Advancing Urban Building Energy Modeling with Satellite-Derived Microclimate Data

Amanda Worthy

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2026

Reading Committee:

Narjes Abbasabaddi, Chair

Julian Marshall

Mehdi Ashayeri

Program Authorized to Offer Degree:

Civil and Environmental Engineering

University of Washington

Abstract

Advancing Urban Building Energy Models with Satellite-Derived Microclimate Data

Amanda Worthy

Chair of the Supervisory Committee:

Narjes Abbasabadi

Department of Architecture

Building energy models typically reference aggregated, non-urban specific weather and climate datasets that overlook urban microclimate conditions. This embedded shortfall leads to substantial differences between the modeled and actual performance of buildings, and thus to inaccurate energy demand-side calculations. Our research aims to close this gap by integrating microclimate information, derived from earth observational and remote sensing datasets, into urban building energy models. We evaluate the value of these products by conducting two complementary studies that examine the integration of satellite-derived microclimate information in both data-driven and physics-based building energy modeling workflows. First, we develop a bottom-up, data-driven urban building energy modeling framework that combines and integrates earth observational microclimate data, spatially interpolated Typical Meteorological Year data, and annual energy usage data, measured by the Seattle Energy Benchmarking Dataset, to capture the impacts of microclimates on urban building energy performance. Using machine learning techniques and Seattle, Washington, USA as a proof of concept, we compare predictive model performance across multiple climate input scenarios; ultimately validating the effectiveness of using earth-observational data inputs to address simulation-to-real modeled uncertainties in microclimate integration. Second, we conduct a hybrid study, that augments physics-based residential building energy consumption insights with earth observational microclimate data using machine learning predictions. This approach produces a spatial heat map of residential energy consumption for a typical family home across Los Angeles County, USA, that is explicitly tailored to reflect urban microclimate

variation. Here, we confirm current building EnergyPlus weather file (EPW) sampling sites to be in lower vulnerability areas with fewer streets and buildings than the city average. This result identifies a mismatch between the environmental conditions observed in dense urban areas and those normally simulated in building energy modeling protocols, further underscoring the structural misrepresentations that are embedded in current frameworks. Throughout this work, we emphasize the value in integrating satellite-derived microclimate products into data-driven building energy studies. However, we cite obstacles in integrating satellite-derived microclimate data directly into physics-based urban building energy models, suggesting future research to explore more spatially and temporally compatible datasets that measure EnergyPlus weather file parameters and respective downscaling opportunities. Despite these challenges, the Landsat thermal band 10, or land surface temperature products, show strong potential in being effective scalable proxies for incorporating microclimate effects into both data-driven and physics-based urban building energy studies. This research advocates for the integration and validation of urban microclimate effects into building energy modeling frameworks, to enable more accurate, just, and precise energy policy and planning.

Table of Contents

Chapter 1: Introduction

Urbanization presents opportunities for achieving more sustainable social, economic, and environmental development [1]. Growing urban populations have caused infrastructure developments to become more centralized, interconnected [2], and vertical [3], leading to substantial shifts in land use patterns and environmental characteristics [4]. Rising network density has altered the surface energy balances, thermal properties [5],[6],[7] native foliage concentrations [6],[7], weather patterns [8], fluid flow characteristics [9],[10], and albedo reflectance's [11] of urban areas; consequently leading to a phenomenon where cities experience greater warming than surrounding less developed areas, formally known as the Urban Heat Island (UHI) effect [12],[13],[14],[15]. These altercations in urban form have exacerbated urban microclimate conditions in cities—defined as locations where differences in outdoor climate are observed between urban and rural areas when referenced in similar geographical areas [10]. Urban microclimate conditions are important from a building energy infrastructure perspective because they influence the environmental conditions outside of buildings, thereby substantially impacting the energy demanded by heating, cooling, and ventilation systems to meet healthy environmental quality metrics indoors [16]. Understanding these microclimate conditions is essential for interpreting the interactions between the built environment and urban energy dynamics, ensuring cities remain sustainable and resilient.

Populations are disproportionately exposed to urban microclimate conditions and the UHI effect, raising sustainable social development concerns for our cities. In the continental United States (U.S.), populations of color face higher levels of UHI exposure compared to non-Hispanic white populations in 96.6% of the largest urban areas [17]. Additionally, low income US census blocks have 15.2% less tree cover and experience temperatures that are 1.5% hotter than high income blocks [18], [19]. Since dense vegetation provides roughly 6% energy savings in buildings, similar to the magnitude of savings from mechanical retrofit interventions [20], these disparities are not only a concern of environmental exposure, but also may impose an increased energy burden on certain communities as they try to maintain comfortable temperatures indoors. In Portland, Oregon, it is documented that the energy burden due to extreme heat disproportionately effects renters [21], populations who are often disadvantaged with lower incomes. Therefore, this research project aims to not only to impose energy infrastructure solutions, but also to identify city planning strategies that reduce energy burdens, promoting equitable thermal comfort and energy consumption across neighborhoods [22].

To investigate the changes in urban climate and their respective implications to human and environmental health, research has broadened its scope to examine the interconnectivities between building energy infrastructure and its interactions within urban spaces. Urban Building Energy Modeling (UBEM) is a powerful tool used to analyze energy consumption patterns and optimize building performance within urban contexts in order to fully understanding energy dynamics in cities [23],[24],[25]. Unlike Building Performance Simulation (BPS) tools, which focus on single buildings, UBEMs examine clusters of buildings, accounting for inter-building connections, and broader urban influences including urban microclimate effects. UBEM methodologies include top-down approaches, which examine cities at the macro-scale, and bottom-up approaches, which consider urban attributes at the microscale of individual units (ex. single buildings), with the latter classified into three categories data-driven (including statistical, machine learning (ML), and artificial intelligence (AI)) models, engineering (physics-based) models, and finally hybrid models, which combine the two [25], [26]. Data-driven models use data techniques such as ML, AI, and statistical modeling methods to provide insight into real world system operation given dynamic data inputs [27]. Whereas physics-based, or simulation-based, models use numerical-based conventions to represent energy consumption through the physical and technical characteristics of buildings [26]. UBEMs rely heavily on input parameters from diverse datasets to holistically characterize energy demands, including building geometry, occupancy data, localized weather and climate, and urban scale parameters [27].

UBEMs support sustainable design, retrofit, planning, policy-making, and resource allocation by generating quantitative energy insights that address energy dynamics across neighborhoods and cities [25],[18], [22]. However, despite significant advancements in recent years, UBEMs face notable limitations in accurately representing urban energy usage, resulting in a substantial gap between the simulation results and actual measured energy usage[26]. Some of these challenges stem from the uncertainties in generalized climate data inputs, which are often comprised of historically aggregated, non-urban specific measurements that are often not representative of climates in urban spaces [16]. [24]. These inputs limit model abilities to capture critical variable interactions between energy systems and the environment at high temporal and spatial precision, exacerbating inaccuracies across modeling results [16]. Additionally, dynamic factors, including socio-economic variables, are often excluded from UBEMs, neglecting necessary considerations for promoting sustainable economic, social, and environmental development [30], [31], [32], [33].

To combat these challenges, this manuscript aims to address and propose solutions for integrating microclimate data, specifically sourced from satellite datasets, into data-driven and simulation-based urban building energy models, while also exploring the respective correlations to infrastructure equity. The manuscript responds to the following objectives:

1. Identifies datasets, tools, methods, limitations, and opportunities for integrating microclimates into Urban Building Energy Modeling (UBEM).
2. Examines the capacity for integrating earth observational climate and weather input datasets into UBEMs using a data-driven statistical approach.
3. Investigates the capacity for integrating GIS-based urban morphology characteristics and earth observational climate data into simulation-based energy models.
4. Addresses future research applications and opportunities for integrating satellite-derived microclimate datasets into building energy models.

In Chapter 2, we conduct a literature review to introduce the datasets, tools, methods, limitations, and opportunities for integrating microclimates into Urban Building Energy Models (UBEMs). In this chapter we underscore the needs of local microclimate data to improve UBEM accuracy, we depict the spatial and temporal tradeoffs of different microclimate datasets, and we tabulate common microclimate parameters that influence building energy performance. Given these contexts, we propose integrating earth observational weather and microclimate datasets into UBEMS to reduce environmental modeling uncertainties and to ultimately make building energy models more accurate. It should be noted that Chapter 2 is a copy of my first-author paper "*Bridging the simulation-to-reality gap: A Comprehensive review of microclimate integration in urban building energy modeling (UBEM)*," which has been published in Energy and Buildings [16].

In Chapter 3, we build off findings from Chapter 2 and conduct a data-driven UBEM study that examines urban building energy model performance improvements given different microclimate datasets. In this study we compare three variable schemas with differing weather and microclimate parameters; including: a variable schema with no climate data inputs, a variable schema with traditionally use TMY3 climate data inputs, and a variable schema that couples both traditionally used TMY3 and satellite derived microclimate data inputs and compare them by referencing different machine learning modeling algorithms. In this chapter, we prove earth observational weather and climate data to be reliable for microclimate integration in data-driven UBEMs. Further, we tabulate the case study specific microclimate parameters that are best at predicting building energy performance and suggest future research to explore their contributions. Chapter 3 is a copy of my first-author paper, "*Leveraging Earth Observational Data and Machine Learning to Enhance Urban Building Energy Modeling with Microclimate Effects*", which has been published in *Sustainable Cities and Society* [34].

In Chapter 4, we build off findings from chapters 2 and 3; and propose a hybrid methodology to reduce UBEM microclimate gaps that augments simulation-based building energy models with satellite-derived microclimate datasets. In order to make recommendations for future sampling placements and procedures, we first determine the urban morphological and climate characteristics at energy plus weather file (.epw) sampling sites in Los Angeles, CA and then compare them to the city average. We find that the TMYx measurements stations are in regions in LA with lower CDC/ATSDR Social Vulnerability Index scores and significantly fewer streets, buildings, and trees compared to the city average. Therefore, to include more urban representative conditions, we call for future placement of TMYx stations to be in locations in the city that are classified high vulnerability, near more buildings, trees, and streets. It should be noted that our augmented framework does not fully resolve the challenge of integrating remote sensing datasets into physics-based building energy modeling tools. We give a larger discussion on these limitations and purpose future research opportunities in Chapter 5. Chapter 4 is a copy of my first-author paper "*A framework for augmenting simulation-based building energy models with earth-observational microclimate data using machine learning predictions*", which has is under review in *Urban Science*.

In Chapter 5, we elaborate on future opportunities for integrating satellite-derived microclimate data into urban building energy models. We consolidate and aggregate the findings from the previous four chapters, discussing the benefits and limitations of earth observational microclimate data in both UBEM data-driven and simulation-based studies. Ultimately, we encourage future scientific studies to explore sustainable urban development and energy infrastructure solutions, promoting a more resilient urban future.

Chapter 2: Bridging the simulation-to-reality gap: A comprehensive review of microclimate integration in urban building energy modeling (UBEM)

Authors: Amanda Worthy[1], Mehdi Ashayeri[2], Julian Marshall[1], and Narjes Abbasabadi[3*]

[1] Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, United States
[2] School of Architecture, Southern Illinois University, Carbondale, IL, United States
[3] Department of Architecture, University of Washington, Seattle, WA, United States

[*]**Corresponding Author:**

Narjes Abbasabadi, Ph.D.

Assistant Professor
Director, Sustainable Intelligence Lab (SILab)
School of Architecture
College of Built Environments
University of Washington
nabbasab@uw.edu

nabbasab@uw.edu
https://arch.be.uw.edu/
https://arch.be.uw.edu/people/

Abstract

Buildings are significant contributors to global energy consumption, necessitating urgent action to reduce energy use and associated emissions. Urban Building Energy Modeling (UBEM) is a critical tool that provides essential insights into citywide building energy dynamics though generating quantitative energy data and enabling holistic analysis and optimization of energy systems. However, current UBEM methodologies and tools are constrained by their reliance on non-urban-specific and aggregated climate data inputs, leading to discrepancies between modeled and actual energy expenditures. This article presents a comprehensive review of the datasets, tools, methodologies, and novel case studies deployed to integrate microclimates into UBEMs, aiming to bridge the modeling gap and to address the uncertainties due to the absence of real-world microclimate data in the models. It expands beyond conventional methods by elaborating on substitutional observational-based and simulation-based datasets, addressing their spatial and temporal tradeoffs. The review highlights that while remote sensing technologies are extensively utilized for building geometric data UBEM inputs, there remains an underexplored potential in reanalysis and observational-based products for environmental data; specifically, for the inclusion of parameters that are conventionally not included in UBEM analysis such as tree canopy coverage and land surface temperature. Furthermore, adopting a hybrid methodology, which combines observational and simulation-based datasets, may be a promising approach for more accurately representing microclimate conditions in UBEMs; as this process would ensure more representative climate parameter inputs and ground-truthing, while effectively managing computational demands across extensive temporal and spatial simulations. This could be achieved through integrating local earth observation datasets with computational fluid dynamics (CFD) tools or by merging local earth observational data with simulation-based reanalysis products and coupling these weather inputs with simulation-based building energy management models. Finally, this review underscores the importance of validating UBEMs with local microclimate weather data to ensure that model results are actionable, reliable, and accurate.

Keywords: Urban Building Energy Modeling (UBEM), Urban Microclimates, Simulation-based and Observational-based Data, Building Energy Demand

Highlights:

- Reviews datasets, tools, and methods for integrating microclimates into UBEMs.
- Underscores the need for local microclimate data to improve UBEM accuracy.
- Depicts spatial and temporal tradeoffs of different microclimate datasets.
- Tabulates microclimate parameters influencing building energy performance.
- Proposes a hybrid UBEM approach to reduce model uncertainties in weather and climate data inputs.

1. Introduction

The percentage of global population residing in cities is projected to increase to 68% by 2050 [35]. This rapid urbanization has led to denser buildings, infrastructure, and increased inter-building connections [2], as well as significant changes in land use patterns in cities [4]. These developments have intensified the Urban Heat Island (UHI) effect [12], [14], characterized by elevated urban temperatures [15], alerted albedo [11], reduced native foliage concentrations [6], [7], and disrupted surface energy balances and thermal properties [5]. Additionally, urban morphological factors, such as taller buildings and higher skylines, have influenced wind patterns and turbulence [9] by creating large shadows and localized thermal eddy trappings [10]. These transformations have led to the emergence of microclimate conditions, where the local climates differ significantly from surrounding environments [36] and have thus impacted the energy demands for heating, cooling, and ventilation systems needed to maintain habitable conditions indoors. Understanding these microclimate conditions is essential for interpreting the interactions between the built environment and urban energy dynamics, ensuring cities remain sustainable and resilient.

Abbreviations

BEM	Building Energy Management
BPS	Building Performance Simulation
CDF	Computational Fluid Dynamics
CDD	Cooling Degree Day
DEM	Digital Elevation Models
EO	Earth Observation
EMR	Electromagnetic Radiation
HDD	Heating Degree Day
HI	Heat Island
IEQ	Indoor Environmental Quality
LES	Large Eddy Simulation
LiDAR	Light Detection and Ranging
LST	Land Surface Temperature
NDVI	Normalized Difference Vegetation Index
NWP	Numerical Weather Prediction
RANS	Reynolds Averaged Navier Stokes
UBEM	Urban Building Energy Modeling
UCM	Urban Canopy Model
UHI	Urban Heat Island
UHII	Urban Heat Island Index
SWIR	Shortwave InfraRed
TIR	Thermal Infrared
TMY	Typical Meteorological Year
VNIR	Visible and Near-Infrared

Urban Building Energy Modeling (UBEM) has emerged as a powerful tool for analyzing energy patterns and optimizing building performance within urban contexts [23], [24], [25]. To fully understand energy dynamics, the scope of energy modeling must extend beyond the individual building level and include interactions between buildings and their urban contexts [37]. Unlike Building Performance Simulation (BPS) tools, which focus on single buildings, UBEM examines clusters of buildings, accounting for inter-building connections, and broader urban influences. UBEM methods vary based on the scale (e.g., block, neighborhood, or city) and temporal scope (e.g., daily, episodic, or annual) [24]. These methods fall into two major bottom-up approaches: data-driven and physics-based models [26], as elaborated on in Section 4. Emerging hybrid approaches combine the two methodologies to enhance the reliability and computational efficiency of UBEMs [26], [38]. UBEM relies on diverse datasets, including building geometry, occupancy data, localized weather and climate, and urban scale parameters, to holistically characterize energy demands [27]. By generating quantitative energy insights and addressing energy dynamics across neighborhoods and cities, UBEM supports sustainable design, retrofit, planning, policy-making, and resource allocation [25],[18], [22].

However, despite significant advancements in recent years, UBEM faces notable limitations in accurately representing urban energy use, resulting in a substantial gap between simulation results and actual measured energy data [26]. These challenges stem from uncertainties in large-scale models, reliance on oversimplified archetypes, and generalized climate inputs. The complexity of urban energy modeling arises not only from thermodynamic systems but also from the nonlinear interactions among diverse and dynamic urban

elements, such as urban contexts and microclimates [39], [40], [41]. Furthermore, dynamic factors, including localized microclimates, occupancy patterns, and socio-economic variables, are often excluded from these models, limiting their ability to capture critical variables such as temperature gradients, wind patterns, and solar radiation variations, all of which significantly influence energy performance [30], [31], [32], [33].

Expansive climatic differences worldwide have been shown to drastically impact building energy demands [42], [43]. Additionally, the proximity of buildings to surrounding infrastructure further contributes to energy performance variations, with urban morphology factors such as building density, interconnections, and the Urban Heat Island (UHI) effect influencing microclimate conditions [24], [44], [45], [46]. The relationships among UHI intensity, urban compactness, and building energy demands are well-documented, emphasizing how urban microclimate dynamics alter energy loads and performance [13], [47], [48], [49], [50]. These challenges are compounded by issues in data resolution, sufficiency, and methodological robustness, which hinder the ability of UBEM to provide accurate and actionable insights [41], [51], [52], [53], [54]. Addressing these gaps is critical for improving the accuracy and applicability of UBEM in evolving urban landscapes, enabling more resilient and sustainable urban energy systems.

The primary objective of this article is to address the simulation-to-reality gap in UBEMs by offering a comprehensive review of datasets, tools, and methodologies for integrating urban microclimates into their frameworks. It aims to enhance the accuracy and applicability of UBEM by examining how localized microclimate variables can effectively inform urban energy dynamics. The article evaluates methods for predicting building energy performance in urban contexts, presenting a detailed analysis of key literature, datasets, and tools. In this context, "datasets" represent critical climate data for energy predictions, divided into simulation-based and observational-based categories, whereas "tools" refer to the software and platforms utilized to simulate and assess the interactions between microclimates and energy systems. Additionally, it introduces a collection of novel case study methodologies, which include approaches and frameworks for integrating these datasets into UBEMs, highlighting their strengths, opportunities, and spatiotemporal trade-offs. Finally, the article provides recommendations for future research directions to enhance UBEM accuracy and performance, specifically by improving the integration of environmental data, thereby bridging the simulation-to-reality gap.

2. Methodology

The outcomes of review papers are heavily influenced by the search criteria for, and selection processes of, novel articles. Systematic procedures are used in this review to ensure an objective and repeatable review processes [55]. This paper utilizes a query-based methodology to find relevant publications within the field of interest using Web of Science (WoS) and Science Direct databases, as highlighted in *Figure 1*. Different search criteria were used for each database to strategically include both case study articles and review articles. For instance, the WoS database returned more articles on case study methodologies and frameworks, whereas Science Direct returned more review article results – thus respective databases' strengths were catered to. It should be noted that WoS and Science Direct databases often returned the same articles under their unique selection criteria, but both databases were used to diversify the sources included in this review. Beyond these databases, additional publications were screened and references, specifically to provide insights to nontrivial technicalities of earth observational technologies. The quantity of articles incorporated into this review that include aspects of UBEM and microclimates is 53 as shown in Section C of *Figure 1*.

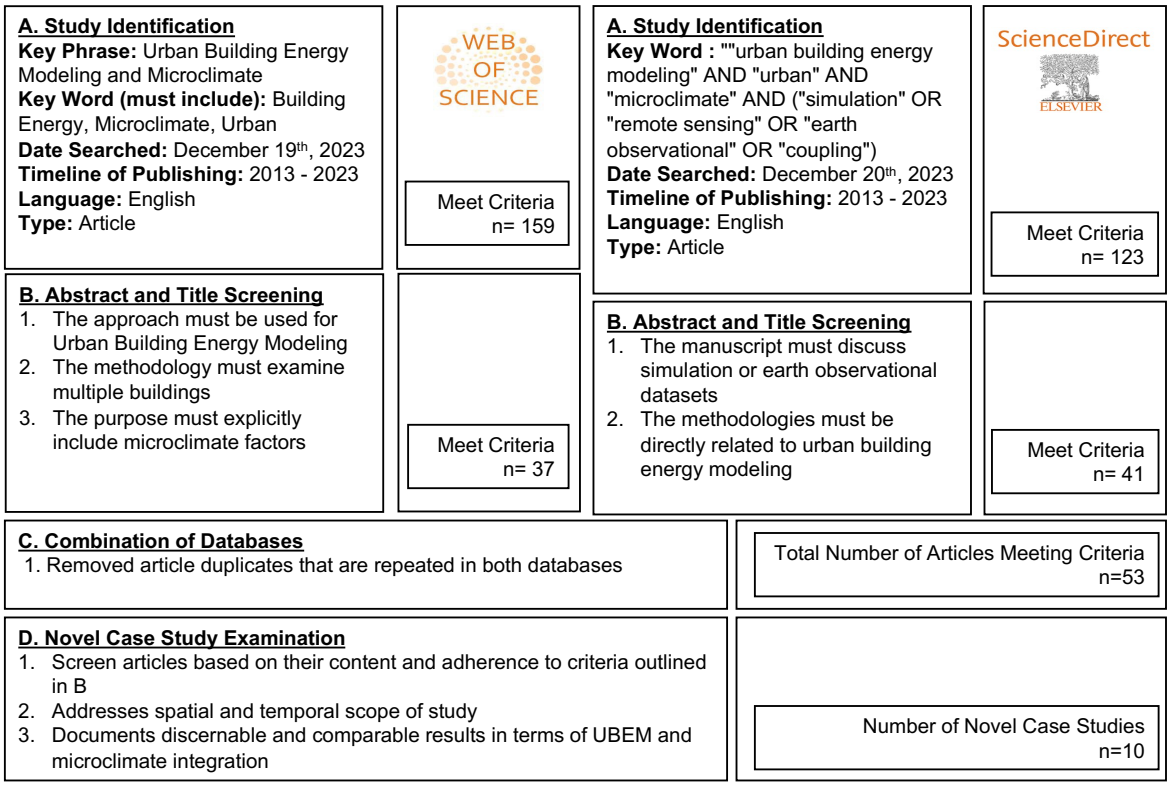

Figure 1: Stages of evaluation for the articles included in this review article.

Ten case studies, noted as unique in the field due to their innovative methodologies for incorporating microclimate data into UBEMs, are elaborated on more explicitly in this literature review and summarized in *Table 5*. The selection criteria for the case studies are characterized in Section D of *Figure 1*. We adhered to common standards used in previous literature reviews published in scientific journals to access the quality of included studies. These standards include peer-review status, citation analysis, clarity and rigor of study design and methodology, sample size and representativeness, transferability of findings, transparency, reproducibility of the study, and the use of standardized measures. This approach enables a rigorous evaluation of the reliability and validity of the studies included in the review.

3. Existing Literature on Microclimates and Energy Modeling

Various existing review articles comment on aspects of both UBEM and urban microclimates within the building energy modeling sector. For instance, [24], [25], [26],[23], [27], expand on urban building energy modeling tools and methodologies, and introduce some key tools and methodologies for including microclimate data. A review of CFD urban microclimate studies expands on applications of simulation-based datasets in the built environment [56]. This literature is complimented by [36], a systematic review article that expands on the applications of microclimate studies. Moreover, simulation based microclimate data and building energy models coupling techniques are well documented and compared in [57]. The integration of local climate data in building energy models is addressed for simulation-based datasets for building level energy models in [58]. Additionally, systematic data-driven energy prediction reviews [59] and [60] tabulate commonly used machine learning methods and climate data inputs. Another recently published UBEM and microclimate hybrid-systematic review article [61] elaborates on key terms in the field, conducts cluster analysis of recent studies, and outlines future research directions specifically focusing on applications with hot climate contexts.

This review article at present stands out due to its novel approach. Unlike previous studies, it explicitly examines urban building energy modeling at larger scale (ex. neighborhood and city) while incorporating a diverse array of microclimate datasets. This review dives deeply into both observational-based and simulation-based sources, and specifically focus on earth observational datasets, which have significant potential in addressing data gaps. This review article is the first of its kind to introduce the application of earth observational datasets in UBEMs, highlighting their benefits, limitations, and potential for future research. *Table 1* further illustrates the distinctions between this review and existing literature, underscoring the novelty and unique contributions of this study to the field.

Table 1: Existing literature review articles in this research domain

Article Name	**Type of Review**	**Application Scale**	**Climate Datasets Discussed**	**Detail**
Addressing the modeling-to-reality gap: A comprehensive review of datasets, tools, and methodologies for integrating microclimates into urban building energy models (UBEMs) [Present paper]	Review of datasets, tools, and methodologies for integrating urban microclimates into UBEMs	urban scale	TMY data, simulation-based datasets (ENVI-met, CityFFD, WFR, Meso-NH, TRNSYS, TEB, Solene-Microclimat,), observational-based datasets (sensor data, downloadable data, earth observational datasets)	This article presents a comprehensive review of the datasets, tools, methodologies, and novel case studies deployed to integrate microclimates into UBEM. It expands beyond conventional methods by elaborating on substitutional observational-based and simulation-based data types, addressing their spatial and temporal tradeoffs.
Ten questions on urban building energy modeling [24]	UBEM overview	urban scale	TMY data, simulation-based datasets (Urban Weather Generator, CFD, Weather Research and Forecasting)	Introduces UBEM methods, applications, challenges, opportunities, and future research directions.
Data acquisition for urban building energy modeling: A review [27]	Review of UBEM data inputs	urban scale	TMY data, observational-based datasets (local weather stations, weather underground), and future weather data (Urban Weather Generator)	This literature review outlines baseline information acquiring all input data for UBEMS. It touches on different aspects of collecting weather and climate data and addresses future weather sources; although weather data is not the preliminary focus of the article.
Urban energy use modeling methods and tools: A review and an outlook [26]	Review of UBEM methodologies (simulation based, data-driven, hybrid)	urban scale	Simulation-based datasets (CFD, Urban Multiscale Environmental Predictor, ENVI-met)	This review article mainly focuses on addressing the differences between data-driven, hybrid, and simulation based UBEM methodologies; highlighting the tradeoffs of each.
How building energy models take the local climate into account in an urban context – A review [58]	Literature review of the simulation-based datasets for building energy models	building scale	Simulation-based datasets (Meso-NH, TEB, WFR, Urban Canopy Model, Building Effect Parameterization, UWG, Canyon Air Temperature, Canopy Interface Model, ENVI-met, Solene-Microclimat)	This literature review gives an overview of Urban Climate Model (UCM) and Building Energy Model (BEM) coupling and chaining strategies; elaborating specifically on 9 different configurations.
A review on the CFD analysis of urban microclimate [56]	Review of CFD microclimate studies for a wide variety of research applications	N/A	Simulation-based microclimate modeling software and equations (ENVI-met -most popular)	This review discusses 183 CFD studies on urban microclimate, addressing key tools, equations, and applications.
A review of data-driven building energy consumption prediction studies [60]	Systematic review of data-driven energy prediction studies	building scale, urban scale	Observational-based datasets (sensor data, downloadable data)	This review conducts a meta-analysis of data-driven energy prediction studies, outlining their applications (heating, cooling, ect), data-driven methods, and metrics for assessing model performance.

Urban microclimate and its impact on built environment – A review [36]	Systematic review of urban microclimate studies for a wide variety of research applications	N/A	Observational-based datasets (field measurmeents), Simulation based datasets (CFD, ENVI-met, Fluent, OpenFOAM)	This article highlights the latest progress in urban microclimate research, addressing traditional methods, field measurements, wind tunnel modeling, CFD, and emerging data driven studies. It reviews 563 publications on urban microclimate.
Correlating the urban microclimate and energy demands in hot climate Contexts: A hybrid review [61]	Hybrid-systematic review on UBEM and microclimate	urban scale	Simulation-based datasets (ENVI-met, OpenFOAM, UWG, WRF-UCM)	This article conducts a systematic review of UBEM and microclimate studies. It elaborates on key terms in the field, conducts cluster analysis of the studies, and outlines future research directions specifically focusing on applications with hot climate contexts.
Urban microclimate and building energy models: A review of the latest progress in coupling strategies [57]	Review of microclimate and energy model coupling strategies	building scale, urban scale	Simulation-based datasets (CityFFD, OpenFOAM, TRNSYS, Fluent, UWG, Green Building Studio, Solene-Microclimat, IES, ENVI-met, ESP-r)	This review article addresses the coupling strategies between urban microclimate and building energy models, elaborating on opportunities, limitations, and future research directions.

4. Microclimate Datasets and Tools for UBEM Inputs

This review examines effective methods for accurately predicting building energy performance in urban contexts through a comprehensive analysis of key datasets and tools for integrating microclimate variables into UBEM frameworks. Microclimates, defined as encompassing a spatial scale of up to 120 meters vertically in the atmosphere [62], [63] and up to 2 kilometers horizontally [64], as illustrated in *Figure 2*, significantly impact building energy demands. These impacts arise from variations in temperature, solar radiation, wind patterns, humidity levels, and precipitation in urban areas compared to rural settings with similar climate zones. In this context, "datasets" refer to climate data critical for UBEM energy predictions, categorized into simulation-based and observational-based types, while "tools" denote the software and platforms used for modeling and analyzing the interactions between microclimates and energy systems.

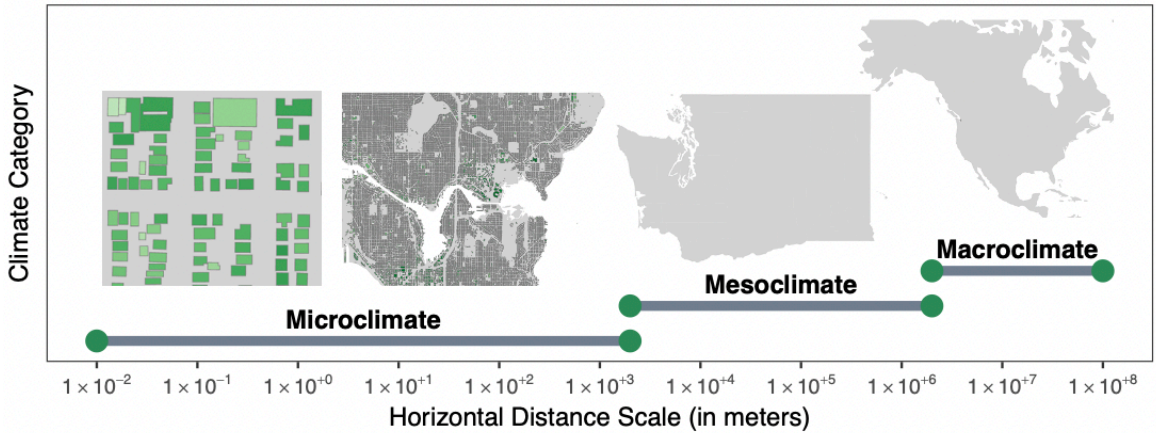

Figure 2: Spatial extent of climate categories: microclimate, mesoclimate, and macroclimate based on horizontal distance scale

Common UBEM models use historically aggregated weather data for their weather and climate component inputs. These datasets, often in form of a Typical Meteorological Year (TMY) [65], are conventionally collected in remote open areas that are normally absent of urban context terrain, [24]. TMY datasets contain measurements of hourly weather observations over only a 12 meteorological month calendar, although collected from multiple years and aggregated into one year [65]. Due to their specific aggregation process,

TMY datasets fail to account for energy estimations during extreme weather events, such as heat waves and cold waves, periods responsible for peak thermal loadings and where infrastructure reliability is pivotal for the protection of human health. Furthermore, these datasets fail to include microclimate parameters known to influence urban building energy consumption such as albedo, vegetation density, water availability in the soils, land surface temperature, etc. Additionally, these datasets do not account for changing climate conditions and do not report inputs at a high enough spatial resolution to describe the changes to climate and weather data in urban environments, ignoring the inferences by physical structures and their corresponding human activities, and thus exacerbating uncertainties within modeling results.

To account for spatiotemporal resolution in climate data, numerous UBEM methods have been developed to substitute, modify, and replace the conventional TMY weather data file inputs. These UBEM climate data inputs can be categorized into observational-based and simulation-based approaches [24], each offering unique tradeoffs in spatial and temporal granularity. These climate variable data inputs are then combined within building energy modeling platforms through various coupling and integration techniques, as elaborated on in the next section. Observational-based approaches include only measured values, such as in-situ data and remote sensing observations; whereas simulation-based approaches use numerical-based conventions to represent local climate data elements, given reference observational based data [37]. In-situ observations have been coupled into UBEM models by simulation-based platforms such as TRNSYS and CityBES. Further, earth observational data has been combined with UBEMs through data-driven techniques. *Figure 3* presents a list of the approaches, and data sources for both observational-based and simulation-based methods, whereas *Table 2* compares these dataset types. The subsequent sections will elaborate on these datasets in greater detail, highlighting the benefits and uncertainties associated with each UBEM climate and weather dataset input.

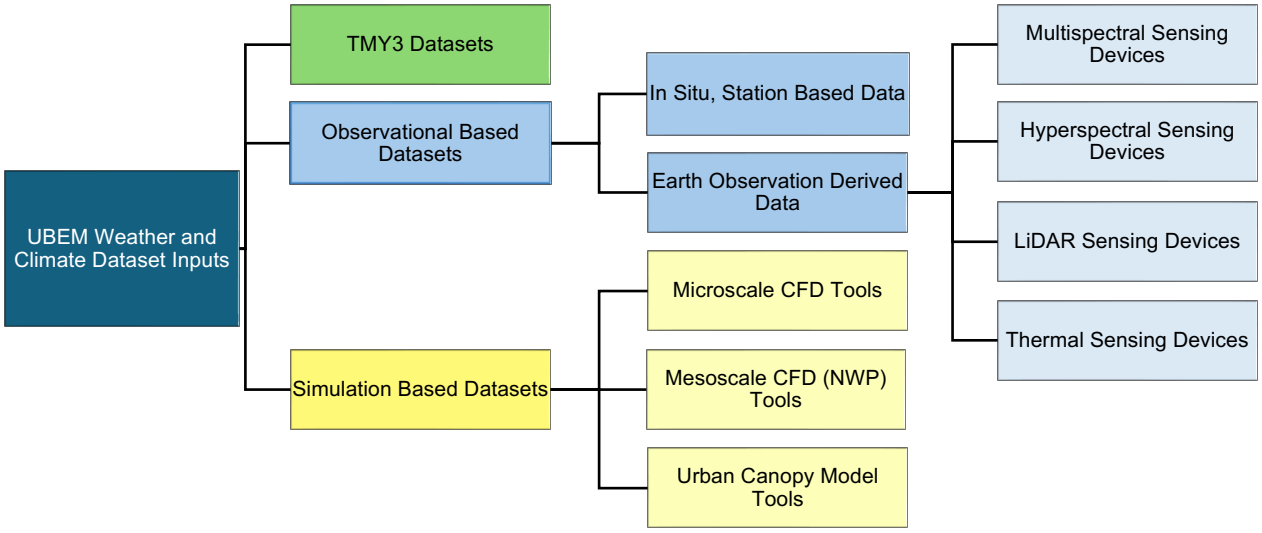

Figure 3: UBEM weather and climate dataset inputs

Table 2: Analysis of Observational-Based and Simulation-Based Climate Dataset Model Inputs with Key Strengths and Limitations Across Categories

Dataset Type		Scalability	Spatial Granularity	Temporal Granularity	Parameter Variety	Reliability	Coupling Feasibility
Observational-Based	In-Situ	–	–	✓	✓	✓	–
	Earth Observational	✓	✓	–	✓	–	–
Simulation-Based		–	✓	✓	–	–	✓

4.1. Observational-Based Microclimate Datasets

In 2020, the quantity of data created, captured, copied, and consumed worldwide was 64.2 zettabytes; by 2025, this value is projected to triple in size due to developments in ubiquitous sensing, Internet of Things (IoT), and machine learning algorithms [66]. That is, observational-based data is becoming increasingly accessible to capture and monitor - thus providing an exciting expansion in scientific progress and discoveries. In the context of UBEM applications, increased accessibility to higher spatial-temporal climate data can help researchers better understand the effects of urban microclimates on building energy performance.

Observational-based climate datasets are classified from two sources, station-based data or Earth Observation (EO) derived data [67]. Station based in-situ data, or ground-based sensor data, observe meteorological parameters from instruments located on the earth's surface. Whereas EO data is sensed with remote sensing techniques, that make observations from sensors onboard satellites and unmanned arterial vehicles located above the earth's surface, subsurface, and atmosphere, and then transmit the recordings remotely to servers on land [68]. Both forms of observational based data types offer unique tradeoffs between accuracy specifications and spatiotemporal granularity, as elaborated on in the subsequent sections.

The quality of observational data is heavily dependent on its collection and recording methods. For all observational-based studies it is important to address quality standards such as; temporal and spatial homogeneity, reliability, and metadata reporting [69]. It is imperative that observational-based data undergoes proper and precise monitoring to ensure its value. This involves adherence to World Meteorological Organization (WMO) standards [70], recording on regular time intervals in consistent locations, and continuously processing through frameworks that record sensor geographical points, instrument specifications, and insights to recording procedures [69]. For all observational recordings, the location of collection, measurement frequency, and sensing tool accuracy play pivotal roles in data quality and reliability. Additional bias mitigation strategies include using accurate (high precision) instruments and placing them in regions that are representative of the surrounding environment (not near heat sources, in sunlight, or in drafts).

4.1.1. Station-Based, In-Situ Data

Station-based, or in-situ microclimate data used in urban building energy modeling studies take on diverse ranges of sourcing patterns and are collected at wide ranges of spatiotemporal resolutions. Due to the diverse availability of devices, there is a spectrum of data collection methods present for in-situ measurement, causing considerable variation in the procedures for deploying instrumentation. Observational-based in-situ data can be accessed using web-based portals, through crowd sourcing and citizen science, or by context specific instrumentation deployment. Some of the most common point source, or localized and stationary, weather observational datasets are hosted by open-source web-based entities. These sites aggregate weather information and forecasts for the majority of the largest US and international cities; Example of sites with free weather download capabilities include Weather Underground [71], Open Sky [72], and the National Oceanic and Atmospheric Administration (NOAAs) Climate Data Online [73]. Other websites such as Cal-Adapt [74] and Open Weather Map [75] host free Application Programing Interfaces (APIs) that facilitate instantaneous data transfers between local and web based servers [76], where observations from existing weather stations can be pulled and updated in real time.

In-situ measurements often face challenges with the routine record keeping at high spatial resolution. Often, sensors with larger temporal time records are deployed for aviation applications, in regions that normally lack urban from [67]. Also, weather data that is readily available may vary it its accuracy specifications or include only a small subset of weather and climate variables (ex. temperature, windspeed, rainfall, pressure). Therefore, if weather parameters are not readily available in the desired spatial and temporal formats, researchers can deploy their own instrumentation to meet their context-specific scientific goals.

The most common in-situ weather variables collected for observational-based data-driven energy prediction studies include: dry-bulb temperature, dew point temperature, relative humidity, global solar radiation, wind speed, wind direction, degree of cloudiness, pressure, rainfall amount, and evaporation [60]. These observations are sometimes transformed for analysis like binned by distribution, such as temperature binning, normalized by heating/cooling thresholds, such as in Heating Degree Days (HDD) calculations and Cooling Degree Days (CDD) calculations, or quantified in terms of human safety, such as the Heat Index (HI) calculation and the Urban Heat Island Index (UHII) calculation [8], [56].

For instance, in-situ measurements were deployed in Rome to address the spatial and temporal granularity of microclimates bellow the urban canopy layer. In this study, air temperature and relative humidity were sampled every 10 minutes, at 5 locations across the city for three years, then coupled and simulated in TRNSYS to access the impact on urban climate for two reference buildings [77]. Furthermore, at a higher spatial resolution and lower temporal resolution, hourly weather data (dry bulb temperature, relative humidity, global horizontal radiation, and wind speed) was collected for 10 years at 27 different monitoring sites in San Francisco and combined into CityBES urban microclimate mapping feature to identify spatial patterns in urban energy expenditures during a 3-day heat wave in 2017 [10].

In synthesis, a downfall to in-situ microclimate data is its inability and lack of feasibility to characterize conditions at high-spatial granularity. For example, in the instance of urban energy studies, it becomes increasingly resource intensive and intricate to install, operate, and maintain environmental data reporting stations at such high spatial granularity (such as the building level) across the city-wide scale. There also pose challenges with the availability of programs and the assumptions made in programs when combining multi-sensor data into UBEMs. Additional challenges with data gathering methods include difficulties with synchronizing sensor sampling rates, sensor drift, sensor communication, and sensor operation [60]. Earth observation-derived data [78] products can address these challenges by providing longitudinal insight to environmental conditions through the respective use of the same sensor for each product domain (ex. air temperature).

4.1.2. Earth Observation-Derived Data

The number of articles published on remote sensing and land surface temperature in 2020 was approximately threefold of that in 2013, indicating a substantial surge in research activity within this field [79]. That is, advancements in knowledge of capturing, spatial coverage, temporal coverage, methods, and frameworks have expanded the scientific applications of remote sensing techniques. In simple terms, Earth Observation (EO)-derived data uses imagery methods sampled above earth to capture information about diverse urban climates and surface variability [80]. Due to its location of sampling, EO-derived data depicts the state of the atmosphere, land, and ocean, offering a unique and expansive set of environmental predictors for urban heat island and energy modeling studies [80]. Earth-derived satellite observations used for climate-related energy modeling studies are measured by multispectral, hyperspectral, Light Detection and Ranging (LiDAR), and thermal remote sensing devices, with each technology offering unique parameters for urban building energy modeling studies, as outlined in *Table 3*.

EO derived data has been heavily explored for building footprint generation in UBEM studies [81]. Although, methodologies for utilization and coupling of EO derived environmental data in urban energy studies are less formulated and established. Dougherty and Jain have published two studies using remote sensing observations for their environmental parameter inputs into data-driven UBEMs [82], [83]. The authors combine EO products and machine learning techniques to determine the effect or urban contexts on building energy demands in New York City. In both [82], [83] data-driven modeling studies, environmental data is easily combined or coupled with the building energy loadings through merging and synchronization of the building location with the environmental sampling location. However, these studies

do not address the pressing gap to couple EO climate data into simulation-based BEM platforms, as the temporal resolution required for BEM platforms does not meet temporal EO revisit periods. Dougherty and Jain's case studies were the first UBEM studies to utilize remote sensing environmental data inputs and advocate for the use of these EO climate datasets for future energy modeling endeavors due to their scalability and resolution.

Collecting earth observational data is very resource intensive, although because measurements are often taken on the global level, case study methodologies are reputable and easy to scale. In general, there is a temporal and spatial granularity trade off among EO data sources, where the devices with the shortest revisit period may have coarser spatial capacity, leading to issues with spatial discontinuity and spatiotemporal incomparability. Furthermore, at present it is difficult to couple climate and weather data EO products with simulation-based BEMs. Compared to in-situ observations remote sensing measurements offer much higher spatial resolution, can act as validation in CFD models [10]. Although, a major disadvantage and uncertainty of this technique is its inability to penetrate clouds, therefore limiting the data availability to timestamps with clear sky conditions. Like other observational data types, EO data suffers from uncertainties from biases in sensors, sensor drift, and retrieval algorithms, calling for the accuracy of measurements to be addressed explicitly [80].

Table 3: Earth observational products for UBEM climate data inputs

Sensor	Available Products	Spatial Resolution	Temporal Resolution
Landsat 8 [84, p. 8]	OLI (9 bands), TIRS (2 bands), Analysis Ready Products: Surface Reflectance, Land Surface Temperature, Surface Water Extent, Fractional Snow Cover, Burned Area, Provisional Actual Evapotranspiration	30 m (visible, NIR, SWIR); 100 m (thermal), 15 m (panchromatic)	16 days
MODIS [85]	Surface Temperature, Cloud Temperature, Atmospheric Temperature, Cirrus Clouds Water Vapor, Cloud Properties, Ozone, Cloud Altitude	250 m (bands 1-2), 500 m (bands 3-7), 1000 m (bands 8-36)	16 days
ASTER [86]	Analysis Ready Products: Brightness Temperature, Emissivity, Surface Reflectance, Surface Kinetic Temperature, Surface Radiance, DEM [87]	15 m (VNIR), 30 m (SWIR), 90 m (TIR)	16 days
Sentinel-2 [88]	BOA reflectance, TOA radiance, Surface Reflectance	10 m (B2, B3, B4, B8), 20 m (B5, B6, B7, B8a, B11, B12) 60 m (B1, B9, B10)	5 days
Sentinel-5P NTRI [89]	Aerosol Index, Aerosol Height, Carbon Monoxide, Formaldehyde, Nitrogen Dioxide, Ozone, Sulfur Dioxide	1113.2 m	daily
ICESat-2 [90]	Land Ice Elevation, Artic Sea Ice Elevation, Land Water Vegetation Elevation, Cloud Characteristics [91]	20 m	91 days
ERA 5 [92]	Divergence, Cloud Cover, Geopotential, Ozone mass mixing ratio, Potential Vorticity, Relative Humidity, Specific Cloud Ice Water Content, Specific Cloud Liquid Water Content, Specific Rain Water Content, Specific Snow Water Content, Temperature, Wind Components, Vertical Velocity, Vorticity	0.25 °x 0.25 °	hourly
MERRA-2 [93]	Air Temperature, Wind Components, Sea Level Pressure, Surface Pressure, Total Precipitable Water Vapor	0.5 °x 0.625 °	hourly
NOAA RTMA [94]	Pressure, Temperature, Dew Point Temperature, Specific Humidity, Wind Speed, Wind Direction, Wind Speed (gust), Visibility, Cloud Cover, Precipitation	2.5 km	hourly
USFS Tree Canopy Cover v2021-4 [95]	Tree Canopy Cover	30 m	N/A
NASA DEM [96]	Digital Elevation	30 m	N/A
USGS National Landcover Database [97]	Landcover	30 m	N/A

4.1.2.1 Multispectral and Hyperspectral Imagery Data

Multispectral and hyperspectral imagery products acquire image layers at different wavelength bands from the same scene. Multispectral high-resolution visible sensors operate with three different primary color (ex. red, green, and blue) band pixel assignments [98] with sensing capabilities between the visible to middle infrared electromagnetic spectrum [99]. Multispectral imagery products such as Landsat 8 OLI [84, p. 8], Landsat 9 OLI-2 [84, p. 9], MODIS (Moderate Resolution Imaging Spectroradiometer) [85], ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) [86], Visible Infrared Imaging Radiometer Suite (VIIRS) [100] and Sentinel-2 [88] offer valuable insights into vegetation, soil, water burned area, and thermal land properties for microclimate UBEM parameter inputs, through products such as Normalized Difference Vegetation Index (NDVI), Land Surface Temperature (LST), ect.. The advantage of these products being their ability to contextualize areas at high spatial resolution (10 m to 250 m); although, with a tradeoff being that their temporal revisits are less frequent (5-16 days). Hyperspectral imagery technology extends beyond the multispectral range and gathers data from a wide spectrum of light [101]. Due to their small band widths and ability to distinguish separate entities in an image, multispectral and hyperspectral imagery can be used as land use and vegetation data classification parameter inputs [99]. For example, Dynamic World is a synthetic dataset that provides regularly updated classifications of global land use and landcover (LULC) and is trained off of and derived from 10 m Sentinel-2 data [102]. Parameters from multispectral and hyperspectral sensing technologies offer valuable insights into urban microclimates, as foliage and land characteristics play an important role in the thermal properties of urban areas [103].

4.1.2.2 LiDAR Data

LiDAR satellites use laser altimeter systems to measure physical distance through pulse travel time [104]. LiDAR measurements provide insights to the vertical profile to the atmosphere on the global scale, the vertical and horizonal distribution of clouds, landscape topography, and heights of ice sheets, land, forests, lakes and urban areas [104]. LiDAR measurements have been used in NWP models, offering perspectives to cloud temperature and formation processes, as well as in Digital Elevation Models (DEMs), providing quantification of surface heights. LiDAR missions tend to have shorter deployment times, longer revisit periods, and higher spatial resolution than imagery missions. In urban building energy and climate applications, LiDAR devices can provide insight to landscape topography, land elevation, building heights, and cloud formation. LiDAR data have been used to extract building geometric data such as building heights and size characteristics [105], and have informed training data that builds off of the Microsoft footprints dataset [106] to create a rasterized building footprint dataset [107], a building characteristic input to urban building energy models.

4.1.2.3 Thermal Imagery Data

Weather, climate, and microclimate parameters used within urban building energy studies are heavily dependent on thermal parameters of sites [108]. Thermal infrared (TIR) sensing provides insight to these characteristics by measuring the amount of radiation from an object, and in terms of satellite data collection, measuring the radiative heat fluxes (RHF) from the earth surfaces [109]. Thermal sensors operate in the emissions part of the earth spectrum, ranging from 8- 14 μm [108]. TIR output data is used for monitoring agricultural drought, land cover, land surface albedo, Normalized Difference Vegetation Index (NDVI), thermal environment, thermal anomaly, and urban surface atmosphere exchange [79], [110]; with the most prominently derived parameter for being land surface temperature (LST); a measurement that is widely used in UHI studies due to its tendency to identifying spatial anomalies and variation in surface energy balances [110]. LST effectively measures surface skin temperature, or the ground radiometric temperature as observed from above, quantifying the energy emitted and reflected from a surface [111]. There are a variety of products that take TIR measurements at different spatial and temporal scales.

High temporal granularity is important for LST measurements to accurately portray UHI effects and have been used to address stability in climate conditions [112]. Although, it should be noted that sensor accuracy is not always consistent, as the MODIS LST product was shown to be accurate at night compared to the daytime [113], [114]. Further, a vertical spatial profile for LST measurements exists and should be considered alongside horizontal variability [115] and may affect building heating and cooling loads, specifically tall rise structures. Also, in dense urban areas, LST measurements may inaccurately represent the surface temperature of rooftops rather than inside of the street canyon [111], emphasizing the differences between land surface temperature measurements and air temperature measurements. Although LST acts as a proxy for air temperature, the observations have different meanings and responses to atmospheric conditions [116]. For use in robust applications, numerous methods have been carried out to derive air temperature from remote sensing observations, such as one that uses statical methods with satellite and weather station data to produce a monthly timeseries air temperature values for 2003 to 2016 [117], and meteorological reanalysis products that provide air temperature measurements resulting from both simulation based and observational based domains.

4.1.2.4 Reanalysis Products

Reanalysis products are derived from global weather forecasting models under observationally constrained scenarios using data assimilation techniques [118]. Because reanalysis models are constrained by weather observational data and combined with physical models, these products embody both observational and simulation-based techniques [117]; offering a product that is most suitable for spatiotemporally consistent environmental analysis [119]. For instance the ERA5 reanalysis product provides hourly estimates for a large number of atmospheric, land, oceanic, and climate variables on a 30 km or 0.1°grid [92]. It combines a large number of observations from in-situ and EO data with an integrated forecast system cycle (Cy41r2) and 4D-Var cost function techniques [119]. MERRA-2 is another popular reanalysis data product for the period 1980 to present with approximate resolution of 0.5° x 0.625° at 1 hour temporal resolution [93]. The NOAA Real Time Mesoscale Analysis (RTMA) product is preferred for urban energy studies due to its spatial granularity [83], and is available at a one hourly frequency for the continental US at 2.5 km spatial resolution [94].

4.2 Simulation-Based Microclimate Datasets and Tools

Simulation-based microclimate modeling methods, conditioned on physical and morphological parameters, are used to explore the effect of urban contexts on building energy performance [120]. In general, urban climate simulation techniques serve to model a surrounding environment, with reference to observational weather station data [57]. These techniques are used for a wide variety of spatial scales, ranging from indoor conditions (less than 10 m) to mesoscale (greater than 200km) [56].

Simulation-based microclimate tools have been implemented to simulate conditions from the building level to the city scale and examine the effect of localized climates on urban energy consumption through microscale Computational Fluid Dynamics (CFD), mesoscale CFD, Numerical Weather Prediction (NWP), and Urban Canopy Model (UCM) tools and models, as elaborated on in *Table 4*. These techniques have been criticized in UBEM applications due to their high computational costs and their resource intensity to scale beyond small scale flow fields (buildings, small campuses) and small temporal (hourly, daily) resolutions [25], [56], [120]. Although, their comparative analysis flexibility and flow parameter iterabilities over entire domains spaces make simulation- based methods an attractive microclimate modeling method. This is especially true when researchers have extensive knowledge and experience in CFD model configuration settings, input parameters, and simulation methods [24], [56]. Model coupling has been adopted to rapidly and accurately combine simulation based microclimate datasets into UBEMs to obtain more realistic results [57]. This approach helps to combat simulation program computational demands through load sharing, although complications may arise while merging program specific time-scales [57] and during generation of comparable formats [111]. As for data accuracy, applying CFD into

microclimate studies may produce conditions for specific locations that are far from reality [57]. Moreover, because this data is fundamentally modeled, simulation-based microclimate datasets may deviate from actual observed measurements and conditions, further propagating errors within building energy model demand estimates. For this reason, it is necessary to ground-truth and validate simulation-based microclimate datasets with real measurements to ensure accurate modeling results.

Table 4: Popular simulation-based microclimate dataset modeling tools for building energy modeling studies

Platform	Microclimate Dataset Modeling Tool Category			Computational Demands	Scalability	Coupling Feasibility	Small Scale Turbulence
	Computational Fluid Dynamics	Numerical Weather Prediction	Urban Canopy Model				
ENVI-met [121]	✓			++	–	–	++
CityFFD [122]	✓			+	+	++	+
Meso-NH [123]		✓		+++	++	++	–
Weather Research and Forecasting Model (WRF) [124]		✓		+++	++	++	–
Town Energy Balance (TEB) [125]			✓	–	+	+	–
Urban Weather Generator (UWG) [126]			✓	–	++	+	–
Solene Microclimat [127]			✓	+	–	+	+

4.2.1 Computational Fluid Dynamics Microclimate Dataset Modeling Tools

The most common simulation-based urban climate modeling techniques for UBEM applications utilize foundational Computational Fluid Dynamics (CFD) principles. CFD models leverage physics-based fluid motion conservation laws (conservation of mass, momentum, and energy) to produce quantitative predictions about fluid flow phenomena [128]. CFD simulates thermal and mass interactions over contextualized obstacles, such as building terrain in urban studies [24] by solving either the Reynolds Averaged Navier Stokes (RANS) or Large Eddy Simulations (LES) fundamental governing equations [56]. The most frequently used micro-scale CFD tool for building energy modeling is ENVI-met, which solves for atmospheric flow and heat transfer with RANS equations [26], [58], [129], followed by CityFFD, a Fast Fluid Dynamics solver with computing algorithms based on the semi-Lagrangian approach. Micro-scale CFD simulations are computationally expensive and often limited to small spatial scale flow fields (buildings and small campuses) and small temporal (hourly, daily) resolution [25], [56], [120]. Within these platforms, temperature, wind velocity, and surface temperature are the most commonly studied and prominent parameters in urban microclimate research [56].

Within UBEM studies, CFD microclimate tools have limitations in their ability to communicate with and exchange information between building energy management (BEM) tools. To combat these challenges, Elnabawi et al. used ENVI-met v4 software to model the most common microclimate parameters over a 6-day summer period for a university building in Bahrain and combined this output with urban weather generator, Meteronorm, to produce a synthetic urban specific weather dataset (USWD). This dataset consisted of a typical year of timeseries data in an hourly format compatible for building energy modeling tool IES-VE [65]. In contrast, Alyakoob et. al. went way with merging microclimate and Building Energy Management (BEM) platforms altogether, by using a machine learning tree-based algorithm approach that examined the impact of ENVI-met produced microclimate conditions at three Arizona State University

(ASU) buildings over three representee summer days [130]. Although, these methodologies were able to combat limitations with platform communication, both augmented UBEM and ENVI-met studies cited programs computational cost as a challenge to expanding their case study region to larger temporal and spatial resolutions.

Coupling CFD with BEM tools is a commonly employed method for considering urban microclimate conditions in building energy research [25], [120]. Coupling, in this context, refers to the combination and exchange of information between independent simulation platforms, a building energy model, and an urban microclimate model, enabling communication and data transfer between them. In downtown Montreal, Katal et al. combine open source building geometric data, building property data, and measured weather station data (air temperature, solar radiation, and wind speed and direction) into CityFFD and CityBEM, using the ping-pong coupling strategy [131] to examine microclimate conditions over a case study region of 225 buildings for a 15 day summer period [132]. Whereas a tool-agnostic semantic schema (in JSON) coupling strategy was used to map building surfaces with CFD grids and transfer data between CityFFD and CityBES pairing nodes, ensuring no double counted or miscounted heat fluxes, over a 97 buildings case study region in northeast San Francisco during a 48-hour summer heat wave [76]. A wide variety of BEM and CFD coupling strategies in UBEM frameworks exist, as well as tool specific optimization strategies, which are well tabulated in this coupling review article [57].

4.2.2 Numerical Weather Prediction Microclimate Dataset Modeling Tools

Numerical Weather Prediction (NWP) tools simulate CFD principles in the atmosphere, using prediction methods to provide insight about atmospheric processes in larger spatial mesoscales [133]. Meso-NH is a noteworthy NWP, or non-hydrostatic mesoscale atmospheric model that considers earths sphericity and is designed to simulate atmospheric processes (ex. motion, moisture, and thermodynamics) at the regional scale [134], [135]. Whereas the World Research and Forecasting (WRF) model is one of the most popular NWP tools, which consists of both a data assimilation system and computational architecture with three coupled Urban Canopy Models [124]; and is used to provide insight for actual atmospheric conditions or idealized atmospheric conditions across a wide range of spatial scales, spanning from tens of meters to thousands of kilometers [124]. An advantage of WRF is its ability to initialize simulations with both statically modeled meteorological conditions and actual measurements, creating many diverse use case applications [136]. Jain et al. used the WRF model to simulate hourly climate conditions over a two-day summer period for 20 buildings in Goose Island Region, Chicago, IL, initialized by the NOAA real-time High-Resolution Rapid Refresh (HRRR) [137] dataset [136]; this output was then coupled through a central data translation engine by data-exchange units with energy plus to examine the effects of microclimate conditions on building cooling loads. The authors acknowledge challenges in their methodology and recommend the use of a finer-scale CFD model to decrease these uncertainties stemming from turbulence around buildings [136].

4.2.3 Urban Canopy Model Microclimate Dataset Modeling Tools

UCMs represent the exchange of momentum, heat, and moister between urban contexts and the atmosphere [138] through performing an energy balance over a given surface. Town Energy Balance (TEB) considers radiation processes and vertical wind profiles of urban geometries in a single layer UCM [58], [139], while Urban Weather Generator (URG) [126] uses four submodules, the rural station model, the vertical diffusion model, the urban boundary-layer (UBL) model and the UCM to provide canopy level urban temperature, facilitating this information exchange in the easy to use ladybug [140] plug in. Further, radiative model Solene Microclimat uses radiosity algorithms to simulate whole solar fluxes reaching urban surfaces, Program output can be coupled with CFD models to holistically describe urban atmospheric conditions [25], [58], [127]. Additionally, NWP models can be combined with Urban Canopy Models (UCMs) to

achieve finer-grain thermal insight at both the city and building resolution. To provide context of the interactions between residential housing energy consumption and urban climate in the Sham Shui Po region of Hong Kong, mesoscale atmospheric model MesoNH and UCM TEB were coupled, creating a multi-layer model that examined energy performance over a 12-day period. The methodology excelled in representing interactions between high density urban areas in the atmosphere at relatively low computational cost over a longer time duration [120].

5. Comparative Analysis of Case Study Datasets, Methodologies, and Results

Ten studies that examine effect of more accurate climate and weather data on UBEM performance were analyzed in this research, where common case study metrics are tabulated in *Table 5*. The studies included in this section use observational-based or simulation-based climate and weather data inputs and analyze their performances in UBEMs. The studies were chosen based off criteria outlined in Section 2 of this paper.

Table 5: Case studies exploring microclimate conditions in UBEMs

	Author (year)	Location	UBEM Methodology Type			Coupling method	Climate Datasets and BEM Tools	Climate Parameters	Detail
			Data-Driven	Simulation-Based	Hybrid				
1	M. H. Elnabawi et al. (2022) [65]	Manama, Bahrain		✓		–	ENVI-met v.4, Meteonorm, IES-VE	AT, MRT, RH	Urban Specific Weather Dataset, University Building, Summer
2	A. Alyakoob et al. (2022) [130]	Tempe, USA			✓	–	ENVI-met	AT, AH, SWR, Shading levels	Tree Based Methods, Machine Learning, University Buildings, Regression, Cooling Loads, Summer
3	A. Katal et al. (2022) [132]	Montreal, Canada		✓		Ping-pong coupling	CityFFD, CityBEM	AT, SR, WS, WD	3D models of buildings, Open datasets, Automated processes, Summer
4	N. Luo et al. (2022) [76]	San Francisco, USA		✓		Tool-agnostic semantic schema (in JSON)	CityFFD, CityBEM	AT, RH, WS, WD, SR, Pressure, Carbon Dioxide	Attached Surfaces, 5 JSON files, Data Pairing, Diverse Buildings, Summer Heat Wave
5	R. Jain et al. (2020) [136]	Chicago, USA		✓		One directional coupling	WFR, HRRR dataset, EnergyPlus	AT, RH, WS, WD, SR, SWR, LWR	Air Nodes, Cooling Loads, Diverse Buildings, Summer
6	S. Liu et al. (2023) [120]	Sham Shui Po, Hong Kong		✓		Multi-layer coupling	Meso-NH, TEB, EnergyPlus	AT, RH, WS, SR	Building Archetypes, Summer Heat Wave, Nighttime differences
7	M. Zinzi et al. (2018) [77]	Rome, Italy		✓		–	5 in-situ measurements, TRNSYS	AT, RH	Building Archetypes, Long Term Study, Seasonality
8	T. Hong et al. (2021) [10]	San Francisco, USA		✓		–	27 in-situ measurements, CityBES, EnergyPlus	AT, RH, SR, WS	DOE Building Archetypes, Summer Heat Wave

9	T. R. Dougherty and R. K. Jain (2022) [83]	New York City, USA		✓	–	Sentinel-2 Level-1C, VIIRS, and NASA's SRTM, NOAA RTMA	SR, CDD, HDD, WS, Nighttime Light Radiance, Vegetation Index, Cloud Cover, Precipitation, Elevation	Remote Sensing, Regression, Linearity, Gas Consumption, Electricity Consumption, Diverse Buildings
10	T. R. Dougherty and R. K. Jain (2023) [82]	New York City, USA		✓	–	Sentinel- 1, Sentinel-2, Landsat 8, CMIP, Dynamic World, SAR, VIIRS, Elevation	AT, ST, WS, WD, RH, SR, SWR, Pressure, Vegetation Index, Cloud Cover, Precipitation, Elevation	Remote Sensing, Tree Based Methods, Machine Learning, Gas Consumption, Electricity Consumption, Nonlinearity, Seasonality, Diverse Buildings

AT: Air Temperature, ST: Surface Temperature, GT: Global Temperature, MRT: Mean Radiant Temperature, CDD: Cooling Degree Days, HDD: Heating Degree Days, RH: Relative Humidity, AH: Absolute Humidity, WS: Wind Speed, WD: Wind Direction, SR: Solar Radiation, SWR: Short Wave Radiation, LWR: Long Wave Radiation

It is evident that there are large disparities in the proposed microclimate datasets abilities to describe spatial and temporal states. The relationship between these resolutions is shown as a scatterplot in *Figure 4*. In essence, simulation-based programs provide fluid flow and thermal measurements at high granularity for smaller spaces (ex. room or building), although these methods are limited by their computational constraints to describe multitudes of buildings in neighborhoods and cities. Observational-based datasets are not limited by these modeling constraints and can span longer temporal study periods, although must be recognized for their limitations in precision for this complex task, specifically because revisit periods (or temporal difference between measurements) may be much less frequent than simulation-based datasets. In other words, simulation-based microclimate datasets are more insightful for instances that examine the effect of microclimates on a small subset of buildings in an urban context or during a specific extreme weather event; whereas observational-based data is best used to study the effects of microclimates on resource consumption over a larger neighborhood or city scale across a longer study period.

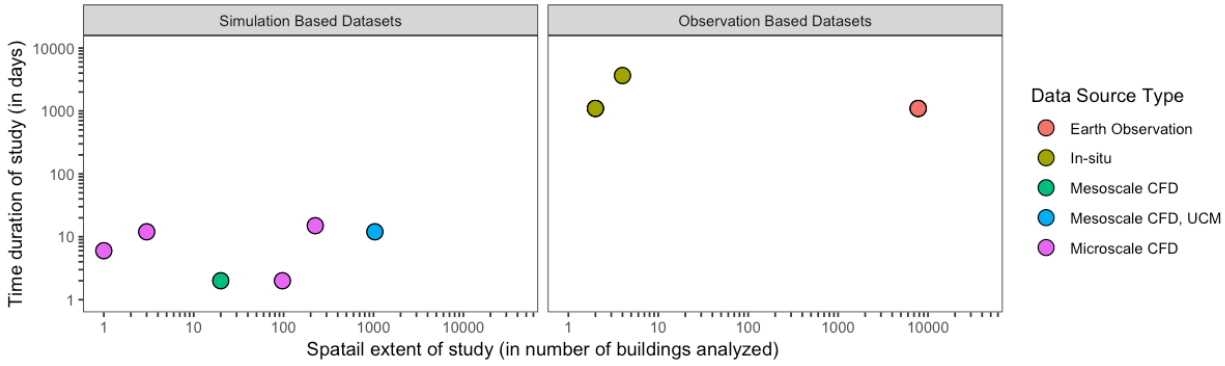

Figure 4: Spatial and temporal scales of the UBEM case studies

From this synthesis of the ten case studies, it is clear that climate dataset selection has a substantial effect on reported resource (energy, cooling, and heating) consumption of buildings. The results of each case study are shown in *Figure 5* and outlined in *Table 5*, along with case study locations, methodologies, coupling methods, climate datasets, climate parameters, and generic keyword details. The most common result reporting method entails the comparison of energy consumption values given both an urban specific microclimate weather dataset and the traditional TMY climate dataset. For the case studies analyzed in this manner, relative changes between microclimate and TMY datasets resulted in 5% and 23% change in

energy consumption, 4.7% to 74% increase in cooling consumption, and a 15% to 20% decrease in heating consumption. That is, for the climate zones examined, cooling loads are more unlikely to be underestimated due to the exclusion of urban microclimate data, than heating loads. Moreover, the inclusion of urban microclimate data substantially impacts UBEM results- thus actions should be taken to include these conditions in UBEMs to decrease the misrepresentations of energy consumption patterns and to ensure accurate and realistic modeling. Due to the relatively small sample size and the differing spatial-temporal scopes of UBEM case studies examined, there is a limitation for creating more transferable concrete conclusions. That is, it should be noted these aggregated summary statistics are inherently case study specific and may not be interpreted to extend beyond the intended climate zone and building types. Additionally, other obvious and unpredictable factors that are not included as UBEM inputs may influence building energy consumption, thus all responsibility of these additional loadings may not soley be attributable to urban microclimate conditions, among these factors include indoor building set points, occupant comfort controls, and variability with resource consumption.

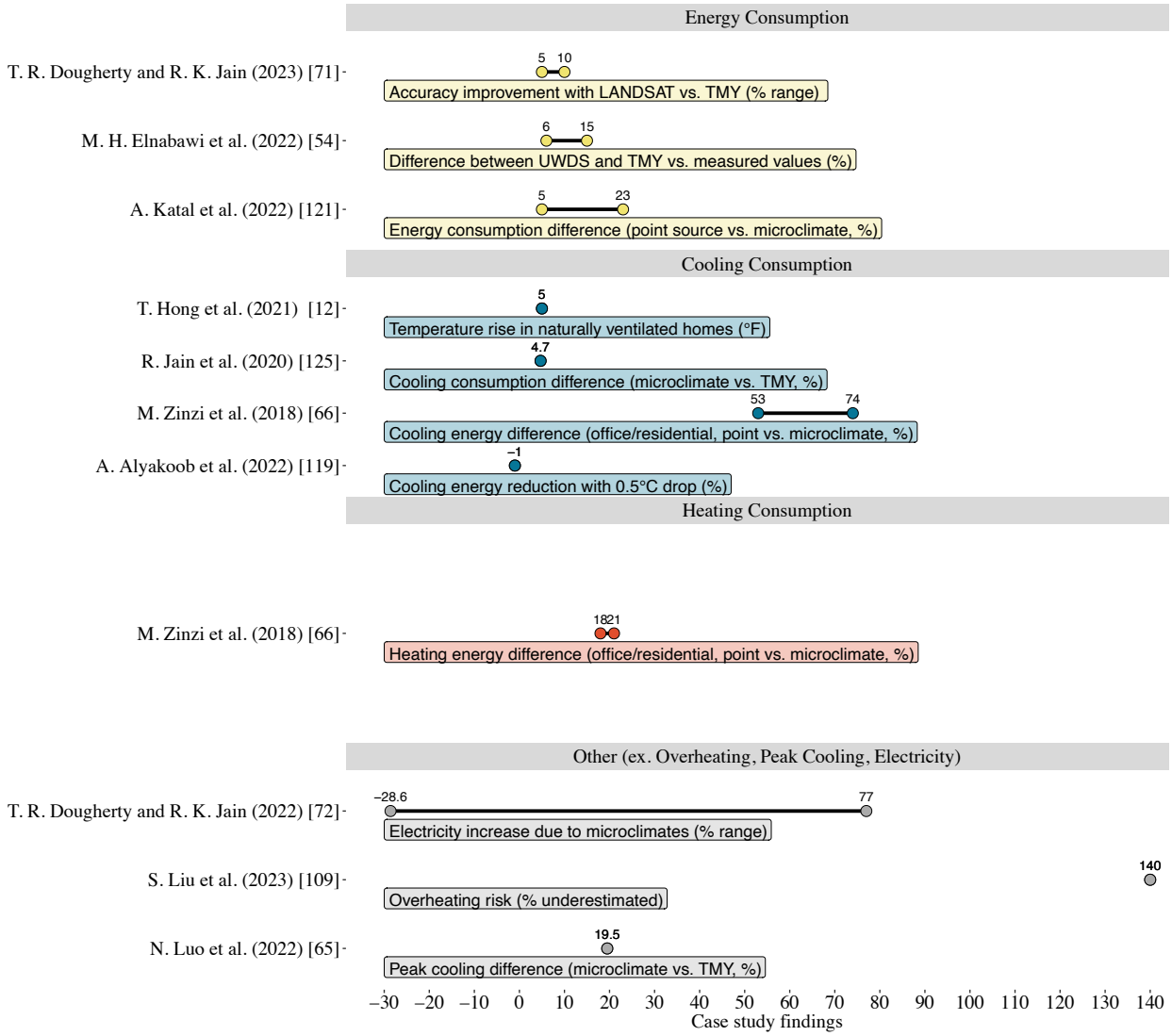

Figure 5: UBEM and microclimate case study findings, highlighting variability in reporting methods and their impacts on energy and temperature metrics.

As illustrated in *Figure 5*, a wide range of result reporting procedures exist for UBEM and microclimate studies, complicating their comparability across frameworks and methodologies. That is, it is difficult to

determine which urban microclimate datasets (and the simulation-based tools to create them) are most accurate, as they are not always validated with real building performance data or real urban microclimate data, and their benchmarking processes often differ in temporal and spatial scope. Therefore, it is recommended for researchers to validate their microclimate model results with real microclimate and energy data to facilitate comparison between models and to examine the reliability of their methodologies. This will not only help with reporting but will also increase model reliability through ensuring ground truthing of the proposed methodologies -- as without this step it is difficult to make applicable, and concrete actions out of study conclusions.

It is well documented that UHIs pose the largest danger to human health during extreme heating periods. That is, for the studies examined, urban building energy was modeled over the summer months. Notably, most of the studies only examined microclimate conditions over the summer, 30% explicitly examined urban microclimate conditions during a summer extreme weather heating event, and 30% conducted research over the entire annual timeframe but identified larger microclimate (and Urban Heat Island) effects throughout the summer months. Therefore, more research should be taken to analyze these conditions in more extreme climate zones and where building cooling demands are increasing throughout summer months. Furthermore, air temperature was the primary microclimate data parameter examined, with traditional variable expansion including humidity, wind speed, solar radiation, precipitation, elevation, pressure, vegetation, cloud cover, and shading.

6. Discussion

Dense urban areas have altered historical land use patterns, habitats, and intrinsically effected environmental conditions; thus, research is needed to account for these changes and their impacts on city infrastructure. Integrating urban building energy models with microclimate weather data that is spatially characterized for urban contexts can aid in accounting for these altercations to further increase urban building infrastructure resiliency. Traditionally, UBEM methodologies rely on historically aggregated TMY weather files for their climate data input sources. However, these files are based on aggregated observations collected as single point measurements which are often located in rural areas; thus, failing to characterize weather and climate in urban regions at high spatial and temporal granularity.

This literature review explores the different climate datasets, tools, and methodologies used to integrate finer grain microclimates in UBEM frameworks. The review characterizes these climate datasets into two types: observational-based and simulation-based. Observational-based datasets contextualize spatial granularity by compiling real measured values at different spatial and temporal resolution within a region of interest. Whereas simulation-based datasets use computational methods to model spatial variability in climate given a reference weather file, through physics-based conservation laws (CFD), atmospheric processes (NWP), and energy balances (UCMs). Combining, or coupling, high-resolution microclimate data into UBEMs remains one of the significant challenges in this research domain. Data-driven UBEMs can facilitate the integration of observational-based and simulation-based microclimate data with building data through merging timestamps and locations. However, this comes with its own set of assumptions and challenges, including limitations with high temporal data availability and quality. Building energy simulation platforms such as TRANYSYS, IES-VE, CityBEM, and EnergyPlus can facilitate the data transfer process for simulation-based microclimate datasets though the utilization of various coupling procedures (such as one-directional and two-directional). Although, complications such as data loss, misalignment of timescales, high computational demands, along with the limitations and assumptions of each underlying software platform arise under these scenarios. Furthermore, because simulation-based datasets are fundamentally modeled- a ground truthing process which ensures that datasets are calibrated to portray actual conditions is necessary to reduce compounding errors in urban-scale energy predictions.

To reduce the uncertainties arising from microclimate modeling, energy modeling, and coupling platforms thereof, a blending of the two methods can be taken. This hybrid, or augmented approach can combine instances of both simulation-based and data-driven techniques and leverage the benefits of both methodologies. For example, Alyakoob et. al. use a data-driven machine learning tree-based algorithm energy model to examine the impact of simulation-based ENVI-met produced microclimate conditions for Arizona State University (ASU) buildings [130]. Through using the ENVI-met microclimate data and a data-driven UBEM model, the authors were able to reduce the computing requirements of building energy simulations, increase the feasibility of coupling the two datasets, and leverage ENVI-met for more granule climate data. As for existing microclimate datasets, reanalysis products embody both simulation-based and observational-based domains, are validated and cleaned by large-scale institutions; and may offer as an alternative to single simulations to contextualize microclimate contexts at the larger city and state level. Due to the simulation-based generative nature and their ability to contextualize given starting conditions, reanalysis products are often reported with higher temporal resolution than conventional earth observational datasets. Dougherty and Jain take a semi-hybrid approach while blending reanalysis products with EO data in their data-driven studies [83], [82], leveraging their ability to contextualize conditions at high temporal resolution. It is recommended that hybrid methodologies are further investigated for integration of microclimates into UBEM models. For instance, there is an opportunity for observational-based microclimate data to be combined with and used as input into coupled CFD and BEM tools in an augmented UBEM methodology [82]. Additionally, methodological opportunities exist for blending earth observational datasets into simulation-based building energy platforms to leverage both the spatial granularity of earth observational datasets and the building granularity of a simulation-based BEM.

Finally, for both simulation-based and observational-based climate data inputs, it is necessary for researchers to recognize and consider the specific limitations of their models during both analysis and result interpretation stages [67]. Additionally, for all contexts, it is important to use local weather data in energy models [26] either within the model validation stage or through use of observational products. Including these higher spatial resolution climate data products that characterize urban climates using a diverse set of environmental features will decrease current UBEM uncertainties. This will further raise model accuracy and produce actionable steps for increasing infrastructure resiliency, ultimately promoting the protection of human health.

7. Conclusion

Currently, without reliable energy modeling predictions, engineers and stakeholders struggle to accurately characterize current and future building energy consumption, leading to discrepancies in environmental, social, and economic performance for their structures. Due to inter-building connections and microclimate conditions, such as the urban heat island effect, these inconsistencies are exacerbated in urban climates. This literature review details the datasets, tools, and methodologies used to incorporate microclimates in UBEMs, focusing to close the simulation-based modeling gap to address environmental modeling uncertainties. In this process, it expands on the traditionally used methods and elaborates on substitutional observational-based and simulation-based data types, detailing the spatial and temporal tradeoffs of each source. It highlights the difficulties of combining both highly temporal and spatially granule microclimate data into building energy platforms, sighting challenges in data merging, scalability, and computation. To reduce the uncertainties associated with each simulation or observational data type, to add more microclimate parameters into models, and to maintain low computational requirements with larger temporal and spatial simulations the review recommends investigating multimethod, or hybrid approaches. Additionally, the review finds that remote sensing technologies have been well explored for building geometric data inputs. Although, there is need to leverage the spatial resolution of these datatypes and explore both reanalysis and observational-based products for environmental data inputs. Further, case study research in this sector primarily uses the air temperature parameter for assessment of urban microclimate

conditions, specifically examining these effects during summer months. Future investigation should be taken to explore the impacts of additional environmental parameters across all climate zones and throughout extreme heating and cooling weather events. Finally, an aggregation of case study findings concludes substantial differences between modeled results with microclimate datasets and with conventional climate dataset inputs, underscoring the importance of including microclimate data into model validation and calibration workflows. There poses a significant challenge to address both building and climate granularity in urban building energy studies. Further research in this domain is essential to bridge these existing gaps between simulated models and real-world scenarios. These advancements will offer more precise recommendations for energy system improvements and guidance for energy resiliency planning, which are specifically critical to combat future infrastructure stresses such as extreme weather events and changing climate conditions on energy systems at the city level.

Chapter 3: Leveraging Earth Observational Data Products and Machine Learning to Enhance Urban Building Energy Modeling (UBEM) with Microclimate Effects

Authors**:** Amanda Worthy[1]*, Mehdi Ashayeri [2], Julian Marshall [1], Narjes Abbasabadi [3]
[1] Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, United States
[2] School of Architecture, Southern Illinois University, Carbondale, IL, United States
[3] Department of Architecture, University of Washington, Seattle, WA, United States

[*]**Corresponding Author:**

Amanda Worthy

PhD Student
Department of Civil and Environmental Engineering
University of Washington
aworthy@uw.edu

Abstract

Urban Building Energy Modeling (UBEM) is a powerful tool used for sustainable design, urban planning, and efficient energy management, as it provides essential insights into the building energy consumption patterns. However, current UBEM methodologies often lack urban-specific microclimate data, leading to discrepancies between modeled and actual energy consumption. This research develops a bottom-up statistical UBEM framework that combines and integrates earth observational climate data, climate reanalysis products, and annual energy usage data, measured by the Seattle Energy Benchmarking Dataset, to capture the impacts of microclimates on urban building energy performance. Using machine learning techniques and Seattle, Washington, USA as a proof of concept, our results demonstrate that incorporating urban-specific microclimate data substantially enhances building energy modeling prediction accuracy. Specifically, three model variable schemas are compared; the optimal model incorporating earth observational data achieved a 0.16 (from 0.55 to 0.71) increase in testing R^2 over the model that did not include any climate data inputs, and a 0.056 (from 0.66 to 0.71) increase in testing R^2, over the model that included TMY3 climate data inputs. These findings validate the use of earth observational datasets for urban building energy modeling to include microclimate effects. Furthermore, machine learning algorithms outperform traditional linear methods, with respective ordered rankings: CATBoost, XGBoost, Random Forest, Decision Trees, and Linear Regression. Our study underscores the importance of integrating microclimate data into UBEM frameworks and advocates for the expanded use of earth observational and remote sensing datasets for mitigation of simulation-to-reality discrepancies; to ultimately inform more accurate energy-driven design and planning strategies at the city level.

Keywords: Urban Building Energy Modeling, Microclimates, Earth Observational Data, Machine Learning

Highlights:

- Demonstrates the reliability of earth observational data for studying urban building energy use.
- Highlights the dependability of machine learning methods for examining building energy consumption
- Tabulates case study specific microclimate parameters influencing building energy performance

1.Introduction

Governments are establishing ambitious sustainability objectives and resiliency adaption plans to make their infrastructure more robust in future and changing climates. These goals are targeted towards every sector, although building infrastructure has become a major focus, given that buildings are responsible for 26% of global energy related emissions [141]. For instance, the United States has pledged to achieve net-zero buildings by 2045 [142], requiring urgent and precise examination of their infrastructure to meet these objectives. Furthermore, 80% of a buildings lifecycle energy consumption can be attributed to operational energy [143], or energy expended while a building is in use; calling for thorough examination of energy system performance throughout this timeframe.

As urban areas tend to grow [35], building systems are becoming more centralized and interconnected [35], through increased cross-building connections [44] and shared urban infrastructure [144]. This rise of network density has also transformed surface energy balances, albedo [11], native foliage concentrations [6], [7] and land use patterns in cities [4]; consequently contributing to elevated temperatures within urban areas [145], [15], or the Urban Heat Island (UHI) effect. Furthermore, urban contexts are have contributed to microclimate conditions in cities [108], [45], characterized by locations where the local climates differ significantly from surrounding environments [36]. Understanding these microclimate conditions is essential for interpreting urban energy contexts, thereby supporting the development of sustainable buildings and resilient cities.

Given a large quantity of infrastructure and great correspondences between structures, sustainable infrastructure planning cannot be achieved through solely focusing on single building stock entities, instead, it requires an urban scale approach that engages various infrastructure and stakeholder groups. Urban Building Energy Modeling (UBEM) [23], [25], [24] examines clusters of buildings within urban environments, attempting to recognize the interactions between them and their urban contexts to fully address energy dynamics in cities. UBEM methods differ in spatial scale and scope from Building Performance Simulation (BPS), which models the energy consumption behavior of a single building [146]. UBEM methodologies include top-down approaches, which examine cities at the macro-scale, and bottom-up approaches, which consider urban attributes through the microscale of individual units (e.g., single buildings), with the latter classified into three categories: statistical (including machine-learning and artificial intelligence) models, engineering (physics-based) models, and finally hybrid models which combine the two [25] ,[26], [147]. Statistical models use data techniques such as artificial intelligence (AI), machine learning (ML) and statistical methods to provide insight into real world system operation given dynamic data inputs [27]. Whereas physics-based models use numerical-based conventions to represent energy consumption through the physical and technical characteristics of buildings [26].

UBEMs provide consumption patterns for a multitude of buildings to gather insights for centralized renovation patterns and to identify resource use hot spots. The models rely substantially on input components to describe building characteristics and functions [27]; so much so that their accuracy depends on and is limited by the availability of data surrounding input components and their respective abilities to precisely contextualize both spatial and temporal contexts [26]. In terms of environmental data inputs, conventional UBEM methodologies utilize historically aggregated and non-urban specific climate and weather data measurements for their weather parameters [24], which fail to account for extreme weather events, changing climate conditions, and urban microclimate conditions; exacerbating uncertainties within modeling results [16]. Conventional UBEM weather data inputs are in the Typical Meteorological Year (TMY) Energy Plus Weather (.epw) file format. These include hourly weather observations over a 12 meteorological month calendar; which are derived from measurements collected during a specific multiple-year time scale (usually 10 years) and selected based on criteria for being “representative” of a typical weather month [65]. TMY files are not always available at high spatial granularity, and thus discount in characterizing climates at smaller scales. For instance, within this studies context, TMY3 datasets are made

available from the EnergyPlus official website [148] for two stations within the City of Seattle limits, and seven stations with 25 miles of Seattle city limits.

Various studies have been conducted to integrate more spatially precise weather and microclimate data into UBEMs. In these studies, urban microclimate data is commonly sourced from two approaches, simulation based and observational-based approaches[24], with each offering unique tradeoffs in their abilities to contextualize different spatial and temporal conditions [16]. Simulation-based approaches use numerical-based conventions to represent local climate data elements [37], although they are limited by their computational demands [26] and often require elaborate coupling schemas to be meaningfully integrated into energy models [57]. Observational-based approaches utilize measured values for environmental data inputs, such as in-situ and remote sensing observations; but are often limited by the larges resource requirements to sample at high spatial-temporal frequencies, and though issues that arise during the data collection processes; such as sensor drift, sensor communication, and sensor operation [26].

Coupling high spatial and high temporal resolution urban microclimate data into UBEMs remains a significant challenge [16]. However, advancements in machine learning and artificial intelligence applications have assisted with combatting these limitations. Statistical models reduce the computational demands for scaling beyond small scale flow fields (buildings, small campuses) and small temporal (hourly, daily) resolution [25], [56], [120], offering an attractive method for examining building energy usage at the urban scale compared to simulation-based platforms. A review on statistical building energy studies identify support vector machines (SVM), artificial neural networks (ANN), decision tress (DTs), multiple linear regression (MLR), ordinary least squares (OLS), and autoregressive integrative moving average (ARIMA) to be the most commonly used algorithms for energy prediction tasks [60]. Whereas simple linear regression (SLR), multi linear regression (MLR), polynomial regression (PR), conditional demand analysis (CDR), Huber M-estimation regression (HMR), ANNs, SVMs, and k-nearest neighbor (kNN) techniques are regularly applied in urban scale building energy use modeling applications [26]. Further, ensemble methods such as random forest (RF), random decision forest (RDF), classification and regression tree (CART), and gradient boosting techniques have been deemed successful in exploring data with noise, heterogeneous features, and correlation, given relatively small datasets [149].

Various statistical building performance studies employ ensemble methods to explore energy consumption at the urban scale. These powerful machine learning algorithms leverage the gradual and additive coupling of several weak learning models to outperform predictions made from a single model [149]. Alyakoob et. al. confirm the validity of ensemble methods in their augmented statistical and simulation based UBEM through using ENVI-met produced microclimate conditions (including data variables such as air temperature, relative humidity, shading levels, and direct shortwave radiation) and tree-based/ boosting algorithms to examine the impact of microclimates on cooling energy demands on Arizona State University (ASU) buildings; predicting a 0.98 accuracy score with their random forest algorithm [130]. In Seattle, categorical boosting (CATBoost) was used to to predict the Weather Normalized Site Energy Use Intensity given energy benchmarking data with a model accuracy R^2 of 0.897 and an outlier detection accuracy of 99.32%. The aforementioned study relied solely on district location and urban density (in number of buildings) as a proxy for environmental conditions [149] and highlights the potential to increase model accuracy given a more diverse array of environmental and climate data in performance-based energy studies. Ahn and Sohn explore the interconnectivities between urban form (horizontal compactness, vertical density, and variation of building heights) on multi-family home residential energy consumption in Seattle. The authors confirm building energy consumption to be spatially dependent and call for further empirical analysis to examine for which and to what do extent urban environmental factors affect building energy performance [150].

Aiming to address urban microclimate conditions in energy models, two statistical studies explore the differences in building energy performance given earth observational (EO) microclimate data and traditional weather files. Dougherty and Jain (2022) [83] employ linear regression model to explore the relative impact of environmental earth observational parameters on electricity and natural gas consumption [83]; finding urban microclimates to decrease a buildings natural gas consumption by as much as 71% and to increase others as much as 221% in other regions. Dougherty and Jain (2023) [82] also explore nonlinear models (random forest and particle swarm optimization algorithms) and a larger breath of remote sensing environmental products, such as the Landsat 8 TIR imagery and the Dynamic World Landuse dataset; and find an 8% reduction of electricity prediction modeled error in the spring and summertime given LANDSAT 8 environmental feature inputs, specifically highlighting the SR 7, Downwelled Radiance, and Upwelled Radiance band contributions [82]. The authors emphasize the importance of including reanalysis products (NOAA RTMA), satellite observations (Sentinel-2 Level-1C products), and LiDAR data (NASA SRTM) into building energy performance models and encourage future investigations to test the robustness of these data types in other cities.

The present study uses a statistical UBEM, as defined in [25] ,[26], [147], [151], to examine the relationships between microclimate and building energy demand in Seattle, Washington, USA. In particular, we aim to address the following questions:

- Can earth observational data enhance Urban Building Energy Models to better assess the impact of urban microclimates on building energy demand and building performance?
- How do Typical Meteorological Year 3 (TMY3) environmental observations compare to earth observational measurements? Are there modeling benefits for including spatially interpolated TMY3 data and earth observational urban microclimate data products into UBEMs?
- Which building geometric, building non-geometric, and environmental parameters are most critical for predicting annual building electricity consumption?
- How can earth observational products be optimized to better support urban energy performance studies?

The methodology at present leverages earth observational environmental data inputs and combines these with city building energy benchmarking data using machine learning algorithms to examine the interactions between urban climate and energy systems; and then compares with conventional methods. Section 2 of the manuscript introduces the materials used within this study, elaborating on the open-source datasets that were compiled for analysis. Section 3 discusses the case study methodology, highlighting specific data preparation and modeling techniques. Section 4 reports project modeling capacities and performances, while Section 5 and 6 discuss implications, limitations, and future research directions.

2. Materials

UBEMs rely extensively on detailed input components to produce meaningful building performance metrics at the urban scale [27]. These data inputs include factors such as building physical features (size, height, type, building systems, building functionality and requirements), occupancy and human behavior patterns, weather and climate data, and community level information (local building code policies, historical redlining maps, urban density characteristics). The following sub-sections outline the datasets and attributes that are included in this study.

2.1 Study Area

Seattle, Washington was chosen for this validation study due to its large urban growth within the past few decades [152] and its increasing exposure to extreme weather events [153], which critically influence

energy demands. Seattle is in the Pacific Northwest region of the United States (47.6° N, 122.3° W), as shown in Figure 6. It is the largest city in the State of Washington and the 18th most populated city in the US [154], having 737,015 people (2020 est) [155]. The city has a land area of 83 square miles [156], with 49% of the total area zoned for single family housing, 14% parks and open space, 11% major institutions and public facilities/ utilities, 8% multifamily housing, 6% commercial and mixed use, 5% industrial, and 5% vacant [156]. The overall climate is moderate with an average annual high temperature of 60° F and an average annual low of 45°F; additionally, the city has an average of 147 days with precipitation per year and 2163 annual hours of solar irradiance per year [157].

Seattle has been experiencing notable extreme heating events in the past few years [153] and is expected to average an additional three heat waves per year by 2030-2059 compared to the extreme weather events observed between 1970-1999 [158]. To combat the adverse health effects caused by these events, the city has increased the resiliency and adaptability of its building infrastructure. For instance, residents are installing air conditioning at a high rate, with a 13% (from 31% to 44%) increase in AC adoption between 2013 and 2019 [159]. These infrastructure adaptions and developments pose increased energy demand concerns to the city, and thus identify an optimal location to conduct a building energy performance case study.

2.2 Building Geometric, Nongeometric, and Energy Usage Data

Building Energy benchmarking data is commonly used for UBEM input or validation data, as it is reported and openly available for many major cities across the United States; including the District of Colombia [160], Chicago [161], Los Angeles [162], New York [163], and Seattle [164]. In Seattle, buildings larger than 20,000 square feet in city limits are mandated by the Seattle Municipal Code 22.920 to track their structures energy performance on an annual basis [164] and report their consumption to the Seattle Energy Benchmarking and Reporting program, who then makes this data open source [164]. In this study, the Seattle 2021 Benchmarking Dataset is the primary dataset used for building input data, reporting building geometric characteristics (ex. number of floors, gross floor area), building non-geometric (ex. building type, year built, energy star score), and building resource consumption (reported in electricity, natural gas, and steam use as well as normalized for building area (EUI), and weather (EUIwn) for more than 3,600 of the largest nonresidential and multifamily buildings within the city limits [165].

Data quality assurance checks, outlier detection, and compliancy checks were implemented on the Seattle Energy Benchmarking dataset to ensure data accuracy in this project. That is, only buildings recorded as "complaint" in the reporting procedure were included in the study. The final cleaned dataset used for analysis in this project contained 2021 annual resource consumption observations for 1553 buildings located in Seattle city limits, as highlighted in *Figure 6*. The data attributes included from the Seattle Energy Benchmarking Dataset are highlighted in Table 7 and distributions among these variables are highlighted for Annual Energy Consumption, Number of Building Floors, Building Gross Floor Area, and Year Built characteristics in histograms in *Figure 7*.

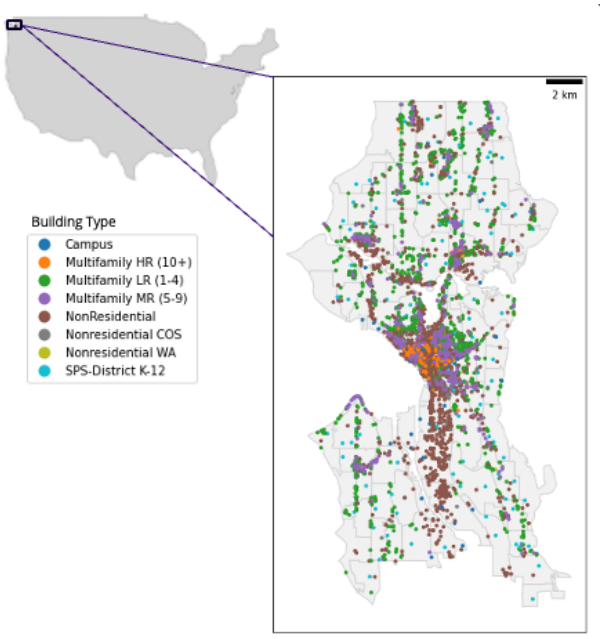

Figure 6: Buildings examined in the Seattle case study, Data Source: Seattle energy benchmarking dataset

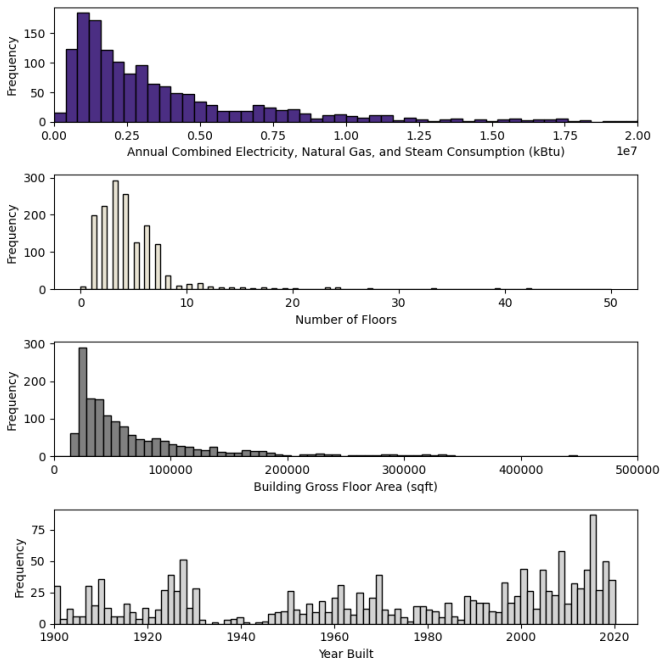

Figure 7: Distributions of specific building characteristics for structures examined in the Seattle case study

2.3 Meteorological Data

Two datasets are formulated and used for UBEM climate and weather data inputs. The first dataset, described in 2.3.1 consists of spatially interpolated TMY3 values. The second dataset, with creation

methodology described in 2.3.2, consists of earth observational weather and microclimate parameters which were gathered from various remote sensing dataset sources.

2.3.1 TMY3 Spatially Interpolated Climate Data

To understand the relationships between microclimates and existing urban building energy modeling methodologies, conventionally used climate datasets were gathered and transformed into comparable weather formats. First, Typical Meteorological Year Three (TMY3) [3] EnergyPlus Weather Files (.epw) were obtained from the EnergyPlus official website [148] for stations within a 25 mile radius of Seattle city limits. This radius value was chosen as a result of examining the nearby stations to Seattle and indicated the optimal distance for locating stations representative to Seattle's climate; for stations outside of the 25-mile boundary are in more mountainous terrain areas, less developed regions, or in harsher, more rainy climates that are not representative of conditions within the city. A total of seven TMY3 station files were captured and used in this analysis; where a map of the stations with boundary file overlay of Seattle city limits is visualized in *Figure 8*. These (.epw) files were then converted to the comma separated value (.csv) format with the EnergyPlus Weather Statistics Conversion Tool [166] and input into Jupyter Notebooks, running on python kernels, for spatial interpolation. Then, TMY3 point data was transformed into a rasterized, spatially interpretable dataset over the Seattle region through cubic interpretation methods with geocube [167], xarray [168], and rasterio [169] python libraries. The TMY3 rasterized spatially interpretable dataset was sampled at every building location, as denoted by the longitude and latitude of all buildings in the energy benchmarking dataset. Finally, the spatially interpolated TMY3 dataset, was merged with the energy benchmarking dataset, embodying a format compatible to that of the earth observational weather and microclimate dataset.

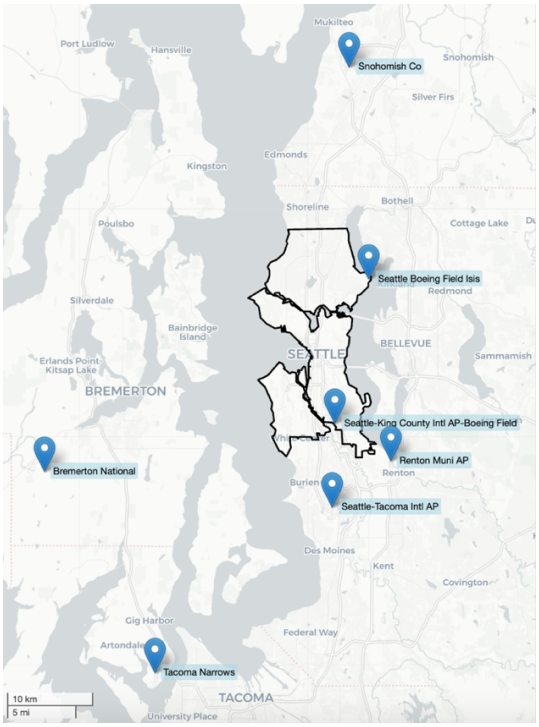

Figure 8: A map of the TMY3 stations within a 25-mile radius of Seattle and respective city limit boundaries

2.3.2 Earth Observational Weather and Microclimate Dataset

To better understand the interactions between spatially precise environmental parameters and building energy performance, earth observational weather and microclimate datasets were sampled at each building location using Google Earth Engine (GEE). First, the GEE Dataset Catalogue (as of 01/2024) was screened for earth observational measurements and reanalysis products relevant to urban microclimates as follows:

A. Observations must be sampled in 2021.
B. Observations must be sampled and collected over the Continental United States of America (CONUS), and specifically over Seattle City Limits
C. In order to maintain data with spatial distinctions across Seattle's 127 km^2 land area, sampling spatial resolution must be lower than 2.5 km [82].
D. To include temporal resolution in the study and to account for seasonality in the data, temporal sampling frequency or revisit must be greater than 30 days.
E. Environmental parameters must be related to urban microclimates as defined in literature, as in [170] [171], [172]..Data must meet quality assurance standards (limited N/As, meet quality assurance pixels, have limited cloud cover), as elaborated on in Section 4.4.

As dataset screening criteria were met, with candidates shown in Table 6, observations taken in 2021 were sampled at each building location (given by longitude and latitude from the Energy Benchmarking dataset) using a scripted automated approach. Then, observations for each environmental parameter were combined and merged to embody the "earth observational weather and microclimate dataset".

Table 6: Microclimate datasets meeting screening criteria, collected from google earth engine

Sensor	Available Products	Spatial Resolution	Temporal Resolution
USGS LANDSAT 8 Level 2, Collection 2, Tier 1 [173]	Ultra Blue, Blue, Green, Red, Near Infrared, Shortwave Infrared, Surface Temperature, Atmospheric Transmittance, Downwelled Radiance, Upwelled Radiance	30 m	16 days
NOAA RTMA [174]	Elevation, Pressure, Temperature, Dew Point Temperature, Wind Speed, Specific Humidity, Wind Direction, Wind Speed (gust), Visibility, Cloud Cover, Precipitation	2500 m	hourly, resampled to daily average
Sentinel- 5P NRTI [175]	Aerosol Index, Aerosol Height, Carbon Monoxide, Formaldehyde, Nitrogen Dioxide, Ozone, Sulfur Dioxide	1113.2 m	resampled to daily average
MOD11A1.061 Terra Land Surface Temperature and Emissivity [176]	Daytime Land Surface Temperature, Nighttime Land Surface Temperature	1000 m	daily
USGS National Landcover Database [97]	Landcover	30 m	2021 instance
USFS Tree Canopy Cover v2021-4 [95]	Tree Canopy Cover	30 m	2021 instance
NASA DEM [96]	Digital Elevation	30 m	2000 tance

3. Methods

In this research, a bottom-up statistical methodology is employed to evaluate how earth observational-based microclimate data effects urban building energy model performance. Different model and variable schemas that incorporate building geometric, building non-geometric, and climate data input parameters are used to predict the annual energy consumption of buildings using machine learning techniques. To spatially contextualize urban microclimates within the city and to examine the influence of climate data on building energy consumption predictions, conventionally used climate and weather datasets are substituted and complimented with earth observational urban microclimate data; and then tested across three distinct model configurations. Two microclimate datasets are created and compared in this analysis, with materials defined in Section 2. The first dataset includes spatially interpolated TMY3 data, which is collected from seven stations within a 25-mile radius of Seattle. The second dataset, coined the 'earth observational weather and microclimate dataset', constitutes of remote sensing weather and climate data. The methodology includes three steps, data processing and merging, feature selection, and model prediction, as defined in *Figure 9*.

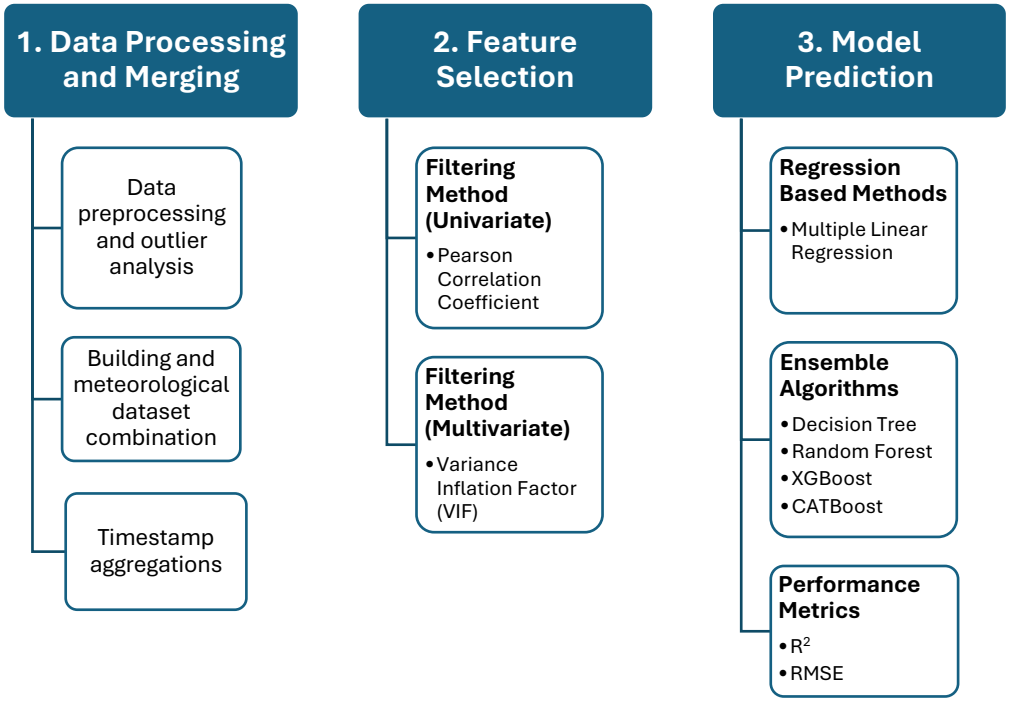

Figure 9: Workflow methods diagram

3.1 Data Preprocessing and Merging

Various pre-processing steps were undertaken to ensure that observations met data accuracy and precision standards. Outlier detection was performed on the energy benchmarking dataset, and buildings with erroneous measurements were removed. Then, TMY3 attributes with more than 20% of observations missing were excluded from the analysis. For the earth observational datasets, data processing steps included filtering for quality assurance bands, filtering for cloud coverage, aggregating data over an annual timestamp, and variable transformations, such as calculating the Normalized Difference Vegetation Index (NDVI). Additional quality assurance analysis steps are elaborated on in Section 4.4.

3.2 Variable Transformation and Feature Selection

Environmental data is often collected at high spatial and high temporal resolutions with closely related attributes, which can lead to significant challenges in data analysis due to correlations. Additionally, selecting the most obvious and relevant features usually yields sub optimally for prediction tasks [177], [178], making feature selection crucial in this analysis. To decrease feature redundancy and model dimensionality in the linear model, feature selection is employed for choosing a subset of the original data attributes in this building energy performance study. Prior to generating linear regression models, the Pearson correlation coefficient was examined, and the Variance Inflation Factor (VIF) techniques were used to select the most optimal features for predicting total annual building resource consumption, calculated by adding electricity, natural gas, and steam use together, the independent variable in this predictive model. Additionally, in order to reduce model biases and dependencies, attributes transformed from energy consumption values or with direct relation to energy consumption were removed from the dependent variable bank, as verified by the VIF analysis, even though these variables are tabulated to bring better model performance [149]. That is, Energy Star Score, Electricity Consumption, Natural Gas Consumption, and EUI attributes were removed from the dependent variable bank. For the machine learning tree-based models (decision tree, random forest, XGBoost, and CATboost), no initial preprocessing step was used for variable selection. Although, obvious relationship dependencies, as noted above, were removed before prior to model training and generation.

3.3 Model Prediction

Three model schemas are generated to examine the impact of urban microclimate data products on building energy prediction performance, as shown in *Figure 10*. The first model schema, S1, includes only building attribute parameters. The second schema, S2, includes both building attributes and TMY3 spatially interpolated climate data. Schema 3, S3, includes building attributes, TMY3 spatially interpolated climate data, and the earth observational weather and microclimate dataset. S1 is the simplest model and is used as a baseline to assess the relative impact of including microclimate and weather data in urban building energy models. S2 is used to access how the weather files from physics-based building energy models perform in terms of microclimate integration. S3 is the most complex schema, it accesses the impact of spatially precise weather and microclimate data in building energy models and is expected to have the best performance given its widespread climate and building attributes.

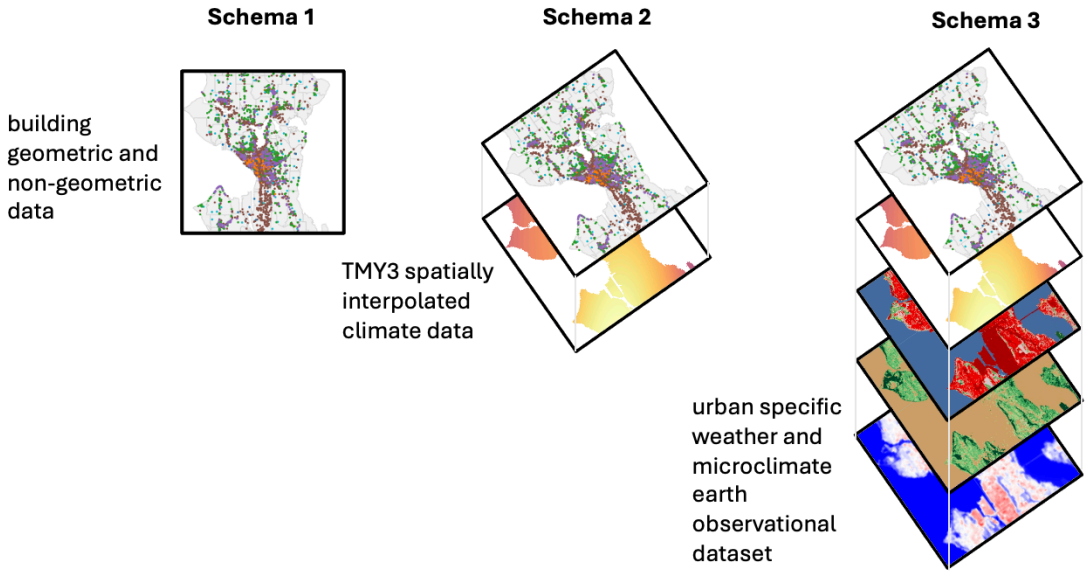

Figure 10: Model variable schemas

Reliable machine learning applications minimize model overfitting and perform well given data they have not seen before. Thus, for each machine learning algorithm examined, training and test dataset are sampled at the same 80% and 20% respective split. Five-fold cross validation (GridSearchCV in the python sklearn library) is then used to hypertune model parameters and optimize prediction accuracy given respective variable schemas. In this research, regression-based methods (linear regression), and tree-based methods (decision tree, random forest, XGBoost, and CATBoost) are explored; as these algorithms have been identified as effective in [60], [149], [130]. Model performance is evaluated under two common statistical indices [179], Root Mean Square Error (RMSE) and the coefficient of determination (R^2); where RMSE measures the standard deviation of model residuals and R^2 expresses the goodness of fit of the model.

Table 7: Description of candidate variables

Category	Variable (GEE Notation)	Unit	Source	Schema
Energy Characteristics	2021 Annual Energy Use (Electricity, Natural Gas, and Steam)	kBtu	Seattle Energy Benchmarking Dataset [180]	S1, S2, S3
Physical Characteristics	Year Built	Year	Seattle Energy Benchmarking Dataset [180]	S1, S2, S3
	Number of Floors	Story		
	Number of Structures on Property	-		
	Gross Floor Area of Building(s)	sft		
	Property Gross Floor Area	sft		
	Building Use Category	-		
	Energy Star Score	Score		
Environmental Characteristics (TMY3 Spatially Interpolated Climate Data)	Dry Bulb Temperature	C	TMY3 Dataset [181]	S2, S3
	Dew Point Temperature	C		
	Relative Humidity	%		
	Atmospheric Pressure	Pa		
	Extraterrestrial Horizontal Radiation	Wh/m_2		
	Extraterrestrial Direct Normal Radiation	Wh/m_2		
	Horizontal Infrared Radiation Intensity from Sky	Wh/m_2		
	Global Horizontal Radiation	Wh/m_2		
	Direct Normal Radiation	Wh/m_2		
	Diffuse Horizontal Radiation	Wh/m_2		
	Global Horizontal Illuminance	lux		
	Direct Normal Illuminance	lux		
	Diffuse Horizontal Illuminance	lux		
	Zenith Luminance	Cd/m_2		
	Wind Direction	deg		
	Wind Speed	m/s		
	Total Sky Cover	.1		
	Opaque Sky Cover	.1		
	Visibility	km		
	Ceiling Height	m		
	Precipitable Water	mm		
	Aerosol Optical Depth	.001		
Environmental Characteristics (Earth Observational Weather and Microclimate Dataset)	Landcover	-	USGS National Landcover Database 2021 Release [97]	S3
	Tree Canopy Cover	%	USFS Tree Canopy Cover v2021-4 [95]	
	Digital Elevation	m	NASA DEM [96]	
	Band 1, Ultra Blue (SR_B1)	-	USGS LANDSAT 8 Level 2, Collection 2, Tier 1 [182]	
	Band 2, Blue (SR_B2)	-		
	Band 3, Green (SR_B3)	-		
	Band 4, Red (SR_B4)	-		

	Band 5, Near Infrared (SR_B5)	-		
	Band 6, Shortwave Infrared 1 (SR_B6)	-		
	Band 7, Shortwave Infrared 2 SR_B7)	-		
	Band 10, Surface Temperature (ST_B10)	C		
	Atmospheric Transmittance (ST_ATRAN)	-		
	Downwelled Radiance (ST_DRAD)	W/(m_2×sr×um)/ DN		
	Upwelled Radiance (ST_URAD)	W/(m_2×sr×um)/ DN		
	Pressure (PRES)	Pa		
	Temperature (TMP)	C		
	Dew Point Temperature (DPT)	C		
	Wind Speed (U component) (UGRD)	m/s		
	Wind Speed (V component) (VGRD)	m/s		
	Specific Humidity (SPFH)	Mass fraction	NOAA RTMA [174]	
	Wind Direction (WDIR)	deg		
	Wind Speed (WIND)	m/s		
	Wind Speed (gust) (GUST)	m/s		
	Visibility VIS)	m		
	Cloud Cover (TCDC)	%		
	Precipitation (ACPC01)	Kg/m_2		
	Absorbing Aerosol Index	-	Sentinel- 5P NRTI [175]	
	Ozone	mol/m_2		
	Daytime Land Surface Temperature	K	MODIS Terra [176]	
	Nighttime Land Surface Temperature	K		
Uncertainty Quantification	Thermal Band Converted to Radiance (ST_TRAD)	W/(m_2×sr×um)/ DN	USGS LANDSAT 8 Level 2, Collection 2, Tier 1 [182]	-
	Pixel Distance to Cloud (ST_CDIST)	km		
	Band 10 Emissivity (ST_EMIS)	-		
	Emissivity Standard deviation (ST_EMSDIS)	-		
	Surface Temperature Band Uncertainty (ST_QA)	K		
	Pixel Quality Attributes (QA_PIXEL)	-		

4. Results

The results of this study are separated into four sections. Notable interactions between environmental variables are first outlined, highlighting the capacity for which different climates and the UHI are observed in Seattle. Second, model and variable schema performances are examined and expanded upon. Third, model variable importance is addressed and the most significant parameters for building energy consumption prediction are highlighted. Finally, data uncertainty are respective project limitations are addressed.

4.1 Variable Interactions

Higher variation between TMY3 and earth observational measurements are observed within the denser metropolitan regions in the city, thus calling for research to examine how microclimate conditions are interconnected with urban infrastructure. This phenomenon is highlighted in *Figure 11* where higher mean annual direct normal radiation measurements are reported by 2021 LANDSAT 8 satellite imagery compared to the spatially interpolated 2021 Annual Meteorological Year (AMY) dataset, derived from the TMY3 dataset using the diyepw [183] library in python. As seen in *Figure 11*, the AMY dataset records a lower mean horizontal direct normal radiation measurement than the Landsat dataset throughout the more developed regions in Seattle. This result emphasizes the need to include higher spatial climate data in for building energy analysis in cities in order to more accurately predict building energy demands (specifically from heating and cooling applications).

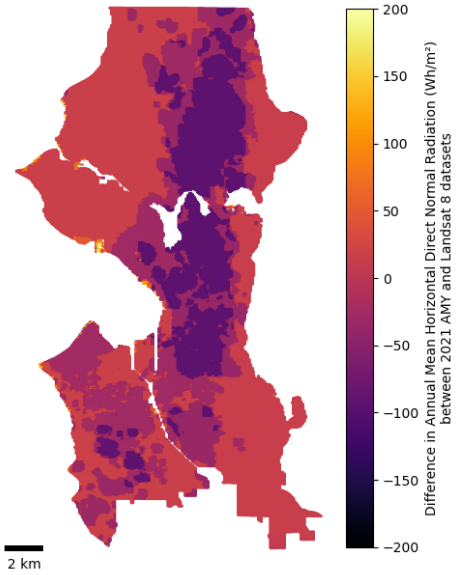

Figure 11: Difference between 2021 mean annual direct normal radiation observations from and spatially interpolated 2021 AMY and 2021 LANDSAT 8 measurements

Buildings sampled in areas identified as USGS Landcover *Developed High Intensity* and *Developed Medium Intensity* categories recorded larger LANDSAT 8 Land Surface Temperatures than the buildings in other landcover categories, as shown in *Figure 12*, *Figure 13*, *Figure 14*, and Table 8. This finding highlights the interactions between urban contexts, high density areas, the UHI, and microclimates, specifically in Seattle, USA and validates the importance of studying their relationships to building energy performance.

Table 8: Average 2021 LANDSAT 8 land surface temperatures reported at buildings in each respective USGS landcover category

USGS Landcover Database Category	**Average 2021 Land Surface Temperature (°C)**
Developed, High Intensity	26.98
Developed, Medium Intensity	26.31
Developed, Open Space	24.91
Developed, Low Intensity	24.26
Open Water	22.64
Barren Land	20.62

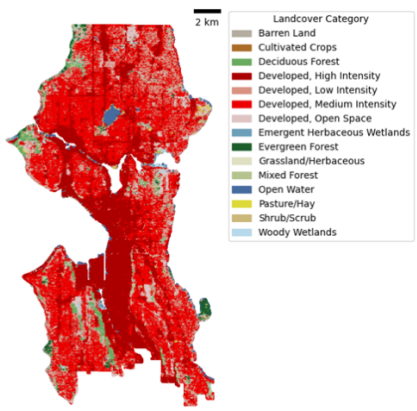

Figure 12: Landcover Category, Data Source: USGS National Landcover Database 2021 Release

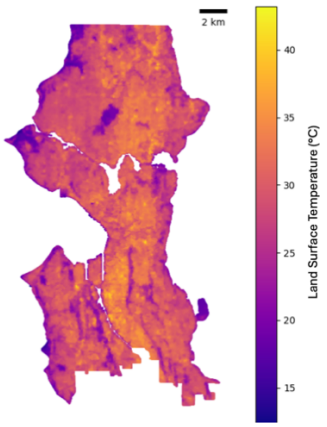

Figure 13: Average 2021 Land Surface Temperature, Data Source: LANDSAT 8 Level 2, Collection 2, Tier 1

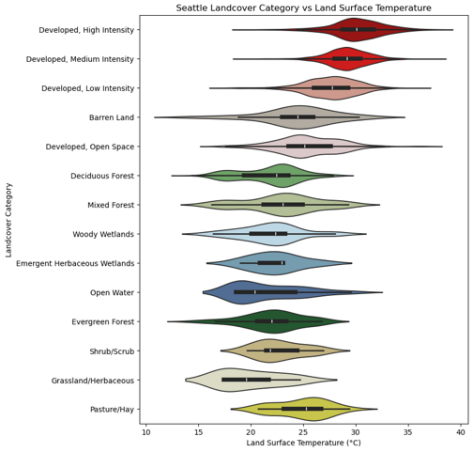

Figure 14: Distribution of Average 2021 Land Surface Temperatures in corresponding USGS Landcover Categoreis

Nighttime land surface temperature is used to study the UHI due to its sensitivity towards urbanization [184]. When sampled at building locations, Daytime LST measurements reported 10.4°C on average higher than Nighttime LST measurements, with mean annual values 19.4 °C and 9.0 °C respectively. Additionally, relationships are well established between land surface temperature, land use land cover, and normalized difference vegetation index [185], highlighting the temporal variability of land surface temperature and its dependence to vegetation and foliage coverage, as shown in *Figure 15*.

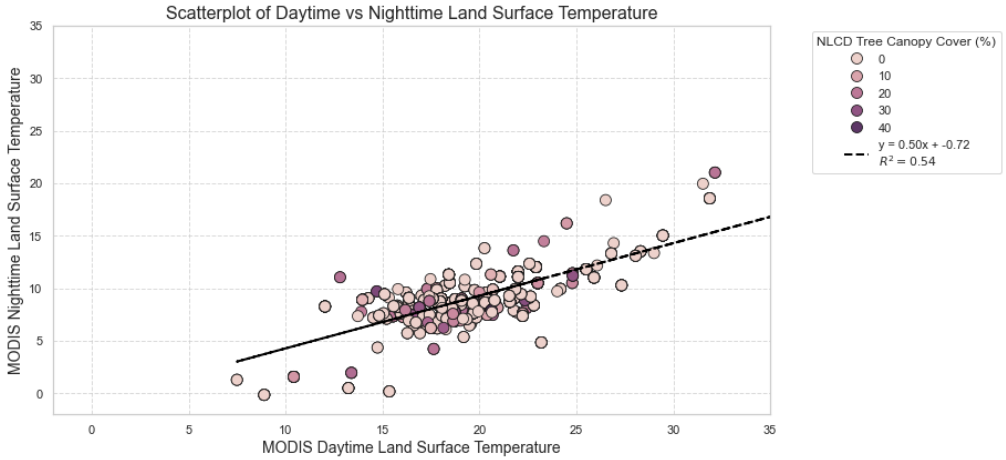

Figure 15: Relationship between MODIS Daytime and Nighttime Surface Temperatures sampled at the building locations

4.2 Model Comparison

To better understand how different data products and modeling techniques effect UBEM performance, different modeling algorithms and variable schemas are used in this analysis. A total of 15 hyperparameter-tuned models are evaluated; where 5 different machine learning models are employed across three independent variable schemas (S1, S2, S3), each incorporating differing levels of climate data inputs. S1 includes no climate data inputs, S2 includes spatially interpolated TMY3 climate data inputs, and S3 includes both TMY3 and earth observational climate data inputs, as shown in *Figure 10* and Table 9. The

best performing model uses S3 climate data inputs and CATBoost ensemble methods to predict building energy consumption; it exhibits a testing R^2 increase of 0.056 (from 0.655 to 0.711) and a 26% testing RMSE decrease over the best performing S2 model, also generated with the CATBoost algorithm. This result highlights the importance of including high spatial resolution earth observational weather and climate data in UBEMs. Moreover, the highest performing earth observational based climate data model in S3, outperforms S1, with no climate data attributes, by a testing R^2 increase of 0.16 (0.55 to 0.71) and a 42% testing RMSE decrease; meaning that UBEMs perform significantly better with reference weather and climate data. These results emphasize the importance of including high spatial resolution microclimate data, specifically earth observational data products in urban building energy modeling analysis.

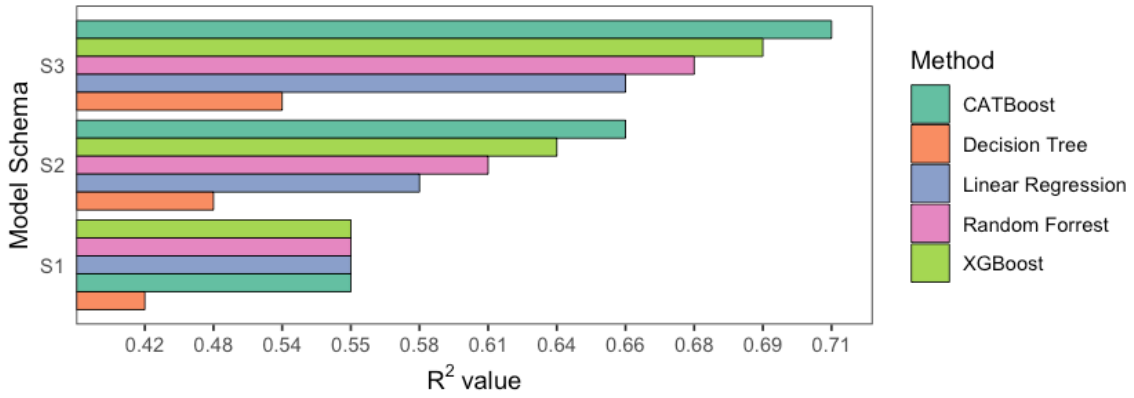

Figure 16: Model R^2 values for the variable schemas and machine learning algorithms examined (testing dataset)

Table 9: Training and testing set performances of different UBEM model schemas and algorithms

Method	Schema	Hyper-Parameters	Training R^2	Testing R^2	Training RMSE	Testing RMSE
CATBoost	S3	od_wait = 10, border_count= 128, max_depth= 7, l2_leaf_reg= 1, iterations=400, learning_rate= 0.07	0.990	0.711	833957	3315082
XGBoost	S3	max_depth=5, random_state=42, n_estimators=400, colsample_bytree = 0.8, gamma= 0,reg_lambda= 3, reg_alpha= 0.01, min_child_weight= 1	0.999	0.690	18266	3437341
Random Forrest	S3	max_depth= 15, min_samples_leaf= 1, min_samples_split=2, n_estimators=200, oob_score= True, max_features= 'auto'	0.949	0.679	1906620	3495887
Linear Regression	S3	NA	0.747	0.666	4242415	3565476
Decision Tree	S3	max_depth= 5, max_features= auto, min_samples_leaf= 1, min_samples_split=5	0.793	0.543	3840763	4173645
CATBoost	S2	bagging_temperature= 0.0, border_count= 128, depth= 6, iterations= 300, l2_leaf_reg= 5, learning_rate= 0.3, random_strength= 0.5	0.980	0.655	1112784	4535810
XGBoost	S2	learning_rate=0.1, max_depth=5, min_child_weight=3, n_estimators= 200, reg_alpha= 0.1, reg_lambda= 1, subsample= 1.0	0.997	0.635	418412	4666079
Random Forrest	S2	max_depth=14, min_samples_leaf= 1, min_samples_split= 2, n_estimators= 200, oob_score=True	0.948	0.611	1791740	4819011
Linear Regression	S2	NA	0.721	0.584	4185207	4984381
Decision Tree	S2	max_depth= 4, max_features= auto, min_samples_leaf= 1, min_samples_split= 2	0.741	0.484	4033558	5552003
Linear Regression	S1	NA	0.780	0.552	4267487	5758214

Random Forrest	S1	max_depth= 6, min_samples_leaf= 1, min_samples_split= 4, n_estimators= 300, oob_score= True	0.855	0.552	3470253	5762292
XGBoost	S1	learning_rate= 0.3, max_depth=3, min_child_weight': 1, 'n_estimators': 50, 'reg_alpha': 0.1, 'reg_lambda': 3, 'subsample': 1.0	0.865	0.551	3342241	5765116
CATBoost	S1	bagging_temperature= 0.0, border_count= 128, depth= 3, iterations= 200, 'l2_leaf_reg= 1, learning_rate= 0.3, random_strength= 0	0.928	0.547	2432869	5794636
Decision Tree	S1	max_depth= 5, max_features= auto, min_samples_leaf= 1, min_samples_split= 10	0.785	0.416	4219167	657546

In this study, nonlinear relationships better represent the interactions between building data, climate data, and energy predictions when given a wide range of variable attributes. Whereas, with limited data feature availability, linear models outperform the complex models. For the model schemas with more variable attributes, such as S2 and S3, ensample-based machine learning algorithms outperform linear and tree-based models, as shown in *Figure 17*. This is potentially due to the algorithms' ability to characterize the data's high dimensionality and nonlinearity through week ensembles chaining and combination processes; thereby improving model prediction accuracy and stability [186]. For model schemas S2 and S3, the gradient boosting CATBoost algorithm outperform the other machine learning algorithms, followed by XGBoost, and random forest. For the least complex model schema, S1, linear methods and random forest methods perform best. In all cases, decision trees have the worst model performance, potentially due to overfitting and their limited ability to contextualize relationships given unseen data.

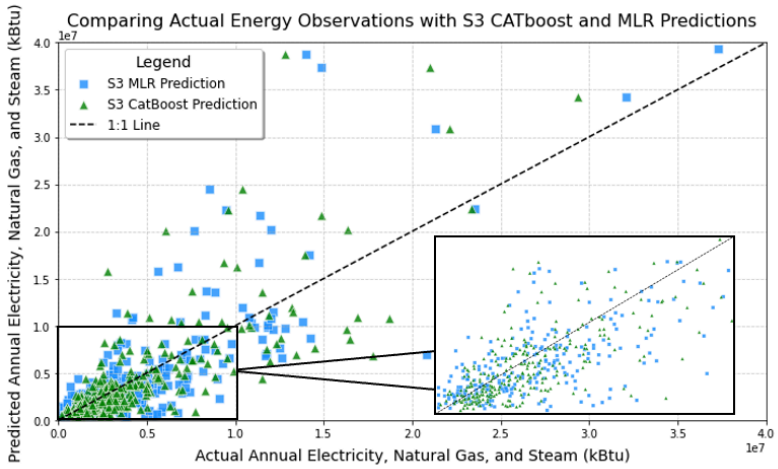

Figure 17: Scatterplot comparing S3 MLR and CATBoost predictions and actual observations (testing dataset)

To better illustrate algorithm performance, *Figure 17* depicts S3 MLR and CATBoost predictions, and compares them with actual energy consumption values, as reported by the Seattle Energy Benchmarking dataset; this plot showcases the wide scatter among predictions and the highlights the performance of the CATboost model. *Figure 18* showcases the residuals from both S3 Linear model and CATBoost predictions. This plot shows that the CATBoost residuals are more frequently centered around smaller values than those of the linear model, with less density throughout the model tails—thereby graphically indicating a better fit and better model prediction from the CATBoost model, compared to the linear model.

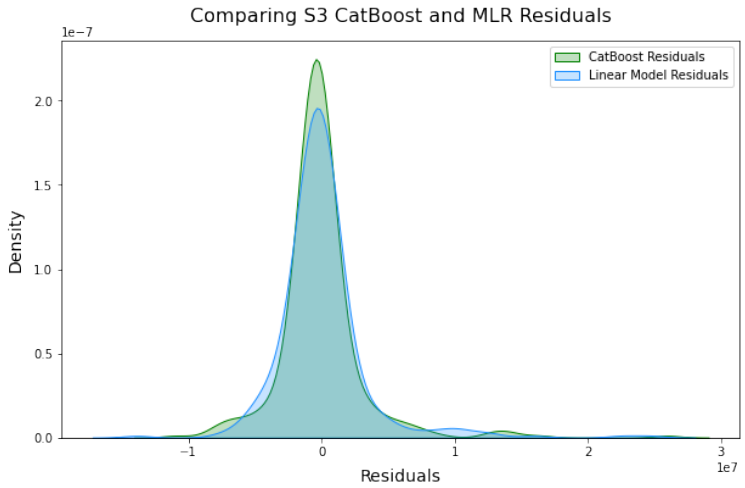

Figure 18: Residuals from the S3 MLR and CATBoost prediction models (testing dataset)

To understand model performance in terms of building energy consumption metrics, the percentage difference between the actual observed energy consumption and the modeled energy consumption was calculated for each individual building through the CATBoost ML algorithm for S1, S2, and S3, over both the training and testing datasets, as shown in *Figure 19*. All three CATBoost models were found to overpredict building energy consumption compared to actual observed energy consumption. At the city scale, the S3 model overpredicted energy consumption by 0.28%, S2 by 0.68%, and S1 by 0.75%. These city scale UBEM performances are lower than many reported UBEM accuracies [187], potentially due to the fact that this study is statistical in nature, where the methods tend standardize the overall residuals around 0. At the single building level, the S3 model overpredicts building energy consumption by 19% on average (5.6% median), the S2 model overpredicts building energy consumption by 21% on average (6.3% median), and the S1 model overpredicts building energy consumption by 28% on average (11% median). These building level discrepancies are within range of well preforming UBEMs, whereas for individual buildings at annual resolution, it is not uncommon for outlier errors to reach up to 1000% [187] . Regarding study takeaways, in all three variable schema instances (S1, S2, and S3), the statistical CATBoost models over predict building energy consumption at the individual building level. This is hypothesized to be due to contributions of individual parameters on model outcomes. Therefore, more elaborate performance indices and overestimation/ underestimation trend are expanded on across specific building attributes in Section 4.3.

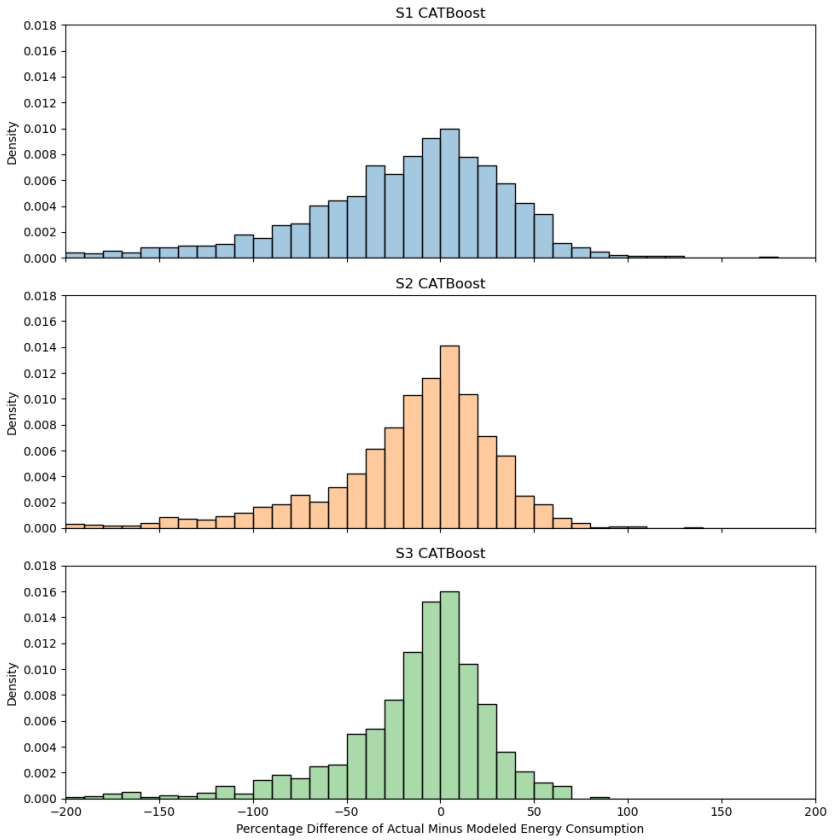

Figure 19:Percentage Difference in Performance of Actual Minus Modeled Energy Consumption at the Individual Building Level (testing and training dataset)

4.3 Feature Importance

Model results reveal the influence of urban microclimate conditions on building infrastructure. In order to relate these results to actionable measures for additional prediction tasks, the most dominant model parameters were identified using the CATBoost feature importance function. This calculation quantifies the average change of model prediction value, given a change in feature value [188]; where larger values of importance indicate greater changes to model prediction. To facilitate comparison between parameters within the same model, each importance is normalized, making the sum of the importances for all the features equal to 100. The relative importance of model features and variables for the best performing CATBoost S1, S2, and S3 models are quantified and plotted in *Figure 20*, *Figure 21*, and *Figure 22* respectively.

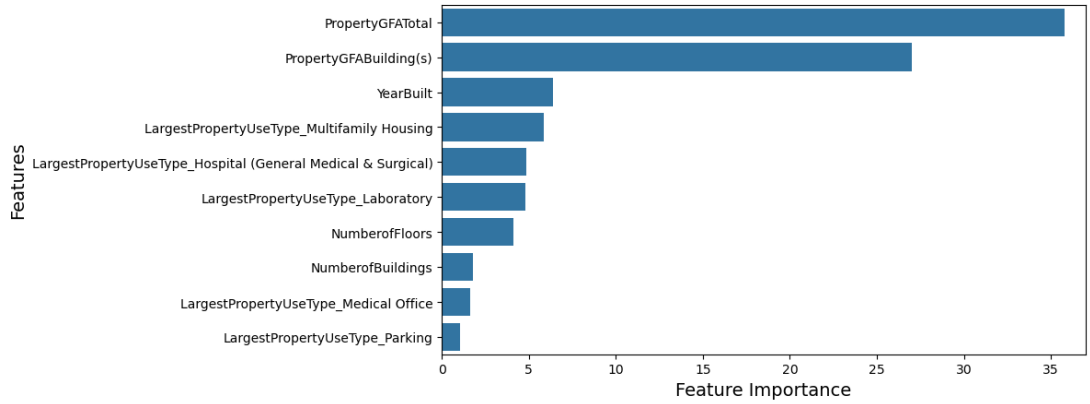

Figure 20: Feature importance for the best performing CATBoost S1 model (testing dataset)

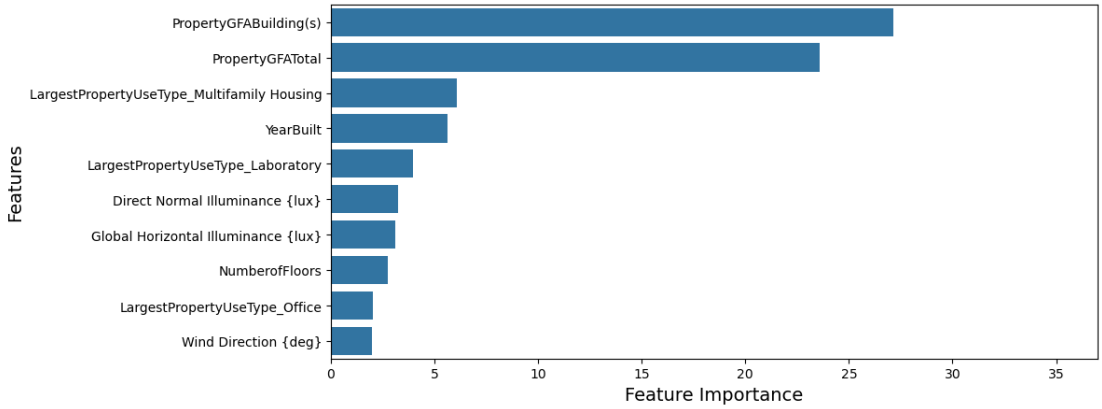

Figure 21: Feature importance for the best performing CATBoost S2 model (testing dataset)

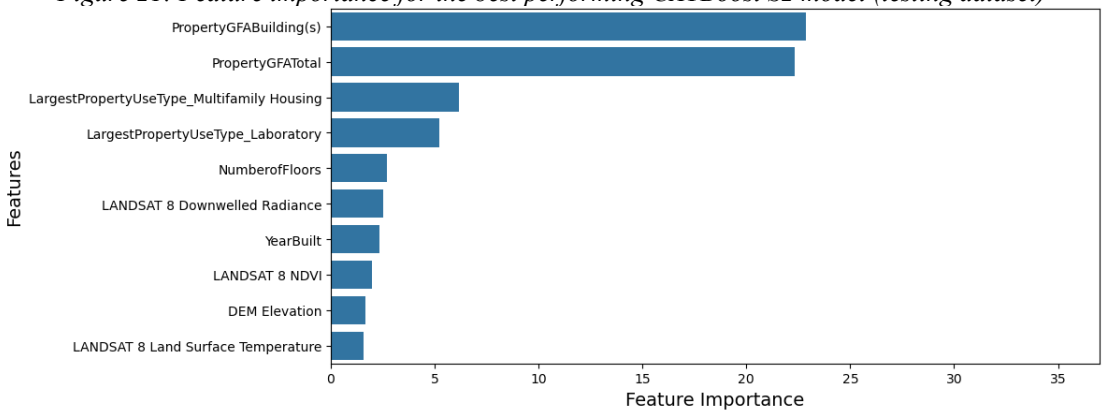

Figure 22: Feature importance for the best performing CATBoost S3 model (testing dataset)

4.3.1 Building Geometric and Non-Geometric Feature Importances

For all three model schemas, total property gross floor area (GFA), building GFA, year-built, number of floors, and building archetype - specifically the multifamily housing archetype, have the largest influence on energy consumption prediction for the building geometric and nongeometric variable inputs, as indicated in the feature importance analysis. Therefore, it is of interest to visualize these parameters and their respective residuals in isolation, to better understand their integrations and individual contributions to the models. In isolation, total property gross floor area (GFA) and building GFA have the largest linear relationship to observed Building Energy Consumption, as shown in *Figure 23*. These results are broadly

in line with the findings from the 2016 Seattle energy benchmarking report, that documents large correlation between the size of the building and the energy consumed [189]. Such a rank for Gross Floor Area in data-driven urban energy use modeling has also been reported for other cities, such as a model developed for Chicago, IL [190]. This result is also consistent with energy policy, as buildings energy intensity, or energy consumption normalized by building area, is a metric often used to benchmark a structures performance [191], [192]. It should be noted that normalizing building energy consumption by gross floor area is known to exhibit a dilution effect [191], and it is not all encompassing to describe a structures performance linearly, thus, it is important to consider other metrics (such as other building characteristic and climate parameters) and non-linear effects when evaluating building efficiency.

Model discrepancies, reported as the percentage difference between the energy observation and predicted value, are highlighted for total property gross floor area (GFA), building GFA, number of floors, and year-built parameters in isolation, in *Figure 23*. These plots depict the boundaries for optimal model performance across different model parameters over both the training and testing datasets. For instance, most of the buildings examined in this study have a property gross floor area less than 250,000 sft (*Figure 7*). However, the CATBoost S3 model performs worse for buildings with properties less than 250,000 sft, overpredicting energy consumption by 20.4% on average. Whereas the CATBoost S3 model performs better for properties greater than 250,000 sft, overpredicting energy consumption by 5.8% on average.

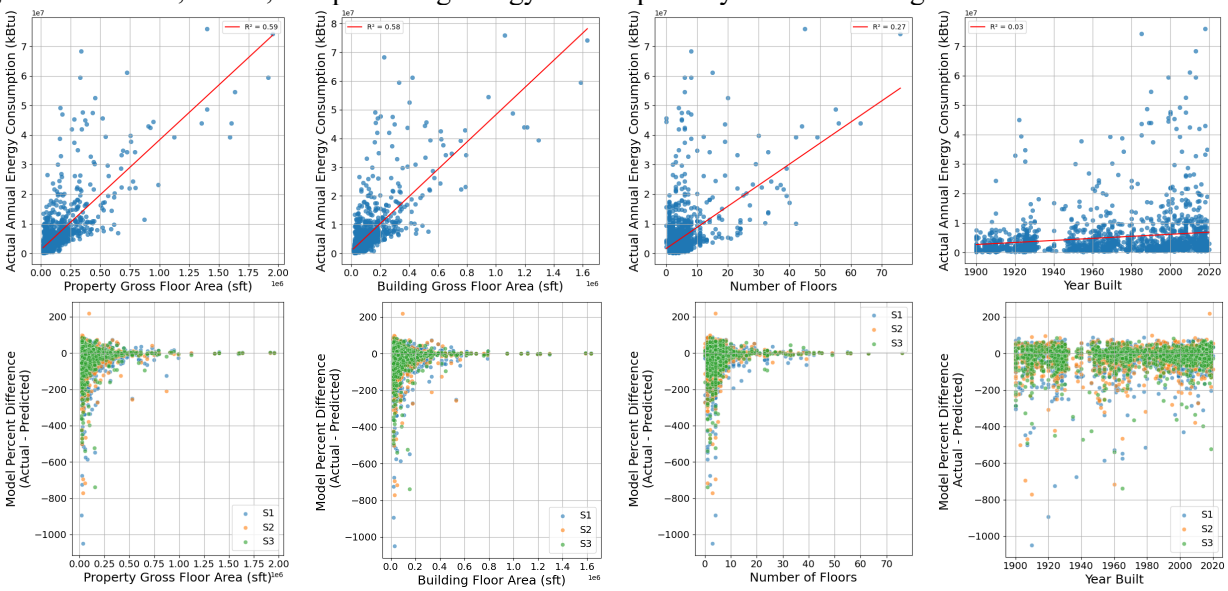

Figure 23: Isolated relationship between the most influential building energy model parameters and actual energy consumption (top) and their respective model discrepancy (bottom)

4.3.2 Weather and Climate Feature Importances

The most influential S2 TMY3 microclimate variable data inputs, as illustrated in *Figure 21* include direct normal illuminance, global horizontal illuminance, and wind direction. Whereas S3 environmental features with the largest model contributions include LANDSAT 8 downwelled radiance, LANDSAT 8 NDVI, DEM Elevation, and LANDSAT 8 Land Surface Temperature variables. Significant downwelled radiance band contributions are consistent with the results tabulated in a building energy and microclimate study in New York City [82]. These feature importance findings are notable because these earth observational data products and parameters are not included in TMY3 datasets and thus traditional UBEM analysis'. These parameters thereby offer a promising potential for improving the accuracy and performance of UBEMs in terms of microclimate data integration. Furthermore, these earth observational products have a higher

spatial resolution than the TMY3 measurements and thus may offer more spatially precise insights to city climate characteristics.

Less linear trends are observed between most influential earth observational weather and climate data parameters and annual building energy consumption, compared to the most influential building geometric and non-geometric features (total property gross floor area (GFA), building GFA, and number of floors), that are elaborated on in Section 4.3.1. Larger average modeling errors (19.3%) are found at less extreme Land Surface Temperatures (between 24 °C and 32 °C) compared to those (15%) at more extreme Land Surface Temperatures (bellow 24 °C and above 32 °C), as shown in *Figure 24*. Whereas most of the buildings examined (90%) had an average land surface temperature recording between 24 °C and 32 °C. This means that the S3 CATBoost model handles model performance given outliers in extreme temperatures; however, the model has more difficulty characterizing consumption where most of the building data is located. The more scatter across buildings in these regions may be caused by building characteristic differences and respective collinearities to land surface temperature that are not explicitly examined in the model (such as correspondence with building material properties, anomalous operation, ect.).

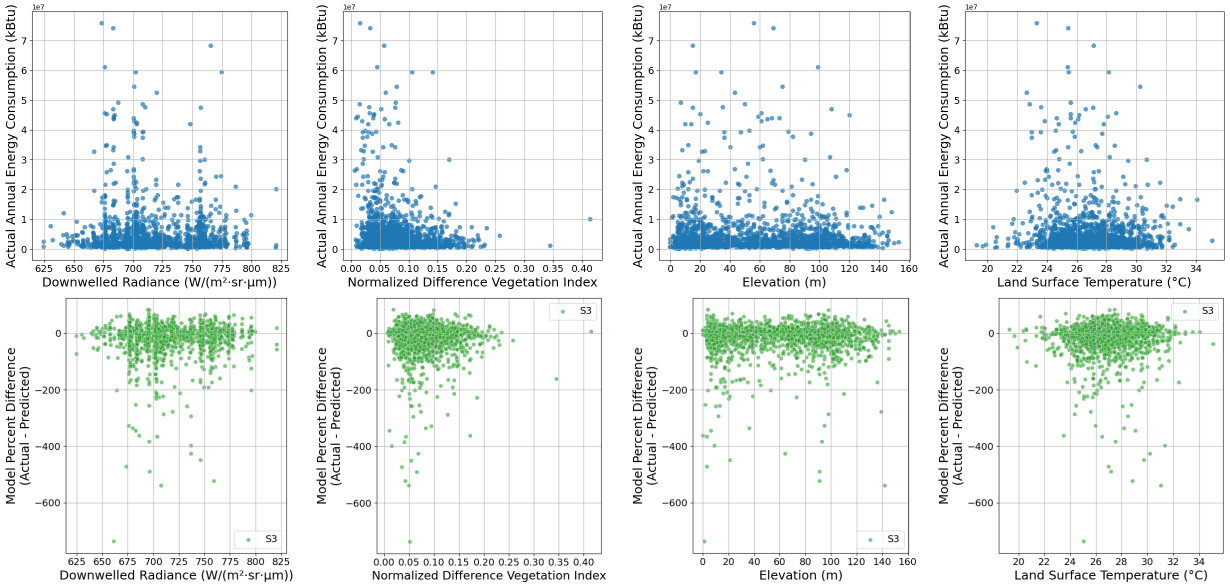

Figure 24: Isolated relationship between the most influential climate building energy model parameters and actual energy consumption (top) and their respective model discrepancy (bottom)

4.4 Data Uncertainty

Earth observational data sensing techniques lack the ability to penetrate clouds, increasing the uncertainties associated with this datatype. Moreover, it is well tabulated that cloud cover can greatly influence Land Surface Temperature (LST) measurements [193]. These limitations has profoundly influenced data quality though restricting data availability to timestamps with clear sky conditions. To better understand the limitations of these datatypes within our studies context, LANDSAT data uncertainty quantification bands, specifically the LANDSAT 8 Land Surface temperature quality assurance bands (ST_QA), as tabulated in the bottom of Table 7, are examined. Throughout this studies timeframe, the average annual LST uncertainty is 3.49 °C. This uncertainty value shall be compared to those reported for the TMY3 dataset, which depends on National Weather Service practices and the instrument used to obtain the data [194]. Further, we find that the largest LST uncertainty measurements are reported from October through January, potentially due to the increased cloud coverage throughout these timestamps, as shown in *Figure 25* and *Figure 26*. Regarding the impact on project results and future research directions, earth observational climate and weather data have higher sensor accuracy in Seattle throughout hotter and low cloud cover

summer periods, thus indicating a better resource to study urban building energy consumption throughout these timeframes.

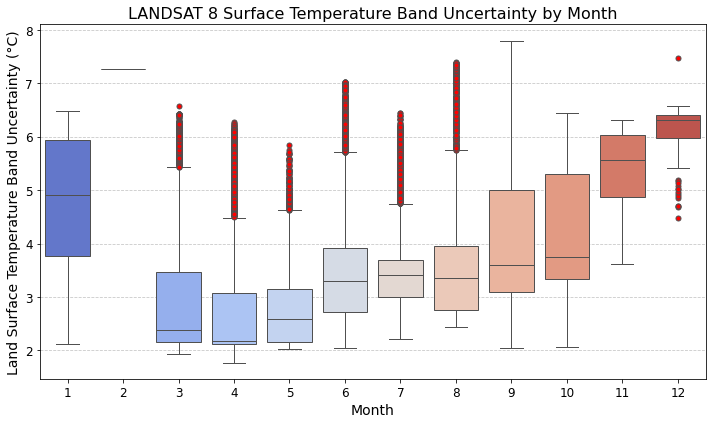

Figure 25: LANDSAT 8 Surface Temperature Band Uncertainty Measurements by Month (reported by the ST_QA band)

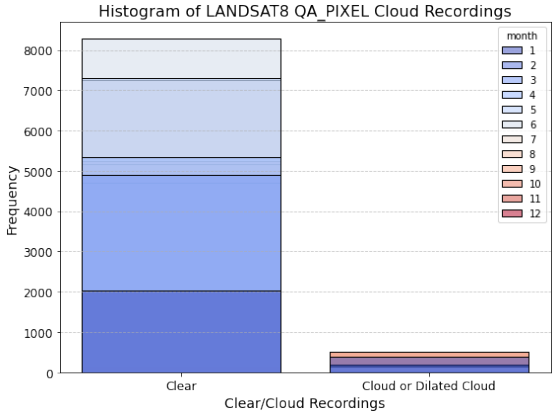

Figure 26: LANDSAT 8 Quality Assurance Cloud Recordings (reported by the QA_PIXEL band)

5. Discussion

In this research, large discrepancies between conventional UBEM weather data inputs (TMY3 files) and higher spatial resolution earth observational measurements are observed within the more developed and urban regions in Seattle. This finding has a direct relationship to building energy demand predictions, as the use of conventional climate data inputs may mischaracterize actual building energy consumption demand by a significant margin. Through methods that compare different data schemas, we find significant building energy modeling improvements, measured by r^2, rmse, and percentage of accuracy discrepancy, given the inclusion of earth observational climate data. Moreover, we demonstrate the capacity for earth observational climate data to spatially contextualize microclimate conditions in UBEMs; thereby validating these datasets for use in UBEMs in order to assist with characterizing annual building energy needs in cities while including microclimate effects.

The ability for conventionally used UBEM climate data to contextually map a city is heavily dependent on the quantity of recording and reporting stations in the urban area of interest. Earth observational data can interpret urban microclimate conditions at a more precise spatial granularity than point source observations; however, should be recognized for its limitations in data quality and uncertainty. As elaborated on in Section 4.4, it should be noted that cloud cover but may characterize some of the uncertainty from earth

observational measurements but may not be all encompassing of these bias. Similar to other observational data types, EO data suffers from uncertainties from biases in sensors, sensor drift, and retrieval algorithms, calling for the accuracy of measurements to be addressed explicitly [80]. Moreover, for application in building energy consumption studies land surface temperature measurements remain a proxy for air temperature measurements but fundamentally measure different properties, therefore cannot be used to replace TMY3 air temperature inputs altogether. For these reasons it is recommended for future research to investigate the blending of earth observational datasets into physics based urban building energy studies; which includes investigation of these datasets in different climate zones and throughout different temporal aggregation conditions (ex. such as during an extreme weather event and at smaller hourly, daily, monthly, timestamps).

A series of limitations exist in this work. First, not necessarily all model performance increases can be attributed to environmental factors. That is, there may be correlation among environmental parameters and other building properties that are not explicitly examined in the model but may be correlated with some model variables thereof. Instances of this phenomena could include the interconnectivities between variable such as surface reflectance and land surface temperature and their relation to building façade type [195], or window to wall ratio [196]. Furthermore, there may be inherit correlation between variables included within this research and analysis. For instance, both property square footage and building square footage were included in this model to account for anomalies in property use, although correlation was intended to be mitigated through integration of nonlinear model algorithms and VIF threshold analysis, this inclusion of parameters could result in excess model uncertainties. Additionally, utilization of all TMY3 parameters may introduce multicollinearity and noise into our S2 and S3 models. However, physics-based models typically utilize all TMY3 parameters in their building simulations, therefore, we hoped to ensure that our results were comparably applicable. Furthermore, buildings are highly complex systems, where occupancy loadings [197], façade type, window to wall ratio [198], operational hours of opening, and other factors are documented to impact performance. Due to data availability limitations, these attributes are not explicitly examined in this work but can play a substantial role in building energy consumption calculations.

This study is limited by the temporal resolution of both energy and climate data. That is, energy measurements included in this study are reported on an annual basis. This is a limitation of our research study, as annual aggregation assumes a mean value of operation and environment and fails to characterize the impact of building energy loadings throughout hourly, daily, and monthly timescales, both in anomalous building operation and extreme weather events. Furthermore, energy benchmarking data is only available for a small subset of cities which are typically located in high-income counties, limiting reproducibility of similar work in different climate zones and in locations with alternative building infrastructure systems and developments.

This study examines the effect of urban microclimate conditions throughout the entirely of Seattle City limits, neglecting the capacity to make informed city design and planning recommendations at more localized city district or neighborhood levels. However, it is well established that populations are exposed disproportionately to urban microclimate conditions and the UHI effect, raising concern in terms of the sustainable social development of our cities. Therefore, it is recommended for future research to not only to discuss engineering solutions, but also to identify strategies that reduce energy burdens that are in result of urban microclimate conditions to promote equitable thermal comfort and energy consumption across neighborhoods.

It should be noted that the typical limitations for UBEM studies are also limitations in this study [199], Moreover, while statistical energy benchmarking studies can be conducted over a large number of buildings, they are often criticized for providing narrow insights into efficiency improvements at the building system level [200], [201]. That is, statistical energy research studies should not replace simulation-

based energy studies but rather compliment them. It should be noted that simulation-based energy platforms predict consumption at an hourly frequency and utilize climate variable inputs that are reported at the same frequency. However, most earth observational sampling procedures do not operate at such a high temporal revisit period, thus posing challenges for augmenting earth observational data in simulation based UBEMs. It should be of interest for researchers to leverage the spatial capacities of this datatype in simulation-based energy models [16]. Thus, we call for future research to investigate the feasibility of and applicability thereof for including earth observational climate data in simulation-based building energy models. This can be done through a hybrid approach, or through creating a schema that augments earth observational satellite products into energy plus weather files directly.

It is of upmost concern to create a robust and resilient energy predictive model that can be transferable to applications greater than the study thereof. This study presents a statistical methodology that accesses the dependability of earth observational microclimate data for urban scale energy modeling, while promoting the use of data-driven decision making to advise energy efficient strategies at the urban scale.

6. Conclusion

Building energy models face notable limitations in accurately representing urban energy use, resulting in a substantial gap between simulation results and actual measured energy data. This shortfall has led to discrepancies in the performance of structures; where modeling inconsistencies are further intensified in urban areas due to inter-building connections and microclimate conditions, such as the urban heat island effect. This study utilizes a bottom-up statistical approach to examine the relationships between microclimates and building energy demand. The methodology leverages earth observational environmental data products and combines these with city building energy benchmarking data, using machine learning algorithms, to examine the interactions between urban climate and energy systems. Substantial reduction in model RMSE, increase in model R^2, and decrease in percentage discrepancy between measured and modeled values are observed when earth observational climate parameters are added to existing climate and building UBEM configurations. Thus, proving earth observational attributes to be reliable data sources for future investigation of energy consumption of urban buildings for the inclusion of microclimate effects. Further, LANDSAT 8 downwelled radiance, LANDSAT 8 NDVI, DEM Elevation, and LANDSAT 8 Land Surface Temperature variables have the largest model contributions from the earth observational microclimate data product inputs, signaling their correlations with building energy consumption and suggesting future research to explore their contributions in UBEM models. This study highlights the importance of addressing both building and climate granularity in urban building energy studies. Although, it is inherently limited by data availability to explore the complex relationships between variables across smaller temporal timescales (ex: daily, monthly, hourly). Future work in this area is encouraged to explore methodologies that adapt a hybrid approach which combine earth observational based climate data into simulation-based building energy platforms, to more accurately represent microclimate conditions in UBEMs while still effectively managing simulation-based computational demands across extensive temporal and spatial simulations. This study underscores the importance of bridging the existing gaps between modeled and real-world energy scenarios through incorporating microclimate data into UBEMs; ultimately advancing our collective goal to foster a more sustainable energy future.

Chapter 4: A framework for augmenting simulation-based building energy models with earth-observational microclimate data using machine learning predictions

Authors: Amanda Worthy[1], Mehdi Ashayeri[2], Julian Marshall[1], Narjes Abbasabadi[3]

[1] Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, United States
[2] School of Architecture, Southern Illinois University, Carbondale, IL, United States
[3] Department of Architecture, University of Washington, Seattle, WA, United States

Abstract

Accurate urban building energy modeling (UBEM) is constrained by mismatches between standard climate inputs and actual urban microclimate conditions. This study introduces a scalable, bottom-up, framework that integrates EnergyPlus building energy modeling simulation outputs with earth observational and GIS-based data, which are enhanced through machine learning techniques to improve energy demand predictions in urban settings. Applied to Los Angeles (LA), CA, we evaluate the representativeness of typical meteorological year (TMYx) sampling sites against actual urban environmental conditions. We find that while satellite-derived surface temperatures show reasonable alignment with average city conditions, significant discrepancies are observed in urban form metrics such as tree cover, street cover, and building density; suggesting that TMYx stations should be placed in denser urban areas. We augment EnergyPlus simulations for 19 single-family buildings, with remote sensing data using machine learning models— to generate city-wide residential energy consumption heatmaps corrected for microclimate conditions. Models capture substantial intra-urban variation, with predicted energy use differing approximately 10% between neighborhoods. Feature importance analysis highlights land surface temperature as a key predictor, underscoring its relevance to building energy research. We also find the majority of TMY3 sampling sites to be in low-vulnerability areas, underscoring the structural mismatch that is embedded in urban form and climate. This framework offers a scalable path for integrating urban microclimate effects into energy modeling to enable more precise and equitable energy policy and planning.

Keywords: Urban Microclimates, Urban Building Energy Use, Earth Observational Data, Infrastructure

Equity

1. Introduction

Urbanization has become a critical challenge that presents obstacles for achieving sustainable social, economic, and environmental development in cities [1]. Growing urban populations have made infrastructure developments more centralized, interconnected [2], and vertical [3], contributing to substantial shifts in land use patterns and environmental characteristics [4]. The rise of network density has altered the surface energy balances, thermal properties [5], [6], [7] native foliage concentrations [6],[7], weather patterns [8], fluid flow characteristics [9], [10], and albedo reflectance's [11] of urban areas; consequently leading to a phenomenon where cities experience local warming greater than surrounding less developed areas, formally known as the Urban Heat Island (UHI) effect [12],[13], [14],[15]. These altercations through urban form have given rise to urban microclimate conditions in cities — defined as locations where differences in outdoor climate are observed between urban and rural areas when referenced in similar geographical locations [10]. Understanding the interactions between these urban conditions and infrastructure is pivotal to achieve more sustainable and resilient development in both present and future cities.

Abbreviations

BPS	Building Performance Simulation
CatBoost	Categorical Boosting
DEM	Digital Elevation Models
DT	Decision Tree
EO	Earth Observational
EUI	Energy Usage Intensity
EPW	Energy Plus Weather File
GEE	Google Earth Engine
GFA	Gross Floor Area
HI	Heat Island
L8	Landsat 8
LiDAR	Light Detection and Ranging
LST	Land Surface Temperature
ML	Machine Learning
MLR	Multiple Linear Regression
NDVI	Normalized Difference Vegetation Index
RMSE	Root Mean Square Error
RF	Random Forest
SRTM	Shuttle Radar Topography Mission
TMY	Typical Meteorological Year
UBEM	Urban Building Energy Modeling
UCM	Urban Canopy Model
UHI	Urban Heat Island
UHII	Urban Heat Island Index
VIF	Variance Inflation Factor
XGBoost	eXtreme Gradient Boosting

Urban microclimate conditions have a profound influence on city infrastructure, specifically energy consumption [13]. That is, urban microclimates have altered thermal balances surrounding buildings, greatly influencing the energy demands required to maintain healthy indoor conditions. Coupled with long term climate changes and increasingly frequent extreme weather events [202], urban microclimate conditions have shaped how energy consumption loads are managed in cities, raising human health and wellbeing concerns.

Populations are disproportionately exposed to urban microclimate conditions and the UHI effect, raising concern in terms of the sustainable social development of our cities. In the continental United States (U.S.), populations of color face higher levels of UHI exposure compared to non-Hispanic white populations in 96.6% of the largest urban areas [17]. Additionally, low income US census blocks have 15.2% less tree cover and experience temperatures that are 1.5% hotter than high income blocks [18], [19]. Since dense vegetation provides roughly 6% energy savings in buildings, similar to the magnitude of savings from mechanical retrofit interventions [20], this is not only an issue of environmental exposure, but a conversation about the energy burden imposed on certain communities to maintain comfortable temperatures indoors. In Portland, Oregon, it is documented that the energy burden due to extreme heat disproportionately effects renters [21], populations who are often more historically disadvantaged and lower income. Therefore, this research project aims to not only to impose engineering recommendations, but also identify city planning strategies that reduce energy burdens, promoting equitable thermal comfort and energy consumption across neighborhoods, as done in similar works including [203],[204],[205],[22].

Urban Building Energy Modeling (UBEM) is a technique used to provide insights into the energy consumption patterns of buildings and energy dynamics in urban areas, capturing the interactions between buildings and their urban contexts (including shading and urban microclimate conditions) [24]. Using bottom-up or top-down methods, while referencing building operational constraint definitions, UBEM models provide insights into building energy consumption through formulating hourly energy demand predictions for a collection of urban buildings [146]. Model inputs include building geometric data (ex. shape, size, height), building non-geometric data (occupancy loadings), and weather data [206].

UBEM climate and environmental data inputs are commonly comprised of hourly weather observations that are sampled over a 12-meteorological month calendar. These are often in Typical Meteorological Year (TMY) files [65], that are made compatible with energy plus engines in energy plus weather, epw, format. Although applauded for their high temporal granularity (hourly), these files source historically aggregated climate and weather data measurements [65], and are often collected in locations that are absent of urban morphology, such as airports. Due to their bias in sampling location, TMY datasets often fail to include urban microclimate parameters that are known to influence building energy demand in cities. This discrepancy often leads to building energy modeling inaccuracies, as these weather and climate data inputs are not representative of the environmental conditions experienced directly at building sites. In addition to TMY datasets, UBEM microclimate inputs has been substitutionally sourced from simulation-based or observational-based datasets [24], which both differ in their capacity to accurately contextualize microclimate conditions [16].

The reliability of weather data for building energy applications depends on the spatial and temporal scale of weather file, and the compatibility of the weather file with simulation-based building energy management platforms [16]. In this context, important tradeoffs exist between simulation-based and observation-based climate data sources. Simulation-based approaches represent local climate conditions using numerical and physics-based models [37] which can be directly coupled with building energy simulations [57]. However, generating high-resolution microclimate datasets through simulation is computationally intensive [26] and because they are synthetic in nature, may not be representative of actual climate conditions. In contract, measured values, such as in-situ and/or remote sensing products, are used for observational based climate data inputs. However, these are constrained by the resource requirements necessary to sample at high spatial-temporal frequencies, by challenges inherent to data-collection processes [26], and by the complex data schemas needed to couple them with building energy modeling tools.

1.1. Relevant Literature

Various UBEM studies have integrated more urban specific and microclimate data into building energy models. To evaluate differences in performance between EPW and suburban weather station files in Nanjing, China, Yang et al. simulate microclimate conditions using Urban Weather Generator, a platform that accounts for the effects of urban morphology on local climate by transforming a typical meteorological year (TMY) weather file through urban canopy parameterization techniques [207]. The authors confirm suburban weather station measurements to be more effective than EPW files in building energy studies, although find non-simulated local weather data to be the most precise, as simulation-based datasets may not be fully representative of actual conditions [8]. Furthermore, Hong et al. access the differences in building performance due to urban microclimate conditions for four DOE prototype buildings in San Francisco, CA. Using 27 in-situ weather measurements and CityBES's urban microclimate mapping feature, they compare building performances across different city conditions [10]; the study finds variations up to 11°C due to urban microclimate conditions—contributing to over 100% difference in annual heating energy use and 65% difference in annual cooling energy use across different stations. These articles both validate the use of point-source in-situ weather observations to conduct building energy demand studies and advocate for similar studies to be conducted throughout more climate zones and regions.

Given recent technological developments, new opportunities exist for integrating earth-observational based microclimate datasets into UBEMs [16], [82], [83]. With the emerging availability of satellite-derived datasets, new studies have leveraged these data products to examine the effect of urban morphology on building energy consumption. For instance, in Montreal, Canada, Katal et al. validate the use of open-source datasets such as OSM/ Microsoft Buildings and Google Earth building height databases for use in UBEM research to generate 3D models of buildings [132]. In addition, He et al. use satellite data for urban building energy modeling reconstruction, to integrate more accurate building geometric data such as building heights into their analysis [36]. Finally, Wang et al extract key building features, such as building footprints, building types, and heights, from a single satellite image – advocating for future use of these datatypes to streamline urban data collection for UBEM [208].

However, microclimate earth-observation and remote-sensing datasets remain underutilized in urban building energy consumption studies, highlighting significant opportunities for deeper integration [209],[210],[16]. Dougherty and Jain's data-driven UBEM research [83],[82] demonstrates the value of incorporating reanalysis products, satellite observations, and LiDAR-based microclimate data into building energy performance models. Additionally, Yu et al. apply machine-learning models on remote-sensing and census datasets to map spatial–temporal disparities and pinpoint energy-vulnerable communities across U.S. census tracts [211]. Collectively, these efforts demonstrate promise for combining remote sensing microclimate data into UBEM to evaluate city-wide energy demands and distributions [212].

1.2. Study Gaps and Opportunities

Satellite microclimate data has been integrated into UBEMs using top-down statistical methods throughout various studies [210]. For example [34], [82], [83], [211] use satellite data to integrate microclimate effects into energy predictions given real measured electricity datasets from city energy benchmarking and census databases. However, to our knowledge, no study has connected satellite-derived microclimate data products to simulation-based building energy models or weather file inputs. Accordingly, our study introduces a bottom-up, hybrid modeling framework to assess the utility of integrating satellite-derived microclimate products into simulation-based building energy models, while also examining equitable infrastructure solutions.

Building on our previous work, [16], [34] which incorporates earth observational microclimate datasets into UBEMs through a data-driven framework. In our present work, we augment physics-based building energy models with satellite-derived microclimate datasets, referencing a workflow that combats challenges with dataset coupling, differing temporal-spatial scales, computational demands, and the spatial availability of EPW compatible TMY data. Our hybrid simulation-based and data-driven framework creates a heatmap that characterizes the energy consumption of a typical residential building throughout different urban contexts in Los Angeles, CA, USA.

First, using the building characteristics of a Department of Energy (DOE) archetype residential building, we leverage the hourly temporal capability of TMYx epw weather files– to simulate annual building energy consumption at all TMYx measurement locations across Los Angeles (LA), CA. Then, we reference earth-observational microclimate and GIS-based urban morphology datasets (which include buildings and streets) to create an urban context model of the city. In phase two of our project, we blend these two products together, using machine learning, to predict annual residential building energy consumption across LA. Moreover, we create an urban energy consumption heatmap that depicts the extent to which microclimate and urban morphological characteristics effect residential building energy consumption. Finally, we compare the urban energy consumption heatmap with existing city planning data to validate its effectiveness, and with CDC/ATSDR social vulnerability data to determine how urban microclimates relate to city policies and infrastructure equity.

1.3 Primary Objectives:

- Proposes a hybrid framework to integrate earth-observational microclimate and GIS-based urban morphology characteristics into simulation-based building energy models.
- Creates a detailed heatmap of residential building energy consumption across LA that accounts for urban context conditions such as urban microclimates and urban morphology characteristics.
- Determines the regions where residential building energy demands are elevated due to urban context conditions and relates these results to zoning characteristics and population vulnerability, thereby informing more equitable infrastructure planning decisions.

2. Methodology

This study employs a two-phase methodology to enhance urban building energy modeling by integrating simulation-based building energy modeling outputs with earth observational and GIS-based microclimate data. *Phase 1* simulates residential building energy consumption using EnergyPlus. The simulations reference DOE archetype models and typical meteorological year (.epw) weather files, with each respective building modeled at designated.epw sampling locations across Los Angeles, CA. These simulations provide baseline energy demand estimates under standardized climate conditions. *Phase 2* constructs a comprehensive urban context model that combines spatial datasets which represent urban morphology (e.g. street cover, building density) and microclimate conditions (e.g., land surface temperature, tree cover). This urban context model is then coupled with Phase 1 simulation outputs using machine learning approaches, powered by Random Forest, CatBoost, Decision Tree, and Multiple Linear Regression algorithms. The final model output predicts energy consumption across varying urban conditions, generating city-wide heatmaps of residential building energy demand.

Finally, the predicted energy consumption patterns are analyzed alongside city zoning and vulnerability data. This model verification step reveals correlations between urban form, microclimate variability, and energy burden, offering insights into spatial and social inequities of energy usage across the city.

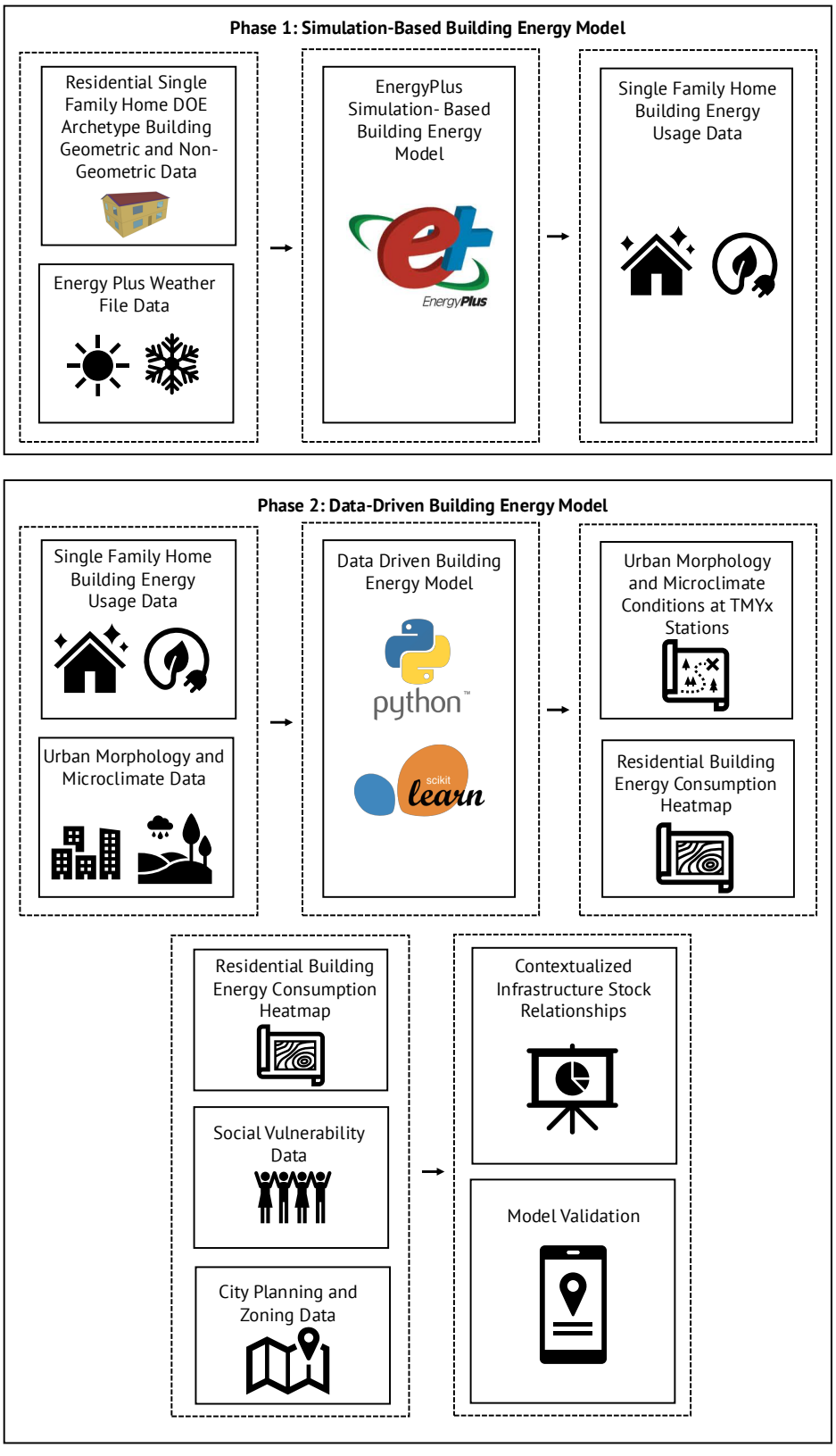

Figure 27: Methodological diagram for the hybrid study

2.1 Phase 1: Simulation-Based Building Energy Model

Phase 1 of these research combines residential building archetype characteristics (.idf) files and respective energy plus weather (.epw) files. It then references EnergyPlus-23-1-0 and the Python eppy library to automate simulation, modeling the energy consumption for a DOE prototype single-family home. Final annual building energy consumption metrics and simulation results are stored alongside their corresponding (.epw) file building site locations and used as data inputs in *Phase 2* of this research.

All EnergyPlus simulations were run using the default gas furnace system configuration, internal load assumptions, occupancy schedules, and infiltration rates, as defined in the DOE Single-Family Residential prototype for climate zone 3C. These defaults include hourly schedules for lighting, plug loads, domestic hot water, and thermostat setpoints, as well as envelope thermal properties and mechanical system efficiencies. No changes were made to the gas furnace schedules or internal gains beyond what is specified in the DOE prototype, ensuring consistent baselines across all weather files.

2.2 Phase 2: Data-Driven Urban Building Energy Model

Phase 2 of this project is separated into three parts. First, it combines urban morphology and urban microclimate datasets to create a geographical urban context map over LA city limits. Then, it couples the urban context model with the *Phase 1* simulation-based outputs using machine learning techniques, to thereby predict residential building energy consumption throughout different urban context conditions. Finally, it examines model results alongside city planning and social vulnerability data, to ultimately uncover relationships between urban contexts, microclimate conditions, infrastructure representation, and residential energy burden.

2.2.1 Urban Context Map

Urban morphology and urban microclimate data products, outlined in Figure 28 and are combined to create an urban context map. To better quantify the spatial influence of each data layer, concentric buffer zones were generated around the EPW sampling locations (defined by longitude and latitude), spanning distances from 10 m to 1000 m in 30 m increments. Values from each data product were then sampled and calculated at all respective buffer radii to evaluate scale-dependent contributions.

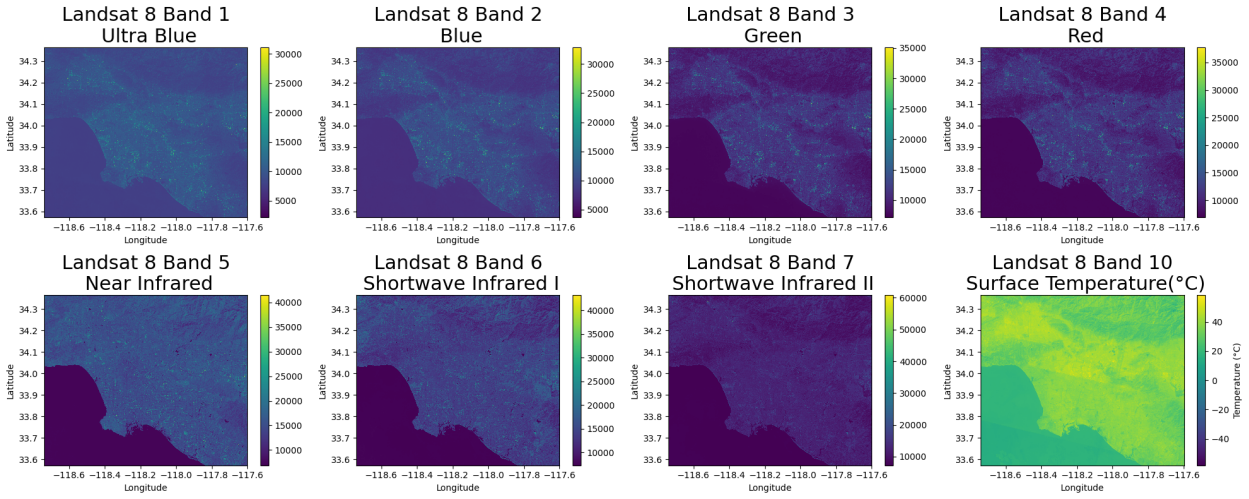

Figure 28: Urban context map final model data layers

2.2.2 Combined Urban Context Map and Energy Consumption Predictive Model

The urban context model and the simulation-based building energy model outputs from *Phase 1* of the project were combined into a single dataset. This dataset includes the location of each building (epw sampling site), building energy use intensity (EUI), and the urban morphological and microclimate data layers sampled at each respective buffer length.

2.2.2.1 Data Buffering

To perform predictive analysis, it is first of interest to identify the best buffer radius for each respective data parameter. This method draws from the Land Use Regression (LUR) modeling framework, which identifies the optimal spatial scale where predictor variables most effectively explain variation in the target outcome [213] and is used to characterize local urban context effects. To perform this step, the Pearson correlation coefficient between annual energy consumption (EUI) and each environmental covariate (e.g., tree cover or land surface temperature) was examined at all respective buffer lengths. Then, the optimal buffer length was chosen based on its capacity to maximize the absolute value of correlation.

2.2.2.2 Feature Selection

The environmental datasets used in this study exhibit strong spatial and temporal autocorrelation, leading to significant challenges in data analysis due to proximity concerns and sampling of closely related features. Therefore, feature selection is employed to decrease feature redundancy and model dimensionality. This process involved evaluating the Pearson correlation coefficient to identify relationships between energy usage and removing derived synthetic data products from analysis to reduce feature redundancy. A refined subset of 8 features was selected, ensuring that the model remained interpretable and statistically robust.

2.2.2.3 Model Prediction

To better access the impact of microclimates on residential building energy consumption, energy use intensity (EUI) is predicted with environmental and urban morphological covariates sampled at their optimal buffer lengths. Training and testing data are sampled at the same 80% and 20% split across all machine learning algorithms. Then, five-fold cross validation (GridSearchCV in the python sklearn library) was used to hypertune model parameters. Both regression-based (Multiple Linear Regression) and tree-based methods (Decision Tree, Random Forest, and CatBoost) are evaluated in this study, having been identified effective in [26], [60], [149], [130]. Model performance is evaluated using two metrics [179]: Root Mean Square Error (RMSE) which measures the standard deviation of model residuals, and the coefficient of determination (R^2), which indicates the goodness of fit of the model.

To visualize energy consumption across heterogeneous urban conditions, the hypertuned models were run to generate residential building EUI heatmaps over a 500m × 500 m spatial grid of the city. For each grid cell, predictor variables were sampled at the cell centroid.

2.3 Model Validation and City Planning Insights

To assess the performance of the building energy model, predicted outputs were compared against existing city planning zoning datasets. In this context, zoning classifications are used as a proxy for urbanization intensity and associated urban heat island effects. We hypothesize that areas designated for higher-density development will exhibit elevated temperatures and, consequently, greater total energy consumption relative to lower-density zones. It should be noted that our city planning zoning dataset was not trained off of earth observational datasets and is thus independent of our models, acting as an appropriate resource for verification.

2.4 Socioeconomic Relationships

To better characterize the energy burden associated with urban microclimates, socioeconomic vulnerability data, sampled at the census tract level, was compared to the residential building energy consumption heatmap. From this, aggregated average consumption levels were calculated across census tract. Then, relationships between residential energy consumption and vulnerability parameters were examined. This analysis helps identify communities that may face disproportionate energy burdens due to their urban context and microclimate conditions.

3. Data

Data inputs were collected for both *Phase 1* and *Phase 2* of the augmented study. *Phase 1* inputs for the simulation-based building energy model include building geometric (.idf), building non-geometric (.idf), and energy plus weather file (.epw) data.

Phase 2 data inputs include earth observational urban microclimate and GIS-based urban morphology datasets, which are compiled to create an urban context map of the city. The urban morphology and microclimate datasets are combined with the simulation-based annual energy consumption outputs from *Phase 1* —and used to predict annual building energy consumption across LA using machine learning techniques. Social vulnerability and city planning data is integrated into the study to determine if energy infrastructure is equitably distributed across different urban neighborhoods.

3.1 Building geometric and nongeometric data

To better understand the impacts of urban contexts and microclimates on residential building energy consumption, DOE prototype residential buildings are referenced [214]. That is, the *US+SF+CZ3C+gasfurnace+crawlspace+IECC_2024.idf* file, which describes building characteristics for a single-family home in the Marine (3C) climate zone, was used to define both geometric and non-geometric building inputs in the EnergyPlus 23-01-01 simulation-based model in *Phase 1* of this study. The baseline energy balance equations referenced in this study are documented in the EnergyPlus Engineering Reference Manual (version 23-01-01 [215].

To enable automated execution of the .idf file using the eppy library in Python, several modifications were required. First, ground heat transfer objects were removed and second, ['BUILDINGSURFACE:DETAILED'] Outside_Boundary_Conditions were changed from 'GroundSlabPreprocessorAverage' to 'Ground'.

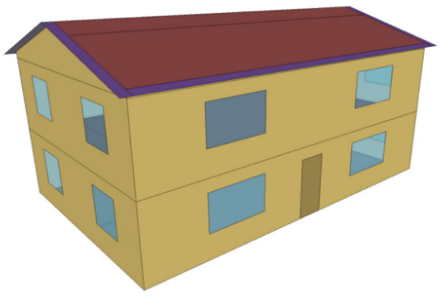

Figure 29: Single Family Home DOE Protype Building, figure from [216]

3.2 Meteorological and Urban Context Data

Two different meteorological and urban context datasets were compiled. The first, outlined in section 3.2.1, comprises of (.epw) files collected throughout LA city limits, sourced from TMYx observations using diyepw, a python package made to facilitate (.epw) file retrieval and generation [183]. These files were used as inputs in the simulation-based building energy model during *Phase 1* of this research. The second dataset, outlined in section 3.2.2 comprises of GIS-based urban morphology and urban microclimate data products, which are used to create an urban context map of the city, and as inputs into the data-driven building energy model during *Phase 2* of this research.

3.2.1 Energy Plus Weather File Climate Data

To integrate weather files into EnergyPlus 23-01-01 software, the weather files must be provided in (.epw) format [166]. Accordingly, all available EnergyPlus weather (.epw) files within the Los Angeles County study-area bounding box were compiled for use as climate inputs in the simulation-based building energy model. A total of 19 Typical Meteorological Year (TMYx) [3] weather files were obtained from Climate.OneBuilding.org [217] and incorporated into the analysis.

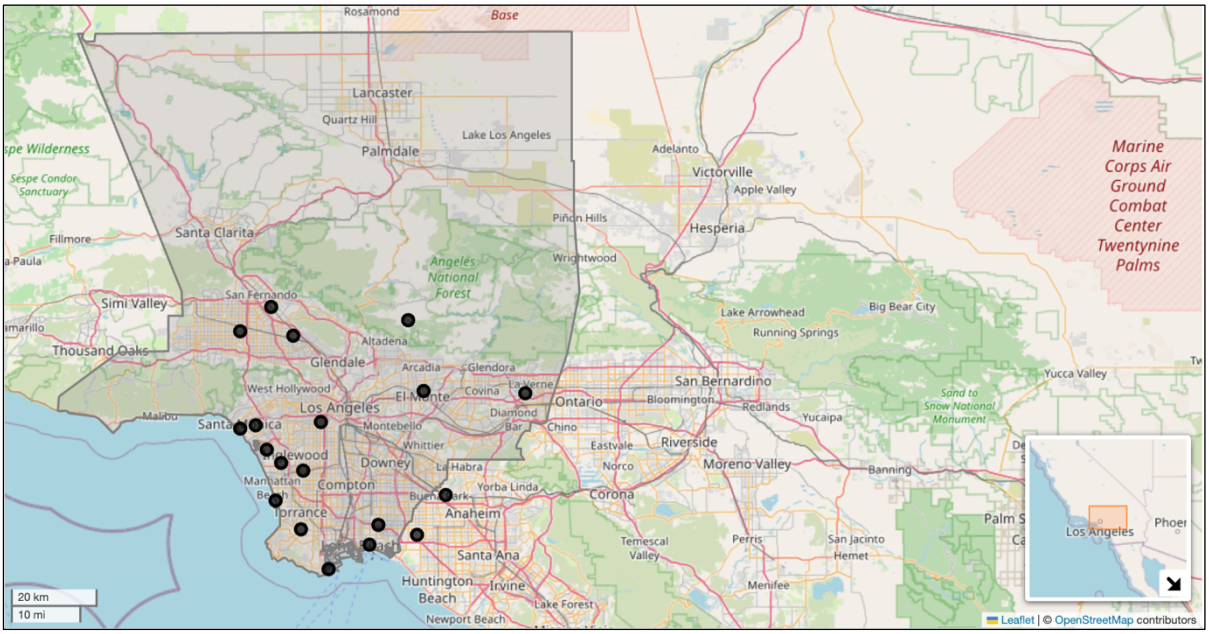

Figure 30: Map of TMYx epw compatible sampling locations (black) in Los Angeles County, CA, USA (grey)

3.2.2 Urban Morphology and Urban Microclimate Data

Different urban morphology and microclimate data products were collected and combined for *Phase 2* of this study. These were layered to create an urban context map of Los Angeles, providing geospatial insights into urban properties, layouts, and forms. The datasets compiled, which are opensource and available in all US cities, are outlined in Table 10 and comprise of high spatial resolution meteorological, environmental, and physical characteristics.

Table 10: Urban morphology and microclimate dataset layers

Data Category	Dataset	Frequency	Duration	Resolution	Attributes	Data Source
Meteorological Characteristics	USGS LANDSAT 8 Level 2, Collection 2, Tier 1	Annual	Averaged across all 2023 instances	30 m	Ultra Blue, Blue, Green, Red, Near Infrared, Shortwave Infrared I, Shortwave Infrared II Atmospheric Transmittance, Downwelled Radiance, Upwelled Radiance, Land Surface Temperature	Google Earth Engine [97]
Environmental Characteristics	NASA DEM	NA	2000 release	30 m	Elevation (m)	Google Earth Engine [96]
Environmental Characteristics	USFS Tree Canopy Coverage	NA	2021 release	30 m	Tree Canopy Cover (%)	Google Earth Engine [95]
Environmental Characteristics	USGS National Landcover Database	NA	2021 release	30 m	Landcover	Google Earth Engine [97]
Physical Characteristics (Streets)	Open Street Map (network= 'drive')	NA	2025 pull	Geometries	Geometry	Open Street Map (python OSMnx api) [218]
Physical Characteristics (Buildings)	Microsoft Buildings Dataset	NA	2020 release	Geometries	Geometry, Area, Centroid	Microsoft Buildings Dataset [106]

3.3 City Zoning Data

To validate the predictive model and inform recommendations for future urban planning strategies, city zoning characteristics and land use classifications within the Los Angeles study area were obtained from the LA Geo Data Hub portal.

Table 11: City zoning and planning datasets

Data Category	Dataset	Frequency	Resolution	Attributes	Data Source
City Planning Characteristics	City of Los Angeles Zoning Polygons	N/A	LA City Limits in Study boundaries	Geometry, Zoning Category	LA Geo Data Hub [219]

3.4 Socioeconomic Data

To assess whether specific communities experience elevated energy burdens as a result of urban context conditions, community characteristics metrics, aggregated at the census tract level, were referenced from the CDC/ATSDR Social Vulnerability Index. This index complies 16 U.S. Census variables from the 5-year American Community Survey into a single metric, combining socioeconomic status, household characteristics, racial and ethnic minority status, housing type, and situation attributes [220].

Table 12: Socioeconomic datasets

Data Category	Dataset	Frequency	Resolution	Attributes	Data Source
Social Characteristics	LA Social Vulnerability Index	Census tract averages	census track, county limits	Geometry, Vulnerability Index	LA Geo Data Hub [221]

4. Results

This section presents the results of predicting residential building energy consumption under varying urban context and microclimate conditions using the augmented modeling framework described in Section 3. We begin by characterizing the urban morphological and geographic properties surrounding each epw weather station location and then compare these averages against average city characteristics. Next, we develop a predictive model to estimate residential building energy consumption across the study region's diverse urban environments. Model outputs are validated using existing city planning datasets and examined in relation to social vulnerability indices.

4.1 Phase 1: Simulation-based energy model

To establish a robust baseline, the study first identifies optimal case study locations— specifically, we look for those containing many EnergyPlus weather file stations and pronounced microclimate variability that is influenced by urban contexts. Among all of the U.S. cities evaluated, including Seattle, San Francisco, New York, and Los Angeles; Los Angeles was identified to meet these criteria best.

To better understand how microclimate conditions effect building energy demand, preliminary energy plus simulations were run with DOE residential building archetype (.idf) files at all given energy plus weather stations inside of the study boundaries. Where the number of independent (.epw) sampling locations, in this case 19, indicates the quantity of buildings simulated and their respective *building locations*. The distribution of energy consumed, simulated for the same building across different city context conditions in Los Angeles, CA, USA are tabulated in Table 4.

Phase 1 results find microclimate variability to influence energy demand across LA by as much as 10.2%, when modeled only at (.epw) sampling locations across the city. Although it is important to note that these results may underestimate true variability, as greater variation may exist in areas not represented by TMYx sampling sites—particularly in highly developed urban regions.

Table 13: Phase 1 building energy simulation summary statistics

City	Longitude Bounding Box	Latitude Bounding Box	Number of .epw sampling locations in bounding box	Annual simulated Energy Use Intensity (EUI) range (kWh/m^2)	Percent difference of simulated energy consumption across study area (%)
Los Angeles, CA, USA	-118.59 to -117.83	33.67 to 34.3	19	36.15 to 39.84	10.2

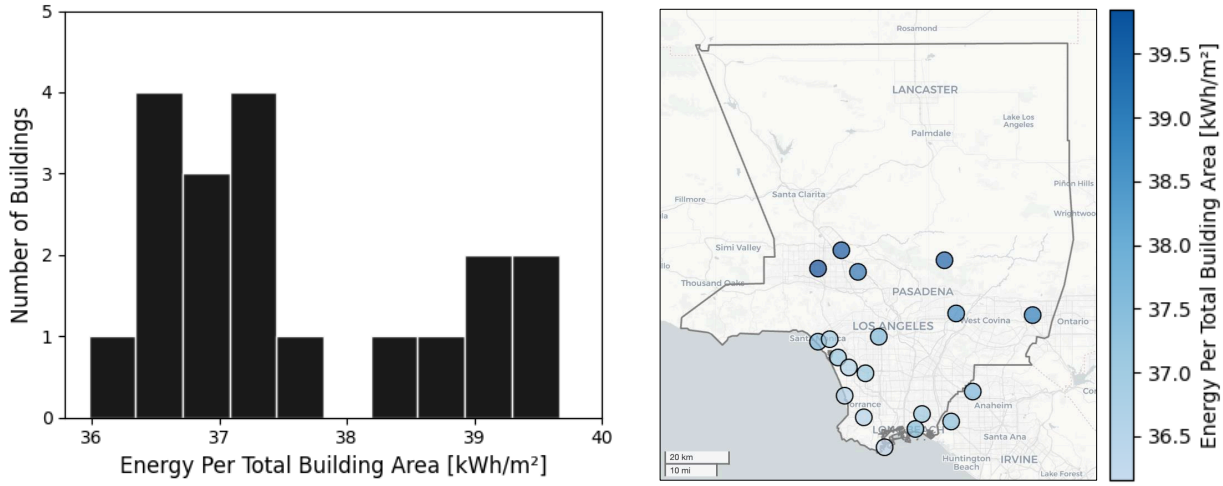

Figure 31: Histogram (left) and Map (right) of Phase 1 Single Family Home EnergyPlus Simulation Outputs

To better interpret total building energy consumption, individual energy end uses are examined separately. The 19 simulated buildings have the same exact interior equipment (10083 kWh), interior lighting (1038 kWh), and exterior lighting loads (211 kWh), encompassing 59 % to 65 % of total building energy demands. However, energy consumption for heating, cooling, fans, and water systems differ respectively for each building, as each is simulated with different epw weather file inputs. Building energy cooling demands range from 5.7% to 9.4% of total building energy consumption, with the highest demands at building 19 in the valley (see *Figure 32*), and the lowest demands at building 9, near the coast. Building energy heating demands are roughly equal in size to the cooling demands, ranging from 4.2 % to 11.2 % of total building energy consumption. The highest heating demands are at building 18, in the mountains, and the lowest heating demands are at building 1, near the coast.

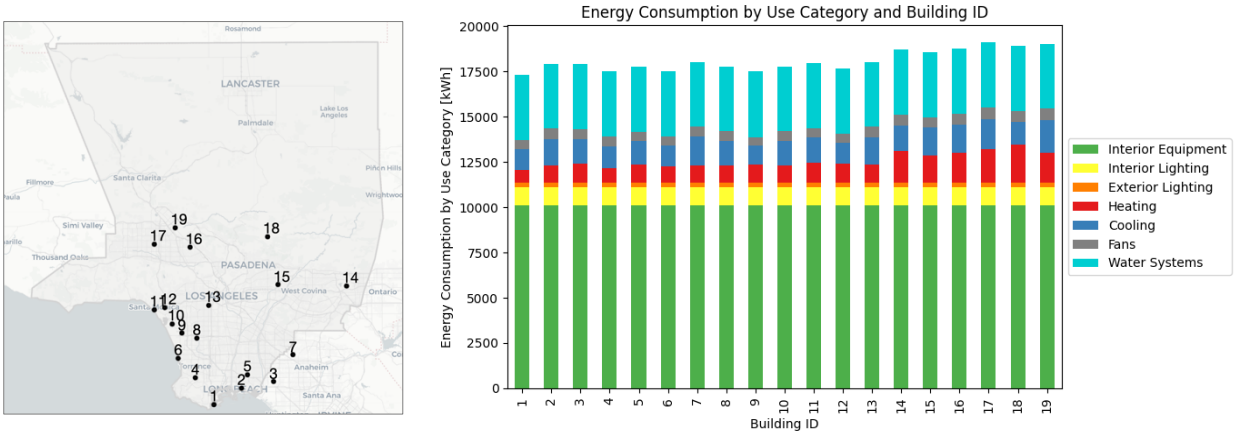

Figure 32: Reference map of buildings, ordered by latitude (left) and Bar Plots of EnergyPlus Simulation Results: Energy Consumption by End Use Category

4.2 Urban Context Model Mapping

To perform predictive analysis, the optimal buffer radius for each respective data parameter is identified. Then, given this optimal buffer length, data parameters are sampled at each respective TMYx sampling site location. After data wrangling, feature selection is employed to reduce covariates through correlation and VIF techniques. The Atmospheric Transmittance (ST_ATRAN), Downwelled Radiance (ST_DRAD), and Upwelled Radiance (ST_URAD) bands are excluded from predictive analysis due to their noise in

observation and visual streaking across the bands. Further, the elevation band is excluded from the predictive analysis, because it is a well-used parameter throughout the EnergyPlus calculations, and we hoped to only include independent variables in our model [222]. The building count parameter, derived from the Microsoft buildings dataset, the USFS tree canopy coverage dataset, and the USGS Landcover dataset were also excluded from the predictive model; as these products themselves are trained on other model covariates and may therefore introduce redundancies and dependencies. Additionally, the street length variable is excluded from the predictive model due to its discrepancy in representation for training datasets parameter characterization—as street length mean values across the TMYx stations were reported 21000000% lower than the Los Angeles average, which is further discussed in section 4.3. A description of final predictive model attributes, having undergone buffer size identification and feature selection, are displayed in Table 14.

Table 14: Description of model covariates and optimal buffer sizes

Category	Variable (GEE Notation)	Buffer Size	Unit	Source
Energy Characteristics	Residential Single Family Home Building Energy Consumption	NA	kWh/m^2	Phase 1 of research
Physical Characteristics	Streets – length *	370	-	Open Street Map (python OSMnx api) [218]
	Buildings – centroid *	790	# of	Microsoft Buildings Dataset [106]
Environmental Characteristics	Landcover *	280	-	USGS National Landcover Database 2021 Release [97]
	Tree Canopy Cover *	190	%	USFS Tree Canopy Cover v2021-4 [95]
	Digital Elevation *	130	m	NASA DEM [96]
	Band 1, Ultra Blue (SR_B1)	40	-	USGS LANDSAT 8 Level 2, Collection 2, Tier 1 [182]
	Band 2, Blue (SR_B2)	40	-	
	Band 3, Green (SR_B3)	40	-	
	Band 4, Red (SR_B4)	40	-	
	Band 5, Near Infrared (SR_B5)	850	-	
	Band 6, Shortwave Infrared 1 (SR_B6)	70	-	
	Band 7, Shortwave Infrared 2 (SR_B7)	70	-	
	Band 10, Surface Temperature (ST_B10)	250	C	
	Atmospheric Transmittance (ST_ATRAN)*	970	-	
	Downwelled Radiance (ST_DRAD)*	970	$W/(m_2 \times sr \times um)/DN$	
	Upwelled Radiance (ST_URAD) *	970	$W/(m_2 \times sr \times um)/DN$	
Uncertainty Quantification	Thermal Band Converted to Radiance (ST_TRAD) *	-	$W/(m_2 \times sr \times um)/DN$	USGS LANDSAT 8 Level 2, Collection 2, Tier 1 [182]
	Pixel Distance to Cloud (ST_CDIST) *		km	
	Band 10 Emissivity (ST_EMIS) *		-	
	Emissivity Standard deviation (ST_EMSDIS) *		-	
	Surface Temperature Band Uncertainty (ST_QA) *		K	
	Pixel Quality Attributes (QA_PIXEL) *		-	

* Parameters that are not included in the final predictive models

4.3 Comparison of conditions in the city and at sampling locations

To validate the use of TMYx files, it is first necessary to determine if the environmental and morphological conditions at (.epw) sampling locations are representative of the actual city conditions. Therefore, the urban

context properties at each TMYx sampling location are calculated and then compared against the average city characteristics (excluding open-water and waterways); these averages are reported in Table 15.

When comparing averages between the average TMYx sampling site and mean LA condition, all Landsat parameter differences fall within 10%. This relationship highlights this datasets competency for model generalization, because these averages are in line with those recorded in the model training dataset. However, differences between street length, tree canopy coverage, elevation, and building count parameters are more pronounced; meaning that for these parameters, the TMYx file sampling site characteristics are not representative of the actual values observed throughout the city. This finding calls for future TMYx stations in Los Angeles to be placed in city regions with greater densities of streets, trees, and buildings.

Table 15: Differences in parameter averages between TMYx sampling site locations and LA county

Feature (buffer length (m))	TMYx Location Mean	Los Angeles Mean	Difference (%)
Street Length (370)	0.014480	3107	21000000
Building Count (790)	560	813	45
Tree Canopy Coverage (190)	2.5	7.1	185
Mode Landcover (280)	24	23	–
Elevation (130)	123	201	64
SR B1 (40)	10570	9834	-6.96
SR B2 (40)	10980	10227	-6.85
SR B3 (40)	11983	11220	-6.35
SR B4 (40)	12402	11497	-7.29
SR B5 (850)	14990	15414	2.83
SR_B6 (70)	15479	14178	-8.4
SR B7 (70)	13973	12487	-10.6
ST B10 (250)	37.043	37.042	-0.002
ST ATRAN (970)	8257	8168	-1.08
ST DRAD (970)	672	699	4.08
ST URAD (970)	1364	1423	4.29

4.4 Energy use prediction model performance and heatmaps

Single building EnergyPlus simulations are enhanced using earth observational datasets and machine learning algorithms (Random Forest, CatBoost, Decision Tree, and Multiple Linear Regression) to create predictive heat maps of residential building energy usage, augmented with urban microclimate conditions, across Los Angeles, USA. Cross validation is simultaneously employed, using the Grid Search CV in Python, to tune model hyperparameters and to ensure robust and generalizable model fits. Table 16 outlines the modeling algorithms referenced, their hyperparameters, their respective modeling performance, and their reported percent difference in range over the predicted LA modeling grid.

Table 16: Performance of different building energy consumption prediction models

Method	Hyper-Parameters	Training R^2	Testing R^2	Training RMSE	Testing RMSE	Percent Difference of range over the LA grid
Random Forest	n_estimators=100, random_state=6, max_depth= 3	0.85	0.56	0.42	0.87	6.3
CatBoost	iterations=30, learning_rate=0.3, depth=3	0.99	0.56	0.11	0.81	8.7

Decision Tree	random_state=6, max_depth=2	0.69	0.61	0.62	0.76	7.4
Multiple Linear Regression	-	0.87	0.35	0.39	1.06	121

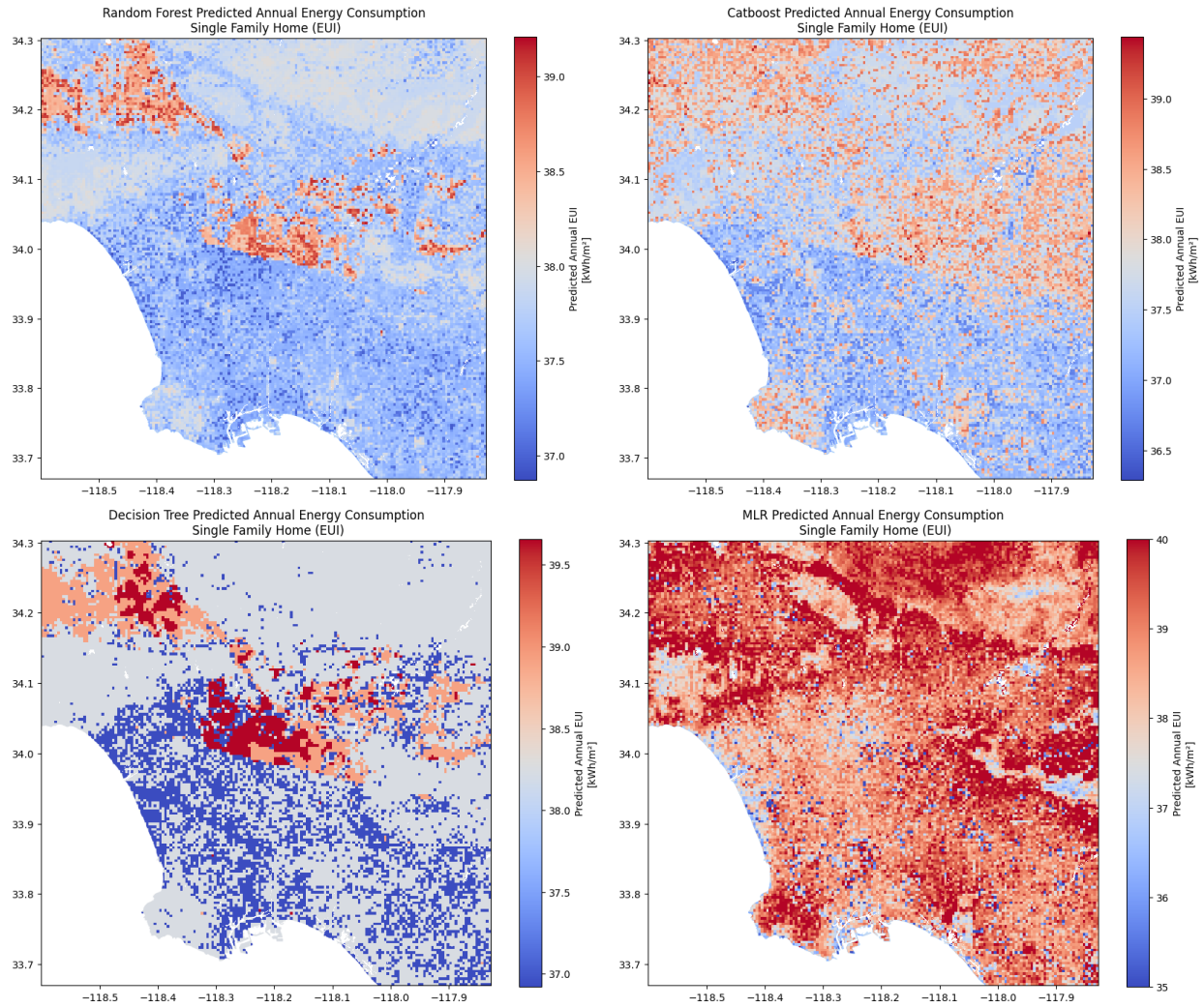

Figure 33: Predicted single family home total energy consumption heatmaps for Random Forest (top-left), CatBoost (top-right), Decision Tree (bottom-left) and MLR (bottom-right) models

Because the machine-learning models were trained on 19 spatially discrete sampling points, some areas of the city fall outside the range of feature values present in the training data. These areas represent extrapolation zones, and their predictions should be interpreted with caution. In contrast, regions with feature values that overlap with those of the training locations represent interpolation zones, where model predictions are expected to be more reliable.

The Random Forest, CatBoost, Decision Tree, and Multiple Linear Regression model heatmaps indicate a 6.3%, 8.7%, 7.4% and 121% respective discrepancy in single family home annual energy usage between the lowest and the highest consuming areas. It should be noted that because our model was trained off data with an overall range discrepancy of 10%, these values should be expected to be around 10%, specifically if data values are interpolated.

In more extreme settings, where city characteristics are not represented by the training dataset, values are then extrapolated and may result in a higher percentage discrepancy. Much larger differences in values are observed for the linear model, this is due to its poor generalizability at extreme conditions, for instance the model has difficulties scaling for data that it did not see in its training stage, such as the higher land surface temperature observations throughout the model testing stage compared to the training stage.

4.5 Feature Importance

Feature importance analysis finds the Landsat 8 land surface temperature band (ST_B10) to be the most influential earth observational microclimate parameter for building energy consumption prediction in the top performing Random Forest, CatBoost, Decision Tree, and Multiple Linear Regression models. Significant land surface temperature band contributions are consistent with the results tabulated in a building energy and microclimate studies in New York City [82] and Seattle [34]. These results imply that land surface temperature products should be further investigated for reference in urban scale energy usage studies, as this remotely sensed parameter is open source, measured at 30m everywhere in the world, and is proven to have direct correlations with building heating and cooling energy demands.

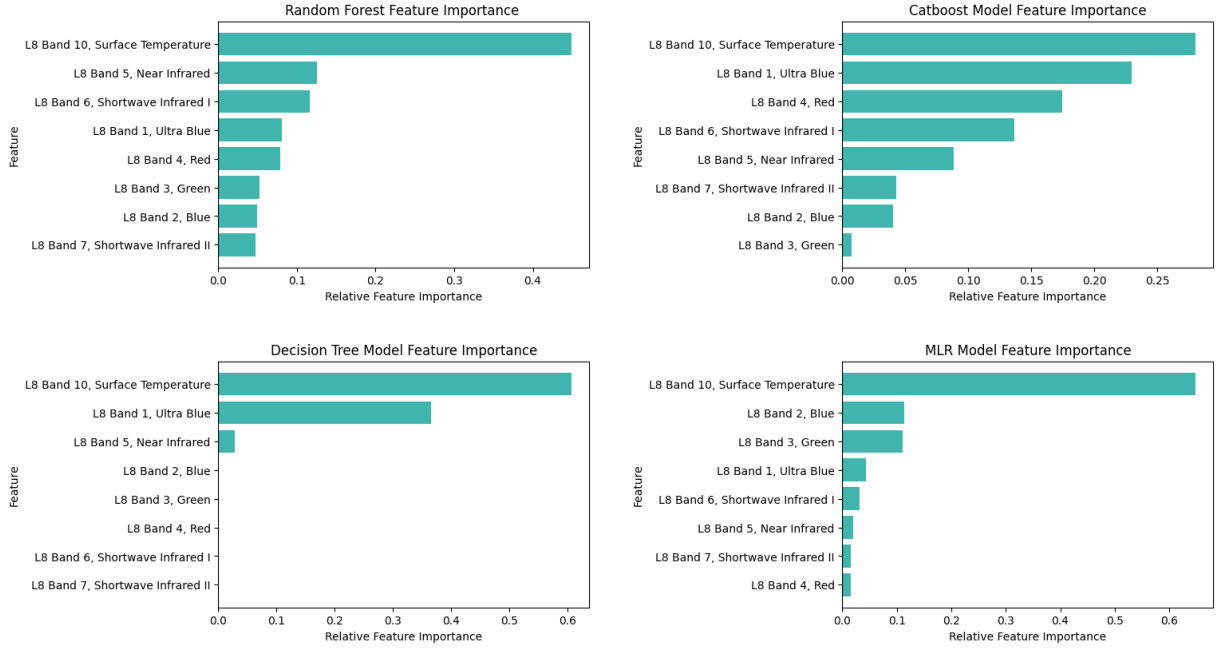

Figure 34: Predictive model variable importance plots for Random Forest (top-left), CatBoost (top-right), Decision Tree (bottom-left) and MLR (bottom-right) models

4.6 City planning insights and model verification

To better understand the impacts of urban microclimates on city energy use patterns, we compared our residential building energy consumption predictions with city mapping and zoning data, collected from the Los Angeles City Planning Department [60].

In Los Angeles, CA, the highest peak electricity demands take place in summer months [223], when elevated temperatures, which are amplified by the urban heat island effect, drive increased building cooling loads. In our analysis, areas designated by the city as higher-intensity development regions, such as parking, industrial, and industrial-mixed zones, show the highest average building energy consumption across all model algorithm predictions, as illustrated in *Figure 35*. Whereas, open space, agricultural, and public facility land areas report lower residential building energy consumption predictions across all model

algorithms. In other words, homes located in more densely developed areas in Los Angeles consume more energy annually than those situated in less developed, natural areas. This finding is consistent with Urban Heat Island effect literature for semi-arid climates and showcases our model effectiveness in identifying resource use hot spots, having great implications for city energy planning concentration efforts.

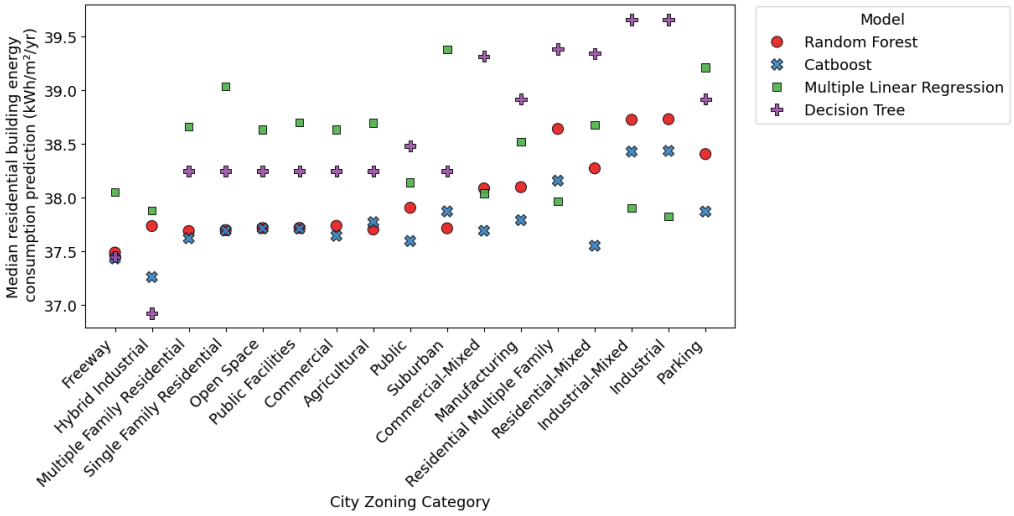

Figure 35: Median residential building energy consumption prediction (annual), aggregated for each Los Angeles city zoning category, showing higher predicted consumption across more industrial and developed zones

4.7 Social vulnerability index correlation analysis

To make sure that infrastructure planning accounts for microclimate variability across neighborhoods, the relationship between single family energy consumption and population vulnerability is examined. That is, the predictive energy consumption heatmaps are merged with social vulnerability index data, which is aggregated at the census track level across LA county, as shown in *Figure 36* and then compared in Figure 37.

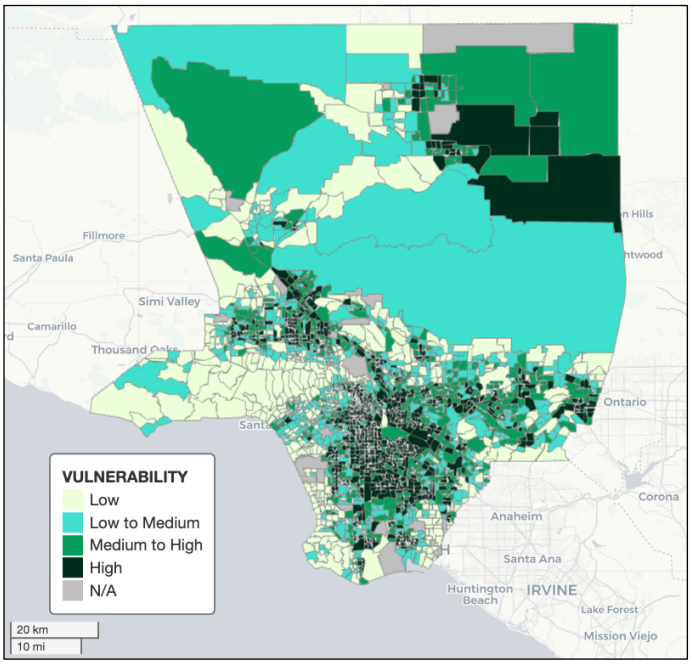

Figure 36: Vulnerability index category of census tracts in LA, data from [220]

Next, energy consumption averages are computed across every census tract and then categorized respective to their census tract vulnerability index, as shown in a boxplot in *Figure 37*. These aggregations show surprisingly similar average single family home EUIs across all predicted models and vulnerability indices.

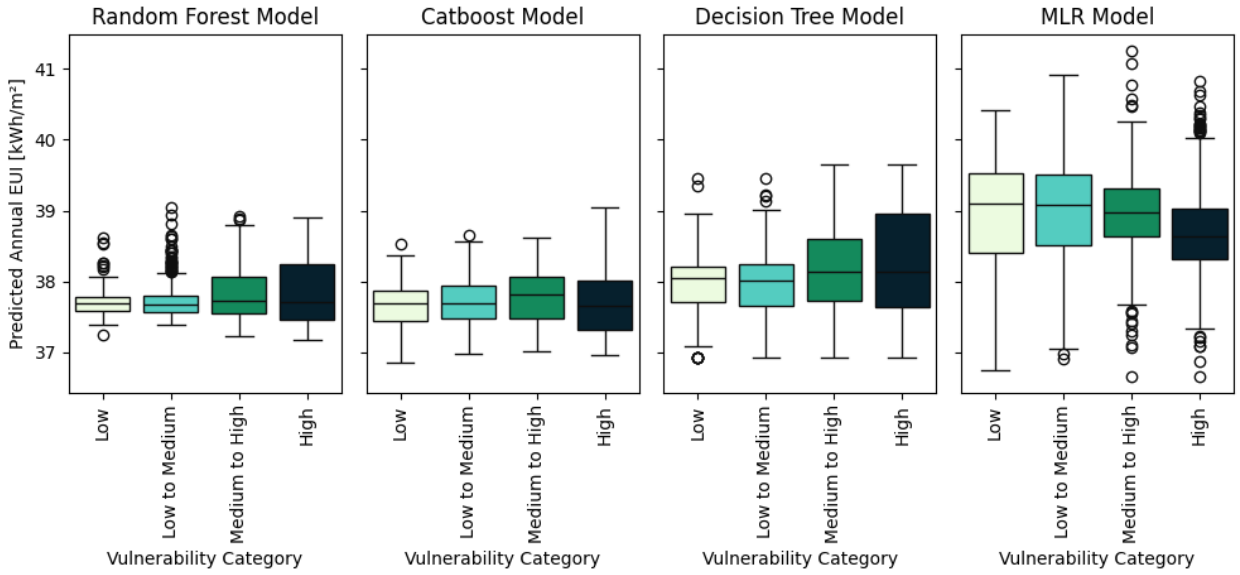

Figure 37: Distribution of predicted single-family home annual EUI by model and vulnerability index

However, we find discrepancies between the distribution of TMYx sampling sites across vulnerability categories in LA. That is, 35% of all census tracts in LA are categorized as high vulnerability, however, only 10% of the TMYx sampling site locations are in these regions, as highlighted in Table 17. Moreover, while 18% of census tracts are in low venerability areas, 26% of TMYx sensing takes place in these locations. This suggests that our training dataset may not be fully representative of the actual conditions observed in high density and high vulnerability areas – limiting our analysis and predictive models to fully contextualize extreme microclimate environments. Therefore, we call for future placement of TMYx sampling sites to be placed in high vulnerability regions, specifically high-density urban areas – which have been documented to experience different environmental conditions than their rural counterparts.

Table 17: Percentage of population, census tracts, and TMYx locations in each vulnerability index category

Vulnerability Index Category	Percentage of land area (%)	Percentage of census tracts in category (%)	Percentage of TMYx sampling locations in category (%)
High	13	35	11
Medium to High	22	25	16
Low to Medium	38	19	21
Low	22	18	26

N/A*	3	1	26 *

* Three sites are in N/A zones, and two sites are out of LA County limits (with no data availability)

5. Discussion

In this research, we create a framework that augments simulation-based building energy models with earth observational microclimate data products using machine learning predictions. We examine the microclimate and urban conditions surrounding current TMYx sampling locations and compare them to the average city conditions. When doing so, we find large discrepancies between the streets, buildings, and trees surrounding the stations compared to the Los Angeles average. This finding has a direct relationship to building energy demand predictions, as it tells us that we need to place future TMYx stations near more streets and buildings if we want to account for the average climate observed within urban areas. Along with this discrepancy, we find that there is less representation of TMYx sensing stations in high vulnerability regions, calling for additional TMYx stations to be placed in these regions to better contextualize how urban microclimate conditions effect household building energy consumption and corresponding energy burdens.

Additionally, we find the Landsat 8 thermal band 10, or land surface temperature products, to be highly important for predicting building energy consumption with microclimate effects. This finding is important because it validates these datasets for including microclimate effects within household building energy consumption predication tasks. Furthermore, it recommends future research to include earth observational microclimate datasets into their urban building energy models – as these features are widely available open source, sensed at high spatial resolution across the entire world, and can be easily pulled into projects with limited GIS and python experience [16], [34].

For locations in borderline mediterranean and semi-arid climates and those similar in climate to Los Angeles, we suggest land use managers, property developers, and city planners to reference local historically aggregated land surface temperature measurements as they perform building heating and cooling load calculations. That is, to incorporate urban microclimate conditions into energy models at the building scale, specifically for new building construction, we recommend that development teams compare the Landsat thermal band 10 observations between their development site of interest and the nearest EPW weather file sampling location, for which their EnergyPlus model is configured to. This additional analysis step provides a proxy to and directional estimate for including urban microclimate conditions into building energy consumption calculations. Although, because every city climate is different, this will only give an approximation of how well your energy plus weather file will extrapolate to your location. If these compared measurements differ greatly, but in the same location, we recommend conducting additional analysis, such as referencing more local data, to access these discrepancies before regarding the EnergyPlus simulations as a ground truth estimates.

There are, however, some noise and biases for using remote sensing and earth observational data in building energy consumption modeling. That is, the reliability of earth observational data can be compromised from cloud cover, sensor drift, and retrieval algorithms, calling for researchers to explicitly address the accuracy of their measurements before model implementation. Various other earth observational and remote sensing indices can be explored for the inclusion of microclimate effects with a few for energy applications included in the following publications [16],[224],[225].

The proposed framework has obvious biases and limitations that should be addressed in future works. One being that because the data-driven model is trained upon (.epw) sampling locations, the model is dependent on the spatial resolution of these stations — therefore limited quantities of these stations over study boundaries could result in underrepresentation of city characteristics, which ultimately hinders model performance. Additionally, the framework is limited by the urban contexts of the that existing (.epw) files.

For instance, the model's generalizability is constrained by the data that it is trained upon— therefore the model will decrease in reliability when needed to extrapolate or predict actual conditions which are outliers from, or not contained in, the training dataset. Moreover, the study framework is optimized for regions where urban context and morphology are hypothesized to be the main factor influencing urban microclimate conditions. That is, this framework has difficulties describing mesoclimate and macroclimate influences – and thus can only be used to predict building energy consumption at smaller scales (ex. city) instead of larger scales (ex. county, state, country). Lastly, in this framework, we are predicting synthetic data. Therefore, the prediction of actual energy consumption data, collected from utilities, and overlayed with microclimate data layers, may give better insights for proper management of energy infrastructure distribution and placement.

Throughout this project, we find the reliability of a weather data file for building energy applications to be dependent of the spatial scale of weather file, the temporal scale of the weather file, and the compatibility of the weather file with simulation-based building energy management platforms. A key limitation of our proposed work is that while it integrates both physics-based simulation and data-driven machine learning techniques, it does not enable direct incorporation of microclimate data into simulation-based building energy models. Earth-observational microclimate datasets lack sufficient temporal granularity whereas conventional simulation models like EnergyPlus operate on an hourly basis. Therefore, while the framework provides valuable insights into urban energy dynamics, it does not fully resolve the challenge of integrating satellite-derived microclimate datasets into physics-based building energy models.

To address the spatial precision of weather data in building energy models, an obvious solution would be to create a schema that augments earth observational microclimate products into (.epw) files directly. However, at the time, this opportunity faces major limitations; these being that earth observational datasets are not temporally precise in their microclimate parameter measurements -- leading to challenges in timestamp aggregation with (.epw) files. Future research directions should be taken to investigate the potential of blending EO datasets directly into energy plus weather (.epw) files, including the exploration of more temporally combatable datasets that measure TMYx parameters and respective downscaling opportunities.

6. Conclusion

In this study, we augment EnergyPlus building simulations with satellite-derived microclimate datasets, ultimately creating heat maps of residential building energy usage that are corrected for urban climate conditions across Los Angeles, CA. We confirm that EnergyPlus Weather (EPW) files, which are commonly used in building energy modeling, rely on non–urban-specific weather inputs. That is, In Los Angeles -- the U.S. city with the greatest number of EnergyPlus-compatible weather files, recording stations are placed in locations with lower CDC/ATSDR Social Vulnerability Index scores and with much fewer streets, buildings, and trees compared to the city average. This discovery calls for future placement of TMYx stations to be in regions where populations are more vulnerable as well as near more streets and buildings, in order to accurately account for the weather and climate conditions observed by all populations within urban areas. We also find the Landsat 8 Band 10, land surface temperature product to be a highly important remote sensing parameter for predicting building energy consumption, validating this dataset for future building energy consumption predication tasks. While our proposed framework provides valuable urban energy dynamic insights, it does not, however, fully resolve the challenge of integrating remote sensing datasets into simulation-based building energy modeling tools. Future work in this area is encouraged to explore software pipelines and workflows that directly combine satellite-derived microclimate data into simulation-based building energy consumption weather files and platforms. This discovery would advance the acknowledgement of microclimate conditions in building energy models while still allowing for users to manage modeling across a single simulation-based energy modeling platform such as EnergyPlus. Our study underscores the importance of bridging the existing gaps between

modeled and real-world energy scenarios through incorporating microclimate data into building energy models, aiming to advance sustainable energy-driven improvements at the city level.

Chapter 5: Conclusions and Future Directions

This manuscript accesses the effectiveness of integrating satellite-derived microclimate datasets into urban building energy modeling frameworks. In chapter 2, it outlines the common methods and datasets used for microclimate integration in building energy performance studies, identifying opportunities to address current modeling gaps with earth observational microclimate datasets. In chapter 3, we reference satellite-derived microclimate datasets in a statistical-based urban building energy modeling study to compare different climate and weather dataset input variable schemas. Results show increasing model performance with earth-observational microclimate datasets over traditionally used TMY3 files, thereby validating their efficacy. In chapter 4, we reference earth observational microclimate datasets in an augmented, data-driven and simulation-based, urban building energy modeling framework, gaining better insights to energy consumption patterns for single family buildings across Los Angeles, USA. Finally, in this chapter, we aggregate the findings from the previous chapters, discussing research takeaways and future research directions.

Throughout this dissertation, we find a gap in, and opportunity for, integrating microclimate effects into urban building energy models for more realistic and actionable urban energy insights. Historically, microclimate integration has been investigated using two different climate datasets—each differing in the origination of microclimate data as well as its temporal and spatial precision. Observational-based datasets, compile real climate and weather data measurements to describe microclimate conditions in cities. Whereas simulation-based microclimate datasets reference sparce weather and climate station data and computational and physics-based modeling to derive environmental insights. In chapter 2, we determine the extent to which each of these products abide with data merging, scalability, computation, and in coupling with building energy modeling platforms. We also find that while remote sensing technologies are extensively referenced for building geometric data UBEM inputs, there remains an underexplored potential to integrate these products as UBEM environmental data inputs. To investigate this gap, we employ two complimentary studies that explore satellite-derived microclimate dataset integration in urban building energy models.

Our first study employs a statistical based UBEM, predicting real building energy consumption measurements using three different climate data parameter schemas, with each differing in their origin of microclimate data, and then compares model performance. Throughout the study deployment across Seattle, WA, USA we find the optimal weather and climate dataset to be comprised of earth observational data and spatially interpolated TMY3 data, achieving a 0.16 (from 0.55 to 0.71) increase in testing R^2 over the model that did included no climate data inputs, and a 0.05 (from 0.66 to 0.71) increase in testing R^2, over the model that included only spatially interpolated TMY3 climate data inputs. This finding confirms the importance of integrating microclimate variables into UBEMs and validates the use of earth observational based datasets for data-driven energy prediction studies.

Our data-driven study guided the goals for our second investigation, a hybrid study, which augments physics-based energy modeling outputs for typical residential buildings with earth observational microclimate data using machine learning predictions; thereby generating a predictive heatmap of energy consumption patterns across city Los Angeles, USA city limits. To better understand interpolation and extrapolation of our augmented model, we examine the microclimate and urban conditions surrounding the current TMYx sampling locations and compare these to the average conditions observed across the city. When doing so, we find a fewer streets, buildings, and trees to surround the TMYx stations compared to the Los Angeles average. This finding helps to inform additional TMYx station placement, suggesting that

we must place future TMYx stations near more streets and buildings if we want to account for the average climate observed within urban areas. We also find that current TMYx stations are placed in census tracks designated as lower social vulnerability areas. This discovery calls for future placement of TMYx stations to be in regions where populations are more vulnerable, in order to accurately account for the weather and climate conditions observed by all populations within urban areas.

Feature importance research results are consolidated from all three studies, to inform more actionable insights for future research. Throughout the literature review, we find air temperature to be the primary microclimate variable examined, with traditional parameter expansion including humidity, wind speed, solar radiation, precipitation, elevation, pressure, vegetation, cloud cover, and shading. In our data-driven energy prediction study, we find Landsat 8 downwelled radiance, Landsat 8 Normalized Difference Vegetation Index (NDVI), Landsat 8 Thermal Band 10 (B10), and LiDAR Digital Elevation Model (DEM) bands to be most relevant remote sensing microclimate parameters for energy consumption prediction. Finally, during our augmented study we find the Landsat 8 land surface temperature band (ST_B10) to be the most influential for microclimate aware building energy consumption prediction. Aggregating these results, under both case study instances, land surface temperature satellite products offer the most valuable microclimate information and should be further considered for application when modeling building energy consumption with microclimate effects to offer novel insights.

Throughout this dissertation, we find the reliability of weather data for building energy applications to be dependent of the spatial scale of weather file, the temporal scale of the weather file, and the compatibility of weather file with simulation-based building energy management platforms. For data driven predictive building energy modeling studies, microclimate data coupling is not a concern. Therefore, we encourage researchers to incorporate satellite derived microclimate datasets in their variable schemas, specifically land surface temperature products or the Landsat 8 Band 10, as these products are widely available open source, sensed at high spatial resolution across the entire world, and can be easily pulled into projects with limited GIS and python experience.

However, due to coupling procedures, combining earth observational microclimate datasets into physics-based building energy modeling platforms is much more complex. Although our proposed augmented framework in Chapter 4 improves the process for making urban scale building energy models more accurate with microclimate integration, it relies on many modeling assumptions, is very timely, resource intensive, and has issues with scalability over larger contexts (ex. county, state, country). To combat these limitations, an obvious solution would be to create a schema that integrates earth observational satellite products directly into epw files. However, at the time, this faces major limitations; these being that earth observational datasets are not temporally precise in their weather and climate data measurements, nor do they measure the exact same climate attributes at TMY3 files, leading to challenges in timestamp aggregation and discrepancies in variable representation. For these reasons, long term future research directions should be taken to investigate the potential of blending earth observational-based microclimate datasets directly into energy plus weather epw files, including the exploration of more temporally combatable (hourly) datasets that measure TMY parameters and respective downscaling opportunities. Moreover, long term technological advancements should increase the revisit period of opensource earth-observational products, making these datasets more compatible for high temporal energy consumption applications.

In larger-scale, physics-based, longitudinal studies, or where high-resolution temporal data is not needed, we identify an opportunity to reference Landsat thermal bands as a scalable proxy to enhance microclimate integration in urban building energy studies. That is, to incorporate urban microclimate conditions into energy models at the building scale, specifically for new building construction, we recommend that development teams compare the Landsat thermal band 10 observations between their development site of

interest and the nearest EPW weather file sampling location, for which their EnergyPlus model is configured to. This additional analysis step provides a proxy to, and directional estimate for, including urban microclimate conditions into building energy consumption calculations. Although, it should be noted that this will only give an approximation of how well the energy plus weather file configures to the development site of interest. In the case that these compared measurements differ greatly, we recommend conducting additional analysis, such as referencing more local data, to access these discrepancies before regarding the EnergyPlus simulations as a ground truth calculations.

For smaller spatial scale simulation-based urban energy studies, or for studies where high temporal data (hourly) is needed, we do not advise using open-source earth-observational climate data, as current technology revisit periods are too long (ex. 16 days), and thus cannot infer actionable insights. Given these circumstances, we encourage researchers to convert local weather station data to epw data directly using new tools and packages such as diyepw [183] or EnergyPlus-MCP [226], which each offer promising and more modular UBEM workflows.

Without reliable energy models, engineers and stakeholders struggle to accurately characterize current and future building energy consumption, leading to discrepancies in environmental, social, and economic performance for their structures. These challenges are further amplified in urban environments, where inter-building interactions and local microclimates, such as the urban heat island effect, introduce additional complexity and uncertainty. Local microclimate integration can substantially reduce these modeling gaps, increasing the reliability of building energy demand calculations. Thus, we encourage future urban energy studies to reference and validate their results with observed microclimate conditions using local data for improved energy insights.

In terms of earth-observational microclimate integration, these products are encouraged for use in data-driven studies and simulation-based building energy studies that address larger city planning and sustainable policy interventions. In particular, to examine general trends across the city or for more aggregated and consolidated longitudinal insights. However, these products fail to provide recommendations for building energy usage studies that have smaller time scales (ex. extreme weather events, hourly insights) – and are still not fully compatible for direct integration in energy plus simulations. Addressing this disconnect requires new technologies, approaches, and software packages that bridge the temporal and spatial gaps between observational datasets and simulation-based building energy modeling platforms. We find promising opportunities for making urban scale building energy models more accurate through local microclimate integration. We therefore call for, and encourage the research community to, develop solutions that more closely align modeled and real-world energy scenarios, advancing our collective goal of fostering a more resilient energy future.

Acknowledgments

My work was supported by the Clean Energy Institute Graduate Fellowship, the Herbold Data Science Fellowship, the Valle Scandinavian Research Exchange, the NASA Washington Space Grant Dissertation Completion Award, the RUA Award, and the National Science Foundation CLDBES grant (award number #1855820).

I would like to thank my advisor, Dr. Narjes Abbasabadi, my committee members, Dr. Kristi Morgansen, and my mentors, Dr. Monika Moskal, Dr. Steve Muench, Dr. Julian Marshal, Dr. Mehdi Asharyeri, Bjørn Aas, Dr. Salvatore Carlucci, and Dr. Hooman Farzaneh for all of their wonderful help and assistance.

I would also like to thank my mom, dad, brother, and many, many friends for supporting me through my graduate journey.

References

[1] I. Kousis, M. Manni, and A. L. Pisello, "Environmental mobile monitoring of urban microclimates: A review," *Renew. Sustain. Energy Rev.*, vol. 169, p. 112847, Nov. 2022, doi: 10.1016/j.rser.2022.112847.

[2] P. RICKWOOD, G. GLAZEBROOK, and G. SEARLE, "Urban Structure and Energy—A Review," *Urban Policy Res.*, vol. 26, no. 1, pp. 57–81, Mar. 2008, doi: 10.1080/08111140701629886.

[3] S. Frolking, R. Mahtta, T. Milliman, T. Esch, and K. C. Seto, "Global urban structural growth shows a profound shift from spreading out to building up," *Nat. Cities*, vol. 1, no. 9, pp. 555–566, Sep. 2024, doi: 10.1038/s44284-024-00100-1.

[4] J. E. Nichol, "High-Resolution Surface Temperature Patterns Related to Urban Morphology in a Tropical City: A Satellite-Based Study," *J. Appl. Meteorol. Climatol.*, vol. 35, no. 1, pp. 135–146, Jan. 1996, doi: 10.1175/1520-0450(1996)035%3C0135:HRSTPR%3E2.0.CO;2.

[5] G. Duveiller, J. Hooker, and A. Cescatti, "The mark of vegetation change on Earth's surface energy balance," *Nat. Commun.*, vol. 9, no. 1, Art. no. 1, Feb. 2018, doi: 10.1038/s41467-017-02810-8.

[6] G. Manoli *et al.*, "Magnitude of urban heat islands largely explained by climate and population," *Nature*, vol. 573, no. 7772, Art. no. 7772, Sep. 2019, doi: 10.1038/s41586-019-1512-9.

[7] J. Schwaab, R. Meier, G. Mussetti, S. Seneviratne, C. Bürgi, and E. L. Davin, "The role of urban trees in reducing land surface temperatures in European cities," *Nat. Commun.*, vol. 12, no. 1, Art. no. 1, Nov. 2021, doi: 10.1038/s41467-021-26768-w.

[8] Y. Yang, Q. Gu, H. Wei, H. Liu, W. Wang, and S. Wei, "Transforming and validating urban microclimate data with multi-sourced microclimate datasets for building energy modelling at urban scale," *Energy Build.*, vol. 295, p. 113318, Sep. 2023, doi: 10.1016/j.enbuild.2023.113318.

[9] Y.-H. Juan, C.-Y. Wen, Z. Li, and A.-S. Yang, "Impacts of urban morphology on improving urban wind energy potential for generic high-rise building arrays," *Appl. Energy*, vol. 299, p. 117304, Oct. 2021, doi: 10.1016/j.apenergy.2021.117304.

[10] T. Hong, Y. Xu, K. Sun, W. Zhang, X. Luo, and B. Hooper, "Urban microclimate and its impact on building performance: A case study of San Francisco," *Urban Clim.*, vol. 38, p. 100871, Jul. 2021, doi: 10.1016/j.uclim.2021.100871.

[11] Z. Ouyang *et al.*, "Albedo changes caused by future urbanization contribute to global warming," *Nat. Commun.*, vol. 13, no. 1, Art. no. 1, Jul. 2022, doi: 10.1038/s41467-022-31558-z.

[12] T. R. Oke, "The energetic basis of the urban heat island," *Q. J. R. Meteorol. Soc.*, vol. 108, no. 455, pp. 1–24, 1982, doi: 10.1002/qj.49710845502.

[13] X. Li, Y. Zhou, S. Yu, G. Jia, H. Li, and W. Li, "Urban heat island impacts on building energy consumption: A review of approaches and findings," *Energy*, vol. 174, pp. 407–419, May 2019, doi: 10.1016/j.energy.2019.02.183.

[14] M. Santamouris, C. Cartalis, A. Synnefa, and D. Kolokotsa, "On the impact of urban heat island and global warming on the power demand and electricity consumption of

buildings—A review," *Energy Build.*, vol. 98, pp. 119–124, Jul. 2015, doi: 10.1016/j.enbuild.2014.09.052.
[15] P. Rajagopal, R. S. Priya, and R. Senthil, "A review of recent developments in the impact of environmental measures on urban heat island," *Sustain. Cities Soc.*, vol. 88, p. 104279, Jan. 2023, doi: 10.1016/j.scs.2022.104279.
[16] A. Worthy, M. Ashayeri, J. Marshall, and N. Abbasabadi, "Bridging the simulation-to-reality gap: A comprehensive review of microclimate integration in urban building energy modeling (UBEM)," *Energy Build.*, vol. 331, p. 115392, Mar. 2025, doi: 10.1016/j.enbuild.2025.115392.
[17] A. Hsu, G. Sheriff, T. Chakraborty, and D. Manya, "Disproportionate exposure to urban heat island intensity across major US cities," *Nat. Commun.*, vol. 12, no. 1, Art. no. 1, May 2021, doi: 10.1038/s41467-021-22799-5.
[18] R. I. McDonald *et al.*, "The tree cover and temperature disparity in US urbanized areas: Quantifying the association with income across 5,723 communities," *PLoS ONE*, vol. 16, no. 4, p. e0249715, Apr. 2021, doi: 10.1371/journal.pone.0249715.
[19] "The tree cover and temperature disparity in US urbanized areas: Quantifying the association with income across 5,723 communities | PLOS ONE." Accessed: Oct. 31, 2024. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0249715
[20] L. E. Excell and R. K. Jain, "Examining the impact of energy efficiency retrofits and vegetation on energy performance of institutional buildings: An equity-driven analysis," *Appl. Energy*, vol. 357, p. 121722, Mar. 2024, doi: 10.1016/j.apenergy.2023.121722.
[21] M. Ashbaugh and N. Kittner, "Addressing extreme urban heat and energy vulnerability of renters in Portland, OR with resilient household energy policies," *Energy Policy*, vol. 190, p. 114143, Jul. 2024, doi: 10.1016/j.enpol.2024.114143.
[22] L. E. Excell, A. Nutkiewicz, and R. K. Jain, "Multi-scale retrofit pathways for improving building performance and energy equity across cities: A UBEM framework," *Energy Build.*, vol. 324, p. 114931, Dec. 2024, doi: 10.1016/j.enbuild.2024.114931.
[23] C. F. Reinhart and C. Cerezo Davila, "Urban building energy modeling – A review of a nascent field," *Build. Environ.*, vol. 97, pp. 196–202, Feb. 2016, doi: 10.1016/j.buildenv.2015.12.001.
[24] T. Hong, Y. Chen, X. Luo, N. Luo, and S. H. Lee, "Ten questions on urban building energy modeling," *Build. Environ.*, vol. 168, p. 106508, Jan. 2020, doi: 10.1016/j.buildenv.2019.106508.
[25] F. Johari, G. Peronato, P. Sadeghian, X. Zhao, and J. Widén, "Urban building energy modeling: State of the art and future prospects," *Renew. Sustain. Energy Rev.*, vol. 128, p. 109902, Aug. 2020, doi: 10.1016/j.rser.2020.109902.
[26] N. Abbasabadi and M. Ashayeri, "Urban energy use modeling methods and tools: A review and an outlook," *Build. Environ.*, vol. 161, p. 106270, Aug. 2019, doi: 10.1016/j.buildenv.2019.106270.
[27] C. Wang, M. Ferrando, F. Causone, X. Jin, X. Zhou, and X. Shi, "Data acquisition for urban building energy modeling: A review," *Build. Environ.*, vol. 217, p. 109056, Jun. 2022, doi: 10.1016/j.buildenv.2022.109056.

[28] F. Johari, G. Peronato, P. Sadeghian, X. Zhao, and J. Widén, “Urban building energy modeling: State of the art and future prospects,” *Renew. Sustain. Energy Rev.*, vol. 128, p. 109902, Aug. 2020, doi: 10.1016/j.rser.2020.109902.
[29] C. F. Reinhart and C. Cerezo Davila, “Urban building energy modeling – A review of a nascent field,” *Build. Environ.*, vol. 97, pp. 196–202, Feb. 2016, doi: 10.1016/j.buildenv.2015.12.001.
[30] D. Wiedenhofer, M. Lenzen, and J. K. Steinberger, “Energy requirements of consumption: Urban form, climatic and socio-economic factors, rebounds and their policy implications,” *Energy Policy*, vol. 63, pp. 696–707, Dec. 2013, doi: 10.1016/j.enpol.2013.07.035.
[31] G. Y. Yun and K. Steemers, “Behavioural, physical and socio-economic factors in household cooling energy consumption,” *Appl. Energy*, vol. 88, no. 6, pp. 2191–2200, Jun. 2011, doi: 10.1016/j.apenergy.2011.01.010.
[32] A. Dagoumas, “Modelling socio-economic and energy aspects of urban systems,” *Sustain. Cities Soc.*, vol. 13, pp. 192–206, Oct. 2014, doi: 10.1016/j.scs.2013.11.003.
[33] N. Abbasabadi, “Understanding Social Dynamics in Urban Building and Transportation Energy Behavior,” in *Artificial Intelligence in Performance-Driven Design*, 1st ed., N. Abbasabadi and M. Ashayeri, Eds., Wiley, 2024, pp. 211–230. doi: 10.1002/9781394172092.ch10.
[34] A. Worthy, M. Ashayeri, and N. Abbasabadi, “Leveraging earth observational data products and machine learning to enhance urban building energy modeling (UBEM) with microclimate effects,” *Sustain. Cities Soc.*, vol. 130, p. 106544, Jul. 2025, doi: 10.1016/j.scs.2025.106544.
[35] “68% of the world population projected to live in urban areas by 2050, says UN | UN DESA | United Nations Department of Economic and Social Affairs.” Accessed: Apr. 08, 2023. [Online]. Available: https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html
[36] S. Yang, L. (Leon) Wang, T. Stathopoulos, and A. M. Marey, “Urban microclimate and its impact on built environment – A review,” *Build. Environ.*, vol. 238, p. 110334, Jun. 2023, doi: 10.1016/j.buildenv.2023.110334.
[37] T. Hong, J. Langevin, and K. Sun, “Building simulation: Ten challenges,” *Build. Simul.*, vol. 11, no. 5, pp. 871–898, Oct. 2018, doi: 10.1007/s12273-018-0444-x.
[38] N. Abbasabadi and M. Ashayeri, “Machine Learning in Urban Building Energy Modeling,” in *Artificial Intelligence in Performance-Driven Design*, John Wiley & Sons, Ltd, 2024, pp. 31–55. doi: 10.1002/9781394172092.ch2.
[39] P. de Wilde, “The gap between predicted and measured energy performance of buildings: A framework for investigation,” *Autom. Constr.*, vol. 41, pp. 40–49, May 2014, doi: 10.1016/j.autcon.2014.02.009.
[40] N. Abbasabadi and M. Ashayeri, “From Tweets to Energy Trends (TwEn): An exploratory framework for machine learning-based forecasting of urban-scale energy behavior leveraging social media data,” *Energy Build.*, vol. 317, p. 114440, Aug. 2024, doi: 10.1016/j.enbuild.2024.114440.

[41] G. Happle, J. A. Fonseca, and A. Schlueter, “A review on occupant behavior in urban building energy models,” *Energy Build.*, vol. 174, pp. 276–292, Sep. 2018, doi: 10.1016/J.ENBUILD.2018.06.030.
[42] L. Pérez-Lombard, J. Ortiz, and C. Pout, “A review on buildings energy consumption information,” *Energy Build.*, vol. 40, no. 3, pp. 394–398, Jan. 2008, doi: 10.1016/j.enbuild.2007.03.007.
[43] M. Auffhammer and E. T. Mansur, “Measuring climatic impacts on energy consumption: A review of the empirical literature,” *Energy Econ.*, vol. 46, pp. 522–530, Nov. 2014, doi: 10.1016/j.eneco.2014.04.017.
[44] K. Javanroodi and V. M. Nik, “Impacts of Microclimate Conditions on the Energy Performance of Buildings in Urban Areas,” *Buildings*, vol. 9, no. 8, Art. no. 8, Aug. 2019, doi: 10.3390/buildings9080189.
[45] J. Allegrini, V. Dorer, and J. Carmeliet, “Influence of the urban microclimate in street canyons on the energy demand for space cooling and heating of buildings,” *Energy Build.*, vol. 55, pp. 823–832, Dec. 2012, doi: 10.1016/j.enbuild.2012.10.013.
[46] M. L. Imhoff, P. Zhang, R. E. Wolfe, and L. Bounoua, “Remote sensing of the urban heat island effect across biomes in the continental USA,” *Remote Sens. Environ.*, vol. 114, no. 3, pp. 504–513, Mar. 2010, doi: 10.1016/j.rse.2009.10.008.
[47] M. Palme, L. Inostroza, G. Villacreses, A. Lobato-Cordero, and C. Carrasco, “From urban climate to energy consumption. Enhancing building performance simulation by including the urban heat island effect,” *Energy Build.*, vol. 145, pp. 107–120, Jun. 2017, doi: 10.1016/j.enbuild.2017.03.069.
[48] A. Boccalatte, M. Fossa, L. Gaillard, and C. Menezo, “Microclimate and urban morphology effects on building energy demand in different European cities,” *Energy Build.*, vol. 224, p. 110129, Oct. 2020, doi: 10.1016/j.enbuild.2020.110129.
[49] E. Erell and B. Zhou, “The effect of increasing surface cover vegetation on urban microclimate and energy demand for building heating and cooling,” *Build. Environ.*, vol. 213, p. 108867, Apr. 2022, doi: 10.1016/j.buildenv.2022.108867.
[50] A. Kamal *et al.*, “Impact of urban morphology on urban microclimate and building energy loads,” *Energy Build.*, vol. 253, p. 111499, Dec. 2021, doi: 10.1016/j.enbuild.2021.111499.
[51] J. A. Fonseca and A. Schlueter, “Integrated model for characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts,” *Appl. Energy*, vol. 142, pp. 247–265, Mar. 2015, doi: 10.1016/j.apenergy.2014.12.068.
[52] C. E. Kontokosta and C. Tull, “A data-driven predictive model of city-scale energy use in buildings,” *Appl. Energy*, vol. 197, pp. 303–317, 2017, doi: 10.1016/j.apenergy.2017.04.005.
[53] C. F. Reinhart and C. Cerezo Davila, “Urban building energy modeling – A review of a nascent field,” *Build. Environ.*, vol. 97, pp. 196–202, Feb. 2016, doi: 10.1016/J.BUILDENV.2015.12.001.
[54] M. C. Silva, I. M. Horta, V. Leal, and V. Oliveira, “A spatially-explicit methodological framework based on neural networks to assess the effect of urban form on energy

demand," *Appl. Energy*, vol. 202, pp. 386–398, 2017, doi: 10.1016/j.apenergy.2017.05.113.
[55] H. Naderi and A. Shojaei, "Digital twinning of civil infrastructures: Current state of model architectures, interoperability solutions, and future prospects," *Autom. Constr.*, vol. 149, p. 104785, May 2023, doi: 10.1016/j.autcon.2023.104785.
[56] Y. Toparlar, B. Blocken, B. Maiheu, and G. J. F. van Heijst, "A review on the CFD analysis of urban microclimate," *Renew. Sustain. Energy Rev.*, vol. 80, pp. 1613–1640, Dec. 2017, doi: 10.1016/j.rser.2017.05.248.
[57] N. Sezer, H. Yoonus, D. Zhan, L. (Leon) Wang, I. G. Hassan, and M. A. Rahman, "Urban microclimate and building energy models: A review of the latest progress in coupling strategies," *Renew. Sustain. Energy Rev.*, vol. 184, p. 113577, Sep. 2023, doi: 10.1016/j.rser.2023.113577.
[58] N. Lauzet *et al.*, "How building energy models take the local climate into account in an urban context – A review," *Renew. Sustain. Energy Rev.*, vol. 116, p. 109390, Dec. 2019, doi: 10.1016/j.rser.2019.109390.
[59] H. Zhao and F. Magoulès, "A review on the prediction of building energy consumption," *Renew. Sustain. Energy Rev.*, vol. 16, no. 6, pp. 3586–3592, Aug. 2012, doi: 10.1016/j.rser.2012.02.049.
[60] K. Amasyali and N. M. El-Gohary, "A review of data-driven building energy consumption prediction studies," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 1192–1205, Jan. 2018, doi: 10.1016/j.rser.2017.04.095.
[61] N. M. Waly, H. Hassan, R. Murata, D. J. Sailor, and H. Mahmoud, "Correlating the urban microclimate and energy demands in hot climate Contexts: A hybrid review," *Energy Build.*, vol. 295, p. 113303, Sep. 2023, doi: 10.1016/j.enbuild.2023.113303.
[62] R. Ooka, "Recent development of assessment tools for urban climate and heat-island investigation especially based on experiences in Japan," *Int. J. Climatol.*, vol. 27, no. 14, pp. 1919–1930, 2007, doi: 10.1002/joc.1630.
[63] J. Li, Y. Mao, J. Ouyang, and S. Zheng, "A Review of Urban Microclimate Research Based on CiteSpace and VOSviewer Analysis," *Int. J. Environ. Res. Public. Health*, vol. 19, no. 8, p. 4741, Apr. 2022, doi: 10.3390/ijerph19084741.
[64] I. Orlanski, "A Rational Subdivision of Scales for Atmospheric Processes," *Bull. Am. Meteorol. Soc.*, vol. 56, no. 5, pp. 527–530, 1975, Accessed: Oct. 12, 2023. [Online]. Available: https://www.jstor.org/stable/26216020
[65] M. H. Elnabawi and N. Hamza, "A Methodology of Creating a Synthetic, Urban-Specific Weather Dataset Using a Microclimate Model for Building Energy Modelling," *Buildings*, vol. 12, no. 9, Art. no. 9, Sep. 2022, doi: 10.3390/buildings12091407.
[66] "Total data volume worldwide 2010-2025," Statista. Accessed: Apr. 09, 2023. [Online]. Available: https://www.statista.com/statistics/871513/worldwide-data-created/
[67] J. M. Colston *et al.*, "Evaluating meteorological data from weather stations, and from satellites and global models for a multi-site epidemiological study," *Environ. Res.*, vol. 165, pp. 91–109, Aug. 2018, doi: 10.1016/j.envres.2018.02.027.
[68] J. Yang *et al.*, "The role of satellite remote sensing in climate change studies," *Nat. Clim. Change*, vol. 3, no. 10, Art. no. 10, Oct. 2013, doi: 10.1038/nclimate1908.

[69] S. Cheval *et al.*, “Meteorological and Ancillary Data Resources for Climate Research in Urban Areas,” *Climate*, vol. 8, no. 3, Art. no. 3, Mar. 2020, doi: 10.3390/cli8030037.
[70] “World Meteorological Organization (WMO),” Library of Congress, Washington, D.C. 20540 USA. Accessed: Oct. 13, 2023. [Online]. Available: https://www.loc.gov/item/lcwaN0010741/
[71] “Weather Underground,” Library of Congress, Washington, D.C. 20540 USA. Accessed: Oct. 12, 2023. [Online]. Available: https://lccn.loc.gov/2004564291
[72] “Welcome to OpenSky | OpenSky.” Accessed: Oct. 12, 2023. [Online]. Available: https://opensky.ucar.edu/
[73] “Climate Data Online (CDO) - The National Climatic Data Center’s (NCDC) Climate Data Online (CDO) provides free access to NCDC’s archive of historical weather and climate data in addition to station history information. | National Climatic Data Center (NCDC).” Accessed: Oct. 12, 2023. [Online]. Available: https://www.ncei.noaa.gov/cdo-web/
[74] “Cal-Adapt.” Accessed: Oct. 12, 2023. [Online]. Available: https://cal-adapt.org/help/faqs/how-do-i-cite-caladapt/
[75] “OpenWeatherMap API guide - OpenWeatherMap.” Accessed: Oct. 12, 2023. [Online]. Available: https://openweathermap.org/guide
[76] N. Luo *et al.*, “A data schema for exchanging information between urban building energy models and urban microclimate models in coupled simulations,” *J. Build. Perform. Simul.*, vol. 0, no. 0, pp. 1–18, Nov. 2022, doi: 10.1080/19401493.2022.2142295.
[77] M. Zinzi, E. Carnielo, and B. Mattoni, “On the relation between urban climate and energy performance of buildings. A three-years experience in Rome, Italy,” *Appl. Energy*, vol. 221, pp. 148–160, Jul. 2018, doi: 10.1016/j.apenergy.2018.03.192.
[78] H.-D. Guo, L. Zhang, and L.-W. Zhu, “Earth observation big data for climate change research,” *Adv. Clim. Change Res.*, vol. 6, no. 2, pp. 108–117, Jun. 2015, doi: 10.1016/j.accre.2015.09.007.
[79] Z.-L. Li *et al.*, “Satellite Remote Sensing of Global Land Surface Temperature: Definition, Methods, Products, and Applications,” *Rev. Geophys.*, vol. 61, no. 1, p. e2022RG000777, 2023, doi: 10.1029/2022RG000777.
[80] A. Anand and C. Deb, “The potential of remote sensing and GIS in urban building energy modelling,” *Energy Built Environ.*, p. S2666123323000685, Jul. 2023, doi: 10.1016/j.enbenv.2023.07.008.
[81] M. Wurm, A. Droin, T. Stark, C. Geiß, W. Sulzer, and H. Taubenböck, “Deep Learning-Based Generation of Building Stock Data from Remote Sensing for Urban Heat Demand Modeling,” *ISPRS Int. J. Geo-Inf.*, vol. 10, no. 1, Art. no. 1, Jan. 2021, doi: 10.3390/ijgi10010023.
[82] T. R. Dougherty and R. K. Jain, “TOM.D: Taking advantage Of Microclimate Data for Urban Building Energy Modeling,” *Adv. Appl. Energy*, p. 100138, Apr. 2023, doi: 10.1016/j.adapen.2023.100138.
[83] T. R. Dougherty and R. K. Jain, “Invisible walls: Exploration of microclimate effects on building energy consumption in New York City,” *Sustain. Cities Soc.*, vol. 90, p. 104364, Mar. 2023, doi: 10.1016/j.scs.2022.104364.

[84] “Landsat 9 | U.S. Geological Survey.” Accessed: Oct. 31, 2023. [Online]. Available: https://www.usgs.gov/landsat-missions/landsat-9
[85] “MODIS Web.” Accessed: Oct. 31, 2023. [Online]. Available: https://modis.gsfc.nasa.gov/about/specifications.php
[86] K. Thome, “ASTER | Terra.” Accessed: Oct. 31, 2023. [Online]. Available: https://terra.nasa.gov/about/terra-instruments/aster
[87] “ASTER Satellite Sensor Specifications | Satellite Imaging Corp.” Accessed: Nov. 21, 2023. [Online]. Available: https://www.satimagingcorp.com/satellite-sensors/other-satellite-sensors/aster/
[88] “Sentinel-2 - Missions - Sentinel Online,” Sentinel Online. Accessed: Nov. 02, 2023. [Online]. Available: https://copernicus.eu/missions/sentinel-2
[89] “Data Products - Sentinel-5P Mission - Sentinel Online,” Sentinel Online. Accessed: Mar. 26, 2024. [Online]. Available: https://copernicus.eu/missions/sentinel-5p/data-products
[90] “Launch Info | ICESat-2.” Accessed: Nov. 01, 2023. [Online]. Available: https://icesat-2.gsfc.nasa.gov/launch-info
[91] J. DiMarzio and D. Hancock, “Ice, Cloud, and Land Elevation Satellite (ICESat-2) Project Algorithm Theoretical Basis Document (ATBD) For ATLAS Level 1A Processing, version 4,” 2021, doi: 10.5067/IGMFO2X80N1E.
[92] H. Setchell, “ECMWF Reanalysis v5,” ECMWF. Accessed: Oct. 25, 2023. [Online]. Available: https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5
[93] R. Gelaro *et al.*, “The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2),” *J. Clim.*, vol. 30, no. 14, pp. 5419–5454, Jul. 2017, doi: 10.1175/JCLI-D-16-0758.1.
[94] N. C. Operations, “NCEP Data Products RTMA/URMA.” Accessed: Nov. 06, 2023. [Online]. Available: https://www.nco.ncep.noaa.gov/pmb/products/rtma/
[95] “USFS Tree Canopy Cover v2021-4 (CONUS and OCONUS) | Earth Engine Data Catalog,” Google for Developers. Accessed: Aug. 08, 2024. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/USGS_NLCD_RELEASES_2021_REL_TCC_v2021-4
[96] “NASADEM: NASA NASADEM Digital Elevation 30m | Earth Engine Data Catalog,” Google for Developers. Accessed: Jul. 30, 2024. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/NASA_NASADEM_HGT_001
[97] “NLCD 2021: USGS National Land Cover Database, 2021 release | Earth Engine Data Catalog,” Google for Developers. Accessed: Jul. 30, 2024. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/USGS_NLCD_RELEASES_2021_REL_NLCD
[98] O. Nicolis and C. Gonzalez, “19 - Wavelet-based fractal and multifractal analysis for detecting mineral deposits using multispectral images taken by drones,” in *Methods and Applications in Petroleum and Mineral Exploration and Engineering Geology*, S. Gaci, O. Hachay, and O. Nicolis, Eds., Elsevier, 2021, pp. 295–307. doi: 10.1016/B978-0-323-85617-1.00017-5.

[99] M. Govender, K. Chetty, and H. Bulcock, “A review of hyperspectral remote sensing and its application in vegetation and water resource studies,” *Water SA*, vol. 33, no. 2, pp. 145–151, Apr. 2007, doi: 10.10520/EJC116430.
[100] “Visible Infrared Imaging Radiometer Suite (VIIRS) - LAADS DAAC.” Accessed: Nov. 06, 2023. [Online]. Available: https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/viirs/
[101] S. Subudhi, R. G. Dabhade, R. Shastri, V. Gundu, G. D. Vignesh, and A. Chaturvedi, “Empowering sustainable farming practices with AI-enabled interactive visualization of hyperspectral imaging data,” *Meas. Sens.*, p. 100935, Oct. 2023, doi: 10.1016/j.measen.2023.100935.
[102] M. M. Dorostkar, “CityFFD – City Fast Fluid Dynamics Model for Urban Microclimate Simulations”.
[103] W. Wang, K. Liu, R. Tang, and S. Wang, “Remote sensing image-based analysis of the urban heat island effect in Shenzhen, China,” *Phys. Chem. Earth Parts ABC*, vol. 110, pp. 168–175, Apr. 2019, doi: 10.1016/j.pce.2019.01.002.
[104] F. Fouladinejad, A. Matkan, M. Hajeb, and F. Brakhasi, “HISTORY AND APPLICATIONS OF SPACE-BORNE LIDARS,” *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XLII-4-W18, pp. 407–414, Oct. 2019, doi: 10.5194/isprs-archives-XLII-4-W18-407-2019.
[105] T. T. Vu, F. Yamazaki, and M. Matsuoka, “Multi-scale solution for building extraction from LiDAR and image data,” *Int. J. Appl. Earth Obs. Geoinformation*, vol. 11, no. 4, pp. 281–289, Aug. 2009, doi: 10.1016/j.jag.2009.03.005.
[106] *microsoft/USBuildingFootprints*. (Apr. 15, 2023). Microsoft. Accessed: Apr. 15, 2023. [Online]. Available: https://github.com/microsoft/USBuildingFootprints
[107] M. P. Heris, N. L. Foks, K. J. Bagstad, A. Troy, and Z. H. Ancona, “A rasterized building footprint dataset for the United States,” *Sci. Data*, vol. 7, no. 1, Art. no. 1, Jun. 2020, doi: 10.1038/s41597-020-0542-3.
[108] C. R. de Almeida, A. C. Teodoro, and A. Gonçalves, “Study of the Urban Heat Island (UHI) Using Remote Sensing Data/Techniques: A Systematic Review,” *Environments*, vol. 8, no. 10, Art. no. 10, Oct. 2021, doi: 10.3390/environments8100105.
[109] A. Sekertekin and N. Arslan, “Monitoring thermal anomaly and radiative heat flux using thermal infrared satellite imagery – A case study at Tuzla geothermal region,” *Geothermics*, vol. 78, pp. 243–254, Mar. 2019, doi: 10.1016/j.geothermics.2018.12.014.
[110] J. A. Voogt and T. R. Oke, “Thermal remote sensing of urban climates,” *Remote Sens. Environ.*, vol. 86, no. 3, pp. 370–384, Aug. 2003, doi: 10.1016/S0034-4257(03)00079-8.
[111] “‘Surface,’ ‘satellite’ or ‘simulation’: Mapping intra-urban microclimate variability in a desert city - Zhou - 2020 - International Journal of Climatology - Wiley Online Library.” Accessed: Oct. 30, 2023. [Online]. Available: https://rmets.onlinelibrary.wiley.com/doi/full/10.1002/joc.6385
[112] E. J. Good, F. M. Aldred, D. J. Ghent, K. L. Veal, and C. Jimenez, “An Analysis of the Stability and Trends in the LST_cci Land Surface Temperature Datasets Over Europe,”

Earth Space Sci., vol. 9, no. 9, p. e2022EA002317, 2022, doi: 10.1029/2022EA002317.
[113] C. J. Tomlinson, L. Chapman, J. E. Thornes, and C. Baker, "Remote sensing land surface temperature for meteorology and climatology: a review," *Meteorol. Appl.*, vol. 18, no. 3, pp. 296–306, 2011, doi: 10.1002/met.287.
[114] G. Rigo, E. Parlow, and D. Oesch, "Validation of satellite observed thermal emission with in-situ measurements over an urban surface," *Remote Sens. Environ.*, vol. 104, no. 2, pp. 201–210, Sep. 2006, doi: 10.1016/j.rse.2006.04.018.
[115] C. Yang *et al.*, "Assessing the effects of 2D/3D urban morphology on the 3D urban thermal environment by using multi-source remote sensing data and UAV measurements: A case study of the snow-climate city of Changchun, China," *J. Clean. Prod.*, vol. 321, p. 128956, Oct. 2021, doi: 10.1016/j.jclepro.2021.128956.
[116] D. Mutiibwa, S. Strachan, and T. Albright, "Land Surface Temperature and Surface Air Temperature in Complex Terrain," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 8, no. 10, pp. 4762–4774, Oct. 2015, doi: 10.1109/JSTARS.2015.2468594.
[117] J. Hooker, G. Duveiller, and A. Cescatti, "A global dataset of air temperature derived from satellite remote sensing and weather stations," *Sci. Data*, vol. 5, no. 1, Art. no. 1, Nov. 2018, doi: 10.1038/sdata.2018.246.
[118] M. N. Mistry *et al.*, "Comparison of weather station and climate reanalysis data for modelling temperature-related mortality," *Sci. Rep.*, vol. 12, no. 1, Art. no. 1, Mar. 2022, doi: 10.1038/s41598-022-09049-4.
[119] H. Zandler, T. Senftl, and K. A. Vanselow, "Reanalysis datasets outperform other gridded climate products in vegetation change analysis in peripheral conservation areas of Central Asia," *Sci. Rep.*, vol. 10, no. 1, Art. no. 1, Dec. 2020, doi: 10.1038/s41598-020-79480-y.
[120] S. Liu, Y. T. Kwok, and C. Ren, "Investigating the impact of urban microclimate on building thermal performance: A case study of dense urban areas in Hong Kong," *Sustain. Cities Soc.*, vol. 94, p. 104509, Jul. 2023, doi: 10.1016/j.scs.2023.104509.
[121] "ENVI-met high-resolution 3D modeling for Climate Adaption." Accessed: Jan. 07, 2025. [Online]. Available: https://envi-met.com/
[122] M. Mortezazadeh, L. L. Wang, M. Albettar, and S. Yang, "CityFFD – City fast fluid dynamics for urban microclimate simulations on graphics processing units," *Urban Clim.*, vol. 41, p. 101063, Jan. 2022, doi: 10.1016/j.uclim.2021.101063.
[123] J. P. Lafore *et al.*, "The Meso-NH Atmospheric Simulation System. Part I: adiabatic formulation and control simulations," *Ann. Geophys.*, vol. 16, no. 1, pp. 90–109, Jan. 1998, doi: 10.1007/s00585-997-0090-6.
[124] "Weather Research & Forecasting Model (WRF) | Mesoscale & Microscale Meteorology Laboratory." Accessed: Jun. 05, 2023. [Online]. Available: https://www.mmm.ucar.edu/models/wrf
[125] "TEB - National Centre for Meteorological Research." Accessed: Jan. 07, 2025. [Online]. Available: https://www.umr-cnrm.fr/spip.php?article199&lang=en
[126] B. Bueno, A. Nakano, and L. Norford, "Urban weather generator: a method to predict neighborhood-specific urban temperatures for use in building energy simulations".

[127] B. Morille, N. Lauzet, and M. Musy, “SOLENE-microclimate: A Tool to Evaluate Envelopes Efficiency on Energy Consumption at District Scale.,” *Energy Procedia*, vol. 78, pp. 1165–1170, Nov. 2015, doi: 10.1016/j.egypro.2015.11.088.
[128] H. H. Hu, “Chapter 10 - Computational Fluid Dynamics,” in *Fluid Mechanics (Fifth Edition)*, P. K. Kundu, I. M. Cohen, and D. R. Dowling, Eds., Boston: Academic Press, 2012, pp. 421–472. doi: 10.1016/B978-0-12-382100-3.10010-1.
[129] P. J. Crank, D. J. Sailor, G. Ban-Weiss, and M. Taleghani, “Evaluating the ENVI-met microscale model for suitability in analysis of targeted urban heat mitigation strategies,” *Urban Clim.*, vol. 26, pp. 188–197, Dec. 2018, doi: 10.1016/j.uclim.2018.09.002.
[130] A. Alyakoob, S. Hartono, T. Johnson, and A. Middel, “Estimating cooling loads of Arizona State University buildings using microclimate data and machine learning,” *J. Build. Eng.*, vol. 64, p. 105705, Apr. 2023, doi: 10.1016/j.jobe.2022.105705.
[131] J. Hensen, “Modelling coupled heat and airflow: Ping-pong versus onions,” *Proc. 16th AIVC Conf.*, pp. 253–262, Jan. 1995.
[132] A. Katal, M. Mortezazadeh, L. (Leon) Wang, and H. Yu, “Urban building energy and microclimate modeling – From 3D city generation to dynamic simulations,” *Energy*, vol. 251, p. 123817, Jul. 2022, doi: 10.1016/j.energy.2022.123817.
[133] P. A. Mirzaei and F. Haghighat, “Approaches to study Urban Heat Island – Abilities and limitations,” *Build. Environ.*, vol. 45, no. 10, pp. 2192–2201, Oct. 2010, doi: 10.1016/j.buildenv.2010.04.001.
[134] “Wiki - Meso-NH - CNRM Open Source Site.” Accessed: Jun. 05, 2023. [Online]. Available: https://opensource.umr-cnrm.fr/projects/meso-nh/wiki
[135] C. Lac *et al.*, “Overview of the Meso-NH model version 5.4 and its applications,” *Geosci. Model Dev.*, vol. 11, no. 5, pp. 1929–1969, May 2018, doi: 10.5194/gmd-11-1929-2018.
[136] R. Jain, X. Luo, G. Sever, T. Hong, and C. Catlett, “Representation and evolution of urban weather boundary conditions in downtown Chicago,” *J. Build. Perform. Simul.*, vol. 13, no. 2, pp. 182–194, Mar. 2020, doi: 10.1080/19401493.2018.1534275.
[137] D. C. Dowell *et al.*, “The High-Resolution Rapid Refresh (HRRR): An Hourly Updating Convection-Allowing Forecast Model. Part I: Motivation and System Description,” *Weather Forecast.*, vol. 37, no. 8, pp. 1371–1395, Aug. 2022, doi: 10.1175/WAF-D-21-0151.1.
[138] “Multi-layer coupling between SURFEX-TEB-V9.0 and Meso-NH-v5.3 for modelling the urban climate of high-rise cities,” *Geosci. Model Dev.*, Jul. 2020, Accessed: Jun. 05, 2023. [Online]. Available: https://www.researcher-app.com/paper/5422518
[139] C. Jansson, P. Samuelsson, and D. Lindstedt, “Using the Town Energy Balance model (TEB) in regional climate simulations over the Netherlands”.
[140] M. Sadeghipour Roudsari, M. Pak, and A. Viola, “Ladybug: A Parametric Environmental Plugin For Grasshopper To Help Designers Create An Environmentally-conscious Design,” presented at the 2017 Building Simulation Conference, Aug. 2013. doi: 10.26868/25222708.2013.2499.
[141] “Buildings - Energy System,” IEA. Accessed: Apr. 25, 2024. [Online]. Available: https://www.iea.org/energy-system/buildings

[142] "Net-Zero Emissions Operations by 2050, including a 65% reduction by 2030 | Federal Sustainability Plan | Office of the Federal Chief Sustainability Officer." Accessed: Apr. 25, 2024. [Online]. Available: https://www.sustainability.gov/federalsustainabilityplan/emissions.html
[143] K. I. Praseeda, B. V. Venkatarama Reddy, and M. Mani, "Life-Cycle Energy Assessment in Buildings: Framework, Approaches, and Case Studies," in *Encyclopedia of Sustainable Technologies*, M. A. Abraham, Ed., Oxford: Elsevier, 2017, pp. 113–136. doi: 10.1016/B978-0-12-409548-9.10188-5.
[144] A. Salvati, H. Coch, and M. Morganti, "Effects of urban compactness on the building energy performance in Mediterranean climate," *Energy Procedia*, vol. 122, pp. 499–504, Sep. 2017, doi: 10.1016/j.egypro.2017.07.303.
[145] S. Falasca, M. Zinzi, A. M. Siani, G. Curci, L. Ding, and M. Santamouris, "Investigating the effects of the greenery increase on air temperature, ventilation and cooling energy demand in Melbourne with the Weather Research and Forecasting model and Local Climate Zones," *Sci. Total Environ.*, vol. 953, p. 176016, Nov. 2024, doi: 10.1016/j.scitotenv.2024.176016.
[146] J. Hensen and R. Lamberts, Eds., *Building performance simulation for design and operation*. London New York: Spon Press, 2011.
[147] L. G. Swan and V. I. Ugursal, "Modeling of end-use energy consumption in the residential sector: A review of modeling techniques," *Renew. Sustain. Energy Rev.*, vol. 13, no. 8, pp. 1819–1835, 2009, doi: 10.1016/j.rser.2008.09.033.
[148] "EnergyPlus." Accessed: Jan. 10, 2025. [Online]. Available: https://energyplus.net/weather-region/north_and_central_america_wmo_region_4/USA/WA
[149] Y. Pan and L. Zhang, "Data-driven estimation of building energy consumption with multi-source heterogeneous data," *Appl. Energy*, vol. 268, p. 114965, Jun. 2020, doi: 10.1016/j.apenergy.2020.114965.
[150] Y. Ahn and D.-W. Sohn, "The effect of neighbourhood-level urban form on residential building energy use: A GIS-based model using building energy benchmarking data in Seattle," *Energy Build.*, vol. 196, pp. 124–133, Aug. 2019, doi: 10.1016/j.enbuild.2019.05.018.
[151] A. Sola, C. Corchero, J. Salom, and M. Sanmarti, "Simulation tools to build urban-scale energy models: A review," *Energies*, vol. 11, no. 12, pp. 3269–3269, Nov. 2018, doi: 10.3390/en11123269.
[152] "Seattle drops out of Top 10 for growth among largest U.S. cities," The Seattle Times. Accessed: Sep. 19, 2024. [Online]. Available: https://www.seattletimes.com/seattle-news/data/seattle-drops-out-of-top-10-for-growth-among-largest-u-s-cities/
[153] "2021-summary-report-heat-watch-seattle-king-county.pdf." Accessed: Apr. 16, 2024. [Online]. Available: https://your.kingcounty.gov/dnrp/climate/documents/2021-summary-report-heat-watch-seattle-king-county.pdf
[154] "Top 50 Cities in the U.S. by Population & Rank | Infoplease." Accessed: May 21, 2024. [Online]. Available: https://www.infoplease.com/us/cities/top-50-cities-us-population-and-rank

[155] “Seattle city, Washington - Census Bureau Profile.” Accessed: May 21, 2024. [Online]. Available: https://data.census.gov/profile/Seattle_city,_Washington?g=160XX00US5363000
[156] “About Seattle - OPCD | seattle.gov.” Accessed: May 21, 2024. [Online]. Available: https://www.seattle.gov/opcd/population-and-demographics/about-seattle#landuse
[157] “Weather averages Seattle, Washington.” Accessed: Jul. 18, 2021. [Online]. Available: https://www.usclimatedata.com/climate/seattle/washington/united-states/uswa0844
[158] Z. Kearl and J. Vogel, “Urban extreme heat, climate change, and saving lives: Lessons from Washington state,” *Urban Clim.*, vol. 47, p. 101392, Jan. 2023, doi: 10.1016/j.uclim.2022.101392.
[159] “Air-Conditioning Was Once Taboo in Seattle. Not Anymore. - The New York Times.” Accessed: Jul. 18, 2021. [Online]. Available: https://www.nytimes.com/2021/06/25/us/western-heat-wave.html
[160] “Building Energy Benchmarking.” Accessed: Apr. 15, 2025. [Online]. Available: https://opendata.dc.gov/datasets/DCGIS::building-energy-benchmarking/about
[161] “Chicago Energy Benchmarking | City of Chicago | Data Portal.” Accessed: Apr. 15, 2025. [Online]. Available: https://data.cityofchicago.org/Environment-Sustainable-Development/Chicago-Energy-Benchmarking/xq83-jr8c/about_data
[162] “Existing Buildings Energy & Water Efficiency (EBEWE) Program | Los Angeles - Open Data Portal.” Accessed: Apr. 15, 2025. [Online]. Available: https://data.lacity.org/City-Infrastructure-Service-Requests/Existing-Buildings-Energy-Water-Efficiency-EBEWE-P/9yda-i4ya/about_data
[163] “NYC Energy Benchmarking Report.” Accessed: Apr. 15, 2025. [Online]. Available: https://www.nyc.gov/site/finance/property/nyc-energy-benchmarking-report.page
[164] “Energy Benchmarking - Environment | seattle.gov.” Accessed: May 21, 2024. [Online]. Available: https://www.seattle.gov/environment/climate-change/buildings-and-energy/energy-benchmarking
[165] “Building Energy Benchmarking Data, 2015-Present | City of Seattle Open Data portal.” Accessed: Apr. 01, 2025. [Online]. Available: https://data.seattle.gov/Built-Environment/Building-Energy-Benchmarking-Data-2015-Present/teqw-tu6e/about_data
[166] “EnergyPlus.” Accessed: May 23, 2024. [Online]. Available: https://energyplus.net/
[167] geocube Contributors, *geocube: Tool to convert geopandas vector data into rasterized xarray data*. Python. Accessed: May 24, 2024. [OS Independent]. Available: https://github.com/corteva/geocube
[168] “Xarray documentation,” xarray. Accessed: May 24, 2024. [Online]. Available: https://docs.xarray.dev/en/latest/index.html
[169] “Rasterio: access to geospatial raster data — rasterio documentation.” Accessed: May 24, 2024. [Online]. Available: https://rasterio.readthedocs.io/en/stable/
[170] S. M. Crum and G. D. Jenerette, “Microclimate Variation among Urban Land Covers: The Importance of Vertical and Horizontal Structure in Air and Land Surface Temperature Relationships,” *J. Appl. Meteorol. Climatol.*, vol. 56, no. 9, pp. 2531–2543, Sep. 2017, doi: 10.1175/JAMC-D-17-0054.1.

[171] D. Ballari *et al.*, “Satellite Earth Observation for Essential Climate Variables Supporting Sustainable Development Goals: A Review on Applications,” *Remote Sens.*, vol. 15, no. 11, Art. no. 11, Jan. 2023, doi: 10.3390/rs15112716.
[172] N. Meili *et al.*, “Tree effects on urban microclimate: Diurnal, seasonal, and climatic temperature differences explained by separating radiation, evapotranspiration, and roughness effects,” *Urban For. Urban Green.*, vol. 58, p. 126970, Mar. 2021, doi: 10.1016/j.ufug.2020.126970.
[173] “USGS Landsat 8 Level 2, Collection 2, Tier 1 | Earth Engine Data Catalog,” Google for Developers. Accessed: Apr. 29, 2025. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC08_C02_T1_L2
[174] “RTMA: Real-Time Mesoscale Analysis | Earth Engine Data Catalog,” Google for Developers. Accessed: Sep. 16, 2024. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/NOAA_NWS_RTMA
[175] “Sentinel-5P NRTI O3: Near Real-Time Ozone | Earth Engine Data Catalog,” Google for Developers. Accessed: Sep. 16, 2024. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S5P_NRTI_L3_O3
[176] “MOD11A1.061 Terra Land Surface Temperature and Emissivity Daily Global 1km | Earth Engine Data Catalog,” Google for Developers. Accessed: Sep. 16, 2024. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD11A1
[177] I. Guyon and A. Elisseeff, “An Introduction to Variable and Feature Selection”.
[178] Q. Qiao, A. Yunusa-Kaltungo, and R. E. Edwards, “Feature selection strategy for machine learning methods in building energy consumption prediction,” *Energy Rep.*, vol. 8, pp. 13621–13654, Nov. 2022, doi: 10.1016/j.egyr.2022.10.125.
[179] D. Chicco, M. J. Warrens, and G. Jurman, “The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation,” *PeerJ Comput. Sci.*, vol. 7, 2021, doi: 10.7717/peerj-cs.623.
[180] “2021 Building Energy Benchmarking | City of Seattle Open Data portal,” Seattle. Accessed: May 21, 2024. [Online]. Available: https://data.seattle.gov/Permitting/2021-Building-Energy-Benchmarking/bfsh-nrm6
[181] “Ladybug Tools | Home Page.” Accessed: Oct. 30, 2023. [Online]. Available: https://www.ladybug.tools/
[182] “USGS Landsat 8 Level 2, Collection 2, Tier 1 | Earth Engine Data Catalog,” Google for Developers. Accessed: Sep. 16, 2024. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC08_C02_T1_L2
[183] A. D. Smith, B. Stürmer, T. Thurber, and C. R. Vernon, “diyepw: A Python package for Do-It-Yourself EnergyPlus weather file generation,” *J. Open Source Softw.*, vol. 6, no. 64, p. 3313, Aug. 2021, doi: 10.21105/joss.03313.
[184] A. Siddiqui, G. Kushwaha, B. Nikam, S. K. Srivastav, A. Shelar, and P. Kumar, “Analysing the day/night seasonal and annual changes and trends in land surface

temperature and surface urban heat island intensity (SUHII) for Indian cities," *Sustain. Cities Soc.*, vol. 75, p. 103374, Dec. 2021, doi: 10.1016/j.scs.2021.103374.
[185] W. Ullah *et al.*, "Analysis of the relationship among land surface temperature (LST), land use land cover (LULC), and normalized difference vegetation index (NDVI) with topographic elements in the lower Himalayan region," *Heliyon*, vol. 9, no. 2, p. e13322, Feb. 2023, doi: 10.1016/j.heliyon.2023.e13322.
[186] Z. Wang and R. S. Srinivasan, "A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models," *Renew. Sustain. Energy Rev.*, vol. 75, pp. 796–808, Aug. 2017, doi: 10.1016/j.rser.2016.10.079.
[187] A. Oraiopoulos and B. Howard, "On the accuracy of Urban Building Energy Modelling," *Renew. Sustain. Energy Rev.*, vol. 158, p. 111976, Apr. 2022, doi: 10.1016/j.rser.2021.111976.
[188] "Feature importance |." Accessed: Apr. 20, 2025. [Online]. Available: https://catboost.ai/docs/en/concepts/fstr
[189] "Seattle Energy Benchmarking Analysis 2016 for web.pdf." Accessed: Jun. 05, 2024. [Online]. Available: https://www.seattle.gov/documents/Departments/OSE/Seattle%20Energy%20Benchmarking%20Analysis%202016%20for%20web.pdf
[190] N. Abbasabadi, M. Ashayeri, R. Azari, B. Stephens, and M. Heidarinejad, "An integrated data-driven framework for urban energy use modeling (UEUM)," *Appl. Energy*, vol. 253, p. 113550, Nov. 2019, doi: 10.1016/j.apenergy.2019.113550.
[191] J. Gao *et al.*, "Dilution effect of the building area on energy intensity in urban residential buildings," *Nat. Commun.*, vol. 10, no. 1, p. 4944, Oct. 2019, doi: 10.1038/s41467-019-12852-9.
[192] C. Filippín, "Benchmarking the energy efficiency and greenhouse gases emissions of school buildings in central Argentina," *Build. Environ.*, vol. 35, no. 5, pp. 407–414, Jul. 2000, doi: 10.1016/S0360-1323(99)00035-9.
[193] B. Abbasi, Z. Qin, W. Du, S. Li, J. Fan, and S. Zhao, "Effects of Cloud on Land Surface Temperature (LST) Change in Thermal Infrared Remote Sensing Images: a Case Study of Landsat 8 Data," in *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, Sep. 2020, pp. 5430–5433. doi: 10.1109/IGARSS39084.2020.9324415.
[194] S. Wilcox and W. Marion, "Users Manual for TMY3 Data Sets," *Tech. Rep.*, 2008.
[195] A. Azarnejad and A. Mahdavi, "Building Façades' Visual Reflectance and Surface Temperatures: A Field Study," *Energy Procedia*, vol. 78, pp. 1720–1725, Nov. 2015, doi: 10.1016/j.egypro.2015.11.277.
[196] C. Marino, A. Nucara, and M. Pietrafesa, "Does window-to-wall ratio have a significant effect on the energy consumption of buildings? A parametric analysis in Italian climate conditions," *J. Build. Eng.*, vol. 13, pp. 169–183, Sep. 2017, doi: 10.1016/j.jobe.2017.08.001.
[197] Y. Zhang, X. Bai, F. P. Mills, and J. C. V. Pezzey, "Rethinking the role of occupant behavior in building energy performance: A review," *Energy Build.*, vol. 172, pp. 279–294, Aug. 2018, doi: 10.1016/j.enbuild.2018.05.017.

[198] L. Troup, R. Phillips, M. J. Eckelman, and D. Fannon, “Effect of window-to-wall ratio on measured energy consumption in US office buildings,” *Energy Build.*, vol. 203, p. 109434, Nov. 2019, doi: 10.1016/j.enbuild.2019.109434.
[199] U. Ali, M. H. Shamsi, C. Hoare, E. Mangina, and J. O’Donnell, “Review of urban building energy modeling (UBEM) approaches, methods and tools using qualitative and quantitative analysis,” *Energy Build.*, vol. 246, p. 111073, Sep. 2021, doi: 10.1016/j.enbuild.2021.111073.
[200] E. H. Borgstein, R. Lamberts, and J. L. M. Hensen, “Evaluating energy performance in non-domestic buildings: A review,” *Energy Build.*, vol. 128, pp. 734–755, Sep. 2016, doi: 10.1016/j.enbuild.2016.07.018.
[201] A. Francisco, N. Mohammadi, and J. E. Taylor, “Smart City Digital Twin–Enabled Energy Management: Toward Real-Time Urban Building Energy Benchmarking,” *J. Manag. Eng.*, vol. 36, no. 2, p. 04019045, Mar. 2020, doi: 10.1061/(ASCE)ME.1943-5479.0000741.
[202] A. F. Krelling, R. Lamberts, J. Malik, W. Zhang, K. Sun, and T. Hong, “Defining weather scenarios for simulation-based assessment of thermal resilience of buildings under current and future climates: A case study in Brazil,” *Sustain. Cities Soc.*, vol. 107, p. 105460, Jul. 2024, doi: 10.1016/j.scs.2024.105460.
[203] M. Ashayeri and N. Abbasaabadi, *Socioeconomic determinants of public health and residential building energy use in Chicago*. 2021. Accessed: Sep. 11, 2025. [Online]. Available: https://www.acsa-arch.org/chapter/socioeconomic-determinants-of-public-health-and-residential-building-energy-use-in-chicago/
[204] M. Ashayeri and N. Abbasabadi, “A framework for integrated energy and exposure to ambient pollution (iEnEx) assessment toward low-carbon, healthy, and equitable cities,” *Sustain. Cities Soc.*, vol. 78, p. 103647, Mar. 2022, doi: 10.1016/j.scs.2021.103647.
[205] M. Ashayeri and N. Abbasabadi, “Unraveling energy justice in NYC urban buildings through social media sentiment analysis and transformer deep learning,” *Energy Build.*, vol. 306, p. 113914, Mar. 2024, doi: 10.1016/j.enbuild.2024.113914.
[206] S. Dabirian, K. Panchabikesan, and U. Eicker, “Occupant-centric urban building energy modeling: Approaches, inputs, and data sources - A review,” *Energy Build.*, vol. 257, p. 111809, Feb. 2022, doi: 10.1016/j.enbuild.2021.111809.
[207] B. Bueno, A. Nakano, and L. Norford, “Urban weather generator: a method to predict neighborhood-specific urban temperatures for use in building energy simulations”.
[208] T. Wang, C. Reinhart, and Y. Q. Ang, “sat2shp: Extracting key building features from a single satellite image for urban building energy modelling and beyond,” *Sustain. Cities Soc.*, vol. 118, p. 106054, Jan. 2025, doi: 10.1016/j.scs.2024.106054.
[209] A. Anand and C. Deb, “The potential of remote sensing and GIS in urban building energy modelling,” *Energy Built Environ.*, vol. 5, no. 6, pp. 957–969, Dec. 2024, doi: 10.1016/j.enbenv.2023.07.008.
[210] Y. Li and H. Feng, “Integrating urban building energy modeling (UBEM) and urban-building environmental impact assessment (UB-EIA) for sustainable urban development: A comprehensive review,” *Renew. Sustain. Energy Rev.*, vol. 213, p. 115471, May 2025, doi: 10.1016/j.rser.2025.115471.

[211] “Mapping Spatiotemporal Disparities in Residential Electricity Inequality Using Machine Learning | Environmental Science & Technology.” Accessed: Nov. 04, 2025. [Online]. Available: https://pubs.acs.org/doi/abs/10.1021/acs.est.4c06093
[212] X. Ma *et al.*, “A comprehensive review of the development of land use regression approaches for modeling spatiotemporal variations of ambient air pollution: A perspective from 2011 to 2023,” *Environ. Int.*, vol. 183, p. 108430, Jan. 2024, doi: 10.1016/j.envint.2024.108430.
[213] M. Ashayeri, N. Abbasabadi, M. Heidarinejad, and B. Stephens, “Predicting intraurban PM2.5 concentrations using enhanced machine learning approaches and incorporating human activity patterns,” *Environ. Res.*, vol. 196, p. 110423, May 2021, doi: 10.1016/j.envres.2020.110423.
[214] “Prototype Building Models | Building Energy Codes Program.” Accessed: Dec. 16, 2024. [Online]. Available: https://www.energycodes.gov/prototype-building-models#Residential
[215] “Engineering Reference”.
[216] J. Kneifel, “Prototype residential buildings for energy and sustainability assessment,” National Institute of Standards and Technology, Gaithersburg, MD, NIST TN 1688, 2011. doi: 10.6028/NIST.TN.1688.
[217] “WMO Region 4 - North and Central America.” Accessed: Jan. 20, 2025. [Online]. Available: https://climate.onebuilding.org/WMO_Region_4_North_and_Central_America/default.html
[218] “OSMnx Paper – Geoff Boeing.” Accessed: Mar. 10, 2025. [Online]. Available: https://geoffboeing.com/publications/osmnx-paper/
[219] “Zoning.” Accessed: Nov. 06, 2025. [Online]. Available: https://geohub.lacity.org/datasets/lahub::zoning/explore
[220] CDC, “SVI Data & Documentation Download,” Place and Health - Geospatial Research, Analysis, and Services Program (GRASP). Accessed: Jul. 15, 2025. [Online]. Available: https://www.atsdr.cdc.gov/place-health/php/svi/svi-data-documentation-download.html
[221] “Social Vulnerability Index 2020.” Accessed: Jul. 15, 2025. [Online]. Available: https://geohub.lacity.org/datasets/lacounty::social-vulnerability-index-2020/about
[222] “Climate Calculations: Engineering Reference — EnergyPlus 9.2.” Accessed: Jul. 23, 2025. [Online]. Available: https://bigladdersoftware.com/epx/docs/9-2/engineering-reference/climate-calculations.html
[223] “Residential Electric Rates | Los Angeles Department of Water and Power.” Accessed: Nov. 17, 2025. [Online]. Available: https://www.ladwp.com/account/understanding-your-rates/residential-electric-rates
[224] D. Bhattarai, A. Lucieer, H. Lovell, and J. Aryal, “Remote sensing of night-time lights and electricity consumption: A systematic literature review and meta-analysis,” *Geogr. Compass*, vol. 17, no. 4, p. e12684, 2023, doi: 10.1111/gec3.12684.
[225] R. Avtar *et al.*, “Exploring Renewable Energy Resources Using Remote Sensing and GIS—A Review,” *Resources*, vol. 8, no. 3, p. 149, Sep. 2019, doi: 10.3390/resources8030149.

[226] H. Li, Y. Xu, and T. Hong, “EnergyPlus-MCP: A model-context-protocol server for ai-driven building energy modeling,” *SoftwareX*, vol. 32, p. 102367, Dec. 2025, doi: 10.1016/j.softx.2025.102367.