urban science

MDPI

Article

A Framework for Augmenting Simulation-Based Building Energy Models with Earth Observational Microclimate Data Using Machine Learning Predictions

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

[1] Department of Civil and Environmental Engineering, University of Washington, Seattle, WA 98195, USA
[2] School of Architecture, Southern Illinois University, Carbondale, IL 62901, USA
[3] Department of Architecture, University of Washington, Seattle, WA 98195, USA
* Correspondence: aworthy@uw.edu

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 geographical-based urban morphology data, which are enhanced through machine learning techniques to improve energy demand predictions in urban settings. Applied to Los Angeles (LA), California, 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 by 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

Check for updates

Academic Editor: Laura Cirrincione

Received: 27 February 2026
Revised: 18 April 2026
Accepted: 14 May 2026
Published: 23 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

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 in network density has altered the surface energy balances, thermal properties [5,6], 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 known as the urban heat island (UHI) effect [12,13], where cities experience local warming greater than that in the surrounding less-developed areas. These altercations through urban form creates 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 urban environments and infrastructure is pivotal to achieve more sustainable and resilient development in both present and future cities.

Urban microclimate conditions have a profound influence on city infrastructure, specifically energy consumption [14]. 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 [15], 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 concerns 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 [16]. Additionally, low-income U.S. census blocks have 15.2% less tree cover and experience temperatures that are 1.5 °C hotter than that in high-income blocks [17]. Since dense vegetation provides roughly 6% energy savings in buildings, similar to the magnitude of savings from mechanical retrofit interventions [18], this is not only an issue of environmental exposure but also 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 affects renters [19], populations who are often more historically disadvantaged and lower-income. Therefore, this research project aims to not only 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 [20,21].

Urban building energy modeling (UBEM) provides 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) [22]. Using bottom-up or top-down methods, while referencing building operational constraint definitions, UBEM models formulate energy demand predictions for a collection of urban buildings [23]. Model inputs include building geometric data (ex. shape, size, and height), building non-geometric data (occupancy loadings), and weather data [24].

UBEM climate and environmental data inputs commonly comprise hourly weather observations that are sampled over a 12-meteorological month calendar. These are often in typical meteorological year (TMY) files [25], which are compatible with energy plus engines in the energy plus weather, epw, format. Although applauded for their high temporal granularity (hourly), these files source historically aggregated climate and weather data measurements [25] 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 the building sites.

In addition to TMY datasets, UBEM microclimate inputs have been substitutionally sourced from simulation-based or observational-based datasets [22], which both differ in their capacity to accurately contextualize microclimate conditions [26]. The reliability

of weather data for building energy applications depends on the spatial and temporal scale of the weather file and the compatibility of the weather file with simulation-based building energy management platforms [26]. 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 [27], which can be directly coupled with building energy simulations [28]; however, they require substantial computational resources to achieve neighborhood-scale resolution [29] 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, can be used as observation-based climate data inputs. However, in situ station-based products are constrained by the resource requirements necessary to sample at both high spatial (ex. 30 m) and high temporal (ex. minutely and hourly) frequencies by challenges inherent to data-collection processes such as misalignment between building locations and sampling locations [29]. In contrast, Earth observation-based or satellite-derived products have insufficient temporal resolution, as satellite revisit periods are inconsistent (i.e., Landsat 8 passes every, on average, 16 days) and by the complex data schemas needed to couple raster-based datasets with building energy modeling tools.

1.1. Study Gaps and Opportunities

Satellite microclimate data has been integrated into UBEMs using top-down statistical methods throughout various studies [30]. For example, [31–34] 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, Refs. [26,31], which incorporated 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. First, using the building characteristics of a Department of Energy (DOE) archetype residential building, we leverage the hourly temporal capability of TMYx epw files—to simulate annual building energy consumption at all TMYx measurement locations across Los Angeles (LA), California. Then, we reference Earth observational microclimate and geographical-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 affect 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.

1.2. Primary Objectives

- Propose a hybrid framework to integrates Earth observational microclimate and geographical-based urban morphology characteristics into simulation-based building energy models.
- Create a detailed heatmap of residential building energy consumption across LA that accounts for urban context conditions such as urban microclimates and urban morphology characteristics.
- Determine the regions where residential building energy demands are elevated due to urban context conditions, and relate these results to zoning characteristics and population vulnerability, thereby informing more equitable infrastructure planning decisions.

2. Materials and Methods

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 geographical-based microclimate data, as shown in Figure 1. 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, California. 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 and building density) and microclimate conditions (e.g., land surface temperature and tree cover). This urban context model is then coupled with Phase 1 simulation outputs using machine learning approaches, powered by Random Forest, CatBoost, and Decision Tree 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.

2.1. Study Area

Los Angeles, California, USA, was chosen for this study as it has the most EnergyPlus weather file stations recorded on an EPWmap [35] of any U.S. city, and the city has notable microclimate variability that is influenced by urban contexts. Los Angeles is situated on the West Coast of the United States. It is the largest city in the State of California and the 2nd most populated city in the U.S. [36], with 3.87 million people (2026 est). The city has a land area of 469.49 square miles [37], and the overall climate is marine moderate with an average annual high temperature of 72 °F, an average annual low of 57 °F, and an average of 30 days with precipitation per year [38].

2.2. Data

Appropriate data inputs were identified to meet the study goals and then tabulated and collected using automated approaches in python 3.12 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 geographical-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 using data-driven methods and were referenced to predict annual building energy consumption across LA using machine learning techniques. Social vulnerability and city planning data is integrated into this study to determine if energy infrastructure is equitably distributed across different urban neighborhoods.

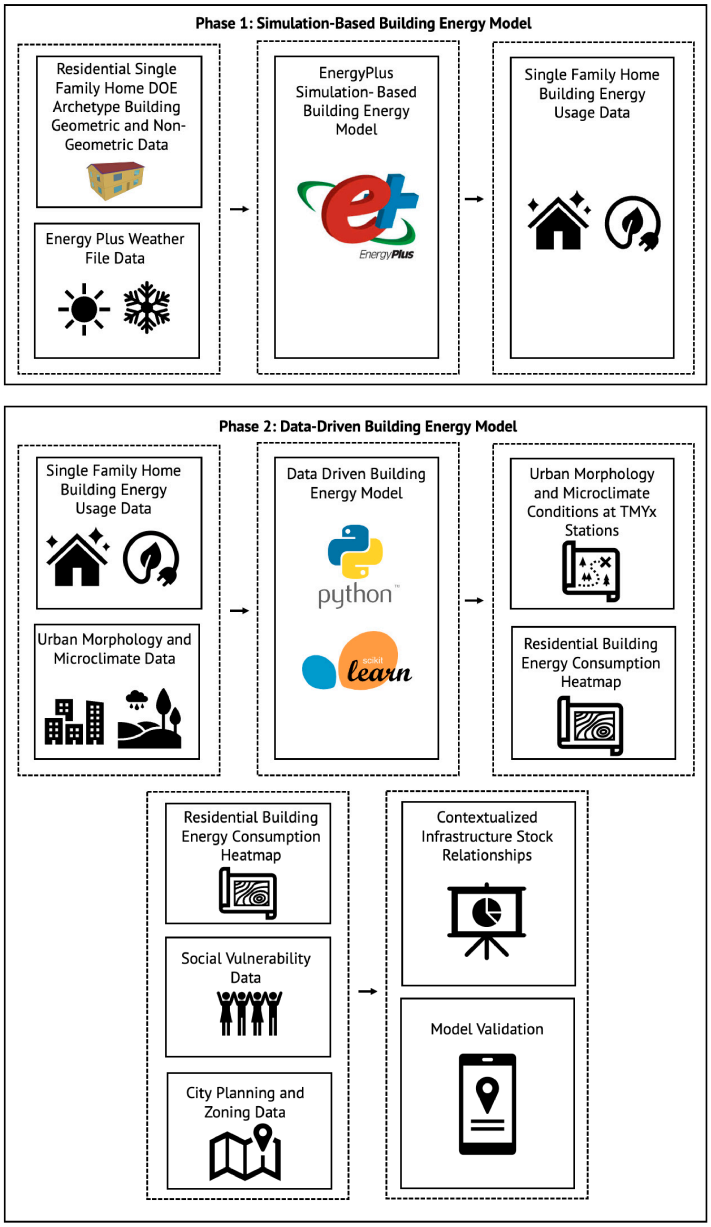

Figure 1. Methodological diagram for the hybrid study.

2.2.1. Building Geometric and Nongeometric Data

To better understand the impacts of urban contexts and microclimates on residential building energy consumption, U.S. Department of Energy (DOE) prototype residential buildings are referenced [39], with the example home shown in Figure 2. The US + SF + CZ3C + gasfurnace + crawlspace + IECC_2024.idf file was used in this study. This file describes building characteristics for a single-family home with a gas furnace as the primary heating element and an attached crawl space as an add on. The building is built to code according to the 2024 International Energy Conservation Code (IECC) requirements, with systems and building enclosure construction compatible for a building in the Marine

(3C) climate zone. The ASHRAE climate zone 3C represents the main climate category for the coastal areas in California, or a mild marine climate, with 99% of temperatures observed contained within 5 °C to 27 °C [40]. This file, which specifies the building geometry, materials, mechanical systems, schedules, zones, and equipment, was used to define both geometric and non-geometric building inputs in the EnergyPlus 23.1.0 simulation-based model in Phase 1 of this study. It should be noted that all default energy balance equations were referenced, as documented in the EnergyPlus Engineering Reference Manual (version 23.1.0) [41].

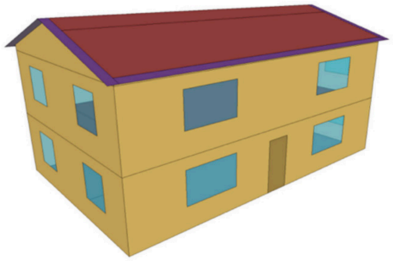

Figure 2. Single-family home DOE protype building, figure from [42].

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'. It should be noted that no other defaults were changed.

2.2.2. Meteorological and Urban Context Data

Two different meteorological and urban context datasets were compiled. The first, outlined in Energy Plus Weather File Climate Data, comprises (.epw) files collected throughout LA city limits, sourced from TMYx observations using diyepw, a python package made to facilitate (.epw) file retrieval and generation [43]. These files were used as inputs in the simulation-based building energy model during Phase 1 of this research. The second dataset, outlined in Urban Morphology and Urban Microclimate Data, comprises geographical-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.

Energy Plus Weather File Climate Data

To integrate weather files into EnergyPlus 23.1.0 software, one of the most widely used simulation-based building energy modeling platforms developed by the U.S. Department of Energy, the weather files must be provided in .epw format [44]. Accordingly, all available EnergyPlus weather (.epw) files within the Los Angeles County study-area bounding box hosted on EPWmap [35] were compiled for use as climate inputs in the simulation-based building energy model. A total of 19 typical meteorological year (TMYx) weather files were obtained from Climate.OneBuilding.org [45] and incorporated into the analysis, as shown in Figure 3.

Urban Morphology and Urban Microclimate Data

Different urban morphology and microclimate data products were collected and combined for Phase 2 of this study. These features were engineered and layered using Python and respective geospatial packages to create an urban context map of Los Angeles that provides geospatial insights into urban properties, layouts, and forms. The datasets compiled,

which are open-source and available in all U.S. cities, are outlined in Table 1 and comprise high-spatial-resolution meteorological, environmental, and physical characteristics.

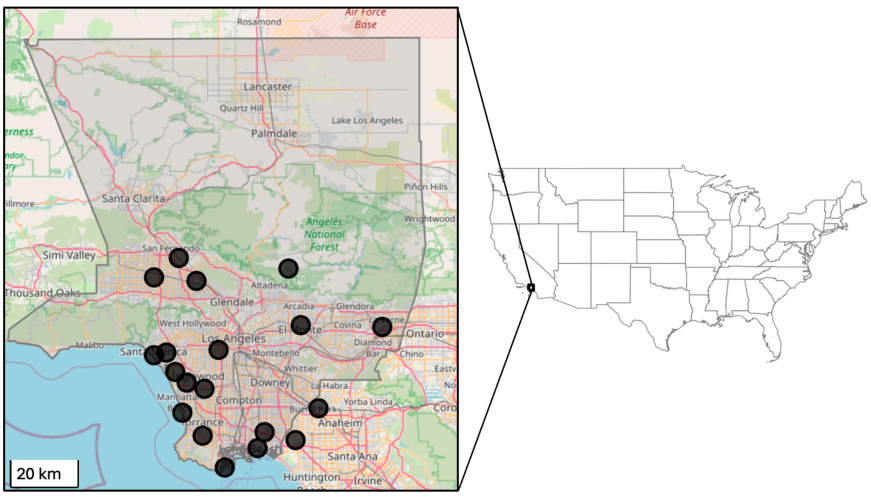

Figure 3. Map of TMYx epw-compatible sampling locations (black) in Los Angeles County, California, USA (gray).

Table 1. 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 [46]
Environmental Characteristics	NASA DEM	NA	2000 release	30 m	Elevation (m)	Google Earth Engine [47]
Environmental Characteristics	USFS Tree Canopy Coverage	NA	2021 release	30 m	Tree Canopy Cover (%)	Google Earth Engine [48]
Environmental Characteristics	USGS National Landcover Database	NA	2021 release	30 m	Landcover	Google Earth Engine [46]
Physical Characteristics (Streets)	Open Street Map (network = 'drive')	NA	2025 pull	Geometries	Geometry	Open Street Map (python OSMnx api) [49]
Physical Characteristics (Buildings)	Microsoft Buildings Dataset	NA	2020 release	Geometries	Geometry, Area, Centroid	Microsoft Buildings Dataset [50]

2.2.3. 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 [51], as tabulated in Table 2.

2.2.4. 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, as tabulated in Table 3. 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 [42].

Table 2. 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 [51]

Table 3. 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 [52]

2.3. Phase 1: Simulation-Based Building Energy Model

Phase 1 of this research combines residential building archetype characteristics from DOE archetype building .idf files and respective energy plus weather .epw files containing typical meteorological year climate and weather information, as tabulated in Section 2.2. 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 ASHRAE 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.

2.4. 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 including Landsat 8, and open-source building and street products, derived as building count and street length values, to create a geographical urban context map over LA city limits using Python geospatial software packages including geopandas, Xarray, shapely, sklearn, and rasterio. Then, it couples the urban context model with the Phase 1 simulation-based outputs using ensemble-based machine learning techniques (CatBoost, Decision Tree, and Random Forest algorithms) to thereby predict residential building energy consumption throughout different urban context conditions over a 500 m by 500 m grid of the city. 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.

2.4.1. Urban Context Map

Urban morphology and urban microclimate data products, outlined in Table 1 and Figure 4, are combined to create an urban context map. To better quantify the spatial influence of each data layer, concentric buffer zones were generated around each of the

EPW sampling locations (defined by a longitude and latitude coordinate), which were obtained from all stations within our study boundary limits, which are defined in the results section. These buffer radii span distances from 10 m to 1000 m in 30 m increments around all EPW stations. Values from each data product were then sampled and calculated at all respective buffer radii to evaluate scale-dependent contributions.

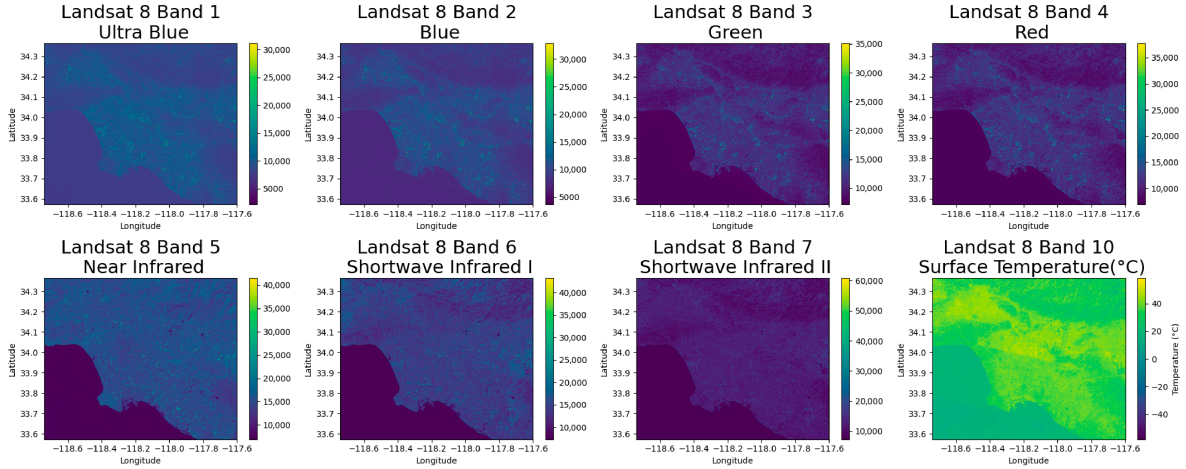

Figure 4. Urban context map final model data layers, representing 2023 annual average observations.

2.4.2. 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) as simulated in Phase 1, and the urban morphological and microclimate data layers sampled at each respective buffer length (including all environmental, physical, and meteorological covariates, as outlined in Table 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 [53] and is used to characterize local urban context effects. To perform this step, the Pearson correlation coefficient is calculated between annual energy consumption (EUI) and each urban morphological and microclimate data covariate (e.g., tree cover or land surface temperature) at all respective buffer lengths. It should be noted that the median value within each buffer was calculated for all Landsat and DEM parameters, whereas the number of buildings and length of streets were calculated for the building and street data layers. Then, the optimal buffer length for each urban morphological and microclimate data covariate was chosen based on the buffer lengths capacity to maximize the absolute value of correlation with annual energy consumption (EUI).

Feature Selection

The 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 involves identifying relationships between energy usage and environmental characteristics and then removing derived synthetic data products from the analysis to reduce feature redundancy.

The Atmospheric Transmittance (ST_ATRAN), Downwelled Radiance (ST_DRAD), and Upwelled Radiance (ST_URAD) bands are excluded from the 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 [54]. The building count parameter, derived from the Microsoft buildings dataset, the USFS tree canopy coverage dataset, and the USGS Landcover dataset, was also excluded from the predictive model, as these products themselves are trained on other model covariates and may therefore introduce redundancies and dependencies.

To validate the use of TMYx files, it is also necessary to determine if the environmental and morphological conditions at the (.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 the results section. The street length variable is excluded from the predictive model due to its discrepancy in representation for training dataset parameter characterization, as street length mean values across the TMYx stations were reported to be 21,000,000% lower than the Los Angeles average, which is further discussed in Section 3.2.

A description of the final predictive model attributes, having undergone buffer size identification and feature selection, is displayed in Table 4. The refined subset has 8 features, including Landsat 8 Band 1—Ultra Blue; Band 2—Blue; Band 3—Green; Band 4—Red; Band 5—Near Infrared; Band 6—Shortwave Infrared 1; Band 7—Shortwave Infrared 2; and Band 10—Surface Temperature.

Table 4. 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) [49]
	Buildings—centroid *	790	# of	Microsoft Buildings Dataset [50]
Environmental Characteristics	Landcover *	280	-	USGS National Landcover Database 2021 Release [46]
	Tree Canopy Cover *	190	%	USFS Tree Canopy Cover v2021-4 [48]
	Digital Elevation *	130	m	NASA DEM [47]
	Band 1, Ultra Blue (SR_B1)	40	-	USGS LANDSAT 8 Level 2, Collection 2, Tier 1 [55]
	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 [55]
	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.

Model Prediction

To better assess 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 (such as max depth, learning rate, and random state) to optimize model generalizability, with the final hyperparameters. Tree-based methods (Decision Tree, Random Forest, and CatBoost) were evaluated in this study, having been identified effective in [29,56,57]. Model performance is evaluated using two metrics [58]: 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 models with the optimal hypertuned parameters, as identified above, were run to generate residential building EUI heatmaps over a 500 m $\times$ 500 m spatial grid of the city. For each grid cell, predictor variables were sampled at the cell centroid.

2.5. 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.

2.6. 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. Results

This section presents the results of predicting residential building energy consumption under varying urban context and microclimate conditions using the augmented modeling framework, which was described in Section 2.

3.1. Phase 1: Simulation-Based Energy Model

Nineteen independent energy plus simulations were run with DOE residential building archetype (.idf) files at all given energy plus weather stations inside of the study boundaries, with the weather station locations (latitude and longitude) representing the simulated building coordinates. The distribution of energy consumed at each building location is tabulated in Table 5 and shown in Figure 5.

Table 5. 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 in 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

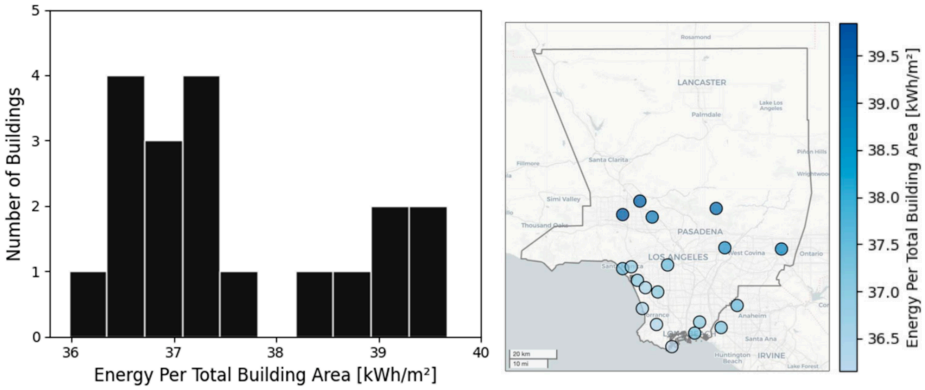

Figure 5. Histogram (**left**) and map (**right**) of Phase 1 single-family home EnergyPlus simulation outputs.

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 very highly urban developed regions.

To better interpret total building energy consumption, individual energy end uses are examined separately. The 19 simulated buildings have the exact same interior equipment (10,083 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 building is simulated with different epw file inputs. Building energy cooling demands range from 5.7% to 9.4% of total building energy consumption, with the highest demands at building ID #19, in the valley (see Figure 6), and the lowest demands at building ID #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 ID #18, in the mountains, and the lowest heating demands are at building ID #1, near the coast.

3.2. Comparison of Conditions in the City and at Sampling Locations

When comparing averages between the average TMYx sampling site and mean Los Angeles conditions observed, all Landsat parameter differences fall within 10%. This relationship highlights this dataset's competency for model generalization, as 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, with these averages reported in Table 6. This finding calls for future TMYx stations in Los Angeles to be placed in city regions with greater densities of streets, trees, and buildings to fully contextualize the environmental conditions that are observed by the general public.

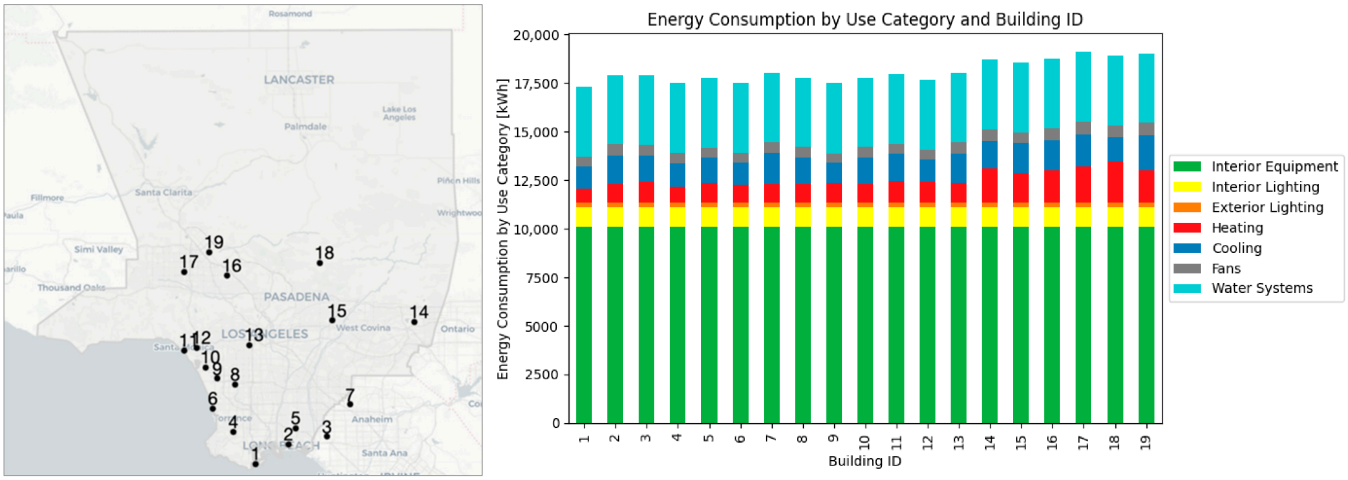

Figure 6. Reference map of buildings with corresponding building ID numbers denoted at their locations, with IDs numbered North to South in latitude (**left**), and bar plots of EnergyPlus simulation results: energy consumption by end use category, ordered by building ID number (**right**).

Table 6. 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	21,000,000
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)	10,570	9834	−6.96
SR B2 (40)	10,980	10,227	−6.85
SR B3 (40)	11,983	11,220	−6.35
SR B4 (40)	12,402	11,497	−7.29
SR B5 (850)	14,990	15,414	2.83
SR_B6 (70)	15,479	14,178	−8.4
SR B7 (70)	13,973	12,487	−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

A green highlight indicates the larger observed value between the TMYx Location Mean and the Los Angeles Mean.

3.3. Energy Use Prediction Model Performance and Heatmaps

Single building EnergyPlus simulations are enhanced using Earth observational datasets and machine learning algorithms (Random Forest, CatBoost, and Decision Tree) to create predictive heatmaps of residential building energy usage, augmented with urban microclimate conditions, across Los Angeles, USA, as shown in Figure 7 with the methodological steps tabulated in Section 2. Table 7 outlines the modeling algorithms referenced for these predictions, their hyperparameters, their respective modeling performance, and their reported percent difference in range over the predicted LA modeling grid.

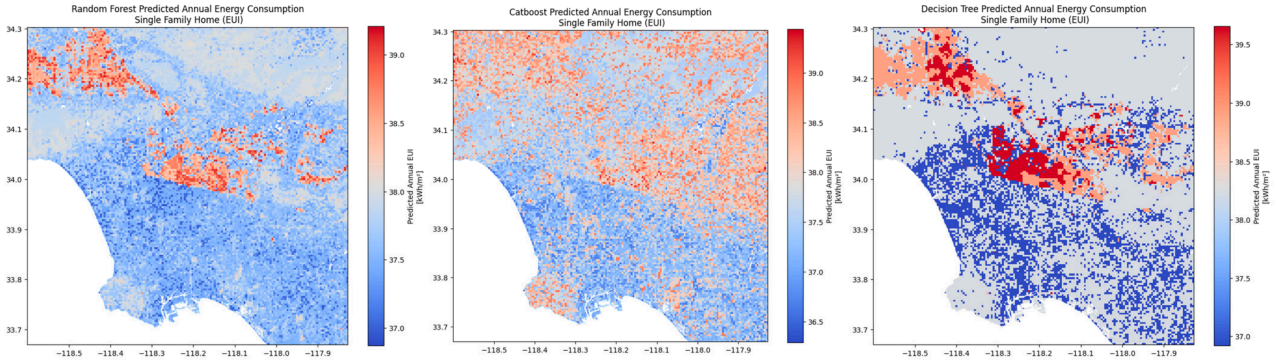

Figure 7. Predicted single-family home total energy consumption heatmaps for Random Forest (**left**), CatBoost (**middle**), and Decision Tree (**right**) models.

Table 7. Performance of different building energy consumption prediction models.

Method	Hyperparameters	Training R^2	Testing R^2	Training RMSE	Testing RMSE	Percent Difference in 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

The energy consumption heatmap predictions contextualize high-energy-usage and medium-energy-usage areas similarly across the Random Forest and Decision Tree algorithms. The high-use areas that are located in the middle of both of these maps geo-correspond directly to downtown Los Angeles, California, which is hypothesized to have higher cooling energy demands due to the urban heat island effect. When examining the Earth observational parameter inputs that derive these predictions (in Figure 4), both of these high-use areas record visually higher Landsat Band 10—or land surface temperature—observations, which is directly related to city landcover and development stage (i.e., park, downtown). It should be noted that the CatBoost model predictions differ quite widely from the tree-based predictions, as these rely less heavily on the Landsat Band 10 (25%) covariate compared to the Random Forest (50%) and Decision Tree (60%) algorithms, as discussed more in Section 3.4.

Moreover, because the CatBoost model draws on a more equally distributed feature set than anchoring predictions to Landsat 8 Band 10, fluctuations across secondary covariates create sharp swings in predicted values across adjacent pixels. This noise, combined with the model's training and testing R^2 values of 0.99 and 0.56, respectively, suggests that the CatBoost model is overtrained relative to the Random Forest (R^2 train: 0.85, test: 0.56) and Decision Tree (R^2 train: 0.69, test: 0.61) models, inferring that it should be interpreted with greater caution than the tree-based alternatives. That is, the Random Forest and Decision Tree models should be considered more reliable models, as their lower training R^2 values indicate reduced risk of overfitting (as shown in Table 7), their predictions are spatially coherent, and their stronger reliance on land surface temperature as a dominant covariate aligns more directly with the literature.

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, and Decision Tree model heatmaps indicate 6.3%, 8.7%, and 7.4% respective discrepancies 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 because we reference ensemble- and tree-based methods that constrain the testing data to the training data limits [59]. While CatBoost produces the largest predicted range (8.7%) and therefore most closely approaches the training data ceiling, Random Forest yields the most conservative estimate (6.3%), consistent with its tendency toward variance reduction through ensemble averaging. These patterns suggest that CatBoost offers the most spatially differentiated predictions, while the Random Forest and Decision Tree models provide more stable but potentially range-compressed predictions.

3.4. Feature Importance

Feature importance analysis finds the Landsat 8 Band 10, 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, and Decision Tree models. This is because the land surface temperature band has the highest relative feature importance, as shown in Figure 8.

It is important to note that, although land surface temperature is related to air surface temperature, these are fundamentally different physical measurements [26]. Despite this distinction, land surface temperature provides valuable context to building energy consumption studies and should be further investigated for reference in urban-scale energy usage, specifically when in situ measurements are limited, as this remotely sensed parameter is open-source, measured at a 30 m spatial scale across the world, and is proven to have direct correlations with building heating and cooling energy demands.

3.5. City Planning Insights and Model Verification

To better understand the impacts of urban microclimates on city energy usage 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 [61], where 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 9, 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 the 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.

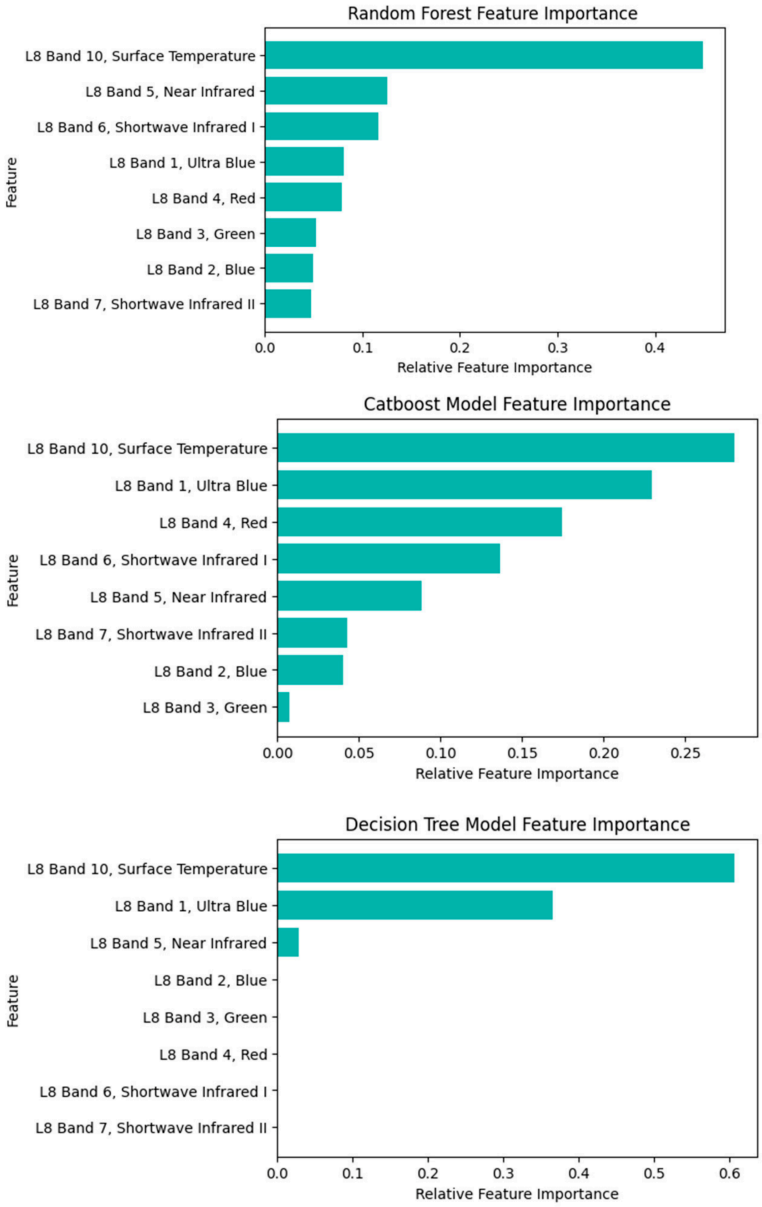

Figure 8. Predictive model variable importance plots for Random Forest (**top**), CatBoost (**middle**), and Decision Tree (**bottom**) models.

Figure 9. 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.

3.6. 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, with the characteristics outlined in Section 2.2.4. 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 10.

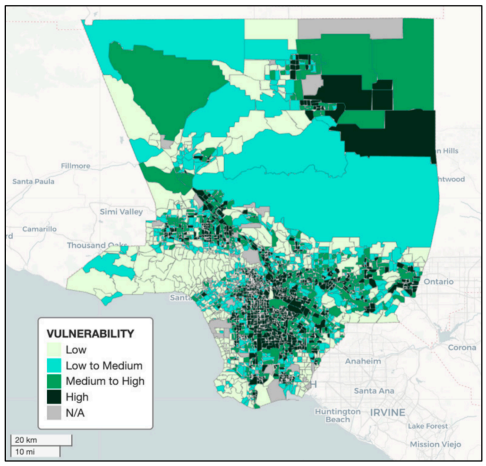

Figure 10. Vulnerability index category of census tracts in LA, data from [52].

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 11. These aggregations show surprisingly similar average single-family home EUIs across all predicted models and vulnerability indices.

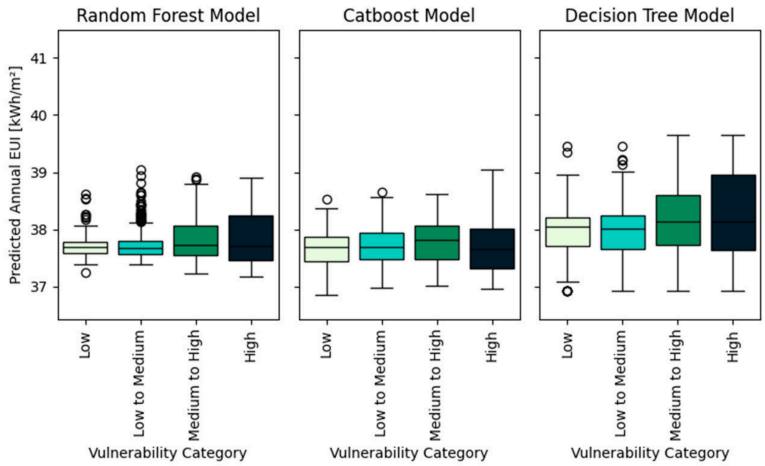

Figure 11. Distribution of predicted single-family home annual EUI by model and vulnerability index.

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 8. Moreover, while 18% of census tracts are in low-vulnerability 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 the ability of our analysis and predictive models to fully contextualize extreme microclimate environments. Therefore, we call for future placement of TMYx sampling sites to be in high-vulnerability regions,

specifically high-density urban areas, which have been documented to experience different environmental conditions than their rural counterparts.

Table 8. 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).

4. 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 with 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. This station placement bias is consistent with longstanding critiques of airport- and rural-sited weather stations, which have been shown to systematically underestimate urban thermal conditions relative to dense city conditions [26,62,63]; however, our study highlights these discrepancies numerically.

We find that there is less representation of TMYx sensing stations in high-population-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 their corresponding energy burden. This finding has especially large implications for cities with large income disparities currently, as present-day income inequality is found to be a stronger predictor of heat burden than the legacy of redlining [64]. Moreover, monitoring network design is a highly important concern for air quality research [59], with even congressional acts to back its importance [65], and should be heavily considered for energy burden research too.

Additionally, we find the Landsat 8 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 for household building energy consumption predication tasks when in situ weather station measurements are limited. 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 relatively limited GIS and python experience [26,31].

Significant land surface temperature band contributions are consistent with the results tabulated in a building energy and microclimate studies in New York City [35] and Seattle [38]. Moreover, in Los Angeles, land surface temperature and median household income have been found to have persistent negative correlations, raising a use case to include these products to also assess energy burden across communities. Collectively, these

efforts demonstrate promise for combining remote sensing microclimate data into UBEM to evaluate city-wide energy demands and distributions [64].

Our Random Forest and Decision Tree models clearly predict the homes in more densely developed areas in Los Angeles (aka downtown) to consume more energy annually than those situated in less developed, natural areas. This result is consistent with findings in Los Angeles, which link a feedback loop of anthropogenic heating from buildings to elevated local air temperatures across dense urban areas [66]. Moreover, these results are consistent with the urban heat island literature, which link increased urban temperatures with increased urban development, and should be of upmost concern as urban heat islands continue to grow [67].

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 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 nearby energy plus weather file will extrapolate to your exact building location. If these compared measurements differ greatly, but are in the same location, we recommend conducting additional analyses, 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 by 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 [26,68,69].

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 existing (.epw) files. For instance, the model's generalizability is constrained by the data that it is trained upon; therefore, the model shows decreased reliability when the actual conditions are not contained in the training dataset (ie: when values are extrapolated). Moreover, the study's 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, and country). Lastly, in this framework, we predict 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 on the spatial scale of the 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 compatible datasets that measure TMYx parameters and respective downscaling opportunities.

5. Conclusions

In this study, we augment EnergyPlus building simulations with satellite-derived microclimate datasets, ultimately creating heatmaps of residential building energy usage that are corrected for urban climate conditions across Los Angeles, California. 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—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. Moreover, recording stations are placed in locations with much fewer streets (21,000,000%), buildings (45%), and trees (185%) 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. Lastly, 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 by incorporating microclimate data into building energy models, aiming to advance sustainable energy-driven improvements at the city level.

Author Contributions: A.W.: Writing—original draft, Conceptualization, Visualization, Methodology, Formal analysis, Data curation, Validation. M.A.: Writing—review and editing, Supervision, Concep-

tualization. J.D.M.: Writing—review and editing, Supervision, Project administration, Conceptualization. N.A.: Writing—review and editing, Supervision, Project administration, Conceptualization. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the University of Washington Clean Energy Institute Graduate Fellowship and the NASA Washington Space Grant Dissertation Completion Award.

Institutional Review Board Statement: Not in applicable.

Informed Consent Statement: Not in applicable.

Data Availability Statement: Data is available upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Kousis, I.; Manni, M.; Pisello, A.L. Environmental Mobile Monitoring of Urban Microclimates: A Review. *Renew. Sustain. Energy Rev.* **2022**, *169*, 112847. [CrossRef]
2. Rickwood, P.; Glazebrook, G.; Searle, G. Urban Structure and Energy—A Review. *Urban Policy Res.* **2008**, *26*, 57–81. [CrossRef]
3. Frolking, S.; Mahtta, R.; Milliman, T.; Esch, T.; Seto, K.C. Global Urban Structural Growth Shows a Profound Shift from Spreading out to Building Up. *Nat. Cities* **2024**, *1*, 555–566. [CrossRef]
4. Nichol, J.E. High-Resolution Surface Temperature Patterns Related to Urban Morphology in a Tropical City: A Satellite-Based Study. *J. Appl. Meteorol. Climatol.* **1996**, *35*, 135–146. [CrossRef]
5. Duveiller, G.; Hooker, J.; Cescatti, A. The Mark of Vegetation Change on Earth's Surface Energy Balance. *Nat. Commun.* **2018**, *9*, 679. [CrossRef]
6. Schwaab, J.; Meier, R.; Mussetti, G.; Seneviratne, S.; Bürgi, C.; Davin, E.L. The Role of Urban Trees in Reducing Land Surface Temperatures in European Cities. *Nat. Commun.* **2021**, *12*, 6763. [CrossRef] [PubMed]
7. Manoli, G.; Fatichi, S.; Schläpfer, M.; Yu, K.; Crowther, T.W.; Meili, N.; Burlando, P.; Katul, G.G.; Bou-Zeid, E. Magnitude of Urban Heat Islands Largely Explained by Climate and Population. *Nature* **2019**, *573*, 55–60. [CrossRef]
8. Yang, Y.; Gu, Q.; Wei, H.; Liu, H.; Wang, W.; Wei, S. Transforming and Validating Urban Microclimate Data with Multi-Sourced Microclimate Datasets for Building Energy Modelling at Urban Scale. *Energy Build.* **2023**, *295*, 113318. [CrossRef]
9. Juan, Y.-H.; Wen, C.-Y.; Li, Z.; Yang, A.-S. Impacts of Urban Morphology on Improving Urban Wind Energy Potential for Generic High-Rise Building Arrays. *Appl. Energy* **2021**, *299*, 117304. [CrossRef]
10. Hong, T.; Xu, Y.; Sun, K.; Zhang, W.; Luo, X.; Hooper, B. Urban Microclimate and Its Impact on Building Performance: A Case Study of San Francisco. *Urban Clim.* **2021**, *38*, 100871. [CrossRef]
11. Ouyang, Z.; Sciusco, P.; Jiao, T.; Feron, S.; Lei, C.; Li, F.; John, R.; Fan, P.; Li, X.; Williams, C.A.; et al. Albedo Changes Caused by Future Urbanization Contribute to Global Warming. *Nat. Commun.* **2022**, *13*, 3800. [CrossRef]
12. Rajagopal, P.; Priya, R.S.; Senthil, R. A Review of Recent Developments in the Impact of Environmental Measures on Urban Heat Island. *Sustain. Cities Soc.* **2023**, *88*, 104279. [CrossRef]
13. Oke, T.R. The Energetic Basis of the Urban Heat Island. *Q. J. R. Meteorol. Soc.* **1982**, *108*, 1–24. [CrossRef]
14. Li, X.; Zhou, Y.; Yu, S.; Jia, G.; Li, H.; Li, W. Urban Heat Island Impacts on Building Energy Consumption: A Review of Approaches and Findings. *Energy* **2019**, *174*, 407–419. [CrossRef]
15. Krelling, A.F.; Lamberts, R.; Malik, J.; Zhang, W.; Sun, K.; Hong, T. 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.* **2024**, *107*, 105460. [CrossRef]
16. Hsu, A.; Sheriff, G.; Chakraborty, T.; Manya, D. Disproportionate Exposure to Urban Heat Island Intensity across Major US Cities. *Nat. Commun.* **2021**, *12*, 2721. [CrossRef]
17. McDonald, R.I.; Biswas, T.; Sachar, C.; Housman, I.; Boucher, T.M.; Balk, D.; Nowak, D.; Spotswood, E.; Stanley, C.K.; Leyk, S. The Tree Cover and Temperature Disparity in US Urbanized Areas: Quantifying the Association with Income across 5,723 Communities. *PLoS ONE* **2021**, *16*, e0249715. [CrossRef]
18. Excell, L.E.; Jain, R.K. Examining the Impact of Energy Efficiency Retrofits and Vegetation on Energy Performance of Institutional Buildings: An Equity-Driven Analysis. *Appl. Energy* **2024**, *357*, 121722. [CrossRef]
19. Ashbaugh, M.; Kittner, N. Addressing Extreme Urban Heat and Energy Vulnerability of Renters in Portland, OR with Resilient Household Energy Policies. *Energy Policy* **2024**, *190*, 114143. [CrossRef]
20. Excell, L.E.; Nutkiewicz, A.; Jain, R.K. Multi-Scale Retrofit Pathways for Improving Building Performance and Energy Equity across Cities: A UBEM Framework. *Energy Build.* **2024**, *324*, 114931. [CrossRef]

21. Abbasabadi, N.; Ashayeri, M. Socioeconomic Determinants of Public Health and Residential Building Energy Use in Chicago. In *Proceedings of the UIA 2021 RIO: 27th World Congress of Architects*; The Association of Collegiate Schools of Architecture (ACSA): Washington, DC, USA, 2021; pp. 707–713.
22. Hong, T.; Chen, Y.; Luo, X.; Luo, N.; Lee, S.H. Ten Questions on Urban Building Energy Modeling. *Build. Environ.* **2020**, *168*, 106508. [CrossRef]
23. Lamberts, R.; Hensen, J.L. *Building Performance Simulation for Design and Operation*; Spoon Press: London, UK, 2011; ISBN 0-415-47414-0.
24. Dabirian, S.; Panchabikesan, K.; Eicker, U. Occupant-Centric Urban Building Energy Modeling: Approaches, Inputs, and Data Sources—A Review. *Energy Build.* **2022**, *257*, 111809. [CrossRef]
25. Elnabawi, M.H.; Hamza, N. A Methodology of Creating a Synthetic, Urban-Specific Weather Dataset Using a Microclimate Model for Building Energy Modelling. *Buildings* **2022**, *12*, 1407. [CrossRef]
26. Worthy, A.; Ashayeri, M.; Marshall, J.; Abbasabadi, N. Bridging the Simulation-to-Reality Gap: A Comprehensive Review of Microclimate Integration in Urban Building Energy Modeling (UBEM). *Energy Build.* **2025**, *331*, 115392. [CrossRef]
27. Hong, T.; Langevin, J.; Sun, K. Building Simulation: Ten Challenges. *Build. Simul.* **2018**, *11*, 871–898. [CrossRef]
28. Sezer, N.; Yoonus, H.; Zhan, D.; Wang, L.L.; Hassan, I.G.; Rahman, M.A. Urban Microclimate and Building Energy Models: A Review of the Latest Progress in Coupling Strategies. *Renew. Sustain. Energy Rev.* **2023**, *184*, 113577. [CrossRef]
29. Abbasabadi, N.; Ashayeri, M. Urban Energy Use Modeling Methods and Tools: A Review and an Outlook. *Build. Environ.* **2019**, *161*, 106270. [CrossRef]
30. Li, Y.; Feng, H. 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.* **2025**, *213*, 115471. [CrossRef]
31. Worthy, A.; Ashayeri, M.; Abbasabadi, N. Leveraging Earth Observational Data Products and Machine Learning to Enhance Urban Building Energy Modeling (UBEM) with Microclimate Effects. *Sustain. Cities Soc.* **2025**, *130*, 106544. [CrossRef]
32. Dougherty, T.R.; Jain, R.K. TOM.D: Taking Advantage Of Microclimate Data for Urban Building Energy Modeling. *Adv. Appl. Energy* **2023**, 100138. [CrossRef]
33. Dougherty, T.R.; Jain, R.K. Invisible Walls: Exploration of Microclimate Effects on Building Energy Consumption in New York City. *Sustain. Cities Soc.* **2023**, *90*, 104364. [CrossRef]
34. Yu, Y.; Li, X.; Hsu, A.; Kittner, N. Mapping Spatiotemporal Disparities in Residential Electricity Inequality Using Machine Learning. *Environ. Sci. Technol.* **2024**, *58*, 19999–20008. [CrossRef] [PubMed]
35. EPW Map. Available online: https://www.ladybug.tools/epwmap/ (accessed on 27 February 2026).
36. World Population Review. Largest US Cities by Population 2026. Available online: https://worldpopulationreview.com/us-cities (accessed on 8 April 2026).
37. City of Los Angeles, Los Angeles County, California. Available online: https://www.laalmanac.com/LA/index.php (accessed on 8 April 2026).
38. Weather Averages Los Angeles, California. Available online: https://www.usclimatedata.com/climate/los-angeles/california/united-states/usca1339 (accessed on 17 November 2025).
39. Building Energy Codes Program. Prototype Building Models. Available online: https://www.energycodes.gov/prototype-building-models#Residential (accessed on 16 December 2024).
40. Climate Zone 3C: Insulation R-Values, Design Temps & HVAC Guide. Available online: https://autohvac.ai/climate-zones/zone/3c (accessed on 27 March 2026).
41. EnergyPlus Version 23.1 Documents. Available online: https://bigladdersoftware.com/epx/docs/23-1/ (accessed on 25 February 2026).
42. CDC SVI Data & Documentation Download. Available online: https://www.atsdr.cdc.gov/place-health/php/svi/svi-data-documentation-download.html (accessed on 15 July 2025).
43. Smith, A.D.; Stürmer, B.; Thurber, T.; Vernon, C.R. Diyepw: A Python Package for Do-It-Yourself EnergyPlus Weather File Generation. *J. Open Source Softw.* **2021**, *6*, 3313. [CrossRef]
44. EnergyPlus. Available online: https://energyplus.net/ (accessed on 23 May 2024).
45. Climate.OneBuilding. WMO Region 4 - North and Central America. Available online: https://climate.onebuilding.org/WMO_Region_4_North_and_Central_America/default.html (accessed on 20 January 2025).
46. NLCD 2021: USGS National Land Cover Database, 2021 Release | Earth Engine Data Catalog. Available online: https://developers.google.com/earth-engine/datasets/catalog/USGS_NLCD_RELEASES_2021_REL_NLCD (accessed on 30 July 2024).
47. NASADEM: NASA NASADEM Digital Elevation 30m | Earth Engine Data Catalog. Available online: https://developers.google.com/earth-engine/datasets/catalog/NASA_NASADEM_HGT_001 (accessed on 30 July 2024).
48. USFS Tree Canopy Cover V2021-4 (CONUS and OCONUS) | Earth Engine Data Catalog. Available online: https://developers.google.com/earth-engine/datasets/catalog/USGS_NLCD_RELEASES_2021_REL_TCC_v2021-4 (accessed on 8 August 2024).

49. Boeing, G. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Comput. Environ. Urban Syst.* **2017**, *65*, 126–139. [CrossRef]
50. Microsoft. US Building Footprints. Available online: https://github.com/microsoft/USBuildingFootprints (accessed on 15 July 2025).
51. Zoning. Available online: https://geohub.lacity.org/datasets/lahub::zoning/explore (accessed on 6 November 2025).
52. Social Vulnerability Index 2020. Available online: https://geohub.lacity.org/datasets/lacounty::social-vulnerability-index-2020/about (accessed on 15 July 2025).
53. Ashayeri, M.; Abbasabadi, N.; Heidarinejad, M.; Stephens, B. Predicting Intraurban PM2.5 Concentrations Using Enhanced Machine Learning Approaches and Incorporating Human Activity Patterns. *Environ. Res.* **2021**, *196*, 110423. [CrossRef]
54. U.S. Department of Energy. *Climate Calculations: Engineering Reference—EnergyPlus 9.2*; U.S. Department of Energy: Washington, DC, USA, 2019.
55. USGS Landsat 8 Level 2, Collection 2, Tier 1 | Earth Engine Data Catalog. Available online: https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC08_C02_T1_L2 (accessed on 16 September 2024).
56. Amasyali, K.; El-Gohary, N.M. A Review of Data-Driven Building Energy Consumption Prediction Studies. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1192–1205. [CrossRef]
57. Alyakoob, A.; Hartono, S.; Johnson, T.; Middel, A. Estimating Cooling Loads of Arizona State University Buildings Using Microclimate Data and Machine Learning. *J. Build. Eng.* **2023**, *64*, 105705. [CrossRef]
58. Chicco, D.; Warrens, M.J.; Jurman, G. The Coefficient of Determination R-Squared Is More Informative than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation. *PeerJ Comput. Sci.* **2021**, *7*. [CrossRef]
59. Chen, X.; Yang, J. Urban Climate Monitoring Network Design: Existing Issues and a Cluster-Based Solution. *Build. Environ.* **2022**, *214*, 108959. [CrossRef]
60. City Planning Services | BuildLA. Available online: https://buildla.lacity.org/city-planning-services.html (accessed on 27 March 2026).
61. Residential Electric Rates | Los Angeles Department of Water and Power. Available online: https://www.ladwp.com/account/understanding-your-rates/residential-electric-rates (accessed on 17 November 2025).
62. Hashemi, F.; Mills, G. On the Impact of Urban Climate and Heat Islands on Building Energy Performance: A Critical Review. *Energy Build.* **2025**, *343*, 115946. [CrossRef]
63. Tang, Y.; Sun, T.; Luo, Z.; Omidvar, H.; Theeuwes, N.; Xie, X.; Xiong, J.; Yao, R.; Grimmond, S. Urban Meteorological Forcing Data for Building Energy Simulations. *Build. Environ.* **2021**, *204*, 108088. [CrossRef]
64. Ma, X.; Zou, B.; Deng, J.; Gao, J.; Longley, I.; Xiao, S.; Guo, B.; Wu, Y.; Xu, T.; Xu, X.; 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.* **2024**, *183*, 108430. [CrossRef] [PubMed]
65. Markey, E.J. S.2476—117th Congress (2021–2022): Environmental Justice Air Quality Monitoring Act of 2021. Available online: https://www.congress.gov/bill/117th-congress/senate-bill/2476/text (accessed on 12 April 2026).
66. Vahmani, P.; Luo, X.; Jones, A.; Hong, T. Anthropogenic Heating of the Urban Environment: An Investigation of Feedback Dynamics between Urban Micro-Climate and Decomposed Anthropogenic Heating from Buildings. *Build. Environ.* **2022**, *213*, 108841. [CrossRef]
67. LaDochy, S. Los Angeles' Urban Heat Island Continues to Grow: Urbanization, Land Use Change Influences. *J. Urban Environ. Eng.* **2021**. [CrossRef]
68. Bhattarai, D.; Lucieer, A.; Lovell, H.; Aryal, J. Remote Sensing of Night-Time Lights and Electricity Consumption: A Systematic Literature Review and Meta-Analysis. *Geogr. Compass* **2023**, *17*, e12684. [CrossRef]
69. Avtar, R.; Sahu, N.; Aggarwal, A.K.; Chakraborty, S.; Kharrazi, A.; Yunus, A.P.; Dou, J.; Kurniawan, T.A. Exploring Renewable Energy Resources Using Remote Sensing and GIS—A Review. *Resources* **2019**, *8*, 149. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.