	Number of above-ground floors
Orientation	The amount of rotation from the north axis
Roof absorptance	The absorptance of the roofing material
Roof U-value	The roof thermal resistance
Topographic projection index	The height of the building relative to neighbors, used to estimate the impact of shading from surrounding buildings
Wall U-value	The wall thermal resistance
Window-to-wall ratio	The ratio of window area as a portion of total wall area
Window U-value	The window thermal resistance

Table 6.3: Parameters relating to building loads including lighting, plug loads, occupancy, and miscellaneous process loads.

Parameter	Description
Daylight control space fraction	Fraction of lighting in a space controlled by a daylighting sensor
Elevator energy use	Energy use for elevators in the building
Exterior lighting power	Exterior lighting power for facade, parking, entrance, and signage lighting
Infiltration Rate	Annual average infiltration rate per exterior wall area
Interior lighting power density	Interior lighting power normalized by floor area
Interior lighting schedule	Equivalent annual full load hours of interior lighting
Occupant density	Number of people normalized by floor area
People schedule	Equivalent annual full load hours of occupants
Plug load power density	Interior equipment power normalized by floor area
Plug load schedule	Equivalent annual full load hours of interior equipment
Refrigeration	Annual energy use used for low and medium temperature refrigeration
Service water heater fuel type	Gas or electric
Service water heating water use	Total annual volume of hot water consumed by the building

Table 6.4: Parameters relating to building HVAC equipment including thermostat setpoints, component efficiencies, and control characteristics.

Parameter	Description
Air system fan minimum flow fraction	Average single-zone and multi-zone air system fan minimum airflow fraction, weighted by annual air flow rate
Air system fan static pressure	Average single-zone and multi-zone air system fan static pressure, weighted by annual air flow rate
Air system fan weighted efficiency	Average single-zone and multi-zone air system fan total efficiency, weighted by annual air flow rate
Air system outdoor airflow fraction	Average annual outdoor air fraction, used for determining DOAS vs. VAV operation
Boiler efficiency	Nominal thermal efficiency of the boiler, if present
Building fraction cooled	Fraction of building floor area that is conditioned for cooling
Building fraction heated	Fraction of building floor area that is conditioned for heating
Chiller efficiency	Annual and design COP of the chiller, accounting for part-load performance
DX cooling coil efficiency	Annual and design COP of DX cooling equipment, accounting for part-load performance
DX heating coil efficiency	Annual and design COP of DX heating equipment, accounting for part-load performance
Gas coil efficiency	Nominal thermal efficiency of gas furnaces

Table 6.5: Parameters relating to building HVAC equipment including thermostat setpoints, component efficiencies, and control characteristics.

Parameter	Description
HVAC system type	The type of HVAC system archetype, indicating fuel types and whether it is single-zone or multi-zone
Has System	Boolean variables for whether a building includes specific HVAC equipment. Separate variables for boilers, chillers, DX cooling, DX heating, zone HVAC system, zone HVAC fans, and multizone HVAC.
Plant loop pump head	Non-service water heating water loop pump head, weighted by plant loop annual mass flow rate
Plant loop pump minimum flow fraction head	Non-service water heating water loop pump minimum flow fraction, weighted by plant loop annual mass flow rate
Plant loop pump motor efficiency	Non-service water heating water loop pump motor efficiency, weighted by plant loop annual mass flow rate
Thermostat setpoint schedule cooling maximum	Cooling setback temperature
Thermostat setpoint schedule cooling minimum	Cooling setpoint temperature
Thermostat setpoint schedule heating maximum	Heating setpoint temperature
Thermostat setpoint schedule heating minimum	Heating setback temperature
Ventilation rate	Design outdoor air rate and average outdoor air fraction, normalized by floor area
Zone HVAC fan minimum flow fraction	Average zone HVAC equipment fan minimum airflow fraction for VAV systems, weighted by design air flow rate
Zone HVAC fan static pressure	Average zone HVAC equipment fan static pressure, weighted by design air flow rate
Zone HVAC fan weighted efficiency	Average zone HVAC equipment fan total efficiency, weighted by design air flow rate

6.1.5 Parameter Importance Calculation Method

The parameter importance method determines how sensitive a given QOI is to the input parameters in the model. In the language of machine learning, the input parameters are features input into a regression model whose output is the QOI. The relative weight of feature importance determines how strongly a given QOI is influenced by that feature. Features can be explicit inputs, implicit inputs, or unit-less features from a cluster or principal components analysis to condense the parameter space [21] [101]. For the regression model considered here, features in the regression model are one-to-one mappings of the model input parameters.

There are many regression models and methods available for sensitivity analysis. Prior work for building sensitivity analysis has typically used the one-at-a-time (OAT) method, also known as the Morris Method, as this is simple and easy to implement [28], [33], [100]. One-at-a-time and other simple linear regression methods are not able to properly capture parameter interaction. While features can be constructed that include interactions of two parameters, inclusion of all such interactions results in too large of a feature space and not enough training data, resulting in model overfitting. Also, as described in the introduction to this section, sensitivity analysis method typically require a pre-defined sample which is computationally prohibitive.

For these reasons, the feature importance method for this project uses an ensemble random forest method constructed by sampling from many decision trees [102]. Decision trees have a set number of branches and leafs, with each branch

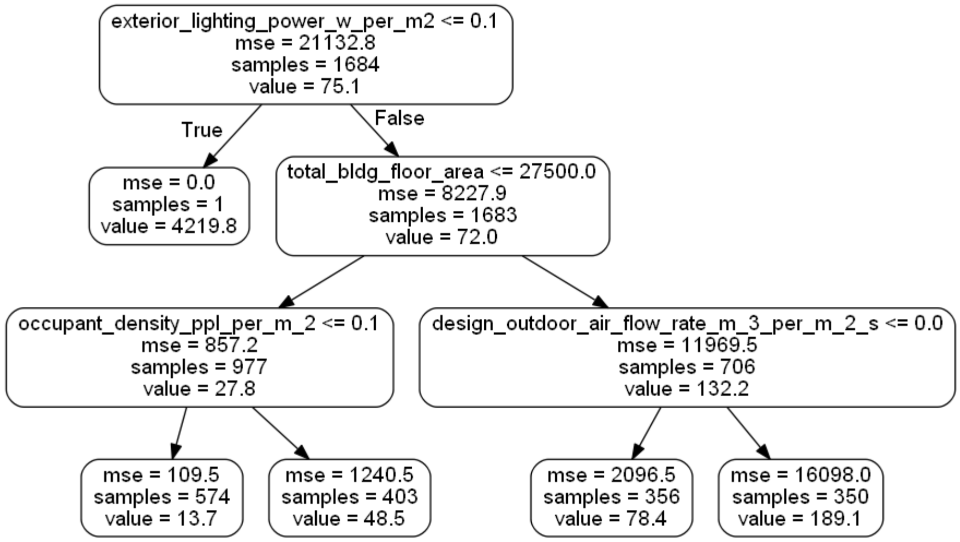

Figure 6.4: A visual example of a decision tree for determining total building energy use.

in the tree determined by a feature (building input parameter). Figure 6.4 gives an extremely simplified decision tree for determining total annual building energy use. At each step, the branch taken is based whether the input parameter is over or under a given value. This continues until it reaches a leaf, which will determine the value of the output variable. The decision tree for calculating quantities of interest is based on 66 feature inputs with 400 estimators.

Decision trees can be very accurate, but tend to overfit their data. By taking an ensemble of decision trees, known as a random forest, it is possible avoid the overfitting issue while still retaining model accuracy [102]. This was used successfully in developing a simplified regression model of a set of office building energy modeling results for the EnergyStar program [103]. To determine model accuracy, 20% of the data is reserved as testing data and the model is trained on the remaining 80%. A large decrease in the accuracy between the training and testing datasets indicates model overfitting. Accuracy is determined by mean square error. The training and testing data split and the random forest regression were implemented using the python scikit-learn package [104].

The relative importance of a decision node in a tree, namely how close to the root it is, will give a relative weight of the importance of that feature. Averaging the relative locations of features in a decision tree will reduce the variance in the feature importance for that parameter, a metric known as mean decrease in impurity [105]. This is the method used in scikit-learn to determine feature importance. The resulting feature importance metric assigns each feature a value of importance relative to other features, with all values summing to 1.

Give how sensitive building energy use is to floor area, that parameter was initially excluded from the model. However, the model has poor accuracy, with a significant drop in the testing set as shown in Table 6.6. Features that are highly correlated with large floor area, such as elevator use, show up highly in the feature importance because of this. To avoid this issue, floor area was included in the model and an additional set of QOIs were added to normalize total kWh and kW values by floor area.

Table 6.6: Comparison of random forest model accuracy including and excluding total building floor area for determining the average top 10 peak summer magnitude (kW).

Model	Training Accuracy	Testing Accuracy
Model excluding total building floor area	91.7%	77.1%
Model including total building floor area	94.5%	95.9%

6.2 Parameter Importance Results

The results of the sensitivity analysis are rankings of feature importance (summing to 1) for each QOI and the random forest regression model training and testing accuracy. The results are based on a sample region of Fort Collins, CO, which is in climate zone 5B. Figure 6.5 shows the model accuracy for different sets of QOIs. QOIs are grouped based on whether they are calculated in relation to the individual building or in relation to the aggregate stock, and whether the values are floor area normalized or not. The training model accuracy is greater than 80% for all QOIs except those that related to peak timing. Summer QOIs have high accuracy, while winter QOIs generally have poor accuracy, especially the winter peaks. This is likely because winter peaks most often coincide with peak heating demand, driven by substantial fan and heat pump energy use. Heating is mostly gas, and the QOIs are all electric, so minor differences in base fan or cold weather heat pump performance can result in substantial differences in energy use.

The random forest models are most accurate at predicting the average daily summer maximum. Feature importance calculated at the building level to determine the average daily maximum use in summer is shown in Figure 6.6. Comparing the feature importance unnormalized vs. normalized by floor area shows several differences. First, building floor area drops from being the most overwhelming significant parameter, though still high, suggesting that there are second-order effects to size that are not captured by other variables. Exterior lighting also drops significantly. It is unclear why this variable ranks so highly in the model for summer QOIs; it could

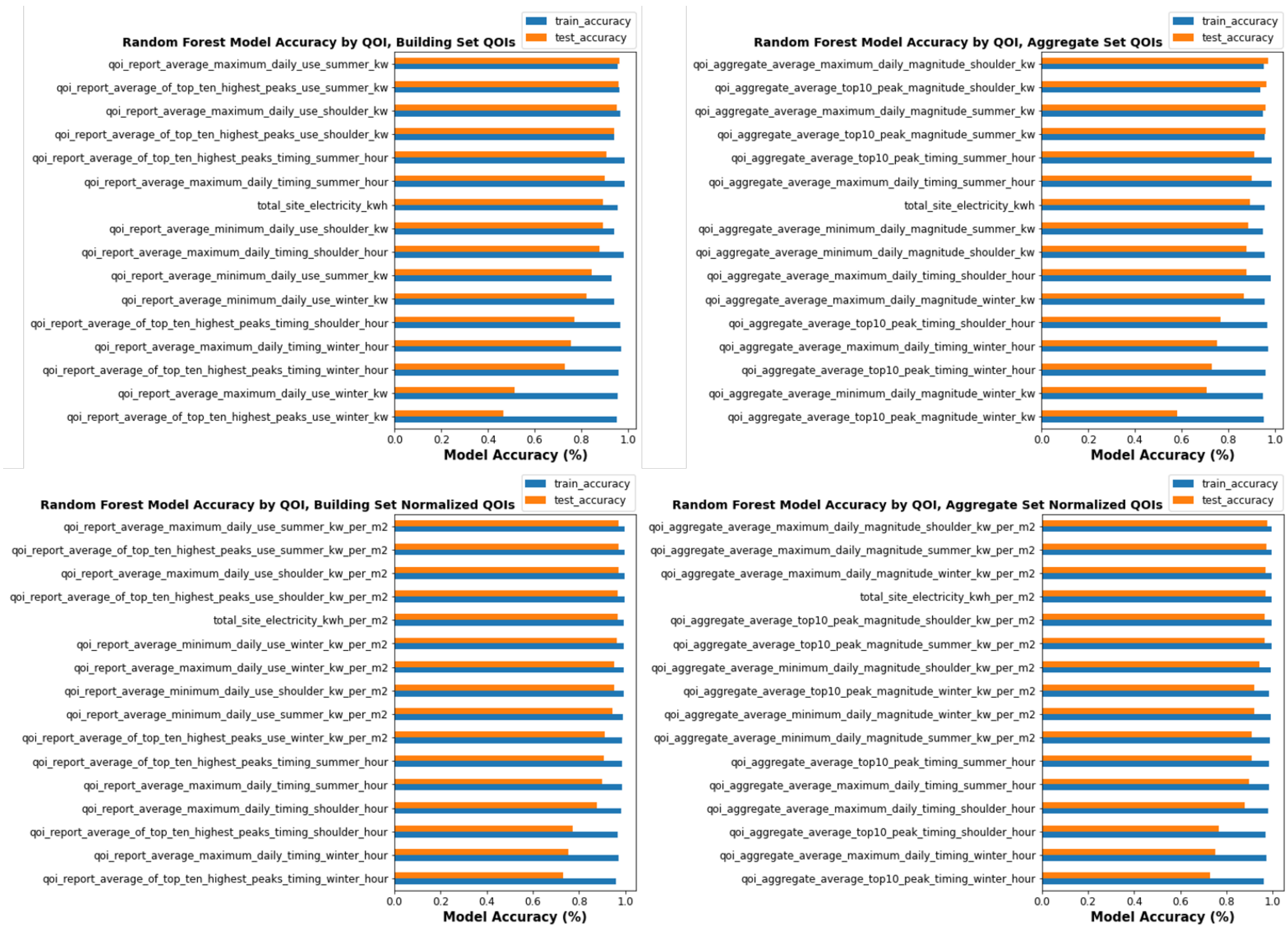

Figure 6.5: The testing and training accuracy of the random forest model for run 6.

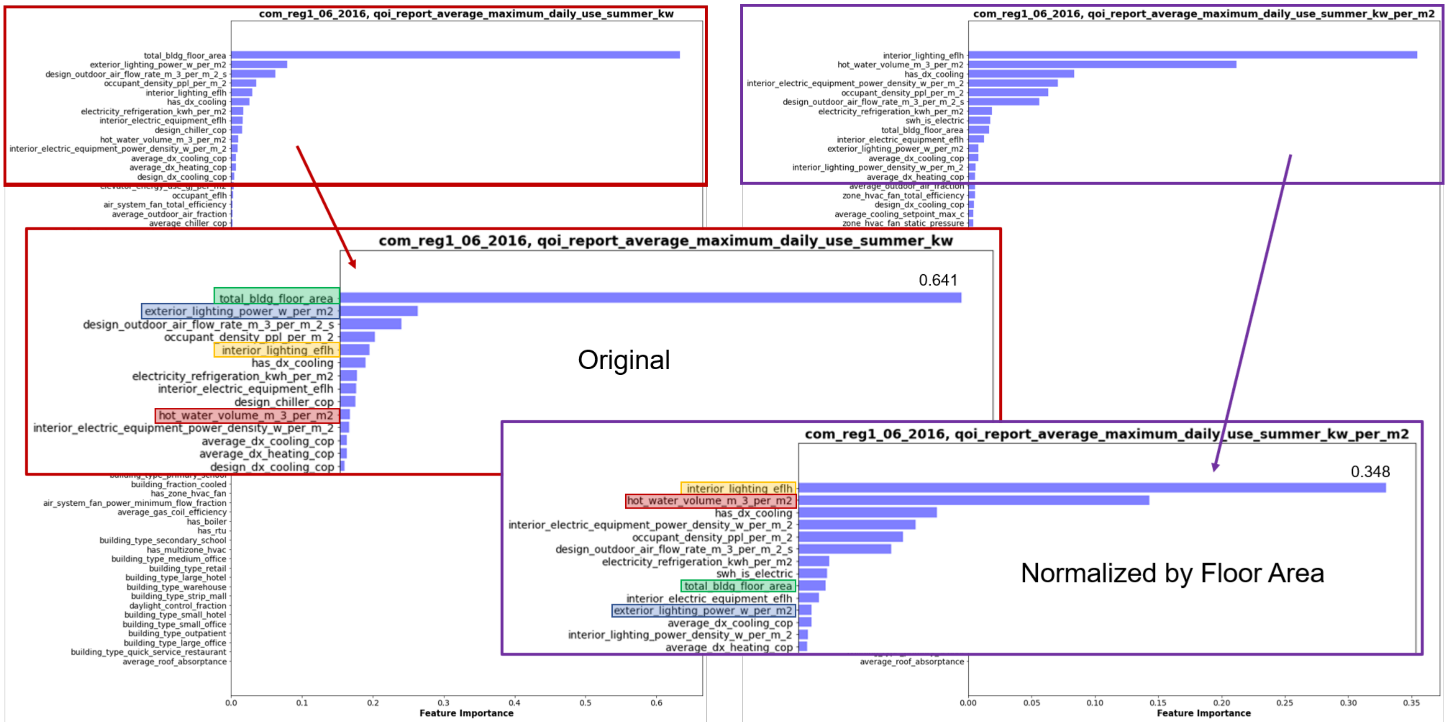

Figure 6.6: The relative feature importance for input parameters in the unnormalized and normalized case for the building average daily maximum use in summer, kW.

be that buildings with high summer peaks are more likely to have exterior lighting (restaurants, retail stores). Once controlling for floor area, hot water use (indicative of restaurants) shows up significantly in the dataset. Other variables appear in roughly similar ranking - outdoor air flow rate, occupant density, the presence of DX cooling equipment, interior lighting schedule full load hours, refrigeration density, and interior plug load density, and are influential for most individual building QOI values. These are parameters to consider when calibrating at the individual building level.

Stock model calibration prioritizes a different set of parameters. While a given

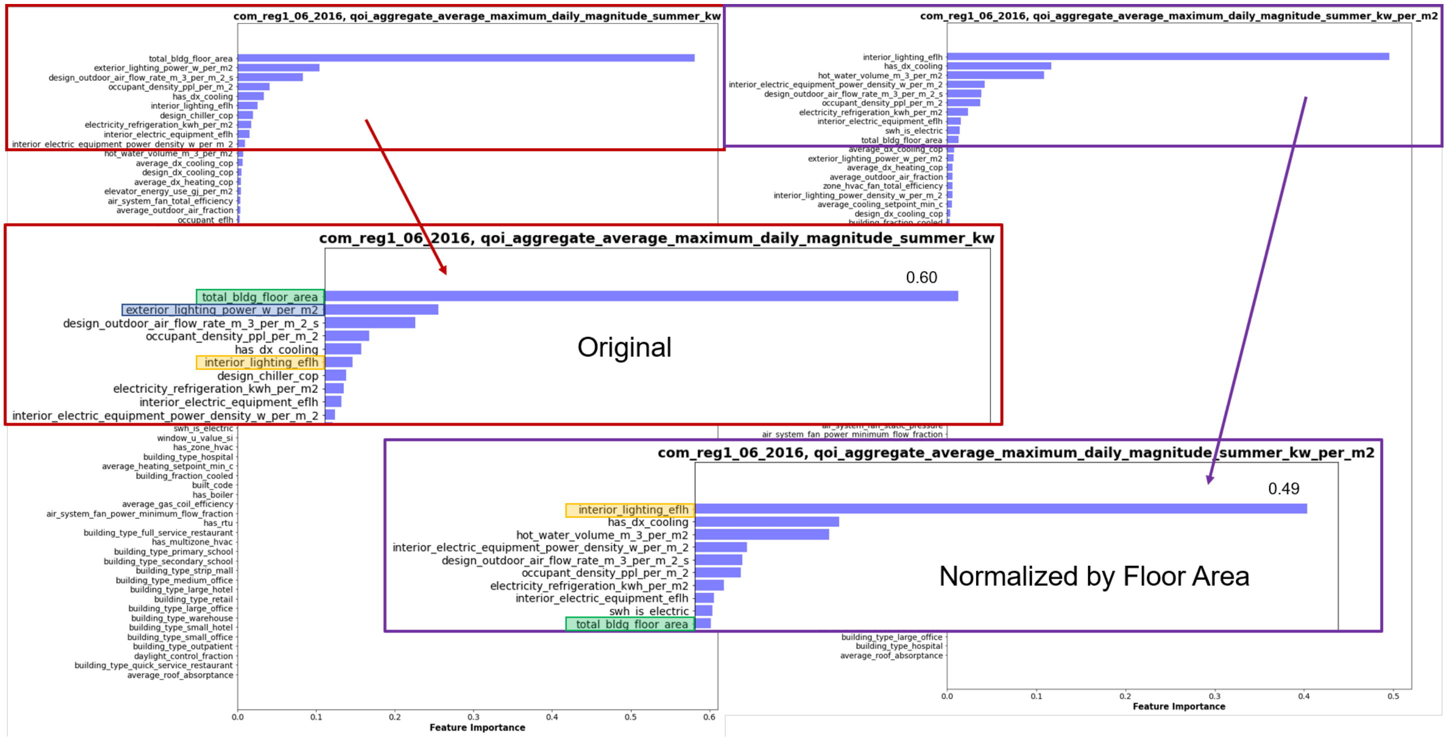

Figure 6.7: The relative feature importance for input parameters in the unnormalized and normalized case for the aggregate average daily maximum use in summer, kW.

parameter may be influential if it significantly contributes to individual building load, such has high hot water use, it may not be significant to the total grid load if there are relatively few buildings with those characteristics. Figure 6.7 shows the same feature importance rankings but in relation to the aggregate commercial building load on the grid, rather than individual building. Notably, the interior lighting schedule full load hours is much more important, indicating it drives total load, and hot water use no longer shows up in the top ten most important parameters.

For the sake of brevity, there is no need to include the ranked feature importance for each QOI here. A few features are significant for most QOIs, so feature importance can be summed across QOIs to determine an overall feature importance. To do this, QOIs are grouped into one of four sets based on whether the QOIs are at the individual building level or in relation to the aggregate, and whether the data is normalized or not. Aggregate unnormalized QOIs are most significant for stock model calibration. To give an overall weighting, the feature importance values for each aggregate QOIs were summed together by parameter. These QOIs are listed in Figure 6.5 and the weighted feature importance is given in Figure A.7. Results for other sets and QOIs are listed in the Appendix. 80% of the feature importance comes from 12 input parameters, and 90% from 20 input parameters. Besides floor area, interior equipment and lighting densities, equipment schedules, outdoor air flow rate, heating setpoints, occupant density, and the presence of rooftop units are the most significant parameters. This suggests special attention should be paid to internal load power densities and schedules, thermostats setpoints, and roof top units for calibrating the Fort Collins, CO stock model.

The full ComStock model simulates buildings across many climate zones, and features that show as important in Fort Collins, CO may not accurately represent the full building stock across the U.S.. To calculate the feature importance for the full U.S. stock, feature importance was calculated for each climate zone, weighted by the number of buildings in that climate zone, and then summed together and normalized to sum to 1. The result is shown in Figure B.3. The major differences between the full ComStock feature importance and Fort Collins, CO are that hot water use and gas coil efficiency show up much higher - indicating heating fan energy use and hot water heating are significant parameters for other regions of

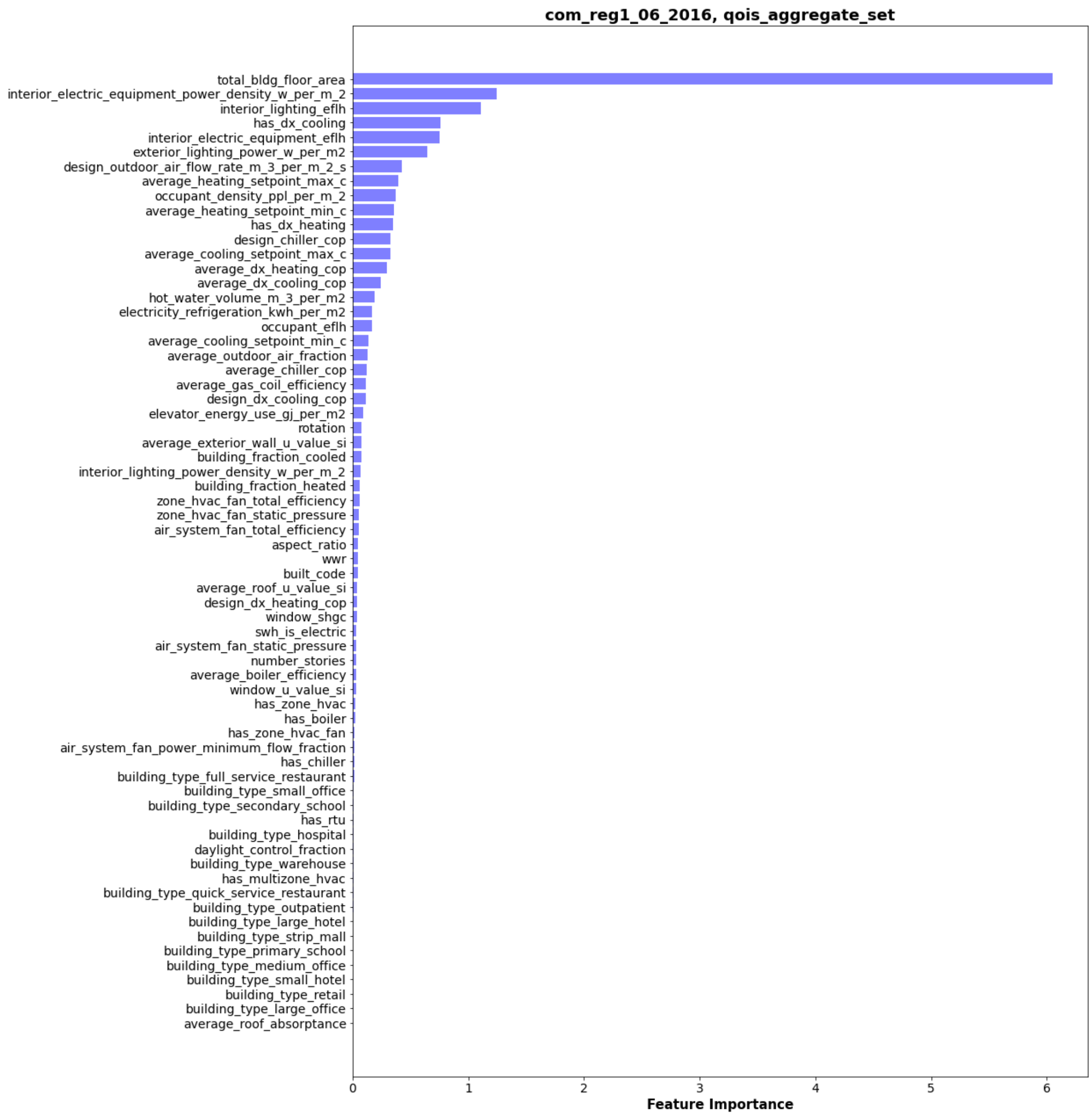

Figure 6.8: The weighted feature importance of input parameters for aggregate unnormalized QOIs.

the country. Hot water use density and gas coil efficiency are likely important not because they have a direct casual influence, but because they are highly correlated with other influential variables, as shown in the correlation matrix in Figure 6.10. Hot water use is highly correlated with equipment power density, indicating that is a strong predictor of high electric kitchen equipment. Equipment power density is largely bimodal, as shown in Figure 6.11. Hot water use is a better indicator of buildings with kitchens that contain high equipment power density than equipment power density directly. Gas coil efficiency has two values in the model, 78% and 80%, depending on the code year, making it the strongest correlate for building code year. Built code year is itself not as high, largely because retrofits of other equipment can occur in the model so the gas coil efficiency is a better predictor of HVAC system age. The variables indicate that a better accounting for kitchen equipment, making it explicit in the model, and making sure we are modeling code year and retrofit frequency accurately are critical for calibrating the full stock model.

6.3 Parameter Importance Conclusions

QOI feature importance for Fort Collins, CO and the full ComStock run showed similar results. Besides floor area, internal loads such as interior equipment power density and lighting schedules were among the most significant parameters in both sets. Also significant were hot water use density, gas coil efficiency, the presence of DX cooling equipment (RTUs), exterior lighting power density, outdoor air flow rates, and thermostat setpoints. Hot water use density, indicative of high

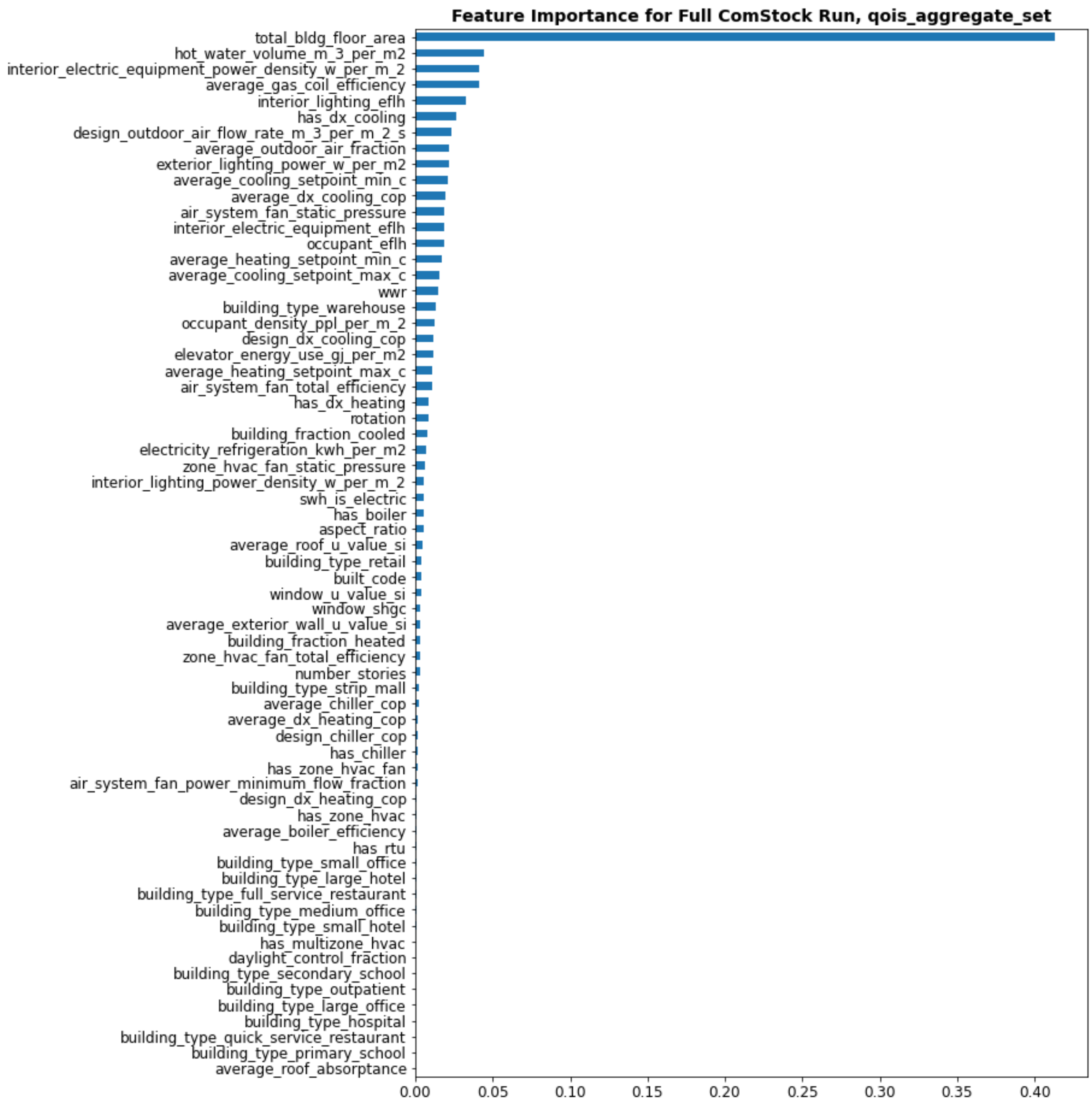

Figure 6.9: The weighted feature importance of input parameters for aggregate QOIs for the full U.S. commercial building stock.

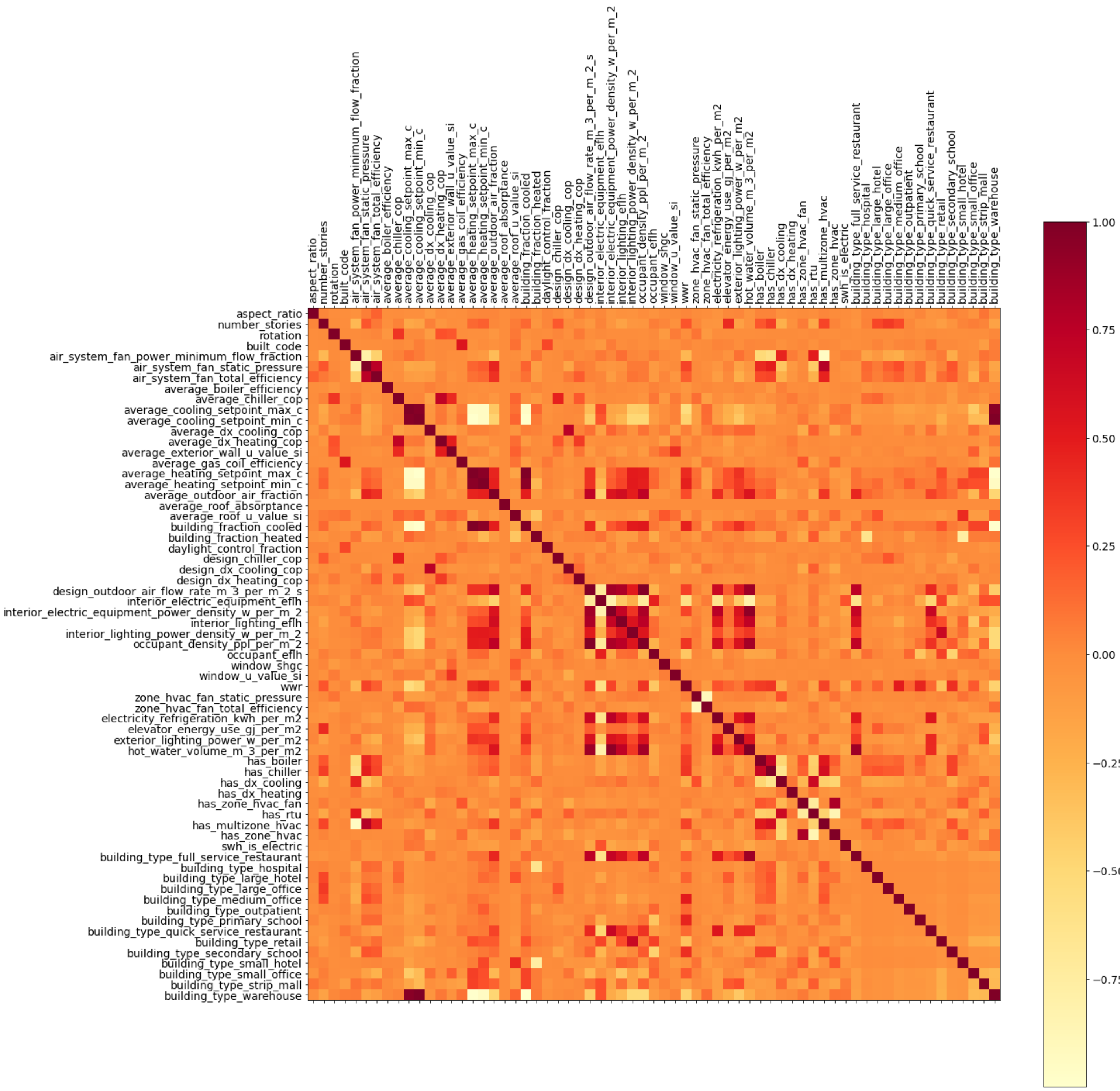

Figure 6.10: A correlation matrix for model input parameters for the full ComStock run.

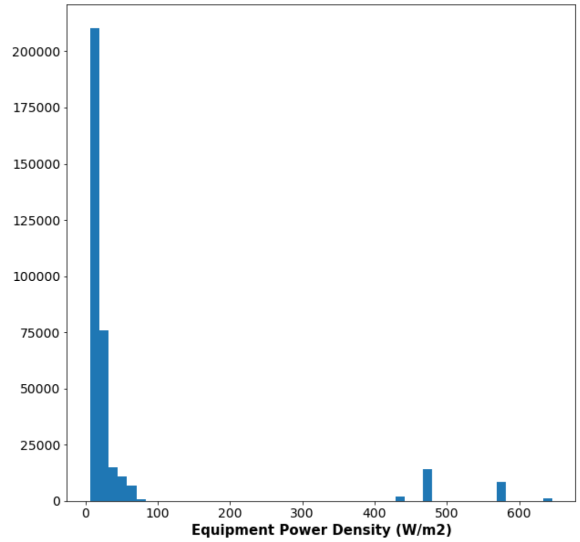

Figure 6.11: A histogram of equipment power density in ComStock Buildings. The highest cluster of the bimodal distribution represents buildings with kitchens.

electric kitchen equipment loads, and gas coil efficiency, indicative of building code year, are indirect parameters. These are calibration priorities to improve model accuracy across all QOIs.

Chapter 7: Stock Model Calibration

This section details the process of calibrating ComStock to utility data and end use meter data for Fort Collins, CO, shown in Figure 7.1.

7.1 Advanced Metering Infrastructure (AMI) data

Utilities in recent years have installed Advanced Metering Infrastructure (AMI) capable of capturing how much electricity a building uses on an interval basis, typically 15 minutes. For this project, AMI data was available for all buildings served by the electric municipal utility in Fort Collins, CO. AMI data is available for years 2014 through 2019, though 2014 and 2015 have sparse data as the metering programming was finishing the AMI roll-out. Weather data is available for 2016, so this project uses 2016 as the calibration year.

Each meter in the AMI serves a premise and is tagged with a unique premise identification (premise id). Several meters (premises) may exist on the same parcel of land, which is what CoStar™ and tax assessor data reference. The mapping of the three datasets - utility data, CoStar™, and the tax assessor database, is shown in Figure 7.2. Unique to Fort Collins, CO, the utility was willing to provide the mapping of premise ids to parcels (parcel ids), allowing us to match specific

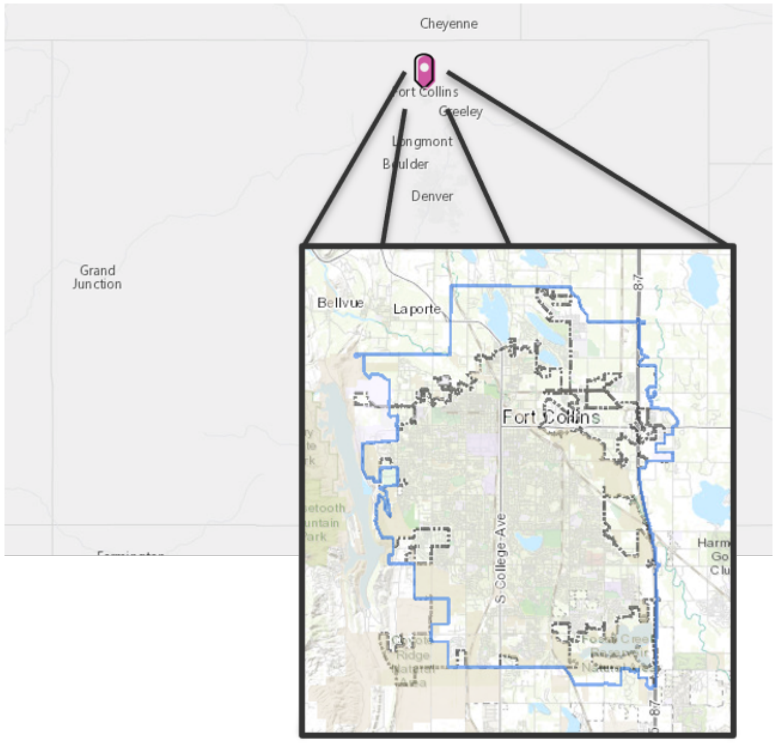

Figure 7.1: The region served by the Fort Collins, CO municipal utility.

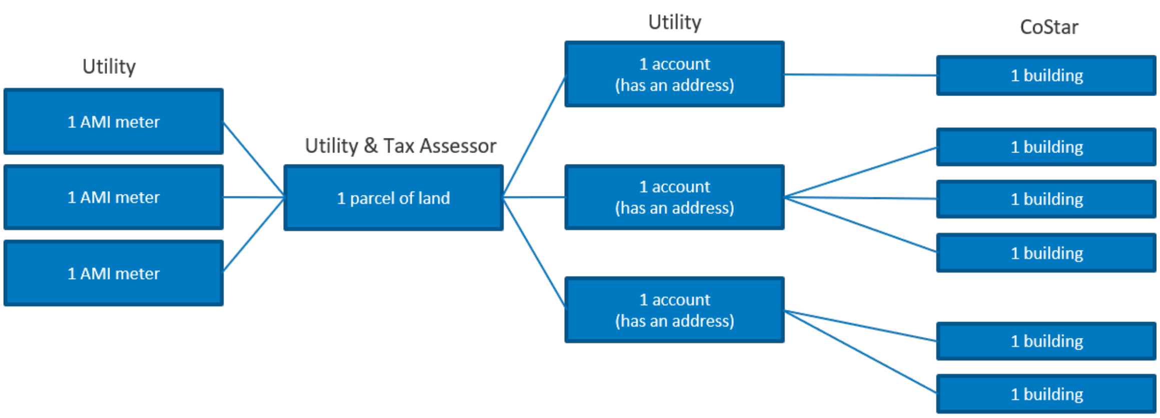

Figure 7.2: A diagram detailing how utility data is matched to building characteristics including size, physical address, and building type.

building meter data to a physical address. In all other regions, only the CoStar™ building classification is known, as the physical addresses can't be persisted and is not available from the utilities. CoStar™ includes a building type classification, which were each mapped to a ComStock building type. This allows us to tag each building with a ComStock building type, and then remove AMI data for buildings that are not represented in the ComStock model. Table 7.1 shows the number of buildings of each ComStock building type in the Fort Collins utility dataset.

Table 7.1: The number of buildings of each ComStock building data available for comparison.

Building Type	AMI Number of Buildings	ComStock Number of Buildings
full service restaurant	72	119
large hotel	8	30
large office	4	26
medium office	28	89
outpatient	96	136
primary school	21	55
quick service restaurant	29	38
retail	181	429
small hotel	7	33
small office	369	771
strip mall	223	626
warehouse	153	933
total	1191	3329

As there are many commercial buildings in the Fort Collins, CO dataset that can't be mapped to a ComStock building type or otherwise classified, this analysis compares energy use on a kWh per floor area basis, with total floor area sampled from CoStar™. To calculate kWh per floor area values, the utility data is summed to generate electricity use timeseries for each parcel id (kWh). Then, all parcel ids belonging to a specific building type are summed and divided by the total area of those parcel ids. The result is an hourly timeseries for each building type of the mean hourly kWh per square foot. The total kWh per square foot for the stock can then be calculated by multiplying each building type timeseries by the floor area of that building type in ComStock, summing the result, and the dividing by the total floor area in the ComStock sample. The ComStock sample is based on CoStar™ counts, as the AMI data represents a fraction of the building stock and is not as representative as CoStar for the relative weights of different building types. This total timeseries can then be used to calculate QOIs for the utility data and compare against the ComStock results. An initial load duration curve comparison, with all 8760 values ordered by magnitude for the utility data and ComStock results, is shown in Figure 7.3. The initial results show poor model fit, with the AMI data several times higher than the ComStock data. The initial AMI data has a total annual average electric EUI of 23.7 kWh/ft^2, or 80.7 kBtu/ft^2, which is very high for commercial buildings. A comparison of CBECS [98] results for the Mountain region, ComStock, and AMI data in Table 7.2 confirms that the AMI data is high, particularly for warehouse buildings. This suggests that AMI data may contain several outliers.

7.1.1 AMI Outlier Filtering Methods

A review of individual building parcel data explains the reasons for the high EUIs and suggests using an outlier filtering method on the AMI data to identify

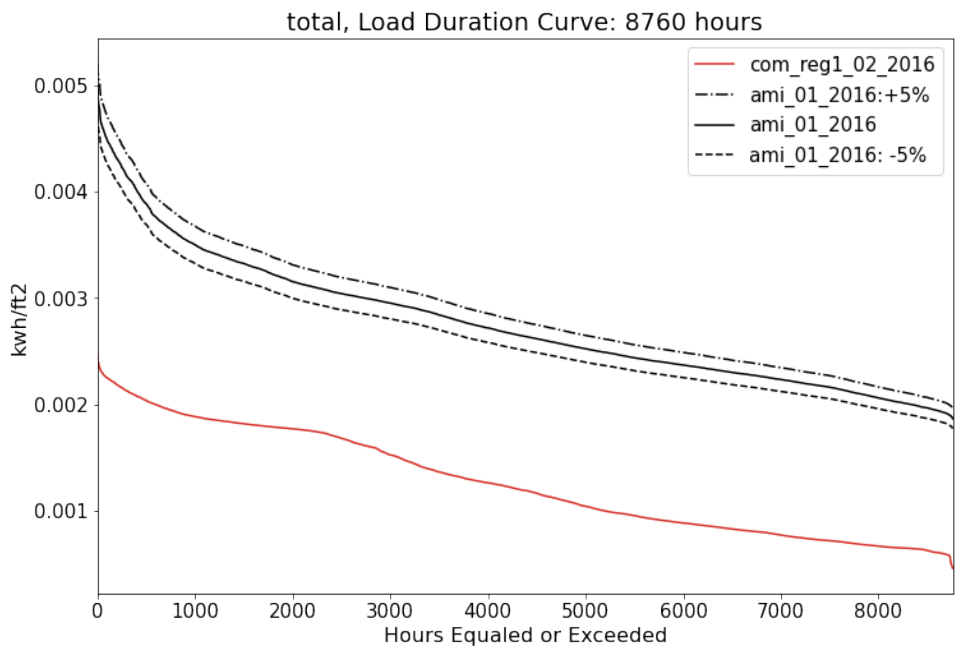

Figure 7.3: The initial load duration curve of uncalibrated ComStock results compared with raw AMI data for all buildings.

misclassified buildings and produce a more representative electricity use timeseries for comparison.

To identify problematic AMI data, the annual electric EUI for each parcel id is calculated and plotted as a distribution for each building type. As an example, Figure 7.4 shows the distribution for warehouses, with one particular building showing an electric EUI of 6600 kWh/ft^2, which turns out to be an electronics manufacturing facility. Investigations of several other high end use outliers show several gas station/convenience stores and restaurants included in the strip mall or retail category, autobody shops in the warehouse category, manufacturing facilities in the small office category, nursery and greenhouses in the retail category, and a few buildings in the quick service restaurants category where high drive-thru service is driving

Table 7.2: A comparison of annual EUI kWh/ft^2 by building type for the CBECS 2012 Mountain region, AMI, and ComStock.

Building Type	CBECS type	CBECS	AMI	Comstock
full service restaurant	food service	35.3	38.1	67.1
large hotel	lodging	18.2	12.1	19.7
large office	office	12.2	13.9	13.8
medium office	office	12.2	13.2	12.1
outpatient	outpatient	14.1	15.3	22.9
primary school	education	10.1	9.6	16.1
quick service restaurant	food service	35.3	95	73.8
retail	retail	18.7	22	14.9
small hotel	lodging	18.2	4.8	24.9
small office	office	12.2	15.5	10.9
strip mall	strip mall	18.3	15.1	15.7
warehouse	warehouse	4.1	**51.1**	5.2
total	all buildings	13.9	23.7	10.8

energy use instead of floor area. To identify outliers, several filtering methods are proposed and described below. The results of each outlier method, including how many buildings are removed are shown in Table 7.3.

7.1.1.1 Outlier Methods

- **No Filter** Keep all data. There are extreme high and low outliers; the high outliers overwhelm the averages and skew the data high.

- **IQR Filter** Remove values if the exceed a metric based on the interquartile range, given in Equation 7.1 [94]. This removes the highest values, but leaves low EUIs as $Q1 - IQR$ is often negative, leaving in vacant and misclassified buildings for some building types.

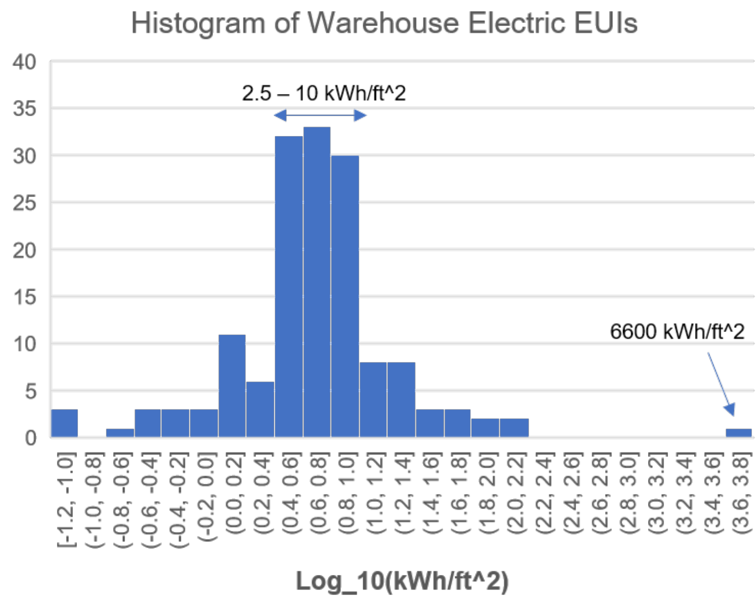

Figure 7.4: The distribution of EUIs for warehouse buildings in the Fort Collins, CO AMI data.

- **Ln IQR Filter** Remove values base on the interquartile range method, but use ln(kwh per sf). EUI distributions tend to follow a log-normal distribution, so this method enables the interquartile range filter to remove lower values. As the IQR is quite large, it leaves in many high, misclassified buildings.

- **Quartile 2nd and 3rd Filter** Keep only values in the second and third quartiles, based on the distribution of EUIs by building type. This is the most restrictive filter, and removes buildings that have well clustered EUIs are properly classified, such as large office and large hotel.

- **5 Times Median Filter** Remove values 5 times greater than or smaller than the median. Only removes extreme cases of high use misclassified buildings.

- **3 Times Median Filter** Remove values 3 times greater than or smaller than the median. Removes extreme high and lower outliers similar to the iqr filter and ln iqr filter methods, but is much less restrictive than the qrt23 filter.

$$\text{outlier } = \begin{cases} true \text{ if } x < Q1 - 1.5 * IQR \\ true \text{ if } x > Q3 + 1.5 * IQR \\ false \text{ else} \end{cases} \tag{7.1}$$

where:

Q1 = median of lower half of the data (x < median)

Q3 = median of upper half of the data (x > median)

IQR = Q3 - Q1

Table 7.3: The result average EUI for each building type and number of buildings remaining, given as EUI kWh/ft^2(# of buildings).

Building Type	none	iqr	ln iqr	qrt23	5x median	3x median
full service restaurant	38.1 (72)	34.2 (69)	39.3 (70)	36 (36)	37.9 (68)	38.6 (62)
large hotel	12.1 (8)	12.1 (8)	12.1 (8)	11.6 (4)	12.1 (8)	12.1 (8)
large office	13.9 (4)	13.9 (4)	13.9 (4)	15.5 (2)	13.9 (4)	13.9 (4)
medium office	13.2 (28)	12.5 (27)	14.4 (26)	13.7 (14)	14.4 (26)	15.7 (23)
outpatient	15.3 (96)	9.2 (92)	11.5 (82)	8.8 (48)	16.9 (84)	10.9 (80)
primary school	9.6 (21)	9.6 (21)	9.9 (19)	8.3 (11)	9.6 (21)	9.9 (19)
quick service restaurant	95.0 (29)	87.9 (27)	91.5 (28)	94.7 (15)	95 (29)	91.5 (28)
retail	22.0 (181)	15.3 (165)	18.9 (175)	12.0 (91)	16.1 (158)	16.0 (124)
small hotel	4.8 (7)	4.8 (7)	4.8 (7)	7.9 (5)	9.4 (6)	9.6 (5)
small office	15.5 (369)	10.4 (341)	12.6 (351)	9.0 (185)	12.2 (343)	10.9 (311)
strip mall	15.1 (223)	12.4 (199)	17.5 (213)	15.8 (113)	17.3 (186)	15.5 (158)
warehouse	51.1 (153)	3.0 (135)	6.4 (135)	5.4 (77)	6.5 (128)	6.3 (108)
total	23.7 (1191)	11.1 (1095)	15.0 (1118)	12.7 (601)	14.9 (1061)	14.0 (930)

Most buildings removed by filtering methods are misclassified, but there are several correctly classified buildings that represent true EUI variability for a given building type. Currently, ComStock has limited sources of variability through changing building system vintage and varying occupancy start time and duration. The wide variability in certain building types suggest that future improvements to ComStock should account for wider variability, particularly in space type ratios and equipment intensity. Some building types, particularly quick service restaurants with drive-thrus, may best be compared using some other metric instead of EUI as energy use is not entirely driven by floor area. For the purposes of generating comparison data for region 1, the 3x median filter method provides the best balance between removing both high and low outliers while including many more buildings than the filtering method that only takes the second and third quartiles.

7.2 Application of Feature Importance to ComStock Calibration

The sensitivity analysis identified several highly important parameters for stock model calibration. This section details the changes and improvements made based on metered data to improve the stock model calibration of ComStock, including results.

7.3 Model Changes

Three major changes were included in ComStock based on the sensitivity analysis: 1) lighting schedules and power densities, 2) equipment schedules and power densities, and 3) thermostat setpoint schedules. New values for these parameters

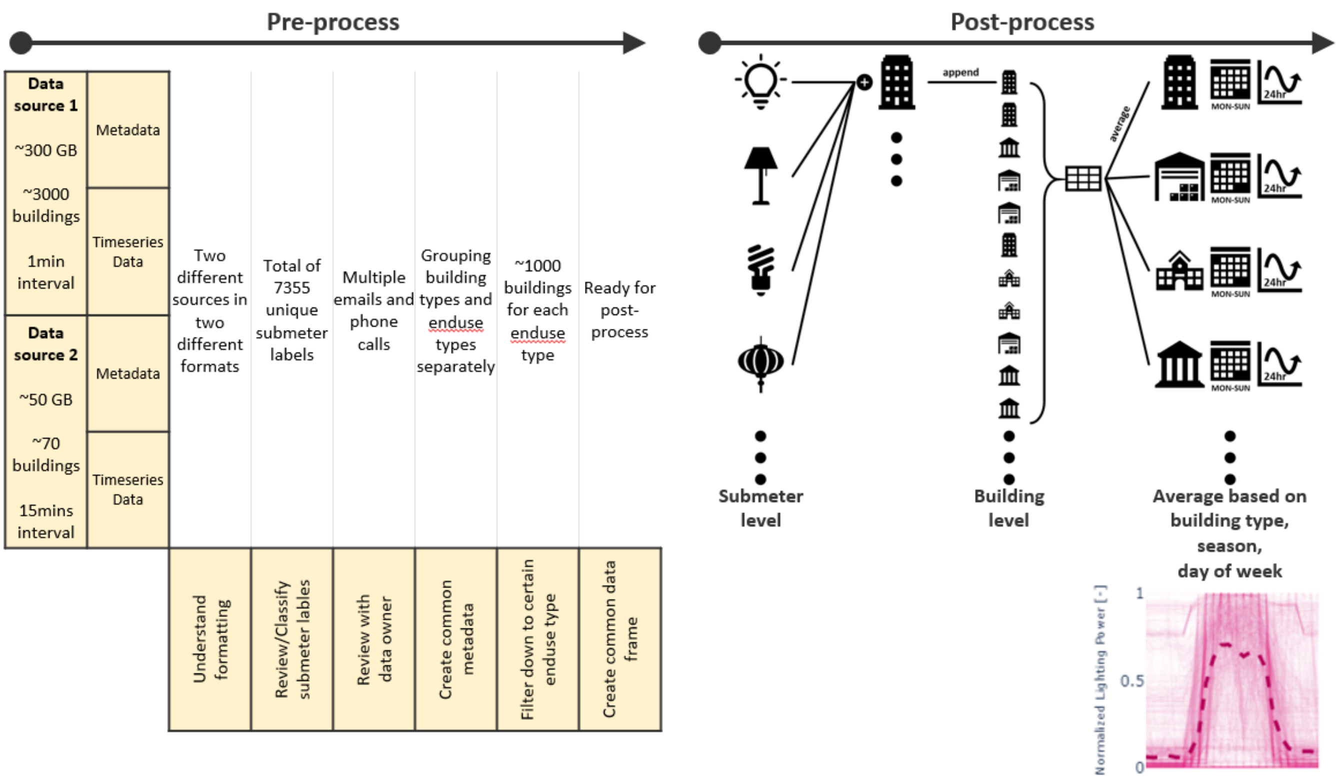

Figure 7.5: A graphical depiction of the method to pre-processes the end use data code by enduse and building type, and then using that submeter data to generated average schedules by building type for lighting, equipment, and thermostat setpoints.

were derived from analysis of two large end use meter datasets from two private submetering companies which requested they not be named in publications. The processing for this is detailed in Figure 7.5. Affected building types and method for calculating updated model inputs are listed below.

7.3.0.1 Description of Model Changes

- **Update lighting schedules**

  Building types: retail, full service restaurant, warehouse, office, primary school, secondary school

Approach: Calculated average daily profiles (hourly interval) for each building type and day of week. School buildings differentiate between the academic period and summer break.

- **Update lighting power density**

  Building types: retail, full service restaurant, warehouse, office

  Approach: The design lighting power density is assumed to be (5%) higher than the peak power in the dataset. The schedules are normalized against this peak power.

- **Update plug load schedules**

  Building types: retail, full service restaurant, grocery, warehouse, office, primary school, secondary school

  Approach: Calculated average daily profiles (hourly interval) for each building type and day of week. School buildings differentiate between the academic period and summer break.

- **Update plug load power density**

  Building types: retail, full service restaurant, grocery, warehouse, office, primary school, secondary school

  Approach: Peak power in dataset used to approximate difference in design peak power and peak power in the normalized schedule. The schedules are normalized against this peak power.

- **Update thermostat setpoint schedules**

  Building types: retail, strip mall, quick service restaurant, full service restaurant, grocery, office

  Approach: Calculated average daily profiles (hourly interval) for each building type and day of week.

Each model change was implemented in ComStock sequentially, starting with lighting schedules, then equipment schedules, and finally thermostat setpoint schedules. Lighting, equipment, and thermostat schedules are not monolithic for each building type. There is variability, for example full service restaurants clustering into those that serve lunch and dinner vs. only dinner. This variability is partly captured in the start time and duration inputs in ComStock, but there is room for improvement in capturing the range of schedule variation in real buildings.

ComStock model runs

- **com reg1 02 2016** The initial ComStock run without any model changes.
- **com reg1 03 2016** Adjust lighting schedules.
- **com reg1 04 2016** Re-normalize lighting schedules to annual peak and adjust lighting power density.
- **com reg1 05 2016** Adjust equipment schedules and equipment power density.
- **com reg1 06 2016** Adjust thermostat setpoint schedules.

7.4 Calibration Results

Assessing calibration results is a mix of quantitative and qualitative, graphical analysis. Qualitative methods compare load duration curves and stacked area graphs of enduse for each building type, broken out by weekday/weekday and QOI season. This shows the major constituents of energy use and shows how well the building matches a particular profile. Quantitative metrics include CVRMSE and NMBE, as well as QOIs. The initial calibration shows a decent match between ComStock and the AMI data 7.6. Peak timings are well matched, but a small amount of nighttime base load is missing, more pronounced during the summer seasons. The first round of changes, runs 3 and 4, adjusted lighting schedules and lighting power densities. Figure 7.7 shows the impact of lighting schedule changes on the daily load profile shapes for retail buildings. The new averaged schedules extend retail lighting into the evening hours, which better matches the load curve for retail buildings. However, exterior lighting turning on in the evenings is causing a bump in energy use that interrupts the evening ramp down of energy use. Figure 7.8 shows the implementation of run 5, updated equipment schedules. Restaurant equipment is now much more stable, and more closely aligns with average restaurants begin open for lunch and dinner but not breakfast. Lastly, Figure 7.9 shows the impact of run 6, updated thermostat schedules, on large offices. There is a slight increase in cooling and fan energy use, but it is not enough to overcome the underestimation of energy use, suggesting ComStock is missing some internal equipment or lighting load.

Figure 7.10 shows the enduse comparison for the total load and Figure 7.11

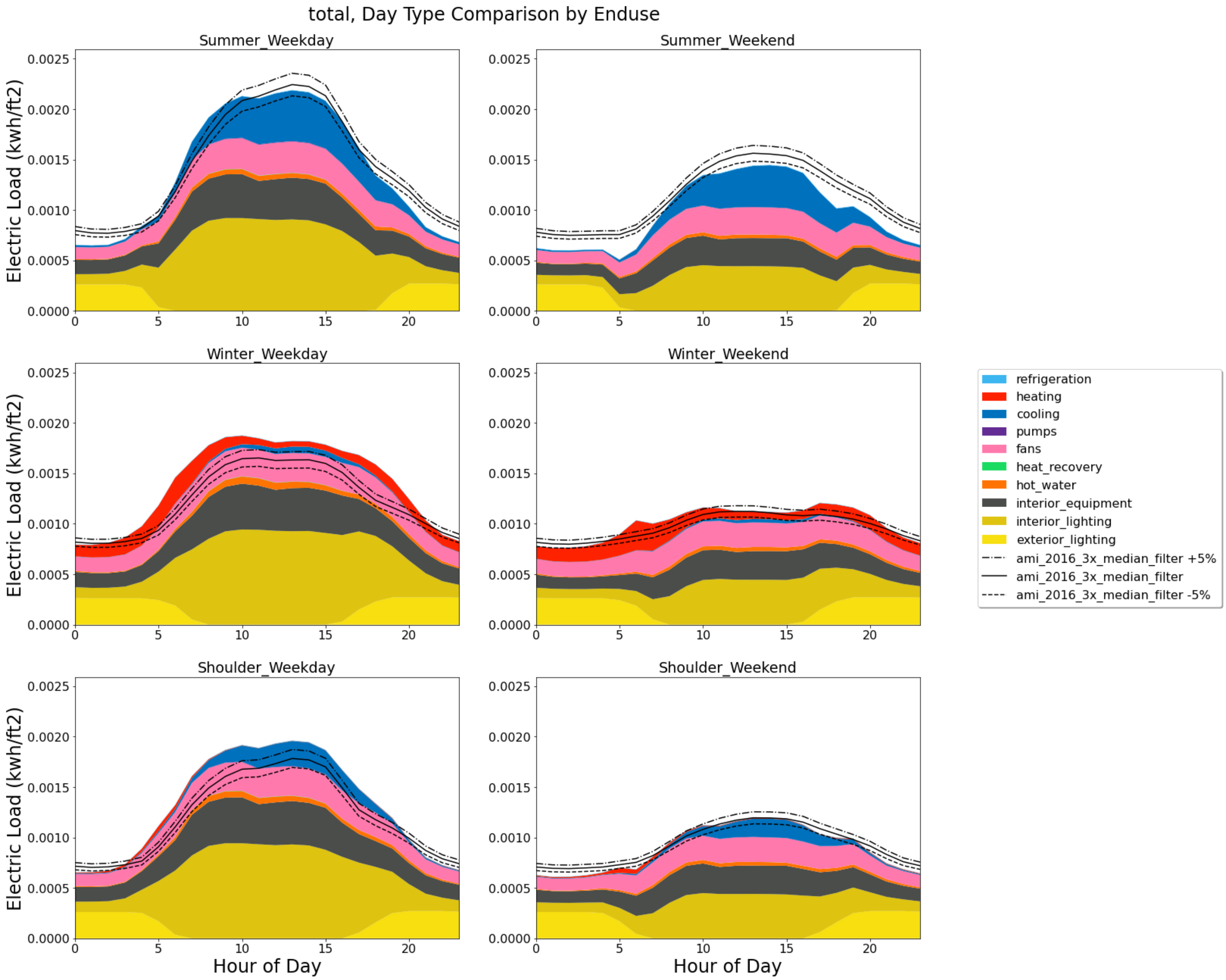

Figure 7.6: The average daily load profiles for weekdays and weekends for each season, broken out by enduse for the baseline ComStock run. The AMI data is shown with +-5% error.

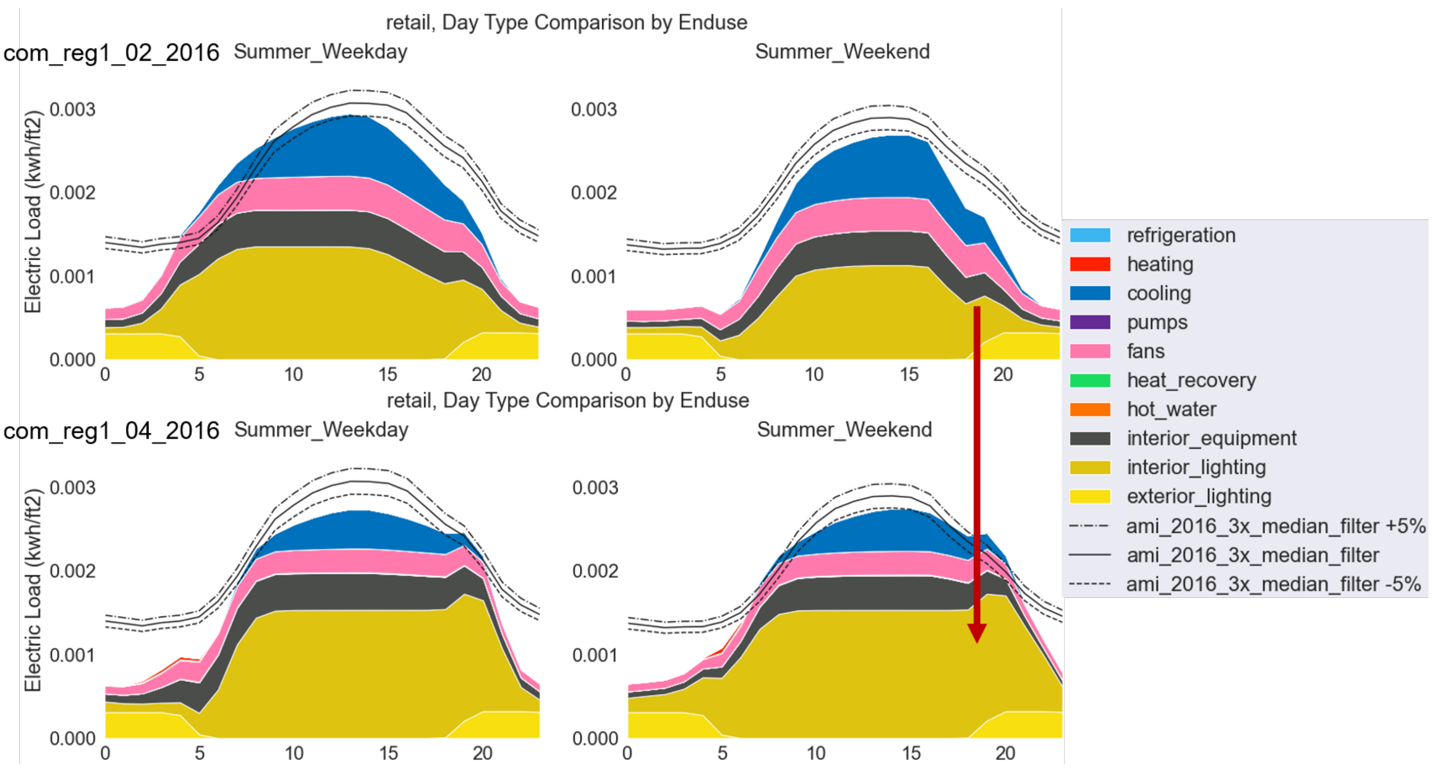

Figure 7.7: Before and after the implementation of updated lighting schedules and power densities for retail buildings.

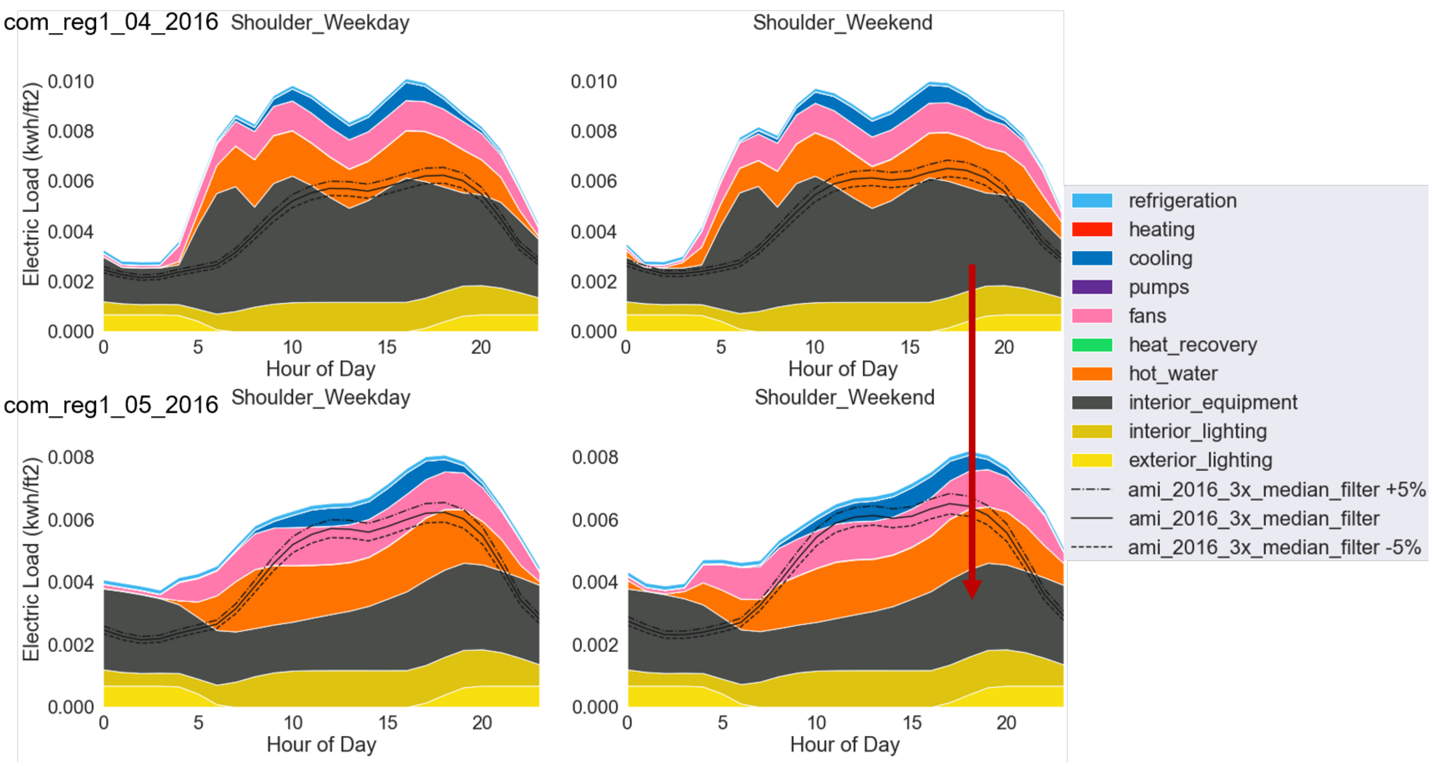

Figure 7.8: Before and after the implementation of updated lighting schedules and power densities for full service restaurants.

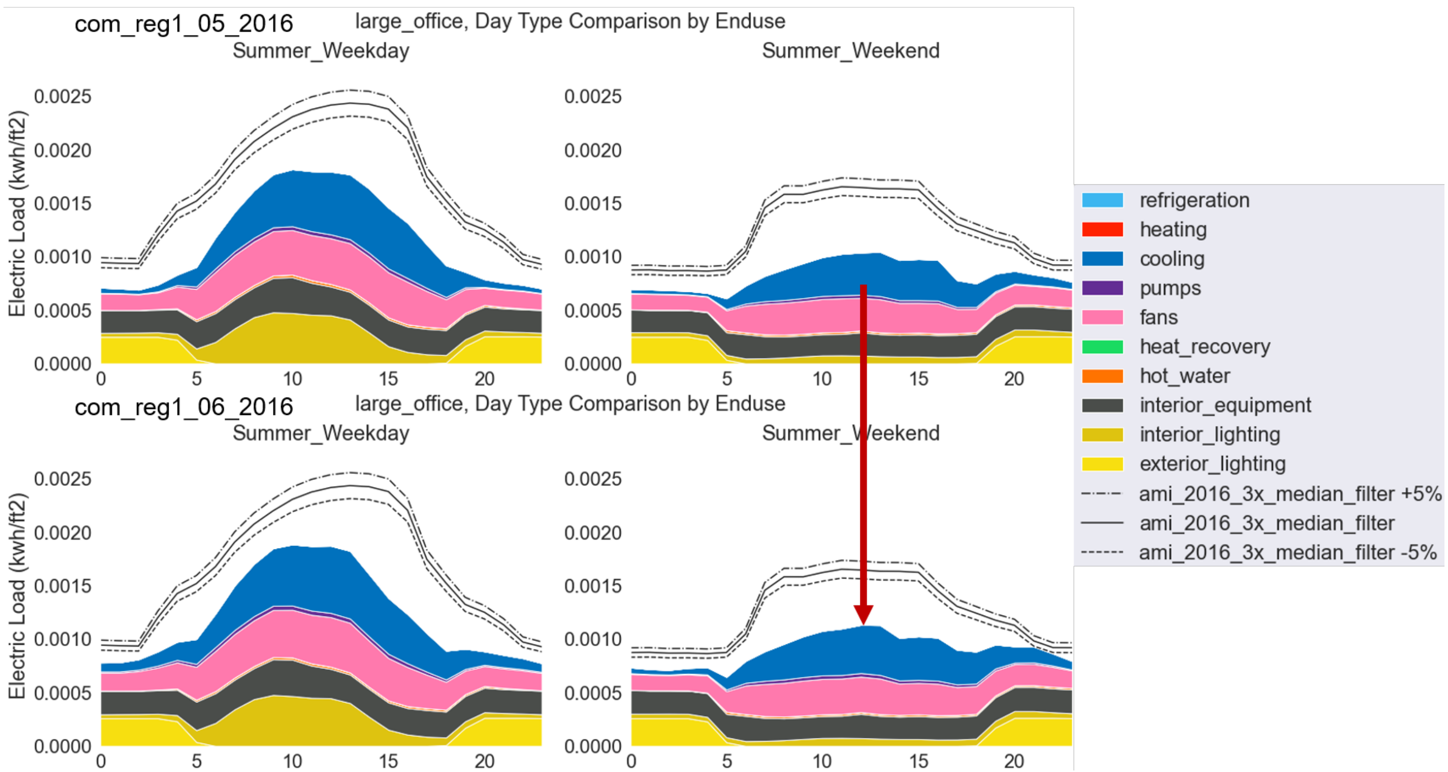

Figure 7.9: Before and after the implementation of updated thermostat setpoint schedules for large office buildings.

compares the load duration curve. There is no longer underestimation of nighttime loads, however some building types appear worse, such as warehouses, Figure 7.12. This suggests that while the total load curve may be accurate, it is important to consider each building type individually to show model fit. The total load evening peak comes from warehouses, which were a tricky building type to classify, and one where there were only 26 buildings in the enduse dataset to use for schedules. It's likely that these weren't representative – the company that provided the enduse data was the one to specify the building type to keep their client confidentiality, and as indicated in the analysis of the AMI data, warehouses are tricky to classify and often get mis-characterized. It could also be the case that warehouses in the enduse dataset are from one chain that has a different use profile, more typical of a distribution center, than what ComStock assumes. In looking at the enduse data, half the warehouses have erratic operation, with some operation only half of the year. The other building types did not have this issue because classification errors were less likely and there were more of them. Because warehouses are a significant portion of the stock, an additional run 7 reverts the changes to teh warehouse equipment and lighting schedule changes to warehouses, which provides a better fit, shown in Figure 7.13.

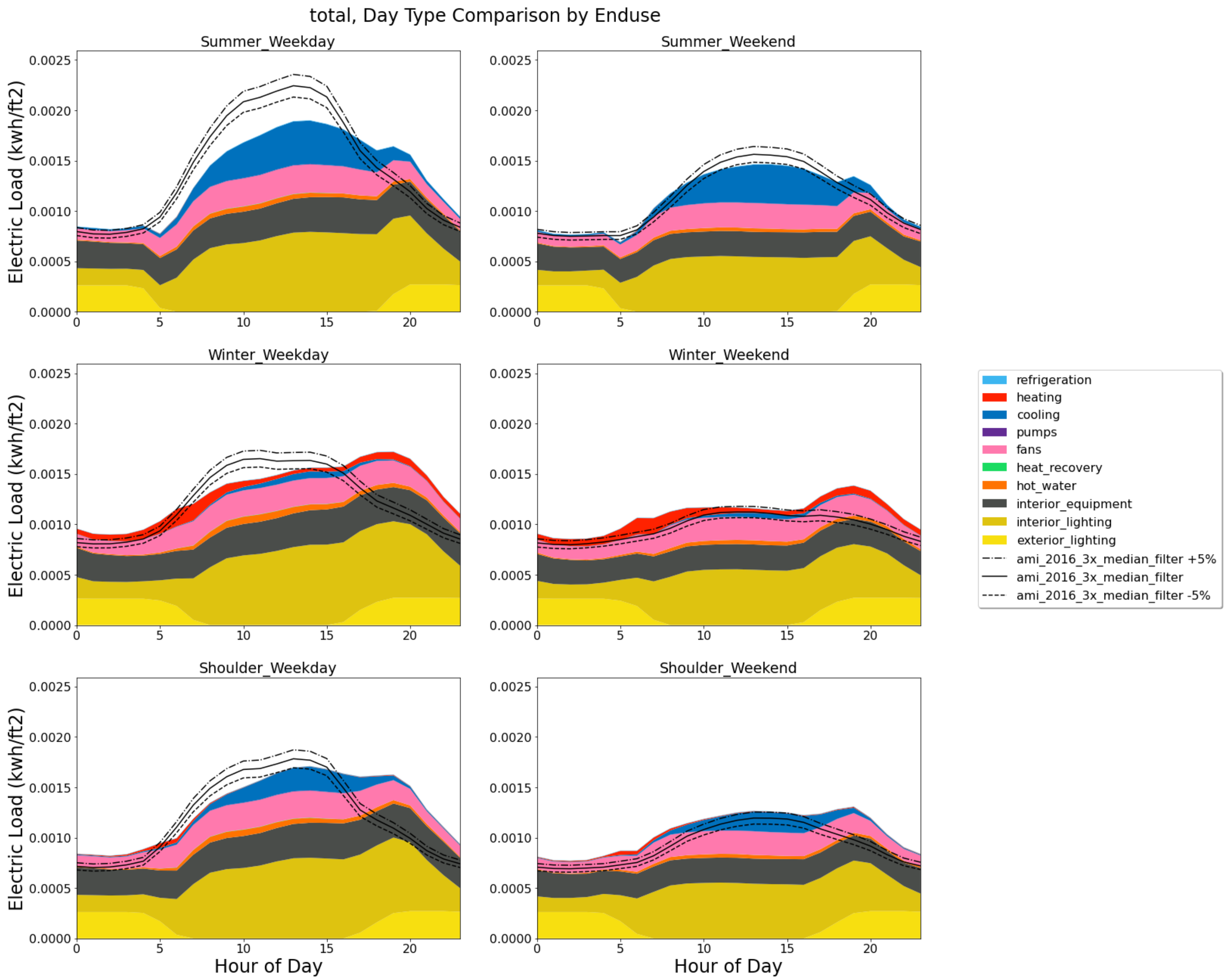

Figure 7.10: The average daily load profiles for weekdays and weekends for each season, broken out by enduse for ComStock run 6.

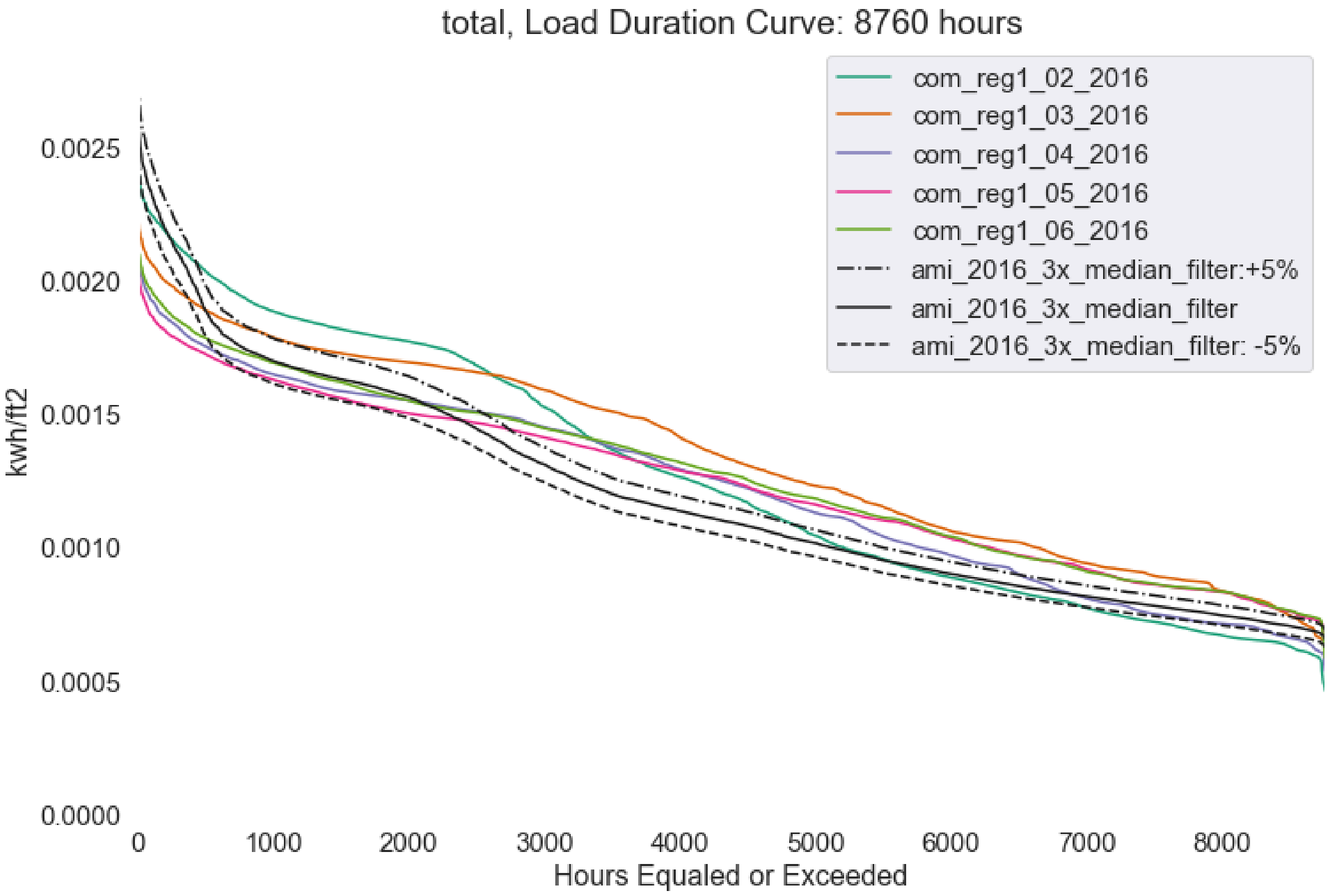

Figure 7.11: The load duration curve for each ComStock run as compared to AMI data.

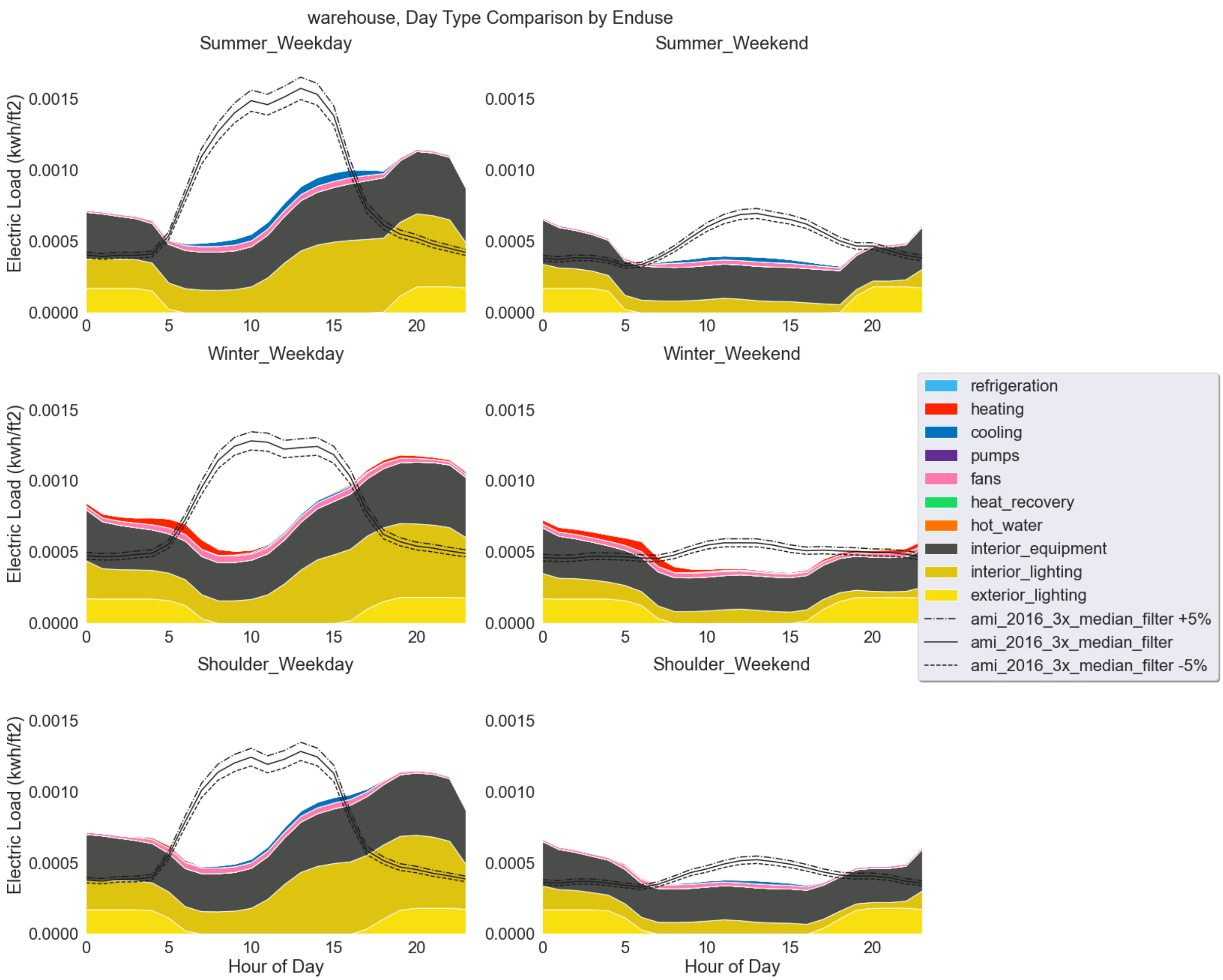

Figure 7.12: The average daily load profile for weekday and weekends for warehouse buildings. The ComStock data is far off, suggesting that warehouses are misclassified in the end use data processing.

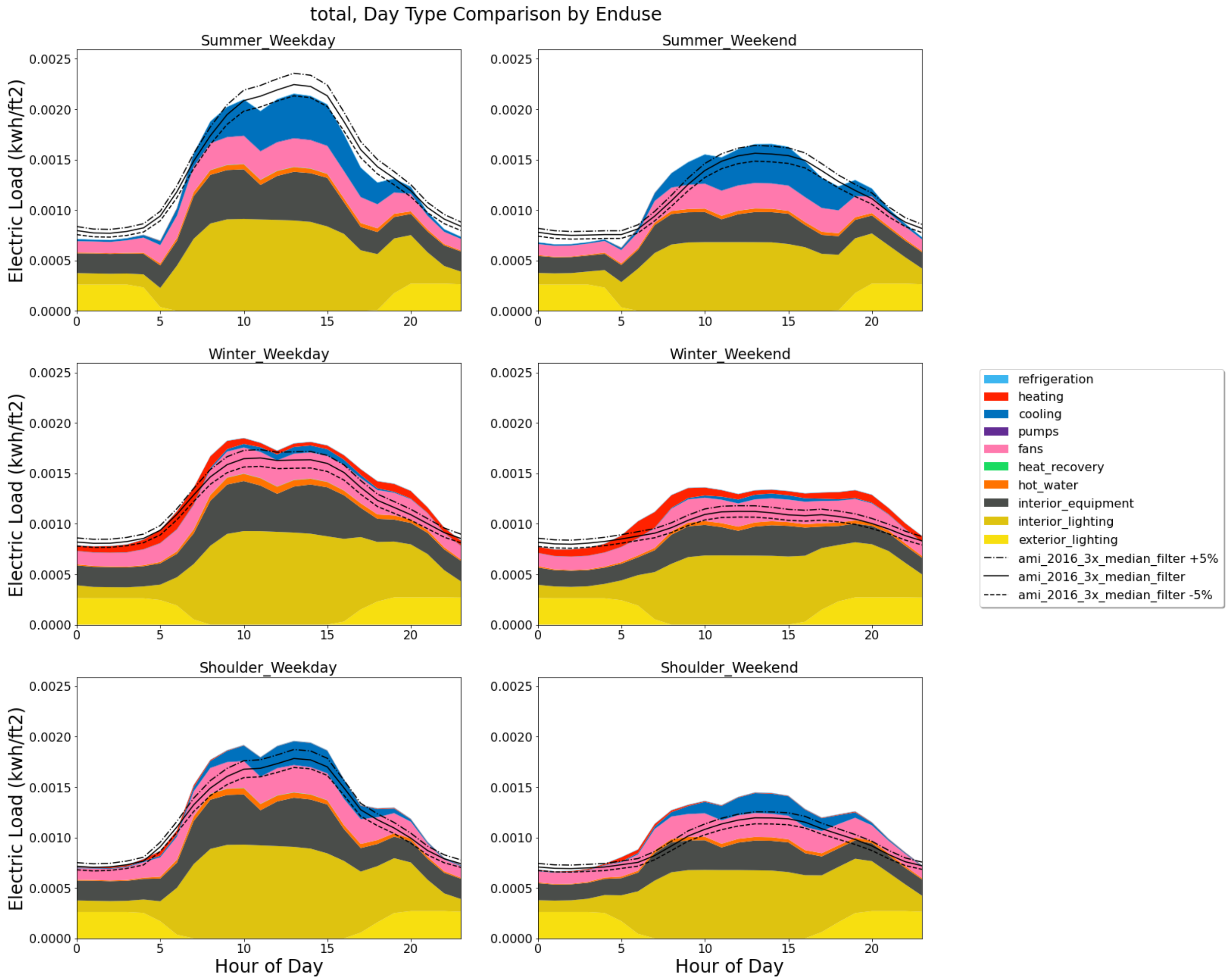

Figure 7.13: The average daily load profiles for weekdays and weekends for each season, broken out by enduse for ComStock run 7.

Table 7.4: Model calibration statistics of total electric load compared to AMI data. Positive error values mean the model overestimates the value. Positive minute values indicated the model peak occurs later than the AMI data.

metric	run 2	run 3	run 4	run 5	run 6	run 7	Improved?
avg. min summer kw err(%)	-18.5	2.1	-15.8	-7.3	-7.2	-19.2	FALSE
avg. max summer kw err(%)	3.7	-2.9	-8.9	-13	-7.3	5.1	FALSE
avg. top10 peak summer kw err(%)	-0.1	-8.4	-13.9	-18.2	-13	-2.7	FALSE
avg. min winter kw err(%)	1.4	13.1	-0.8	14.4	14.6	0.2	TRUE
avg. max winter kw err(%)	23.7	22	15.2	15.5	16.5	25.8	FALSE
avg. top10 peak winter kw err(%)	27.4	23.1	17.3	17.2	16.1	24.2	TRUE
avg. min shoulder kw err(%)	-6.5	15.2	-0.9	13.7	13	-1	TRUE
avg. max shoulder kw err(%)	17.2	14.1	6.5	4.4	8.8	22	FALSE
avg. top10 peak shoulder kw err(%)	11.1	3	-3.8	-9.1	-3.5	8.7	TRUE
avg. min timing summer hour (err minutes)	-47	-20	37	89	98	90	FALSE
avg. max timing summer hour (err minutes)	17	46	57	127	81	21	FALSE
avg. top10 peak timing summer hour (err minutes)	11	30	41	68	68	-6	TRUE
avg. min timing winter hour (err minutes)	180	167	60	124	132	119	TRUE
avg. max timing winter hour (err minutes)	85	214	256	308	328	108	FALSE
avg. top10 peak timing winter hour (err minutes)	57	305	312	396	396	143	FALSE
avg. min timing shoulder hour (err minutes)	113	137	59	92	82	126	FALSE
avg. max timing shoulder hour (err minutes)	32	67	116	263	217	39	FALSE
avg. top10 peak timing shoulder hour (err minutes)	36	36	54	99	60	8	TRUE
annual electricity use kwh per sf err(%)	5.96	12.4	2.13	3.93	6.23	6.92	FALSE
cvrmse (%)	16.46	20.19	16.91	21.8	20.79	16.55	FALSE

Table 7.4 gives the calibration statistics for each run, showing the largest improvements came from updates to the lighting schedules. Lastly, Figure 7.14 shows the percentage error in daily minimum and maximum kW per ft^2 for each season, as well as the error in the timing of seasonal peaks, represented in minutes.

While the total load may show decent model fit, some building types are still far off from the AMI dataset, especially warehouses. This presents a risk of Simpsons paradox, where several poorly fitting building type comparisons could sum together to show a good model fit overall. Warehouses were the only building type to show a worse model fit after calibration however, and reversal of the changes

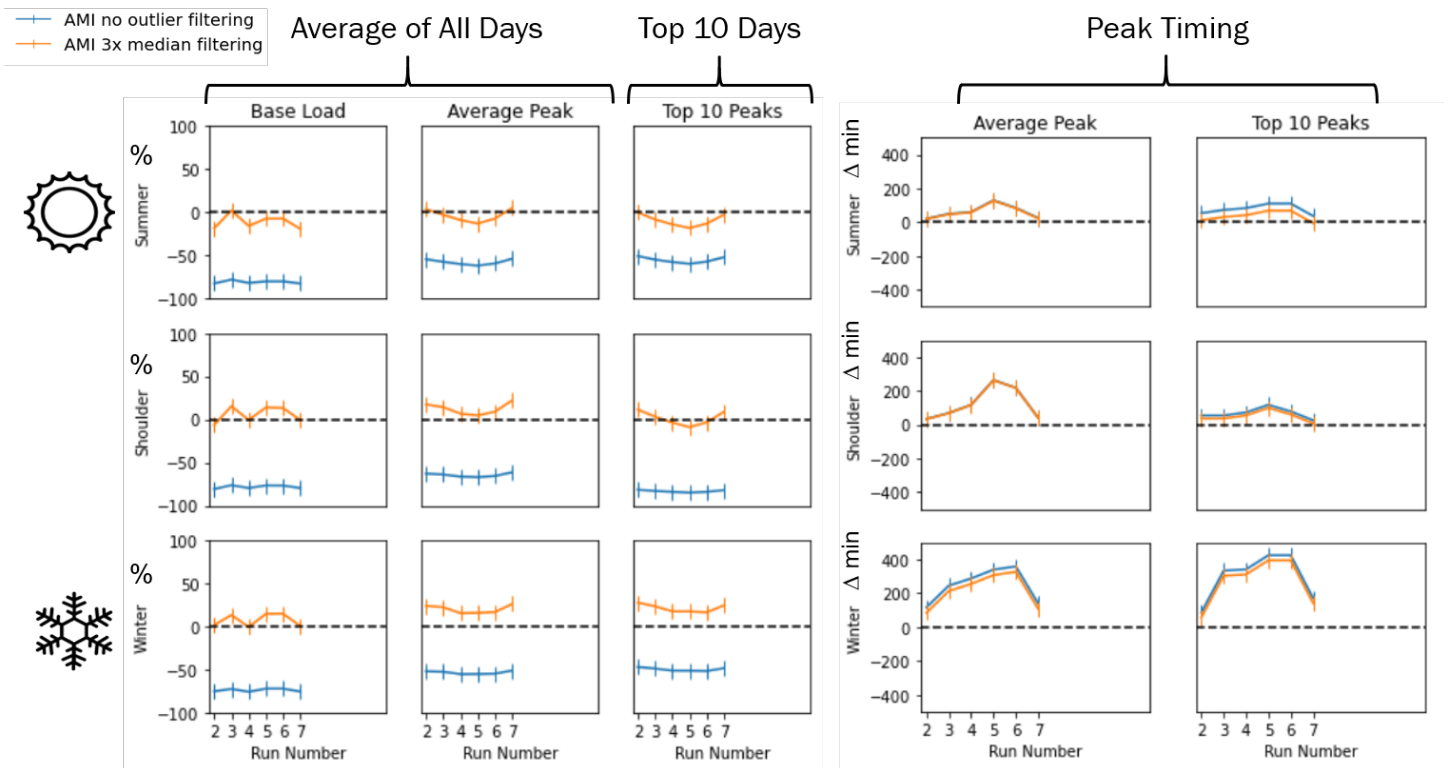

Figure 7.14: Calibration Quantities of Interest for run 6 compared against the AMI data with no outlier removal and the 3x median filter outlier removal method.

to warehouses would likely improve total model accuracy, suggesting there is not a Simpsons paradox here. As to why the warehouse model fit is worse, the enduse data used to generate updated warehouse equipment and lighting schedules included 26 buildings and the classification of these buildings as warehouses was determined by the data providers. The subset of warehouses from this dataset is likely not representative of warehouse buildings in general and contains different use cases than the warehouses in Fort Collins. It could also be the case that warehouse buildings in the AMI dataset include several buildings that act more like offices, such as auto body shops. In either case, there is a need for additional studies of the range of use cases for warehouses, their operating schedules, and typical equipment and lighting profiles. Retail, restaurant, office, and school buildings are easily classified and tend to have more homogeneous uses compared to warehouses. There were also more of them in the enduse datasets, so they did not suffer the same issues of being poorly representative of that building type in ComStock.

7.5 Calibration Conclusions

The model changes show little improvement to the total load fit but better performance for some specific building types, particularly retail, restaurants, school, and offices. The most significant changes came not from changes to ComStock, but rather establishing a filter to improve classification in the AMI data set. ComStock model adjustments changed total energy use by less than 20%, while AMI data changes were typically 10% to 15%, up to 60% depending on how extreme values were identified. This suggests that future calibration efforts should work to ensure proper building type classification in truth data sets and establish larger error bound for calibration. The final model has a 6.92% normalized mean bias error (NMBE) and 16.55% coefficient of variation of root mean square error (CVRMSE) based on normalized annual energy per floor area. Typical calibration statistics given in the literature review suggests a calibration goal of hourly CVRMSE less than 20 or 30% and NBME less than 5 or 10%. All runs meet the CVRMSE criteria, and several meet the NMBE criteria, however run 7 is above 5% NMBE, suggesting further room for model improvement. For the QOIs, seasonal maximum and minimum magnitudes were under 20%, except for peak winter magnitude at 258% error. Peak timing was within 2 hours except winter peak time which was 2.5 hours off. Investigation of the enduse plots provides some direction on future work to improve ComStock calibration.

7.5.1 Areas for Further ComStock Model Improvement

- **Code Adoption**

  Building efficiency levels in ComStock are currently set assuming the most recent ASHRAE Standard 90.1 [86] when the building was constructed, with updates to building systems following a retrofit frequency approach. In reality, the version of ASHRAE 90.1 is adopted on a state-by-state basis, with some states still still using 90.1-2004 and 90.1-2007 in 2020. This most significantly influences envelope performance in states that are slow to adopt new versions of 90.1.

- **RTU Efficiency Degradation**

  ComStock consistently underestimates peak energy use, typically occurring during peak summer cooling. Most cooling is served by rooftop DX cooling equipment. Currently, ComStock assumes the same performance characteristics for the HVAC equipment over its lifetime, when in reality there is minor performance drop each year from coil fouling and refrigerant leakage. An intern at NREL in the summer of 2020 analyzed performance data in DX RTU units and those assumptions for performance degradation will be incorporated into future versions of ComStock.

- **Exterior Lighting**

  Exterior lighting is a sizeable portion of nighttime load, and causes a bump in evening electricity use that is not seen in the AMI data. Exterior lighting is applied in ComStock as a mix of parking lot and facade lighting with allowances in 90.1 for each building type. As these are allowances, not installed lighting, there is likely room for improvement in estimating representative exterior amounts for each building type.

- **Warehouse End Use Data**

  The warehouse enduse data is likely not representative and serious classification issues remain in both the enduse and AMI datasets. Separating warehouse building types (differentiating between predominantly storage vs. sorting and distribution centers) and seeking additional enduse data is likely necessary to improve ComStock accuracy.

- **Restaurant Kitchen Equipment**

  Restaurant kitchen equipment, including hot water use from commercial dishwashers, is the most significant end use in full service restaurant buildings. Likewise kitchen equipment is the most significant use in quick service restaurants, where energy use is not directly tied to floor area. ComStock overestimates late evening electric equipment use in full service restaurants and underestimates electric equipment use in quick service restaurants, suggesting a need to break out major electric equipment explicitly in the model and separating energy use for kitchen and dining areas.

- **Variability**

  ComStock current uses the same averaged lighting and equipment schedule for each building, adjusted to building start times and duration. In reality, there are several distinct schedule patterns for equipment and lighting use within a given building type, and these can be incorporated to show proper end use distribution between buildings of the same type.

Calibration of ComStock is an ongoing project. This chapter covers calibration for the first region of ComStock, and is a foundation to for improving calibration in other regions. The next region will focus on a hot, humid climate in the southeastern U.S.. ComStock region 1 metrics will continued to be tracked with future model improvements.

Chapter 8: Results and Conclusions

8.1 Conclusions

This body of work sought to improve the foundation on which energy models are generated and calibrated for the purpose of building energy retrofits. An analysis of a case study building found that for energy retrofits typically executed through energy savings performance contracts, measure costs and capital availability are the most significant constraints on long term financial performance and greenhouse gas emission reduction potential. Greenhouse gas prices, colloquially referred to as the social cost of carbon, do not influence decisions until they reach several hundreds of dollars per metric ton, which is far beyond the range considered in current policy making. Furthermore, there was little difference in measure ordering - the idea of placing load reduction measures first to downsize HVAC equipment - and that this savings was well within the noise of typical measure performance uncertainty. This suggests that energy savings measures should be targeted with most cost effective and reliable savings first, and that lower measure costs and access to cheap capital will be necessary to retrofit the building stock. Currently there are two ways to access cheap capital: through direct government investment and in the private sector by greatly reducing the financial risk involved in energy savings projects. Reducing risk means reducing energy savings uncertainty, which is where energy model calibration is needed, especially for projects that will address multiple building systems at once.

To reduce calibration uncertainty, this work programmatically generated calibrated end-use models of the commercial building stock to reduce parameter uncertainty and serve as a starting point for calibration efforts for individual buildings. As the current metrics for demonstrating annual hourly calibration (NMBE and CVRMSE) are insufficient to properly calibrate several enduses, this work developed a set of Quantities of Interest to calibrate buildings to an hourly basis for each season. A random forest regression model is able to accurately determine which model parameters matter most for each QOI, which allows energy modelers to focus on improving model fit specific to each building type, season, and climate zone. Only a subset of the 66 model features are important; when taken at the stock level across all U.S. climate zones, 25 parameters explain 90% of the stock energy use and are critical for model calibration. These include direct, causal parameters such as lighting and equipment schedules, as well as a few indirect parameters such as hot water use and gas heating coil efficiency which are proxies for electric kitchen equipment and HVAC system age respectively.

These important model parameters were determined for a stock model of Fort Collins, CO and used to calibrate the stock model to utility AMI data. Building classification, that is choosing how to best represent existing buildings with meter data as a specific building type, is the largest source of calibration uncertainty. For Fort Collins, CO, classification uncertainty caused changes from 10% to 60%

depending on how many extreme misclassified buildings were removed, while the energy model changes were all less than 20%. A 3x median filter approach can best help identify outlier and misclassified buildings in the AMI data to provide a proper point of comparison for calibrating the stock model. The final stock model has a 6.92% NMBE and 16.55% CVRMSE based on normalized annual energy per floor area, and QOIs show decent agreement with the exception of winter peak magnitude and timing. Analysis of enduses during times of large model disagreement suggests several areas for ComStock improvement.

8.2 Contributions

To enable use by the building energy modeling community, the software developed as part of this dissertation is open source and publicly available on Github.

The details of the retrofit methodology to analyze sources of external uncertainty and findings were shared in a primary author paper, *Building Energy Retrofits Under Capital Constraints and Greenhouse Gas Pricing Scenarios* [49], and available at (`https://github.com/CITY-at-UMD/retrofitLCC`). This methodology does not apply to deep energy retrofits, which typically involve a complete gut retrofit of a building, replacing the exterior facade, installing a new HVAC system, and often conversion to an all-electric system. These projects are relatively rare and are more appropriately considered as new construction.

The method for generating reduced-order models was pursued through two paths, initially through Virtual Pulse, described in a co-authored paper *Demonstra-*

tion of reduced-order urban scale building energy models [106], and then later with much more rigorous treatment through development of the openstudio-standards ruby gem [26], available at (`https://github.com/NREL/openstudio-standards`). Though not detailed here, this work required substantial development, debugging, and validation of building energy models generated through openstudio-standards. These methods are then accessed through the *Create Typical Building Measure*, available on the OpenStudio Building Component Library (`https://bcl.nrel.gov/`) and summarized in a co-authored conference presentation *Automatic Generation of Highly Customizable Energy Models from High Level Input Data* [107]. This work has been used extensively by a range of projects, including a co-authored report *ENERGY STAR for Tenants: An Online Energy Estimation Tool for Commercial Office Building Tenants* [103], an analysis of load forecasts for the Los Angeles Department of Water and Power, the forthcoming ASHRAE Net Zero Energy Multifamily Design Guide [108] [109], and several commercial building energy modeling software interfaces. ComStock is available at (`https://github.com/NREL/ComStock`) and a ComStock output web viewer is available on NREL's website. Hopefully building energy modeling practitioners will find this software useful in improving the built environment.

A forthcoming article will share the QOIs and how to apply them to building energy model calibration. If there is one major take-away from this dissertation, it should be to include seasonal daily load profile comparisons, ideally broken out by end use, to assist with building energy model calibration. The feature importance results from ComStock can be atomized to the climate zone, building type, and season level, allowing building modeling practitioners to focus on the input parameters most likely to influence the energy savings for their particular retrofit project.

Lastly, the 3x median filter can be applied to other AMI datasets where direct address matching is not available to provide reasonable truth data for comparison when calibrating stock models. This filter method will be applied to assist in calibrating future regions of ComStock. The identified areas of model disagreement provide a detailed list of further enduse data needs and areas for future research to accurately represent the building stock.

8.3 Future Work

This dissertation laid the groundwork for calibrating stock energy models and prioritized data collection efforts. The most immediate next steps are to continue calibration efforts for additional regions, adding in model improvements summarized in Chapter 7. Most important of these is re-assessing building classification to improve truth data, potentially adding additional building types and space type distributions within a given building type, as well as including additional schedule and load variability. Once the stock model has been calibrated to a reasonable level, it enables a wide range of future work in the building energy model space.

First, utilities and research and development agencies, like NREL and the DOE Building Technologies Office, can use ComStock to prioritize building technologies for further investment. ComStock is currently being used by the Los Angeles Department of Water and Power to prioritize research into new building technologies that will help the utility reach its goal of 100% renewable energy. This summer 2020, the author supervised an intern that analyzed the national potential for behind-the-meter thermal energy storage (BTMS), identifying retail building RTUs as the greatest market for BTMS and determining storage rate and capacity requirements. Coupling ComStock with a cost model will allow building load to be explicitly expressed in NREL's dispatch models for the electric grid, allowing building technologies to compete directly as grid services. Utility load resource planning currently treats building energy efficiency as a small, static, uniform reduction in load. Explicitly breaking out commercial building load in grid models and being able to predict reliable reductions in load through efficiency will open up inexpensive grid infrastructure capital to building energy efficiency projects, removing a significant limitation to building energy retrofits as demonstrated in Chapter 4.

The second major use case is to use an ensemble of building energy models from ComStock as a starting point for individual building model calibration and savings estimation. As noted in Chapter 2, calibration efforts are hampered by model overfitting and the lack of prior parameter distributions for model inputs. The range of climate zones, building types, HVAC systems and other major factors can greatly influence which parameters are significant for a given building, and using a subset of models generated from ComStock down-selected to specific building characteristics will provide more accuracy than generalized case studies or practitioner inference. This is critical for the success of Bayesian calibration approaches, which while promising, are dependent on accurate prior distributions [62]. For smaller commercial buildings, where having a practitioner perform energy auditing, data collection, and model calibration may be prohibitively expensive given the available energy savings, an ensemble modeling approach may provide an alternative means to estimate efficiency savings with much better accuracy than current deemed savings approaches.

Hopefully better energy savings estimation from building energy model calibration will create a positive feedback loop by encouraging greater adoption of building energy efficiency leading to further investment in efforts to reduce energy savings uncertainty.

Appendix A: Feature Importance for ComStock run 6, Fort Collins, CO

Feature importance for select QOIs and QOI sets for ComStock run 6, Fort Collins, CO.

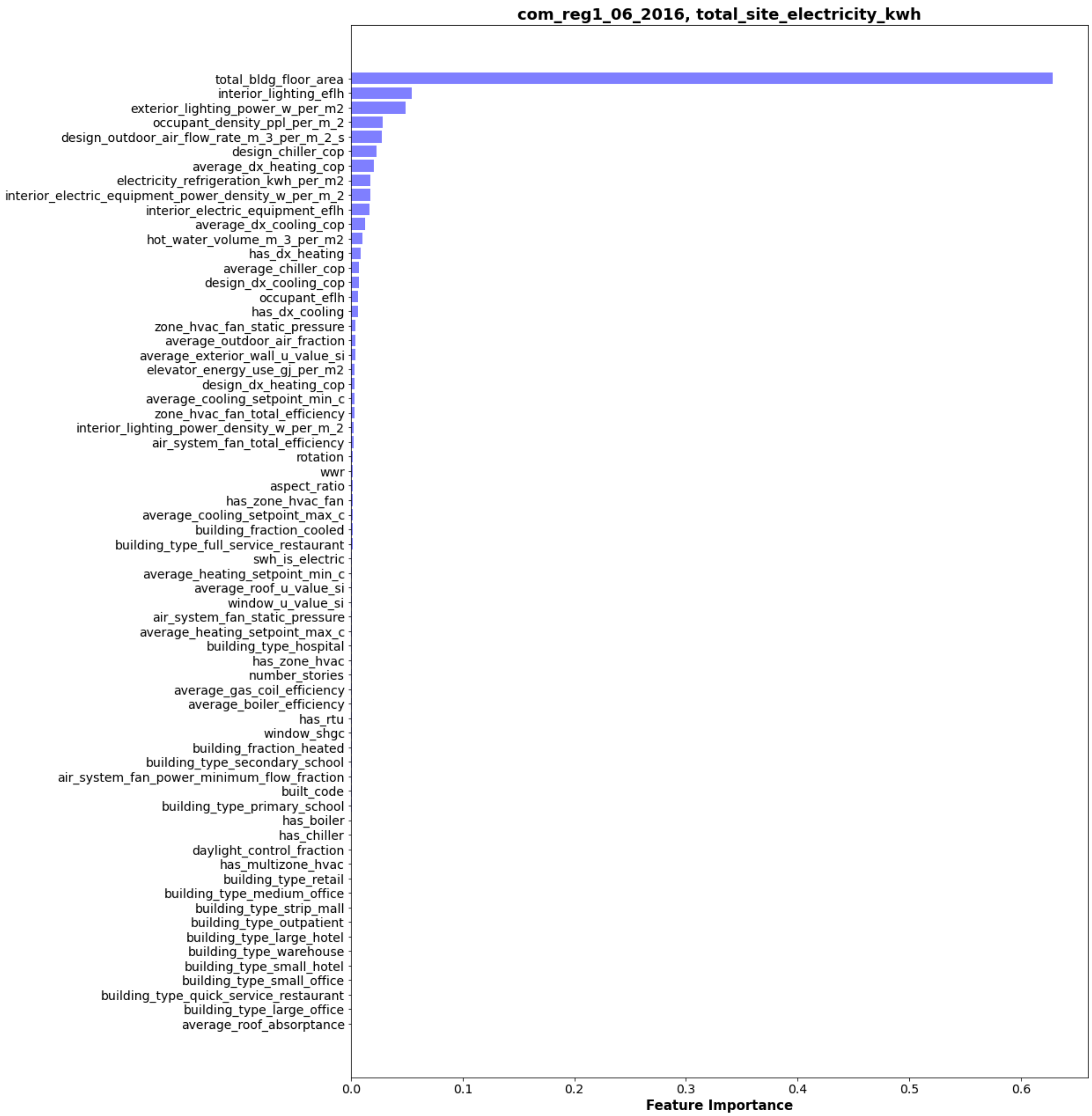

Figure A.1: The ranked feature importance for total site electricity for buildings in relation to the total commercial building load for ComStock run 6.

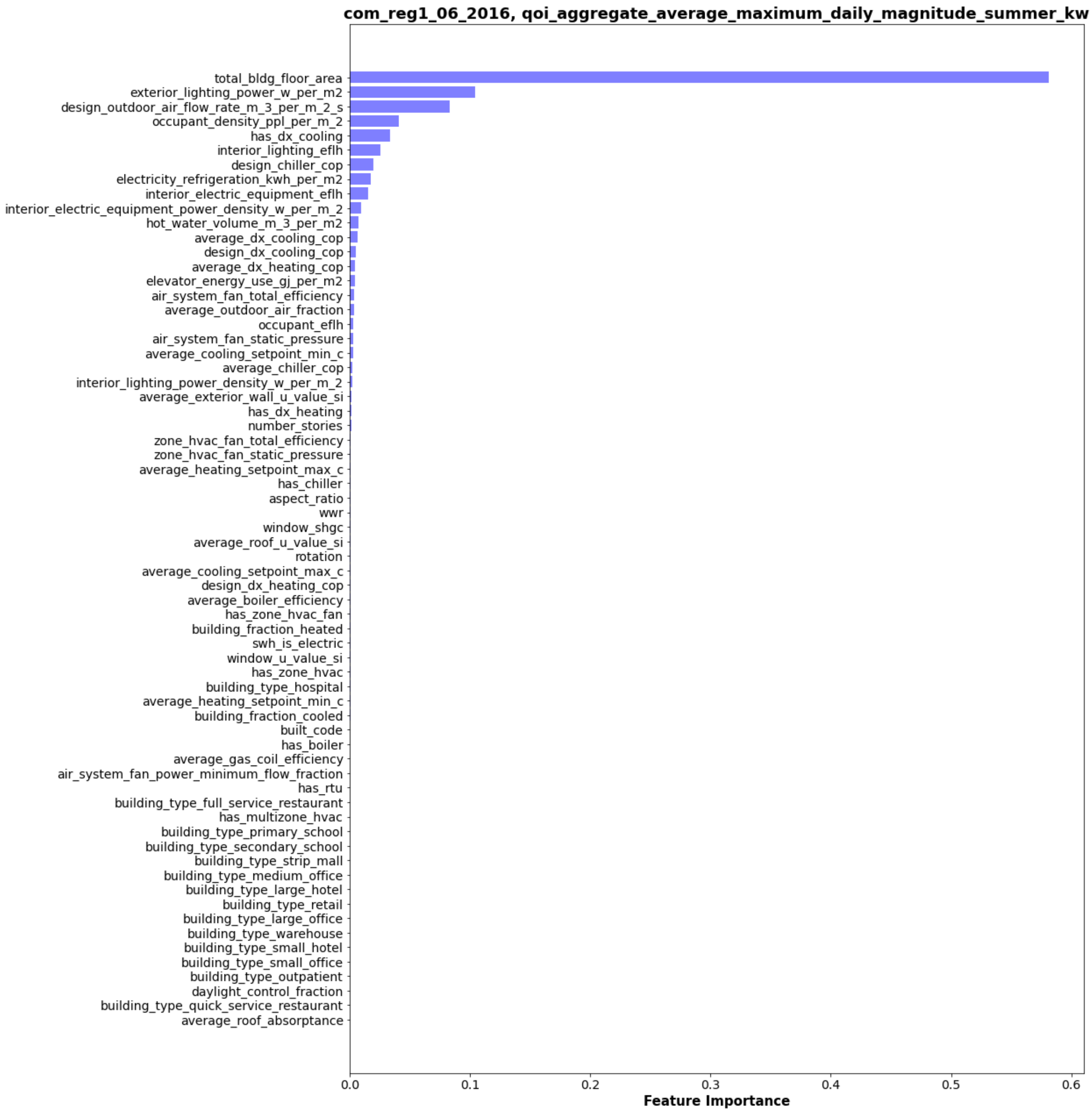

Figure A.2: The ranked feature importance for maximum summer demand for buildings in relation to the total commercial building load for ComStock run 6.

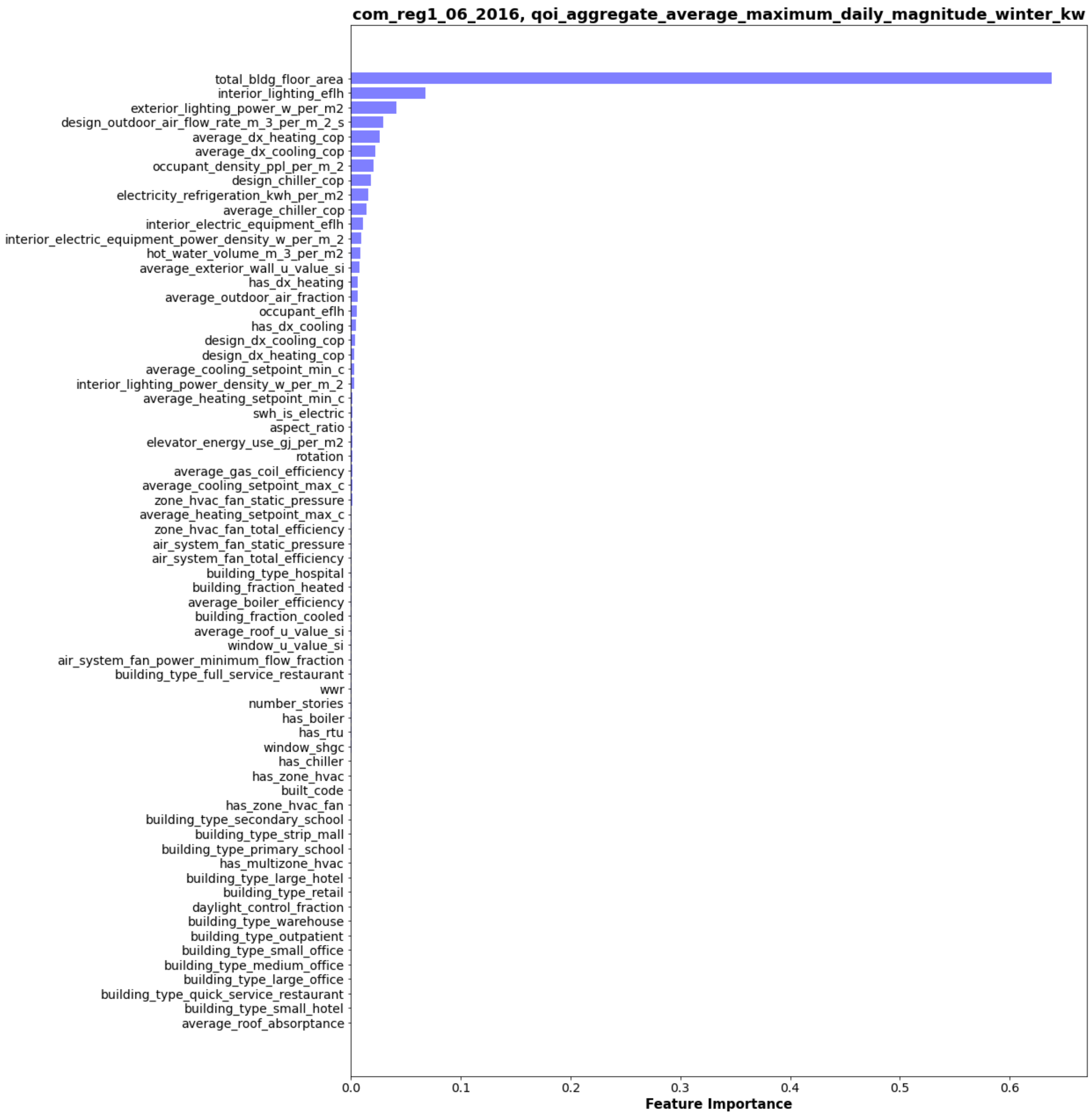

Figure A.3: The ranked feature importance for maximum winter demand for buildings in relation to the total commercial building load for ComStock run 6.

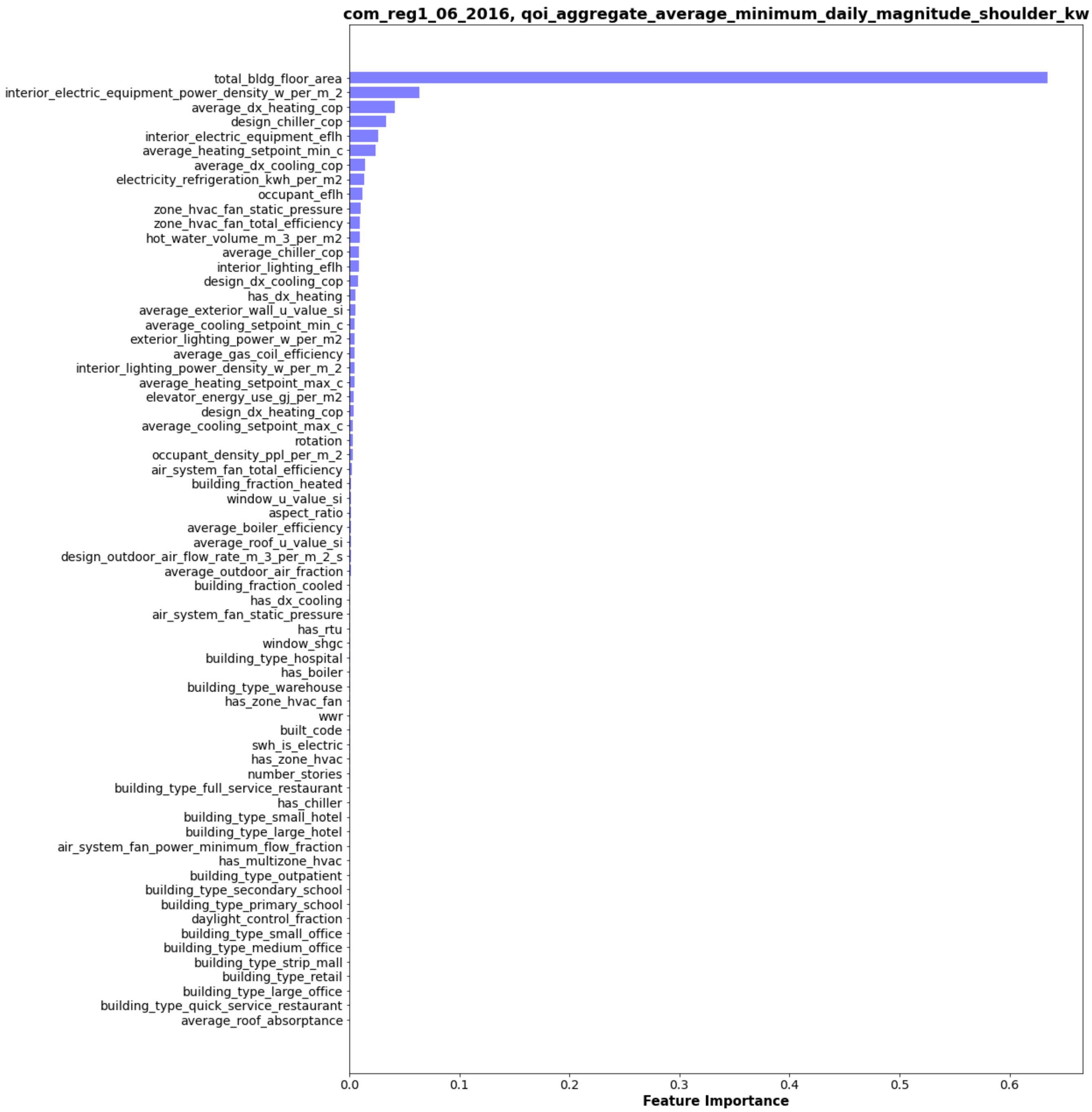

Figure A.4: The ranked feature importance for minimum shoulder demand for buildings in relation to the total commercial building load for ComStock run 6.

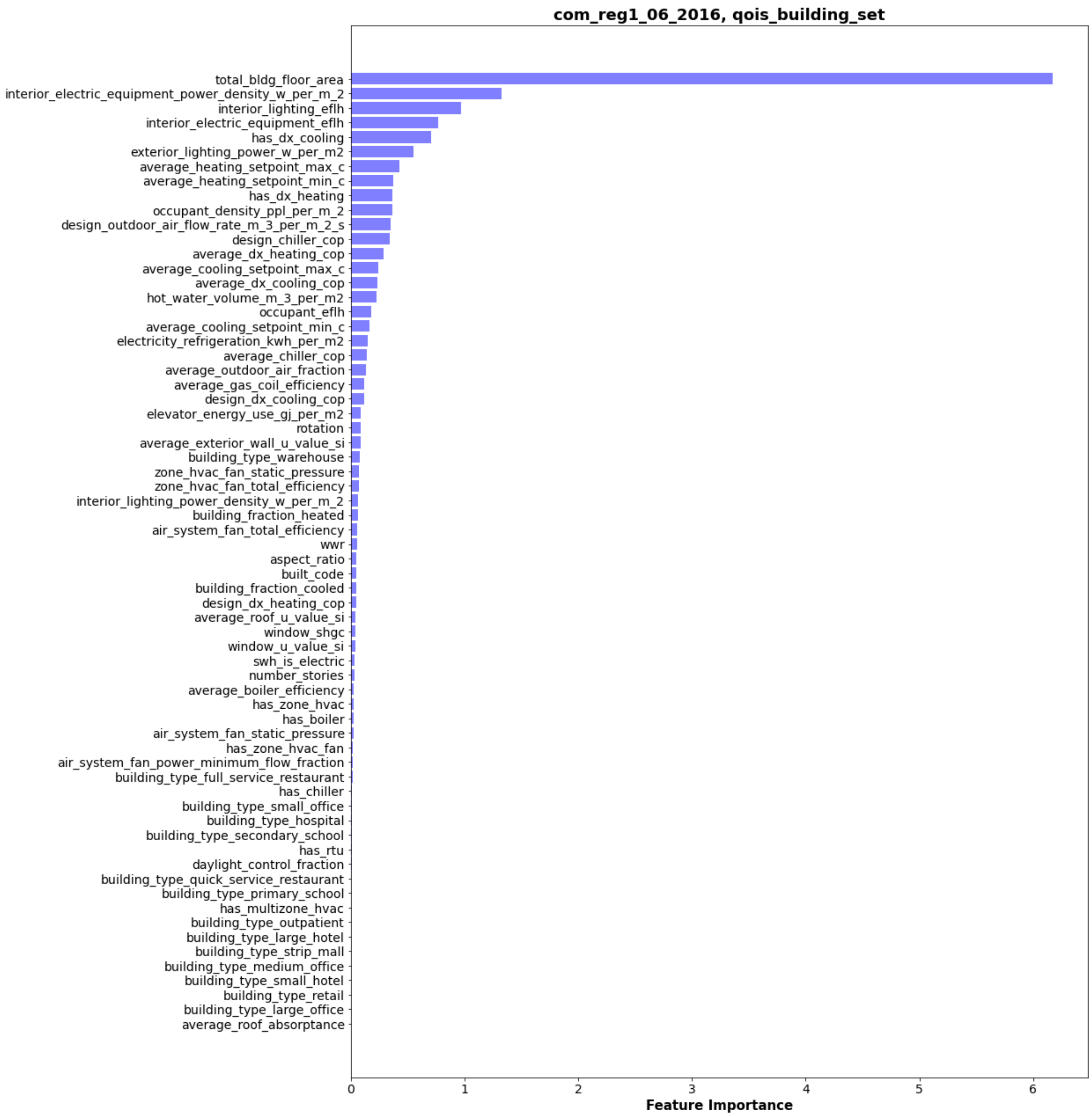

Figure A.5: The ranked feature importance for individual buildings for ComStock run 6.

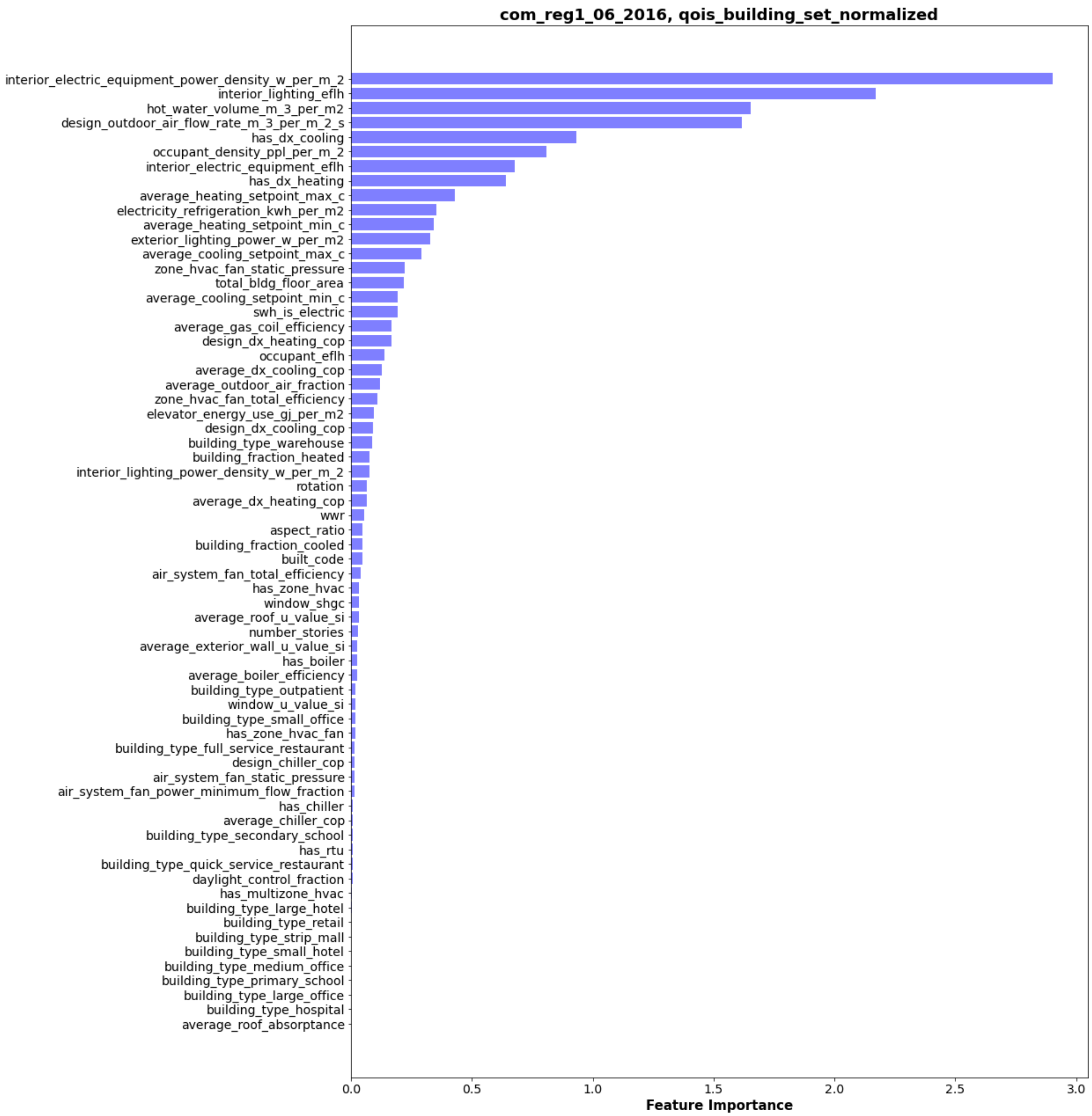

Figure A.6: The ranked feature importance for individual buildings normalized by floor area for ComStock run 6.

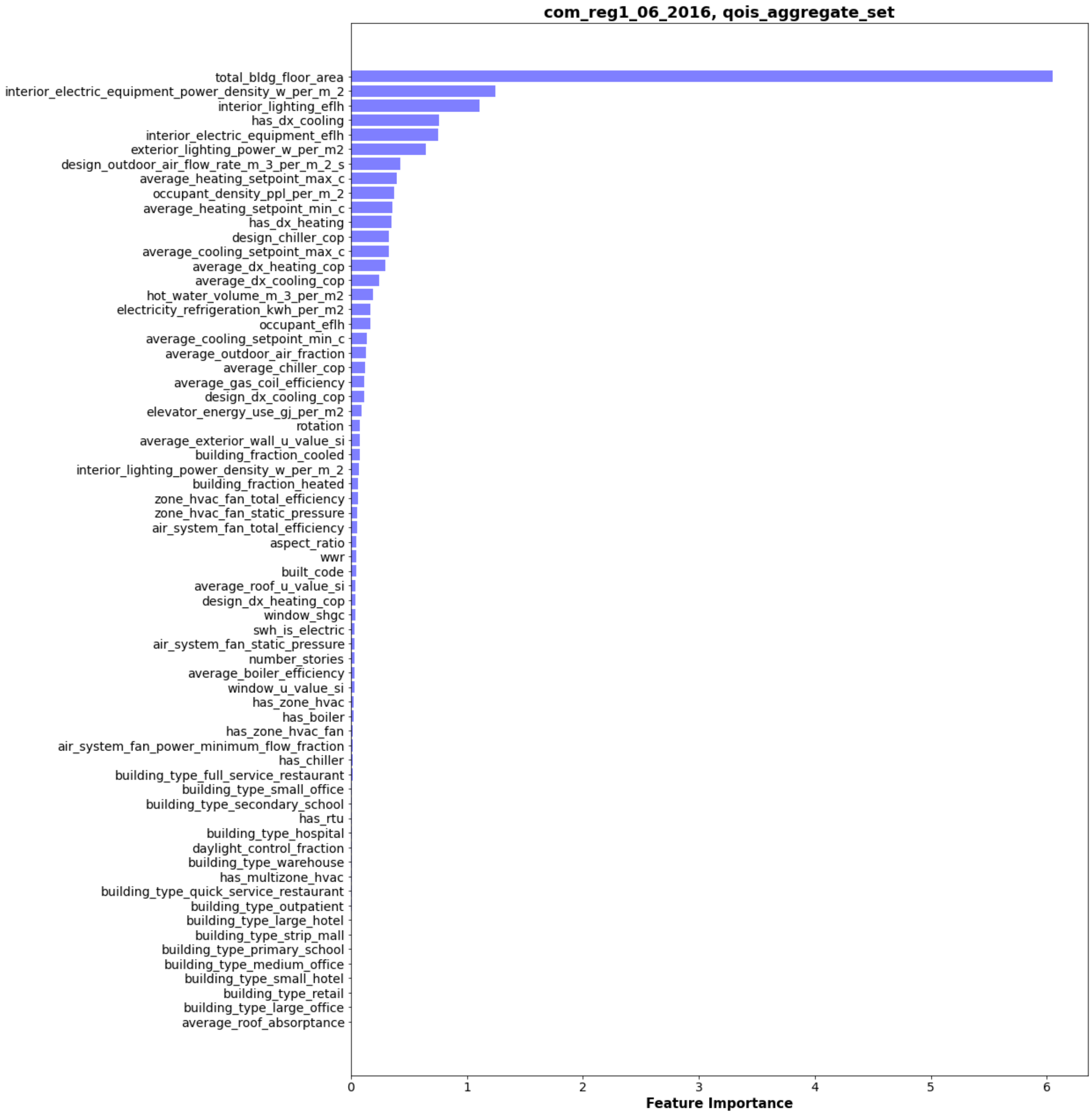

Figure A.7: The ranked feature importance for buildings in relation to the total commercial building load for ComStock run 6.

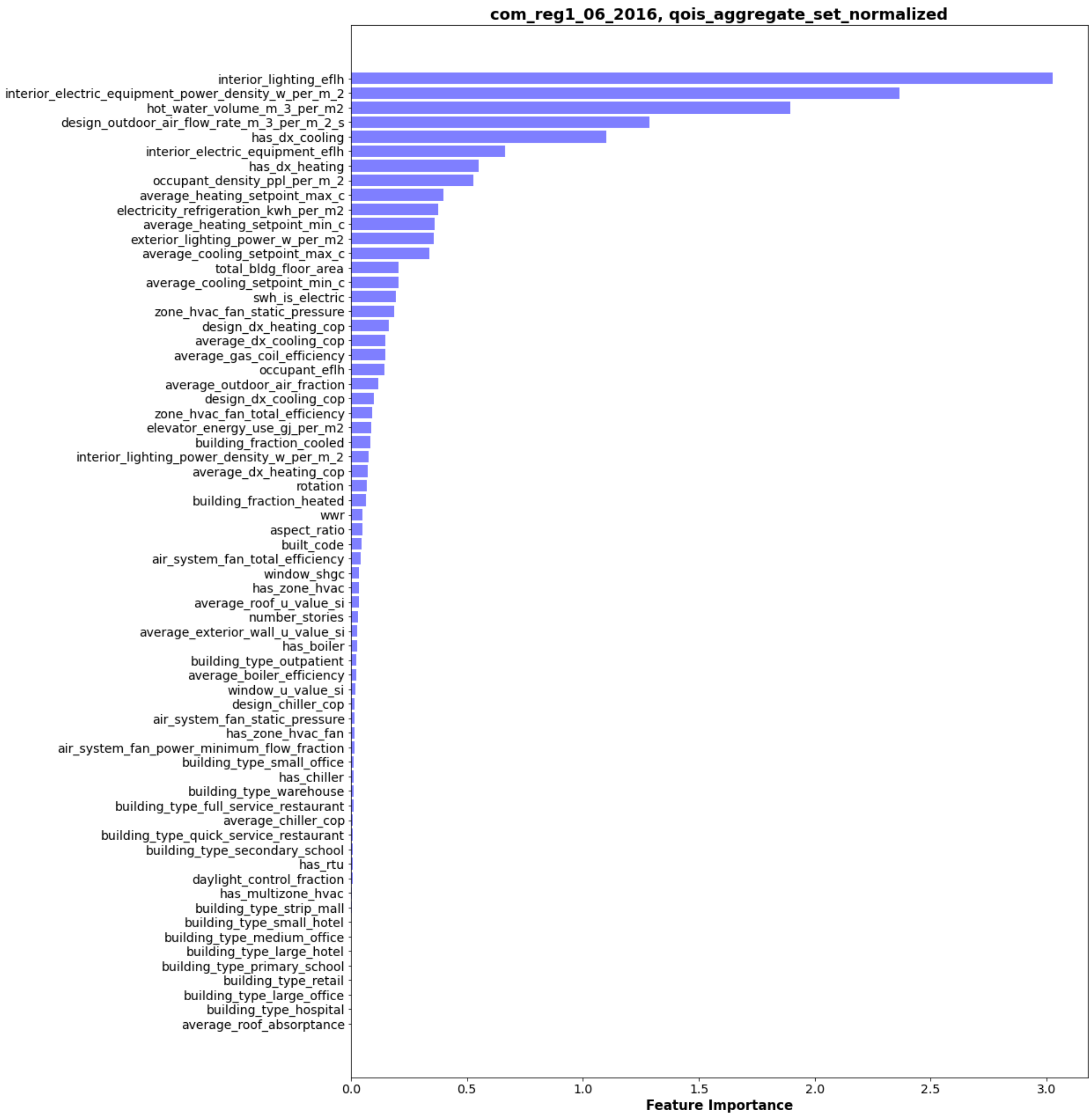

Figure A.8: The ranked feature importance for buildings in relation to the total commercial building load normalized by floor area for ComStock run 6.

Appendix B: Feature Importance for ComStock, entire U.S.

Feature importance for individual QOIs and QOI sets in individual climate zones are available upon request.

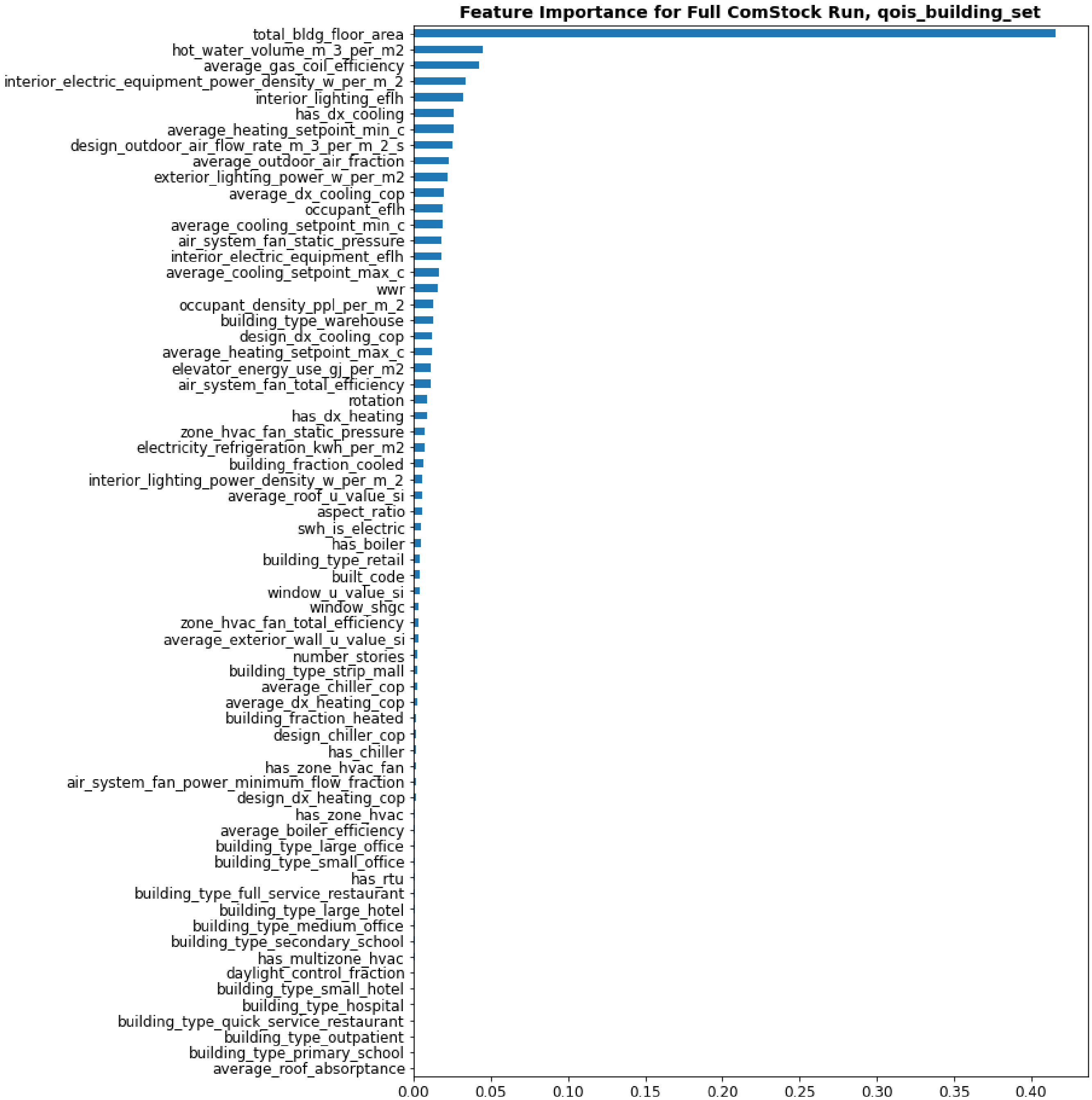

Figure B.1: The ranked feature importance for individual buildings for ComStock nationally.

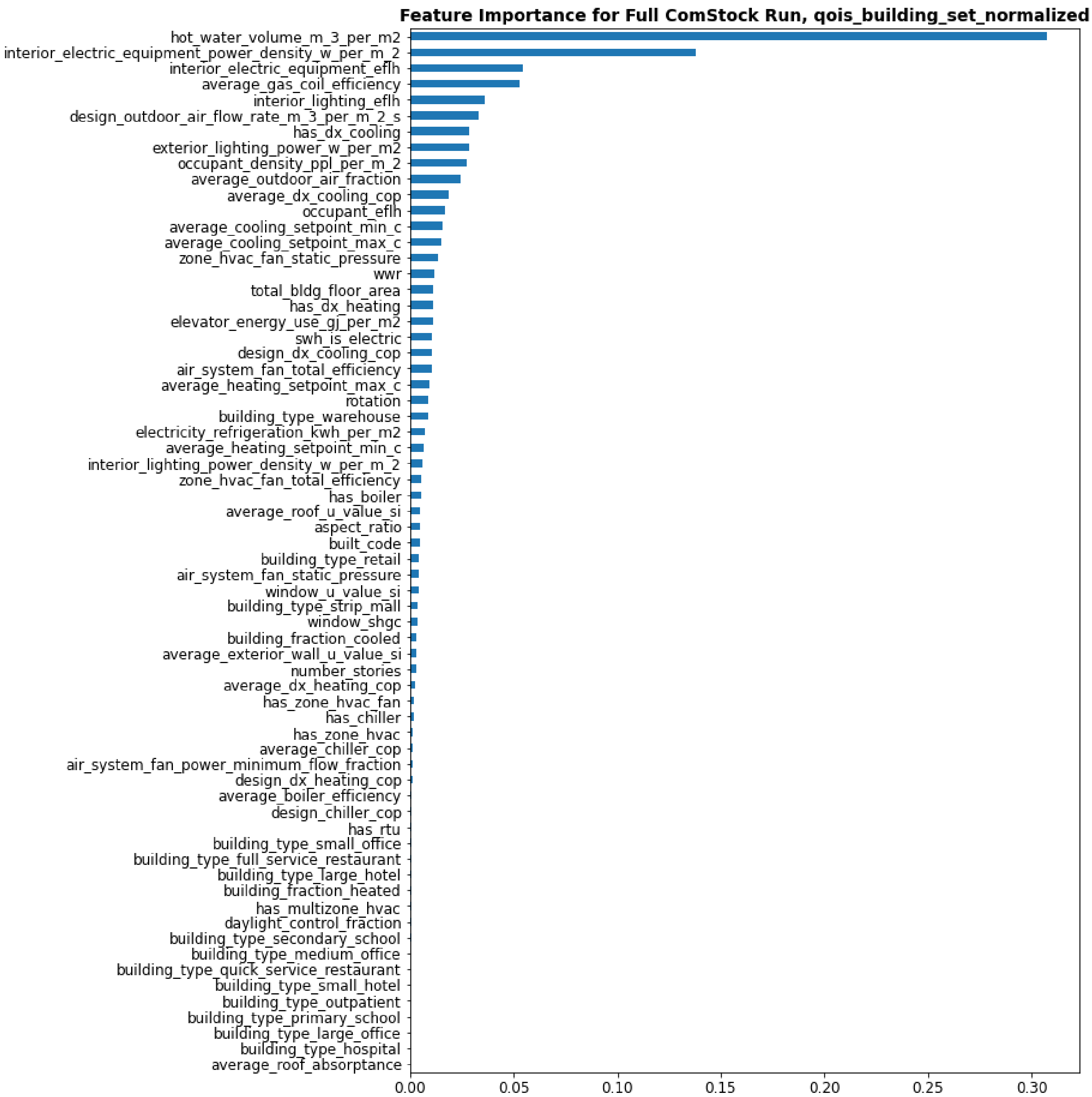

Figure B.2: The ranked feature importance for individual buildings normalized by floor area for ComStock nationally.

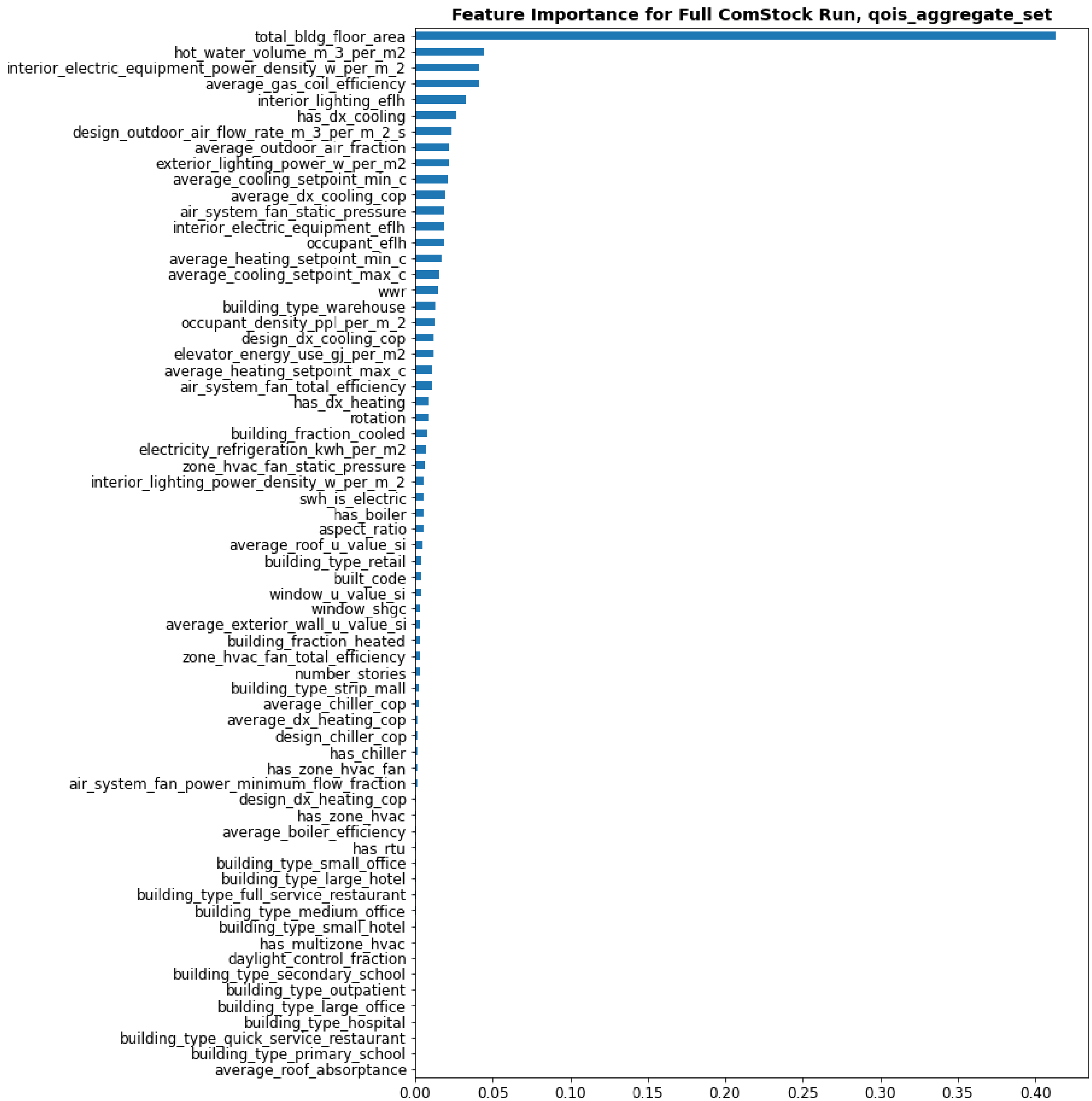

Figure B.3: The ranked feature importance for buildings in relation to the total commercial building load for ComStock nationally.

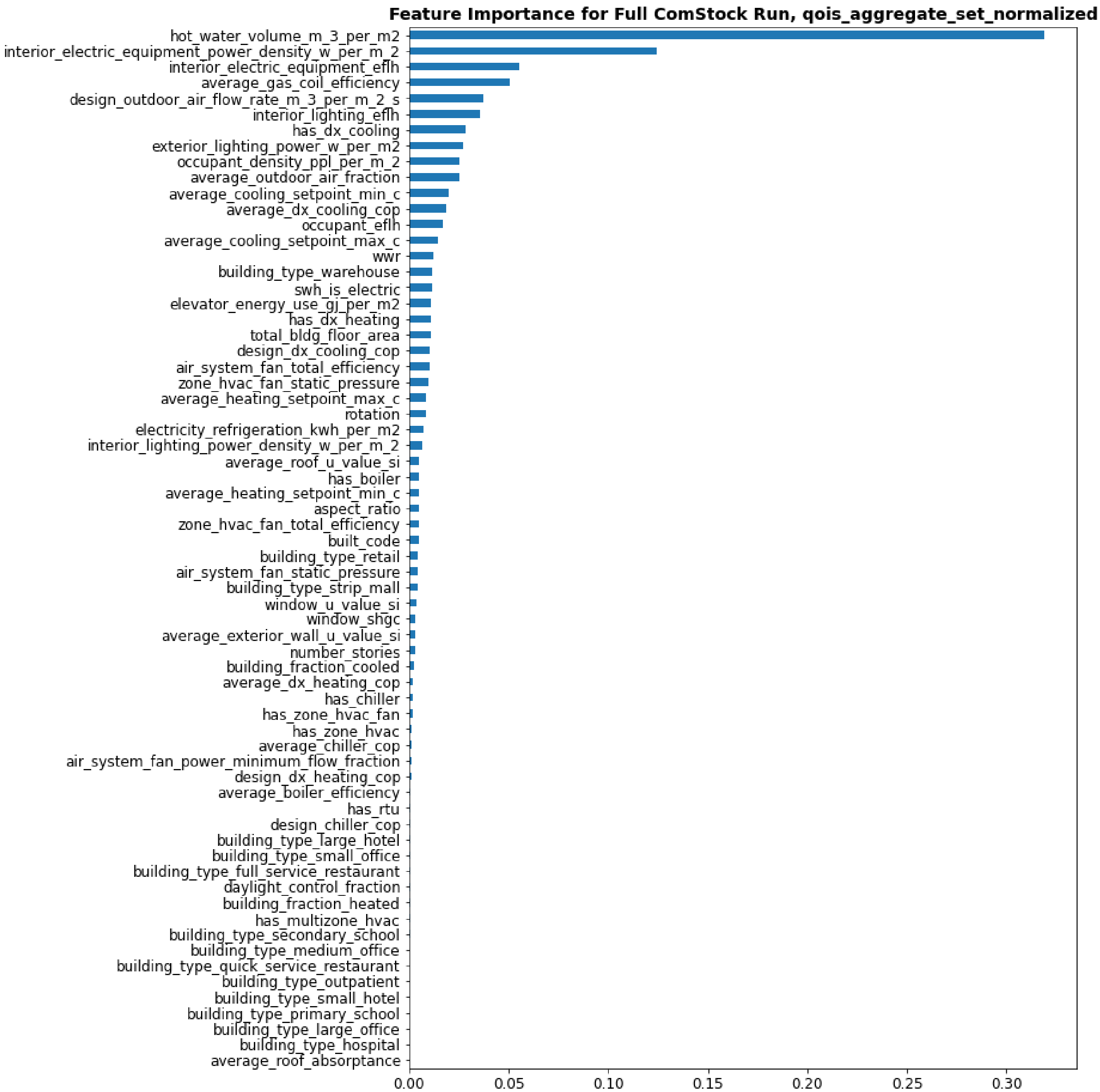

Figure B.4: The ranked feature importance for buildings in relation to the total commercial building load normalized by floor area for ComStock nationally.

Appendix C: Building Type Calibration Results for ComStock run 7, Fort Collins, CO

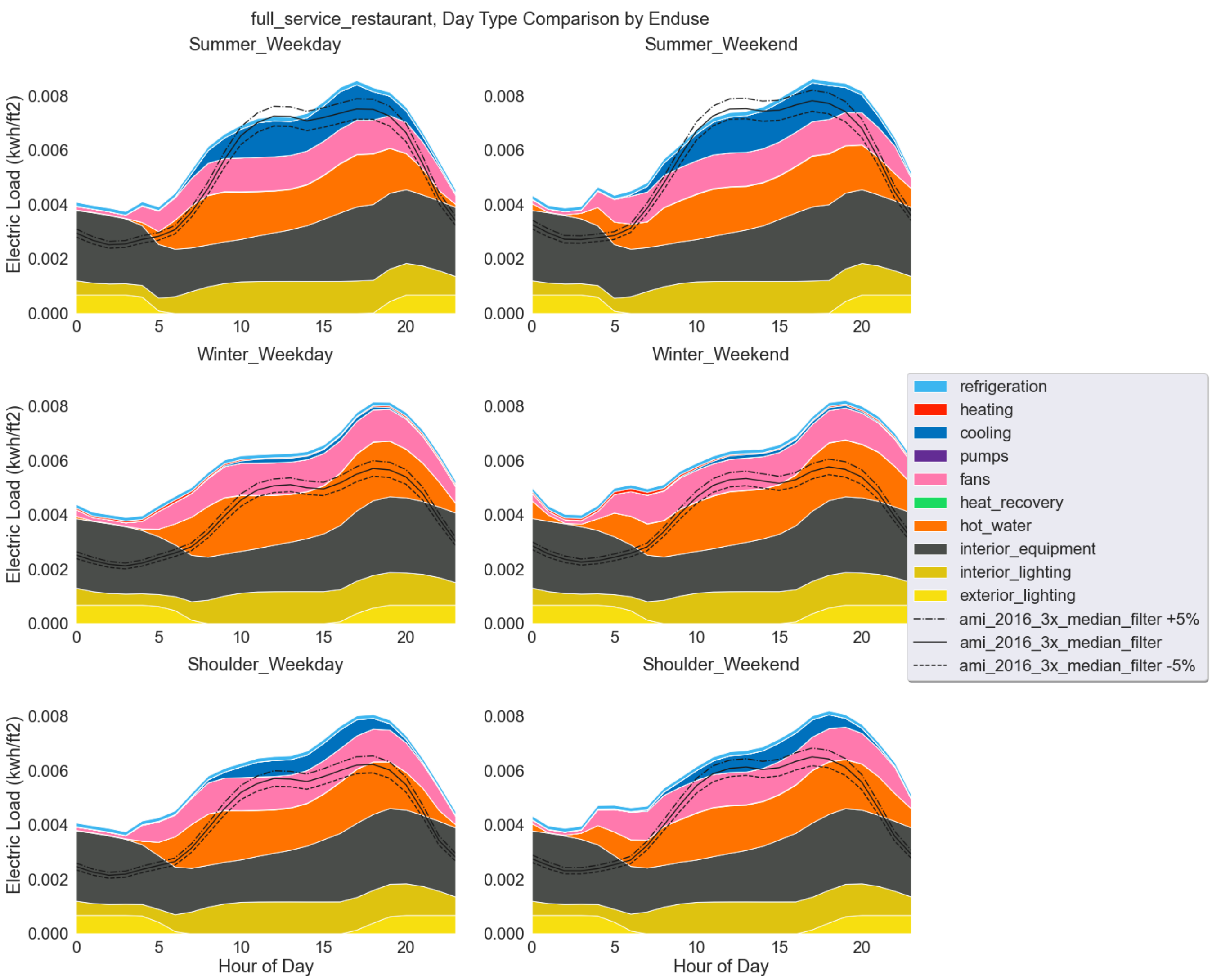

Figure C.1: Stacked area enduse plots by season and day type for full service restaurants, ComStock run 7, Fort Collins, CO.

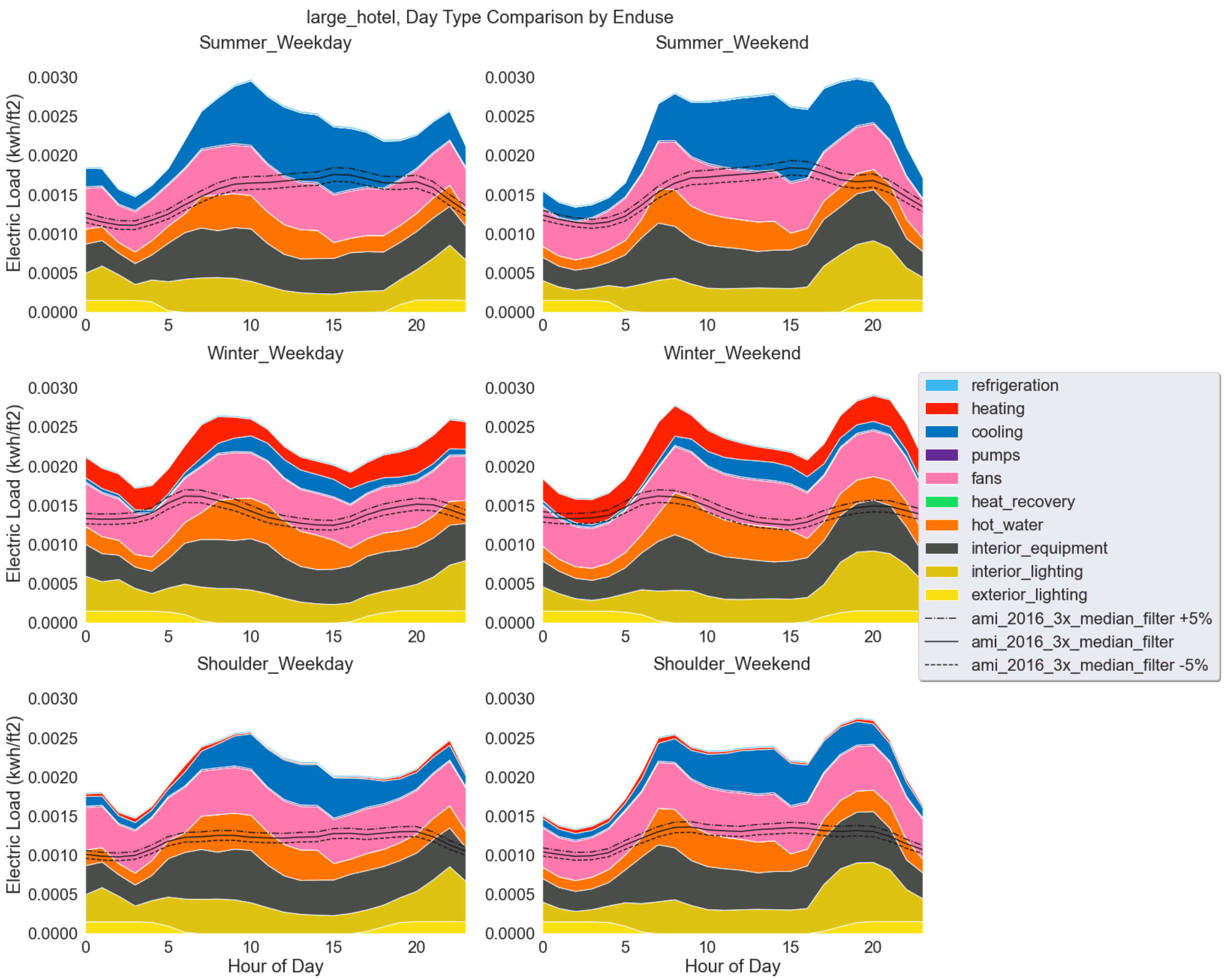

Figure C.2: Stacked area enduse plots by season and day type for large hotels, ComStock run 7, Fort Collins, CO.

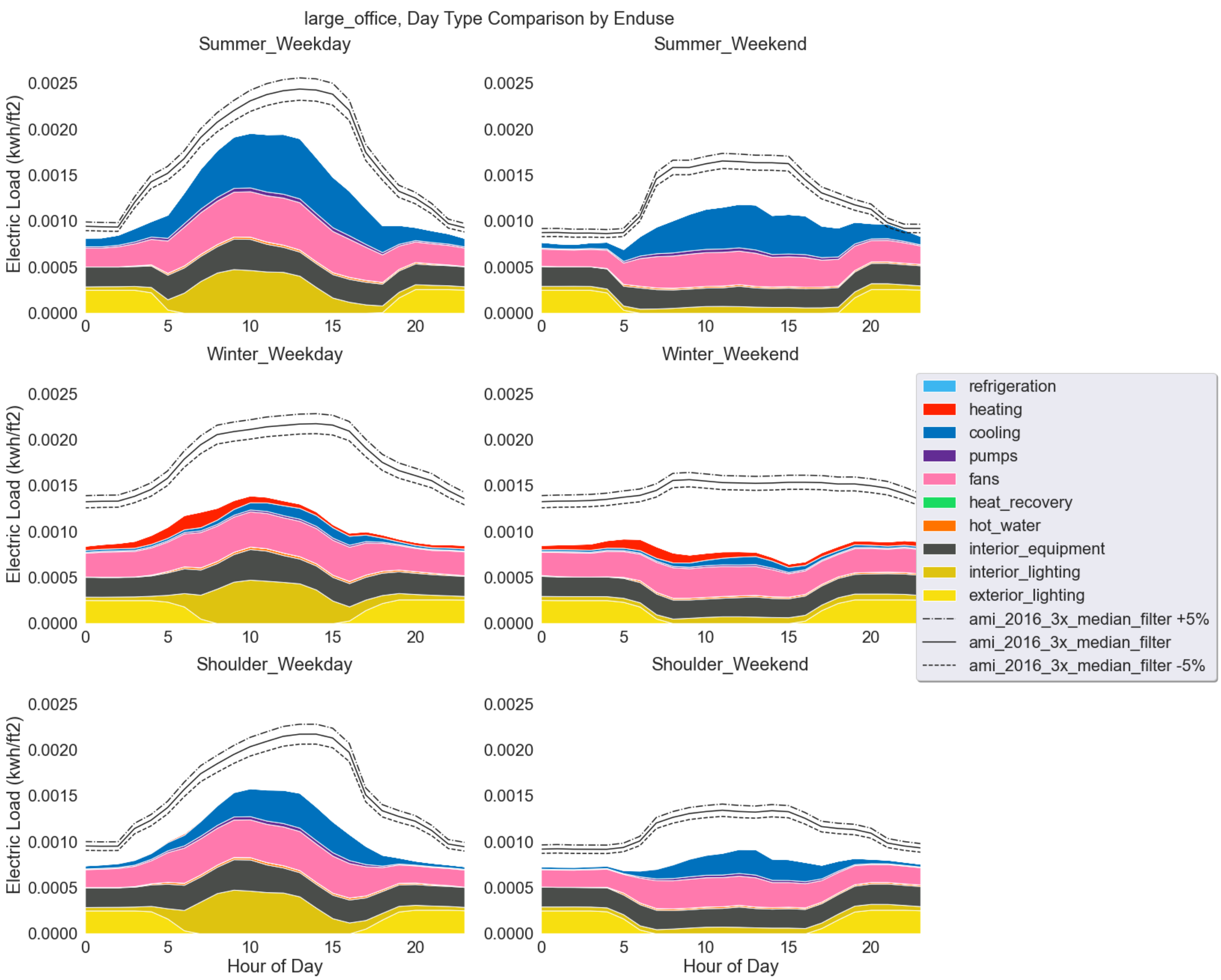

Figure C.3: Stacked area enduse plots by season and day type for large offices, ComStock run 7, Fort Collins, CO.

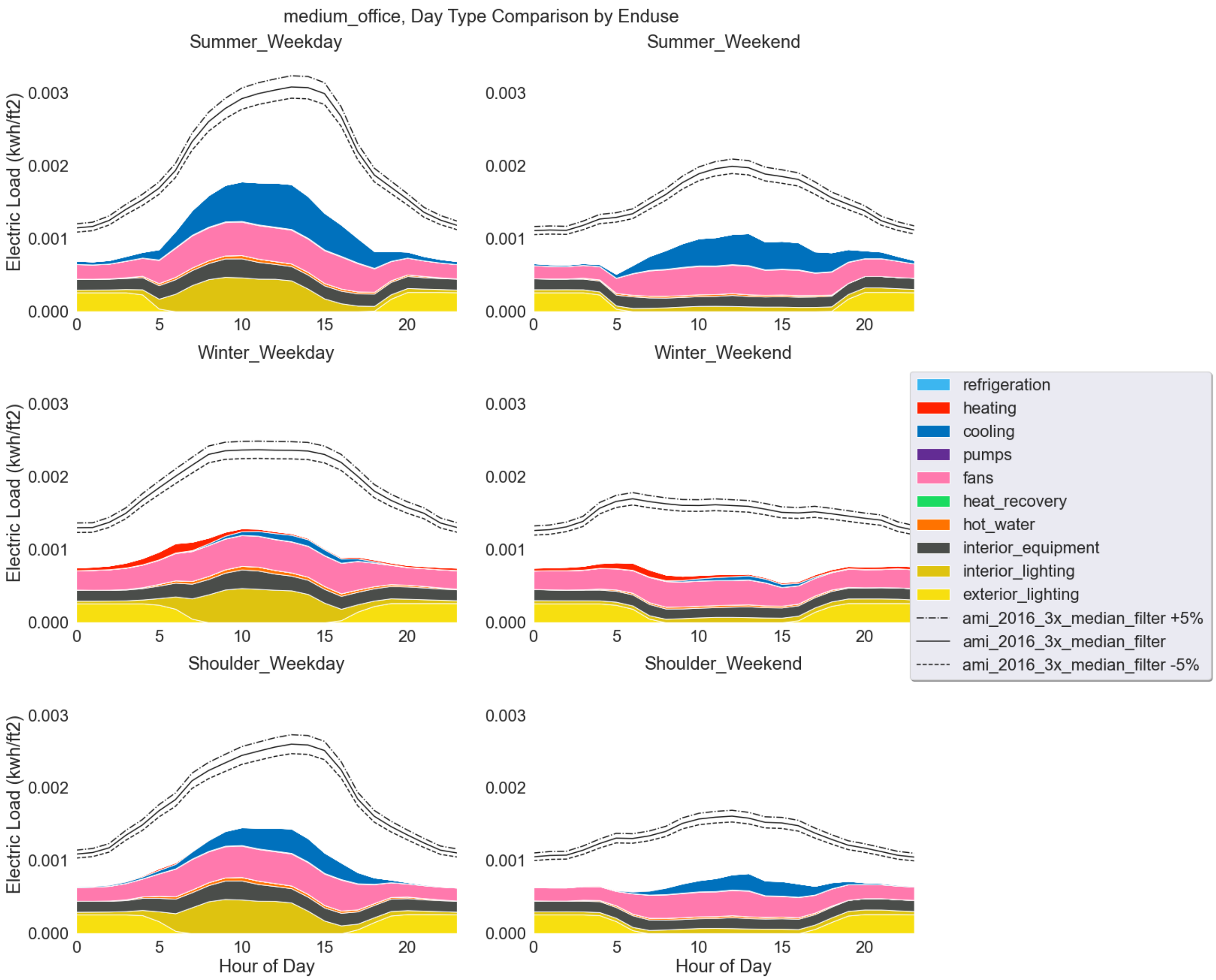

Figure C.4: Stacked area enduse plots by season and day type for medium offices, ComStock run 7, Fort Collins, CO.

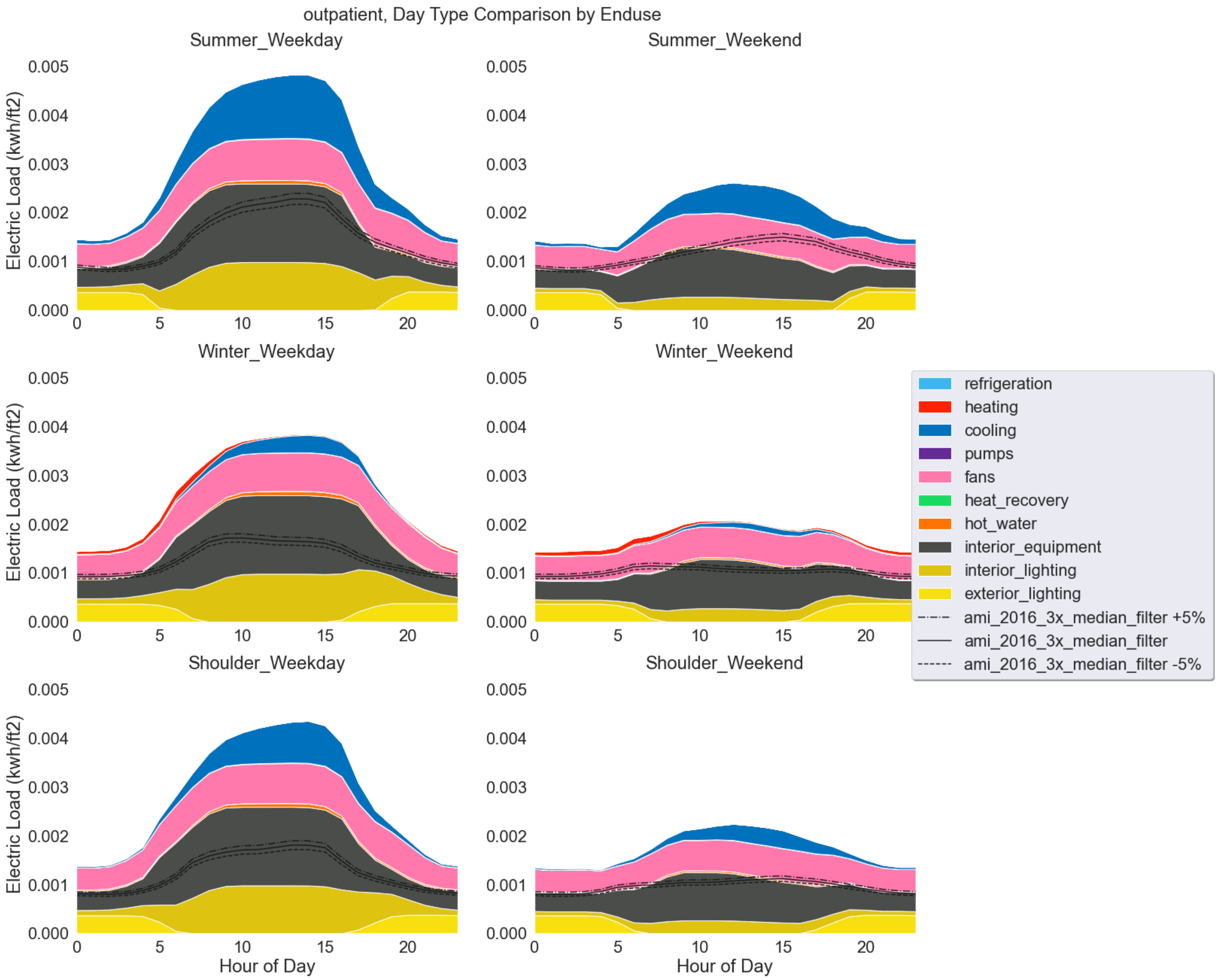

Figure C.5: Stacked area enduse plots by season and day type for outpatient, ComStock run 7, Fort Collins, CO.

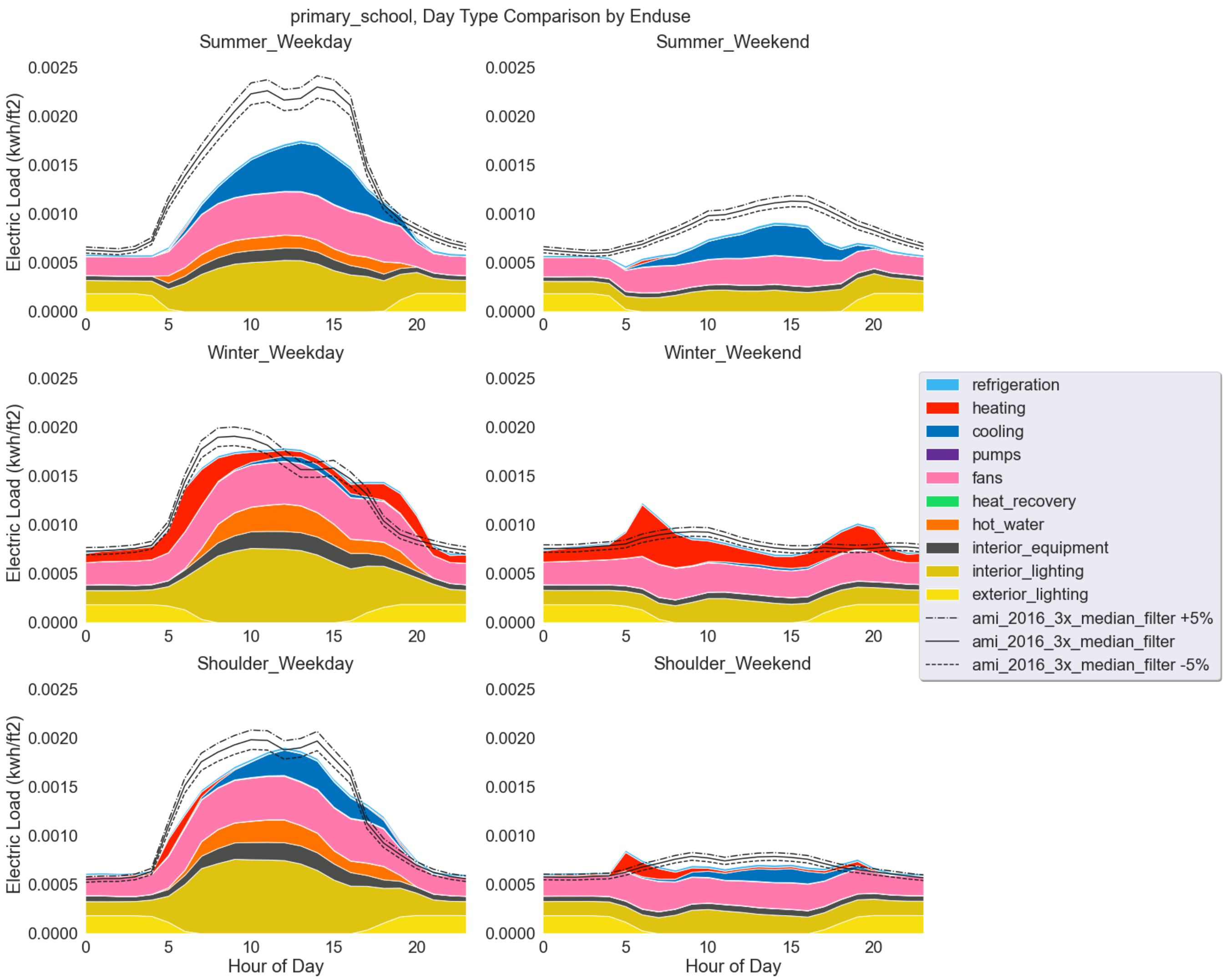

Figure C.6: Stacked area enduse plots by season and day type for primary schools, ComStock run 7, Fort Collins, CO.

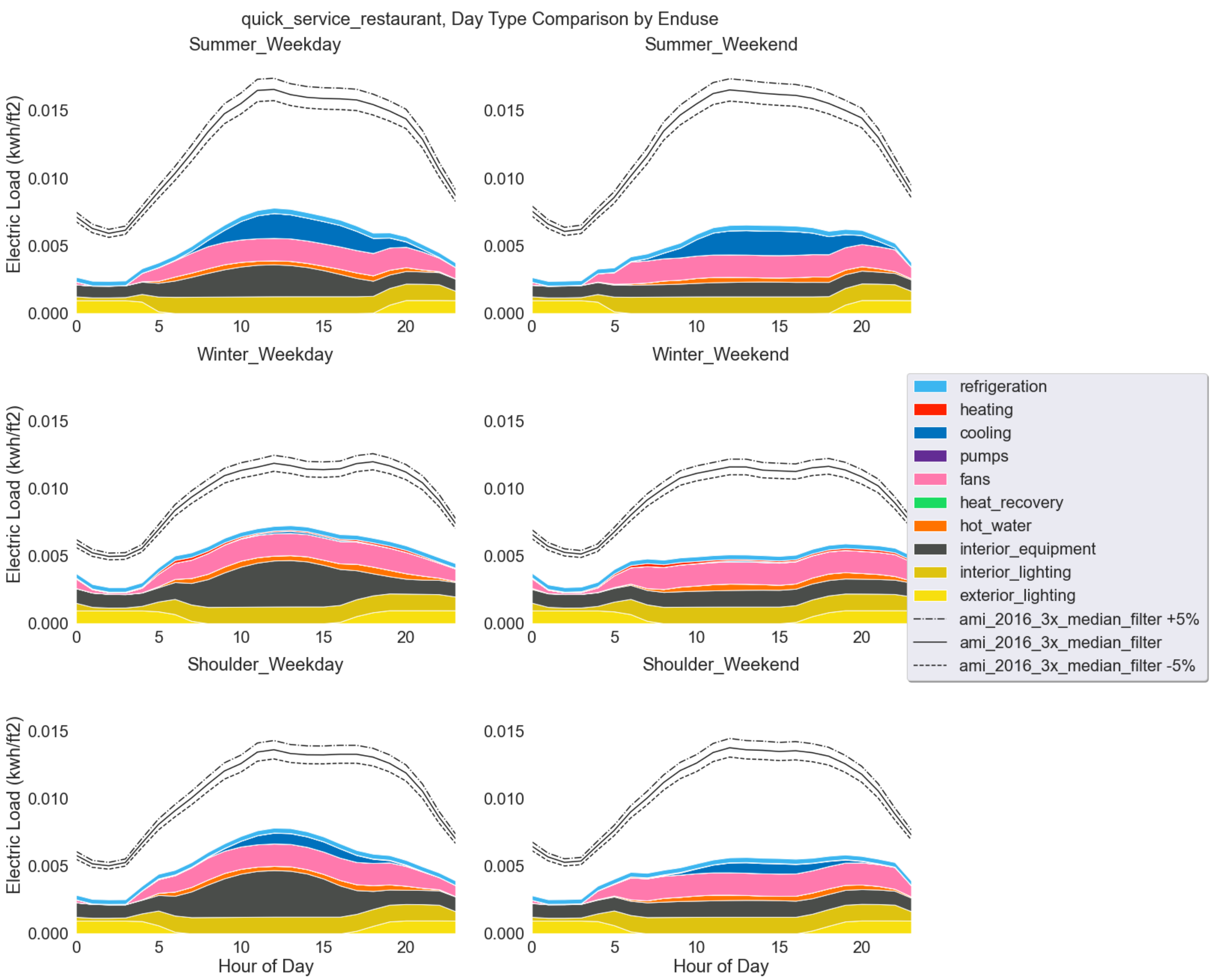

Figure C.7: Stacked area enduse plots by season and day type for quick service restaurants, ComStock run 7, Fort Collins, CO.

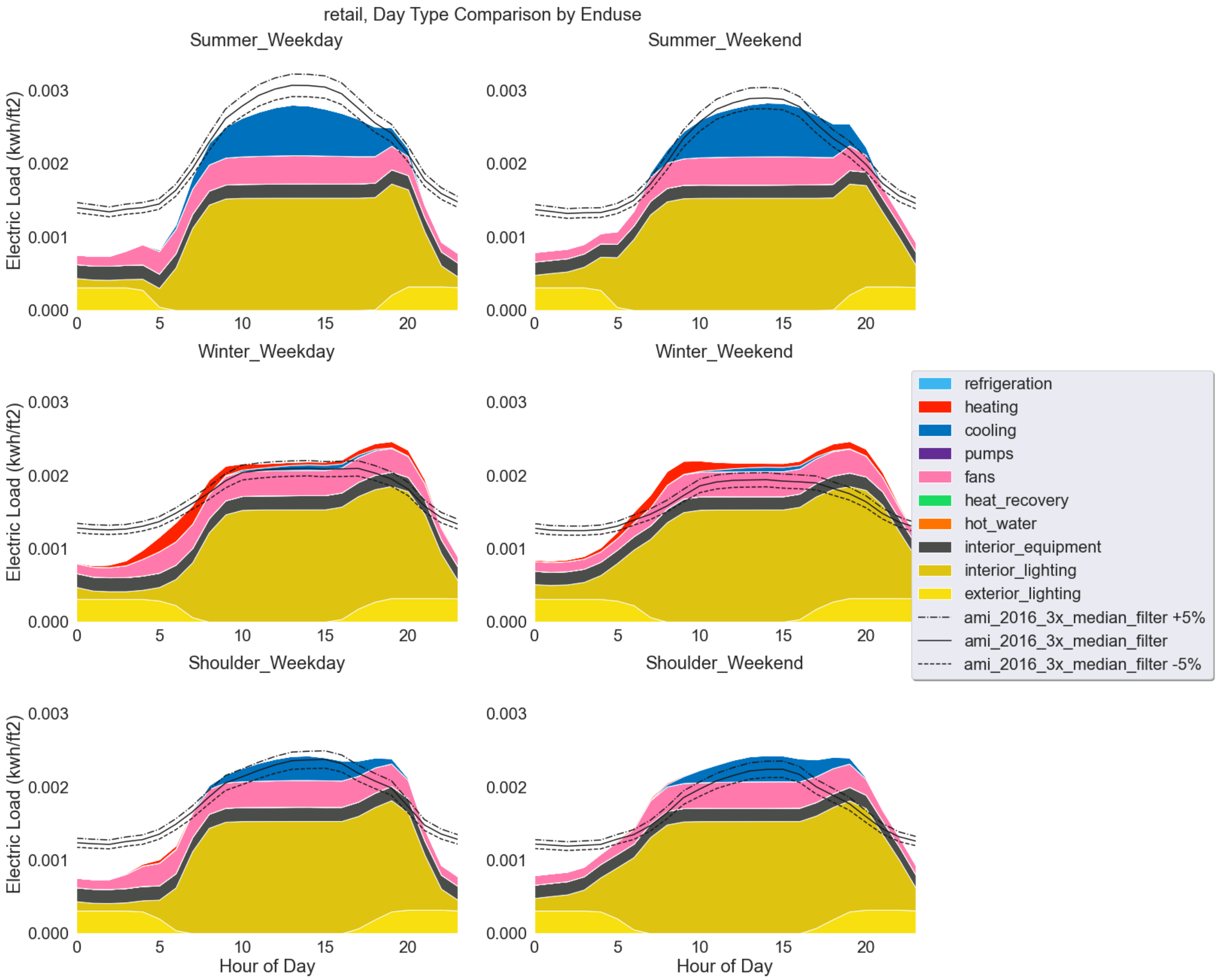

Figure C.8: Stacked area enduse plots by season and day type for retail, ComStock run 7, Fort Collins, CO.

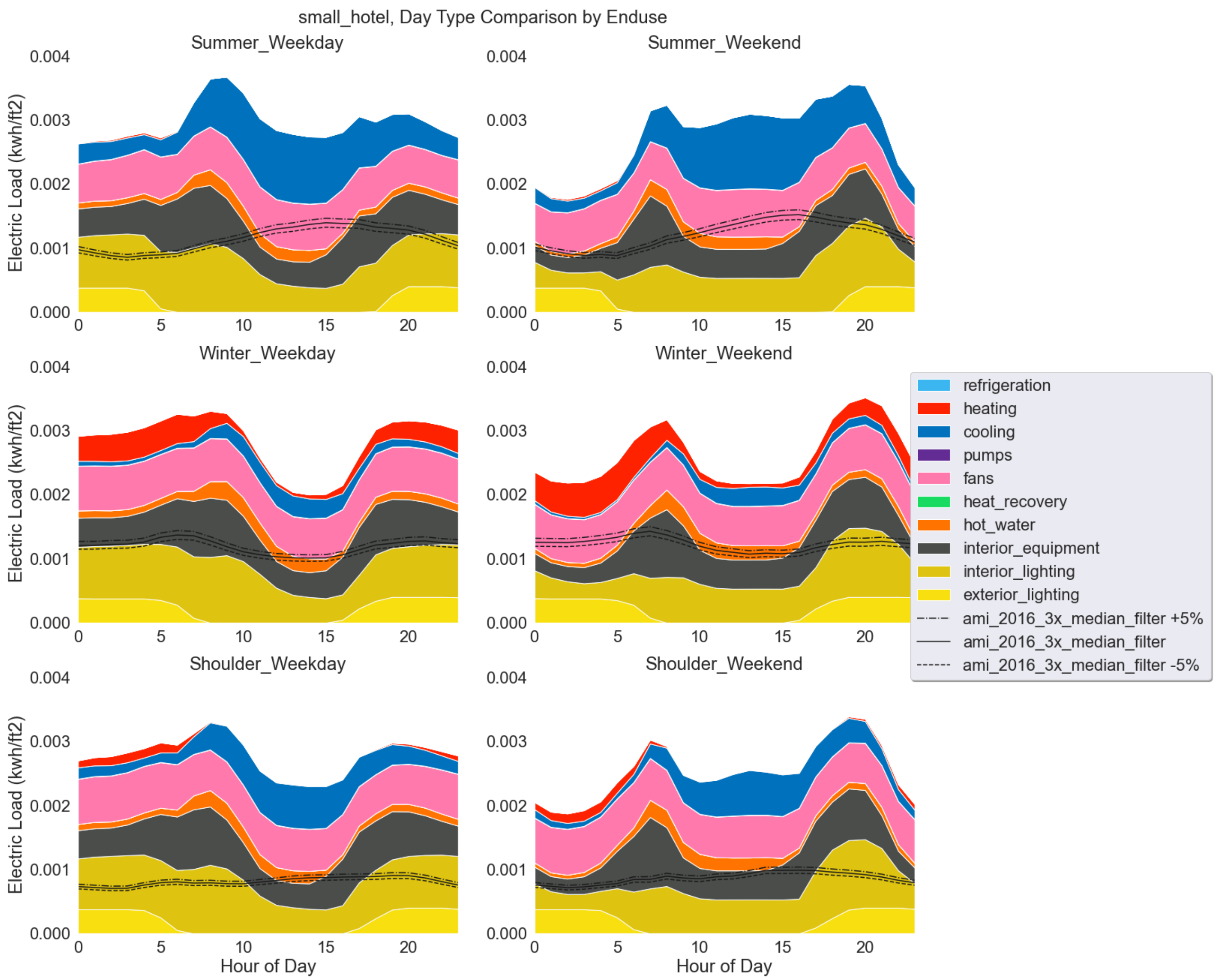

Figure C.9: Stacked area enduse plots by season and day type for small hotels, ComStock run 7, Fort Collins, CO.

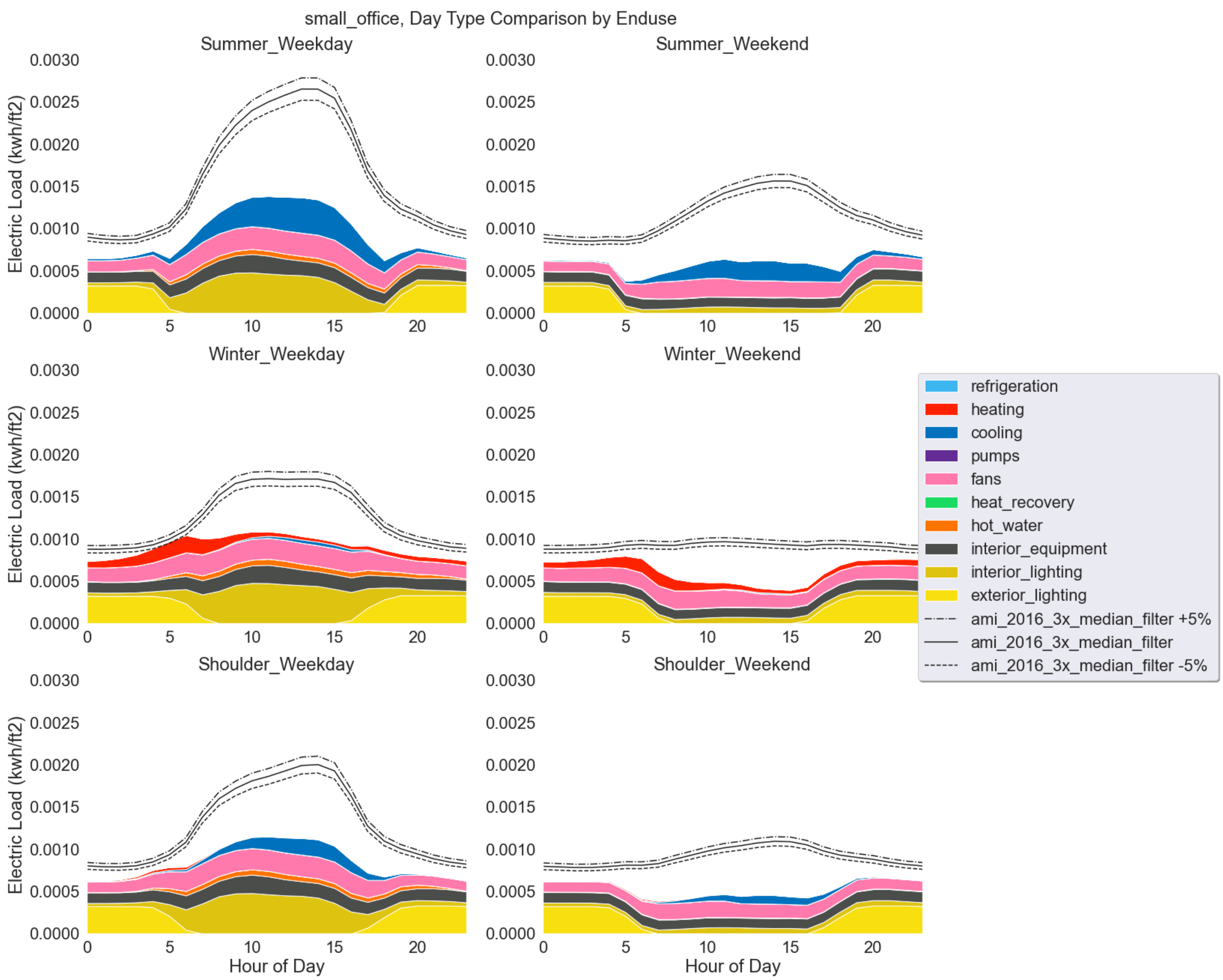

Figure C.10: Stacked area enduse plots by season and day type for small offices, ComStock run 7, Fort Collins, CO.

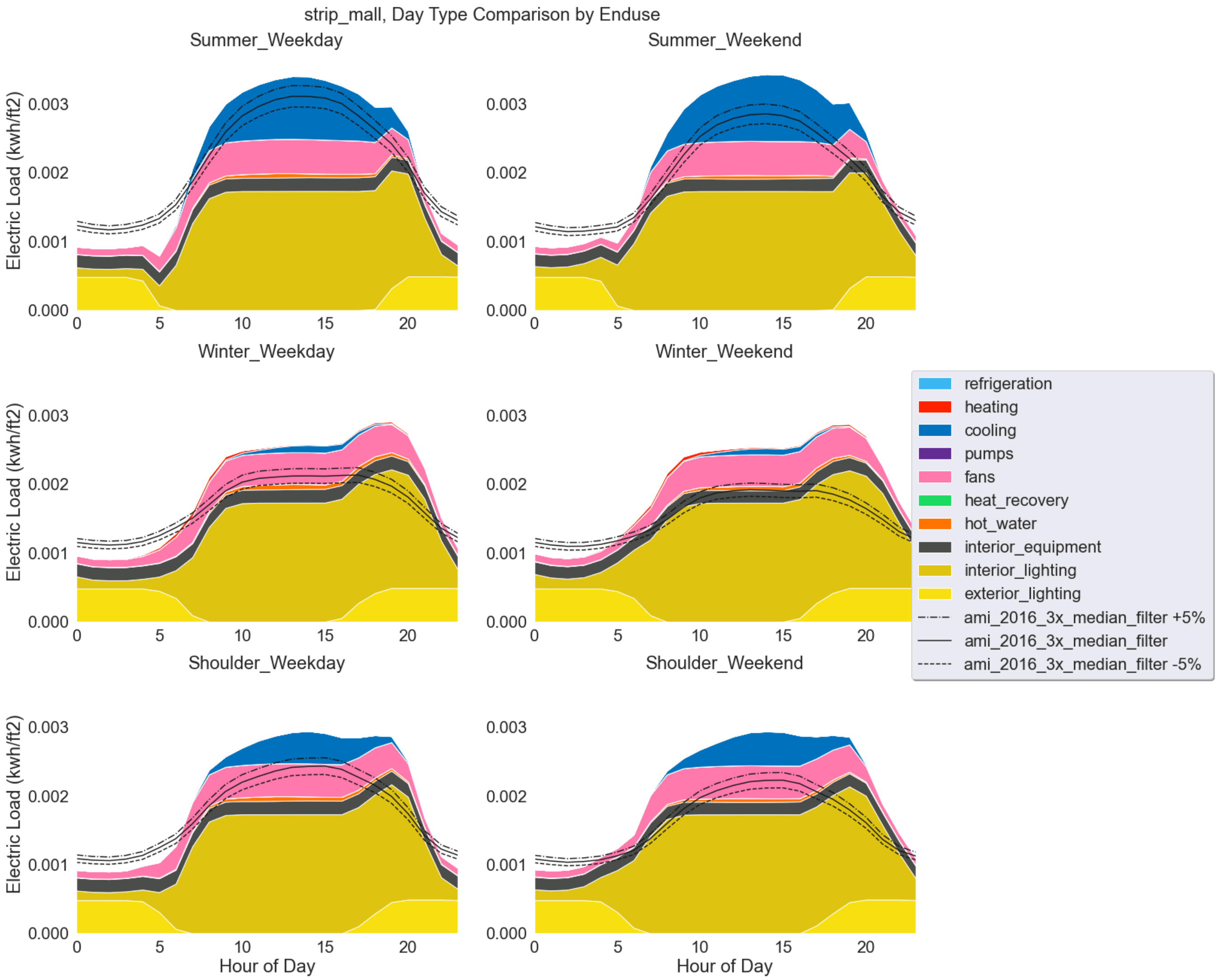

Figure C.11: Stacked area enduse plots by season and day type for strip malls, ComStock run 7, Fort Collins, CO.

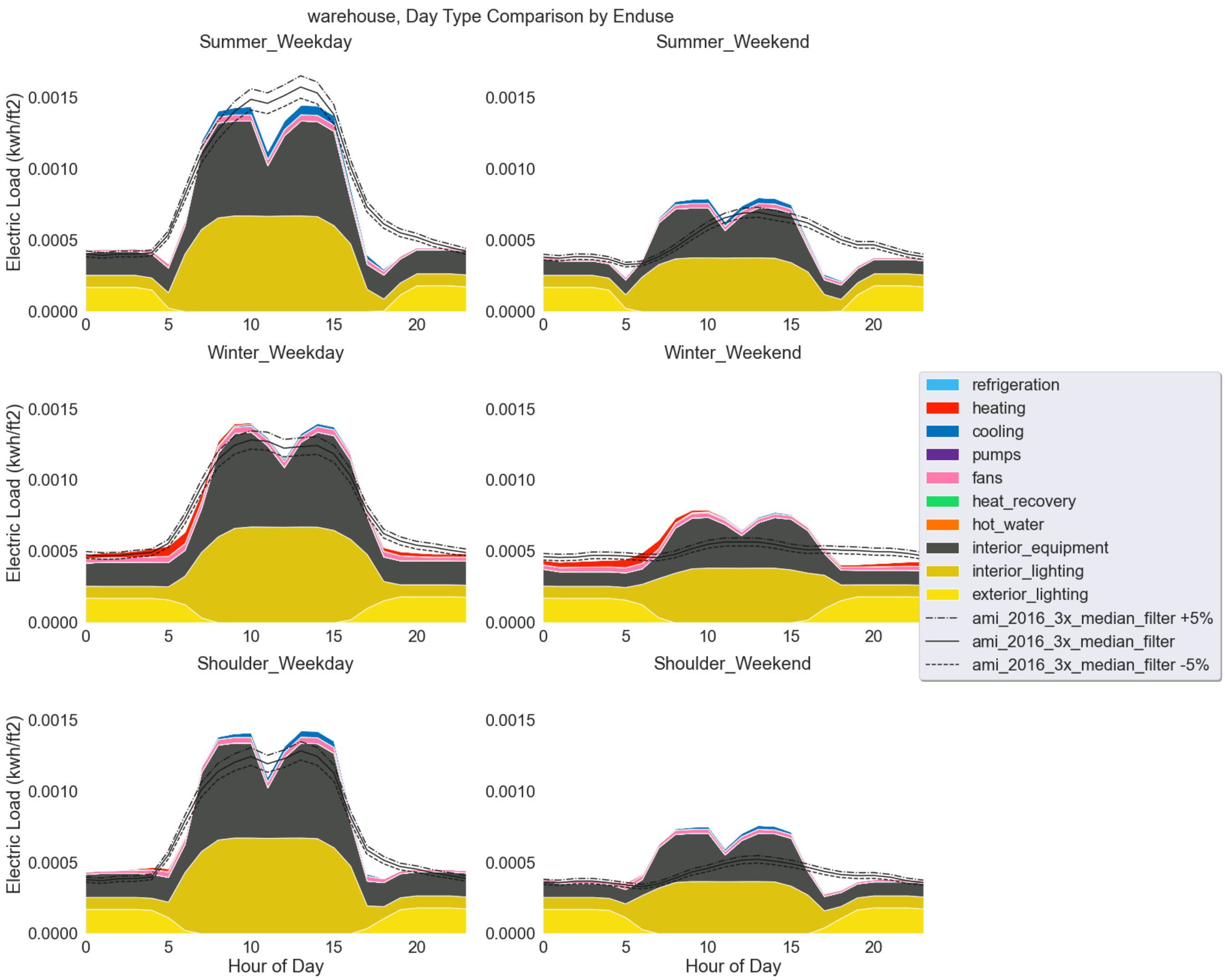

Figure C.12: Stacked area enduse plots by season and day type for warehouses, ComStock run 7, Fort Collins, CO.

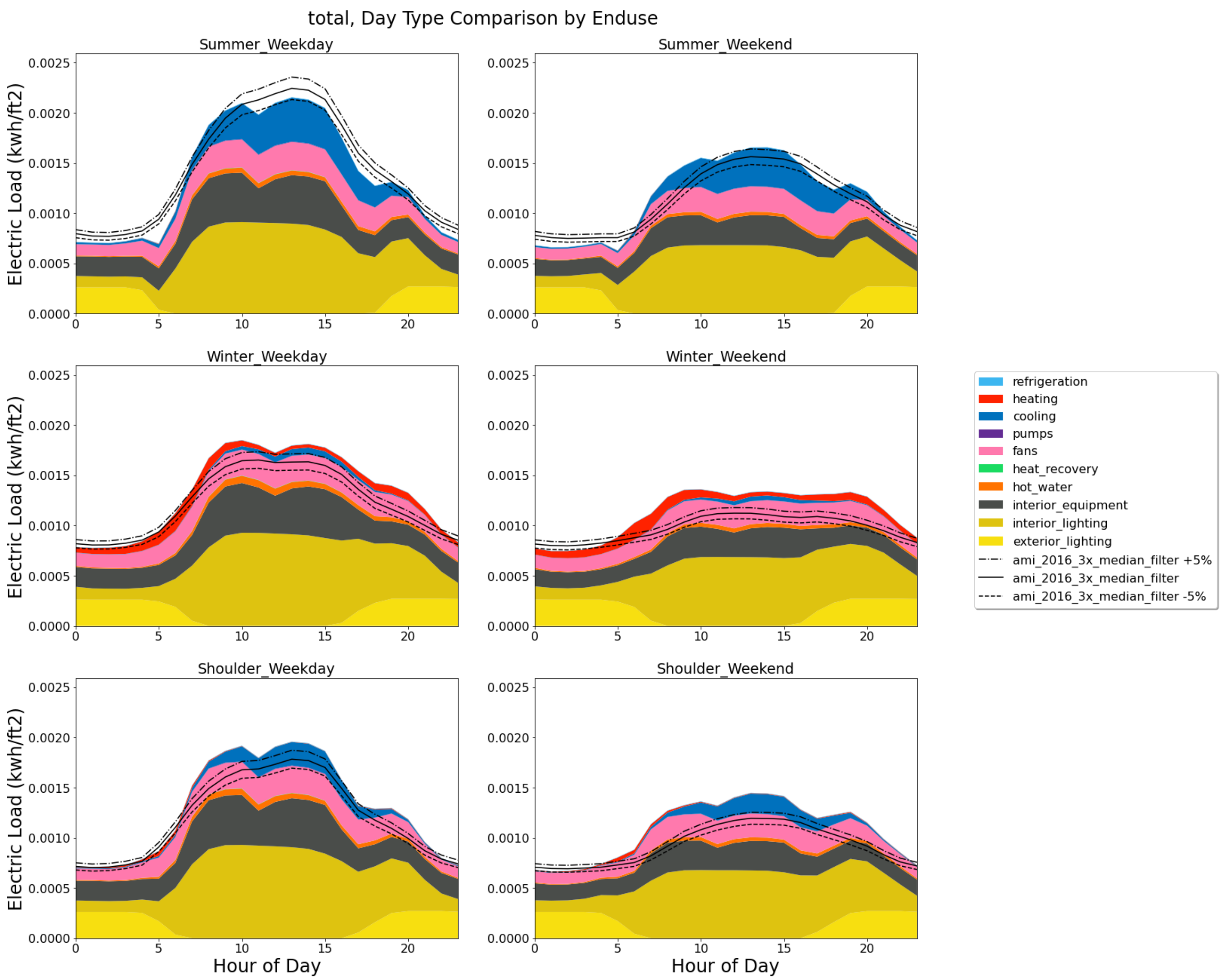

Figure C.13: Stacked area enduse plots by season and day type for total of all buildings, ComStock run 7, Fort Collins, CO.

Bibliography

[1] Hannah Choi Granade, Jon Creyts, Anton Derkach, Philip Farese, Scott Nyquist, and Ken Ostrowski. Unlocking Energy Efficiency in the U.S. Economy. Technical report, McKinsey&Company, July 2009.

[2] Philip Farese, Rachel Gelman, and Robert Hendron. A Tool to Prioritize Energy Efficiency Investments. Technical Report NREL/TP-6A20-54799, National Renewable Energy Laboratory, 15013 Denver West Parkway Golden, Colorado 80401, August 2012.

[3] U.S. DOE. Buildings Energy Data Book. Table 3.2.7. Technical report, U.S. Department of Energy, 2011.

[4] Nora Wang, Cody Taylor, and Molly McCabe. DOE Commercial Building Energy Asset Rating: Market Research and Program Direction. In *2012 ACEEE Summer Study on Energy Efficiency in Buildings, Pacific Grove, CA*, 2012.

[5] Victor Olgyay and Cherlyn Seruto. Whole-Building Retrofits: A Gateway to Climate Stabilization. *ASHRAE Transactions*, 2010.

[6] D. Urge-Vorsatz, K. Petrichenko, and A.C. Butcher. How far can buildings take us in solving climate change? a novel approach to building energy and related emission forecasting. In *ECEEE 2011 Summer Study, Belambra Presqu'ile de Giens, France*, pages 1343–1354, 2011.

[7] NEEA. NEEA Study: Examples of Deep Energy Savings in Existing Buildings. Technical report, New Buildings Institute, 2011.

[8] PNNL. Advanced Energy Retrofit Guide: Office Buildings. Technical report, Pacifc Northwest National Laboratory, September 2011.

[9] Jane Harris, Jane Anderson, and Walter Shafron. Investment in energy efficiency: A survey of Australian firms. *Energy Policy*, 28(12):867 – 876, 2000.

[10] Luis M. Abadie, Ramon A. Ortiz, and I. Galarraga. Determinants of energy efficiency investments in the us. *Energy Policy*, 45(Supplement C):551 – 566, 2012.

[11] EEB Hub. *BP2 Final Report*. Energy Efficiency Buildings Hub, Philadelphia, PA, 2013.

[12] Constantine Kontokosta, Danielle Spiegel-Feld, and Sokratis Papadopoulos. The impact of mandatory energy audits on building energy use. *Nature Energy*, 5(4):309 – 316, 2020.

[13] Zhenjun Ma, Paul Cooper, Daniel Daly, and Laia Ledo. Existing building retrofits: Methodology and state-of-the-art. *Energy and Buildings*, 55(Supplement C):889 – 902, 2012. Cool Roofs, Cool Pavements, Cool Cities, and Cool World.

[14] A.M. Rysanek and R. Choudhary. Optimum building energy retrofits under technical and economic uncertainty. *Energy and Buildings*, 57(Supplement C):324 – 337, 2013.

[15] Gürkan Kumbaroğlu and Reinhard Madlener. Evaluation of economically optimal retrofit investment options for energy savings in buildings. *Energy and Buildings*, 49(Supplement C):327 – 334, 2012.

[16] EVO. *International Performance Measurement and Verification Protocol: Concepts and options for determining energy and water savings volume 1*. Efficiency Valuation Organization, Vienna, Austria, 2012.

[17] ASHRAE. ASHRAE Guideline 14-2014: Measurement of Energy and Demand Savings. Technical report, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA, 2014.

[18] Daniel Coakley, Paul Raftery, and Marcus Keane. A review of methods to match building energy simulation models to measured data. *Renewable and Sustainable Energy Reviews*, 37(Supplement C):123 – 141, 2014.

[19] Nelson Fumo. A review on the basics of building energy estimation. *Renewable and Sustainable Energy Reviews*, 31(Supplement C):53 – 60, 2014.

[20] Aaron Garrett and Joshua New. Suitability of ASHRAE Guideline 14 Metrics for Calibration. *ASHRAE Transactions*, 122:469–477, January 2016.

[21] Jessica Granderson, Phillip N. Price, David Jump, Nathan Addy, and Michael D. Sohn. Automated measurement and verification: Performance of public domain whole-building electric baseline models. *Applied Energy*, 144(Supplement C):106 – 113, 2015.

[22] Jessica Granderson, Samir Touzani, Samuel Fernandes, and Cody Taylor. Application of automated measurement and verification to utility energy efficiency program data. *Energy and Buildings*, 142(Supplement C):191 – 199, 2017.

[23] J.M. Ayres and E. Stamper. Historical development of building energy calculations. In *American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) winter meeting and exhibition, Chicago, IL*, volume 101, page 1517. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta, GA (United States), Aug 1995.

[24] Drury B. Crawley, Jon W. Hand, Michaël Kummert, and Brent T. Griffith. Contrasting the capabilities of building energy performance simulation programs. *Building and Environment*, 43(4):661 – 673, 2008. Part Special: Building Performance Simulation.

[25] Drury B. Crawley, Linda K. Lawrie, Frederick C. Winkelmann, W.F. Buhl, Y.Joe Huang, Curtis O. Pedersen, Richard K. Strand, Richard J. Liesen, Daniel E. Fisher, Michael J. Witte, and Jason Glazer. Energyplus: creating a new-generation building energy simulation program. *Energy and Buildings*, 33(4):319 – 331, 2001. Special Issue: Building Simulation'99.

[26] NREL. *OpenStudio*. National Renewable Energy Laboratory, 15013 Denver West Parkway Golden, Colorado 80401, 2015.

[27] Paul Raftery, Marcus Keane, and James O'Donnell. Calibrating whole building energy models: An evidence-based methodology. *Energy and Buildings*, 43(9):2356 – 2364, 2011.

[28] T. Agami Reddy. Literature Review on Calibration of Building Energy Simulation Programs: Uses, Problems, Procedures, Uncertainty and Tools. *ASHRAE Transactions*, 112(1), January 2006.

[29] Nelson Fumo, Pedro Mago, and Rogelio Luck. Methodology to estimate building energy consumption using energyplus benchmark models. *Energy and Buildings*, 42(12):2331 – 2337, 2010.

[30] Lincoln C. Harmer and Gregor P. Henze. Using calibrated energy models for building commissioning and load prediction. *Energy and Buildings*, 92(Supplement C):204 – 215, 2015.

[31] Paul Raftery, Marcus Keane, and Andrea Costa. Calibrating whole building energy models: Detailed case study using hourly measured data. *Energy and Buildings*, 43(12):3666 – 3679, 2011.

[32] Young-Jin Kim and Cheol-Soo Park. Stepwise deterministic and stochastic calibration of an energy simulation model for an existing building. *Energy and Buildings*, 133(Supplement C):455 – 468, 2016.

[33] Zheng Yang and Burcin Becerik-Gerber. A model calibration framework for simultaneous multi-level building energy simulation. *Applied Energy*, 149(Supplement C):415 – 431, 2015.

[34] Iain Alexander Macdonald. *Quantifying the Effects of Uncertainty in Building Simulation*. PhD thesis, University of Strathclyde, July 2002.

[35] Yang-Seon Kim, Mohammad Heidarinejad, Matthew Dahlhausen, and Jelena Srebric. Building energy model calibration with schedules derived from electricity use data. *Applied Energy*, 190(Supplement C):997 – 1007, 2017.

[36] Gaurav Chaudhary, Joshua New, Jibonananda Sanyal, Piljae Im, Zheng O'Neill, and Vishal Garg. Evaluation of "autotune" calibration against manual calibration of building energy models. *Applied Energy*, 182(Supplement C):115 – 134, 2016.

[37] Elaine Hale, Lars Lisell, David Goldwasser, Daniel Macumber, Jesse Dean, Ian Metzger, Andrew Parker, Nicholas Long, Brian Ball, Marjorie Schott, Evan Weaver, and Larry Brackney. Cloud-Based Model Calibration Using OpenStudio. In *eSim 2014, Ottawa, Canada*, 2014.

[38] Wei Tian, Yeonsook Heo, Pieter de Wilde, Zhanyong Li, Da Yan, Cheol Soo Park, Xiaohang Feng, and Godfried Augenbroe. A review of uncertainty analysis in building energy assessment. *Renewable and Sustainable Energy Reviews*, 93:285 – 301, 2018.

[39] M. Rois Langner, Gregor P. Henze, Charles D. Corbin, and Michael J. Brandemuehl. An investigation of design parameters that affect commercial high-rise office building energy consumption and demand. *Journal of Building Performance Simulation*, 5(5):313–328, 2012.

[40] Kathrin Menberg, Yeonsook Heo, and Ruchi Choudhary. Sensitivity analysis methods for building energy models: Comparing computational costs and extractable information. *Energy and Buildings*, 133(Supplement C):433 – 445, 2016.

[41] Ke Xu. *Assessing the Minimum Instrumentation to Well Tune Existing Medium Sized Office Building Energy Models*. PhD thesis, The Pennsylvania State University, December 2012.

[42] T. Agami Reddy, Itzhak Maor, and Chanin Panjapornpon. Calibrating Detailed Building Energy Simulation Programs with Measured Data—Part I: General Methodology (RP-1051). *HVAC&R Research*, 13(2):221–241, 2007.

[43] Brian Coffey. A Development and Testing Framework for Simulation-Based Supervisory Control With Application to Optimal Zone Temperature Ramping Demand Response Using a Modified Genetic Algorithm. Master's thesis, Concordia University, 6 2008.

[44] Daniel Macumber, Brian Ball, and Nicholas Long. A Graphical Tool for Cloud-Based Building Energy Simulation. In *2014 ASHRAE/IBPSA-USA Building Simulation Conference, Atlanta, GA*, 2014.

[45] T. Agami Reddy, Itzhak Maor, and Chanin Panjapornpon. Calibrating detailed building energy simulation programs with measured data—part ii: Application to three case study office buildings (rp-1051). *HVAC&R Research*, 13(2):243–265, 2007.

[46] Yeonsook Heo, Godfried Augenbroe, and Ruchi Choudhary. Quantitative risk management for energy retrofit projects. *Journal of Building Performance Simulation*, 6(4):257–268, 2013.

[47] Herman Carstens, Xiaohua Xia, and Sarma Yadavalli. Low-cost energy meter calibration method for measurement and verification. *Applied Energy*, 188(Supplement C):563 – 575, 2017.

[48] Qi Li, Li Gu, Godfried Augenbroe, C.F. Jeff Wu, and Jason Brown. Calibration of Dynamic Building Energy Models with Multiple Responses Using Bayesian Inference and Linear Regression Models. *Energy Procedia*, 78(Supplement C):979 – 984, 2015. 6th International Building Physics Conference, IBPC 2015.

[49] Matthew Dahlhausen, Mohammad Heidarinejad, and Jelena Srebric. Building energy retrofits under capital constraints and greenhouse gas pricing scenarios. *Energy and Buildings*, 107(Supplement C):407 – 416, 2015.

[50] Marc C. Kennedy and Anthony O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.

[51] Yeonsook Heo. *Bayesian Calibration of Building Energy Models for Energy Retrofit-Decision-Making Under Uncertainty.* PhD thesis, Georgia Institute of Technology, December 2011.

[52] Adrian Chong. *Bayesian Calibration of Building Energy Models for Large Datasets*. PhD thesis, Carnegie Mellon University, May 2017.

[53] Filippo Monari. *Sensitivity Analysis and Bayesian Calibration of Building Energy Models*. PhD thesis, University of Strathclyde, February 2016.

[54] Matthew Riddle and Ralph Muehleisen. A Guide to Bayesian Calibration of Building Energy Models. In *2014 ASHRAE/IBPSA-USA Building Simulation Conference, Atlanta, GA*, 2014.

[55] Yeonsook Heo, Diane J. Graziano, Leah Guzowski, and Ralph T. Muehleisen. Evaluation of calibration efficacy under different levels of uncertainty. *Journal of Building Performance Simulation*, 8(3):135–144, 2015.

[56] Y. Heo, R. Choudhary, and G.A. Augenbroe. Calibration of building energy models for retrofit analysis under uncertainty. *Energy and Buildings*, 47(Supplement C):550 – 560, 2012.

[57] Christina Johanna Hopfe. *Uncertainty and sensitivity analysis in building performance simulation for decision support and design optimization.* PhD thesis, Technische Universiteit Eindhoven, June 2009.

[58] Yuming Sun. *Closing the Building Energy Retrofit Performance Gap by Improving our Predictions.* PhD thesis, Georgia Institute of Technology, August 2014.

[59] Yuming Sun, Li Gu, C.F. Jeff Wu, and Godfried Augenbroe. Exploring HVAC system sizing under uncertainty. *Energy and Buildings*, 81(Supplement C):243 – 252, 2014.

[60] Ralph Muehleisen and Joshua Bergerson. Bayesian Calibration - What, Why And How. In *4th International High Performance Buildings Conference at Purdue*, July 2016.

[61] Wei Tian, Song Yang, Zhanyong Li, Shen Wei, Wei Pan, and Yunliang Liu. Identifying informative energy data in bayesian calibration of building energy models. *Energy and Buildings*, 119(Supplement C):363 – 376, 2016.

[62] Adrian Chong, Khee Poh Lam, Matteo Pozzi, and Junjing Yang. Bayesian calibration of building energy models with large datasets. *Energy and Buildings*, 154(Supplement C):343 – 355, 2017.

[63] Qi Li, Godfried Augenbroe, and Jason Brown. Assessment of linear emulators in lightweight bayesian calibration of dynamic building energy models for parameter estimation and performance prediction. *Energy and Buildings*, 124(Supplement C):194 – 202, 2016.

[64] Hyunwoo Lim and Zhiqiang John Zhai. Comprehensive evaluation of the influence of meta-models on bayesian calibration. *Energy and Buildings*, 155(Supplement C):66 – 75, 2017.

[65] Sungmin Yoon and Yuebin Yu. A quantitative comparison of statistical and deterministic methods on virtual in-situ calibration in building systems. *Building and Environment*, 115(Supplement C):54 – 66, 2017.

[66] M. Kavgic, A. Mavrogianni, D. Mumovic, A. Summerfield, Z. Stevanovic, and M. Djurovic-Petrovic. A review of bottom-up building stock models for energy consumption in the residential sector. *Building and Environment*, 45(7):1683 – 1697, 2010.

[67] Eric Wilson, Craig Christensen, Scott Horowitz, Joseph Robertson, and Jeff Maguire. Energy efficiency potential in the u.s. single-family housing stock. Technical Report TP-5500-68670, National Renewable Energy Laboratory, December 2017.

[68] C.F. Reinhart and C. Cerezo Davila. Urban building energy modeling - a review of a nascent field. *Building and Environment*, 97:196–202, 2016. cited By 249.

[69] Elaine Hale, Henry Horsey, Brandon Johnson, Matteo Muratori, and Eric Wilson. The demand-side grid (dsgrid) model documentation. Technical Report TP-6A20-7149, National Renewable Energy Laboratory, 2018.

[70] Stéphane Bertagnolio. *Evidence-Based Model Calibration For Efficient Building Energy Services* . PhD thesis, University of Liège, June 2012.

[71] J. Cipriano, G. Mor, D. Chemisana, D. Pérez, G. Gamboa, and X. Cipriano. Evaluation of a multi-stage guided search approach for the calibration of building energy simulation models. *Energy and Buildings*, 87(Supplement C):370 – 385, 2015.

[72] Young-Jin Kim, Seong-Hwan Yoon, and Cheol-Soo Park. Stochastic comparison between simplified energy calculation and dynamic simulation. *Energy and Buildings*, 64(Supplement C):332 – 342, 2013.

[73] Zheng O'Neill and Bryan Eisenhower. Leveraging the analysis of parametric uncertainty for building energy model calibration. *Building Simulation*, 6(4):365–377, Dec 2013.

[74] Fernando Simon Westphal and Roberto Lamberts. Building Simulation Calibration Using Sensitivity Analysis. In *Buidling Simulation 2005, Montréal, Canada*, 2005.

[75] Ying Ji and Peng Xu. A bottom-up and procedural calibration method for building energy simulation models based on hourly electricity submetering data. *Energy*, 93(Part 2):2337 – 2350, 2015.

[76] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[77] S.E. Chidiac, E.J.C. Catania, E. Morofsky, and S. Foo. A screening methodology for implementing cost effective energy retrofit measures in canadian office buildings. *Energy and Buildings*, 43(2):614 – 620, 2011.

[78] NREL. *Building Component Library*. National Renewable Energy Laboratory, 15013 Denver West Parkway Golden, Colorado 80401, 2014.

[79] Nicholas Long, Brian Ball, Larry Brackney, David Goldwasser, Andrew Parker, Jennifer Elling, Oliver David, and Dale Kruchten. Leveraging OpenStudio's application programming interfaces. In *13th Conference of International Building Performance Simulation Association, Chambéry, France*, page 1095–1102, 2013.

[80] Guopeng Liu and Mingsheng Liu. A rapid calibration procedure and case study for simplified simulation models of commonly used hvac systems. *Building and Environment*, 46(2):409 – 420, 2011.

[81] Payam Delgoshaei, Ke Xu, Scott Wagner, Richard Sweetser, and James Freihaut. Hourly Plug Load Measurements and Profiles for a Medium Office Building - a Case Study. In *AEI 2013*, pages 827–836, 2013.

[82] ASHRAE. ASHRAE Standard 189.1-2014 - Standard for the Design of High-Performance Green Buildings. Standard, American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Atlanta,GA, 2014.

[83] Matthew Dahlhausen. Staging Building Energy Retrofits. Master's thesis, The Pennsylvania State University, January 2014.

[84] A. Dasgupta, H. Henderson, R. Sweetser, and T. Wagner. Building Monitoring System and Preliminary Results for a Retrofitted Office Building. In *International High Performance Buildings Conference, Purdue University*, 2012.

[85] Public Works & Government Services Canada. Air Leakage Control: Retrofit Measures for High-Rise Office Buildings. Technical report, Public Works & Government Services Canada, 1993.

[86] ASHRAE. ASHRAE Standard 90.1-2010 - Energy Standard for Buildings Except Low-Rise Residential Buildings. Standard, American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Atlanta,GA, 2010.

[87] ASHRAE. Advanced energy design guide for small to medium office buildings. Technical report, American Society of Heating, Refrigerating, and Air-Conditioning Engineers, 2011.

[88] Trane. *Split Systems 20 to 120 tons*, 2014.

[89] N.A. Desai, R.D. Taylor, S. Narayanan, and T. Wagner. Deep Retrofit System Solution Assessment for Philadelphia Navy Yard Office Buildings, Building Monitoring System and Preliminary Results for a Retrofitted Office Building. In *International High Performance Buildings Conference, Purdue University*, 2012.

[90] Reed Construction Data Inc. *RSMeansOnline Version 5.0.3*, 2014.

[91] ASHRAE. 2015 ASHRAE Handbook of Fundamentals. Technical report, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA, 2015.

[92] Federal Energy Management Program. M&V Guidelines: Measurement and Verification for Performance-Based Contracts Version 4.0. Technical report, U.S. Department of Energy, November 2015.

[93] J. Bohadel. Building 101 Energy Audit Report. Technical report, Dome-Tech, Inc., 2011.

[94] Mohammad Heidarinejad, Matthew Dahlhausen, Sean McMahon, Chris Pyke, and Jelena Srebric. Cluster analysis of simulated energy use for leed certified u.s. office buildings. *Energy and Buildings*, 85(Supplement C):86 – 97, 2014.

[95] Mohammad Heidarinejad. *Relative Significance of Heat Transfer Processes to Quantify Tradeoffs Between Complexity and Accuracy of Energy Simulations With a Building Energy Use Patterns Classification, A Dissertation in Architectural Engineering*. PhD thesis, The Pennsylvania State University, 2014.

[96] Michael Deru, Kristin Field, Daniel Studer, Kyle Benne, Brent Griffith, Paul Torcellini, Bing Liu, Mark Halverson, Dave Winiarski, Michael Rosenberg, Mehry Yazdanian, Joe Huang, and Drury Crawley. U.S. Department of Energy Commercial Reference Building Models of the National Building Stock. Technical Report NREL/TP-5500-46861, National Renewable Energy Laboratory, 15013 Denver West Parkway Golden, Colorado 80401, February 2011.

[97] Mohammad Heidarinejad, Nicholas Mattise, Krishang Sharma, and Jelena Srebric. Creating Geometry with Basic Shape Templates in OpenStudio. *Procedia Engineering*, 205:1990 – 1995, 2017. 10th International Symposium on Heating, Ventilation and Air Conditioning, ISHVAC2017, 19-22 October 2017, Jinan, China.

[98] U.S. Energy Information Administration. Commercial buildings energy consumption survey, 2012.

[99] Nicole Buccitelli, Clay Elliott andSeth Schober, and Mary Yamada. 2015 u.s. lighting market characterization survey. Technical report, U.S. Department of Energy Office of Energy Efficiency and Renewable Energy, November 2017.

[100] Kee Han Kim and Jeff S. Haberl. Development of methodology for calibrated simulation in single-family residential buildings using three-parameter change-point regression model. *Energy and Buildings*, 99(Supplement C):140 – 152, 2015.

[101] Jessica Granderson, Samir Touzani, Claudine Custodio, Michael D. Sohn, David Jump, and Samuel Fernandes. Accuracy of automated measurement and verification (m&v) techniques for energy savings in commercial buildings. *Applied Energy*, 173(Supplement C):296 – 308, 2016.

[102] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2011.

[103] Marlena Praprost, Katherine A Fleming, and Matthew Dahlhausen. Energy star for tenants: An online energy estimation tool for commercial office building tenants. 2 2020.

[104] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[105] Gilles Louppe. *Understanding Random Forests: From Theory to Practice.* PhD thesis, University of Liège, 2014.

[106] Mohammad Heidarinejad, Nick Mattise, Matthew Dahlhausen, Saber Khoshdel Nikkho, J. Liu, Stefan Gracik, Kai Liu, S. Krishang, Haoyue Zhang, Josh Wentz, M. Sadeghipour Roudsari, George Pitchurov, and Jelena Srebric. Urban scale modeling of campus building using virtual pulse. *International Building Performance Simulation Association (IBPSA)*, 2015.

[107] David Goldwasser, Matthew Dahlhausen, and Marlena Praprost. Automatic generation of highly customizable energy models from high level input data. Presented at the Building Performance Analysis Conference Denver, CO, 2019.

[108] Rois Langner, Paul Torcellini, Matthew Dahlhausen, David Goldwasser, Joe Robertson, and Sarah Zaleski. Transforming New Multifamily Construction to Zero: Strategies for Implementing Energy Targets and Design Pathways. In *2020 ACEEE Summer Study, Virtual*, 2020.

[109] Matthew Dahlhausen and David Goldwasser. Integrating residential and commercial modeling for zero energy mixed-use multifamily building design. Presented at the Building Performance Analysis Conference, Virtual, 2020.