Accelerating Urban Building Energy Modeling

by

Zoe Le Hong

BASc
The University of British Columbia, 2020

and

Samuel Wolk

AB
Harvard College, 2017

Submitted to the Department of Architecture
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE IN BUILDING TECHNOLOGY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2024

Authored by:

Zoe Le Hong
Department of Architecture
April 29, 2024

Authored by:

Samuel Wolk
Department of Architecture
April 29, 2024

Certified by:

Christoph Reinhart
Professor of Building Technology, Thesis Advisor

Accepted by:

Leslie K. Norford
Professor of Building Technology
Chair, Department Committee on Graduate Students

Accelerating Urban Building Energy Modelingby

Zoe Le Hong

and

Samuel Wolk

Submitted to the Department of Architecture
on April 29, 2024 in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN BUILDING TECHNOLOGY

ABSTRACT

Enabling data-driven decision-making in the built environment is critical to achieving ambitious and urgent decarbonization goals. In the building sector, urban building energy models (UBEMs) have become a valuable tool for jurisdictions to develop evidence-based retrofitting policies, but dynamically exploring solutions is hampered by the computational expense and organizational overhead of physics-based building energy models. In order to address these challenges, we present a fast, flexible, and comprehensive UBEM methodology which can be used to reduce identified barriers to time-sensitive decision-making in building stock decarbonization spheres. The methodology combines the speed of current data-driven approaches with the flexibility of computationally intensive, but accurate, engineering models. Identifying machine learning methods as a viable approach, we implement convolutional neural networks (CNNs) which embed time-series from hourly weather data and building schedules; the embeddings are then combined with static building characteristics and projected to monthly heating and cooling loads. The proposed approach allows for programmatic flexibility and robustness to unique hourly weather conditions globally, while contextual abstraction enables geometric independence. A dataset of over 1 million detailed thermodynamics-based simulations was constructed to train and validate the surrogate model. Model results at the individual shoebox, building, and urban scales compare favorably to traditional numerical methods and meet accepted error bounds under national energy simulation standards. Additional validation at the urban- and national-scales are performed using public building simulation datasets. We then demonstrate expanded applications, which leverage the reduced computational cost of the framework to make traditionally infeasible analysis modes tractable and deployable. The methodology presented is intended to be utilized for both very-large-scale systematic analysis and near-real-time interactive explorations. In developing this framework, we aim to provide new mechanisms for key stakeholders in the decarbonization effort to quickly generate actionable insights and engage in iterative discussions to develop evidence-based policy across global building stocks.

Thesis Advisor: Christoph Reinhart
Title: Professor of Building Technology

Acknowledgments

This work would not have been possible without the support, contributions, guidance, and patience of our advisor, professors, colleagues, friends, and family.

First, we would like to express our utmost gratitude for our thesis advisor and mentor, Christoph Reinhart, whose never-ending fountain of enthusiasm and optimism for all things building science has propelled us to play what role we can in designing a more sustainable world. To Caitlin Mueller and Renaud Danhaive, for helping us imagine the potential of this work, as well as the rest of the Building Technology faculty, Les Norford, Leon Glicksman, and John Ochsendorf for your teaching, mentorship and example.

Thank you to all the brilliant members of the MIT Sustainable Design Lab who have worked alongside us for the past two years and have made every late night and windowless lunch a delight. Thank you to Jenni, for the lunches and cheerful chats.

Sam Wolk and Zoë Le Hong

To Aoife Leonard, I can't express enough the importance of your incredible support throughout the countless late nights and hair falling out over failed simulations. None of this would be possible without you, and to my family, who have always inspired me to push on to the next challenge, I thank you for your constant belief in me.

Sam Wolk

To my family and friends, I owe my deepest gratitude for being the support and joy that keeps me going. Your belief has been indispensable and I would not be here without it. To my mother, for her example and brilliance in expression and research. Thank you for grounding me and guiding me. To my father, for helping me navigate my presentations and career. To Xav, for his curiosity and joyful presence - I look forward to seeing you excel. To Amelie, thank you for the countless hours of silent focus that kept us both on track. And to Georg, my partner and support system, thank you for being my North star.

Zoë Le Hong

Contents

List of Figures

List of Tables

All images created by authors unless otherwise noted.

Hurtling towards uncharted complex system dynamics, cascading failures and runaway feedback loops, we find ourselves collectively engaged in a race against time with a singular goal: reduce carbon emissions, or, less euphemistically, prevent climate collapse. The built environment has a limited carbon budget, and so as architects and engineers, we have to make every ton of carbon count. None of us are so easily classified though: we are all also computational designers, programmers, digital storytellers, toolmakers. With these other intersecting digital identities comes an analogous responsibility to make every floating-point operation (FLOP) count. Each FLOP is a lever: spend some carbon on "compute", and in doing so, hopefully, somewhere further down the line, prevent more carbon from being emitted than was spent. We have an imperative to make every single FLOP effective, just as we ought to strategically engage every gram of material within a structure. This responsibility has two axes of activity: on one hand, how can we extract the maximum amount of actionable information from every FLOP in the shortest amount of time, and, on the other hand, how do we intentionally transduce this virtual activity into real-world impact.

Chapter 1

Introduction

The global building stock must undergo significant changes to minimize and mitigate the effects of climate change. Current buildings' operational energy use accounts for approximately one-third of global energy consumption and carbon emissions, which are concentrated in urban areas [1]. Without efforts to decarbonize and reduce the energy use of both new and existing buildings' operations, emissions will continue to rise, especially as urban areas age, grow, and increase in density.

Thus far, we have seen that fossil fuel emission reduction targets such as those set out by the Intergovernmental Panel on Climate Change (IPCC) and the Paris Agreement have not been addressed with the necessary urgency [1], [2]. Many global, national, and municipal goals set objectives for reductions at least 10 to 20 years in the future (often 2050). As such, efforts are typically focused on top-down planning, relying on future leadership and natural technological advancement for actual implementation, and focusing on increasing efficiency standards for new buildings rather than improving existing ones. New buildings can be built to near- or net-zero standards by design, but the decarbonization of existing buildings (a decarbonization lever which is vital) comes with additional challenges [3], [4]. Weber et. al. estimate that approximately 51% of the global floor area needed in 2050 already exists (68% in North America, and 72% in Europe) [5]. With a current retrofit rate optimistically estimated to be 1% of global floor area per year (most of which is focused in high-GDP countries), this must increase to 2.5-5.6% per year to meet the IPCC's Shared Socioeconomical Pathways 2050 reduction targets [1], [5]. So, while the challenge of designing and building energy efficient and sustainable new buildings is of great importance, we must also tackle the challenge of decarbonizing our existing building stock, faster.

Technical solutions exist to reduce emissions of buildings through energy upgrade renovations, or retrofits. In this work, we define a retrofit to be the modification of an existing building, specifically designed to reduce the building's energy use. Discussed further in 1.1, some common retrofits include: installing LED lights, upgrading to more efficient windows, increasing insulation levels, sealing leaky windows and doors, implementing occupancy-driven lighting and ventilation controls, installing low-flow faucets, and upgrading to more efficient HVAC systems.

However, implementations of these solutions are not siloed as engineering challenges but are complex and societal [6]. Electrification can increase pressure on cities' electrical systems [7]; urban heat islands and rising temperatures increase health risks related to overheating and air quality [8]; and due to buildings' long lifetimes, decisions made today will impact urban emissions for decades [1]. Challenges within and across systems must be addressed in an integrated manner to progress towards increasingly stringent and necessary emissions targets. This systems-based

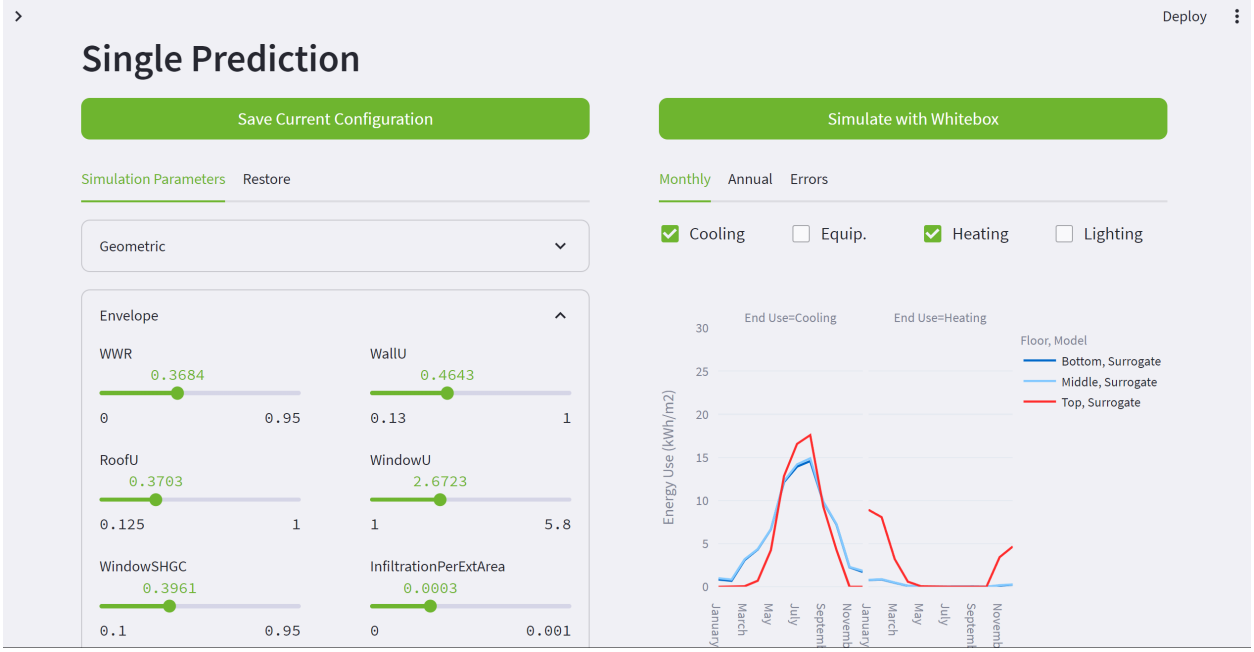


Figure 1.1: Prototype web application deployed with surrogate modeling building simulation capacities.

approach is especially important to ensure just and equitable energy transitions which minimize the imbalanced impacts of climate change on vulnerable populations [6], [9]. Yet, there is a direct conflict between the long-term impacts of building-related policy and the time and resource constraints of current policymakers, developers, and political cycles [10]. Therefore, in the realm of building energy modeling applied to the study of and planning for top-down change, there is a consistent struggle to balance accurate findings with rapid results.

This thesis attempts to address these challenges though the development and validation of a tool which has been specified to empower decision-making for urban decarbonization. The ultimate goal is to incorporate this work into accessible tools for a wide range of stakeholders, such as that which is shown in Figure 1.1, depicting a prototype deployment in an interactive dashboard. Our proposed methodology successfully generates results for urban building energy models at a rapid pace, with novel flexibility and detail for the observed computation speed. The development and validation of this solution is described in this thesis, forming the basis for a wide range of future applications and additional work within sustainable policy-making, design, and engineering.

1.1 The use and reduction of operational energy in buildings

Most fossil fuel emissions of conventional and existing buildings are due to the use of a building over its operational lifetime, through the direct or indirect use of fossil fuels for heating, cooling, and electricity needs (as opposed to *embodied* emissions from construction and demolition). The largest sources of energy use in buildings globally are space heating, (increasingly) cooling, appliance and equipment usage, water heating, and lighting. Other energy needs include fans, pumps, and other active system demands [1]. These end-uses are largely dependent on the type of building,

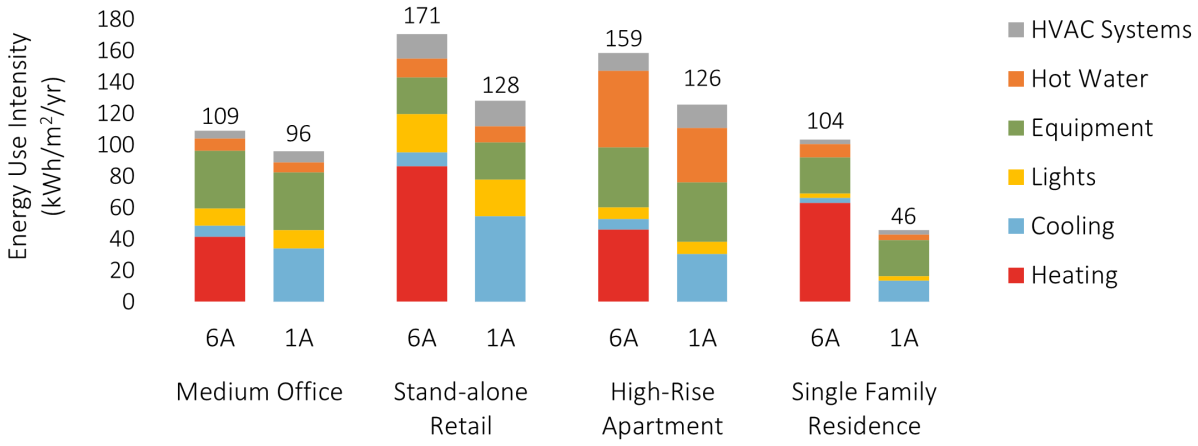


Figure 1.2: Energy use intensity of typical North American commercial building typologies in cool (6A) and hot (1A) climate zones [11].

or use *patterns* and *intensity*, as well as the local climate. Figure 1.2 describes typical breakdowns of energy end-use for several common building types in North America, in a cool and warm climate zone, illustrating the significance of a building's use and climatic context, along with how it is constructed. Annual energy use varies greatly due to unique building profiles; it is therefore important to understand a building's energy use at a baseline prior to determining which energy efficiency interventions can have an impact.

Retrofits can be implemented to reduce buildings' total fossil fuel emissions in four main ways: (1) reducing reliance on fossil fuels through electrification of fossil-fuel-intensive systems; (2) increasing the efficiency of building systems and equipment to prevent losses from fuel- and electricity-intensive systems; (3) reducing heating, cooling, and electricity demands by altering the construction and operation of the building itself; and, (4) generation of on-site green energy to reduce reliance on the grid (such as rooftop photovoltaics). In this way, most buildings are able to achieve near-zero operation with (often significant) retrofits, making buildings an achievable but expensive lever for meeting jurisdictional emissions reduction goals.

However, buildings are very complex systems, and, combined with a lack of data surrounding *current* energy use, understanding precisely what is needed to reduce building energy use is a difficult task, even at a single-building scale. End-use demands for heating and cooling are particularly complex, as they depend on an intricate interplay between the thermal properties of the building envelope, variable outside environmental conditions, desired indoor conditions, the type of conditioning system used, and the dynamics of interior occupancy patterns. So, measuring and predicting the impact of interventions such as retrofits on energy use is not trivial.

1.1.1 The thermal response of buildings

To maintain comfortable and healthy indoor conditions, a building must respond to constantly changing internal and external environments, such as weather or sunlight, use of appliances, or operation of windows. The intervention of HVAC systems (active or passive) for heating, cooling, and conditioning are not simply a response to instantaneous changes in the system, but interact

with the construction of the building itself. The building envelope, walls, roof, windows, and floor, act to thermally separate a building's interior from the exterior. The level of insulation, thermal mass, and airtightness of the envelope influences the interaction between the interior and exterior, and therefore how much additional energy a building will need to achieve desired conditions.

- Insulation acts as a barrier between internal and external temperature differences by slowing down how quickly heat can transfer across the building envelope. Thermal resistance refers to a material's or construction's ability to impede the flow of heat. In cold climates for example, higher insulation acts to slow down the loss of heat from the interior of a building, reducing the heating demands and therefore energy usage of the building compared to an envelope with a low thermal resistance.

- Thermal mass is like a thermal battery or capacitor, storing and subsequently releasing heat when exposed to thermal fluctuations. A building has a certain thermal *inertia*, which is affected by both the envelope materials (which directly separate internal and external conditions), as well as internal constructions and objects, which can accumulate thermal energy from internal heat gains and external solar radiation. High thermal mass materials like brick or concrete exhibit a time-lag in responding to external temperature changes, smoothing out internal fluctuations in temperature. For example, a cool slab in the morning will maintain cool temperatures during a hot day, while storing that daytime heat to be released during a cool night. This can influence the energy use of a building by reducing the need for constant and instantaneous intervention from heating and cooling systems.

- An envelope's airtightness will override the resistance and capacitance of the building envelope if air can flow freely between the interior and exterior environment. While this is very useful to reduce cooling demands through natural ventilation when conditions allow, a "leaky" envelope with high levels of infiltration can be the cause of a great amount of heat loss (or gain), increasing energy use dramatically.

Acting simultaneously, insulation, inertia, and infiltration interact in nonlinear ways, and with the added stochastic behavior of internal use and external conditions (which can be considered as boundary conditions), the thermal response and operation of a building is very complex. Building energy models attempt to represent these thermodynamics algorithmically in order to make inferences about building performance and better understand energy use.

1.2 Building energy modeling (BEM)

Building energy models estimate end-use demands given characteristic inputs based on building use and construction, along with climatic conditions to capture the nonlinear behavior of energy demands. While lighting and equipment energy usage are relatively simple to calculate given adequate data (a function of energy intensity and usage), as discussed, the thermodynamics of heating and cooling loads are much more complicated. Several methods exist to estimate heating and cooling energy demands. At a building scale, three main modeling approaches are commonly used:

- *White-box*, or engineering models utilize explicit mathematical equations to simulate a building's thermodynamic response, derived from fundamental engineering principles of conduction, convection, and radiation. Although white-box models are the most accurate, they are highly computationally intensive and are cumbersome to use in applications requiring multiple models or model iterations.

- Approaches to simplify physics-based models are categorized as *grey-box*, or reduced order models. These methods lump parameters such as overall thermal resistance and capacitance by thermal zone, floor, or building to summarize a white-box model while preserving the key underlying transient dynamics. Commonly, grey-box models utilize an electrical engineering allegory, representing a building envelope as a circuit of resistors and capacitors, and internal and external environments as current and voltage sources or sinks.

- *Black-box*, or data-driven models are used to balance speed and input detail. These are statistics- or AI-based models that capture the relationship between selected model parameters and a building's system response. They often require building-level energy data or detailed and contextual modeling data, creating models that are not flexible and often require energy meter data that is not readily available for all buildings in urban areas. All modeling methods are discussed further in 2.2.2.

Each modeling approach has unique operational characteristics, to which the application of the building energy model is highly sensitive. When developing strategies to reduce building stock emissions at urban scales, modeling speed is very important. In many cases, a slow but detailed model may render modeling individual buildings infeasible or unnecessary at urban scales, so choosing an appropriate energy model is crucial, and discussed further in the following sections.

1.3 Urban building energy modeling (UBEM)

Now, considering the energy demands of a city, community, or real estate portfolio from a comprehensive perspective, the technical problem of existing building decarbonization transforms into research- and policy-related questions that focus on macro-scale dynamics of retrofit diffusion. These are related, but not limited to, aggregate carbon emissions, costs, labor, and occupant health [4]. Buildings are not considered individually, and therefore top-down statistical approaches have been common. Drawing from overall energy use, trends at urban scales can appear attractive due to the simplicity and alignment with real aggregate urban energy use. However, while these can estimate the outcome of converting certain groups of building types to another, models rely on extrapolation and interpolation of existing conditions and may not be the most appropriate choice when assessing more intricate energy supply-demand scenarios or when considering the impacts of new technologies. They have also been found to underestimate total energy usage [12], [13]. Thus, urban *building* energy modeling has been established as a key planning tool for modeling future conditions of building stocks. Simply put, UBEMs are a bottom-up urban energy model, using information at an individual building scale to simulate building energy models at urban scales. UBEMs can be applied in areas such as decarbonization target planning, district energy system design, peak modeling, and urban overheating that require problem-solving related to spatially detailed analysis, unseen future conditions, or building-scale alterations [14].

1.3.1 Decarbonizing the building stock

When applying UBEMs to urban decarbonization strategy for building stock emission reduction, the first step, as discussed, is to understand current building conditions and baseline energy use [15]. Once existing conditions are known, plausible retrofits for achieving decarbonization goals can be determined. A multi-family residential building in Florida may benefit from the installation of high-efficiency appliances, while a mid-century office in Minnesota may need to install additional envelope insulation and new, triple-pane windows to see measurable reductions in energy use. The latter is a much more intensive and expensive retrofit than the former, involving significant changes to the building as a whole and focusing on demand reduction to have a more significant impact on energy use. This would be considered a *deep retrofit*, which involves significantly renovating or replacing several elements in a building to radically reduce energy use. On the other hand, *shallow retrofits* are small-scale alterations of aspects, which can occur gradually over time [16]. Some common shallow retrofits are the replacement of fluorescent lights with LEDs or upgrading to a high-efficiency boiler. While shallow retrofits are cheaper and easier, they often do not achieve needed reductions [16]. In general, deep retrofits are needed to attain net-zero status for buildings across North American communities, which involves several barriers with regards to implementation, especially at large scales [1].

Evidently, the cost of deep building retrofits is a primary challenge. Particularly with regards to the distribution of costs and benefits between operators, owners, and tenants [17]. With many cities considering the use of financial incentives like rebates to support widespread retrofits, a detailed understanding of the scope and scale of retrofits (and which retrofits to implement) is necessary to plan for resource allocation [15], [16]. At the policy-scale, operational questions become relevant in this estimation and are all dependent on quantifying the retrofits that must occur to reach high-level decarbonization goals. One example is understanding which buildings should be targeted first and how to phase program roll-outs, or determining labor requirements and gauging gaps in workforce numbers and industry knowledge [17]. Large cities have the resources to commission large-scale modeling for data-driven policy recommendations, but smaller communities (which are much more numerous across the globe) often do not have the same capacity [15].

1.3.2 The need for speed

Even at a single building scale, energy models are computationally complex, posing issues of scalability when applied to urban contexts. So, while UBEMs are necessary for the development of widespread and specific decarbonization policy [4], the modeling process can be tedious and cumbersome due to the time needed to complete a great number of building energy simulations. First, a UBEM must be built and (manually or algorithmically) calibrated to aggregate metered data at annual and jurisdictional scales to create a baseline case, which can require a great number of iterations. Next, the bottom-up benefits of UBEMs are realized through the exploration of “what-if” scenarios of unknown future technological solutions, new retrofit combinations, diffusion scenarios, and detailed policy decisions.

Further iterations can be necessary to address issues of risk. Developing a UBEM requires data and contextual knowledge about buildings, and validating potential pathways require information about construction market trends, which have historically been very difficult to come by without targeted market research and widespread energy audits. So, when considering the risk of poten-

tial emissions reduction pathways, especially in tandem with the long-term commitment of capital improvements, iterations of both baseline and upgrade scenarios can be highly beneficial [18].

Finally, systemic challenges call for systems-based questions and integrated approaches to the problem definition process, a key first step towards solution development. If a UBEM takes several hours to simulate, an iterative and collaborative design process is virtually impossible, especially if jurisdictions wish to explore interactions across different systems and models, such as the electricity grid or transportation networks. With many emissions reduction targets targeting change within the next 25 years, we are running out of time for implementation, let alone planning. *Rapid* building energy modeling is therefore needed at urban scales.

1.4 Thesis statement and concept

From a desire to make tools that increase the accessibility and effectiveness of building stock decarbonization planning, the focus of this work is to develop a foundational model for accelerating urban building energy modeling. The kernel of such a framework is the ability to accurately and quickly estimate seasonal energy use of buildings at an urban scale and beyond, anywhere in the world. With a novel combination of speed and robustness in an energy modeling engine, new opportunities for analysis methodologies become available to the energy modeler, while at the same time, urban building energy modeling can become accessible to a wider range of users. The thesis is centered around three core aims that span across policy, usability, and technology:

1. To understand current barriers and bottlenecks in developing actionable solutions and pathways for urban building energy use reduction and related policies;

2. to specify the technological needs of a UBEM to empower collaborative and systemically informed urban building stock decarbonization policy-making; and,

3. to develop a data-driven accelerated BEM method that is agnostic to geometry, climate, and building program, which can reduce the cost and time-to-solution for UBEMs by orders of magnitude.

In 2, we outline the state of current research and data-driven methods to approach urban building energy modeling for large scale building decarbonization interventions. We identify several gaps in research to be addressed through this work. In 3, we establish the functional needs of the proposed methodology, drawing from past work, experience, and tacit knowledge to present a set of guiding technical specifications for the model and its proposed design. In 4 and 5, we describe the methods in which the model and methodology were developed, built, and tested through a rigorous validation process across multiple scales. A final validation process is completed with a large national dataset of realistic residential building energy use profiles. We benchmark the model against data drawn from distributions which are known to accurately reflect the real building stock and provide a test of robustness against model misalignment. Finally, in 6 we discuss the performance of the model as it fits with the core goal of enabling decarbonization decision-making, and present several detailed use-cases for future work and extended applications.

Chapter 2

Current state of research

This chapter provides a review of influential bodies of literature that are foundational and relevant to this work, as well as highlights current gaps in research. Use-cases and validations of UBEMs are first summarized. Then, an overview of several approaches to accelerate building energy models at several scales is presented to demonstrate the widespread need and applicability of improved and rapid methods.

2.1 Validating the UBEM approach

If there is adequate information about the composition and distribution of the building stock, UBEMs are deemed a reliable method for testing scenarios at urban scales in unknown conditions (such as changing building use and construction characteristics or future climate scenarios) [19]. However, not all UBEMs have the same modeling methodology. The simplest UBEM follows a "floor-area" approach evaluating energy use intensity (EUI) for building groups. This is accomplished with an energy model of a simple south-facing "shoebox," to be multiplied out by the total urban floor area. While this may be adequate for rough estimation, applications often demand energy models which are sensitive to more than building program and scale [14]. In particular, urban context has been found to be an important aspect of building energy use, including building geometry, building-to-building interactions, and self-shading. [20] evaluated the importance of shading on thermal energy demand in a Mediterranean climate. They found that an

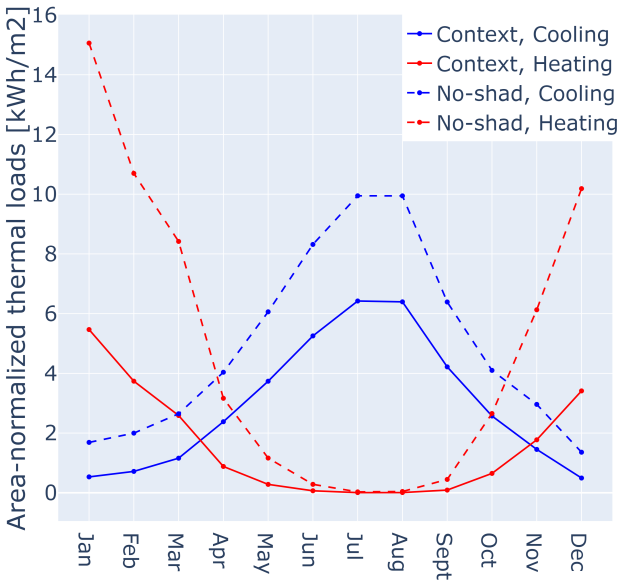


Figure 2.1: Monthly loads with and without shading context. Both scenarios utilize full energy models which include building geometry (as opposed to an area-based method).

isolated building energy model (with no shading considered) resulted in errors of up to 18.0%, while the consideration of inter-building effects reduced the error to 2.5%, concluding that inter-building shading effects should be modeled in urban (dense) contexts. This aligns with [21], and the author's own experimentation of a UBEM test region located in a temperate climate. Illustrated in Figure 2.1, the impacts of context are clearly seen year-long, increasing both heating and cooling peak seasonal loads due to shading. Therefore, geometric context along with known construction and use information point towards a need to maximize detail in UBEMs. Yet, their usefulness as an interactive and iterative tool demands quick computation, as discussed in 1.3.2. To date, the authors are not aware of an efficient enough building energy modeling methodology nor the data availability to accurately accomplish this, emphasizing that bottom-up models are a balancing act between fidelity and usability. The combination of data scarcity and the computational burden of UBEMs requires the use of model complexity reduction techniques and approximations. Accordingly, selecting the correct building energy model becomes a delicate balance between accuracy and allowable computation time. Some common approaches in BEMs and UBEMs are discussed below.

2.2 Model simplifications

Complexity reduction of building energy modeling is a large area of work. Spanning across the urban and building scale, several methods of simplification and acceleration are discussed in this section.

2.2.1 Archetype segmentation

A pervasive (and often necessary) approach for reducing the complexity of UBEMs is through the development of building archetypes. Due to market trends, building codes and other macroeconomic factors, building constructions are similar across time periods and geographic regions [11]. This property can be leveraged through the clustering of buildings into representative programmatic groups. Illustrated in Figure 2.2, this process increases model efficiency and addresses incomplete data, utilizing characteristic similarities of buildings [22], [23]. Divisions are often (and ideally) based on program and age, but can increase in complexity with data availability. For example, floor area, basement and attic characteristics, and census-related data have been found to affect energy use [24].

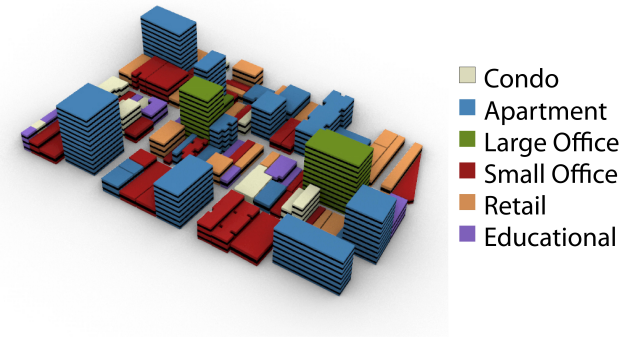


Figure 2.2: Urban archetype segmentation by typology, illustrated.

This approach has been standardized for widespread regional use through the development of regional archetypes, such as the European TABULA project and the US National Renewable Energy Laboratory (NREL) commercial reference buildings [11], [23]. These datasets and build-

ing typologies have been widely used for UBEM applications across Europe and North America. Broadly following historic construction standards and events, the US NREL commercial reference buildings are representative of more than 60% of existing commercial buildings, informed by historical releases of national building standards [11].

Archeypal segmentation is useful and necessary in UBEMs, both due to data limitations at the development stage and for ease of policy intervention planning at the analysis phase [17]. Additionally, [25] found that in many cases, highly detailed UBEMs are unnecessary, especially when considering a largely homogeneous building stock in a heating- or cooling-dominated climate, where large chunks of the building stock are very similar, and therefore call for similar retrofits. This can simplify the retrofit selection process, as heating- or cooling-dominated climates tend to have clearer interventions related to reducing one particular energy end-use across all buildings, regardless of type. However, when these conditions are not present, archetypal segmentation at a finer resolution can be called for.

Additionally, when clustering a large number of buildings into a subset of archetypes, identical building conditions can result in a strong bias towards specific energy end-use distributions, and identical usage schedules applied across the entire building stock can result in amplifications of peak loads [14]. Therefore, a key component for realistic UBEM analysis is increased stochasticity of building occupancy and use schedules, which in turn increases the number of individual energy simulations needed for a UBEM model [26].

With increasing computational resources and access to building stock data, UBEMs can be modeled at increasingly higher levels of detail [27]. For example, NREL's ResStock and ComStock databases characterize the national building stock by construction decade, census zone, building type (program), stories, mechanical system, and construction type [28]. Models are created by sampling conditional probability distributions for building characteristics throughout the country, resulting in unique building profiles in almost all modeling scenarios. Even though this high granularity increases computational time and model complexity [24], [27], the underlying assumption is that simulation results will become more accurate. So, there is a need for more computationally efficient ways of running BEMs capable of granular expression of the wide range of thermodynamics seen in buildings.

2.2.2 The building energy modeling engine

Choosing an appropriate building energy model is crucial in determining the accuracy, scope, detail, and importantly, the usability of a UBEM [14], [19], [29]. White-box models have the advantage of functioning with essentially any inputs of building geometry and construction, and can be used to model the energy demands of any building in any climate. A prominent physics-based model is the U.S. Department of Energy's EnergyPlus, developed as a multi-zone heat balance model based on thermal dynamics principles [30]. Other commonly used physics-based simulation engines are TRNSYS, IDA ICE, and Modelica, which are implemented across several UBEM studies and tools [31], [32]. However, the computational requirements of bottom-up approaches can inhibit large-scale modeling and complex building geometry due to time constraints, and simulations with short time steps due to instability [13].

The heat-balance modeling methodology

EnergyPlus is based on a heat-balance and state-space method, in which all loads in all zones and systems at each timestep are calculated using a heat balance method [33]. From thermodynamic principles, the heat balance method assumes that the rate of heat gain is equal to the rate of heat loss at the given point in time, across a given building element or surface, or equivalently add to zero. Heat transfer is physically multi-dimensional, but in whole-building energy models, it is assumed to be unidirectional across each model element. The factors affecting heat transfer modeled in EnergyPlus include conduction through the element (q''_{ki}), convective exchange between the element's interior surface and the air (q''_{conv}), internal radiant exchange (q''_{LWX}), solar radiation through windows (q''_{sol}), and radiant heat gain from internal lights (q''_{SW}) and equipment (q''_{LWS}). This is captured in a heat balance equation for the internal face of a given surface at a single point in time:

$$q''_{LWX} + q''_{SW} + q''_{LWS} + q''_{ki} + q''_{sol} + q''_{conv} = 0$$

Simultaneously, the interior state of a zone must follow the energy balance method as well, interacting with all surfaces enclosing the zone as well as the elements within it (lights, people, and equipment) and the active systems interacting with it (ventilation, heating and cooling).

Incorporating the transient nature of the system, conduction through building elements, q''_{ki}, is modeled as an infinite series of temperature histories at the interior and exterior surface, with

$$q''_{ki}(t) = \sum_{j=0}^{\inf} Y_j T_{i,t-j\delta} - \sum_{j=0}^{\inf} X_j T_{o,t-j\delta}$$

Where T is the element temperature on the inside, i, or outside, o, surface, at time, t, and history, j, for timestep, δ. The coefficients X and Y are response factors and must be calculated as properties of the construction and fluctuating boundary conditions. Of course, an infinite number of terms is not desirable, so there are several ways to approach this calculation. A main method utilized in EnergyPlus are Conduction Transfer Functions (CTFs), which enable the approximation of the infinite series based on methods proposed by [34]. Other approaches may use a variety of numerical methods, such as finite element analysis (FEA), Laplace transforms, or the finite difference method (FDM); the latter is prominently integrated into EnergyPlus as a secondary solver and can be particularly useful in the modeling of phase-change materials [33], [35]. So, while there are multiple ways to capture the detailed thermal response of building elements, all of these take time, both literally and figuratively, to solve.

Simplifying the heat-balance method

To address the efficiency of thermodynamic calculations for white-box models, the radiant time series method is commonly used to calculate heat transfer through the building envelope. It utilizes predefined and calculated conduction and solar heat gain coefficients to speed up transient calculations, limiting the temperature "history" of a building element to 24 terms (usually hours). However, this approach is constrained to certain simpler building materials [36]. [35] proposed the use of power series expansions to calculate the CTF coefficients in a much more efficient way. The Spawn-of-EnergyPlus model was designed to support HVAC controls modeling. Built upon the

fundamentals of EnergyPlus, the modeling process is broken into a set of components that can be calculated in parallel, utilizing existing complex system modeling frameworks [37]. This prototype allowed for up to a 46% decrease in computation time.

As discussed in 1.2, multiple methods have been utilized to reduce the computational effort required by engineering models, while maintaining their robustness. Reduced-order models simplify building thermal processes which can be calculated at a much faster rate than energy-balance methods. For example, the Resistor-Capacitor model. Some applications in UBEMs are CitySim solver, ISO standard 13790, and FastBuildings [38]. Reduced order model calculations are based on state-space calculations, transfer function models, and finite difference models [33], [39]. While reduced-order models are useful in many ways to determine instantaneous energy loads, there are limitations as to their accuracy in predicting the impact of thermal mass on energy use, as well as the calculation of solar gains under varying incident angles [13], [39].

On the other hand, black-box models have proven to capture the transient response of buildings, but lack intepretability and flexibility. Current models often require building-level energy data or detailed and contextual modeling outputs, creating models that are immutable and reliant on data that is not widely available [13], [31]. For this reason, most applications of black-box models have been limited to specific geographic and programmatic contexts [40]. On top of this, even if a model is developed with climate or construction-based flexibility in mind, current grey- or black-box models are constrained to certain building forms and layouts, thus limiting their applicability to be repeated for large numbers of unique buildings. In this case, the white-box method of modeling heat transfer through individual building elements is a robust method that allows for flexibility in virtually any building form, use, location, and construction, when scaled up to urban models, this approach becomes prohibitively slow and cannot effectively be used in interactive exploratory analyses. So, recent work in geometric abstraction has allowed grey- and black-box models to achieve this geometric flexibility in other ways.

2.2.3 Geometric abstractions

Core-perimeter models

The use of multi-zone building energy models has been established for decades: floor plans are divided into a series of representative thermal zones, ideally following realistic room layouts. When detailed internal layouts are unknown or modeling specific internal zones is deemed unnecessary, buildings are divided into "core" and "perimeter" sections, as per national standards [41]. Many calculation methods (including those used by EnergyPlus) assume perfect mixing within thermal zones. So, modeling buildings with multiple thermal zones is needed to eliminate the assumption that an entire building floor is thermally "well mixed". Divisions between perimeter and core areas reflect the differences between demand in internal core areas compared to areas with high access to climatic and solar interactions (facade-adjacent) [42]. Especially, when a floor plate contains partitions and objects which constrain convection. The perimeter regions, typically 4.5-6 m (15-20 feet), are divided based on orientation, to capture the temporal aspects of solar radiation. The core zone(s) are simply modeled as the leftover internal spaces. This algorithmic process of splitting up core and perimeter areas is often referred to as "autozoning," and is illustrated in Figure 4.2 [29].

Shoeboxing algorithms

Multi-zone models can become incredibly complex with irregularly shaped and numerous buildings and are currently impractical at urban scales. The shoeboxer algorithm was originally developed to reduce function evaluation time and enable urban building energy modeling while maintaining the benefits of zone division [43]. Rather than running detailed energy models for entire buildings, each building is represented as a linear combination of representative shoebox energy models. Each has a single core and perimeter zone, as seen in Figure 4.3. For a single building, the shoeboxing algorithm typically involves reducing full building geometry into a selected number of representative "shoeboxes" with the following steps: (1) Discretize the floor plate into $n + 1$ zones, where n is the number of perimeter zones, typically 1 per facade edge. (2) Compute summary features for unique floorplates, e.g. core-to-perimeter ratios, facade-to-floor area ratios. (3) Place a sensor grid of nodes over the building's facade to represent candidate locations for shoebox allocation. (4) Run solar irradiation analysis for each facade face's sensor grid. (5) Cluster sensor nodes based on annual irradiation. (6) Select one representative sensor node per cluster. (7) Allocate one shoebox at each cluster's representative node. (8) Derive shoebox geometry according to summary floorplate features. (9) Identify the shading context for each shoebox. (10) Simulate shoeboxes. (11) Compute shoebox weighting factors according to sensor node counts per cluster and building. (12) Aggregate shoebox results according to weighting factors per building. After reducing buildings to a standard geometric shape, further clustering based on annual irradiance levels allows for a significant reduction in computations needed for a UBEM. This methodology has been used in several urban and building energy modeling contexts in [44], [45], with similar approaches applied by [46]. Although this approach allows for significant computational savings, with an optimal acceleration of 150 times, large urban models can still result in long computation times. However, geometric abstraction can be a key factor in building a robust model, explained in 4.

2.2.4 On the accuracy of UBEMs

At building scales, the ANSI/ASHRAE Standard 140, the *Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs*, is widely used to measure simulation error in a directly comparative manner. Errors within 15% CVRMSE and ±5% nMBE are deemed acceptable [47]. To measure error in the first place, an accepted "ground truth" must be determined. In recent literature, EnergyPlus is deemed a reliable baseline. ASHRAE 104 was used to validate EnergyPlus itself, with the BESTEST method, which validated EnergyPlus against several other white-box simulations in a variety of contexts [48]. We follow a similar approach, with an even wider scope, by evaluating the proposed methodology in a variety of geometric and climatic contexts.

Due to the frequency of near-zero demands in energy models, [49] investigate the adequacy of BEM statistical metrics for system-level analysis. The authors suggest that the ASHRAE-suggested coefficient of variance of root mean squared error (CVRMSE) and mean bias error (MBE) are not sufficient alone, and that the addition of R^2 and the range normalized root mean squared error (RN_RMSE) can be more meaningful in evaluating the performance of energy models. However, in the evaluation of a new energy model, particularly for use with near-zero energy buildings, these proposed metrics could put too high of an emphasis on the accuracy of high values or overall scale of values rather than directly evaluating model accuracy in a comparative manner. [50] alterna-

tively find that CVRMSE is the most robust approach for evaluating white-box energy models in calibration applications. Therefore, in this thesis, we follow both the recommendations of ASHRAE 140's calibration convergence tolerance (with some modifications due to the project scope) in the validation and testing of the proposed methodology.

In general, there is little alignment nor recently updated standards on acceptable tolerances for error across urban building energy modeling [51]. [51] found that in some widely accepted UBEM methodologies, a single building's energy use may have a percent error upwards of 1000%, but when aggregated to urban scales, error can be as low as 1%. This uncertainty and lack of standardization is in part, due to the fact that detailed data on building energy use has been historically difficult to obtain (nor possible for new and unseen contexts) [40], but also because error tolerance is largely due to application [1]. [51] note that, in addition to the selection of appropriate measures of error, UBEM accuracy lies across the axes of spatial and temporal resolution. For example, at city-scales, sub-hourly energy usage is not relevant in the determination of viable retrofit packages, but it may be for electricity demand modeling. [52] explores the importance of accuracy in early-stage design of low-carbon buildings through the use of a *similarity threshold*, finding that simplified grey-box models can be a viable energy modeling engine for the design of sustainable architectural strategies in early-stage design. Similarly, context can be very important. When developing retrofit pathways at urban scales, [25] notes that in building stocks with homogeneous building archetypes and a clear heating- or cooling-dominated climate, identifying solutions to reduce energy use can be much simpler than more nuanced climates and contexts. A temperate climate may introduce trade-offs between heating and cooling energy conservation measures, for example, triple-pane windows with a higher solar heat gain coefficient may lower heating demands, but could also increase cooling demands in the summer without adequate ventilation and shading. So, findings within literature highlight the importance of calibration and sensitivity testing of bottom-up models [51], [53], [54]. Thus providing further incentive to increase the speed of calculations which drive UBEMs. But, as functionality is, to date, highly sensitive to computation time, there must be a trade-off between error tolerance and speed in UBEMs. In this work, we attempt to address this barrier through data-driven methods.

2.3 Machine learning in BEM and UBEM

With advancements in machine learning (ML) and increasing data availability, ML surrogate models have been increasingly implemented for rapid calculations of UBEMs. An ML surrogate, referred to simply as "surrogate model" in this thesis, is a black-box model designed to reproduce the outcomes of a system given a set of design parameters when engineering models cannot be readily used, often due to computational burden [40], [55]. In UBEMs, surrogates may replace the building energy modeling engine, which is either input- and physics-informed, like EnergyPlus [56], or output-based, trained on real energy meter data for specific sets of buildings [57], [58]. [59] takes a combined approach. As ML models have difficulties extrapolating outside of the scope of training data, often producing irregular results in regions of sparse data, the use of meter-based models is limited to existing conditions, which lack the flexibility needed at the core of UBEM studies: unseen geometric,

[1] It is important to note that there is an additional layer of error when comparing to *real* building energy use, due to lack of data about building stock. However, as this thesis considers these factors to be inputs (e.g. building envelope composition and occupancy), this source of error is not considered.

compositional, and climatic changes [12], [55], [60].

In the interest of developing a flexible model, we focus on physics-based surrogates since they have the theoretical opportunity to model any building and climate condition without the need to rely on high-quality data availability, so long as the desired building conditions are expressible within the design space's representational framework. Several ML architectures have been explored in UBEM applications, the most common being neural networks, support vector machines, gradient boosting, and random forest models with promising accuracy [40], [53], [58], [59], [61], [62]. While several studies have successfully modeled multi-use and multi-zone UBEMs, most rely on site-specific training datasets with static geometry and climate. General limitations related to ML models are also observed, such as a lack of interpretability, especially as it relates to correlations between inputs, overfitting, and a need for standardization in data-driven energy model inputs, model selection, and feature engineering [12], [61], [63].

In an attempt to address these limitations of immutability, [56], present a climate-independent surrogate for annual energy use. Employing a deep temporal convolutional neural network to process multivariate hourly weather data, the proposed network extracts relevant features for estimating heating or cooling demand. These location-specific weather features are then combined with building design parameters and fed into a single surrogate model (a feed-forward neural network). The study utilized 569 weather files from various Canadian locations, across 5 ASHRAE climate zones, demonstrating less than 3% error in the surrogate model's predictions for annual heating demand for new building designs outside of the training data set. [56] note that the generalizability of a surrogate model relies on the parametrization of the problem; more detailed inputs (geometry, window sizes, construction parameters, climatic variables, etc.) allow for more flexible applications, but also require significantly more effort and data for training. Therefore, similar to the use of archetypes in UBEMs, it can be beneficial to utilize abstractions and simplifications which minimally impact the accuracy of an energy model prior to training. For example, the use of geometric abstraction into similar modules with the shoeboxer algorithm, as is applied in the proposed methodology.

2.4 Identified research gaps

Through a comprehensive review of the state of current research, several research gaps were identified, which provide a basis for future work. These are summarized below:

- Many UBEM approaches do not adequately consider urban context, especially with regards to geometry and shading.
- However, UBEMs that are detailed tend to be so computationally and data-intensive that they limit their own usability.
- There is a lack in standardization across UBEM and BEM validation and methodology development.
- Simplification of urban building energy models through archetypal clustering can introduce bias in baseline energy results, which can be transferred to selections of upgrades in building-stock decarbonization pathway development and cause the overestimation of peak loads, suggesting a need for UBEMs to support high-resolution timeseries inputs.

- Current engineering-based (bottom-up) BEMs are too computationally expensive to be used in urban contexts.
- On the other hand, black-box methods are valid only within the contexts and bounds in which they have been trained, often being overly specific to certain use cases and therefore not reusable or applicable for widespread adoption (especially for modeling future scenarios).

While the potential impact of rapid building energy models is vast, the parameterization and scope of the black-box method should be considered alongside accuracy and is critical to the development of tools that have a potential for widespread adoption. This thesis attempts to address limitations of current methods with a novel shoebox energy surrogate model that is valid in a variety of climate and building parameter contexts while maintaining granularity of results at a monthly level and accuracy comparable to physics-based models at the urban scale. Taking a similar approach to [56], we utilize a multi-modal architecture for processing the time-series portions of the input vector separately from the unstructured data in the initial layers of the surrogate, introducing the incorporation of building use and operation schedules and a vigorous synthetic data generation methodology. The surrogate will integrate into, and accelerate existing urban modeling methodologies, such as the UBEM.io workflow [19].

Chapter 3

Establishing a need for rapid monthly modeling

Instead of approaching this methodology from a "maximalist" perspective (attempting to maximize detail, accuracy, and speed simultaneously), we approached the concept of robustness from a "user-first" perspective. Understanding the roles and questions of key stakeholders first through existing work, tacit knowledge, and personal experience, the needs and goals of the urban retrofit planning process were converted into specifications to define the scope of this model. The needs of the urban building energy modeling team can be used to generate the demands on a new accelerated foundational UBEM framework.

Discussed in the above sections, an ability to run models quickly is imperative for efficient and comprehensive retrofit studies as well as collaborative work. If, during the exploration of retrofit pathways a UBEM simulation takes multiple hours, the policy-making process and the ability to iteratively explore options will be greatly limited. Not only are there an essentially limitless number of interventions that can be tried and tested for a UBEM, but there are a multitude of stakeholders which are needed to develop and evaluate the proposed options in the UBEM process. So, the ability to implement changes or scenarios and immediately see their impact on urban-scale results is highly important in enabling collaborative decision-making. In an effort to increase the accessibility of UBEM tools for cities to develop decarbonization plans, [15] has developed an 8-step framework for cities to build, validate, and explore a baseline building energy model and potential retrofit pathways at a fraction of the cost of traditional methods. Within this work, the author has identified three key roles needed in the UBEM and retrofit pathway process: a project champion, driving the decarbonization agenda; a Geographic Information System (GIS) specialist, providing city building characteristics and geometric data; and an energy modeler, who is well versed in building energy modeling as well as local historic energy policy. Each role is critical to developing (and importantly, scaling and implementing) an urban decarbonization initiative, but their level of expertise with regards to building systems and technical approaches to decarbonization are varied, presenting a unique challenge in the development of tools and methods for retrofit modeling methodologies.

Through workshops with cities across the globe, [16], [64] achieve an understanding that the process of developing an effective retrofit plan (effectively) is highly collaborative and iterative. As mentioned in 1, building decarbonization policy-making cannot be siloed in engineering solutions, but must be accessible to those involved in implementation and regulation. Tools should be intuitive accommodate the appropriate level of detail for the diverse set of stakeholders; comprehensive

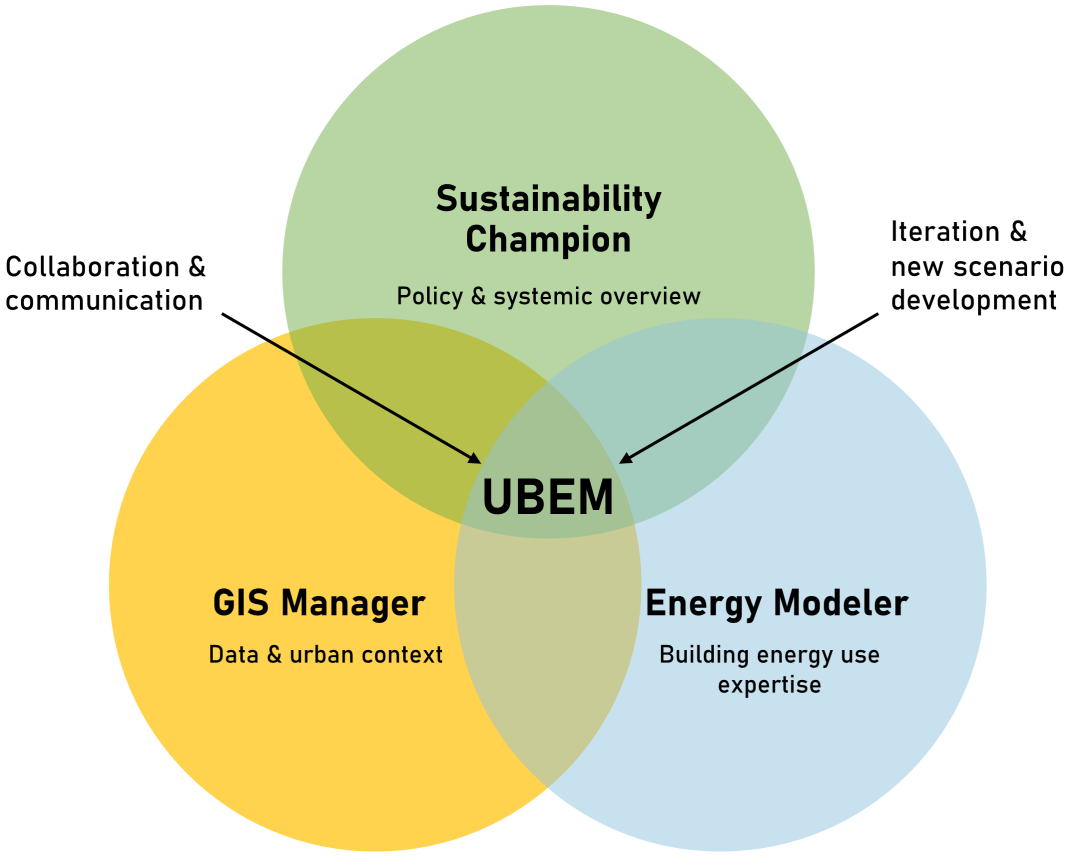


Figure 3.1: Typical roles present in developing and utilizing a UBEM for decarbonization policy and retrofit planning.

enough to allow skilled energy modelers and engineers to implement detailed retrofits at a building scale; flexible enough to function easily in a wide range of climates and contexts; and fast enough for testing and learning from results to be an iterative process, leaving time to focus on the wider impacts and systems at play. From an engineering or energy modeler perspective, the ability to model multi-use buildings or consider behavioral change (for example, occupancy sensors for lighting, or peak-load reduction of equipment with load shifting) have been a previously difficult task in UBEMs, as these are programmatic and have hourly variations. Most current black-box models do not allow for changes in use patterns (or schedules), being tied to a particular representative program (like a residence or office), therefore eliminating the possibility of directly measuring savings from these interventions. When they do incorporate programmatic variations as part of the design space, it might typically be as categorical parameters, due to the fact that the underlying model inputs are high dimensional time series. This does not allow for the full range of behaviors, new behaviors, or the elimination of coincident peaking.

From a policy perspective, the ability to understand macro-scale results of engineering decisions, as well as develop strategic groups for staged and organized implementation which acknowledge the unique and diverse composition of cities is critical in the UBEM process [15]. So, with this level of detail desired, the ability to introduce speed into this process is only useful if model

results can be validated. However, there is a point at which additional detail could be considered excessive. For example, in the development of archetypes, simplification can prove highly useful in UBEMs. Archetypes can simultaneously make the modeling process more efficient while also streamlining interpretation of the model's results, and can enable bottom-up energy models even when there is limited data. In finding a balance between detail and speed, it is important to establish (1) the maximum delay in model simulations and iterations, and (2) the minimum viable output to make informed decisions about retrofits. Both act to define and determine the structure of the proposed methodology. [52] found that in many cases, a simple shoebox can provide accurate results in exploring low-carbon design options, particularly with regards to shading, but the temporal response of the building envelope is significant. This approach of trying to maintain a high information-to-model complexity ratio underlies most urban-scale and policy-driving methodologies. For instance, the authors have previously made the decision to trade higher fidelity for higher speed when developing accelerated algorithms for photovoltaic potential estimation which is sensitive to urban contexts [65].

A focus on selecting the appropriate level of detail for the urban modeling context guides many key decisions and architectural choices detailed through this thesis. As such, we begin with an attempt to review some of the functional needs of potential end users and the resulting technological requirements. We question the need for detail where possible to distill the urban energy model into its most functional parts. For example, discussed in 2.1, shading has a significant impact on energy use. But while most approaches utilize time-consuming radiation calculations or ray tracing methods, the nature of black-box models can be leveraged to simplify inputs precisely to what is easily accessed by and intuitive to users, such as simple geometric definitions of shading maps.

In a similar vein, many UBEM approaches utilize annual energy use, or annual end-use demand to evaluate building performance, as hourly values are not particularly necessary for selecting building upgrades. However, the seasonality of building use and thermal response is highly important across several facets of the urban energy model [7]. It is particularly important in the development of a baseline model when calibrating the selected archetypal representations of the building stock. Described further in 6.1 in the context of detailed single-building autocalibration and in [53], calibration involves the refinement or selection of unknown building parameters (such as insulation levels or lighting power density) through a fitting process to overall energy or fuel use. Usually, multiple parameters are calibrated at once and there are therefore multiple solutions to a building definition which converge to a given energy use profile. But, in many cases, detailed energy data is not known at urban scales, yet annual values are not suitable for selecting building characteristics [14]. For example, lowering internal loads from lights or equipment can reduce heating loads, but so can higher levels of insulation. Annual fuel use would not show a difference between the two solutions, but with monthly data, internal loads would increase cooling needs in the summer, while insulation would decrease it. Calibration is always an ill-posed problem, but having access to the seasonal peaks and contours helps reduce the size and complexity of the target's inverse design solution space. Similarly, the selection of upgrades can be sensitive to seasonality as well, so a monthly timestep for results is deemed appropriate for urban scale studies.

Table 3.1 summarizes the identified needs for UBEMs in retrofit decision-making, and their translation into specifications in the development of a tool, which were considered throughout the process of defining and developing the methodology presented in the following sections.

Challenge or Functional Need	**Technological Specification**
Iterate and collaborate	Near-real time response
Geometric context	Consider context shading
Geometric flexibility	Representation of buildings as syntheses of discretely modeled parts
Programmatic flexibility	Use schedules and intensities data as input (occupancy, lighting, equipment)
Future climate and location-independent	Hourly weather data as input
Accessible	Cheap-to-free, easily deployed and consumed in cloud architectures
Actionable level-of-detail	Monthly fidelity, urban accuracy

Table 3.1: Functional needs translated into technological specifications.

Chapter 4

Methods

Developing a model for rapid UBEM evaluation which can be deployed globally presents challenges which are not present in implementing typical function approximations for single-building energy models. In addition to incorporating common building energy model variables (like envelope construction, power densities, and thermostat setpoints), the model must be robust to features which might normally be fixed. In particular, the features which describe program and form must be flexible to accommodate the wide range of geometry and programmatic uses present in a city, while environmental conditions (including both urban shading context and climatic timeseries data) must also be incorporated into the model's domain to enable inference anywhere in the world.

We leverage a geometric abstraction approach which represents a building as a linear combination of discretized shoebox energy models, extending aspects of existing urban building energy models developed in [43]. While developing a fixed-length geometric parameterization of building geometry which can represent any building in the world would be a significant challenge, individual shoebox geometry can be represented with 8 intuitive parametric features. In this implementation, shoebox generation and allocation is then driven by the building footprint and height alone.

The challenge of achieving climatic robustness is due, in-part, to the non-linear responses of buildings to their environments over the course of a year. Typical weather files for energy models include several timeseries channels of 8760 hourly values over a year, rendering many function approximation techniques infeasible due to mathematical complexity which may grow with input size. While some approaches might rely on feature engineering to reduce weather data to summary statistics or other compressed representations, we instead elect to follow a similar approach to [56], using deep convolutional neural networks to process the raw timeseries. This architecture additionally enables incorporating varying building programs independently; not as categorical inputs but instead by using full timeseries schedules for occupancy, operation of equipment, and use of electric lights. As discussed in 2.2.2, heating and cooling are the most complex to estimate due to numerical method bottlenecks, making up the bulk of the computational cost of bottom-up building energy models. Therefore, the proposed methodology focuses only on these end-uses, to be later combined with energy use due to lighting, equipment, and other linear functions. It is recognized that values such as lighting and equipment have effects on heating and cooling loads due to internal heat gains, so these values are considered as inputs.

A dataset of 1.25 million shoeboxes is generated to train the neural network surrogate to predict monthly cooling and heating loads for shoebox energy models, which is to be used as a replacement for EnergyPlus simulations within the larger urban building energy modeling pipeline.

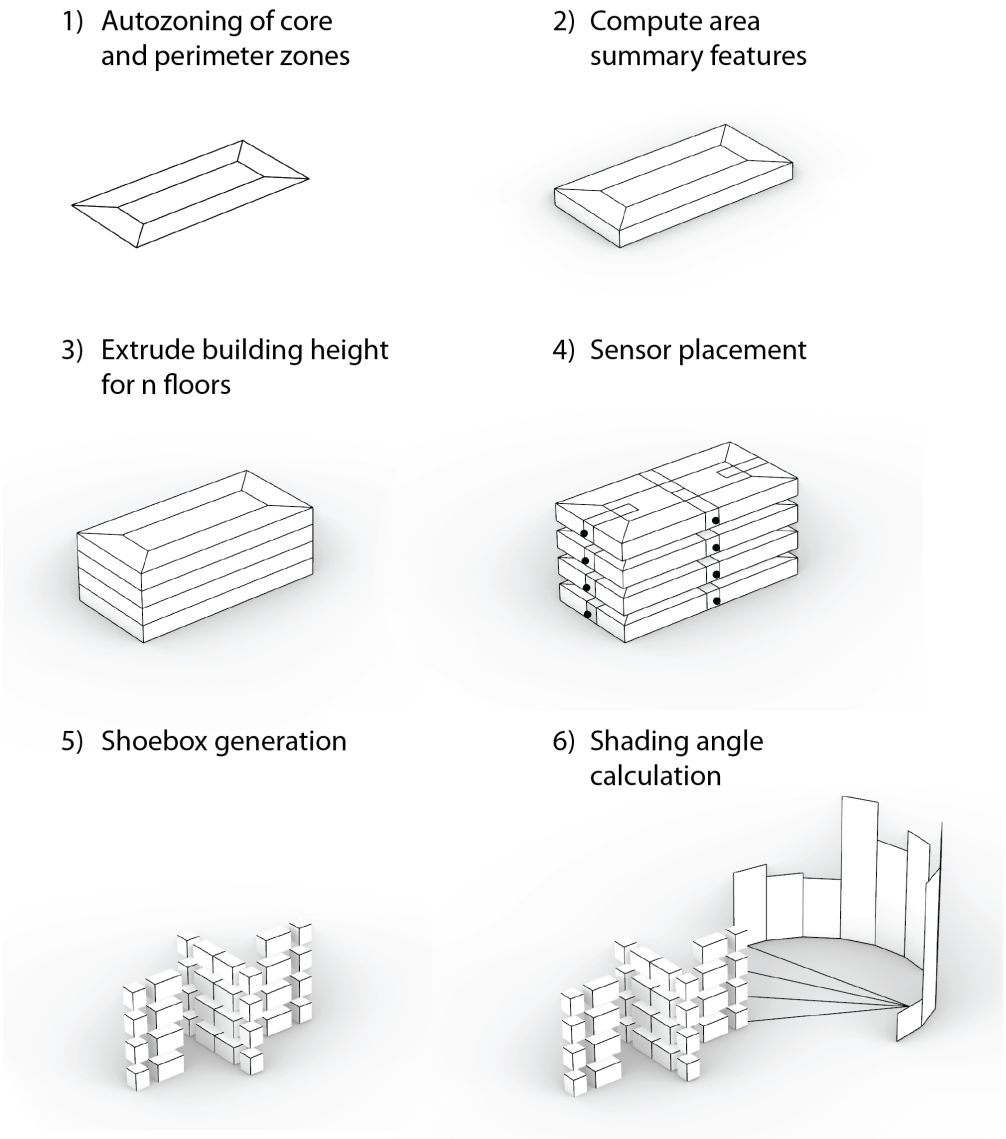


Figure 4.1: Shoebox allocation process

4.1 Representing buildings

We integrate and extend components of previous shoeboxing algorithms to develop a methodology that is encodable for a neural network while simultaneously capable of expressing the geometric dynamism of urban contexts. Irradiation-based clustering seen in [43] and other shoebox-based approaches are primarily motivated by the need to reduce the number of function evaluations, as total time-to-solution without clustering quickly becomes infeasible with multiple buildings. Although distributed computing tools unavailable at the time the algorithm was developed can now be used to run thousands of shoeboxes in parallel and at least reduce the real-world time to solution, total compute time (and thus cost) remains high, even if time-to-solution is reduced. However, with the multiple orders-of-magnitude performance factor of the surrogate model developed in subsequent sections, we elect to drop the shoebox clustering analysis and follow a simpler and more granular allocation strategy which enables placing vastly more shoeboxes per facade. Dropping clustering also facilitates some of the extended applications discussed in 6.1, including allocating multi-program buildings within UBEMs and performing energy demand intensity mapping over facade discretizations in early-stage design.

4.1.1 Shoebox weighting

Given a building with gross floor area G, for each floorplate with area A_i we assign a weight $a_0, a_1, \ldots, a_{m-1}$, where m is the number of floorplates and,

$$a_i = \frac{A_i}{G}$$

When a building is represented as a simple extrusion of a 2D footprint as is common in GIS-to-UBEM workflows, the following can be calculated once per building. The weights are reduced to a constant for all building floors:

$$a_i = \frac{\text{footprint area}}{G}$$

For each of the n_i edges on the i-th floorplate with total perimeter L_i, we sub-divide the j-th edge into $p_{i,j}$ arbitrary segments of length $L_{i,j,0}, L_{i,j,1}, \ldots, L_{i,j,p-1}$ and assign a weight $b_{i,j,k}$, with i indexing floorplates, j indexing edges, and k indexing the subdivisions per edge:

$$b_{i,j,k} = \frac{L_{i,j,k}}{L_i}$$

We allocate one shoebox $S_{i,j,k}$ per edge subdivision at the edge midpoint, with a final weight given by the following:

$$c_{i,j,k} = a_i * b_{i,j,k}$$

When no subdivisions per edge are used (as demonstrated in subsequent sections), the $b_{i,j,k}$ weight simplifies to a simple edge weight.

Given an energy model, $\mathcal{F}$, which operates over shoeboxes, $S_{i,j,k}$, and returns area-normalized energy usage, we can define the entire area-normalized energy usage of the building, E, as a linear combination of the constitutive shoeboxes.

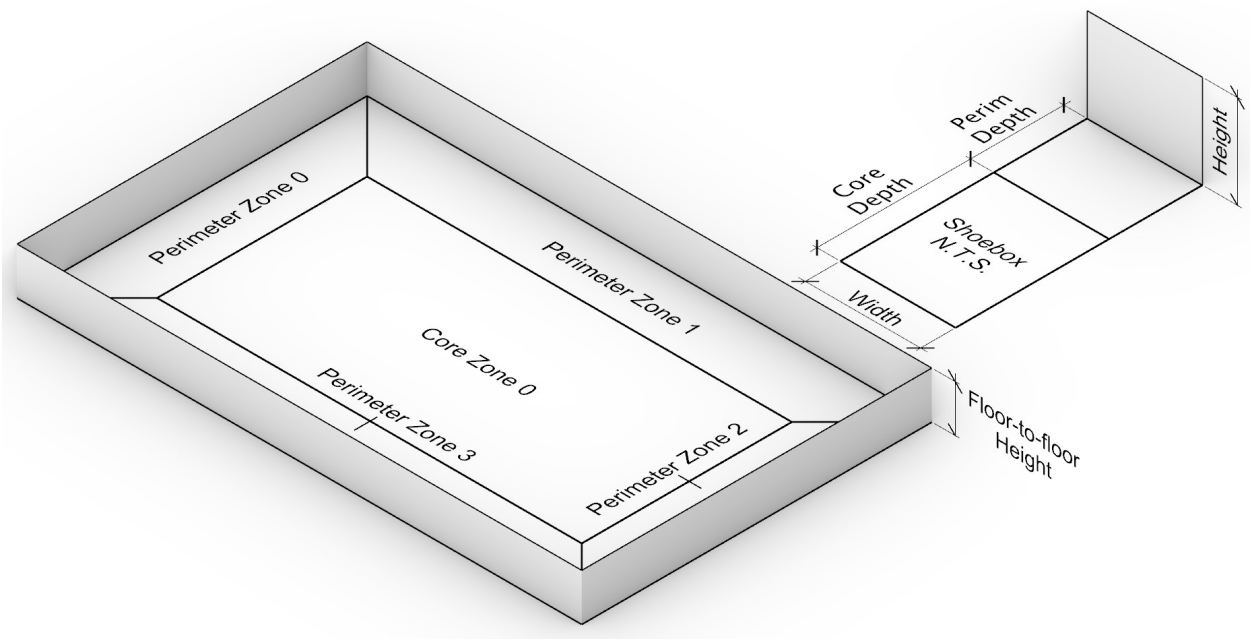


Figure 4.2: Shoebox geometry extraction from auto-zoned floor plan

The energy model $\mathcal{F}$ first returns area-normalized energy for the perimeter and core zones; then these are weighted by perimeter and core area to yield the whole shoebox's area-normalized values before the edge weighting in this section and shown in 4.1. As the shoeboxes are rectangular prisms, these weights are necessarily the depth of each zone over the total shoebox depth.

$$\begin{aligned} E &= \sum_{i=0}^{m} \sum_{j=0}^{n_i} \sum_{k=0}^{p_{i,j}} c_{i,j,k} \mathcal{F}(S_{i,j,k}) \\ &= \sum_{i=0}^{m} \sum_{j=0}^{n_i} \sum_{k=0}^{p_{i,j}} \frac{A_i}{G} \cdot \frac{L_{i,j,k}}{L_i} \cdot \mathcal{F}(S_{i,j,k}) \end{aligned} \tag{4.1}$$

4.1.2 Shoebox geometry specification

The geometric parameters which define a shoebox are as follows:

$\theta_{i,j,k}$:= shoebox facade normal azimuthal angle

$w_{i,j,k}$:= shoebox facade width

$h_{i,j,k}$:= shoebox height

$\omega_{i,j,k}$:= shoebox window-to-wall ratio

$d^{p}_{i,j,k}$:= shoebox perimeter zone depth

$d^{c}_{i,j,k}$:= shoebox core zone depth

$\rho_{i,j,k}$:= shoebox roof surface adiabatic fraction

$\sigma_{i,j,k}$:= shoebox ground surface adiabatic fraction

While typical UBEMs require the height and WWR parameters to be shared across a building facade to enable clustering, the evaluation of the surrogate allows for more granular shoebox allocation with unique parameters per shoebox, allowing for richer design expressivity. Orientation is driven directly by edge orientation, while shoebox width is typically set by users to a value between 3 and 5 meters.

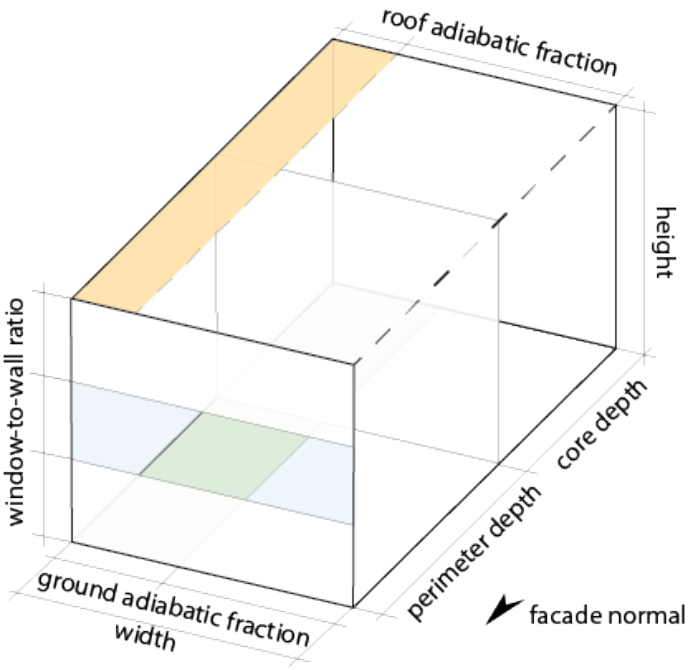


Figure 4.3: Shoebox geometry parameters

The remaining parameters are driven by the floorplate autozoning so that each individual shoebox maintains key characteristics of the floorplate. After the ith floorplate has been autozoned into n_i perimeter zones and one core zone, the entire floorplate's *core area*-to-*perimeter area* ratio η_i is computed. Similarly, the floorplate's entire *perimeter area*-to-*facade area* ratio ξ_i is computed. These together drive the shoebox perimeter zone depth $d^{p}_{i,j,k}$ and shoebox core zone depth $d^{c}_{i,j,k}$ for a given shoebox width $w_{i,j,k}$ and height $h_{i,j,k}$. The product of the floorplate perimeter area-to-facade area ratio and the shoebox facade area $\xi_i \cdot w_{i,j,k} \cdot h_{i,j,k}$ gives the shoebox perimeter zone area; dividing by the shoebox width returns the desired shoebox perimeter depth as in 4.2. The shoebox core zone depth is found from the perimeter depth and η_i as in 4.3:

$$d^{p}_{i,j,k} = \xi_i \cdot h_{i,j,k} \tag{4.2}$$

$$d^{c}_{i,j,k} = \eta_i d^{p}_{i,j,k} \tag{4.3}$$

Finally, the roof and ground surfaces of the shoebox are set to be partially adiabatic according to ratios $\rho_{i,j,k}$ and $\sigma_{i,j,k}$ to preserve heat loss form factors for the building. When using many shoeboxes, these ratios can be set to fully adiabatic for middle floors, and fully exposed for roofs on the top floor or fully exposed for grounds surfaces on the bottom floor; alternatively, when using fewer shoeboxes, they can be set according to a building's heat loss form factor. The side and rear walls of the shoeboxes are also set to be adiabatic, as shoeboxes are placed per facade edge. Roof and facade constructions are set according to the non-geometric features, as discussed in sec. 4.1.4.

While this scheme for shoebox geometry allocation is used and validated in this work, focus was placed on the development of the surrogate, and its rapid evaluation opens the door for alternative shoebox allocation strategies to be more easily developed and tested in an iterative manner in future work.

4.1.3 Shading and context geometry

To accommodate urban context effects via building-to-building shading and self-shading, a compact representation of the context geometry is required. To achieve this, each shoebox discretizes the surrounding context (including itself) in azimuthal increments of $\phi = \pi/12$ radians and casts rays into the scene for each azimuthal angle; the top edges of intersecting surfaces are used to determine an elevational angle for the corresponding azimuthal angle. These elevational angles become the inputs which together describe a *shading mask* for the corresponding shoebox. In the simple example in Figure 4.4, one of the twelve shading elevation angles is drawn. Although an urban scene may take on a more complex form, determining the shading mask for a shoebox reduces the entire scene to an equivalent shading mask which would visually follow the form shown in Figure 4.4. While this step might seem computationally expensive due to the raytracing, custom GPU kernels extending previous work by the authors were written to exploit algorithmic shortcuts which arise from the 2.5D nature of the problem and to exploit the massive potential for parallelism, and is still a fraction of the compute of a radiation analysis [65]. These shading masks are also easily cached and only need to be computed once. The algorithm developed to solve this is a variant of the Amanatides-Woo voxel traversal adapted to the 2.5D GIS context [66]. However, as shown in 5.4, even using simple rules of thumb to estimate shading context can be sufficient to avoid this step all together by replacing explicitly computed shading masks with reasonable assumptions.

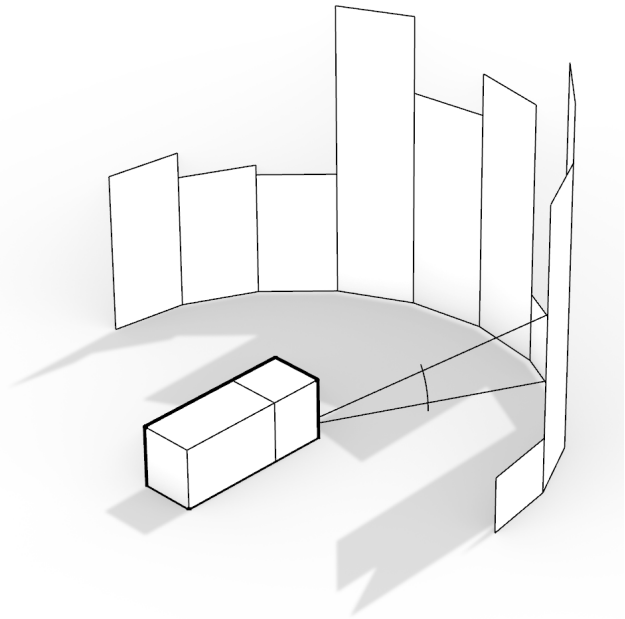


Figure 4.4: Context shading representation for a shoebox module.

4.1.4 Non-geometric features and schedules

In addition to the geometric features, each shoebox has a variety of non-geometric features representing the building's profile and usage. A complete list of static features is reproduced in Table 4.1. In addition to the static non-geometric features, three schedules of 8760 independently variable timesteps are used to describe hourly building usage of occupancy, equipment, and lights, represented as fractions of a given people or power density.

4.2 Synthetic data framework

This framework requisites the creation of a comprehensive synthetic dataset of physics-based shoebox models on which a neural network-based surrogate is trained. The dataset plays a pivotal role in accurately capturing the complexity and variability of energy use across different climates and building typologies and is freely published in accompanying materials to facilitate open research. Over 1.25 million unique shoeboxes were generated, spanning the globe and covering all ASHRAE climate zones equally. EnergyPlus is used to simulate each of the shoeboxes using Ideal Loads Air System HVAC input, reporting monthly heating and cooling loads. The dataset is split into three segments, *training* (80%), *validation* (10%), and *testing* (10%).

4.2.1 Shoebox sampling and generation

Each shoebox simulation in our dataset represents a simplified building model with specific geometric and non-geometric features. These features are uniformly and randomly sampled from a predefined design space, ensuring a wide variety of building types and characteristics, including unusual parameter combinations which are uncommon in the real world. Uniform random sampling was chosen because of this intent to learn a complete function approximation of the particular mathematical model of reality over the whole design space domain, as opposed to an approach which aims to learn from the non-i.i.d. (independent and identically distributed), features. In other words, the sampling strategy seeks to aid in learning the *physics*, rather than the *data*.

The geometric sampled features include parameters such as facade orientation, window-to-wall ratio, and zone depths, as defined in 4.1.2, while the non-geometric features cover building characteristics, including insulation levels, thermal properties of building materials, power densities, and more, detailed in 4.1.4. The azimuthal shading context was uniformly randomly sampled as well. Bounds for all parameters are listed in Table 4.1.

4.2.2 Schedule generation

To realistically model building operation, unique schedules for equipment, lighting, and occupancy were generated for every sample in the dataset. These schedules are fractional usage schedules spanning 8760 hours, corresponding to a full year. The schedules were generated according to a stratified strategy. For each schedule, there was a 50% probability of selecting pure noise, simulating unpredictable usage patterns with the intent to increase robustness to arbitrary schedules and assist in capturing impulse responses. The remaining half of the schedules were based on 15 different building use programs (e.g., residential, commercial, educational) from the US Department of Energy's representative commercial buildings dataset [11]. These were chosen with

equal probability, with white noise perturbations added at every timestep. The amplitude of these perturbations varied uniformly randomly up to 0.3, introducing realistic variability in building usage.

Table 4.1: Shoebox parameterization and training dataset bounds

Parameter Name	Min	Max	Units
Lighting power density	0	30	W/m^2
Equipment power density	0	60	W/m^2
People density	0	0.5	ppl/m^2
Heating setpoint	14	24	$^\circ C$
Cooling setpoint	22	30	$^\circ C$
Infiltration	0.0	1	$L/s/m^2$ (envelope area)
Ventilation per area	0.0	5	$L/s/m^2$ (floor area)
Ventilation per person	0	15	$L/s/person$
Ventilation mode	-	-	off/on/demand-controlled
Economizer	-	-	none/differential enthalpy
Heat recovery	-	-	none/sensible/latent
Facade mass	-	-	steel/wood/masonry/concrete
Roof mass	-	-	steel/wood/masonry/concrete
Facade R-value	0.3	15	m^2K/W
Roof R-value	0.3	18	m^2K/W
Slab R-value	0.3	15	m^2K/W
Window U-value	0.3	7.0	W/m^2K
Window SHGC	0.05	0.99	-
Occupancy schedule	0	1	fractional, 8760 timesteps
Lighting schedule	0	1	fractional, 8760 timesteps
Equipment schedule	0	1	fractional, 8760 timesteps
Normal azimuthal angle	0	2π	radians
Facade width	2	8	m
Height	2.5	6	m
Window-to-wall ratio	0	0.9	-
Perimeter zone depth	1.25	12	m
Core zone depth	1.25	60	m
Roof adiabatic fraction	0	1	-
Ground adiabatic fraction	0	1	-
Shading $1, \ldots 12$	0	$\pi/2$	radians

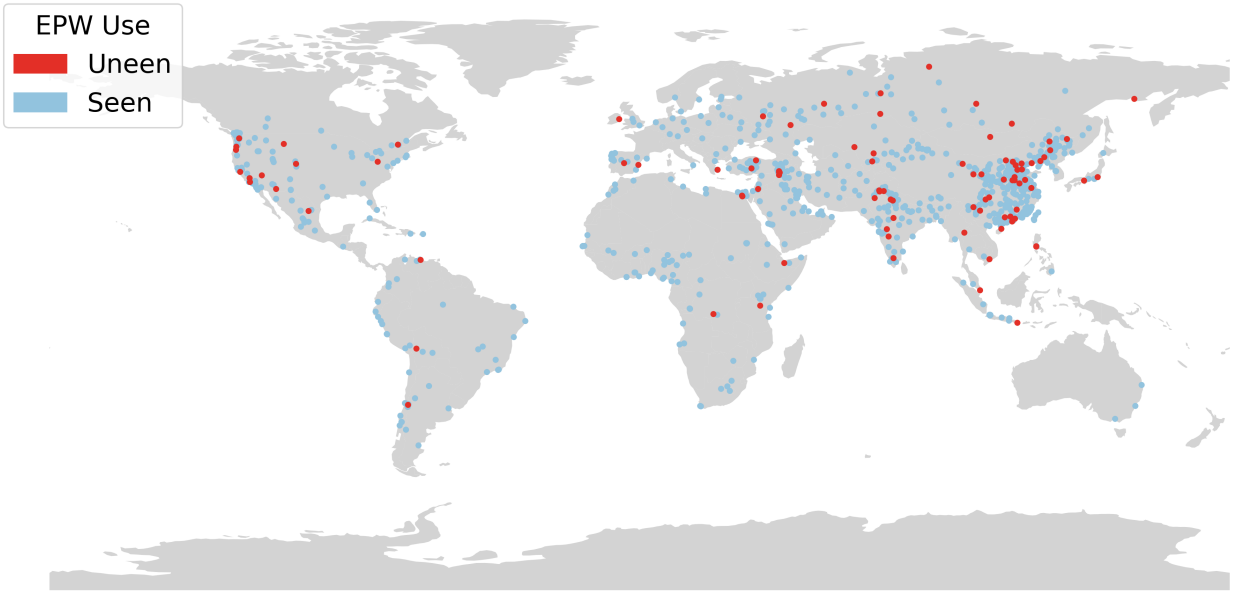


Figure 4.5: Weather file locations

4.2.3 Climate zone representation and weather file assignment

A critical aspect of our dataset is the representation of diverse climatic conditions. We select 850 EnergyPlus Weather (EPW) files from the 1000 most populous cities globally, while ensuring a balanced representation across all ASHRAE climate zones. These files are divided into two sets: *seen* and *unseen* EPWs, with an 80%/20% splitting ratio dividing the EPWs within each climate zone into the two *seen* and *unseen* buckets respectively. Shoeboxes in the *training* segment exclusively use weather files from the *seen* collection of weather files, while the *validation* and *testing* shoebox segments exclusively use *unseen* weather files. This separation allows for rigorous testing of the model's generalizability to new climatic conditions.

Each shoebox is randomly assigned to an ASHRAE climate zone, followed by a random selection of a weather file from that climate zone's respective set of seen or unseen EPW files. This approach ensures equal probability for each climate zone in each segment of the dataset, enabling the model to learn and adapt to a broad range of environmental conditions without overfitting in a certain region of the world.

4.2.4 Physics-based simulation data

Each shoebox sample's design vector is simulated with EnergyPlus to report monthly heating and cooling loads by zone, illustrated in Figure 4.6. All 30 conditioning, construction, geometry, and building use parameters are defined as model inputs. All other components of the energy models are held constant between samples in the dataset. More details can be found in Appendix A.1, and the parameters are listed in Table 4.1.

Simulations were executed with massive parallelism on AWS cloud infrastructure. This approach leveraged AWS Batch compute on Fargate, facilitating efficient and scalable processing of the large dataset.

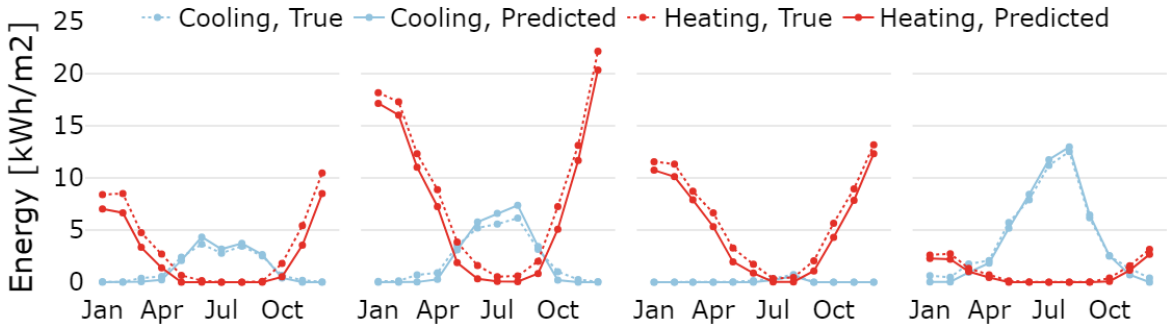


Figure 4.6: Illustration of energy model monthly heating and cooling requirements for a selection of sampled shoeboxes in Seattle (4C). Each shoebox has unique construction characteristics and programs resulting in differing monthly profiles.

4.3 Surrogate architecture

In this section, we define the architecture of the surrogate model used to predict monthly heating and cooling for shoebox energy models. Given our energy model $\mathcal{F}$ and a dataset of shoeboxes $\mathcal{S}$, we seek a surrogate model $\hat{\mathcal{F}}$ parameterized by θ which is defined as the solution to a minimization problem of a loss function $\mathcal{L}$ over the training dataset $\mathcal{S}$:

$$\mathcal{L}(\theta) = \frac{1}{|\mathcal{S}|} \sum_{\vec{x} \in \mathcal{S}} \left(\hat{\mathcal{F}}(\vec{x}; \theta) - \mathcal{F}(\vec{x}) \right)^2 \tag{4.4}$$

$$\theta^* = \arg\min_{\theta} \mathcal{L}(\theta) \tag{4.5}$$

4.3.1 Preliminaries

Neural networks are well-known for their highly expressive function approximation capabilities and amenability to efficient training and inference through easily parallelized operations. Appendix A.2 provides a brief review of neural networks, including convolutional neural networks.

We incorporate *temporal convolutional layers* which operate over multivariate timeseries. Given an input sequence for layer k with m_k channel dimensions (e.g. dry bulb temperature and relative humidity) and a kernel size of d_k timesteps, each output channel j is computed with a *filter* matrix $\mathbf{W_{k,j}} \in \mathbb{R}^{m_k \times d_k}$ applied as a dot product over a sliding window of the input sequence with length d_k:

$$\vec{x}_{k+1,j,t} = \phi(\mathbf{W}_{k,j} \cdot \vec{x}_{k,:,t:t+d_k} + b_{k,j}) \tag{4.6}$$

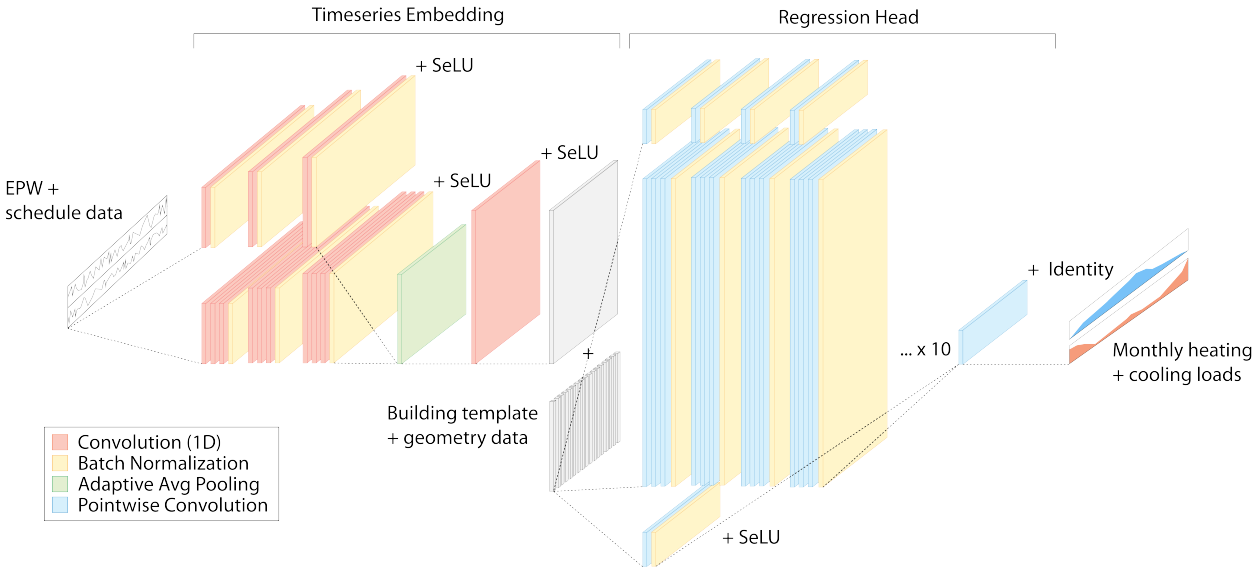


Figure 4.7: Two-part model architecture.

4.3.2 Model architecture

The model is composed of two networks which follow a typical ResNet-like architecture [67]. The first is responsible for processing the hourly annual timeseries data (including weather and schedules) and generating a latent representation of learned features at a monthly resolution, called the *timeseries latent projector*. The second, called the *regression head* is responsible for combining the learned features with the static input features (both geometric and non-geometric) to regress monthly heating and cooling demands. Both networks employ the same modular structure, differing only in the dimensionality configurations of each block. The basic modular block is composed of a chain of alternating temporal convolutions and non-linear activations, followed by a batch normalization layer and a skip connection which adds the input to the output of the block (including a channel size transformation applied as a pointwise 1D convolution). The skip connection is used to mitigate the vanishing gradient problem and enable training of very deep networks [68].

4.3.3 Timeseries latent projection

The timeseries latent projection is composed of three blocks, where each block contains three temporal convolution layers and a skip connection. In each of the three blocks, the three kernel sizes for each of the temporal convolution layers are 16,9,4. The first block has 16 output channels, the second has 16, and the third has 32. When the input channel dimension differs from the output channel dimension, the first layer in the block performs the channel size transformation via an increased number of filters. *SELU* activation functions are used. All striding is set to 1 and zero-padding is employed, preserving dimensionality along the time-axis across each layer. The only change in the temporal dimension occurs at the end of the timeseries latent projection via an average pooling layer which reduces the temporal dimension from 8760 to 360 followed by a learned convolution with kernel size 30 and stride of 30 to bring the temporal dimension down to 12, effectively acting

as a learned weighted averaging layer which reduces timeseries to monthly resolutions. This final transformation also expands the channel dimension to 4 times the number of static channels which will be included in the subsequent regression head.

4.3.4 Regression head

The regression head concatenates the static features with the output timeseries latent projections. It converts each of the static features into a timeseries of length 12 and concatenates these pseudo-timeseries with the projections of the original timeseries features along the channel axis. It is composed of 12 blocks each with 3 alternating temporal convolution layers and non-linear activations, followed by batch normalization and a skip connection. Unlike the timeseries latent projection, here the kernel sizes are all set to 1 with 512 output channels, effectively acting as a pointwise convolution, or, in other words, typical fully-connected layers applied over the channel dimension and parallelized over time. A final pointwise convolution reduces the channel count to 4, representing monthly heating and cooling for the perimeter and core zones separately, totaling 48 values in the output vector.

4.4 Surrogate data encoding

Weather timeseries channels extracted from EPW files were *dry bulb temperature*, *dew point temperature*, *relative humidity*, *wind direction*, *wind speed*, *direct normal radiation*, *diffuse horizontal radiation*, and *diffuse horizontal radiation*. Additionally, for each weather file, *solar azimuth* and *solar elevation* were computed for each hour of the year using PVLib [69]. *Latitude* and *longitude* were also included as constant timeseries. Schedules were included as additional channels in the timeseries latent projection's input layer. Each timeseries channel was normalized using the distributions of the data as present in the *seen* weather file dataset. All continuous static features (non-geometric and geometric) were min-max normalized to the range [0,1] as they were drawn from randomly and uniformly sampled distributions over the design space bounds (see 4.2). Categorical static features were one-hot encoded. Each of the static features were repeated into timeseries of length 12 before concatenation with the timeseries latent vectors along the channel axis in preparation for input to the regression head. Simulated monthly perimeter and core heating and cooling area-normalized thermal demand results were all normalized to a reference range of [0 $kWh/m^2 2$, 100 kWh/m^2].

4.5 Hyperparameters and training

The model was trained using the Adam optimizer with a learning rate of 0.01 and a batch size of 128 and the *MSE*-loss function (see equation 4.4). A learning rate scheduler was used to decay the learning rate by 0.95 for each epoch. 4x NVIDIA Tesla V100 GPUs were used for 53 epochs, taking approximately 8 hours, with approximately 1 million samples per epoch and shuffling performed at the end of each.

Chapter 5

Validation

The novel rapid UBEM evaluation pipeline was evaluated in three stages. First, the surrogate model was validated against the withheld data from the EnergyPlus simulated dataset (see 4.2) for approximately 125,000 shoeboxes. After validating the shoebox surrogate, validation of the shoeboxer algorithm was performed by creating 10,000 autozoned buildings (as in ASHRAE 90.1) with the same non-geometric feature parameterization as the shoeboxes. These buildings were all simulated in 13 locations around the world with EnergyPlus, then discretized and predicted using the shoeboxing algorithm and surrogate engine as detailed in 4.1. After validating the performance of the engine on individual buildings, the results for each of the 13 locations were aggregated into urban-scale results for both the autozoned multi-level EnergyPlus models and the neural network shoebox models.

The validation datasets generated by the authors were designed to minimize inductive bias between the building-scale validation models and the underlying shoebox models which the framework originally used to generate data and which the neural network regresses. Validating against data drawn from the defined design space was done to isolate any failure modes in the actual surrogate and shoeboxing process, as opposed to simply failing because the validation dataset was drawn from a distribution too disjoint from the supported parameters and their scopes. However, the real world of course is not perfectly captured by this design space, but we do still hope to be robust to it. As such, we additionally validate from the building-scale to the state-scale and even up to the national scale using the ResStock open dataset, which uses considerably more detailed white-box energy models than the shoebox models used here [70]. These pre-existing ResStock datasets have been generated from an entirely different energy modeling pipeline with different geometric parameterizations, and fully modeled HVAC equipment. The dataset, having gone through its own extensive validation to reflect real distributions of building characteristics and demand across the United States, crucially provide individual samples which act as a proxy for real-world conditions (and dramatically increases the risk of inductive bias of the entire shoeboxing algorithm).

Several error metrics are used in validation: Root Mean Squared Error (RMSE) and its normalized variant the Coefficient of the Variation of the RMSE (CVRMSE), the Mean Bias Error (MBE) and its normalized variant (nMBE), Mean Absolute Error (MAE), and Percent Error. Percent error is an easily understandable metric and therefore widely used but is highly sensitive to zero and near-zero conditions. RMSE is a commonly used metric representative of error in physical units between modeled and baseline values but does not indicate the relative magnitude of errors compared to the scale of the problem. Therefore, CVRMSE, the RMSE normalized by the dataset's true mean, can

provide clarity on the predictive performance of models due to its expressive clarity. MAE serves a similar purpose as RMSE but with no additional penalty incurred by larger magnitude errors with the removal of the quadratic. MBE sums the prediction residuals, giving an overall consideration of bias to over or underestimation through a signed value. However, as MBE is calculated across the dataset, cancellation of over and underestimated values can give the impression of overly optimistic results [49]. However, some steps of the framework rely on aggregating outputs of previous steps (e.g. from building to urban scale), so MBE is still useful: the nMBE at one scale becomes a percent error at the next scale. The validation process is described further in the following sections.

In each section, metrics are often reported segmented by end-use demand (heating or cooling), timeframe (monthly or annual), scale (shoebox, building, or urban), and ASHRAE climate zone. For example, reporting the CVRMSE of annual building-scale heating demand intensity in climate zone 1A. As the model predicts monthly thermal demand intensity, careful attention must be paid to aggregation. The terms before the word "scale" will always be used to refer to the aggregation, while the second set of terms will be used to refer to the segmentation. In the example above then, the buildings are segmented by climate zone, with the CVRMSE computed using each climate zone's respective true heating demand intensity mean as the normalizing factor, resulting in one error metric result for each climate zone present in the data (in this example). For an urban-scale example, the total energy of all buildings would be first aggregated and then divided by the total floor area of all buildings. When monthly CVRMSEs are computed and reported, they are computed over the entire timeseries, rather than computing 12 independent timeseries as is sometimes done.

$$\text{RMSE}(\hat{\mathcal{F}}, \mathcal{F}, \mathcal{S}) = \sqrt{\frac{1}{|\mathcal{S}|} \sum_{\vec{x} \in \mathcal{S}} \left(\hat{\mathcal{F}}(\vec{x}) - \mathcal{F}(\vec{x}) \right)^2}$$

$$\text{CVRMSE}(\hat{\mathcal{F}}, \mathcal{F}, \mathcal{S}) = \frac{\text{RMSE}(\hat{\mathcal{F}}, \mathcal{F}, \mathcal{S})}{\frac{1}{|\mathcal{S}|} \sum_{\vec{x} \in \mathcal{S}} \mathcal{F}(\vec{x})}$$

$$\text{MAE}(\hat{\mathcal{F}}, \mathcal{F}, \mathcal{S}) = \frac{1}{|\mathcal{S}|} \sum_{\vec{x} \in \mathcal{S}} |\hat{\mathcal{F}}(\vec{x}) - \mathcal{F}(\vec{x})|$$

$$\text{MAPE}(\hat{\mathcal{F}}, \mathcal{F}, \mathcal{S}) = \frac{1}{|\mathcal{S}|} \sum_{\vec{x} \in \mathcal{S}} \frac{|\hat{\mathcal{F}}(\vec{x}) - \mathcal{F}(\vec{x})|}{\mathcal{F}(\vec{x})}$$

$$\text{MBE}(\hat{\mathcal{F}}, \mathcal{F}, \mathcal{S}) = \frac{1}{|\mathcal{S}|} \sum_{\vec{x} \in \mathcal{S}} \hat{\mathcal{F}}(\vec{x}) - \mathcal{F}(\vec{x})$$

5.1 Shoebox-scale

5.1.1 Testing framework

The neural network model was validated on a test segment of 125,000 randomly generated shoeboxes exclusively using weather files that were not included in the training segment (see 4.2.3). Stratified sampling over the climate zones was used in the training dataset to ensure that the model had ample variety to learn from, as well as in the test set to ensure that error metrics could be com-

puted on a per-climate zone basis with equally sized segments. Heating and cooling loads are evaluated separately.

5.1.2 Validation

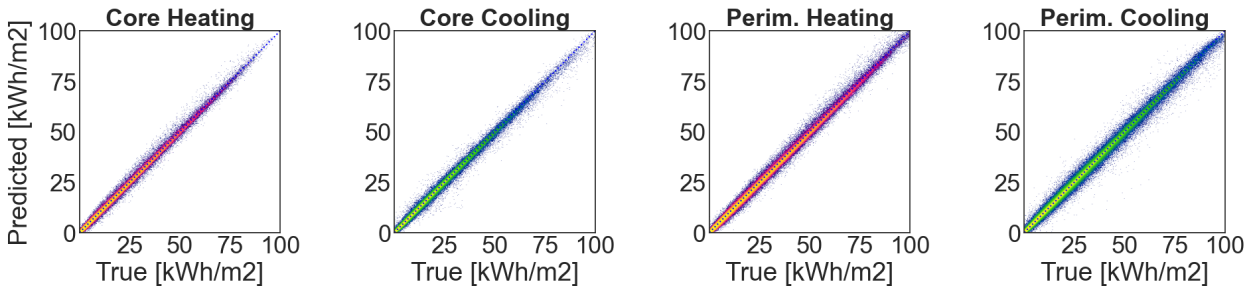


Figure 5.1: Shoebox monthly model fit

The model achieved a CVRMSE and RMSE for shoebox-scale monthly loads segmented at the zone level of 9.6% and 0.6 kWh/m2 respectively. To better understand behavior at the zone and end-use level, errors were also computed independently for heating and cooling in both zones, representing four error axes (see Figure 5.1). Results are shown in the Table 5.1. Given the ASHRAE recommendations of 15% CVRMSE and ±5% for nMBE for monthly predictions, the model is deemed sufficient for usage at both the zone-scale and shoebox-scale.

Table 5.1: Disaggregated shoebox-scale errors in unseen weather files

Period	Zone	End Use	CVRMSE [%]	nMBE [%]	RMSE [kWh/m2]
Monthly	Core	Cooling	7.5	-1.1	0.5
-	-	Heating	15.2	-0.6	0.3
-	Perimeter	Cooling	7.4	-0.3	0.8
-	-	Heating	11.7	-0.9	0.5
Annual	Core	Cooling	4.6	-1.1	3.6
-	-	Heating	8.5	-0.6	2.3
-	Perimeter	Cooling	4.2	-0.3	5.5
-	-	Heating	6.4	-0.9	3.6

Errors were also computed for each shoebox's monthly demands with segmentation over climate zones and individual months to ensure adequate performance. Complete error summaries are provided in tables B.2 - B.9. In all cases, both nMBE and CVRMSE were within the desired bounds for cooling. For heating, select cases exceeded the bounds for CVRMSE, as illustrated in 5.2, however the true mean was near zero and the RMSE was in line with or better than the other cases. Essentially, in certain climate zones or months where the expected value of the heating demand is exceedingly small, there is a degree of noisiness in the predictions, but given the low RMSE and nMBE values, this is considered acceptable. Users are encouraged to consult tables B.2 - B.9 provided for evaluating model performance according to their particular use-case.

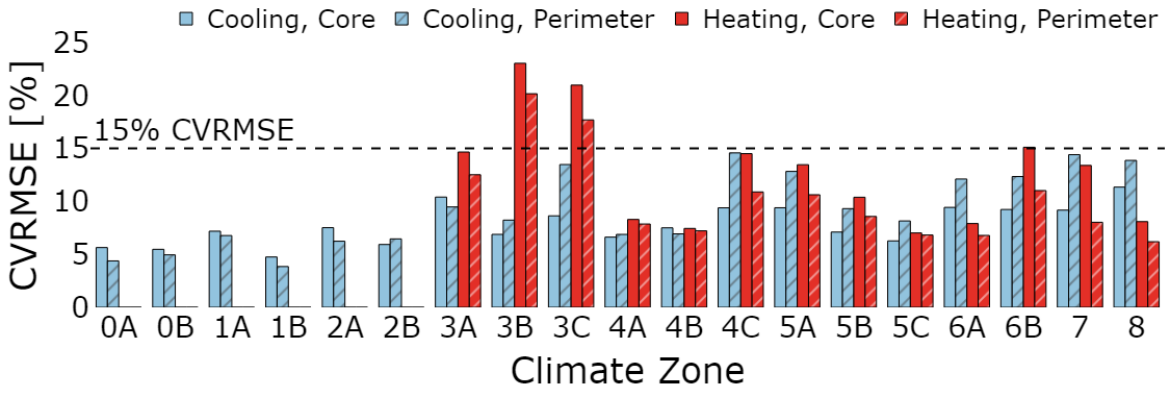


Figure 5.2: Monthly CVRMSE by climate and thermal zone (no-heating climates omitted)

5.1.3 Comments

The neural network is deemed sufficiently accurate for any use case which directly relies on explicitly reporting the individual results of shoeboxes, such as early-stage design exploration or parametric facade analysis (see 6.1) so long as the white-box model would be appropriate for that application as well. Algorithms which require specific aggregations of shoeboxes using computed weights (including the one detailed in 4.1) must still be validated independently of the neural network surrogate, as is done in the next section. However, it is clear that the neural network architecture and training pipeline are performing effectively.

5.2 Building-scale

5.2.1 Testing framework

To validate the performance of the surrogate in conjunction with the algorithm in 4.1 for representing buildings as linear combinations of shoebox discretizations, 10,000 buildings were generated by sampling from a uniform distribution over a modified version of the original design space that the surrogate was trained on (for instance, replacing shoebox-specific geometry features with building geometry features like number of floors or floorplate size). The design space was also restricted to remove regions with trivial solutions like extremely wide setpoint dead-bands which yield no demand. Complete sampling bounds for building generation are listed in B.1.

Each design vector from the new dataset was used to generate a multi-story building. This building was then autozoned into a multi-zone energy model, as described in 4.1, resulting in a whole-building energy model (rather than shoeboxes) per ASHRAE 90.1. Each whole-building model was run in EnergyPlus with 13 different weather files, totaling 130,000 test cases over all cities. These were then discretized into shoeboxes as discussed in 4.1.2 and processed by the surrogate before re-aggregation to the building-scale according to equation 4.1. Error metrics for heating and cooling demand intensities were then computed.

Additionally, whole-building EUI was also estimated from the thermal demands. To convert the monthly heating and cooling requirements predicted by the surrogate and shoeboxing algorithm into energy usage, each building was assigned a heating type (*heat-pump electrified* or *non-heat pump*)

followed by sampling a coefficient of performance (COP) between 0.92 and 0.97 for non-heat pump heating and between 2 and 4 for heat-pump electrification. Cooling COPs were sampled from the same range as heat pump electrified heating. The buildings were sampled with 85% probability of using non-heat pump heating and 15% heat-pump electrification. For each of the 10,000 autozoned buildings, this was performed 100 times to ensure no sensitivity to the coefficient of performance assignments. After assigning and applying coefficients of performance, the values were added together along with the total electricity usage from equipment and lighting loads. For the equipment and lighting loads, the electricity was directly reported from the EnergyPlus autozoned models as ground truth. This was compared to a summation of the fractional usage schedules with power density for the surrogate approach, returning equivalent results as expected.

5.2.2 Validation

Thermal demands

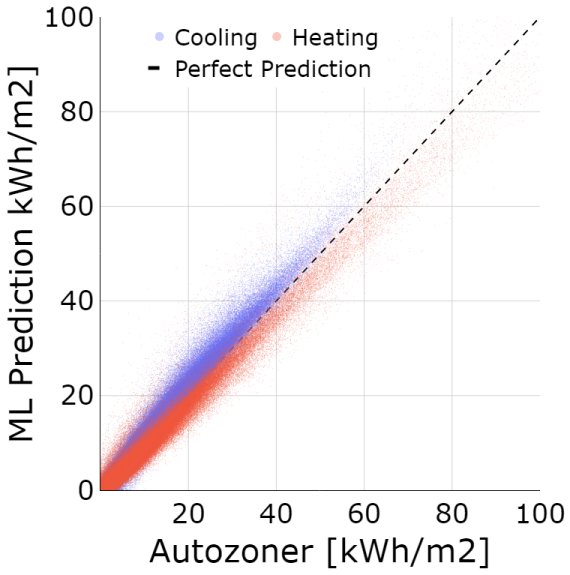


Figure 5.3: Building level heating and cooling demand monthly model fit.

Heating and cooling demand reported by EnergyPlus and the proposed algorithm are plotted against each other in Figure 5.3, showing strong biases in opposite directions for heating and cooling. This is also reflected in the normalized mean bias errors of 9.5% for cooling and -12.0% for heating on an annual basis. Error metrics for each thermal demand end use and timescale are given in table 5.2. Complete error metrics at the building-scale for each end use in each city on monthly and annual bases are given in tables B.10 - B.13. While the errors are higher than desired for building-scale predictions, further inspection reveals that the model performs better as thermal demand increases, as illustrated in figs 5.4 and 5.5, which depict errors with respect to different magnitudes of monthly thermal demand. The model's prediction error rate grows significantly slower than the monthly thermal demands' true values: when demand is low, noise in the predictions dominates, but when demand is high, the signal emerges from the noise floor and the model correctly captures the thermal demand. This is clearly illustrated anecdotally in 5.6, where monthly thermal demand predictions and true values for four different buildings are shown for three different climates (Mumbai, Seattle, and Oslo). While the general morphologies and scales of predictions are correctly identified, heating tends to be underestimated and cooling tends to be overestimated. This ability to correctly identify the scale and general morphology is proven by the high R2-scores for each individual month, which are above 0.92 for all months except for heating in June-July, which are still above 0.82, as shown in table 5.3.

Table 5.2: Building-scale thermal demand errors

Period	End Use	CVRMSE [%]	nMBE [%]	MAE [kWh/m2]	RMSE [kWh/m2]
Monthly	Cooling	25.9	+9.5	0.9	1.6
-	Heating	42.0	-12.0	0.6	1.5
Annual	Cooling	21.4	+9.5	9.4	16.3
-	Heating	31.6	-12.0	7.2	13.5

Table 5.3: Building-scale thermal demand R2 scores

Month	Cooling	Heating
Jan	0.966	0.972
Feb	0.963	0.975
Mar	0.967	0.965
Apr	0.966	0.934
May	0.970	0.920
Jun	0.958	0.846
Jul	0.953	0.821
Aug	0.961	0.882
Sep	0.960	0.936
Oct	0.966	0.933
Nov	0.970	0.975
Dec	0.959	0.974

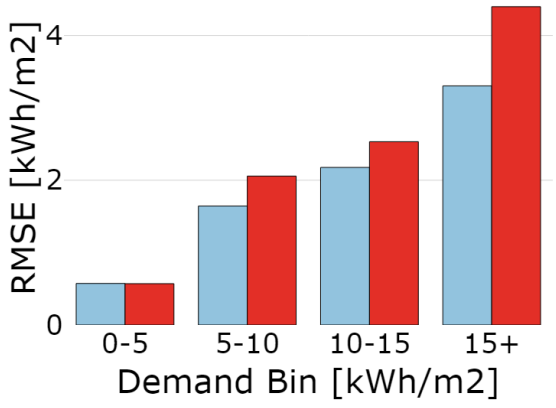


Figure 5.4: Building-scale monthly RMSE by demand

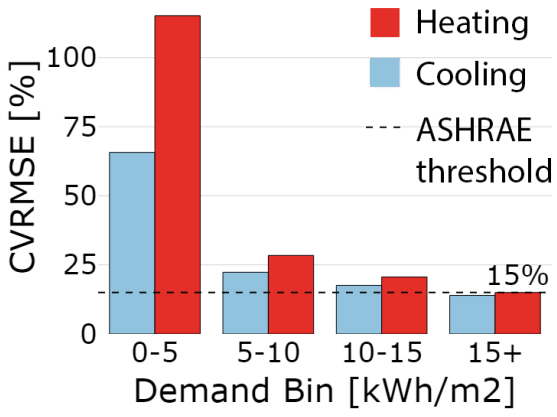
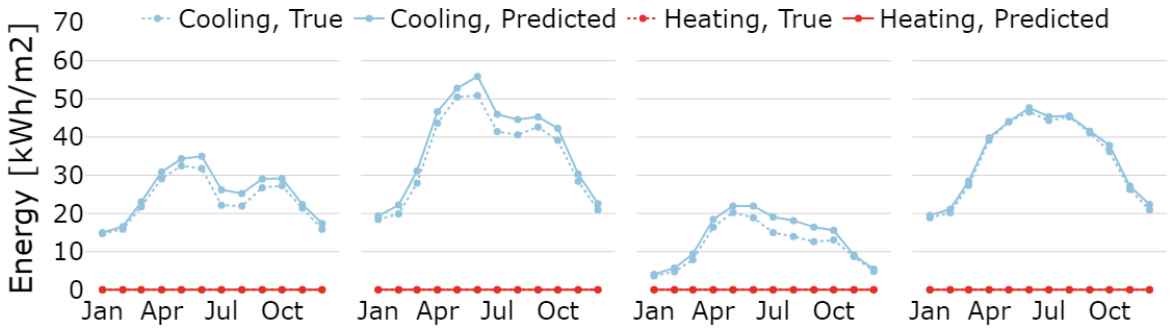
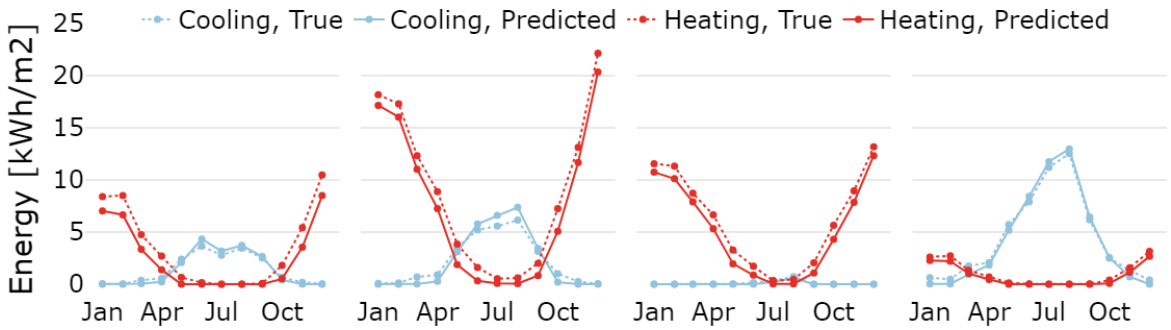


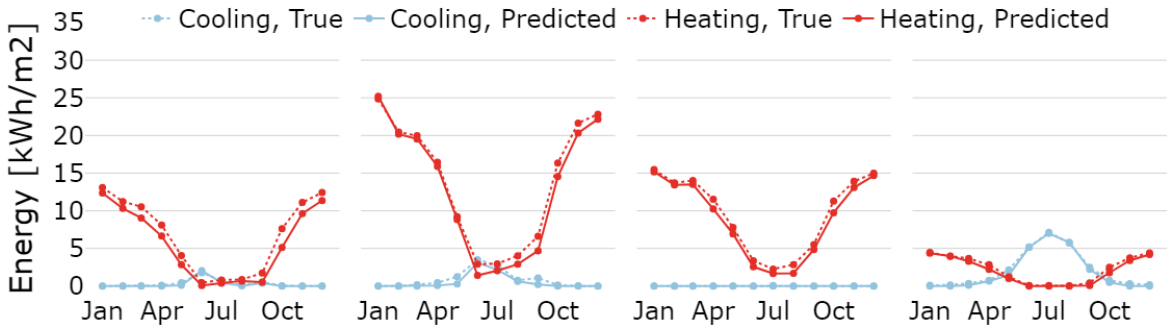
Figure 5.5: Building-scale monthly CVRMSE by demand



(a) Mumbai (0A).



(b) Seattle (4C).



(c) Oslo (6A).

Figure 5.6: Monthly thermal demands for four buildings in 3 different climates

Energy use intensity

Operating over the entire test suite (10,000 buildings ×13 locations ×100 COP assignments per building), the shoeboxing surrogate pipeline achieved annual individual building total area-normalized energy RMSE of 7.1 kWh/m2 and a CVRMSE of 5.6%. When evaluating error metrics separately per location, CVRMSE was less than 8.5% in all locations, as shown in Figure 5.8a. Complete error metrics for building-level total energy use are provided in Table B.14. Figure 5.7 illustrates RMSE values for heating and cooling as compared to mean normalized loads for heating and cooling individually. As expected, with higher load demands, the magnitude of RMSE increases. This propagates further into urban-scales as seen in Figure 5.8a. Future applications could extend the usability of the surrogate in extreme climates with a single dominating load end-use with further training.

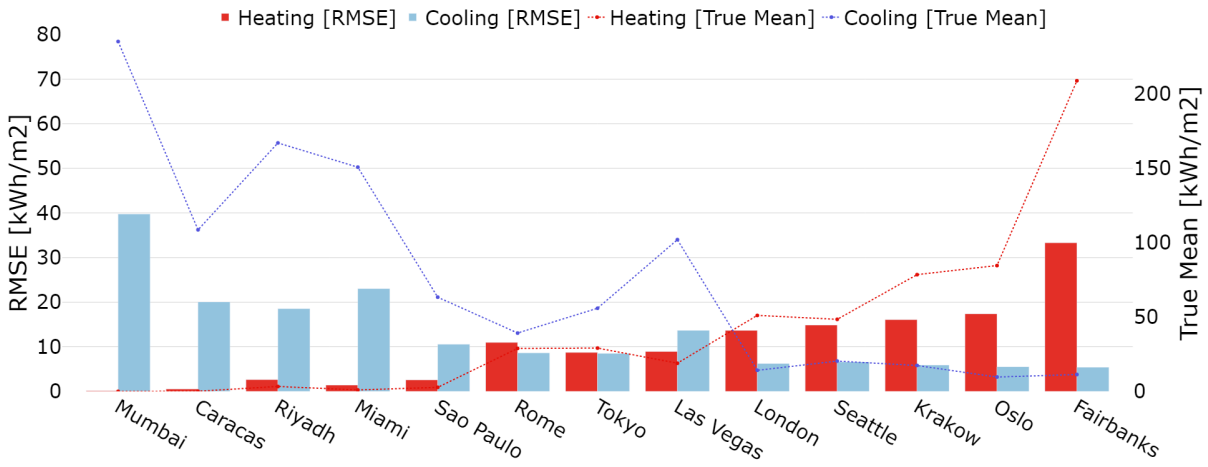


Figure 5.7: RMSE and mean normalized load for annual whole-building thermal demand by city and end-use.

5.2.3 Comments

The normalized mean bias error present in heating and cooling pointing in opposite directions is a clear indication that the shoeboxing process is introducing a misalignment with validation data that appears as a clear thermodynamic effect with a variety of possible explanations: more heat is being represented in the approximation of the building than in the simulated case. As a secondary goal of this project is to be able to use the same foundational model made for the UBEM context in single building contexts, it will be important to identify this source of error, or otherwise establish simple weighting factors to account for it as it appears to be present across climate zones. In the interim, it is recommended to use this model with caution in cases that depend on disaggregated demands at the building scale, as it will tend to over-predict cooling and under-predict heating; additionally, it is recommended to be aware of the fact that disaggregated predictions at the monthly scale are noisy when demand is low, but for identifying demand in the most thermally intense months, it is considered acceptable depending on the climate zone.

Possible causes of biased predictions

A possible explanation for bias in the predictions is the handling of infiltration in the shoeboxer algorithm. The effective infiltration of the shoeboxes resulting from the assigned infiltration rates and adiabatic proportions of each shoebox's roof and facade could end up too small; an artifact of how the adiabatic proportions are derived from the number of floors. This could result in insufficient heat loss in the winter, and potentially "free cooling" in transitional months.

Another possibility is that the core and perimeter zones are not weighted optimally. The core and perimeter zones are plainly area weighted with their depths determined to preserve the building's core area-to-perimeter area ratio. The two zones are otherwise identical with the exception of the facade in the perimeter zone and the parallel partition at the core zone's boundary. This means that all internal gains are identical on a per square meter basis, as are all conductive and infiltration rates for the roof, and side walls between the two zones. However, the core zone cannot exchange heat through its rear wall, instead only exchanging heat with the perimeter zone through the partition. Similarly, the perimeter zone can additionally exchange heat with the environment through its facade, including the window. In the winter, when the environment is too cold, the perimeter zone will likely be losing a large amount of heat through the window and facade infiltration which the core zone has no analog for. This ought to result in higher heating demand for the perimeter zone than the core zone for a shoebox at most heating timesteps, with the exception of some south-facing facades which receive solar gains at certain times of the day. If the core zone was given too much weight in the building level aggregation, then it is possible that the lower heating of the core zone becomes over-represented in the aggregated building's prediction. Similarly in the spring and fall, there may be many days should result in lower or no cooling demand as the perimeter zone more closely tracks the environment, so overweighting the core could result in excess cooling in the final estimation.

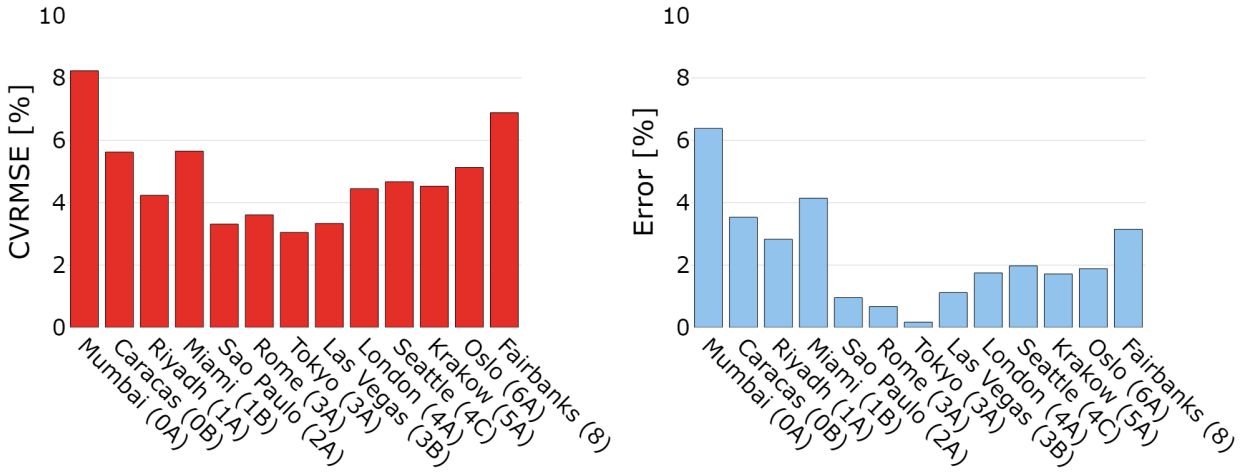
Additional work is warranted to further refine the core and perimeter depth weighting schemes so that they better reflect the thermodynamic process of the multi-zone whole-building model and some caution should be used in specific applications which rely on the disaggregated heating and cooling demands of particular buildings, e.g. in the autocalibration workflows detailed in 6.1. One natural approach for future work will be to use the rapid function evaluation of the surrogate to generate many different combinations of shoeboxes for a variety of buildings and separately learn correction factors for heating and cooling artifacts in the shoebox aggregation scheme.

5.3 Urban scale

As this methodology has been developed to analyze and model building energy use at urban scales, a representative method was developed to initially validate this use-case.

5.3.1 Testing framework

Using Rhino and Grasshopper, a "city" generation script was written to rapidly generate thousands of test buildings with shading context, which were evaluated in the same city's weather context. The area-normalized thermal demand intensity results for the individual buildings for each city were multiplied by each building's gross floor area (GFA) to give the total thermal demand of the city,



(a) Building-scale CVRMSE

(b) Urban-scale percent error

Figure 5.8: Error by city for shoeboxing algorithm with surrogate for whole-building annual energy usage, as compared to an autozoned EneryPlus model at the building-scale (left) and aggregated at the urban scale (right).

summed, and re-normalized by the UBEM's total GFA to give a final heating and cooling thermal demand intensity. As in the building-scale case, a series of trials where COPs were assigned to each building were conducted in order to then compute total energy use intensities for each city. Additional urban scale (and beyond) testing using ResStock data is detailed in 5.4.

5.3.2 Validation

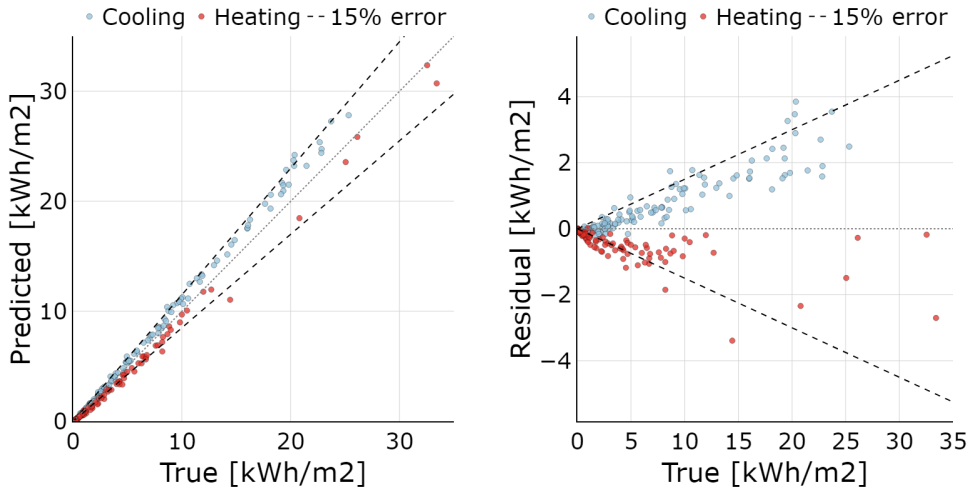
At the urban scale, in almost all cases the predicted monthly thermal demand was within 15% of the true thermal demand, with the exception of heating in select months and locations where the demand was less than 2.5 kWh/m2; in these cases, the error was almost always within 0.75 kWh/m2 (as shown in figs. 5.9a and 5.9b). For annual thermal demand intensity, the maximum error was within 15% in almost all cases except for select heating annual demands in cooling dominated climates as expected, and is illustrated in figure 5.10.

For each city's total energy use intensity, our approach achieved a percent error compared to the autozoned UBEM of less than 6.5% in all 13 locations and a percent error of 2% or better in 8 locations, shown in Figure 5.8b. Again, in climates with high loads of a single end-use, we see higher error. We also observe an overall decrease in error with aggregation at the urban scale. Although this is expected (and a key mechanism of the applicability of shoeboxer algorithms), error at the building-scale calls for future work on the shoeboxer methodology.

5.3.3 Comments

The hypothesis that at urban scales total error will decrease was validated, but we continue to observe higher levels of error in heating- and cooling-dominated climates. This is partially due to the fact that loads are simply higher for these areas, as is depicted in Figure 5.7, but also highlights that

any bias at a building scale will be emphasized when aggregated at urban scales. Nevertheless, the overall results are deemed acceptable for both monthly and annual predictions of both thermal demand intensities and total energy use intensities.



(a) monthly thermal demand fit (b) monthly thermal demand residuals

Figure 5.9: Urban-scale monthly thermal demand model fit (left) and residuals (right) for all cities

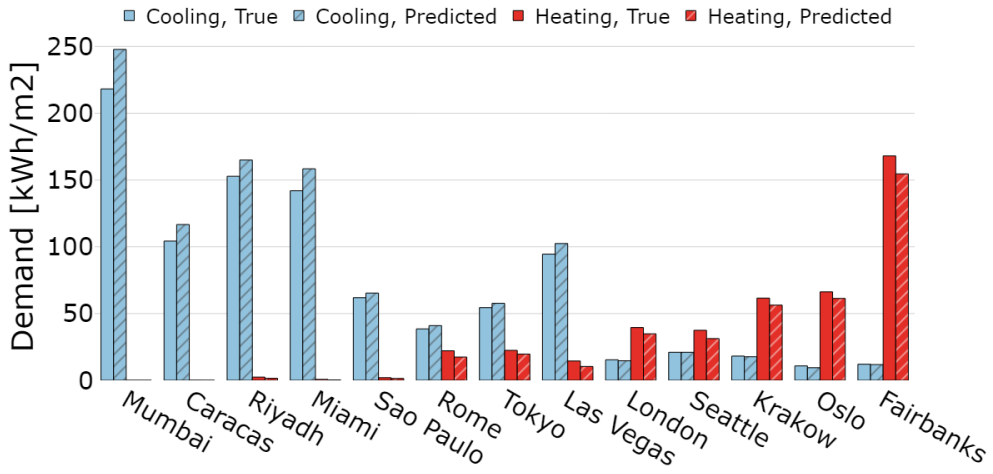


Figure 5.10: Predicted versus true annual thermal demand intensities at the urban scale

5.4 Validating against ResStock: from building to national scale

We additionally perform an alternate form of validation from the building-scale to the state-scale and even up to the national scale using the ResStock open dataset, which uses considerably more detailed white-box energy models. ResStock, developed by the US NREL and Department of Energy (DOE), is a methodology and dataset which enables unprecedented granularity in modeling the USA national building stock's residential buildings. Utilizing datasets and findings from several sources of building stock surveys and studies, the framework leverages statistical sampling and large datasets of stochastic sub-hourly building schedules to generate hundreds of thousands of detailed EnergyPlus building simulations set across thousands of counties across the United States with unique weather data. Developed, in large part, to explore retrofit pathways at large scales, ResStock has been rigorously tested and validated for this application, providing an ideal dataset to validate our proposed methodology.

In these ResStock test cases, we challenge the inductive bias of our model by using this pre-existing dataset generated from an independently developed energy modeling pipeline. The ResStock pipeline has entirely different parameterizations, fully modeled HVAC equipment, and crucially, individual design vectors generated to reflect the real distribution of building characteristics across the United States. While validating against the previous datasets reveals how the accelerated shoeboxing pipeline performs at regressing randomly generated data from the same design space, here we aim to prove that the entire stack is not the fruit of the poisoned tree (sometimes known as "junk in, junk out"). Although comparing performance to a uniformly random distribution (as done in the previous sections) is reasonable when no appropriate prior is available, ResStock fortunately provides an excellent repository of high-quality, locally cohesive individual buildings which can be used for large-scale urban validation.

5.4.1 Testing framework

The ResStock 2022 TMY3 1.1 release was used [70]. 257 counties from across the United States were selected to generate equal representation of all states. These counties were selected based off of the number of buildings present in the ResStock representation of each county to make estimating county-scale energy usage meaningful. Up to 2,000 buildings were randomly selected from each county if available; if fewer buildings were available, all were selected. A mapping from ResStock's input parameters was developed into the drastically simplified design space of the surrogate model, detailed in Appendix C.1. This mapping provides a strong test of the robustness of the model: how well can it perform despite its significantly reduced dimensionality, lack of fully detailed HVAC systems, and black-box nature? After mapping into the surrogate's design space, each building's design vector was then converted into shoeboxes which were predicted by the surrogate model and aggregated into a building scale prediction. These results were then compared to the ground truth data from ResStock. County-scale aggregations were computed in addition to state- and national-scale aggregations. As in previous sections, the aggregations were computed as the total energy demand for that scale (e.g. all buildings in a single county) over the total GFA for that scale, as expected.

5.4.2 Validation

Error metrics for the building, county, and state scale are provided in table 5.4. First, we examine the most granular predictions: monthly thermal demand for all buildings. A model fit scatter plot with 107,000 buildings is shown in Figure 5.11, and clearly captures the limitations of this approach. At the monthly level, the surrogate pipeline fails to provide sufficiently accurate predictions, likely due to the drastically different design spaces. At the same time, it is clear that both heating and cooling predictions are at least responsive to demand scales, as indicated by the R2 scores, and while the predictions are biased, they are not drastically biased (both within ±10%).

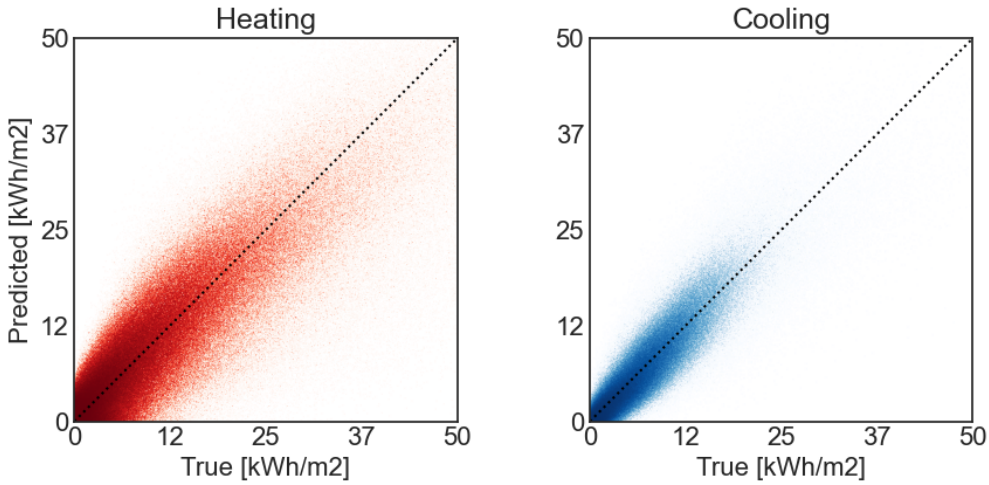


Figure 5.11: Resstock monthly thermal demand model fit (n=107,000 bldgs)

However, when we aggregate to the county level, the story changes: after converting the thermal demand intensities to thermal demand for each county, summing, and then returning back to normalized thermal demands, the CVRMSE for both annual heating and cooling demand intensities fall below 15%, and the R2 scores are all above 0.93 for both monthly and annual predictions. indicating that the surrogate model pipeline would give comparable predictions to NREL's ResStock dataset without the complexity of hundreds of parameters. This pipeline's simplifications may result in differences to ResStock at the building scale, but these differences tend to cancel out when aggregating to larger scales. Interestingly, the signs of the normalized bias errors flipped, indicating that smaller buildings and larger buildings must behave differently: at building-scale, these are weighted equally, but at the county-scale, the larger buildings within a county influence the error more.

This trend continues at state-scale, where the annual CVRMSEs of both heating and cooling demand intensities separately fall below 12.5% and the R2 scores rise above 0.95 for both monthly and annual predictions. This methodology yields incredibly similar results for entire states as compared to using highly detailed white-box models, despite the fact that it uses a vast amount of geometric and parametric abstraction and was trained with highly simplified energy models vis-a-vis ResStock. At the national scale, with over 100,000 unique buildings modeled by ResStock across 48 states, the percent errors in annual heating and cooling demand were 2% and 5%; monthly demands are shown in Figure 5.12.

Table 5.4: Annual and monthly errors for ResStock thermal demand intensity

Scale	Timestep	End Use	CVRMSE [%]	RMSE [kWh/m2]	nMBE [%]	R2
Building	Monthly	Heating	70.3	5.2	-8.9	0.820
-	-	Cooling	57.1	2.0	-3.3	0.879
-	Annual	Heating	48.9	43.2	-8.9	0.700
-	-	Cooling	38.5	16.3	-3.3	0.842
County	Monthly	Heating	19.8	1.5	+2.4	0.974
-	-	Cooling	32.2	0.8	+7.2	0.956
-	Annual	Heating	11.7	11.0	+2.4	0.935
-	-	Cooling	14.6	4.6	+7.2	0.967
State	Monthly	Heating	15.3	1.2	+2.6	0.983
-	-	Cooling	28.1	0.8	+7.3	0.963
-	Annual	Heating	8.4	7.9	+2.6	0.957
-	-	Cooling	12.1	3.9	+7.3	0.971

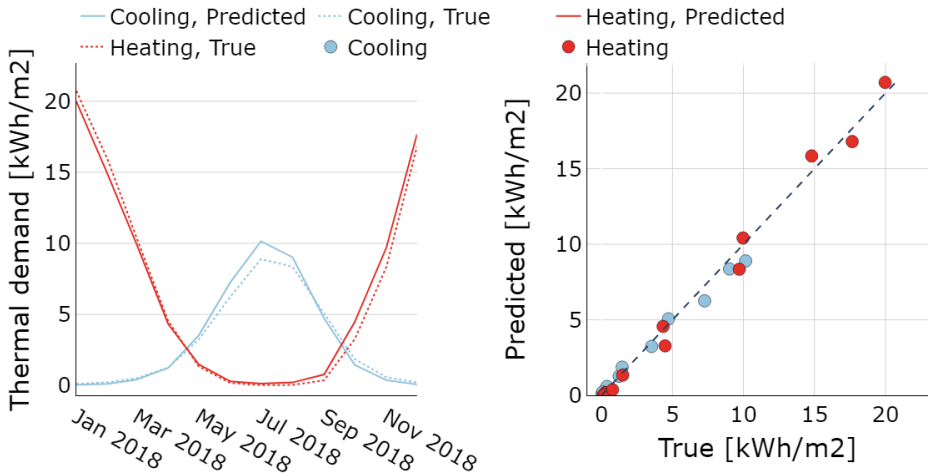


Figure 5.12: National-scale monthly thermal demand

5.4.3 Comments

While part of the value of high-resolution datasets like ResStock and ComStock comes with modeling and sizing systems where transients dominate design decisions, for instance in peak load estimations, these datasets also can play a key role as benchmark evaluation suites for prototypes of new building energy modeling methodologies. Although the model presented in this thesis is not yet robust enough to capture all the nuances of ResStock's design space or reach its high-resolution at the building-scale, this is by design: these levels of detail are not always necessary for the policy and planning contexts in which UBEMs are often deployed. The strong parity between the surrogate pipeline and ResStock at the county scale demonstrates the philosophy of UBEM: building-level accuracy can take a backseat when urban or county-scale predictions are desired.

5.5 Speed and savings

Full compute comparisons are detailed in table 5.5. For both urban-scale predictions, the black-box models do not use any archetypal segmentation. The white-box urban-scale prediction was made leveraging the shoeboxer algorithm with clustering and archetypal segmentation.

Table 5.5: Speed comparison between models

Scale	white-box	black-box	Performance Factor	Comments
shoebox	5s	<1ms	9,000x	1,800 shoeboxes/s
building	30s	11ms	2,700x	avg. 20 shoeboxes/bldg.
urban	83.3 hrs [multi-zone]	1.9 min	2,700x	10k bldgs.
urban	10.3 hrs [shoeboxed]	1.9 min	325x	10k bldgs.

Each EnergyPlus autozoned building energy model takes approximately 30 seconds on an AWS Fargate 1vCPU/2048MB machine; the EnergyPlus models for all 13 test locations were run massively in parallel on AWS in approximately 30 minutes (limited by a max parallelism of 5,000 simultaneous simulations). The serial compute for the autozoned models was approximately 83 hours for each of the 13 test cities. The actual time to solution was 5 minutes by leveraging distributed computing, at a total cost of $3.33, compared to a batched solution-time of 2 minutes and $0.05 total per location (10,000) simulations for the neural network shoebox engine when running on a serverless Nvidia A5000 GPU endpoint, representing substantial performance, cost, and energy improvements. At a throughput of between 1500 and 2000 shoebox predictions per second, a large-scale UBEM with 100,000 buildings in a city, which previously without distributed cloud computing may have taken days to weeks to complete depending on the stability of the user's platform could now be completed in under 20 minutes with individual results for every building, or in under a minute when using archetypal segmentation. For UBEM seed models, typically comprised of 100 buildings and used iteratively to define calibrated baseline templates, the modeling becomes nearly real-time at 1s-2s maximum.

Chapter 6

Discussion

In this manuscript, we set out to create a flexible and fast surrogate model combined with a shoeboxer algorithm that allows users to model monthly energy use and/or savings from common retrofit upgrades in multiple buildings in a fraction of the computational effort required by conventional BEM-based approaches. When combined with retrofit cost models and census data, the method can be applied in cities worldwide to formulate socioeconomic technology pathways to help municipal governments plan for the energy transition, as discussed in 6.1 and [16]. In reflecting on it's applicability in this wider context, we question: Is the methodology accurate enough to replace existing BEM workflows, flexible enough to be used in different geographic contexts and, most importantly, are these time savings crucial for the advancement of UBEM?

Starting off with the question of accuracy. There are no existing prescriptions for UBEM accuracy, leaving consensus on acceptable error bounds to the discretion of the user. For example, much of the focus of this work has been surrounding usability within the particular use-case of UBEMs for retrofit planning. Therefore, while validating the surrogate at it's minimal scale (the shoebox) was imperative, the building-scale results are less demanding in terms of absolute accuracy therefore allowing for savings in development and computation time. At the building scale, eliminating bias is of utmost importance so that urban aggregations remain accurate, which is the focus of this work, and will require further work on the existing shoeboxer algorithm. This perspective holds in considering the surrogate in future use cases, as described in 6.1, whereby the acceptable bounds of error are flexible and dependent on an identified minimum viable scope and result. These are particularly acceptable when the modeling focus lies on predicted savings from overall retrofit measures across different building typologies, if differences between the surrogate and physics-based model are sustained and under ASHRAE 140 recommendations at annual scales. When testing against 13 locations comprised of 10,000 buildings each (where each building was generated with multi-zone energy models and a shading context), monthly thermal demand errors were less than 15% in almost all locations and months, excepting heating loads in a few cooling-dominated climates where the demand was near-zero. Total annual energy use intensity errors were all less than 7%. indicating that the surrogate framework is providing comparable accuracy to an urban-scale model which is even more detailed than typical shoebox+archetype-based UBEMs.

When testing against the NREL dataset with 107,000 buildings spread across 257 counties from 48 states, at the county-scale, we found an nMBE of ±2.4% and ±7.2% for heating and cooling demand intensities respectively and CVRMSEs of 19.8% and 32.2% monthly heating and cooling demand intensities or 11.7% and 14.6% for annual demands compared to the corresponding

aggregations of ResStock simulation results [70], [71]. Considering the high level of detail in ResStock simulations, the necessarily flawed translations from ResStock data vectors into the surrogate pipeline, and the surrogate pipeline's many simplifications, this relative alignment is a promising sign for the validity of a real-time urban scale model as discussed in 5.4. It should be added that in its current stage, our method will not be able to provide reliable peak load predictions. However, the same is true for existing BEM-based methods for urban-scale predictions. One can however argue that fast surrogate UBEMs will be better equipped to tackle such problems going forward than their conventional alternatives since the computational cost of introducing stochastically modified schedules that reflect differences in occupant behavior is easily accomplished by our method.

Can the model be used for socioeconomic technology pathway analysis anywhere? A key feature of the model is its demonstrated ability to predict monthly energy use across all climate zones, saving the need to retrain the model for different geographic regions. Another key requirement for UBEM is of course the availability of reliable template libraries for different building stocks. Fortunately, there are efforts underway worldwide to compile such datasets, as both surrogate and physics-based UBEMs, equally rely on this simulation input. Additionally, with the cheap function evaluation, it becomes much easier to perform analyses which stochastically perturb various parameters or mutate them to better understand the actual behavior of newly developed archetype templates under a wide range of conditions. Along those lines, it becomes possible to introduce probabilistic analysis into modeling workflows: rather than predicting outcomes under a single set of conditions, populations of scenarios can easily be generated by randomly sampling over ranges of unknown parameters in building templates.

How crucial is the speed improvement offered by surrogate models? In a series of recent workshops with a total of 24 municipal representatives the authors have trained groups of sustainability champions, GIS managers, and energy modelers to develop customized building retrofit strategies for their communities [15]. Until now, these workshops have had the character of a consultation, where the sustainability champions initially lay out their concerns and aspirations for their community followed by two days of training the energy modelers and running a small number of upgrade scenarios. With the new, faster method in place, these workshops can both be shortened and evolve into design charette type gatherings where various community stakeholders propose different upgrades and incentive schemas that can then be tested in near real time. For example, a key stage in the development of city-wide UBEMs is the use of a "seed model," typically consisting of about 100 buildings [15]. The seed model is used to develop and calibrate baseline archetypes for a specific urban context through trial and error, and it is then used to test out different retrofit pathways. This stage of a workshop typically takes a full day, with each iteration taking up to an hour to run, even if only a single parameter has changed (for example, changing an R-value by 0.5-RSI for a single archetype). With the surrogate, each iteration is no more than two seconds, allowing for more nuanced iterative design (and more time spent on critical analysis). This shift will lead to multi-way conversations where more personal, community-specific solutions can be identified with greater buy-in from all parties involved. In terms of technical opportunities, such brainstorming sessions can be further supported by targeted optimization studies to fine-tune energy transition plans. From structural design to daylighting simulations, the building technology field has repeatedly demonstrated how rapid feedback and interactivity can lead to a more fluid and effective design process [72], [73]. With the surrogate framework in this thesis, UBEMs can now enjoy these same benefits.

6.1 Extended applications

We have touched on several applications within the subject of urban building energy modeling. However, with adequate accuracy at the shoebox and building-scale, there are several further applications for the proposed model and methodology in which the authors see potential for impact. By leveraging the accelerated nature of function evaluation, previously computationally expensive analysis methodologies or difficult-to-scale applications become feasible. These may have wide-ranging impact, ranging from injecting data-driven decision-making into early-stage design processes, autocalibrating building energy models to determining governmental strategies for heat-pump retrofit funding allocation policy. Cutting across building and urban scales, these uses-cases suggest starting points for developing end-to-end modeling pipelines to accelerate global decarbonization of the building stock.

6.1.1 Strategic retrofit optimization

A major barrier in retrofit adoption is a lack of understanding and uncertainty with regards to how best to balance the high financial burden of deep retrofits with the most "effective" interventions. Currently, the selection of retrofit packages for a building are a result of an energy audit and a complicated iterative energy modeling process based on a subjective and experienced-based list of potential interventions presented by the energy auditor or modeler. Because of complex interactions between building systems and subsystems, as well as between a building and its contextual environment, detailed surveying and energy modeling is needed to determine the applicability of selected retrofit measures for a specific building [74], [75]. The audit itself can be time consuming and costly, with rates between $0.02 to $0.5 per square foot in 2012 [76]. However, auditing is just the first step of the process. During project planning, the results of the energy audit are used to identify the specific retrofit measures that will be implemented. This includes determining the scope of the work, establishing a timeline, and identifying any potential obstacles or challenges. So, while selecting potential retrofits is an intensive process to begin with, the process of developing an integrated project construction timeline and financing plan can also have an impact on financial and operational feasibility. For example, given a series of annual cash flows, the implementation of higher-efficiency mechanical systems may be cheaper than an envelope upgrade, but if insulation levels are increased prior to purchasing (and therefore sizing of) new mechanical systems, the overall size of components can be reduced. Financial analysis can be used to determine the costs and benefits associated with the

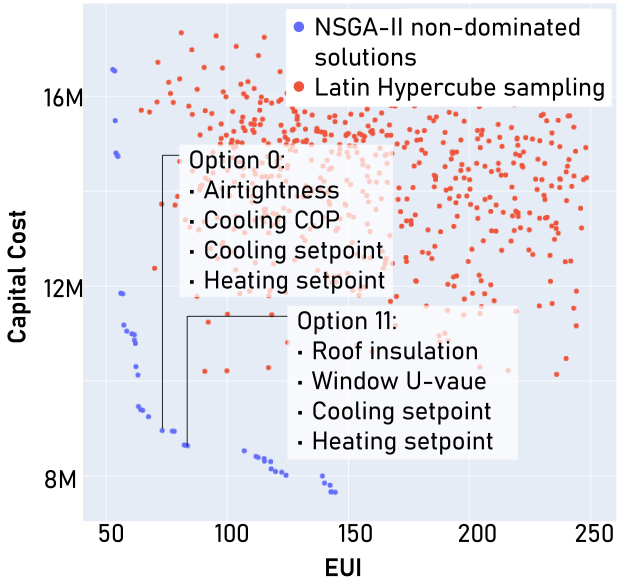


Figure 6.1: An illustration of retrofit pathway options analysis using the NSGA-II algorithm. Capital cost and energy use intensity are jointly reduced for a school located in Boston, MA.

proposed retrofit measures, and often results in an iterative process, to refine and re-select energy conservation measures that fit within financial needs [75].

Therefore, holistic energy auditing and retrofit implementation planning can be used to computationally determine optimal retrofit pathways in an efficient manner. Given the complexity of buildings, there can be a multitude of possible options and pathways for achieving desired emissions reduction goals, a decision process which can be fittingly addressed through multi-objective and linear optimization. Figure 6.1 illustrates an output of the optimization process, with each point representing a series of retrofit combinations which attempt to minimize both capital cost and energy use intensity. Being a non-differentiable optimization problem, optimal solutions are not immediately definable and must be determined through an iterative search, such as the NSGA-II algorithm. So, the ability to generate potential pathways is limited by the efficiency of the energy model. By lowering function evaluation time, surrogate models can be utilized to optimize retrofit plans through efficient design space exploration and search algorithms.

6.1.2 Connecting to cost models and funding policy: heat-pump sizing estimates

Accurately estimating thermal loads at the building scale can play a significant role in influencing the equitable and effective deployment of government resources in support of decarbonizing the building stock beyond simply using the thermal loads to estimate total emissions. It can also play a role in large, end-to-end pipelines. While evaluating which retrofit pathways are effective from a carbon perspective, it is important to recognize that actually getting building owners to adopt retrofits is a completely separate policy and economic challenge. Various socioeconomic factors can influence the likelihood of building owners actually adopting retrofit measures, particularly in the residential context [77]. Governments from the municipal to federal scale can play a role in increasing retrofit rates with incentives such as tax rebates, upfront cash, or match restricted savings accounts. However, allocating this funding so that various goals can be met, ranging from maximum retrofit rates and minimum carbon emissions to equitable distribution or prioritizing healthy living environments.

A climate robust surrogate can play a crucial role in an end-to-end pipeline which models from the bottom up the effects of various policy decisions concerning retrofit incentives. The key steps are to model the energy use and savings potential for each home, estimate the cost of the retrofits, estimate the willingness to pay for each homeowner in conjunction with the economic policy decisions, and then model cost and emissions outcomes.

The surrogate model can of course directly yield baseline energy savings. If retrofit costs and energy savings are known, payback periods can be calculated. Together with census-tract level data about socioeconomic statuses, these parameters can be used with data-driven models to estimate homeowner's willingness to accept sizes and types of financial incentives provided by governments [78]. However, estimating costs of retrofits can be challenging. We have recently contributed to work which demonstrates that using homeowner surveys, total installation cost (including all equipment and labor for heat pump retrofits) can be modeled with reasonable accuracy using just the year of construction (which serves as a proxy for the envelope quality) and size of the home [79]. Adding system size as an input to the cost model significantly improves the model's accuracy, even using simple coarse bins of small, medium, and large systems. This suggests a need for peak load estimation, however, even a monthly surrogate may be able to use peak winter

and summer months as binning proxies for categorization, though future work is needed to validate the binning accuracy.

With cost estimations in hand along with energy savings, the willingness-to-pay model can be executed with a given economic policy as an input affecting either upfront or payback time cost reductions. With homeowners' willingness to pay represented probabilistically, technology adoption curves can be modeled as distributions, as implemented by the authors on the UBEM.io platform. The technology adoption curves can in turn be used to drive emissions (using energy results from the surrogate), as well as cost projections to both homeowners and cost to the funding government selecting the incentivization strategy. Taken all together, these inputs can be used to drive a high-level economic policy planning tool which can explore how and where money is effectively allocated, which allows the user to identify various key priorities and goals, as shown in [80]. This represents a complete end-to-end policy planning pipeline that goes beyond simple savings estimations.

6.1.3 Granular and stochastic UBEMs

Archetypal segmentation continues to be an effective manner of simplifying and summarizing building characteristics, particularly from an envelope perspective in the event of limited data. However, the realities of building program and stochastic use remain limitations to this approach. Simply, people use buildings daily at different times and in different ways, and today's buildings often contain multiple programs, often across different levels, such as a retail and restaurants podium below an office and residential high-rise. Currently, this additional detail would need to be considered through a "best guess" average increase in program-related factors such as equipment and people densities, and use-schedules. Even if this process works well for one building in an urban area, these factors are highly dependent on the size and proportion of building programs, and therefore do not scale. Clearly, an ability to decouple overall building characteristics from internal use and program is ideal, which becomes feasible with rapid surrogates. Now, with the ability to allocate many more shoeboxes, each potentially with fully unique parameterizations, clustering by archetype is not as necessary to produce a feasible UBEM. Stochasticity can uniquely be introduced to building's schedules, allowing for improved modeling of peak loads, and several architectural programs can be simulated at a floor plan or even a room-by-room level.

6.1.4 Low-energy and parametric early-stage design

In architectural workflows, early-stage design decisions can make some of the biggest impacts on final building emissions, especially when considering embodied carbon. Real-time updates on operational energy use during early-stage architectural workflows are enabled with surrogate models such as this through the discretization of a single building design into shoeboxes. Additionally, sensitivity and parameter interaction analyses performed on single shoeboxes can be easily executed, aiming to build intuition in the designer about how thermal demands respond to parameter relationships given the climatic boundary conditions and how various design decisions result in trade-offs between heating and cooling demands or any other post-processed analysis metric therein.

While traditional physics-based shoebox models can be used for iterative design exploration relatively effectively and quickly (e.g. 4-5 seconds per single run), they still often require specialized software to run which a designer or architect may not have access to or working knowledge of. Engineers with the ability to set up such models and analyses may only be joining projects after

early-stage analysis is complete. Deploying such tools in web environments with simple, intuitive interfaces aims to increase the accessibility for usage by early-stage stakeholders and thus inject data-driven decision-making earlier into the design process. However, deploying cloud-native energy models in an effective manner comes with its own set of infrastructure and maintenance challenges, even at a runtime of 6 seconds per white-box model.

Using a surrogate model (which can easily be deployed in a minimum maintenance serverless inference architecture) can cut the costs associated with simulation by nearly the same factor as the performance gain or more by leveraging batching decisions. In the case of the surrogate model detailed in this thesis, that entails several orders of magnitude in cost savings per simulation to maintain a cloud-hosted, free-to-use tool. One such prototype application was developed and deployed using a combination of serverless inference providers as a proof-of-concept for such an application, and a screenshot of the user interface is shown in Figure 1.1.

Additionally, using white-box models, even with short run-times, makes systematic sensitivity and parameter interaction analyses necessarily low resolution and low dimensional. With a rapidly evaluated shoebox, generating high-resolution parameter grid sweeps and sensitivity analyses become feasible, while low resolutions ones become near-real-time. For instance, a 32x32 grid search (1024 simulations) completes in less than a second on a consumer GPU, while a 256x256 grid search (approx. 65,000 simulations) completes in under 40 seconds, compared to 102 minutes of serial compute and 110 hours respectively for the white-box model. Figure 6.2 demonstrates one such grid sweep for the familiar window-to-wall ratio versus orientation relationship.

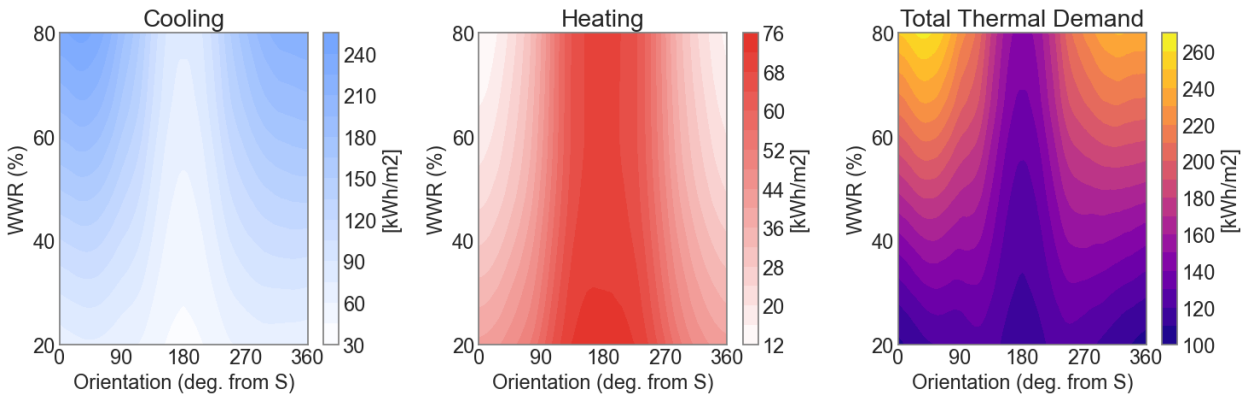


Figure 6.2: Contour maps of thermal demand intensities as a function of window-to-wall ratio and orientation

Floor plan optimization

Looking at single sections or floors of buildings, the structure of the shoebox could allow floor plans to be represented as a series of adjoining shoeboxes. [81] has explored automatic floor plan generation as a viable method for interactive and parametric design of low-carbon buildings, connecting structural material efficiency to energy-related factors, such as daylight availability and the placement of less-conditioned zones to act as buffers to highly conditioned zones thus reducing loads on active systems.

6.1.5 Facade energy analysis in early-stage design

By discretizing facades into unit cells and allocating one shoebox per cell (rather than lower-order shoebox allocations as in the UBEM case), it becomes feasible to run hundreds or even thousands of shoebox energy models along a facade to identify regions where additional architectural or engineering solutions may be necessary to control loads beyond typical irradiation analyses. This process is comparable to placing a sensor grid over a facade to perform solar irradiance analysis.

While using solar maps tells part of the story and can be used to guide decision-making about design and engineering strategies to manage gains, it necessarily cannot give the designer feedback about how iterative changes are impacting the actual ultimate energy usage directly, instead only giving feedback on how the input to one such energy model is changing. Select test case shoeboxes would still need to be run for each design scenario, with the designer needing to wait for results each time. Even using just 10 shoebox models per facade and a simulation time of 6 seconds per shoebox, this still results in waiting a full minute between runs, presenting a barrier to iterative design even at very low spatial resolutions. With a rapidly evaluated surrogate which is sensitive to the surrounding context and solar conditions, as presented in this thesis, shoeboxes can be placed across high-resolution facade cells to better understand how the surrounding context and shading affect thermal performance.

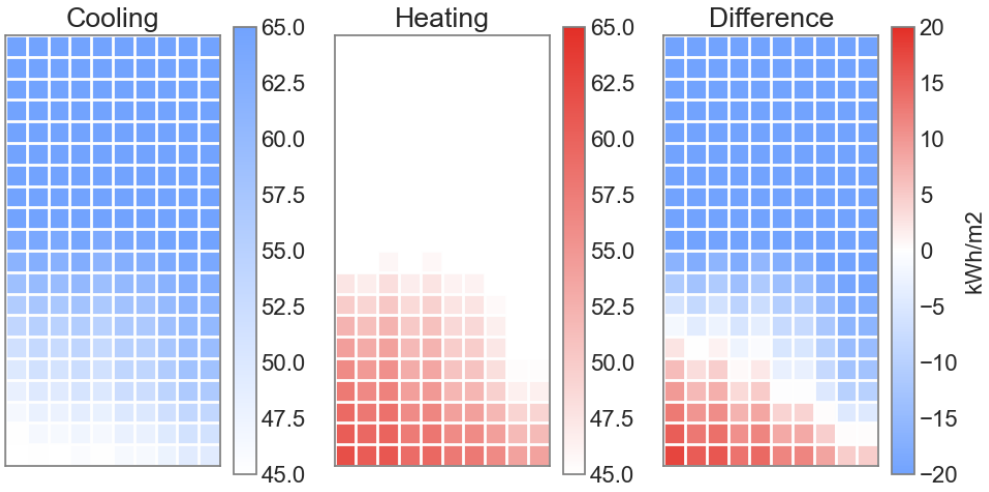


Figure 6.3: Elevational view of facade thermal demand intensity analysis for a 60m x 20m south-facing facade, 3m x 3m cells

Figure 6.3 demonstrates this analysis mode. A south-facing 60m x 20m facade is discretized into cells 3m tall and 3m wide, with one shoebox placed behind each facade and resulting in 10 shoeboxes for each of the 20 floors. All 200 shoeboxes were modeled with identical input parameters to represent a baseline starting point for the design except for the shading mask. The shading mask was calculated from the surrounding context independently for each shoebox, and then the heating and cooling demands for the perimeter zones were predicted using the surrogate model, taking under a second to return results. At this speed, the analysis can be used interactively in near-real-time, greatly increasing the potential for iterative design while allowing the form and forces to interact with each other in a feedback loop.

6.1.6 Autocalibration of building parameters

Given the importance of building parameter inputs for bottom-up modeling, ML models can be used to improve calibration methods for building energy models against metered data. BEM calibration is an inverse design or model parameter estimation process in which a set of energy simulation inputs is sought such that the modeled output aligns with target data, typically monthly utility bills. Once calibrated, then counterfactuals can be explored with the surrogate: how would the building perform under different conditions? Which retrofits will be the most cost- or carbon-effective? A globally robust surrogate can be used to accelerate the autocalibration process by replacing multiple different steps within an end-to-end calibration pipeline, though the most natural choice would be to directly use the surrogate within an optimization loop which aims to find populations or *ensembles* of viable designs which all yield the same target energy bill; this differs from one of the definitive treatments of autocalibration given by [53], in which a separate model is trained per building for the objective function itself, as opposed to a single surrogate model shared between all buildings, as is proposed here.

Chapter 7

Conclusion

We have described the project scoping definition, methodology, data collection, and validation process for the development of a novel black-box building energy model for use in UBEM applications, as well as presented several cases for expanded applicability. The two-part neural network architecture and chosen parameters enable a climate- and program-agnostic ML surrogate model, while the use of geometric abstraction with the shoeboxer algorithm permits geometric flexibility. The presented data collection, training, and validation methodologies also contribute to enabling cheap cloud-based pipelines for building- and urban- scale analyses in multiple applications. Currently capable of being deployed in a serverless web application for monthly evaluation of heating and cooling loads at urban scales, the inclusion of context through additional building energy loads will allow it to be applied in decision-making processes for policymakers and urban planners for decarbonization policy and retrofit adoption strategizing. Given the format of the model's timeseries inputs, this methodology can easily be implemented in conjunction with detailed urban microclimate data and stochastic building use and occupancy schedules, along with a wide array of identified (and unidentified) extended use cases.

7.1 Limitations and future work

While the proposed approach attempts to encompass a wide-ranging set of conditions and parameters and is deemed acceptable for use in UBEM applications, limitations are present. First, although the model performs well at both the shoebox scale and the urban scale, at the building scale, it demonstrated systematic error in a key building-level test case which must be addressed. Although context is included, window coverings or overhangs are not directly considered. In the validation methodology, validated geometries were equirectangular. However, given the accuracy of individual shoebox results, we suggest that irregular building shapes can be addressed by the surrogate given that the shoeboxer algorithm have previouslt been shown to appropriately handles irregular geometric conditions as well [43]. HVAC system information (for which COP is deemed a proxy), natural ventilation, and operational settings such as electric light dimming, heating and cooling setpoint hourly schedules, or ventilation/infiltration schedules are not considered. Future inclusion of these variables would allow for increased detail in programs across and within buildings and zones, which becomes particularly useful when considering operational retrofits and more detailed timesteps. Accordingly, a prominent use case for UBEMs is the analysis of peak loads with

increasing electrification. We have addressed monthly loads given their importance for urban emissions reduction planning, however, future work will address hourly values or seasonal peak analysis. Similarly, issues of overheating and thermal comfort are of growing concern. An extension to the surrogate could be altered to include thermal parameters as outputs.

7.2 Concluding remarks

This thesis presents a data-driven methodology to rapidly estimate the monthly energy end use of a detailed urban building energy model from the bottom up. A focus on usability, flexibility, and future impact drove the scoping and specification of the proposed surrogate model approach. Geometric abstraction with a shoeboxing algorithm and a two-part model architecture enable the handling of high-fidelity timeseries data and therefore a highly robust UBEM approach. After training the model and validating the methodology at shoebox, building, and urban scales, an additional validation process was carried out with a large national dataset of realistic data and highly detailed building simulations. Comparative to current approaches, the proposed method can be applied in the context of urban building energy modeling to balance speed, accuracy, accessibility, and applicability.

But, it is important to note that it is not just the speed or scale of the computation, or how detailed the information generated is, but also how that information is made useful and actionable; that is to say, how it is embedded into tools which enable designers, policymakers, educators, and other engineers to make critical decisions and implement meaningful carbon reduction actions. This methodology therefore aims to enable dialogues that engage the intersectional nature of implementing existing and effective technological solutions into the complexity and diversity of real systems. In the hopes that, eventually, these attempts to tackle complexity may propel efforts for a more sustainable future to catch up to a world already hurtling towards a climate emergency.

Appendix A

Methodological details

A.1 Physics-Based Simulation

Each shoebox sample's design vector is automatically translated into an EnergyPlus Input Data Format (IDF) model and simulated with EnergyPlus v22.2.0 [82]. These energy models use the Zone Ideal Loads Air System and report monthly heating and cooling loads. Economizers and heat recovery ventilation are optionally added according to each shoebox's sampled design vector. Ventilation per floor area and person are set using the design vector, as is the ventilation mode (always off, always on, demand controlled ventliation). Infiltration is set to the design vector's value representing flow per exterior area. EnergyPlus Simple Glazing Systems are used to represent window constructions by solar heat gain coefficient (SHGC) and U-Value as drawn from the design vector. Envelope construction is determined by first selecting from a pre-determined architectural section according to the design vector's categorical mass type (steel, woodframe, masonry, concrete) and then setting the thickness of the insulation layer according to the specified R-Value in the design vector. Constant dual setpoint thermostats are used. Lighting power density, equipment power density, and people density are set according to the design vector and controlled by the fractional schedules sampled in the design vector (see 4.2.2). All other components of the energy models are held identical between all samples in the dataset.

A.2 Convolutional Neural Networks

Neural networks are generally defined by composed layers of linear transformations $f_k : \mathbb{R}^{n_k} \rightarrow \mathbb{R}^{n_{k+1}}$ followed by non-linear activations $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ evaluated elementwise in $R^{n_{k+1}}$. Given the output of the previous layer $\vec{x}_k$, the output of the kth layer $L_k(\vec{x}_k) = \vec{x}_{k+1}$ is given by

$$L_k(\vec{x}_k) = \sigma(f_k(\vec{x}_k))$$
$$f_k(\vec{x}_k) = \mathbf{W}_k \vec{x}_k^T + \mathbf{b}$$

where the *weight matrix* $\mathbf{W}$ is a learnable $n_{k+1} \times n_k$ dimensional parameter and $\mathbf{b}$ is a learnable $n_{k+1} \times 1$ bias parameter.

A challenge faced by neural networks is the time and space complexity growth that scale with both the input and output dimensions of a layer, n_k and n_{k+1}. Dealing with high-dimensional data can quickly become infeasible. *Convolutional neural networks* (CNNs) are commonly deployed architectures which alleviate complexity explosion by assuming an underlying structure in the data which exhibits some form of temporal and/or spatial organization. Following this assumption, a much smaller linear transformation $\hat{f}$ (known as *kernel*) can be used in each layer, operating over only a subset of each input vector per application but parallelized over time and/or space, known as *striding* (when the input subsequence is non-contiguous, the kernel is said to be *dilated*, extending the window of time or space over which relationships can be made, though this is not used in this paper). The output vector of a convolution is reconstructed as a concatenation of the parallel results of kernel application. Convolutional parameters can be thought of as a weight-sharing scheme, vastly reducing the memory footprint. CNNs have been shown to be effective in wide-ranging applications alongside energy modeling, from computer-vision to audio processing to multi-variate timeseries sequence modeling.

Appendix B

Supplemental tables

Table B.1: Whole building validation sampling

Parameter Name	Min	Max	Units
Lighting power density	0	15	W/m^2
Equipment power density	0	30	W/m^2
People density	0	0.2	ppl/m^2
Heating setpoint	14	24	$^\circ C$
Cooling setpoint	22	30	$^\circ C$
Infiltration	0.0	1	$L/s/m^2$ (envelope area)
Ventilation per area	0.0	1	$L/s/m^2$ (floor area)
Ventilation per person	0	5	$L/s/person$
Ventilation mode	-	-	off/on/demand-controlled
Economizer	-	-	none/differential enthalpy
Heat recovery	-	-	none/sensible/latent
Facade mass	-	-	steel/wood/masonry/concrete
Roof mass	-	-	steel/wood/masonry/concrete
Facade R-value	0.3	15	m^2K/W
Roof R-value	0.3	18	m^2K/W
Slab R-value	0.3	15	m^2K/W
Window U-value	0.3	7.0	W/m^2K
Window SHGC	0.3	0.6	-
Building use type	0	15	(determines 8760 hour schedules)
Orientation	0	2π	radians
Building width	10	50	m
Building length	10	50	m
Perimeter zone offset	4.5	4.5	m
Floor-to-floor height	3	5	m
Number of floors	1	10	-
Window-to-wall ratio	0.1	0.8	-
Shading $1, \ldots 12$	0	$\pi/2$	radians

Table B.2: Shoebox monthly errors: core zone cooling

Month	CVRMSE [%]	nMBE [%]	RMSE [kWh/m2]	True Mean [kWh/m2]
01/Jan	10.4	-1.5	0.4	3.5
02/Feb	10.2	-1.8	0.4	3.4
03/Mar	8.3	-1.1	0.4	4.8
04/Apr	7.4	-1.3	0.4	5.9
05/May	6.8	-1.2	0.5	7.9
06/Jun	6.2	-0.6	0.6	9.5
07/Jul	5.9	-1.0	0.7	11.3
08/Aug	6.3	-1.0	0.7	10.7
09/Sep	6.5	-1.0	0.5	7.9
10/Oct	7.7	-1.3	0.4	5.9
11/Nov	9.2	-1.4	0.4	4.2
12/Dec	10.3	-1.4	0.4	3.6

Table B.3: Shoebox monthly errors: perimeter zone cooling

Month	CVRMSE [%]	nMBE [%]	RMSE [kWh/m2]	True Mean [kWh/m2]
01/Jan	14.0	-0.7	0.6	4.6
02/Feb	12.2	-0.4	0.6	4.8
03/Mar	9.1	-0.5	0.7	7.2
04/Apr	7.6	-0.4	0.7	9.7
05/May	6.7	-0.4	0.9	13.6
06/Jun	5.8	0.1	1.0	17.0
07/Jul	5.3	-0.3	1.1	20.7
08/Aug	5.5	-0.3	1.1	19.5
09/Sep	6.2	-0.2	0.8	13.7
10/Oct	7.8	-0.4	0.7	9.5
11/Nov	10.6	-0.9	0.6	6.1
12/Dec	12.9	-0.5	0.6	4.7

Table B.4: Shoebox monthly errors: core zone heating

Month	CVRMSE [%]	nMBE [%]	RMSE [kWh/m2]	True Mean [kWh/m2]
01/Jan	7.8	-0.7	0.5	7.0
02/Feb	10.5	0.1	0.5	5.1
03/Mar	12.1	-0.5	0.4	3.2
04/Apr	25.1	-1.2	0.3	1.1
05/May	68.2	-2.6	0.3	0.4
06/Jun	79.1	-0.1	0.1	0.2
07/Jul	61.1	-2.5	0.1	0.1
08/Aug	71.6	0.2	0.1	0.1
09/Sep	53.7	0.6	0.1	0.2
10/Oct	30.3	-1.3	0.3	0.9
11/Nov	11.6	-1.0	0.4	3.2
12/Dec	9.5	-0.6	0.6	5.8

Table B.5: Shoebox monthly errors: perimeter zone heating

Month	CVRMSE [%]	nMBE [%]	RMSE [kWh/m2]	True Mean [kWh/m2]
01/Jan	6.2	-1.2	0.9	14.4
02/Feb	8.1	-0.5	0.8	10.5
03/Mar	9.6	-0.5	0.6	6.5
04/Apr	18.2	-0.8	0.4	2.2
05/May	49.3	-2.0	0.3	0.7
06/Jun	68.7	0.0	0.2	0.3
07/Jul	59.4	-2.3	0.1	0.2
08/Aug	62.8	0.3	0.1	0.2
09/Sep	40.3	-0.3	0.2	0.5
10/Oct	21.1	-1.6	0.4	2.0
11/Nov	9.0	-1.2	0.6	6.7
12/Dec	7.2	-1.0	0.9	12.1

Table B.6: Shoebox monthly errors by climate zone: core zone cooling

Climate Zone	CVRMSE [%]	nMBE [%]	RMSE [kWh/m2]	True Mean [kWh/m2]
0A	5.6	-0.5	0.8	14.9
0B	5.5	-1.1	0.7	12.8
1A	7.2	-1.4	0.7	9.5
1B	4.8	-0.7	0.6	12.2
2A	7.5	-0.7	0.7	9.1
2B	5.9	-1.2	0.5	8.3
3A	10.4	-1.3	0.6	5.9
3B	6.9	-1.2	0.4	5.7
3C	8.6	-1.7	0.3	3.8
4A	6.6	-0.9	0.4	5.9
4B	7.5	-1.1	0.4	5.4
4C	9.4	-1.3	0.3	3.2
5A	9.4	-1.7	0.3	3.5
5B	7.1	-1.6	0.3	4.6
5C	6.3	-1.3	0.3	4.3
6A	9.4	-1.3	0.3	3.4
6B	9.2	-2.2	0.4	3.8
7	9.2	-1.9	0.3	3.3
8	11.4	-1.7	0.4	3.2

Table B.7: Shoebox monthly errors by climate zone: perimeter zone cooling

Climate Zone	CVRMSE [%]	nMBE [%]	RMSE [kWh/m2]	True Mean [kWh/m2]
0A	4.4	0.1	1.3	28.8
0B	5.0	-0.2	1.2	23.6
1A	6.8	-0.7	1.1	16.7
1B	3.9	0.1	0.9	22.3
2A	6.2	0.2	1.0	16.1
2B	6.5	-0.5	0.9	14.4
3A	9.5	-0.4	0.9	9.7
3B	8.2	-0.4	0.8	9.8
3C	13.5	-1.0	0.7	5.4
4A	6.9	-0.1	0.6	9.4
4B	6.9	-0.2	0.6	8.7
4C	14.6	-0.9	0.6	4.2
5A	12.8	-0.9	0.6	4.5
5B	9.3	-0.8	0.7	7.1
5C	8.2	-0.4	0.5	6.4
6A	12.1	-0.8	0.5	4.4
6B	12.3	-1.5	0.6	5.2
7	14.4	-1.4	0.6	3.8
8	13.9	-1.2	0.5	3.6

Table B.8: Shoebox monthly errors by climate zone: core zone heating

Climate Zone	CVRMSE [%]	nMBE [%]	RMSE [kWh/m2]	True Mean [kWh/m2]
0A	-	-	0.0	0.0
0B	101.7	-7.5	0.1	0.1
1A	139.8	7.6	0.1	0.1
1B	54.0	-2.8	0.1	0.2
2A	47.8	-6.2	0.3	0.6
2B	38.4	-3.1	0.2	0.4
3A	14.7	-0.1	0.3	2.2
3B	23.0	-3.3	0.4	1.5
3C	21.0	-2.4	0.3	1.4
4A	8.3	-1.5	0.2	2.9
4B	7.5	-1.2	0.2	3.0
4C	14.5	-2.5	0.5	3.6
5A	13.5	0.0	0.5	3.6
5B	10.4	0.5	0.4	3.3
5C	7.0	-1.0	0.2	3.2
6A	7.9	-1.6	0.4	4.6
6B	15.1	1.3	0.6	3.8
7	13.4	1.8	0.7	5.3
8	8.1	0.5	0.4	5.2

Table B.9: Shoebox monthly errors by climate zone: perimeter zone heating

Climate Zone	CVRMSE [%]	nMBE [%]	RMSE [kWh/m2]	True Mean [kWh/m2]
0A	-	-	0.0	0.0
0B	97.7	-8.8	0.1	0.1
1A	127.3	9.4	0.2	0.1
1B	50.7	-3.3	0.1	0.2
2A	38.4	-6.4	0.4	1.0
2B	35.4	-3.5	0.2	0.7
3A	12.5	0.1	0.5	3.8
3B	20.2	-3.2	0.5	2.6
3C	17.7	-1.9	0.4	2.4
4A	7.9	-1.6	0.4	5.5
4B	7.2	-1.5	0.4	5.9
4C	10.9	-1.9	0.7	6.7
5A	10.6	-0.5	0.8	7.0
5B	8.6	0.0	0.6	6.6
5C	6.8	-1.3	0.4	6.4
6A	6.8	-2.0	0.7	10.6
6B	11.0	0.6	0.9	8.4
7	8.0	0.3	1.0	13.0
8	6.2	-0.4	0.8	12.2

Table B.10: Building-scale end use monthly load error: heating

City	CVRMSE [%]	nMBE [%]	RMSE [kWh/m2]	True Mean [kWh/m2]
Mumbai	-	-	0.0	0.0
Caracas	-	-	0.1	0.0
Riyadh	147.6	-40.6	0.4	0.3
Miami	239.9	-59.6	0.2	0.1
Sao Paulo	127.4	-33.3	0.3	0.2
Rome	51.2	-22.4	1.2	2.4
Tokyo	42.8	-13.7	1.0	2.4
Las Vegas	71.9	-29.7	1.1	1.6
London	32.4	-13.2	1.4	4.3
Seattle	37.6	-17.7	1.5	4.0
Krakow	25.6	-9.4	1.7	6.5
Oslo	23.9	-8.7	1.7	7.1
Fairbanks	22.3	-8.4	3.9	17.4

Table B.11: Building-scale end use monthly load error: cooling

City	CVRMSE [%]	nMBE [%]	RMSE [kWh/m2]	True Mean [kWh/m2]
Mumbai	18.9	13.7	3.7	19.6
Caracas	20.9	12.6	1.9	9.1
Riyadh	14.1	7.7	2.0	13.9
Miami	18.5	11.8	2.3	12.6
Sao Paulo	20.0	5.9	1.1	5.3
Rome	32.2	7.4	1.1	3.3
Tokyo	26.8	6.5	1.3	4.7
Las Vegas	18.7	8.4	1.6	8.5
London	57.6	-2.6	0.7	1.2
Seattle	44.9	1.7	0.8	1.7
Krakow	48.4	-1.0	0.7	1.4
Oslo	76.1	-10.6	0.6	0.8
Fairbanks	75.4	0.6	0.7	0.9

Table B.12: Building-scale end use annual load error: heating

City	CVRMSE [%]	nMBE [%]	RMSE [kWh/m2]	True Mean [kWh/m2]
Mumbai	-	-	0.1	0.0
Caracas	566.4	-68.2	0.5	0.1
Riyadh	80.0	-40.6	2.6	3.3
Miami	147.0	-59.6	1.4	1.0
Sao Paulo	95.0	-33.3	2.6	2.7
Rome	37.9	-22.4	11.0	28.9
Tokyo	29.9	-13.7	8.7	29.1
Las Vegas	47.2	-29.7	8.9	19.0
London	26.7	-13.2	13.7	51.1
Seattle	30.7	-17.7	14.9	48.4
Krakow	20.5	-9.4	16.1	78.5
Oslo	20.5	-8.7	17.4	84.6
Fairbanks	15.9	-8.4	33.3	208.9

Table B.13: Building-scale end use annual load error: cooling

City	CVRMSE [%]	nMBE [%]	RMSE [kWh/m2]	True Mean [kWh/m2]
Mumbai	16.9	13.7	39.7	235.3
Caracas	18.5	12.6	20.0	108.6
Riyadh	11.1	7.7	18.5	167.1
Miami	15.3	11.8	23.0	150.8
Sao Paulo	16.7	5.9	10.6	63.3
Rome	22.0	7.4	8.6	39.2
Tokyo	15.2	6.5	8.5	55.9
Las Vegas	13.4	8.4	13.7	102.0
London	44.1	-2.6	6.2	14.1
Seattle	32.3	1.7	6.6	20.4
Krakow	33.9	-1.0	5.9	17.4
Oslo	57.3	-10.6	5.5	9.6
Fairbanks	47.7	0.6	5.4	11.3

Table B.14: Building-scale total energy error

City	True mean [kWh/m2]	RMSE [kWh/m2]	CVRMSE [%]	MBE [kWh/m2]
Mumbai	167.9	13.8	8.2	-11.1
Caracas	124.0	6.9	5.6	-4.7
Riyadh	145.4	6.1	4.2	-4.0
Miami	138.9	7.8	5.6	-5.9
Sao Paulo	109.2	3.6	3.3	-0.9
Rome	109.8	4.0	3.6	1.3
Tokyo	115.6	3.5	3.1	0.2
Las Vegas	128.2	4.3	3.3	-1.0
London	108.7	4.9	4.5	2.5
Seattle	109.9	5.2	4.7	2.9
Krakow	119.1	5.4	4.6	2.6
Oslo	118.5	6.1	5.2	2.9
Fairbanks	161.4	11.2	6.9	6.0

Appendix C

Pseudocode

C.1 NREL data conversion

The parameters and conversion methodology used to convert ResStock building definitions into inputs for the surrogate methodology are detailed below. Each variable has a related parameter or parameters from ResStock which are used to calculate the input to the model.

Because EPD and LPD values are not presented directly in ResStock, and it is not immediately apparent which loads are considered internal gains from a thermal perspective, we derived EPDs and LPDs from the provided timeseries. Additionally, because our underlying EnergyPlus shoebox models used in training do not include natural ventilation within their design space, we used double the number of expected shoeboxes for each building: half were modeled with a higher infiltration rate to try to capture the effects of natural ventilation within the shoebox predictions.

```
ResStock_building_definition = {
    "area": {
        ResStock Parameter: ["sqft"],
        Type: int,
        Conversion Method: lambda x: x / 10.7, # Convert to meters
    },
    "climate_zone": {
        ResStock Parameter: ["ashrae_iecc_climate_zone_2004"],
        Type: string,
        Conversion Method: None,
    },
    "state": {
        ResStock Parameter: ["state"],
        Type: string,
        Conversion Method: None,
    },
    "county": {
        ResStock Parameter: ["county"],
        Type: string,
        Conversion Method: None,
    },
```

```
    "city": {
        ResStock Parameter: ["city"],
        Type: string,
        Conversion Method: None,
    },
    "floors": {
        ResStock Parameter: ["geometry_stories"],
        Type: int,
        Conversion Method: None,
    },
    "width": {
        ResStock Parameter: None,
        Type: int,
        Default Value: 3,
    },
    "height": {
        ResStock Parameter: None,
        Type: int,
        Default Value: 3,
    },
    "perim_depth": {
        ResStock Parameter: None,
        Type: int,
        Default Value: 3,
    },
    "core_depth": {
        ResStock Parameter: ["sqft"],
        Type: int,
        Conversion Method: core_to_perimeter_area_ratio *
           perimeter_depth,
    },
    "roof_2_footprint": {
        ResStock Parameter: [
            "geometry_building_horizontal_location_sfa",
            "geometry_building_horizontal_location_mf",
            "geometry_attic_type",
        ],
        Type: boolean,
        Conversion Method:
            if residence is on a middle floor:
                no exposed roof
            else:
                exposed roof
    },
    "ground_2_footprint": {
        ResStock Parameter: None,
        Type: boolean,
```

```
        Default Value: exposed slab,
    },
    "orientation": {
        ResStock Parameter: ["orientation"],
        Type: float,
        Conversion Method: lambda x: convert orientation to radians,
    },
    "wwr": {
        ResStock Parameter: ["window_areas"],
        Type: float,
        Conversion Method: convert to percentage,
    },
    "cop_cooling": {
        ResStock Parameter: ["hvac_cooling_type"],
        Type: float,
        Conversion Method: look up cooling COP based on type,
    },
    "cop_heating": {
        ResStock Parameter: ["heating_fuel"],
        Type: float,
        Conversion Method: look up heating COP based on type,
    },
    "HeatingSetpoint": {
        ResStock Parameter: ["heating_setpoint"],
        Type: float,
        Conversion Method: convert to celcius,
    },
    "CoolingSetpoint": {
        ResStock Parameter: ["cooling_setpoint"],
        Type: float,
        Conversion Method: convert to celcius,
    },
    "LightingPowerDensity": {
        ResStock Parameter: ["lighting"],
        Type: float,
        Conversion Method: estimate lighting power density from use of
           overall lighting energy; infer lights schedule from hourly
           usage,
    },
    "EquipmentPowerDensity": {
        ResStock Parameter: ["plug_loads"],
        Type: float,
        Conversion Method: estimate equipment power density from use of
            overall lighting energy; infer equipment schedule from
           hourly usage,
    },
    "PeopleDensity": {
```

```
        ResStock Parameter: ["occupants", "sqft"],
        Type: float,
        Conversion Method: None,
    },
    "Infiltration": {
        ResStock Parameter: [
            "infiltration",
            "sqft",
            "geometry_stories",
        ],
        Type: float,
        Conversion Method: lambda x: calculate infiltration per square
           meter of envelope area assuming a square building
    },
    "VentilationPerArea": {
        ResStock Parameter: None,
        Type: int,
        Default Value: 0,
    },
    "VentilationPerPerson": {
        ResStock Parameter: None,
        Type: int,
        Default Value: 0,
    },
    "VentilationMode": {
        ResStock Parameter: None,
        Type: int,
        Default Value: 0,
    },
    "FacadeMass": {
        ResStock Parameter: ["insulation_wall"],
        Type: int,
        Conversion Method: look up mass group (low, med, high, very
           high) based on roof material,
    },
    "RoofMass": {
        ResStock Parameter: ["roof_material"],
        Type: int,
        Conversion Method: look up mass group (low, med, high, very
           high) based on roof material,
    },
    "FacadeRValue": {
        ResStock Parameter: ["insulation_wall"],
        Type: float,
        Conversion Method: None,
    },
    "RoofRValue": {
```

```
        ResStock Parameter: ["insulation_roof", "insulation_ceiling"],
        Type: float,
        Conversion Method: lambda x: None,
    },
    "SlabRValue": {
        ResStock Parameter: [
            "insulation_slab",
            "insulation_rim_joist",
            "insulation_foundation_wall",
        ],
        Type: float,
        Conversion Method: None,
    },
    "WindowUValue": {
        ResStock Parameter: ["windows"],
        Type: float,
        Conversion Method: look up as-built U-values based on type of
           window,
    },
    "WindowShgc": {
        ResStock Parameter: ["windows"],
        Type: float,
        Conversion Method: look up as-built SHGC based on type of
           window,
    },
    "EconomizerSettings": {
        ResStock Parameter: None,
        Type: int,
        Default Value: 0,
    },
    "ShadHeight": {
        ResStock Parameter: ["neighbors", "geometry_stories"],
        Type: int,
        Conversion Method: calculate angle of shading to right and left
            based on distance to neighbours, assuming they are of equal
            building height,
    },
}
```

References

[1] IEA, "Buildings," IEA, Paris, Tech. Rep., 2022.

[2] UNFCCC, "The Paris Agreement - Publication," ser. COP 21, Nov. 2018.

[3] Y. Lai, S. Papadopoulos, F. Fuerst, G. Pivo, J. Sagi, and C. E. Kontokosta, "Building retrofit hurdle rates and risk aversion in energy efficiency investments," en, *Applied Energy*, vol. 306, p. 118 048, Jan. 2022. DOI: 10.1016/j.apenergy.2021.118048.

[4] Y. Q. Ang, Z. M. Berzolla, and C. F. Reinhart, "From concept to application: A review of use cases in urban building energy modeling," en, *Applied Energy*, vol. 279, p. 115 738, Dec. 2020. DOI: 10.1016/j.apenergy.2020.115738.

[5] R. Weber, C. Mueller, and C. Reinhart, "Building for Zero, The Grand Challenge of Architecture without Carbon," en, *SSRN Electronic Journal*, Oct. 2021. DOI: 10.2139/ssrn.3939009.

[6] M. Wetter and M. Sulzer, "A call to action for building energy system modelling in the age of decarbonization," en, *Journal of Building Performance Simulation*, pp. 1–11, Nov. 2023. DOI: 10.1080/19401493.2023.2285824.

[7] J. J. Buonocore, P. Salimifard, Z. Magavi, and J. G. Allen, "Inefficient Building Electrification Will Require Massive Buildout of Renewable Energy and Seasonal Energy Storage," en, *Scientific Reports*, vol. 12, no. 1, p. 11 931, Jul. 2022. DOI: 10.1038/s41598-022-15628-2.

[8] C. Tuholske, K. Caylor, C. Funk, A. Verdin, S. Sweeney, K. Grace, P. Peterson, and T. Evans, "Global urban population exposure to extreme heat," eng, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 118, no. 41, e2024792118, Oct. 2021. DOI: 10.1073/pnas.2024792118.

[9] A. Mastrucci, E. Byers, S. Pachauri, and N. D. Rao, "Improving the SDG energy poverty targets: Residential cooling needs in the Global South," *Energy and Buildings*, vol. 186, pp. 405–415, Mar. 2019. DOI: 10.1016/j.enbuild.2019.01.015.

[10] T. Dixon and M. Eames, "Scaling up: The challenges of urban retrofit," *Building Research & Information*, vol. 41, no. 5, pp. 499–503, Oct. 2013. DOI: 10.1080/09613218.2013.812432.

[11] M. Deru, K. Field, D. Studer, *et al.*, "U.S. Department of Energy Commercial Reference Building Models of the National Building Stock," en, Tech. Rep. NREL/TP-5500-46861, 1009264, Feb. 2011, NREL/TP–5500–46 861, 1 009 264. DOI: 10.2172/1009264.

[12] Y. Sun, F. Haghighat, and B. C. Fung, "A review of the-state-of-the-art in data-driven approaches for building energy prediction," en, *Energy and Buildings*, vol. 221, p. 110 022, Aug. 2020. DOI: 10.1016/j.enbuild.2020.110022.

[13] Y. Li, Z. O'Neill, L. Zhang, J. Chen, P. Im, and J. DeGraw, "Grey-box modeling and application for building energy simulations - A critical review," en, *Renewable and Sustainable Energy Reviews*, vol. 146, p. 111 174, Aug. 2021. DOI: 10.1016/j.rser.2021.111174.

[14] C. Reinhart and C. C. Davila, "Urban building energy modeling," en, in *Building Performance Simulation for Design and Operation*, 2nd ed., Abingdon, Oxon ; New York, NY: Routledge, Apr. 2019, pp. 696–722.

[15] Z. Berzolla, Y. Q. Ang, S. Letellier-Duchesne, and C. Reinhart, "An Eight-Step Simulation Based Framework to Help Cities Reach Building-Related Emissions Reduction Goals," en, *SSRN Electronic Journal*, 2023. DOI: 10.2139/ssrn.4528659.

[16] Y. Q. Ang, Z. M. Berzolla, S. Letellier-Duchesne, and C. F. Reinhart, "Carbon reduction technology pathways for existing buildings in eight cities," en, *Nature Communications*, vol. 14, no. 1, p. 1689, Apr. 2023. DOI: 10.1038/s41467-023-37131-6.

[17] G. Trencher, V. Castán Broto, T. Takagi, Z. Sprigings, Y. Nishida, and M. Yarime, "Innovative policy practices to advance building energy efficiency and retrofitting: Approaches, impacts and challenges in ten C40 cities," *Environmental Science & Policy*, vol. 66, pp. 353–365, Dec. 2016. DOI: 10.1016/j.envsci.2016.06.021.

[18] J. Milford, M. Henrion, C. Hunter, E. Newes, C. Hughes, and S. F. Baldwin, "Energy sector portfolio analysis with uncertainty," *Applied Energy*, vol. 306, p. 117 926, Jan. 2022. DOI: 10.1016/j.apenergy.2021.117926.

[19] Y. Q. Ang, Z. M. Berzolla, S. Letellier-Duchesne, V. Jusiega, and C. Reinhart, "UBEM.io: A web-based framework to rapidly generate urban building energy models for carbon reduction technology pathways," en, *Sustainable Cities and Society*, vol. 77, p. 103 534, Feb. 2022. DOI: 10.1016/j.scs.2021.103534.

[20] F. Ascione, N. Bianco, T. Iovane, M. Mastellone, and G. M. Mauro, "Is it fundamental to model the inter-building effect for reliable building energy simulations? Interaction with shading systems," en, *Building and Environment*, vol. 183, p. 107 161, Oct. 2020. DOI: 10.1016/j.buildenv.2020.107161.

[21] Y. Han, J. E. Taylor, and A. L. Pisello, "Exploring mutual shading and mutual reflection inter-building effects on building energy performance," *Applied Energy*, Clean, Efficient and Affordable Energy for a Sustainable Future, vol. 185, pp. 1556–1564, Jan. 2017. DOI: 10.1016/j.apenergy.2015.10.170.

[22] C. Cerezo, J. Sokol, C. Reinhart, and A. Al-Mumin, "Three methods for characterizing building archetypes in urban energy simulation. A case study in Kuwait city.," en, in *BS2015*, vol. 14, Hyderabad, India, 2015.

[23] T. Loga, B. Stein, and N. Diefenbach, "TABULA building typologies in 20 European countries," en, *Energy and Buildings*, vol. 132, pp. 4–12, Nov. 2016. DOI: 10.1016/j.enbuild.2016.06.094.

[24] C. S. Monteiro, A. Pina, C. Cerezo, C. Reinhart, and P. Ferrão, "The Use of Multi-detail Building Archetypes in Urban Energy Modelling," en, *Energy Procedia*, vol. 111, pp. 817–825, Mar. 2017. DOI: 10.1016/j.egypro.2017.03.244.

[25] Z. Le Hong, Z. Berzolla, and C. Reinhart, “The more the better? Archetype segmentation in urban building energy modelling,” en, *Journal of Physics: Conference Series*, vol. 2600, no. 8, p. 082 004, Nov. 2023. DOI: 10.1088/1742-6596/2600/8/082004.

[26] L. Zhang, S. Plathottam, J. Reyna, *et al.*, “High-resolution hourly surrogate modeling framework for physics-based large-scale building stock modeling,” en, *Sustainable Cities and Society*, vol. 75, p. 103 292, Dec. 2021. DOI: 10.1016/j.scs.2021.103292.

[27] J. Langevin, J. Reyna, S. Ebrahimigharehbaghi, *et al.*, “Developing a common approach for classifying building stock energy models,” en, *Renewable and Sustainable Energy Reviews*, vol. 133, p. 110 276, Nov. 2020. DOI: 10.1016/j.rser.2020.110276.

[28] J. L. Reyna and M. V. Chester, “The Growth of Urban Building Stock: Unintended Lock-in and Embedded Environmental Effects,” en, *Journal of Industrial Ecology*, vol. 19, no. 4, pp. 524–537, 2015. DOI: 10.1111/jiec.12211.

[29] T. Dogan, C. Reinhart, and P. Michalatos, “Autozoner: An algorithm for automatic thermal zoning of buildings with unknown interior space definitions,” *Journal of Building Performance Simulation*, vol. 9, no. 2, pp. 176–189, Mar. 2016. DOI: 10.1080/19401493.2015.1006527.

[30] D. Crawley, F. Winkelmann, C. Pedersen, and L. Lawrie, “EnergyPlus: A New-Generation Building Energy Simulation Program,” en, US Department of Energy, Jun. 2011, pp. 58–59.

[31] X. Li and J. Wen, “Review of building energy modeling for control and operation,” en, *Renewable and Sustainable Energy Reviews*, vol. 37, pp. 517–537, Sep. 2014. DOI: 10.1016/j.rser.2014.05.056.

[32] F. Johari, J. Munkhammar, F. Shadram, and J. Widén, “Evaluation of simplified building energy models for urban-scale energy analysis of buildings,” en, *Building and Environment*, vol. 211, p. 108 684, Mar. 2022. DOI: 10.1016/j.buildenv.2021.108684.

[33] EnergyPlus Development Team, *Documentation Engineering Reference*, 2022.

[34] J. Seem, “Modeling of heat transfer in buildings,” Ph.D. dissertation, University of Wisconsin–Madison, 1987.

[35] K. Jaber, “FastCTF: A Robust Solver for Conduction Transfer Function Coefficients and Thermal Response Factors,” *Energy and Buildings*, vol. 253, p. 111 461, Dec. 2021. DOI: 10.1016/j.enbuild.2021.111461.

[36] J. D. Spitler, D. E. Fisher, and C. O. Pedersen, “The Radiant Time Series Cooling Load Calculation Procedure,” en, p. 14, 1997.

[37] M. Wetter, T. Stephane Nouidui, C. Brooks, E. A. Lee, D. Lorenzetti, and A. Roth, “Prototyping the Next Generation EnergyPlus Simulation Engine,” en, Dec. 2015. DOI: 10.26868/25222708.2015.2419.

[38] M. Ferrando, F. Causone, T. Hong, and Y. Chen, “Urban building energy modeling (UBEM) tools: A state-of-the-art review of bottom-up physics-based approaches,” en, *Sustainable Cities and Society*, vol. 62, p. 102 408, Nov. 2020. DOI: 10.1016/j.scs.2020.102408.

[39] P. R. Armstrong, S. B. Leeb, and L. K. Norford, “Control with Building Mass— Part I: Thermal Response Model,” en, *ASHRAE Transactions*, p. 14, 2006.

[40] P. Westermann and R. Evins, "Surrogate modelling for sustainable building design – A review," en, *Energy and Buildings*, vol. 198, pp. 170–186, Sep. 2019. DOI: 10.1016/j.enbuild.2019.05.057.

[41] ASHRAE, *ANSI/ASHRAE Standard 90.1 Energy Standard for Sites and Buildings Except Low-Rise Residential Buildings*, New York, 2022.

[42] Y. Chen and T. Hong, "Impacts of building geometry modeling methods on the simulation results of urban building energy models," *Applied Energy*, vol. 215, pp. 717–735, Apr. 2018. DOI: 10.1016/j.apenergy.2018.02.073.

[43] T. Dogan and C. Reinhart, "Shoeboxer: An algorithm for abstracted rapid multi-zone urban building energy model generation and simulation," en, *Energy and Buildings*, vol. 140, pp. 140–153, Apr. 2017. DOI: 10.1016/j.enbuild.2017.01.030.

[44] F. Battini, G. Pernigotto, and A. Gasparella, "A shoeboxing algorithm for urban building energy modeling: Validation for stand-alone buildings," en, *Sustainable Cities and Society*, vol. 89, p. 104 305, Feb. 2023. DOI: 10.1016/j.scs.2022.104305.

[45] N. Buckley, G. Mills, S. Letellier-Duchesne, and K. Benis, "Designing an Energy-Resilient Neighbourhood Using an Urban Building Energy Model," en, *Energies*, vol. 14, no. 15, p. 4445, Jan. 2021. DOI: 10.3390/en14154445.

[46] P. Zhu, D. Yan, H. Sun, J. An, and Y. Huang, "Building Blocks Energy Estimation (BBEE): A method for building energy estimation on district level," en, *Energy and Buildings*, vol. 185, pp. 137–147, Feb. 2019. DOI: 10.1016/j.enbuild.2018.12.031.

[47] ASHRAE, *ANSI/ASHRAE Standard 140 Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs*, 2017.

[48] R. H. Henninger and M. J. Witte, "EnergyPlus Testing with ANSI/ASHRAE Standard 140-2001 (BESTEST)," en, Lawrence Berkeley National Laboratory, Berkeley, California, Tech. Rep., Jun. 2004.

[49] D. Chakraborty and H. Elzarka, "Performance testing of energy models: Are we using the right statistical metrics?" *Journal of Building Performance Simulation*, vol. 11, no. 4, pp. 433–448, Jul. 2018. DOI: 10.1080/19401493.2017.1387607.

[50] S. Martínez, P. Eguía, E. Granada, A. Moazami, and M. Hamdy, "A performance comparison of multi-objective optimization-based approaches for calibrating white-box building energy models," *Energy and Buildings*, vol. 216, p. 109 942, Jun. 2020. DOI: 10.1016/j.enbuild.2020.109942.

[51] A. Oraiopoulos and B. Howard, "On the accuracy of Urban Building Energy Modelling," *Renewable and Sustainable Energy Reviews*, vol. 158, p. 111 976, Apr. 2022. DOI: 10.1016/j.rser.2021.111976.

[52] A. Y. Arsano, "CLIMATE—CARBON—EQUITY: Making Sustainable Design Concepts Accessible for All," Ph.D. dissertation, MIT, Cambridge, MA, Feb. 2022.

[53] S. Nagpal, C. Mueller, A. Aijazi, and C. F. Reinhart, "A methodology for auto-calibrating urban building energy models using surrogate modeling techniques," en, *Journal of Building Performance Simulation*, vol. 12, no. 1, pp. 1–16, Jan. 2019. DOI: 10.1080/19401493.2018.1457722.

[54] K. Menberg, Y. Heo, and R. Choudhary, "Sensitivity analysis methods for building energy models: Comparing computational costs and extractable information," en, *Energy and Buildings*, vol. 133, pp. 433–445, Dec. 2016. DOI: 10.1016/j.enbuild.2016.10.005.

[55] S. Fathi, R. Srinivasan, A. Fenner, and S. Fathi, "Machine learning applications in urban building energy performance forecasting: A systematic review," English, *Renewable & Sustainable Energy Reviews*, vol. 133, p. 110 287, Nov. 2020. DOI: 10.1016/j.rser.2020.110287.

[56] P. Westermann, M. Welzel, and R. Evins, "Using a deep temporal convolutional network as a building energy surrogate model that spans multiple climate zones," en, *Applied Energy*, vol. 278, p. 115 563, Nov. 2020. DOI: 10.1016/j.apenergy.2020.115563.

[57] Y. Zhang, B. K. Teoh, M. Wu, J. Chen, and L. Zhang, "Data-driven estimation of building energy consumption and GHG emissions using explainable artificial intelligence," en, *Energy*, vol. 262, p. 125 468, Jan. 2023. DOI: 10.1016/j.energy.2022.125468.

[58] Z. Wang, T. Hong, and M. A. Piette, "Building thermal load prediction through shallow machine learning and deep learning," en, *Applied Energy*, vol. 263, p. 114 683, Apr. 2020. DOI: 10.1016/j.apenergy.2020.114683.

[59] A. Nutkiewicz, Z. Yang, and R. K. Jain, "Data-driven Urban Energy Simulation (DUE-S): Integrating machine learning into an urban building energy simulation workflow," en, *Energy Procedia*, vol. 142, pp. 2114–2119, Dec. 2017. DOI: 10.1016/j.egypro.2017.12.614.

[60] Y. Li, C. Wang, S. Zhu, J. Yang, S. Wei, X. Zhang, and X. Shi, "A Comparison of Various Bottom-Up Urban Energy Simulation Methods Using a Case Study in Hangzhou, China," en, *Energies*, vol. 13, no. 18, p. 4781, Sep. 2020. DOI: 10.3390/en13184781.

[61] W. Tian, C. Zhu, Y. Sun, Z. Li, and B. Yin, "Energy characteristics of urban buildings: Assessment by machine learning," en, *Building Simulation*, vol. 14, no. 1, pp. 179–193, Feb. 2021. DOI: 10.1007/s12273-020-0608-3.

[62] J. Vazquez-Canteli, A. Dilsiz Demir, J. Brown, and Z. Nagy, "Deep neural networks as surrogate models for urban energy simulations," en, *Journal of Physics: Conference Series*, vol. 1343, no. 1, p. 012 002, Nov. 2019. DOI: 10.1088/1742-6596/1343/1/012002.

[63] V. Todeschi, R. Boghetti, J. H. Kämpf, and G. Mutani, "Evaluation of Urban-Scale Building Energy-Use Models and Tools—Application for the City of Fribourg, Switzerland," en, *Sustainability*, vol. 13, no. 4, p. 1595, Jan. 2021. DOI: 10.3390/su13041595.

[64] Z. Berzolla, Y. Q. Ang, S. Letellier-Duchesne, and C. Reinhart, *A framework for city-scale modeling of carbon reduction pathways for existing buildings*, en, 2023.

[65] S. Wolk, Z. Berzolla, L. Carethers, and C. Reinhart, "Accelerating photovoltaic potential simulations for urban building energy modeling to inform policymakers," Sep. 2023. DOI: 10.26868/25222708.2023.1389.

[66] J. Amanatides and A. Woo, "A fast voxel traversal algorithm for ray tracing," *Proceedings of EuroGraphics*, vol. 87, Aug. 1987.

[67] K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, 2015.

[68] A. E. Orhan and X. Pitkow, *Skip connections eliminate singularities*, 2018.

[69] K. S. Anderson, C. W. Hansen, W. F. Holmgren, A. R. Jensen, M. A. Mikofski, and A. Driesse, "Pvlib python: 2023 project update," *Journal of Open Source Software*, vol. 8, no. 92, p. 5994, 2023. DOI: 10.21105/joss.05994.

[70] S. Horowitz, M. Pathak, E. Wilson, *et al.*, *ResStock™*, Aug. 2019.

[71] ASHRAE, *Guideline 14-2014, Measurement of Energy,Demand and Water Savings.* Atlanta, GA, 2014.

[72] N. L. Jones and C. F. Reinhart, "Effects of real-time simulation feedback on design for visual comfort," *Journal of Building Performance Simulation*, vol. 12, no. 3, pp. 343–361, May 2019. DOI: 10.1080/19401493.2018.1449889.

[73] R. Danhaive and C. Mueller, "Combining parametric modeling and interactive optimization for high-performance and creative structural design," in *Proceedings of the International Association for Shell and Spatial Structures (IASS)*, Amsterdam, Aug. 2015.

[74] M. Bendewald, D. Miller, and S. Muldavin, "How to Calculate and Present Deep Retrofit Value: A Guide For Investors," Rocky Mountain Institute, Boulder, CO, USA, Tech. Rep., Apr. 2015, p. 122.

[75] P. N. N. Laboratory, "Advanced Energy Retrofit Guides: Office Buildings," U.S. Department of Energy, Tech. Rep., Sep. 2011, Contract DE-AC05-76RLO 1830.

[76] Z. Ma, P. Cooper, D. Daly, and L. Ledo, "Existing building retrofits: Methodology and state-of-the-art," en, *Energy and Buildings*, Cool Roofs, Cool Pavements, Cool Cities, and Cool World, vol. 55, pp. 889–902, Dec. 2012. DOI: 10.1016/j.enbuild.2012.08.018.

[77] Z. Berzolla, T. Meng, and C. Reinhart, "Homeowners' Willingness to Pay for Residential Building Retrofits," en, *SSRN Electronic Journal*, 2023. DOI: 10.2139/ssrn.4536734.

[78] Z. Berzolla, Y. Q. Ang, and C. Reinhart, "Combining Urban Building Energy Models with Retrofit Adoption Models for Time-Dependent Carbon Emission Projections and Equity Implications," en, Asilomar, CA, USA, Aug. 2022.

[79] A. Kirkeby, "How much does it really cost? A dynamic approach to building retrofit costs for decarbonization pathways," M.S. thesis, Massachusetts Institute of Technology, 2024.

[80] Z. De Simone, "Developing frameworks for an equitable future: From building decarbonization to generative modeling," M.S. thesis, Massachusetts Institute of Technology, 2024.

[81] R. E. Weber, C. Mueller, and C. Reinhart, "Automated floorplan generation in architectural design: A review of methods and applications," *Automation in Construction*, vol. 140, p. 104 385, Aug. 2022. DOI: 10.1016/j.autcon.2022.104385.

[82] *EnergyPlus™, version 22.2.0*, Sep. 2022.