Sustainable Cities and Society 90 (2023) 104364

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs

Invisible walls: Exploration of microclimate effects on building energy consumption in New York City

Thomas R. Dougherty, Rishee K. Jain *

Urban Informatics Lab, Department of Civil & Environmental Engineering, Stanford University, 473 Via Ortega, Stanford, CA 94305, United States

ARTICLE INFO

Keywords:
Microclimate
Urban energy consumption
Remote sensing

ABSTRACT

Reducing greenhouse gases from buildings forms the cornerstone of policy to mitigate the effects of climate change. However, the automation of urban-scale building energy modeling systems required to meet global urban demand has proven challenging due to the bespoke characteristics of each city. One such point of uniqueness between cities is that of urban microclimate, which may play a significant role in altering the performance of energy efficiency in buildings. This research proposes a way to rapidly collect urban microclimate data through satellite readings and climate reanalysis, enabling researchers to study the invisible walls of microclimate, which play a critical role in the buildings' energy consumption.

We demonstrate the utility of this data by composing an analysis against three years of monthly building energy consumption data from New York City. Our study highlights the significance of urban microclimates in decreasing the gas consumption of some buildings in New York by 71% and increasing the gas consumption in others by as much as 221%. Microclimates also seem to be responsible for the decrease of electricity consumption by up to 28.6% in regions or increases of up to 77% consumption in others.

1. Introduction

The built environment accounts for around 20% to 40% of the energy consumed in developed countries, with similar statistics for their proportion of greenhouse gas emissions (Pérez-Lombard et al., 2008). To meet climate goals in the Paris Agreement, urban areas worldwide have adopted policies to reduce their carbon emissions. However, large-scale energy modeling attempts are laden with uncertainties associated with building modeling, occupancy behavior, and socio-technical factors (Ali et al., 2021). A particularly challenging branch of modern research is to blend our understandings between the built environment and the natural world (Palmer et al., 2005), which directly segues into urban microclimates. In this vein, the influence of urban microclimates on energy consumption has been gaining attention in the research community (Özyavuz, 2013). Preliminary studies show potential deviations of up to 100% heating or 65% due entirely to urban microclimates (Hong et al., 2021).

This research explores the potential advantages of blending high-resolution climate reanalysis with remote sensing data into a statistical model of urban energy consumption for New York City. In doing so, we identify the key features from the data sets and estimate their effects on regional energy consumption within the city. This work results in a newfound capacity for high-quality, readily available data to enable cities with scarce building energy consumption data to estimate the effects of microclimate on building energy inefficiency. As New York City serves as a testbed for our analysis, the results of this study may be suitable for mid to large-sized cities with cold climates. Finally, we use the analysis results to define a set of "energy microclimates" (EMC) which offer an extended definition of urban microclimates to include features most pertinent to building energy consumption and with a higher time resolution. We use these microclimate zones to explore how buildings might shift between microclimates throughout the year.

2. Background

Building energy modeling has been utilized as a valuable tool in the design phase of buildings for its capacity to estimate the building's energy consumption, carbon emissions, and peak electricity demand (Reinhart & Cerezo Davila, 2016). Large-scale energy analysis often disregards the unique aspects of each building in favor of macroscopic grid planning, potentially missing valuable information (Cetin, 2013; Cetin, 2015). More regional models proposed instead attempt to inject pertinent information about each building by utilizing archetype buildings in a physics-based model (Cerezo et al., 2017). While recent works have shown temperature to be a significant variable in both data-driven modeling and physics-based modeling (Santamouris et al., 2015;

* Corresponding author.
E-mail address: rishee.jain@stanford.edu (R.K. Jain).

https://doi.org/10.1016/j.scs.2022.104364
Received 23 August 2022; Received in revised form 16 December 2022; Accepted 16 December 2022
Available online 22 December 2022

Luo et al., 2020; Torabi Moghadam et al., 2018), the incorporation of highly localized weather data is often neglected in practice. Instead, Typical Meteorological Year (TMY) data is used as a surrogate for the climate conditions likely to interact directly with the buildings (Xu et al., 2022; Reinhart & Cerezo Davila, 2016).

However, recent work has shown that TMY poorly tracks localized effects like urban heat islands (Weclawiak, 2022), and a lack of integration between building energy models and climate models is noted as one of the significant engineering obstacles to the continued improvement of UBEM (Craig et al., 2022). Recent research has begun to shed light into the true implications of localized weather on our energy infrastructure. One study couples the localized cooling effects of trees with building energy simulations, discovering that buildings directly adjacent to the park have a cooling demand which is 13.9% lower than that of the same buildings further away from the park (Toparlar et al., 2018).

Another study estimated that the removal of waste heat from cooling systems would result in a 1∘C drop in temperature and 6% decrease in cooling energy consumption for the region of Ootemachi, Tokyo (Kikegawa et al., 2003). More recent work in San Francisco used simulation coupled with high-resolution climate sensors to estimate the impact of microclimate on building energy consumption (Hong et al., 2021). They find that the failure to consider localized weather may result in up to 100% differences in annual heating, 65% difference in annual cooling, and up to 30% difference in peak cooling electricity demand. Microclimate may play a significant role in driving energy consumption, although it has yet to reach mainstream adoption in energy models. To achieve a higher-quality mesh of climate variables throughout the city, the researchers used interpolation between the sensors as a proxy for missing spatial data. Data collection with sensors is common for researchers within the domain and is often a limiting factor in building a more powerful analysis. However, given the difficulty of deploying and maintaining a fleet of sensors, researchers often opt to study a single microclimate effect for each analysis (Srebric et al., 2015; Li et al., 2019; Yang et al., 2020; Yang et al., 2021).

The general lack of adequate, high-resolution data was demonstrated in recent air-quality research, highlighting that existing data sources may rapidly fluctuate and are potentially susceptible to localized sources of error (Apte et al., 2017). Given that microclimate features have similar spatio-temporal dynamics, it is likely that urban microclimate readings suffer from similar issues.

The capacity to study the influence of climate on urban spaces has historically been limited to the resolution of TMY data. While the past 50 years have seen a continued march of general climate models towards a higher spatial resolution (Sellers, 1969; Uppala et al., 2005; Hersbach et al., 2020), the diversity of interaction effects within urban spaces mitigates the capacity to generalize the results of one study to a new region. Preliminary research has attempted to curate a holistic model of urban climate (Salvati et al., 2020, and Ma & Cheng, 2016), which has grown in parallel with computational fluid dynamics (CFD) models of urban regions (Toparlar et al., 2017). While CFD models may provide valuable insights into a city's heat and wind distributions, they are challenging to build and validate due to the lack of ground truth data.

This work provides a potential solution for the deep integration between sources of high-resolution microclimate data and building energy modeling. Specifically, we propose integrating high-frequency satellite data with climate reanalysis for our underlying dataset. By using these rapidly accessible data sources to curate an energy analysis, we aim to demonstrate the impact of microclimate on building energy usage in New York City.

3. Data

For context, a borough map of New York City has been provided in Fig. 1 for your reference, as various regions will be discussed throughout the analysis. Our data is split into three primary categories. The first is energy data, the second is building features, and the final category is environmental data captured from satellites or reanalysis models.

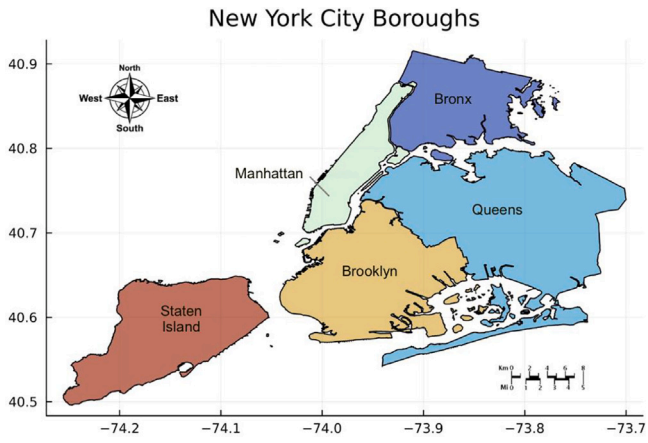

Fig. 1. Overview of Boroughs in New York City.

The environmental data collected from climate reanalysis and remote sensing was captured using Google Earth Engine. The coordination of various spatial resolutions of data is handled internally by Google Earth Engine, which provides a mechanic of upsampling lower-resolution data to accommodate requests for higher-resolution geometries. The process of collecting environmental parameters from Google Earth Engine is as follows. First, the convex hull of the building footprint is computed and used in place of the actual building footprint. The substitution of simple geometry is due to the complexity of actual building footprints, which both inflate file sizes and dramatically increase the computational complexity of subsequent analysis. Next, a buffer radius of 100 meters is applied to the building's convex hull, which will serve as the region of capture for all subsequent data. The 100-meter buffer was selected based on visual inspection of the satellite data, which may be collected adjacent to the structure. This radius was designed to capture all potential features adjacent to the building that might have regular interactions with the structure. As per Google's documentation on image reduction (Gorelick et al., 2017), the incoming data source will attempt to partition into the specified mesh resolution, which will then be averaged over the area of the buffered region. A flowchart of the process for curating and cleaning building footprints, energy consumption, and microclimate features can be found in Fig. 2.

A note on the final quality of the aggregate data set: integrating each remote sensing instrument into the inference model brings new insights into the urban microclimate, improving the quality of the model which can be constructed. However, the requirement of multiple data streams will introduce new opportunities to embed missing data. In aggregate, 3.9% of the original data is removed from the analysis due to missing data from remote sources.

3.1. Building data

New York City was chosen as the test bed for this analysis due to the abundance of publicly reported data. One of the primary sources which serve as a foundation of this analysis is the monthly building energy consumption. Privately owned buildings over 25,000ft^2 (2322m^2) and city-owned buildings over 10,000ft^2 (929m^2) are required by Local Law 84 to report their monthly electricity and gas consumption. Each building can be described with a unique ID in the form of a borough, block, and lot (BBL) number and a building identification code (BIN). This analysis links each additional source of building-specific data through the building's BBL and BIN.

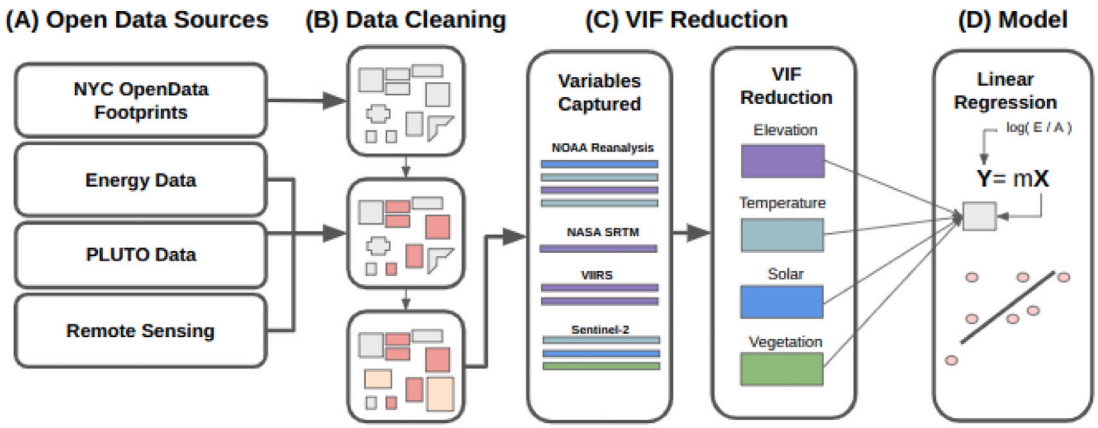

Fig. 2. Data Pipeline: (A) The collection of openly accessible data from New York City building-specific features, energy data, and remote sensing. (B) The cleaning of this open data for further processing based on availability, followed by (C) identifying highly correlated terms and reducing features based on the variance inflation factor. Finally, (D) shows the regression of the reduced terms against the logarithmic energy per unit area.

New York City also provides two other critical sources of information about the built environment. The first is building footprints. Building footprints for the 1.08 million buildings in New York City are captured and provided by New York City's OpenData platform. The capture mechanism for these footprints is a mixture of photogrammetry and manual orthophotography, with details of edge case scenarios listed in the public disclosure of data quality provided alongside the data set.

The Primary Land Use Tax Lot Output (PLUTO) is the final New York data set used to collect building features. PLUTO contains over 70 features particular to the structure, including zoning information, zip code, building use type, and wall/floor dimensions. Of note, our study intentionally leaves out several terms from the PLUTO dataset which are specific to New York City. The exclusion of the terms is done to encourage the generalization of our research into new urban regions of similar climates. The terms selected for the study from PLUTO are the tax value assessments (assesstot), useable building area, the year of construction (yearbuilt), and metadata used to link PLUTO data to existing data sources. The tax value metrics are captured through the Assessed Total Value and the Assessed Land Value. The usable building floor area is captured through the Building Area metric, which is procured through the collection of sources like the Property Tax System (PTS), the Computer Assisted Mass Appraisal (CAMA) or inferred by using the number of stories and building footprint (PLUTO Data Dictionary May 2022 (22v1), 2022).

At the time of writing in 2022, the monthly energy consumption of 9732 buildings was reported throughout the years 2018, 2019, and 2020. Assuming these buildings report energy data for each month, we should have 350,352 electricity data points and 350,352 gas data points. After filtering for duplicated IDs and pairing the energy consumption data of the 9732 buildings with those of the building footprints, the total number of unique buildings falls to 9250, with the number of valid monthly energy consumption points falling to 273,600.

3.2. Mesoscale reanalysis data

Climate reanalysis, often known through the phrase "maps without gaps", blends historical climate observations with that of modern climate models (Uppala et al., 2005). In doing so, reanalysis attempts to provide a complete picture of the earth's climate history for all locations on the planet at hourly intervals. The grid of data points provided by reanalysis is three-dimensional, spanning the planet's surface and extending into space. As cities tend to spread more horizontally than vertically, climate models' most pertinent measurements should be those close to the surface (Kondo et al., 2005).

Without much consideration to high altitude weather patterns, mesoscale models often only need to consider climate activity within the first 120 meters from the ground (Shi et al., 2016). In contrast, the mesh resolution adjacent to the surface becomes a more significant feature of interest with modern urban canopy models able to achieve a planar resolution of at least 1 kilometer (Ooka, 2007). Attempts have been made to transition single-layer urban canopy models to multi-layer urban canopy models, but they were shown to provide no appreciable benefit for accurately measuring urban climate features (Kusaka et al., 2001). As an illustration of scale, New York City's 99th percentile of building heights is only 26 m.

As such, we put more weight into selecting a climate reanalysis with high mesh resolution at the planet's surface. This work selects NOAA's National Weather Service RTMA for the collection of reanalysis data (Caldwell, 0000). The NOAA RTMA offers historical meteorological data with a resolution of 2.5 kilometers per cell over the contiguous United States since 2011 at an hourly interval. NOAA RTMA provides the highest spatial resolution among climate reanalysis models, with comprehensive coverage of our data set.

In considering alternatives, ERA5 climate reanalysis was considered for its global spatial coverage, which may extend this research to more regions (Hersbach et al., 2020). ERA5 was additionally found to have a broad temporal reach. Reporting data as early as 1979, utilization of the ERA5 data set in energy analysis enables almost all internationally reported building energy data to be included as part of a global data set. However, ERA5's surface level resolution is roughly 30 km at the equator, making it less ideal for interpreting microclimate causes of energy consumption. As ERA5 is often used for climate modeling research, one of its primary benefits is the high-resolution perpendicular to the planet's surface. The lowest altitude cell is registered at 1hPa of pressure, roughly 80 meters above sea level. While this is not necessarily a limiting factor in its implementation, the increased surface resolution of the NOAA reanalysis made it the more comprehensive and appropriate option for the analysis.

3.2.1. Remote sensing

Three remote sensing products were used as part of the data pipeline for this project: Sentinel-2 Level-1C, VIIRS, and NASA's SRTM. These data sets were selected due to their prominence in scientific discourse, open access, and high-resolution data.

The Sentinel satellite provides imagery at a spatial resolution of up to 10 meters per pixel, capturing wavelengths between 0.44 and 2.2 μm with a revisit interval of 5 days. The raw data captured by the Sentinel satellite is processed through a series of quality control systems internal to the European Space Agency (ESA) before releasing the data to the public. The ESA released the data used in this analysis as a Level-1C product, which provides top-of-atmosphere reflectances in cartographic geometry. Before building any analysis, cloud masking was applied to the sentinel data using the provided bitwise mask band, QA60.

Table 1
Final variables description.

	Variable	Description	Unit	Resolution	Data source	VIF
1	avg_rad	Nighttime Light Radiance	nW sr^{-1} cm^{-2}	500 m	JPSS VIIRS/DNB	1.45
2	B1	Aerosols (443 nm)	W sr^{-1} m^{-2}	60 m	Sentinel-2 Level-1C	2.61
3	B11	Shortwave Infrared (1612 nm)	W sr^{-1} m^{-2}	20 m	Sentinel-2 Level-1C	4.00
4	NDVI	Vegetation Index	–	10 m	Sentinel-2 Level-1C	2.09
5	WIND	Average Wind Speed	m s^{-1}	2.5 km	NOAA RTMA	1.60
6	TCDC	Total Cloud Cover	%	2.5 km	NOAA RTMA	2.81
7	ACPC01	Total Precipitation	kg m^{-2}	2.5 km	NOAA RTMA	1.60
8	hdd	Heating Degree Days	–	2.5 km	NOAA RTMA	3.37
9	cdd	Cooling Degree Days	–	2.5 km	NOAA RTMA	3.23
10	elevation	Elevation	m	30 m	NASA SRTM	1.16
11	assesstot	Total Property Value	$	–	PLUTO	1.15
12	yearbuilt	Year of Construction	–	–	PLUTO	1.03

Because of its high spatial resolution, the Sentinel-2 Level-1C product provides hyper-local, regional features like vegetation and impervious surfaces, which may significantly influence the structure's energy consumption (Robineau et al., 2022). The vegetation was measured using the Normalized Difference Vegetation Index (NDVI), a common method of locating vegetation growth based on the reflection of low frequency infrared and absorption of red light during photosynthesis as seen in Bhandari et al. (2012).

Urban nightlights were the next source of exploration in remote sensing data as they are shown to provide insights into the spatial connectivity of urban spaces (Small et al., 2013). Urban night lights additionally provide hints into economic activity (Dasgupta, 2022; Määttä et al., 2021). We hypothesize that the high saturation of night lighting may indicate higher rates of commercial activity, which may serve as a valuable feature for prediction. The Earth Observation Group provided the night lights data at the Payne Institute for Public Policy, Colorado School of Mines (Elvidge et al., 2017). They capture radiance using the VIIRS instrument on the Joint Polar Satellite System (JPSS) between the wavelengths of 412 nm and 12 µm.

Finally, ground elevation data was folded into the analysis through the incorporation of NASA's Shuttle Radar Topography Mission (SRMT) in 2000. The static map created by this endeavor provides a 30-meter resolution map of global elevation.

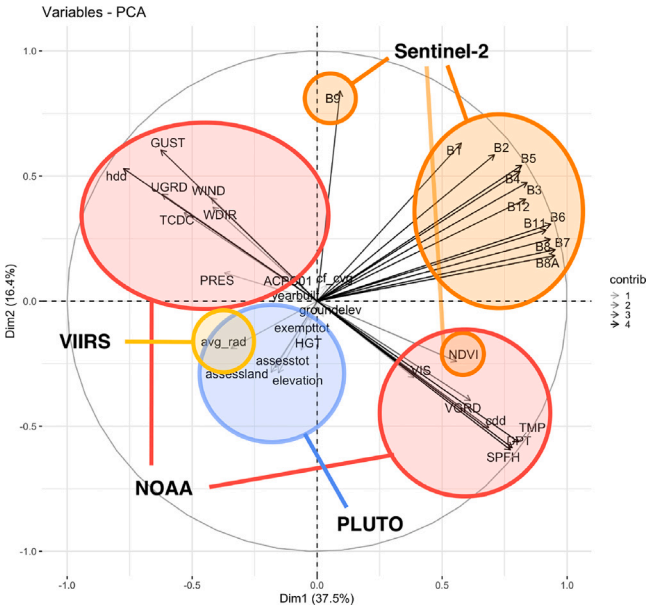

Fig. 3. Visualized Correlation of Variables.

4. Methods

Having conducted a general survey of potential climate models and satellite data, we proceed with analysis by conducting linear regression against the final variables found in Table 1.

4.1. Endogenous terms

The endogenous terms (Y) are defined by the equation $Y_i = ln(E_i/A_i)$, with $i \in B$ and B defined by the set of all buildings in the data set. Two variables are selected as the endogenous terms (E): the monthly electricity consumption and the monthly gas consumption, both in units of MWh. These terms are divided by the useable floor area (A) of the buildings (B), provided by the PLUTO dataset as sq.ft and translated to square meters. A logarithm is then applied to this area normalized energy consumption, reducing the influence of right skew and permitting analysis of each variable's effect size per unit increase in the term.

4.2. Exogenous terms

Of particular interest for this inference-based study is the issue of multicollinearity, a thorough analysis of which is captured in the study by Kim (2019), which may cause the coefficients of our regression parameters to fluctuate if the exogenous terms are too highly correlated wildly. The terms are evaluated by computing their variance inflation factors (VIF), which indicate the degree of multicollinearity by quantifying the capacity of the exogenous terms to act as predictors for another term. The plot in Fig. 3 is a correlation matrix of the exogenous variables, which are sequentially removed until all VIF terms are below 5.

A single feature from these highly related groups serves as a representative for each correlated group. For example, bands B1–B12, except B11, were represented by the single band of B1. The specific humidity and dewpoint temperature are captured using temperature measurements from the NOAA reanalysis and represented using heating degree days and cooling degree days, commonly used terms in the building energy domain (D'Amico et al., 2019). Gust, directional wind speeds, pressure, and visibility were all highly correlated with wind speed and thus removed. Finally, land value was removed for its high correlation with total property value and redundant elevation terms.

The final variables selected, the data set of origin, and its computed VIF value can be found in Table 1.

4.3. Regression construction

For each regression constructed, two variations are proposed. The first analysis is created by regressing endogenous energy consumption against mean-centered exogenous terms. By choosing not to normalize unit variance prior to regression, we can estimate the percentage change in energy per unit area given a unit increase in the exogenous terms. The second regression is instead constructed by mean centering and normalizing the exogenous terms, with the intention of interpreting the magnitude of significance for the term, given that some features might have more significant variance throughout the city.

Table 2
Linear model — mean centered.

	Dependent variable:	
	log(Electric/Area)	log(Gas/Area)
avg_rad	0.002*** (0.00004)	−0.003*** (0.0001)
B1	0.060 (0.114)	3.007*** (0.276)
B11	−2.001*** (0.061)	3.971*** (0.148)
NDVI	−0.754*** (0.029)	0.007 (0.070)
WIND	−0.041*** (0.003)	0.096*** (0.008)
TCDC	−0.002*** (0.0003)	0.006*** (0.001)
ACPC01	0.783*** (0.028)	−0.133* (0.069)
hdd	−0.015*** (0.0004)	0.097*** (0.001)
cdd	0.047*** (0.001)	−0.060*** (0.003)
elevation	−0.003*** (0.0001)	−0.0001 (0.0003)
assesstot	0.000*** (0.000)	−0.000*** (0.000)
yearbuilt	0.001*** (0.00003)	0.0001 (0.0001)
Constant	−8.651*** (0.002)	−8.769*** (0.004)
Observations	220,820	217,883
R^2	0.116	0.171
Adjusted R^2	0.116	0.170
Residual Std. Error	0.748 (df = 220807)	1.794 (df = 217870)
F Statistic	2,408.626*** (df = 12; 220807)	3,731.927*** (df = 12; 217870)

Note: *p<0.1; **p<0.05; ***p<0.01.

4.4. Regression and microclimate identification

After removing all zero terms from the endogenous variable and corresponding exogenous features, relationships are discovered through linear regression. The results of the mean-centered regression can be found in Table 2, while the results of the normalized regression can be found in Table 3. The regression is utilized as a point of information for subsequent clustering to help identify microclimate regions within the city.

Standard linear regression looks to minimize the sum of mean squared loss $L(\beta) = \|\mathbf{X}\beta - \mathbf{Y}\|^2$. $\mathbf{X}\beta$ thus becomes a matrix–vector product for an array of samples. The dimensionality of $\mathbf{X}$ is NxM, where N is the number of samples while M is the number of features used to describe each building. As β represents the coefficients from the regression, its dimensionality is $Mx1$, and thus the matrix–vector product results in the prediction $p_i = \sum_{j=1}^{M} \mathbf{X}_{ij}\beta_j$. If instead the summation is removed and an augmented matrix $\mathbf{A}$ of dimensionality NxM is created such that $\mathbf{A}_{ij} = \mathbf{X}_{ij}\beta_j$, then clustering against $\mathbf{A}$ will uncover microclimate communities which are particularly significant for energy consumption. Two $\mathbf{A}$ matrices will be constructed, one for gas - $\mathbf{A}_g$ and one for electricity $\mathbf{A}_e$.

As underlying land use changes often link environmental parameters, it may be unlikely that a single environmental variable will ever be shifted without modification to others. Thus for a more accurate simulation of potential modifications to the urban microclimate, we propose that shifts between clusters instead be simulated. For this analysis, gaussian mixture models were explored as potential clustering mechanics for $\mathbf{A}$.

5. Results

The results of the regression may be found in Tables 2 and 3 below. Our results indicate that localized microclimate indeed seems to play a significant role in urban energy consumption, and many of the pertinent microclimate features may be captured using the resolution of data provided by modern climate models and satellite imagery.

5.1. Electricity

The microclimate causes of electricity consumption in New York seem relatively heterogeneous, as seen in the regression results from Table 3. The most significant microclimate features to drive consumption habits are cooling-degree days, heating-degree days, and B11 readings. B11 captures light at a wavelength of 1610 nanometers, classified as short-wave infrared radiation. An increase in measured values of short-wave radiation might indicate that the material can dissipate thermal energy, whereas lower readings might indicate a level of thermal trapping that the urban canopy may cause.

Night light emissions (avg_rad) are also significant in the regression model. While night light emissions may be a direct indicator of electricity consumption due to indoor lighting, it is perhaps more likely that buildings with higher levels of night-time emissions correlate to commercial regions in the city.

We may also use these results to explore localized microclimate effects and estimate the influence of various urban features. For example, with a coefficient of −0.754, the mean-centered regression indicates that an increase of one unit in NDVI would correspond to a decrease of 75% in electricity consumption. In practice, the typical NDVI in New York City might swing between −0.05 and 0.2 depending on the region's season and density of vegetation. For example, the average NDVI values in Midtown (middle of Manhattan) are −0.04. With a regression coefficient of −0.754, NDVI is expected to modify the electricity consumption in Midtown by $100 \cdot (exp(-0.754 \cdot -0.04) - 1) = 3.06\%$.

Buildings adjacent to Central Park on the Upper West Side (northern Manhattan) have average readings closer to 0.11. The estimated impact from vegetation on electricity is thus $100 \cdot (exp(-0.754 \cdot 0.11) - 1) = -7.96\%$. This example result indicates that an identical building next to Central Park on the Upper West Side is likely to have 11% lower electricity consumption than in Midtown due to the difference in vegetation, which is similar to the results found by Toparlar et al. in their analysis of cooling demand in Antwerp, Belgium (Toparlar et al., 2018). The same method of computation may be applied to any of the regression coefficients in Table 2 to estimate the feature's impact on the endogenous term.

5.2. Spatial analysis — Electricity

In general, the more densely packed urban areas in New York suffer from the adverse effects of urban microclimate concerning electricity consumption. The lower B11 readings in Manhattan indicate that trapping of low-frequency light might disproportionately impact the structures' efficiency. Fig. 4 shows both the overall estimated impact of environmental features on electricity consumption and the specific regional influence of vegetation on electricity consumption.

The same maps can be used to explore the regional effects of vegetation on electricity consumption. For example, a regional effect due to vegetation can be found on the periphery of Central Park, as seen

Table 3
Linear model — mean centered & normalized.

	Dependent variable:	
	log(Electric/Area)	log(Gas/Area)
avg_rad	0.115*** (0.111, 0.119)	−0.158*** (−0.167, −0.149)
B1	0.001 (−0.004, 0.006)	0.068*** (0.055, 0.080)
B11	−0.104*** (−0.110, −0.098)	0.207*** (0.192, 0.222)
NDVI	−0.060*** (−0.064, −0.055)	0.001 (−0.010, 0.011)
WIND	−0.026*** (−0.030, −0.022)	0.062*** (0.052, 0.071)
TCDC	−0.016*** (−0.021, −0.010)	0.053*** (0.040, 0.065)
ACPC01	0.055*** (0.051, 0.059)	−0.009* (−0.019, 0.0002)
hdd	−0.097*** (−0.102, −0.091)	0.638*** (0.624, 0.651)
cdd	0.128*** (0.122, 0.134)	−0.166*** (−0.180, −0.152)
elevation	−0.049*** (−0.053, −0.046)	−0.002 (−0.010, 0.006)
assesstot	0.073*** (0.070, 0.077)	−0.256*** (−0.264, −0.248)
yearbuilt	0.082*** (0.079, 0.086)	0.003 (−0.005, 0.012)
Constant	−8.651*** (−8.654, −8.648)	−8.769*** (−8.776, −8.761)
Observations	220,820	217,883
R^2	0.116	0.171
Adjusted R^2	0.116	0.170
Residual Std. Error	0.748 (df = 220807)	1.794 (df = 217870)
F Statistic	2,408.626*** (df = 12; 220807)	3,731.927*** (df = 12; 217870)

Note: *p<0.1; **p<0.05; ***p<0.01.

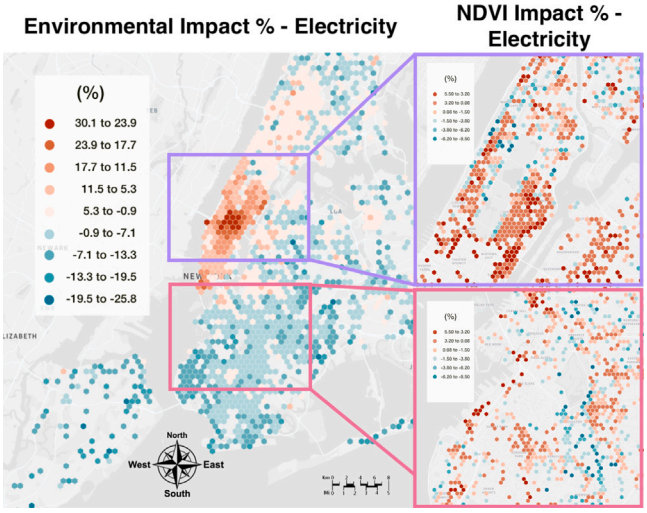

Fig. 4. Environmental Effects — Electricity.

in the upper right quadrant of Fig. 4. The higher levels of vegetation from the park are estimated to reduce the electricity consumption of adjacent buildings by 5%–7% per unit area compared to their peers in the same neighborhood.

5.3. Gas

The gas data is observed to have much more homogeneous sources of deviation. This source diversity is evidenced in the regression results found in Table 2. The regression is dominated by the heating degree days, with a coefficient of 0.638. B11 readings from Sentinel-2, capturing short-wave radiation, are again identified as a significant indicator. For the gas regression, we find that higher B11 readings seem to indicate higher gas consumption. As higher B11 readings would indicate that the region is less effective at trapping thermal energy, this matches our intuition that regions with a greater capacity to trap heat may need less heating energy in the form of gas. The summation of the absolute value coefficients for B1, B11, NDVI, WIND, TCDC (cloud coverage), and ACPC01 (precipitation) is 0.49. While heating degree days provide significant insights into the behavioral mechanics of gas consumption, nearly half of the potential sources for deviation are missed when alternative environment variables are not included.

Regarding social indicators, the night light radiance may provide an essential clue for the discrimination of building class. The average radiance coefficient of −0.158 indicates that higher night-time lighting may indicate lower gas consumption. Regions with a higher intensity of night lighting may correspond to commercial zones, which may not require heat throughout the night. Additionally, the total property value of the building serves as a more significant indicator for gas consumption, likely as more expensive buildings may have transitioned away from gas.

5.4. Spatial analysis — Gas

Regions in the south of New York City that interface the ocean seemed to have the most significant likely increase in gas consumption, as seen in Fig. 5. This difference is unlikely to be due to discrepancies with vegetation, as the vegetation of the city seems to have almost no appreciable effect on gas consumption. The greenest regions of New York are likely to see a marginal increase of 0.11% in gas consumption due to vegetation. While prior work has shown trees typically provide a cooling effect in the summer months from evapotranspiration, they are dormant in the winter when heating may be necessary (Kleerekoper et al., 2012).

The distribution of heating degree days in New York is reasonably flat, with an average monthly spread between 6 and 8, as seen on the leftmost image in Fig. 5. This distribution map indicates that given temperature alone, we might expect the typical building in the Bronx to have a 10% greater gas consumption than those of Brooklyn. However, we do not see this expressed in the regression for the complete prediction of environmental effects, as evidenced by Fig. 5.

Prior work has demonstrated that the effect of infiltration is more pronounced in older buildings (Antretter et al., 2007). Since Brooklyn has a high density of older buildings, we might expect the wind to impact their overall gas consumption significantly. Indeed, the wind significantly influences gas consumption in Brooklyn, with the coastal wind playing a dominant role among the potential causes for increased gas consumption. In the coastal regions of southern Brooklyn and Staten Island, the impact of wind on gas consumption rivals that of heating degree days. We find that in the southern part of New York City, the wind is regularly estimated to increase gas consumption by 5%–15% per unit area, as seen in Fig. 5.

Wind speeds drop off when they interface with larger bodies of vegetation, which likely increases the surface roughness of the urban texture. The reduced wind speeds in Midwood, Flatbush, and Staten Island leads the predictions for environmental effects to hover between a 0 and 15% increase in gas consumption. The results for the more interior parts of Brooklyn live in stark contrast to the nearby region

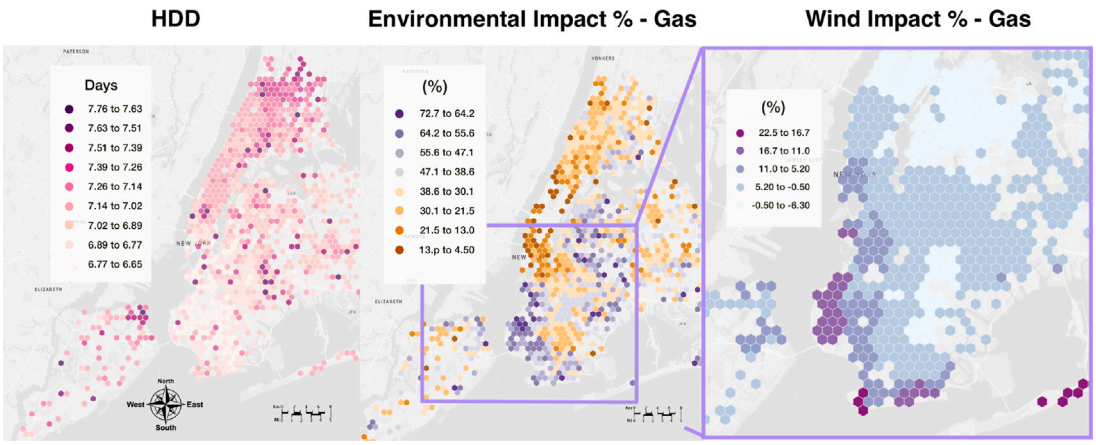

Fig. 5. Environmental Effects — Gas.

of Sunset Park, in which buildings regularly may expect between a 20%–30% increase in gas consumption primarily due to increased wind speeds.

Another interesting topic is the role of B11 measurements from Sentinel-2. With a positive coefficient, higher readings of B11 indicate the likelihood of a greater quantity of gas consumption per unit area. Higher B11 readings indicate that the region is less effective at naturally trapping thermal energy, which may be detrimental to the building's energy efficiency throughout the cold winter months.

5.5. Microclimate transitions

A microclimate is commonly framed as inducing a consistent environmental impact on a small city region. Urban Heat Island, one of the commonly cited forms of urban microclimate, is often portrayed as the consistent addition of heat to a region due to modified heat capacitance, changes in land use, or urban canyon effects (Yin et al., 2018). As microclimate analysis is typically conducted on the scale of a year, less granular temporal readings for land surface temperature are nonetheless likely to provide valuable information when developing policy for urban decarbonization (Reinhart & Cerezo Davila, 2016).

The utilization of monthly energy consumption data also permits the exploration of how a building might transition between environmental microclimates throughout the year. To explore the dynamics of microclimate transitions, the regressed microclimate significance matrix **A** is clustered into ten unique groups for illustration. To maintain consistent terminology, we will use the term "energy microclimate" (EMC) to describe the results of the clustering against **A**. EMCs are perhaps most similar to Homogeneous Urban Zones (HUZ), which are introduced in the urban land use classification (López-Moreno et al., 2022). A notable difference between HUZs and EMCs is that EMCs are curated against statistically significant microclimate features conditioned using energy consumption. In practice, EMC groups represent the distinct environmental conditions that significantly influence the energy consumption of buildings.

Notably, almost every building in New York City will experience transitions through at least eight of the ten potential energy microclimates. The high transition rate speaks to the dominance of seasonal weather patterns, as the partitioning of microclimate features overlaps with seasonal weather trends. The plot of unique microclimate counts in Fig. 6 shows the presence of two potentially significant factors to drive the behavior of New York City's urban climate. The first is the potential presence of an *urban core*, which exists in southern Manhattan and dips into the western part of Queens. The more climatically stable urban core is likely related to a lack of vegetation, which in other parts of the city has a consistent annual growth pattern. This theory is validated by the high number of transient regions stretching into Long Island, which is often much greener than the rest of the city.

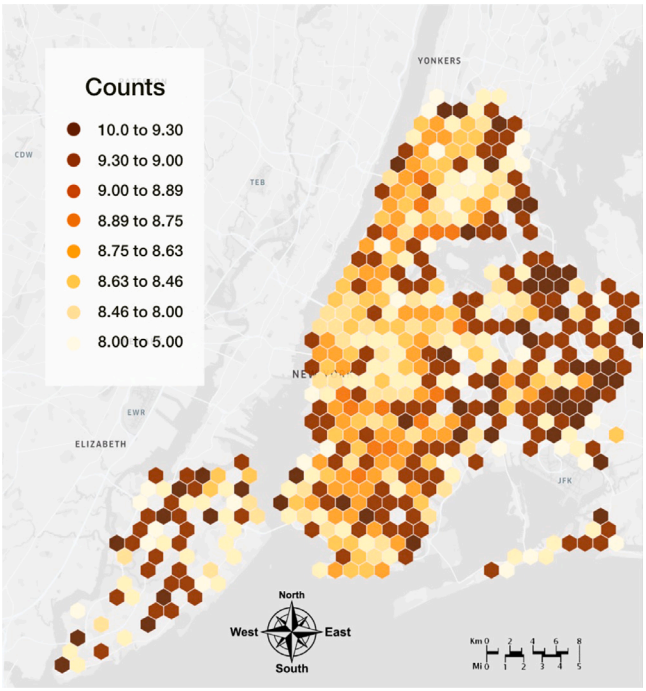

Fig. 6. Count of Unique Microclimates.

The second interesting point of note is coastal climates, which also have lower transition rates between urban microclimates. These regions may be seen on the eastern coast of Staten Island and the southern coast of Brooklyn in Fig. 6. While our intuition might be that high rates of microclimate transition would lead to higher energy consumption, these southern regions cast doubt on this theory. With more predictable, high-speed coastal winds, buildings on this southern interface regularly consume more gas.

6. Discussion

The process laid out in this research demonstrates that the energy consumption of buildings may be coupled with readily available environmental data to interpret the effect sizes of various microclimate events on the urban scale. In recognizing the influence of microclimate on building energy consumption, we propose that the city may be spatially split into "zones", which serve as a more compact way to reference significant microclimate archetypes. Approaches to cluster urban spaces into representative regions have been conducted before, with particular emphasis on urban morphology (Joshi et al., 2022), building

archetypes (Dogan & Reinhart, 2017), or urban heat island (Shen et al., 2021). The clustering approach in this analysis instead slots into the gap of research in the building energy domain, providing both a method of automatically collecting data and clustering buildings based on critical microclimate variables to explain their energy performance.

In providing this pipeline for cities with access to building energy data, this research provides a mechanism for utilizing building energy data to monitor the interaction effects of bespoke urban microclimates on building energy efficiency.

A city with disclosed building energy consumption data may now apply our methodology to identify the primary environmental features driving the inefficiency of its structures. Additionally, this research provides a significant step forward for cities with sparse resources to monitor their structures' interactions with urban microclimates. Without building energy data, the city may utilize the coefficients discovered in our study, which may provide insights for other cities around the latitude of 40.7N within the United States and Canada.

Some of the unique cultural features of New York City may also introduce challenges to the generalization of our research to new environments. For example, the distribution and intensity of urban night lights captured by VIIRS may be a unique signature for New York City's vibrant commercial activity. New York City additionally provides a uniquely high-quality dataset describing the geometries of its buildings. Cities without a comprehensive database of building footprints may have limited capacity to generate a comprehensive building energy model of their city. While there is growing potential to automate the task of extracting building footprints (Microsoft, 2018), the current quality of building footprints in dense urban areas is not suitable for analysis.

To our knowledge, energy benchmarking for buildings, a branch of energy-related research which seeks to rank buildings against their peers, does not consider the implications of urban microclimate in establishing relational metrics to score buildings. Given that prior work has demonstrated other microclimate phenomena like surface urban heat island (SUHI) may not have an equitable impact on all members of society (Hsu et al., 2021), we believe that the automatic curation of environmental parameters using our system will play a valuable role in promoting more comprehensive metrics for benchmarking research.

7. Limitations and future work

While this analysis demonstrates the potential value of using remote sensing systems combined with climate reanalysis, there are some notable limitations regarding the scope of potential applications. As our system was not curated with the primary purpose of predicting building energy consumption, it likely misses meaningful nonlinear relationships between variables. For example, building age and urban night lighting may host a wealth of information about the construction quality of buildings and their occupancy patterns. However, a city region with one-fifth of the urban night lighting compared to the archetypical commercial region does not necessarily have one-fifth of the commercial activity. Nightlights are thus likely best expressed through nonlinear relationships with energy consumption.

The 100-meter buffer around buildings as a region of data collection may be a point of further investigation, which may be approached based on the results of techniques like saliency mapping (Dougherty et al., 2021). Additionally, the relationship between regions in the city needs to be explicitly incorporated as a mechanism within this model. Our results from the spatial analysis of gas consumption found that high wind speeds were predicted to increase gas consumption on the city's southern coast. However, as our system cannot predict the change of environmental variables associated with modifications to the region's microclimate, it cannot simulate the impact of something like efforts to reduce the wind speed in southern Brooklyn.

8. Conclusion and implications

This study is one of the first examples of the direct impact of urban microclimate on building energy consumption based on real-world, historical monthly data over three years. In this study, we demonstrated a method of rapidly extracting hyper-localized environmental features around buildings using a variety of satellite imagery and high-resolution reanalysis data. The tools used to capture this data will be made available for all researchers as an open-source tool, enabling the rapid collection of historical microclimate data for all research related to the built environment.

This study demonstrates the potential value of high-resolution, historical microclimate data with a small case study in New York City. Through this case study, we show the significant effects of wind on gas consumption in Brooklyn. We estimate that wind in Brooklyn may increase the gas consumption of buildings in coastal regions by 10%. We also note the effect of parks on electricity consumption, as buildings directly adjacent to Central Park are predicted to consume 5%–7% less electricity.

Upon aggregating the effects of various microclimate features, we show that localized microclimate effects on a building may decrease gas consumption by as much as 71% or increase it by as much as 221%. We additionally recognize the potential of microclimate effects to drop the electricity consumption of a typical building in New York City by 24.4% or increase the electricity consumption by 55.2%.

In summary, modern research in urban energy consumption has shown urban microclimate to be a significant hurdle for improved accuracy of urban energy modeling. This study demonstrates a mechanism of rapidly collecting high-resolution urban microclimate data, utilizing it in a brief case study to explore microclimate effects on energy consumption in New York City. We also explore correlations between environmental microclimate features, using a cleaned version of the data to compose a definition of Energy Microclimates (EMCs) which enables microclimate archetypes to be defined as it pertains to building energy consumption. Finally, we show that the urban microclimate in New York is expressed through dynamic, competing forces which may have significant localized effects both in time of year and spatial arrangement. The process in this research may be used to rapidly collect microclimate data for accelerated research into urban microclimate design strategies.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data used is publicly available. The code to collect the data and process it will be made openly available. Additionally, the final data used in the analysis may be made available upon request.

Acknowledgments

The research presented in this paper was supported in part by a Terman Faculty Fellowship and the U.S. National Science Foundation (NSF) under Grant No. 1941695. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of U.S. NSF.

References

Ali, U., Shamsi, M. H., Hoare, C., Mangina, E., & O'Donnell, J. (2021). Review of urban building energy modeling (UBEM) approaches, methods and tools using qualitative and quantitative analysis. *Energy and Buildings*, *246*, Article 111073. http://dx.doi.org/10.1016/j.enbuild.2021.111073.

Antretter, F., Karagiozis, A., TenWolde, A., & Holm, A. (2007). Effects of air leakage of residential buildings in mixed and cold climates. Retrieved July 20, 2022 from URL http://www.fs.usda.gov/treesearch/pubs/29794.

Apte, J. S., Messier, K. P., Gani, S., Brauer, M., Kirchstetter, T. W., Lunden, M. M., Marshall, J. D., Portier, C. J., Vermeulen, R. C., & Hamburg, S. P. (2017). High-resolution air pollution mapping with Google street view cars: Exploiting big data. *Environmental Science and Technology*, *51*(12), 6999–7008. http://dx.doi.org/10.1021/acs.est.7b00891, Publisher: American Chemical Society.

Bhandari, A. K., Kumar, A., & Singh, G. K. (2012). Feature extraction using normalized difference vegetation index (NDVI): A case study of Jabalpur City. *Procedia Technology*, *6*, 612–621. http://dx.doi.org/10.1016/j.protcy.2012.10.074.

Caldwell, D. (0000). Real-time mesoscale analysis. p. 5.

Cerezo, C., Sokol, J., AlKhaled, S., Reinhart, C., Al-Mumin, A., & Hajiah, A. (2017). Comparison of four building archetype characterization methods in urban building energy modeling (UBEM): A residential case study in Kuwait City. *Energy and Buildings*, *154*, 321–334. http://dx.doi.org/10.1016/j.enbuild.2017.08.029.

Cetin, M. (2013). Landscape engineering, protecting soil, and runoff storm water. In *Advances in landscape architecture*. IntechOpen.

Cetin, M. (2015). Consideration of permeable pavement in landscape architecture. *Journal of Environmental Protection and Ecology*, *16*(1), 385–392.

Craig, M. T., Wohland, J., Stoop, L. P., Kies, A., Pickering, B., Bloomfield, H. C., Browell, J., De Felice, M., Dent, C. J., Deroubaix, A., Frischmuth, F., Gonzalez, P. L. M., Grochowicz, A., Gruber, K., Härtel, P., Kittel, M., Kotzur, L., Labuhn, I., Lundquist, J. K., ... Brayshaw, D. J. (2022). Overcoming the disconnect between energy system and climate modeling. *Joule*, *6*(7), 1405–1417. http://dx.doi.org/10.1016/j.joule.2022.05.010.

D'Amico, A., Ciulla, G., Panno, D., & Ferrari, S. (2019). Building energy demand assessment through heating degree days: The importance of a climatic dataset. *Applied Energy*, *242*, 1285–1306. http://dx.doi.org/10.1016/j.apenergy.2019.03.167.

Dasgupta, N. (2022). Using satellite images of nighttime lights to predict the economic impact of COVID-19 in India. *Advances in Space Research*, *70*(4), 863–879. http://dx.doi.org/10.1016/j.asr.2022.05.039.

Dogan, T., & Reinhart, C. (2017). Shoeboxer: An algorithm for abstracted rapid multi-zone urban building energy model generation and simulation. *Energy and Buildings*, *140*, 140–153. http://dx.doi.org/10.1016/j.enbuild.2017.01.030.

Dougherty, T. R., Huang, T., Chen, Y., Jain, R. K., & Rajagopal, R. (2021). SCHMEAR: Scalable construction of holistic models for energy analysis from rooftops. In *proceedings of the 8th ACM international conference on systems for energy-efficient buildings, cities, and transportation* (pp. 111–120). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3486611.3486666.

Elvidge, C. D., Baugh, K., Zhizhin, M., Hsu, F. C., & Ghosh, T. (2017). VIIRS night-time lights. *International Journal of Remote Sensing*, *38*(21), 5860–5879. http://dx.doi.org/10.1080/01431161.2017.1342050, Publisher: Taylor & Francis _eprint: DOI: 10.1080/01431161.2017.1342050.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. http://dx.doi.org/10.1016/j.rse.2017.06.031.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., ... Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. http://dx.doi.org/10.1002/qj.3803, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3803.

Hong, T., Xu, Y., Sun, K., Zhang, W., Luo, X., & Hooper, B. (2021). Urban microclimate and its impact on building performance: A case study of San Francisco. *Urban Climate*, *38*, Article 100871. http://dx.doi.org/10.1016/j.uclim.2021.100871.

Hsu, A., Sheriff, G., Chakraborty, T., & Manya, D. (2021). Disproportionate exposure to urban heat island intensity across major US cities. *Nature Communications*, *12*(1), 2721. http://dx.doi.org/10.1038/s41467-021-22799-5, Number: 1 Publisher: Nature Publishing Group.

Joshi, M. Y., Rodler, A., Musy, M., Guernouti, S., Cools, M., & Teller, J. (2022). Identifying urban morphological archetypes for microclimate studies using a clustering approach. *Building and Environment*, *224*, Article 109574. http://dx.doi.org/10.1016/j.buildenv.2022.109574.

Kikegawa, Y., Genchi, Y., Yoshikado, H., & Kondo, H. (2003). Development of a numerical simulation system toward comprehensive assessments of urban warming countermeasures including their impacts upon the urban buildings' energy-demands. *Applied Energy*, *76*(4), 449–466. http://dx.doi.org/10.1016/S0306-2619(03)00009-6.

Kim, J. H. (2019). Multicollinearity and misleading statistical results. *Korean Journal of Anesthesiology*, *72*(6), 558–569. http://dx.doi.org/10.4097/kja.19087.

Kleerekoper, L., van Esch, M., & Salcedo, T. B. (2012). How to make a city climate-proof, addressing the urban heat island effect. *Resources, Conservation and Recycling*, *64*, 30–38. http://dx.doi.org/10.1016/j.resconrec.2011.06.004.

Kondo, H., Genchi, Y., Kikegawa, Y., Ohashi, Y., Yoshikado, H., & Komiyama, H. (2005). Development of a multi-layer urban canopy model for the analysis of energy consumption in a big city: Structure of the urban canopy model and its basic performance. *Boundary-Layer Meteorology*, *116*(3), 395–421. http://dx.doi.org/10.1007/s10546-005-0905-5.

Kusaka, H., Kondo, H., Kikegawa, Y., & Kimura, F. (2001). A simple single-layer urban canopy model for atmospheric models: Comparison with multi-layer and slab models. *Boundary-Layer Meteorology*, *101*(3), 329–358. http://dx.doi.org/10.1023/A:1019207923078.

Li, X., Zhou, Y., Yu, S., Jia, G., Li, H., & Li, W. (2019). Urban heat island impacts on building energy consumption: A review of approaches and findings. *Energy*, *174*, 407–419. http://dx.doi.org/10.1016/j.energy.2019.02.183.

López-Moreno, H., Núñez Peiró, M., Sánchez-Guevara, C., & Neila, J. (2022). On the identification of homogeneous urban zones for the residential buildings' energy evaluation. *Building and Environment*, *207*, Article 108451. http://dx.doi.org/10.1016/j.buildenv.2021.108451.

Luo, X., Hong, T., & Tang, Y.-H. (2020). Modeling thermal interactions between buildings in an urban context. *Energies*, *13*(9), 2382. http://dx.doi.org/10.3390/en13092382, Number: 9 Publisher: Multidisciplinary Digital Publishing Institute.

Ma, J., & Cheng, J. C. P. (2016). Estimation of the building energy use intensity in the urban scale by integrating GIS and big data technology. *Applied Energy*, *183*, 182–192. http://dx.doi.org/10.1016/j.apenergy.2016.08.079.

Määttä, I., Ferreira, T., & Leßmann, C. (2021). Nighttime lights and wealth in very small areas. *Review of Regional Research*, http://dx.doi.org/10.1007/s10037-021-00159-6.

Microsoft (2018). USBuildingFootprints: Computer generated building footprints for the United States. *Microsoft*.

Ooka, R. (2007). Recent development of assessment tools for urban climate and heat-island investigation especially based on experiences in Japan. *International Journal of Climatology*, *27*(14), 1919–1930. http://dx.doi.org/10.1002/joc.1630, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.1630.

Özyavuz, M. (2013). *Advances in landscape architecture*. Rijeka: IntechOpen, http://dx.doi.org/10.5772/51738.

Palmer, M. A., Bernhardt, E. S., Chornesky, E. A., Collins, S. L., Dobson, A. P., Duke, C. S., Gold, B. D., Jacobson, R. B., Kingsland, S. E., Kranz, R. H., Mappin, M. J., Martinez, M. L., Micheli, F., Morse, J. L., Pace, M. L., Pascual, M., Palumbi, S. S., Reichman, O. J., Townsend, A. R., & Turner, M. G. (2005). Ecological science and sustainability for the 21st century. *Frontiers in Ecology and the Environment*, *3*(1), 4–11, Publisher: Ecological Society of America.

Pérez-Lombard, L., Ortiz, J., & Pout, C. (2008). A review on buildings energy consumption information. *Energy and Buildings*, *40*(3), 394–398. http://dx.doi.org/10.1016/j.enbuild.2007.03.007.

PLUTO data dictionary May 2022 (22v1). (2022). (p. 55).

Reinhart, C. F., & Cerezo Davila, C. (2016). Urban building energy modeling – A review of a nascent field. *Building and Environment*, *97*, 196–202. http://dx.doi.org/10.1016/j.buildenv.2015.12.001.

Robineau, T., Rodler, A., Morille, B., Ramier, D., Sage, J., Musy, M., Graffin, V., & Berthier, E. (2022). Coupling hydrological and microclimate models to simulate evapotranspiration from urban green areas and air temperature at the district scale. *Urban Climate*, *44*, Article 101179. http://dx.doi.org/10.1016/j.uclim.2022.101179.

Salvati, A., Palme, M., Chiesa, G., & Kolokotroni, M. (2020). Built form, urban climate and building energy modelling: Case-studies in Rome and Antofagasta. *Journal of Building Performance Simulation*, *13*(2), 209–225. http://dx.doi.org/10.1080/19401493.2019.1707876, Publisher: Taylor & Francis _eprint:.

Santamouris, M., Cartalis, C., Synnefa, A., & Kolokotsa, D. (2015). On the impact of urban heat island and global warming on the power demand and electricity consumption of buildings—A review. *Energy and Buildings*, *98*, 119–124. http://dx.doi.org/10.1016/j.enbuild.2014.09.052.

Sellers, W. D. (1969). A global climatic model based on the energy balance of the earth-atmosphere system. *Journal of Applied Meteorology and Climatology*, *8*(3), 392–400. http://dx.doi.org/10.1175/1520-0450(1969)008<0392:AGCMBO>2.0.CO;2, Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology.

Shen, P., Liu, J., & Wang, M. (2021). Fast generation of microclimate weather data for building simulation under heat island using map capturing and clustering technique. *Sustainable Cities and Society*, *71*, Article 102954. http://dx.doi.org/10.1016/j.scs.2021.102954.

Shi, Y., Ren, C., Zheng, Y., & Ng, E. (2016). Mapping the urban microclimatic spatial distribution in a sub-tropical high-density urban environment. *Architectural Science Review*, *59*(5), 370–384. http://dx.doi.org/10.1080/00038628.2015.1105195, Publisher: Taylor & Francis _eprint:.

Small, C., Elvidge, C. D., & Baugh, K. (2013). Mapping urban structure and spatial connectivity with VIIRS and OLS night light imagery. In *Joint urban remote sensing event 2013* (pp. 230–233). http://dx.doi.org/10.1109/JURSE.2013.6550707.

Srebric, J., Heidarinejad, M., & Liu, J. (2015). Building neighborhood emerging properties and their impacts on multi-scale modeling of building energy and airflows. *Building and Environment*, *91*, 246–262. http://dx.doi.org/10.1016/j.buildenv.2015.02.031.

Toparlar, Y., Blocken, B., Maiheu, B., & van Heijst, G. J. F. (2017). A review on the CFD analysis of urban microclimate. *Renewable and Sustainable Energy Reviews*, *80*, 1613–1640. http://dx.doi.org/10.1016/j.rser.2017.05.248.

Toparlar, Y., Blocken, B., Maiheu, B., & van Heijst, G. J. F. (2018). Impact of urban microclimate on summertime building cooling demand: A parametric analysis for Antwerp, Belgium. *Applied Energy*, *228*, 852–872. http://dx.doi.org/10.1016/j.apenergy.2018.06.110.

Torabi Moghadam, S., Toniolo, J., Mutani, G., & Lombardi, P. (2018). A GIS-statistical approach for assessing built environment energy use at urban scale. *Sustainable Cities and Society*, *37*, 70–84. http://dx.doi.org/10.1016/j.scs.2017.10.002.

Uppala, S. M., KÅllberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., ... Woollen, J. (2005). The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, *131*(612), 2961–3012. http://dx.doi.org/10.1256/qj.04.176, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1256/qj.04.176.

Weclawiak, I. A. (2022). *Viability and accessibility of urban heat island and lake microclimate data over current TMY weather data for accurate energy demand predictions* [Ph.D. thesis], Cleveland State University, Retrieved July 25, 2022 from https://etd.ohiolink.edu/apexprod/rws_olink/r/1501/10?clear=10&p10_accession_num=csu1656429382303039.

Xu, L., Tong, S., He, W., Zhu, W., Mei, S., Cao, K., & Yuan, C. (2022). Better understanding on impact of microclimate information on building energy modelling performance for urban resilience. *Sustainable Cities and Society*, *80*, Article 103775. http://dx.doi.org/10.1016/j.scs.2022.103775.

Yang, X., Peng, L. L. H., Jiang, Z., Chen, Y., Yao, L., He, Y., & Xu, T. (2020). Impact of urban heat island on energy demand in buildings: Local climate zones in Nanjing. *Applied Energy*, *260*, Article 114279. http://dx.doi.org/10.1016/j.apenergy.2019.114279.

Yang, J., Wang, Y., Xue, B., Li, Y., Xiao, X., Xia, J. C., & He, B. (2021). Contribution of urban ventilation to the thermal environment and urban energy demand: Different climate background perspectives. *Science of the Total Environment*, *795*, Article 148791. http://dx.doi.org/10.1016/j.scitotenv.2021.148791.

Yin, C., Yuan, M., Lu, Y., Huang, Y., & Liu, Y. (2018). Effects of urban form on the urban heat island effect based on spatial regression model. *Science of the Total Environment*, *634*, 696–704. http://dx.doi.org/10.1016/j.scitotenv.2018.03.350.