Sustainable Cities and Society 72 (2021) 102978

Contents lists available at ScienceDirect

Sustainable Cities and Society

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

Context-aware Urban Energy Analytics (CUE-A): A framework to model relationships between building energy use and spatial proximity of urban systems

Ranjitha Shivaram [a], Zheng Yang [b], Rishee K. Jain [b,*]

[a] Emmett Interdisciplinary Program in Environment and Resources, Stanford University, 473 Via Ortega, Stanford, CA 94305, United States
[b] Urban Informatics Lab, Department of Civil & Environmental Engineering, Stanford University, 473 Via Ortega, Stanford, CA 94305, United States

ARTICLE INFO

Keywords:
Building energy use
Urban energy use
Data analysis
Energy efficiency
Spatial proximity
Multivariate multiple regression

ABSTRACT

Cities account for over 75% of primary energy use in the world, with buildings making up a significant share of this energy use. Previous *simulation-based* research has established that building energy use can be greatly impacted by surrounding urban systems such as other buildings, vegetation, and roads. Understanding these relationships is thus critical to enhancing the efficiency of energy-intensive urban environments. Taking advantage of the recent profusion of urban data, this paper proposes a novel Context-aware Urban Energy Analytics (CUE-A) framework to *empirically* extract and quantify the relationships between building energy use and the spatial proximity of multiple surrounding urban systems. We apply the CUE-A framework to a case study of 477 buildings in a mid-size U.S. city to demonstrate its merits and the statistical significance of explored relationships. Results show that spatial proximity of other buildings, trees, and roads is associated with varied and significant changes in both the central tendency and variability of building energy use, indicating that empirical frameworks, which are a growing field of work, can serve as useful complements to existing simulation models. Further, our paper demonstrates that energy-aware urban planning and design has the potential to unlock energy efficiency and low-carbon pathways for cities around the world.

1. Introduction

Cities account for more than 75% of global energy use, with the majority of this consumption coming from buildings (United Nations, 2015). Cities also account for more than 60% of global greenhouse gas emissions, making the energy efficiency of the urban built environment critical to meeting the world's climate change and sustainability goals. Due to the development of new information technologies, massive amounts of data are now being collected on an array of urban systems, defined here as systems of urban elements such as buildings, trees, and roads (each building in an urban area, for instance, is an urban element, and the collection of all buildings forms the urban system of buildings). These data are opening up opportunities to optimize urban energy use in ways that were previously not possible.

A building's energy use often depends on a variety of building characteristics such as built area, age, construction materials, HVAC characteristics, occupant behavior, and building operation and maintenance (O&M). More importantly, urban building energy use can be impacted not just by building characteristics, but also by surrounding urban systems, including surrounding buildings (e.g., due to heat islands, mutual shading and reflection, and occupant dynamics), roads (e.g., due to street surfaces and heat from traffic), and trees (e.g., due to shading and air flow) (Chen, Hong, & Piette, 2017; Liu, Heidarinejad, Gracik, & Srebric, 2015; Toparlar et al., 2015). These relationships are complex and context-specific. For instance, shade from surrounding trees can reduce cooling loads in the summer but increase heating loads in winter for buildings in some climates and contexts (Akbari, Kurn, Bretz, & Hanford, 1997). In other contexts, trees might shelter buildings from strong cold winds and reduce the heating load in winter depending on the tree type.

While previous work has acknowledged that surrounding systems can affect urban building energy use, there is a need to empirically quantify the relationship between building energy use and spatial proximity of surrounding urban systems. The availability of building energy consumption data with greater spatiotemporal granularity and the ability to extract spatial relationships between urban features have

* Corresponding author.
E-mail addresses: ranjshiv@stanford.edu (R. Shivaram), zhengy@stanford.edu (Z. Yang), rishee.jain@stanford.edu (R.K. Jain).

https://doi.org/10.1016/j.scs.2021.102978
Received 15 June 2020; Received in revised form 2 November 2020; Accepted 24 April 2021
Available online 30 April 2021
2210-6707/

Table 1
Summary of commonly used input and output variables in prior research.

CATEGORY	VARIABLES
INPUT VARIABLES	
Urban contextual features	
Buildings	Mutual shading, Mutual reflection (Chen et al., 2017; Han et al., 2017; Li & Wong, 2007; Liu et al., 2015; Pisello et al., 2012; Pisello, Castaldo, Taylor et al., 2014)
Roads	Surface reflection, Heat from vehicles (Ihara et al., 2008; Karan et al., 2016; Lee & Kim, 2015; Mohammadi & Taylor, 2017)
Vegetation	Shading, Air flow (Akbari et al., 1997; Djedjig et al., 2016; Morakinyo et al., 2016; Perini & Magliocco, 2014; Wang, 2014)
Microclimate and local climate features	Relative humidity, Wind speed and direction, Temperature, Precipitation, Radiation, Anthropogenic heat sources, Evapotranspiration, Heating degree days, Cooling degree days (Barry & Blanken, 2016; Bouyer et al., 2011; Heidarinejad et al., 2016; Lauzet et al., 2019; Toparlar et al., 2015)
Building features	
Building geometry	Surface characteristics, Building height, Building shape, Window to wall ratio (WWR), Orientation, Location, Floor-to-area ratio (FAR) (Abanda & Byers, 2016; Anton & Tănase, 2016; Barry & Blanken, 2016; Bouyer et al., 2011; Heidarinejad et al., 2016; Pacheco et al., 2012; Samuelson et al., 2016; Toparlar et al., 2017; Tuhus-Dubrow & Krarti, 2010; Yi & Malkawi, 2009)
Building type	Type of building/ type of land use (e.g. residential/ commercial/ industrial) (Aksoezen et al., 2015; Porse et al., 2016)
Building condition	Age, Operation and maintenance (O&M) characteristics, Occupancy characteristics (Aksoezen et al., 2015; Porse et al., 2016)
OUTPUT VARIABLES	
Energy use features	Building energy use, Building energy use intensity
Energy-relevant features	Temperature, Air flow, Thermal comfort

enabled researchers to explore empirical methods (in addition to simulations) to understand how urban context - defined here as the spatial proximity of different urban systems such as buildings, roads, and vegetation - can influence building energy use.

This paper proposes a Context-aware Urban Energy Analytics (CUE-A) framework to investigate and empirically quantify the relationships between building energy use and the spatial proximity of other urban systems. The CUE-A framework contains four steps: (1) Extract features on spatial proximity for each building in an urban area through a series of physical relationship learning algorithms, (2) Extract building energy use features from granular energy consumption data, (3) Model the relationships between building energy use features and spatial proximity features through multivariate multiple regression, and (4) Apply learned relationships in data-informed urban energy decision making. The physical relationship learning algorithms, developed in (Gupta, Yang, & Jain, 2019), consist of three modules: a pre-processing module for urban data structure standardization, a learning module to search and calculate the proximity relationships of urban elements, and a management module to reconstruct urban data in a graph database based on learned proximity relationships (for an in-depth understanding of the model, see (Gupta et al., 2019)).

The rest of the paper is organized as follows: Section 2 presents the motivation for this research through a succinct review of existing literature and the identification of gaps. Section 3 lays out the generalized CUE-A framework, including both the theoretical grounding and the process by which analysis can be conducted using the framework. To validate the CUE-A framework, a real-world case study on 477 buildings in a mid-size U.S. city is presented in Section 4, where we find that building energy use has varied relationships in both magnitude and direction to the spatial proximity of buildings, roads, and trees. Section 5 discusses the implications of this work, while Section 6 lays out the limitations and offers insights into future work.

2. Literature review

In the following section, we offer a succinct review and classification of prior work on the relationships between urban building energy use and the spatial proximity of urban systems, with the goals of identifying research gaps and situating the current work.

Prior work (see full list of citations in Table 1) has, through modeling and simulation, explored the impacts of a building's own characteristics as well as urban contextual features on aspects of building energy use. Table 1 summarizes some common input and output variables employed in previous models, wherein research has typically explored relationships between one or two input variables and an output variable. We classify input variables from prior work into urban contextual features, microclimate/ local climate features and building features. The spatial proximity of contextual features is central to their effects on building energy use. For instance, the closer the surrounding buildings are, the greater the mutual shading potential from surrounding buildings and its effect on building energy use. The output variables are classified into two types - energy use features which directly quantify building energy use, and energy-relevant features that can influence (and be influenced by) building energy use such as temperature, air flow, and thermal comfort of occupants. In the following paragraphs, we summarize prior research by key input variable categories listed in the table.

2.1. *Relationships between urban contextual features and building energy use*

In urban areas, buildings, roads, and vegetation are clustered together in close spatial proximity. This adjacency and closeness can affect the energy use of urban buildings through mechanisms that are detailed in the following subsections.

2.1.1. *Buildings*

The higher proximity of buildings in dense urban areas can impact building energy use through mutual shading and reflection, together referred to as inter-building effects (IBE). Modeling these inter-building effects allows researchers to understand the mutual impacts of nearby buildings on building energy use instead of considering individual buildings as stand-alone objects. For instance, Pisello et al. find that neglecting mutual shading and reflection can result in substantial inaccuracies (up to 42% in summer and up to 22% in winter) of heating and cooling load prediction (Pisello, Taylor, Xu, & Cotana, 2012). Mutual shading can also affect lighting energy demand (Li & Wong, 2007; Pisello, Castaldo, Taylor, & Cotana, 2014) and the energy performance of building retrofits (Chen et al., 2017). Further, the relative magnitude of inter-building effects is often context-dependent. For instance, Han, Taylor, and Pisello (2017) find that shading has a greater effect than reflection on energy consumption in an urban building network in Perugia, Italy. Inter-building effects can also vary with different urban layouts. For instance, using a simulation approach, Liu et al. (2015) find that building cooling loads increase and heating loads decrease with increasing plan area density due to convective heat transfer from exterior surfaces that alter building temperatures.

2.1.2. Roads

Recent work has begun to establish clear connections between transportation networks and building energy use (Karan, Mohammadpour, & Asadi, 2016; Mohammadi & Taylor, 2017). For instance, Lee and Kim (2015) find that transportation accounts for a significant share of anthropogenic heat emissions in the urban Gyeong-In region, Korea, which can affect building energy use. Incorporating the effect of exhaust heat from automobiles into a building energy model, Ihara, Kikegawa, Asahi, Genchi, and Kondo (2008) find that reducing waste heat from automobiles can result in reduced cooling loads in buildings. However, the explicit modeling of the effects of road proximity on building energy use is quite nascent and presents an opportunity for further research.

2.1.3. Vegetation

Research has long recognized, through detailed simulations, that urban vegetation such as parks, trees, and green roofs can reduce energy use in buildings by moderating shade and air flow (Akbari et al., 1997). The magnitude of these effects depends on the degree of proximity of vegetation, their relative abundance in urban areas, and local climate conditions. Perini and Magliocco (2014) find that urban vegetation can mitigate summer temperatures and decrease the indoor cooling demand. Morakinyo, Dahanayake, Kalani, Adegun, and Balogun (2016) find that tree shade on a building can result in lower indoor and outdoor temperatures and higher relative humidity. Similarly, using Monte Carlo simulation, Wang (2014) establishes that shade from trees can lower temperatures of a street canyon and reduce energy use in buildings. Focusing on the interfaces between buildings and roads, Djedjig, Bozonnet, and Belarbi (2016) find that green walls can reduce building energy use and street air temperature.

Together, the clustering of buildings, roads, and other human-made elements in a dense urban environment contributes to the urban heat island effect (UHI) (Arnfield, 2003; Howard, 1833), defined by higher local air temperatures in dense urban environments compared to surrounding rural areas (Kim, Gu, & Kim, 2018; Oke, 1982; Santamouris, 2001), which can in turn affect building energy use. However, since the case study in this paper explores intra-urban relationships in a small downtown area of a single city, a detailed consideration of local variations in the heat island effect is outside the scope. Such variations will, however, need to be considered in greater depth for larger study areas.

2.2. Microclimate/ local climate conditions

Extensive literature in urban climatology has established that a variety of local climate factors – such as solar irradiance (Bouyer, Inard, & Musy, 2011; Heidarinejad et al., 2016), temperature (Toparlar et al., 2015), pressure, and air flow – can significantly impact building energy use. While definitions vary, *local climate* is typically defined at a neighborhood/ city scale while *microclimates* are confined to smaller areas, such as a street canyon between two rows of buildings (Barry & Blanken, 2016) (city and regional scales are often called 'meso' scales). It is thus important to account for these variables since oversimplifying or neglecting microclimatic effects can lead to substantial inaccuracies in building energy models (Lauzet et al., 2019). For a detailed review of microclimatic effects on building energy use, including recent developments in computational fluid dynamics (CFD) methods, see the comprehensive review by Toparlar, Blocken, Maiheu, and van Heijst (2017).

2.3. Effects of building features

A building's own characteristics can not only impact energy use, but also mitigate or exacerbate the effects of other urban systems. For instance, older buildings with less efficient HVAC systems and poor insulation might be affected more by heat from traffic than newer buildings with more efficient HVAC systems and better insulation. It is thus important to account for inherent effects of building characteristics when studying the effects of urban context on building energy use. The following subsections briefly review key building features that can impact building energy use.

2.3.1. Building geometry and orientation

A wide range of energy modeling and simulation studies have established that the shape, size, and orientation of buildings can affect building energy use. For instance, Samuelson, Claussnitzer, Goyal, Chen, and Romo-Castillo (2016) find that various early-stage parameters for building design such as shape, window to wall ratio (WWR), envelope, and shading can affect the energy use of a residential building. For a comprehensive review, see Pacheco, Ordóñez, and Martínez (2012). Many algorithms have been developed to optimize building geometry for enhanced energy efficiency (Tuhus-Dubrow & Krarti, 2010; Yi & Malkawi, 2009), including recent work on parametric building modeling and simulation (Anton & Tănase, 2016). Building orientation can impact the internal solar gain of a building, affecting its energy use (Abanda & Byers, 2016).

2.3.2. Building type and condition

The type of building - commercial, industrial, residential, or other - can impact its energy use due to variations in building characteristics and the relative energy intensity of activities being performed. Commercial and industrial buildings are typically larger in area and have more energy intensive activities than residential buildings. Porse, Derenski, Gustafson, Elizabeth, and Pincetl (2016) find that complex interrelated social and structural factors can influence building energy use, including age, area, and land use type. Building age, often used as a proxy for building condition, is an important but mixed indicator of energy use, since older buildings might have lower energy efficiency measures but are typically more compact than newer dwellings (Aksoezen, Daniel, Hassler, & Kohler, 2015).

In conclusion, we find that a large variety of factors influence building energy use. In particular, Section 2.1 highlighted that the spatial proximity of urban systems can influence building energy use through a range of complex and often interactive mechanisms. Since this is the relationship of interest explored in this paper, the microclimate and building features described in Sections 2.2 and 2.3 are treated as critical control variables that must be accounted for in quantifying the relationship between building energy use and spatial proximity of other urban systems.

2.4. Addressing gaps in prior research

The literature review reveals that there is a wealth of simulation-based methods in prior work that help us understand relationships between building energy use and the spatial proximity of urban systems.

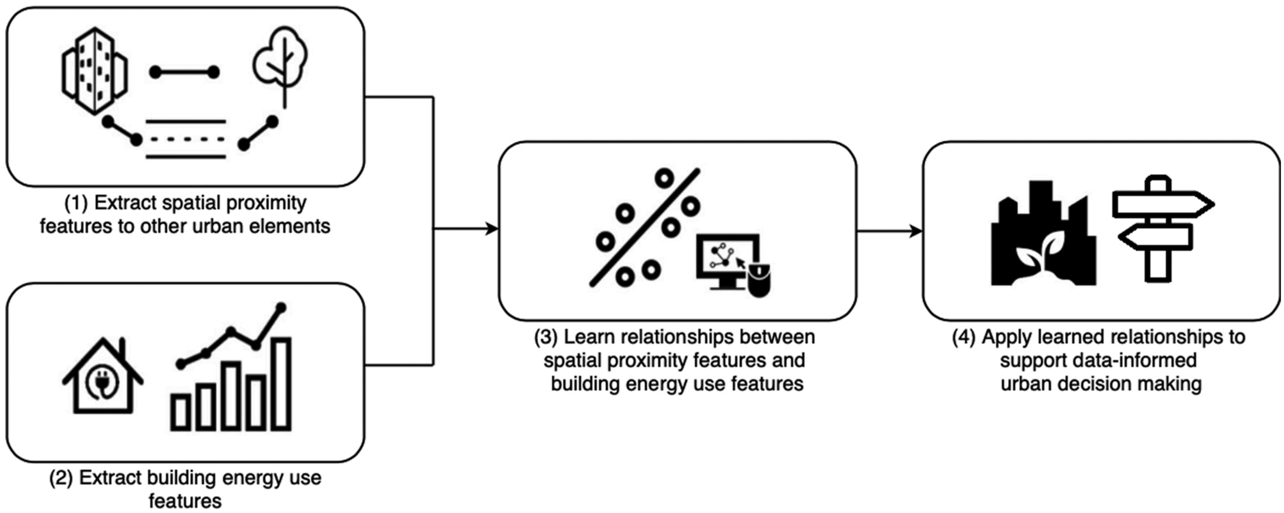

Fig. 1. Overview of the CUE-A framework.

Simulation approaches offer significant advantages – substantial control over the urban environment being modeled, the ability to isolate confounding factors, and greater confidence in results subject to specific assumptions. These advantages also pose three specific challenges that can be addressed by empirical methods as detailed below, especially as simulation-based models are time-consuming and computationally expensive.

First, empirical methods enable researchers to model real-world urban environments as they actually exist in cities today. Past work, especially when simulation-based, uses either just a single test building (Bouyer et al., 2011), or a test building in combination with simplified urban grids of building blocks that are often 10 m cubes (Han et al., 2017; Liu et al., 2015). Only in a few cases, some complexity is introduced by varying the urban context. For instance, Samuelson et al. (2016) use stochastic urban context models generated using the software Grasshopper, and Toparlar et al. (2015) generate an actual geometrical model of a section of Rotterdam using municipal building data. So, by extracting spatial proximity features that represent buildings, roads and vegetation in *actual* cities, we can begin to address problems associated with simplification of urban layouts in existing research.

Second, while building layouts are still varied to some extent in prior research, other urban features such as trees, cars, and other anthropogenic heat sources are often neglected or not considered concurrently in a single integrated model. A significant opportunity thus exists to simultaneously consider the relationships between the proximity of building, street and vegetation features and building energy use, capturing their interactions.

Third, recent benchmarking policy efforts have made energy consumption data at the individual building level readily available (Chen et al., 2017). So, we now have the opportunity to employ data on actual energy use in buildings as an output variable in empirical models, enabling real-world validation of the hypothesized/ simulated effects of urban context on building energy use. Integrating data streams on urban systems from multiple sources (e.g., municipal records, in-situ sensors, surveys) and the extraction of relationships between them is also a growing area of research. Prior work (Gupta et al., 2019) classifies existing data integration methods into three categories - domain-centric, demand-centric and data-centric. To address deficiencies in these methods, (Gupta et al., 2019) develops an Urban Data Integration (UDI) framework that uses a series of proximity relationship learning algorithms to automatically integrate urban data in a graph database. In this paper, we utilize the output of the UDI framework to extract information on urban context, i.e., spatial proximity of urban systems.

Table 2
Spatial proximity features extracted for each building.

Spatial proximity features	Feature category	Feature
Number of proximate elements	–	Number
Distance of proximate elements	Central tendency	Mean distance
		Median distance
	Variability	Minimum distance
		Maximum distance
		Standard deviation
		Interquartile range of distance
	Distribution	Skewness of distance
		Kurtosis of distance

3. CUE-A framework

The generalized CUE-A framework comprises of the following steps (Fig. 1): (1) extract features to represent the spatial proximity of buildings to other urban systems (other buildings, vegetation and roads); (2) extract features to depict the characteristics and patterns of building energy use; and (3) learn relationships between spatial proximity features and energy use features, while accounting for confounding variables. The ultimate goal of this process is to offer data-based evidence to complement existing simulation models in informing urban planning and design decisions (Step 4). The following subsections describe these steps in detail. A case study that employs this framework for a city is presented in Section 4.

3.1. Step 1: feature extraction and spatial proximity determination for urban context

The spatial proximity features used in previous studies to represent urban context can be divided into three categories: features related to buildings, features related to vegetation, and features related to roads. The features related to buildings generally describe the spatial relations among buildings, such as the number of nearby buildings and the distance between buildings (Pearlmutter, Berliner, & Shaviv, 2007; Santamouris et al., 2001). The features related to vegetation represent the spatial connections between buildings and vegetation that are either on the buildings (e.g., green roofs) or along the buildings (e.g., density of vegetation around a building) (Perini & Magliocco, 2014). The features related to roads indicate the spatial relations between buildings and roads such as the distance between a building and nearby roads (Allegrini, Dorer, & Carmeliet, 2012; Tong, Chen, Malkawi, Adamkiewicz, &

Spengler, 2016).

Our proposed CUE-A framework builds on this previous work by first extracting features that depict spatial proximity between buildings in an urban area and the surrounding buildings, roads and trees through a series of physical relationship learning algorithms introduced in (Gupta et al., 2019). The algorithms rely on spatial data on the locations and geometry of buildings, roads, and trees in an urban area that can be obtained through open data from cities or extracted from map providers such as OpenStreetMap (OSM), satellite imagery such as from NASA, commercial data providers such as SafeGraph, or obtained from photogrammetry. Based on their geometric properties, buildings are abstracted to polygonal urban elements, roads are abstracted to linear urban elements, and trees are abstracted to point urban elements. Then, the relationship learning algorithm utilizes an iterative process to determine the degree to which spatial proximity exists between any two urban elements (see Appendix A for a brief explanation and (Gupta et al., 2019) for details). The algorithm iterates through all the buildings in the dataset and the results are aggregated to construct a set of spatial proximity features for each building with respect to other buildings, trees, and roads. Table 2 shows suggested features that could be extracted to represent the spatial proximity of each building to every distinct surrounding urban system.

3.2. Step 2: feature extraction for building energy use

The first step in feature extraction for building energy use is to choose key performance indicators (KPIs) that are easy to compute and interpret. Commonly used KPIs in prior work include (1) total energy use, (2) heating and cooling loads, and (3) energy use intensity (EUI). Each KPI can be of different types, such as site KPI to indicate the unit primary and secondary energy directly used at the building, source KPI to express site KPI plus delivery and production energy losses, weather-normalized site KPI to indicate the site KPI if the building experienced 30-year average outside temperature, and weather-normalized source KPI to account for the weather-normalized site KPI plus losses for total building energy use intensity (New York City Mayor's Office of Sustainability, 2016; US EPA, 2017). The temporal granularity of KPIs can be hourly, daily, monthly, and yearly, depending on the study (ASHRAE, 2002; IPMVP Committee and others, 2002; US DOE, 2008). (Note that building source energy use intensity (source EUI), where building energy use is normalized by building area, is recommended by the Portfolio Manager in EnergyStar (Scofield, 2013). Users of the CUE-A framework can choose KPIs that are appropriate to their specific context of study. In the case study presented in this paper, building area can be added as an independent variable in the regression model and so, we utilize total energy use as the primary KPI.

Next, statistics can be calculated to capture trends and patterns in the KPIs, especially if the temporal granularity of energy use data is small. These statistics become the building energy use features. Commonly used statistics can be employed to represent the central tendency (median, mean), variability (standard deviation, interquartile range (IQR), mean absolute deviation, and median absolute deviation), distribution (skewness and kurtosis) as well as kernel density, categorization and percentage (Kontokosta, 2012).

3.2.1. Control variables

In Table 1, we discuss how a variety of microclimatic features and building characteristics can impact building energy use. In order to disentangle the specific relationships between the spatial proximity of urban systems and building energy use, it is important to account for these potentially confounding characteristics. In the CUE-A framework, this can be done by adding microclimate (heating and cooling degree days, relative humidity etc.) and building characteristics (building geometry, type, and condition) as control variables to the regression model based on contextual relevance and data availability.

Fig. 2. Location of case study city in California, United States. (Source: OpenStreetMap).

3.3. Step 3: Multivariate multiple regression model

The CUE-A methodology models the relationships between spatial proximity features and building energy use through a multivariate multiple regression, wherein there are several independent variables ('multivariate') and several dependent variables ('multiple'). This regression model allows us to adequately capture time-series energy use data through a set of energy use features that describe not just the central tendency but also the variability and distribution of energy use, which are then used as independent variables ('multivariate'). Then, spatial proximity features which describe the adjacency and closeness of various urban elements (buildings, trees and roads) to buildings are employed as dependent variables ('multiple'). The spatial proximity features are assumed to be time-invariant, although a future version could model changes in the urban environment over time, providing a basis for causal deductions. Note that including multiple urban elements simultaneously enables us to investigate their co-effects on building energy use, offering a useful addition to prior models that typically account for each type of urban element individually. In addition, there are control variables that account for microclimatic features and building characteristics that could potentially confound the effects of the predictors on the outcomes.

As a result, the data has P independent variables $x_1, x_2..., x_P$ that are spatial proximity features and Q dependent variables $y_1, y_2..., y_Q$ that are building energy use features as detailed in Section 3.2. The objective is to model the statistical mapping from building spatial proximity features to energy use features, which is represented by the nonzero entries in the P by Q coefficient matrix $\mathbf{B} = (\beta_{pq})$ from the N i.i.d. buildings:

$$y_q = \sum_{p=1}^{P} x_p \beta_{pq} + \varepsilon_q \ (q = 1, \ \ldots Q)$$

The error term ε has a joint distribution of mean 0 and covariance $\Sigma\varepsilon$. With the normality assumption, β_{pq} is interpreted as proportional to the conditional correlation $Cor(y_q, x_p|x_{-(p)})$, where $x_{-(p)} : \{x_{p'} : 1 \leq p' \neq p \leq P\}$. To identify the significant impacts of proximity features on building energy use and improve the generaliz-

Fig. 3. Map of buildings (centroids shown as dots in red color), roads (lines in grey color) and vegetation (i.e., trees) (centroids shown as dots in green color) in case study area.

ability of the proposed method to other urban areas with more features, an l_1-norm is imposed to coefficient matrix **B** as regularization in order to control for overfitting and the overall sparsity of the multivariate multiple regression model. It is assumed that there exist master regulators of building characteristics and spatial proximity features for building energy use. Therefore, the objective of the model could be specified to find these master regulators:

$$L(\boldsymbol{B};\lambda) = \frac{1}{2}\left\|\boldsymbol{Y} - \sum_{p=1}^{P} X_p B_p\right\|_F^2 + \lambda \sum_{p=1}^{P} ||B_p||_1$$

Where $\boldsymbol{Y}$ is the N by Q dependent variable matrix, and $\boldsymbol{X}$ is the N by P independent variable matrix ($X_p = (x_p^1, x_p^2, \ldots, x_p^N)^T$ is the sample of p^{th} independent variable). $\boldsymbol{B}$ represents the impacts of different building characteristic features and spatial proximity features on energy use features (B_p is the p^{th} row of $\boldsymbol{B}$). $||\cdot||_F$ and $||\cdot||_1$ indicate the Frobenius norm and the l_1-norm, respectively. The final computed coefficient matrix is $\widehat{\mathbf{B}}(\lambda) = \mathrm{argmin}_B L(B;\lambda)$**.** Cross-validation estimation is applied to calculate the best hyperparameter. R-squared values are applied to denote how closely the multivariate multiple regression model fits the data.

4. Empirical validation: Case Study

4.1. Case study data and feature extraction

To validate the CUE-A framework, a real-world case study is conducted for the downtown area of Palo Alto, a mid-size city in California, USA. Fig. 2 shows the location of the city and Fig. 3 shows a map of the buildings, trees, and roads in the case study area. The case study area contains 477 buildings of varying sizes and mixed functions, including residences, offices, stores, and restaurants. The buildings studied are a mix of 43.4 % commercial, 43.1 % multi-family residential and 13.5% single-family residential building types (the 43.4% commercial category includes four land use designations employed by the City of Palo Alto - Community Commercial (CC) (39.4%), Neighborhood Commercial (CN) (1.0%), Major Institution/Special Facilities (MISP) (0.4%), and South of Forest Area (SOFA) (2.6 %)). Both the data assembly (extraction of spatial proximity features, energy use features, and control variables) and the multivariate multiple regression were conducted in Python (version 3.6).

4.1.1. Feature extraction for urban context

Spatial proximity features were computed for each building in the case study area with respect to other buildings, roads, and trees. The algorithms to extract spatial proximity rely on open data on building, street, and vegetation obtained from the City of Palo Alto (City of Palo Alto, 2020). Fig. 4 describes these spatial relationships through discrete probability distributions for the number (top row) and mean distance (bottom row) of surrounding buildings, trees, and roads for the buildings in the case study area. While the number and mean distance of

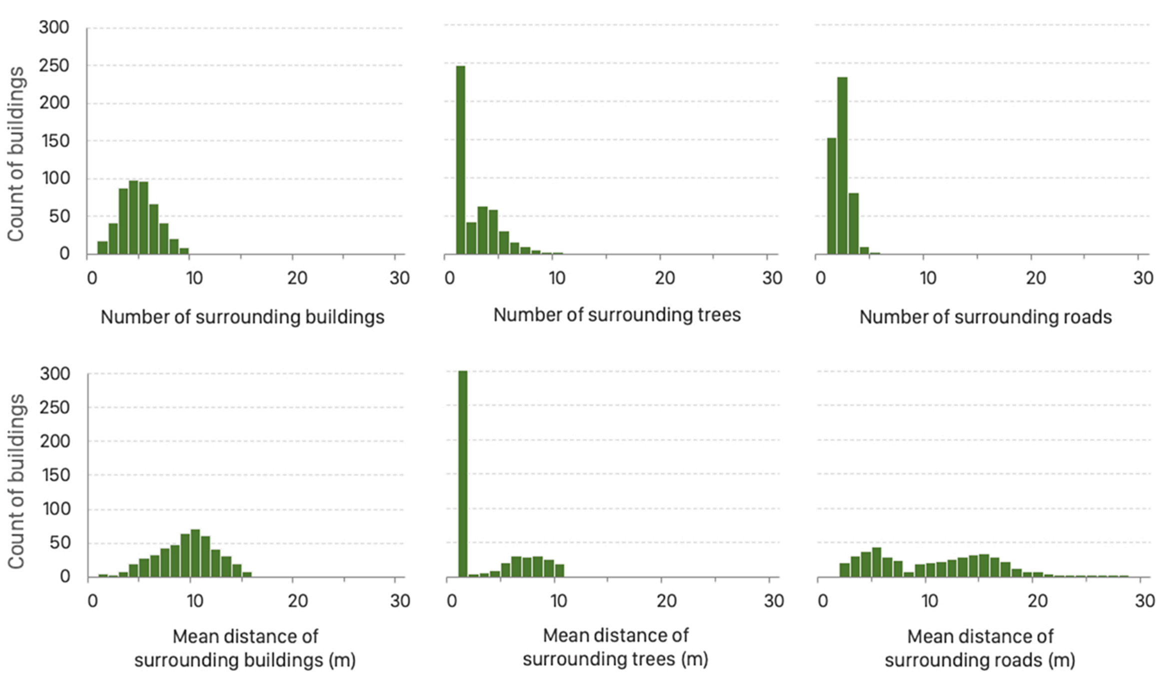

Fig. 4. Number (top row) and mean distance (bottom row) of surrounding buildings, trees, and roads for the buildings in the case study area.

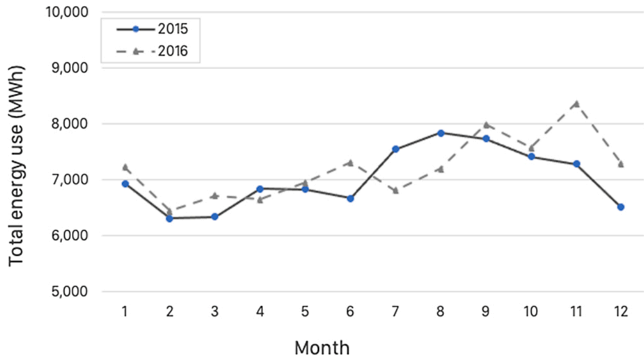

Fig. 5. Total building energy use for all buildings in the case study area by month and year.

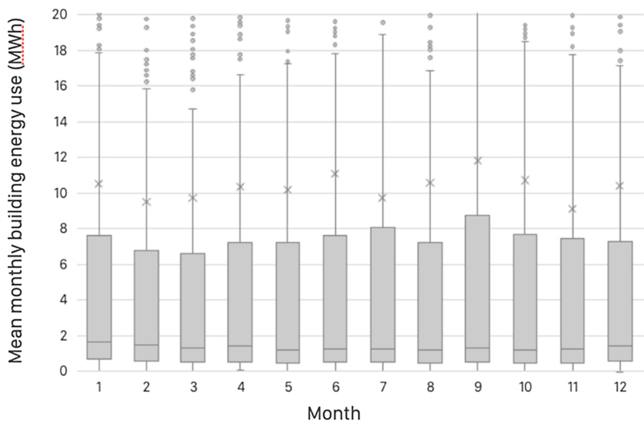

Fig. 6. Mean monthly building energy use in the case study area.

surrounding buildings are approximately normally distributed, the number of surrounding trees and roads have unimodal distributions and the distance of surrounding roads and trees have bimodal distributions. A key insight from Fig. 4 is that the majority of buildings in the case study have four to five surrounding buildings, one surrounding tree, and one to two surrounding roads. Similarly, most buildings have a mean distance of around 10 m from other buildings, of 1−2 m or 6−8 m from trees, and of 5-6m or 15-16m from roads.

4.1.2. Feature extraction for building energy use

Monthly metered energy use data for 477 buildings in the case study area for two years (2015 and 2016) was obtained from the City of Palo Alto. The building energy use features extracted from this data are the mean and standard deviation of monthly building energy use for each building, representing the central tendency and variability of energy use respectively. These features are extracted for total energy use, cooling seasonal energy use, and heating seasonal energy use for each building. For Palo Alto, June to October are assumed to be cooling months and November to May are heating months, based on heating and cooling degree day data as shown later in Fig. 7. Note that, as part of data cleansing, we discarded 53 buildings with missing energy consumption from an initial dataset of 530 buildings.

Fig. 5 shows that the total monthly energy use of all buildings in the case study area does not show significant seasonal variation by year, which can be explained by the less varied weather conditions in Palo Alto. Note that the building energy use is the sum of electricity use measured in kWh and gas consumption measured in thm (therms) and converted to kWh. Fig. 6 shows the mean monthly energy use for all buildings in the case study area. The energy use is skewed as there are some buildings with large amount and variation of energy use. The mean monthly energy use across buildings is 10,729.5 kW h, with a mean standard deviation of 2087.5 kW h and a mean range of 6802.9 kW h. Single-family and multi-family residential buildings have a mean monthly energy use of 1916.41 kW h and 2320.75 kW h respectively. In contrast, commercial buildings have higher monthly energy use, with an average of 20830.11 kW h, necessitating the incorporation of the building type variable in our regression model.

4.1.3. Control variables

As stated in Section 3, there are two primary types of control variables to be accounted for, with some customization based on the context of analysis - microclimatic features and building characteristics. In terms of microclimatic features, the city of Palo Alto has a warm-summer Mediterranean climate (Köppen Climate Classification system), with low variation in temperature, humidity and precipitation. Fig. 7 shows the Heating Degree Days (HDD) and Cooling Degree Days (CDD) aggregated to the monthly level for each of the years 2015 and 2016. Heating and cooling degree days measure the degrees by which the daily average outdoor temperature is lower or higher than the baseline temperature (65°F in the US) respectively, serving as a useful proxy for the amount of energy needed to heat or cool buildings for thermal comfort in that context. As the figure shows, there is no significant variation in Palo Alto in HDD and CDD between the two years. We thus exclude these microclimatic variables in our regression analysis for total energy use but note that these variables may become salient to case studies in other locations with significant seasonal and annual variations in weather conditions. In addition, we run the same regression model as total building energy use for the heating and cooling energy use separately to capture any variation across seasons.

The next step is to control for building characteristics. Despite the profusion of urban data, it still remains difficult to assemble a

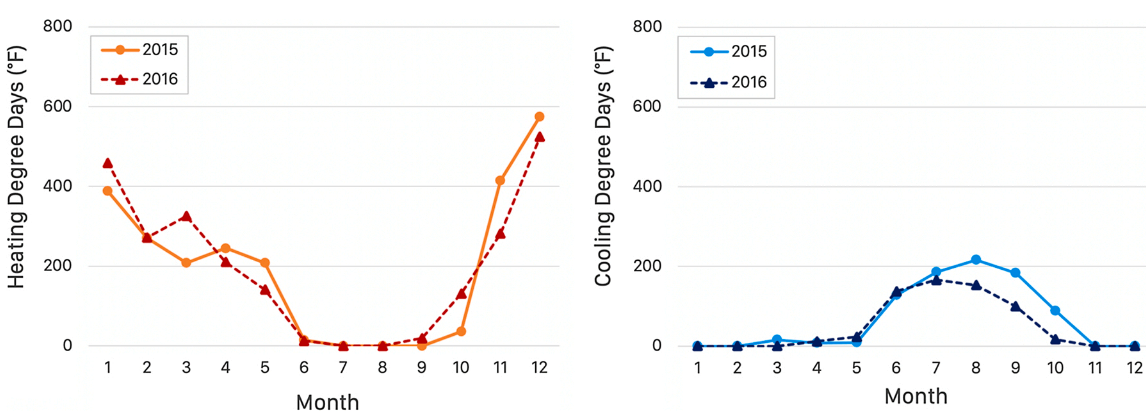

Fig. 7. Heating Degree Days (HDD) (left) and Cooling Degree Days (CDD) (right), aggregated by month for the case study area (2015-2016).
Source: NOAA National Centers for Environmental Information; Data for NOAA Division "California CD 4. Central Coast Drainage".

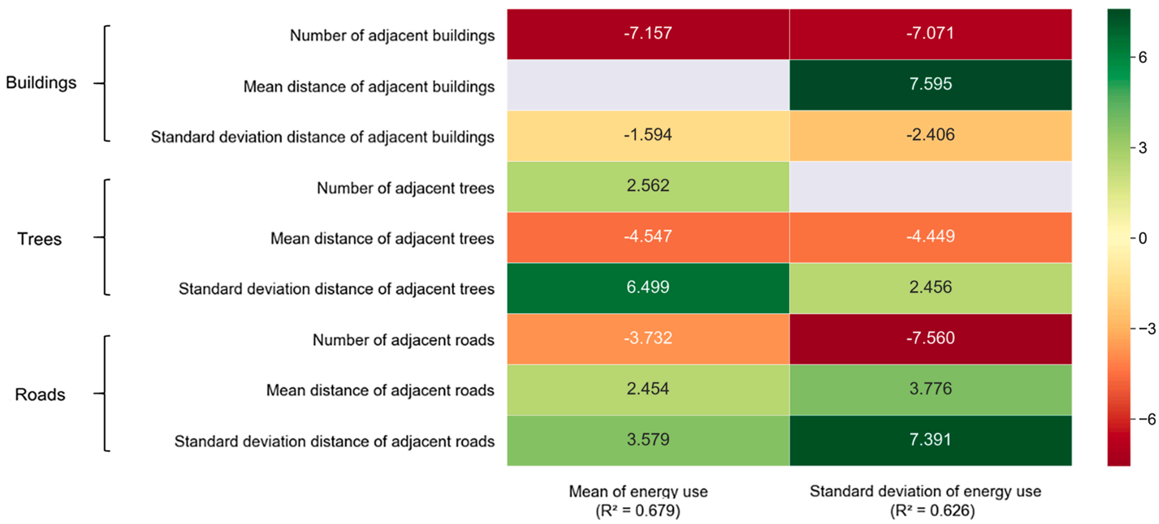

Fig. 8. Multiple multivariate regression coefficients for the relationship between building energy use features and spatial proximity features in the case study area. Note: Since the regression was conducted with log-transformed dependent variables, regression coefficients are exponentiated and converted to percent change in the figure to facilitate easier interpretation.

comprehensive dataset of characteristics for all buildings in an urban area from public data sources (Gupta et al., 2019). Due to this limitation, we accounted for a subset of building characteristics, focusing on the most critical features suggested by prior literature: building geometry and orientation, building type, and building condition. Building geometry is represented by building height and roof area, which is assumed to be a proxy for the typical floor area. Adding in building area as a control variable accounts for the variation in building size that results from building type (e.g. commercial buildings are typically larger than residential buildings). Building orientation is a categorical variable with four values representing the axis along which the longest side of the building is facing one of the four directions (0 degrees Solar North, 45 degrees West Solar North, 90 degrees Solar North, 135 degrees West Solar North). Building type is identified from parcel-level land use data obtained for the city. There are three types of buildings in the case study area – single-family residential buildings, multi-family residential buildings, and commercial buildings. Finally, we utilize building age as a proxy for building condition and construction practices given the fact that the California building code has progressed to more aggressive energy efficiency standards over the years.

4.2. Case study results: Multivariate multiple regression model

After the relevant features for each building were prepared, multivariate multiple regression was implemented to understand the relationship between building energy use and spatial features. The regression uses log-transformed dependent variables to account for skewness. Using L1 regularization, shrinkage is implemented to shrink coefficients to zero, producing a simple and sparse model that displays only the statistically impactful variables.

Fig. 8 shows a heatmap representation of the regression coefficients, where the x-axis shows building energy use features (dependent variables, namely the mean and standard deviation of building total energy use) and the y-axis shows spatial proximity features for surrounding buildings, trees, and roads (independent variables, namely the number, mean and standard deviation of distance of proximate elements). The numeric values represent percent change in the dependent variable per unit change in the independent variable (log-transformed regression

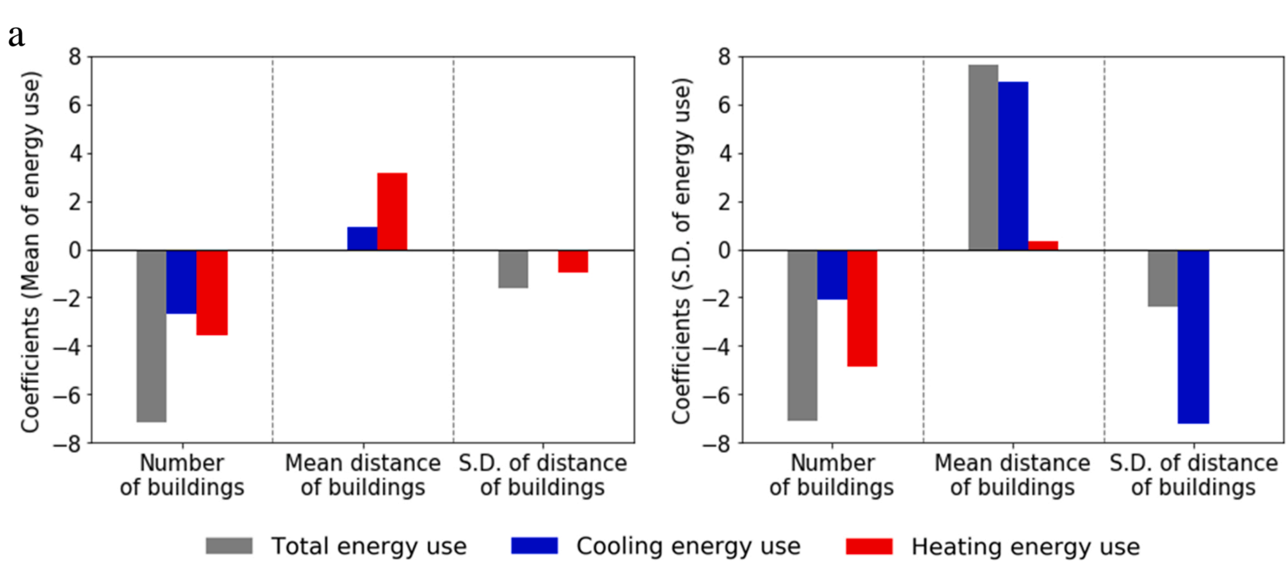

Fig. 9a. Significant coefficients for building energy use and spatial proximity of other buildings. (b) Significant coefficients for building energy use and spatial proximity of trees. (c) Significant coefficients for building energy use features and spatial proximity of roads.

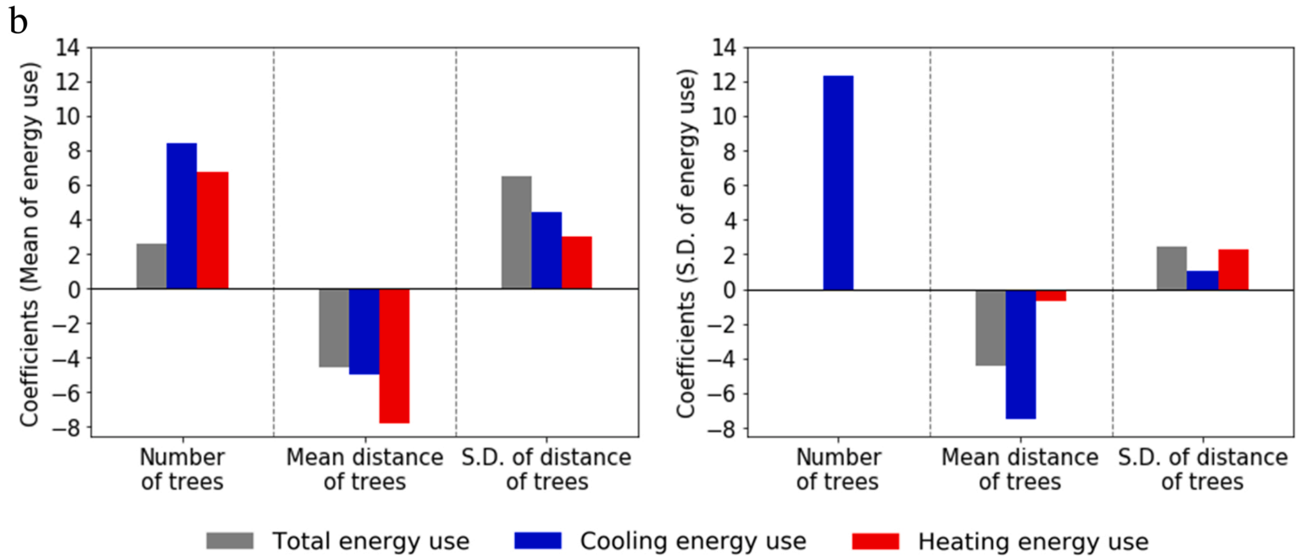

Fig. 9b. Significant coefficients for building energy use and spatial proximity of trees.

coefficients are exponentiated and converted to percent change to facilitate easier interpretation). The colors show the variation in coefficient values from large and positive (deep green) to large and negative (deep red). The subsequent Fig. 9a–c in this section further describe the coefficients for the mean and standard deviation of building cooling seasonal energy use and building heating seasonal energy use in addition to the building total energy use (for the full regression table with control variables, refer to Appendix B (Table B1)).

We see that building energy use has varied relationships in both magnitude and direction to the spatial proximity of buildings, trees, and roads. Having selected for the most impactful variables, we see that nearly every spatial proximity feature is associated with statistically significant changes in the central tendency (mean as seen in column 1) and variability (standard deviation as seen in column 2) of building energy use, pointing to the need to account for these variables in urban building energy models. The R-squared values for the mean of energy use and the standard deviation of energy use are 0.679 and 0.626 respectively.

A general guide to the interpretation of coefficients is as follows: coefficients for the numbers of buildings, trees, and roads indicate the percentage change in the mean or standard deviation of building energy use associated with a unit increase in the number of surrounding proximate elements. The coefficients for the mean distance of buildings, trees and roads indicate the percentage change in mean or standard deviation of energy use associated with a unit increase in the mean distance of surrounding proximate elements. Intuitively, it describes the association between energy use and surrounding urban elements being farther away. Coefficients for the standard deviation of distance of buildings, trees and roads indicate the percentage change in mean or standard deviation of energy use associated with a unit increase in the standard deviation of distance of surrounding elements. As standard deviation of distance increases, there is reduced concentration of surrounding elements around the mean distance of that specific element, so these urban elements are more varied in their distance from buildings. Thus, it describes the association between energy use and more spread-out surrounding urban elements.

Looking across column 1 of the heat map, we see that the standard deviation of the distance of trees has the largest positive coefficient (+6.499) and is associated with the largest positive percent change in the mean of energy use controlling for all other variables. We can thus expect building energy use to increase when trees are more varied in their distance from the buildings. The number of proximate buildings has the largest negative coefficient (−7.157), and therefore is associated with the largest negative percent change in mean of energy use. In other words, as more buildings are clustered together, we can expect building energy use to decrease. Looking across column 2 of the heat map, we see that the mean distance of buildings has the largest positive coefficient (+7.595), while the number of roads has the largest negative coefficient (−7.560), controlling for all other variables. As a result, we can expect building energy use to be more varied as surrounding buildings are farther away and to be less varied as the number of surrounding roads increases.

Fig. 9a–c offer further detail on the statistically significant regression coefficients for total building energy use, adding in the coefficients for building cooling and heating seasonal energy use for comparison (for the full regression table, refer to Appendix B (Table B1)). Note again that coefficients have been regularized to improve interpretability. In the following paragraphs, we interpret, in the same order, the association between the coefficients for the number, mean distance and standard deviation of distance of each urban element and the mean and standard deviation of total, cooling seasonal, and heating seasonal energy use.

Digging deeper into coefficients for surrounding buildings in Fig. 9a , we see that an additional proximate building is associated with a decrease in the mean of building total energy use by 7.157 percent and with a decrease in the standard deviation of building total energy use by 7.071 percent, controlling for all other variables. Intuitively, the clustering together of a larger number of buildings is associated with lower mean and lower variation in building energy use. This could point to mutual reflection and shading effects in larger clusters of buildings resulting in lower energy use across all seasons in temperate Palo Alto, and is consistent with earlier literature (Li & Wong, 2007; Pisello et al., 2012). The direction of these effects is the same for both cooling and heating seasonal energy use, although the magnitude of the coefficients is higher for heating seasonal energy use. A possible explanation is that,

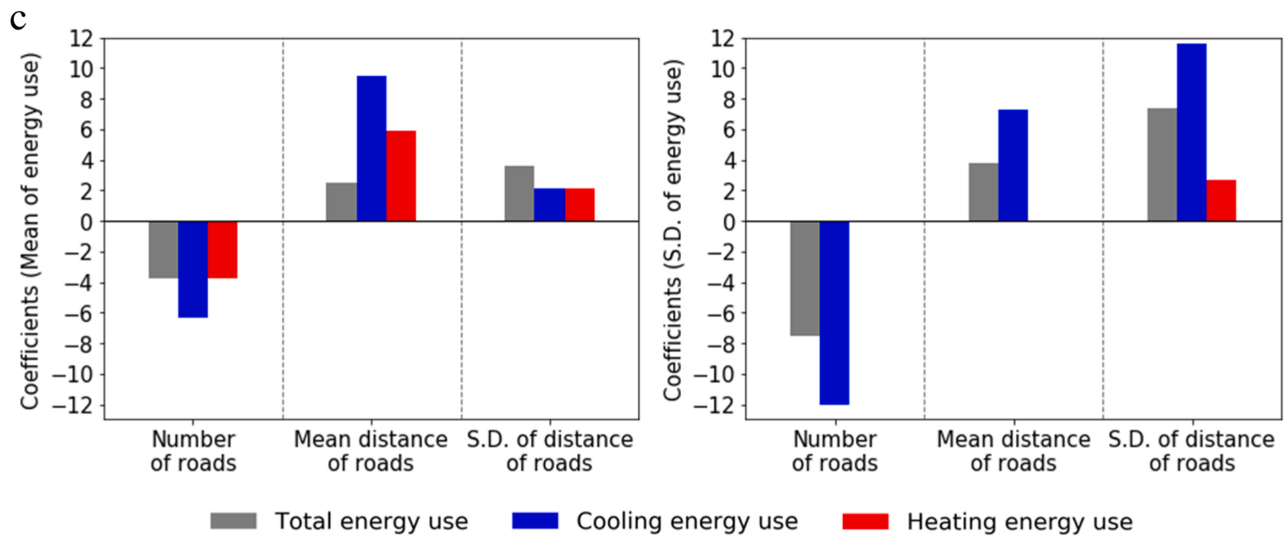

Fig. 9c. Significant coefficients for building energy use features and spatial proximity of roads.
Note: S.D. of distance refers to the standard deviation of distance of buildings.

in Palo Alto, the number of heating degree days is higher than cooling degree days, which means more heating is required. Urban heat island effects during the heating season could reduce the heating load and associated energy use.

Surprisingly, every additional unit of mean distance of surrounding buildings is associated with no significant change in mean building total energy use, indicating its effects may be offset by the effects of number and standard deviation of surrounding buildings. It is, however, associated with small changes in mean building cooling and heating seasonal energy use, indicating that such seasonal loads are more impacted by inter-building effects (Pisello et al., 2012). Every additional unit of mean distance of surrounding buildings is associated with a 7.595 percent increase in the standard deviation of total energy use, and 6.908 and 0.313 percent increases in the standard deviation of cooling and heating seasonal energy use respectively. This indicates that surrounding buildings being further away is associated with greater variation in building energy use. This could be explained by reduced mutual shading and reflection or by spatial separation of land uses in the U.S. context (Huang, Zhang, & Li, 2007), whereby energy-intensive commercial and industrial buildings are placed farther away from residential buildings. Finally, every additional unit of standard deviation of distance of surrounding buildings is associated with a decrease of 1.594 percent in mean of building total energy use and with a decrease of 2.406 percent in the standard deviation of building total energy use. This implies that, the more spread out the surrounding buildings are, the lesser the possible effects on energy use of other buildings.

Looking at coefficients for trees in Fig. 9b, we see that an additional proximate tree is associated with an increase in mean of total energy use by 2.562 percent and with no significant change in the standard deviation of total energy use, controlling for all other variables. Our results indicate a relationship between having more trees and greater mean total building energy use as well as cooling and heating seasonal energy use. Additionally, it is worth noting that the number of surrounding trees has the largest impact on the variation of cooling seasonal energy use. This result points to the notion that shading provided by surrounding trees can impact cooling seasonal energy in varying degrees. This finding is a counter-intuitive finding since earlier literature has pointed to the moderating effects of surrounding trees on building energy use during both hot and cold seasons (Akbari et al., 1997). However, there are other explanations for this effect. First, this regression explores the co-effects of multiple urban system elements on building energy use, which means that there could be interactions between effects of different surrounding urban systems. Further, this relationship could be an artifact of sprawling American land use patterns whereby larger commercial buildings are located on tree-lined avenues while smaller residential buildings might have more sparse surrounding vegetation (Huang et al., 2007). Similarly, every additional unit of mean distance of surrounding trees is associated with a decrease of 4.547 percent in the mean of building energy use, and with a 4.449 percent decrease in the standard deviation of energy use. As trees are farther away, building energy use is lower and less varied with the impact especially pronounced for cooling seasonal energy use. This finding contrasts with existing literature. We note that this result could be due to an artifact of land use patterns in our case study city in which trees are more prevalent in business districts that have buildings with higher energy use. However, further cross-analysis and comparison with other cities with different land use patterns is required to ascertain if this is the case. Finally, every additional unit of standard deviation of distance of trees is associated with an increase of 6.199 percent in mean of building energy use and with an increase of 2.456 percent in the standard deviation of building energy use. The more spread out surrounding trees are, the lower the ability of the trees to buffer external environment loads.

Looking at specific coefficients for roads in Fig. 9c, we see that an additional proximate road is associated with a decrease in mean building energy use by 3.732 percent and with a decrease in the standard deviation of energy use by 7.560 percent, controlling for all other variables. Having more roads nearby is associated with lower mean building energy use, probably due to a larger number of narrow roads surrounding residential buildings, while a smaller number of larger avenues and boulevards often surround larger commercial buildings. This relationship holds for the cooling and heating seasonal energy use regressions as well. Similarly, every additional unit of mean distance of surrounding roads is associated with a 2.454 percent increase in the mean of building energy use, and with a 3.776 percent increase in its standard deviation. Finally, every additional unit of standard deviation of distance of trees is associated with increases of 3.571 and 7.391 percent in mean and the standard deviation of building energy use respectively. These findings deviate from previous studies and could be explained by heat from traffic offsetting the heating loads during the heating season and reducing energy use, given that there are more heating degree days than cooling degree days in the case study area.

5. Discussion and implications

In this paper, we propose the CUE-A framework to empirically quantify the relationships between building energy use and spatial proximity to other urban systems, and then test its applicability using a case study. The significance of the case study results indicates that using empirical data in the proposed framework can help validate the estimated impacts of urban context on building energy use and extend findings from previous simulation-based studies.

First, let us consider the spatial proximity features extracted for different urban elements. While previous studies have often considered buildings as stand-alone elements (Pisello, Castaldo, Poli, & Cotana, 2014) or employed simplified spatial relationships between a building and one proximate urban element, we are able to generate empirical data on the proximity relationships that exist not just between a building and a *single* urban system, but between buildings and *multiple* surrounding urban systems. This allows us to conduct analysis on relationships between urban building energy use and proximity of multiple urban systems at the *building* scale in an urban area, building on previous studies on the impact of urban form on energy use at the *city* scale (Güneralp et al., 2017; Yamaguchi, Shimoda, & Mizuno, 2007).

Second, beyond just characterizing spatial proximity for multiple urban systems, the CUE-A framework incorporates that data into building energy analysis, presenting a step forward in considering the real-world urban context in urban energy efficiency research. The framework employs a novel application of the multivariate multiple regression model that allows us to simultaneously incorporate time-series building energy use data and “static” urban contextual data in a single model. Having multiple urban systems in a single model yields some insights that confirm previous studies and some that contrast existing work, indicating the value of investigating these phenomena in tandem as opposed to individually. For instance, the case study shows that having additional proximate buildings nearby is associated with lower energy use while controlling for all other variables including distance of surrounding buildings. At the same time, an increase in the central tendency and variation of distance of surrounding buildings is associated with decreases in the central tendency and variation in building energy use as well, while controlling for all other variables including the number of surrounding buildings. These results agree with previous work, where the clustering of buildings is associated with reduced energy use due to mutual shading and reflection (Pisello et al., 2012). On the other hand, the presence of additional proximate trees is associated with an increase in building energy use while having additional proximate roads is associated with decreased building energy use, controlling for all other variables. Both of these results are surprising and in contrast to previous literature. Further research is necessary to delve into how various planning and zoning regimes across different cities impact the relationships between urban systems and energy use to corroborate our empirical work.

Third, the CUE-A framework is scalable and extensible to different cities in future work. Running the algorithms to extract spatial proximity relationships for other cities, large or small, is straightforward. It is also possible to experiment with different thresholds for what is considered “spatial proximity” between urban elements. For instance, proximity in a sprawling metropolis like Los Angeles, U.S.A. might be defined very differently from proximity in a dense neighborhood in Jakarta, Indonesia. Our proposed CUE-A methodology could be applied by other cities to understand more city-specific relationships between urban context on building energy use, and to compare urban energy efficiency across cities. More importantly, while the current framework offers correlational insights, a future version of CUE-A could substitute time-series urban contextual data in place of the time-invariant data assumed in this paper, allowing us to model changes in the urban environment over time and providing a basis for causal deductions.

Finally, CUE-A offers a meaningful empirical complement to urban building energy models based on simulation, where significant progress has been made towards developing workflows (Reinhart & Cerezo Davila, 2016). The empirical approach that CUE-A proposes can, by no means, replace complex and vividly detailed simulation models that establish causal relationships. The CUE-A framework can, however, offer a useful “screening” model for more complex and time-consuming simulation models by identifying critical spatial proximity features associated with significant changes in building energy use in a specific urban context, so that they can be incorporated into a simulation model. It can also help identify proximity features that are not as important in a particular context, saving time and resources spent on accurately integrating those features into as simulation model. We thus envision CUE-A complementing the ever-growing body of simulation approaches by (1) offering empirical validation, (2) serving as “screening models” to develop more parsimonious simulation models, (3) serving as add-ons to physics-based models to decipher the level of detail needed for specific urban features, and (4) paving the path for hybrid models in the future.

6. Limitations and future work

This paper represents a first step in creating a scalable and extensible framework to empirically quantify the relationships between building energy use and spatial proximity of urban systems. However, our methodology has several limitations, both in the framework and the case study.

First, the framework has limitations that offer many opportunities for future work. To specify spatial proximity, we employ simplified and abstracted two-dimensional features, while more informative features such as three-dimensional proximity properties could be explored to develop more comprehensive and generalizable relationships. Additionally, we assume that the relationships between spatial proximity of urban systems and building energy use are linear, and therefore multivariate multiple regression is appropriate. More advanced algorithms (e.g., conditional random fields) could be employed to examine whether non-linear relationships exist. Similarly, while our goal was primarily descriptive, i.e. understanding the relationships between energy use and spatial proximity, there is room for various machine learning algorithms to be employed in future work aimed at prediction. However, we point out that utilizing such advanced methods introduces challenges to interpretability and may limit the applicability of results.

Second, our case study has some limitations. In this paper, we limit the case study to the downtown area of a mid-size city in the temperate climate of Northern California. While this serves as a useful small-scale demonstration of the framework, there is a need for more case study cities of varying sizes and planning paradigms to corroborate the relationships uncovered in this work and to point out how relationships differ across cities. Further, while the case study results offer key *correlational* insights between spatial proximity and energy use, there is an opportunity for further in-depth *causal* research on the mechanisms and driving factors underlying these phenomena through an expansion of the CUE-A framework for randomized control trials and natural experiments. For instance, a future study could utilize a natural experiment on road closures to understand changes in building energy use using this framework. This could shine a light on whether the traffic intensity is a driving factor behind the observed association between road proximity and building energy use.

Finally, there is a need to understand the application of this analysis in urban planning and decision-support. Future work aims to investigate the translation of analytical findings into actionable insights on sustainable urban design, management, and operations. For example, concrete reasoning rules (e.g., what-if-else) can be developed to formulate the relationships between building spatial proximity and energy use for data-informed decision-making. The results also point to the need to explore how existing planning paradigms can affect these relationships. This framework can also complement simulation studies by identifying key relationships to be explored further through high-fidelity and detailed simulation models.

7. Conclusion

With the profusion of new data on urban systems and the ability to extract insights using novel analytical techniques, there is now tremendous potential to bring empirical evidence to test prevailing heuristics- or simulation-based approaches for energy analysis in urban design, planning and systems-level analysis. The objective of this paper was to propose an analytical framework (CUE-A) to operationalize and empirically quantify relationships between spatial proximity of urban systems and building energy use, enabling the inclusion of multiple urban elements in a single analysis. Further, by conducting a case study that utilizes this framework, we aimed to demonstrate the merits of using empirical data analysis to study these relationships.

We find that the spatial proximity of other buildings, roads and trees is associated with varied and significant changes in both the central tendency and variability of building energy use. Our findings substantiate some insights from previous simulation-based studies, while offering some novel results that begin to account for varying planning paradigms and urban design factors that determine the urban form of a city. For instance, having additional proximate buildings nearby is associated with lower energy use, in agreement with previous studies on mutual shading and reflection. But, having additional proximate trees and roads is associated with increased and decreased building energy use respectively, controlling for all other variables, in contrast to previous literature, showing that more work is needed to quantify and parse the effect of the urban design and land-use planning of the city. Our results are consistent for heating and cooling seasonal energy use in addition to total building energy use, which offers confidence on the robustness of the studied relationships.

In conclusion, through the application to this scalable and extensible framework to other contexts, we aim to further catalyze data-driven research at the intersection of urban design, land-use planning and building energy use. Ultimately, insights from such data-driven methods will inform how we (re)design our energy-intensive cities to transition them to a more sustainable future.

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.

Acknowledgments

The material presented in this manuscript is based in part upon work supported by the Stanford Terman Faculty Fellowship, the Stanford Interdisciplinary Graduate Fellowship, the Emmett Interdisciplinary Program in Environment and Resources, and the U.S. National Science Foundation (NSF) under Grant Nos. 1642315, 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.

Appendix A

The relationship learning algorithm utilizes an iterative process to determine the degree to which spatial proximity exists between any two urban elements. For example, the relationship between two buildings is learned by calculating the projection of a building on a side of a target building as presented in the following vector equation:

$$r = a + t(b - a) = a + tv \tag{A.1}$$

Where $v = b - a$. a and b are the vertices of the side (given as vectors) and t is a scaling parameter. The equation of a line segment which is perpendicular to the side ab and passes through a vertex (e.g., p_1) of building B is:

$$r_1 = p_1 + t_2 v' \tag{A.2}$$

Where t_2 is a scaling parameter and v' is a vector perpendicular to v. The point of intersection of r and r_1 is found, which gives a unique value of t. If $0 < t < 1$, the point of intersection of r and r_1 lies on the side ab. Therefore, the projection of vertex p_1 on the side ab is given by $i_1 = t$. Similarly, for other vertices $p_2, p_3, \dots p_n$ of the building A, the projections $i_2, i_3, \dots i_n$, can be calculated. Let $max_p = max(i_1, i_2, i_3, \dots i_n)$ and $min_p = min(i_1, i_2, i_3, \dots i_n)$; The projection of the building B on the side ab of building A element is defined as intersection of interval $[min_p, max_p]$ with interval $[0, 1]$. The side of building A on which building B has the largest projection is identified as the side spatially proximate to building B. If the building B has the same projection on multiple sides of building A, then we pick the side closer to building B. A spatial proximity relationship is then defined between building

B and the proximate side of building A if the distance between them is less than a predefined threshold as specified by the user. Similarly, vegetation and roads that are within the threshold distance of the building A can be identified along with the side of building B upon which they are located (refer to (Gupta et al., 2019) for a detailed explanation of the methodology).

Appendix B

Table B1
Results of the Multivariate Multiple Regression analysis.

Variable	Total Energy Use		Cooling Energy Use		Heating Energy Use	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Spatial proximity feature						
Surrounding Buildings						
Number	-7.157	-7.071	-2.669	-2.072	-3.587	-4.882
Mean distance	0.000	7.595	0.926	6.908	3.126	0.313
Standard deviation of distance	-1.594	-2.406	0.000	-7.239	-0.993	0.000
Surrounding Trees						
Number	2.562	0.000	8.411	12.330	6.742	0.000
Mean distance	-4.547	-4.449	-4.981	-7.535	-7.793	-0.736
Standard deviation of distance	6.499	2.456	4.367	1.001	3.025	2.298
Surrounding Roads						
Number	-3.732	-7.560	-6.384	-12.014	-3.720	0.000
Mean distance	2.454	3.776	9.516	7.259	5.872	0.000
Standard deviation of distance	3.579	7.391	2.112	11.576	2.092	2.711
Control variables						
Building type						
Commercial	401.473	325.867	530.001	468.718	375.778	234.244
Multi-family residential	0.000	0.000	0.000	0.000	0.000	-1.794
Single-family residential	0.000	0.000	0.000	0.000	-6.064	0.000
Building geometry & orientation						
Height of building	30.511	24.040	63.981	29.077	30.736	23.063
Floor area of building	71.281	79.908	66.644	66.126	87.450	70.861
Building orientation (N)	0.000	0.000	0.000	0.000	0.000	0.000
Building orientation (NW 45)	-11.560	-2.012	-17.275	-5.287	-10.946	0.000
Building orientation (NW 90)	0.000	0.000	0.000	0.000	0.000	0.000
Building orientation (NW 135)	0.000	0.000	0.000	0.000	0.000	0.000
Building condition						
Age of building	-19.367	-5.077	-17.207	-17.303	-13.350	-5.918
R^2	**0.679**	**0.626**	**0.627**	**0.638**	**0.728**	**0.628**

Note: The regression was implemented using a logarithmic transformation of dependent variables, which are then reconverted here.

References

Abanda, F. H., & Byers, L. (2016). An investigation of the impact of building orientation on energy consumption in a domestic building using emerging BIM (Building Information Modelling). *Energy, 97*, 517–527. https://doi.org/10.1016/j.energy.2015.12.135

Akbari, H., Kurn, D. M., Bretz, S. E., & Hanford, J. W. (1997). Peak power and cooling energy savings of shade trees. *Energy and Buildings, 25*, 139–148. https://doi.org/10.1016/S0378-7788(96)01003-1

Aksoezen, M., Daniel, M., Hassler, U., & Kohler, N. (2015). Building age as an indicator for energy consumption. *Energy and Buildings, 87*, 74–86. https://doi.org/10.1016/j.enbuild.2014.10.074

Allegrini, J., Dorer, V., & Carmeliet, J. (2012). Influence of the urban microclimate in street canyons on the energy demand for space cooling and heating of buildings. *Energy and Buildings, 55*, 823–832. https://doi.org/10.1016/j.enbuild.2012.10.013

Anton, I., & Tănase, D. (2016). Informed geometries. Parametric modelling and energy analysis in early stages of design. *Energy Procedia, 85*, 9–16. https://doi.org/10.1016/j.egypro.2015.12.269

Arnfield, A. J. (2003). Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island. *International Journal of Climatology, 23*, 1–26. https://doi.org/10.1002/joc.859

ASHRAE. (2002). ASHRAE Guideline 14-2002. Measurement of Energy and Demand Savings. *Vent Air Cond Eng Atlanta, Georgia*. In this issue.

Barry, R. G., & Blanken, P. D. (2016). *Introduction. Microclimate and local climate* (pp. 1–8). Cambridge University Press. https://doi.org/10.1017/CBO9781316535981.002

Bouyer, J., Inard, C., & Musy, M. (2011). Microclimatic coupling as a solution to improve building energy simulation in an urban context. *Energy and Buildings, 43*, 1549–1559. https://doi.org/10.1016/j.enbuild.2011.02.010

Chen, Y., Hong, T., & Piette, M. A. (2017). Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis. *Applied Energy, 205*, 323–335.

Djedjig, R., Bozonnet, E., & Belarbi, R. (2016). Modeling green wall interactions with street canyons for building energy simulation in urban context. *Urban Climate, 16*, 75–85. https://doi.org/10.1016/j.uclim.2015.12.003

Güneralp, B., Zhou, Y., Ürge-Vorsatz, D., Gupta, M., Yu, S., Patel, P. L., et al. (2017). Global scenarios of urban density and its impacts on building energy use through 2050. *PNAS, 114*, 8945–8950. https://doi.org/10.1073/pnas.1606035114

Han, Y., Taylor, J. E., & Pisello, A. L. (2017). Exploring mutual shading and mutual reflection inter-building effects on building energy performance. *Applied Energy, 185*, 1556–1564. https://doi.org/10.1016/j.apenergy.2015.10.170

Heidarinejad, M., Gracik, S., Sadeghipour Roudsari, M., Khoshdel Nikkho, S., Liu, J., Liu, K., et al. (2016). Influence of building surface solar irradiance on environmental temperatures in urban neighborhoods. *Sustainable Cities and Society, 26*, 186–202. https://doi.org/10.1016/j.scs.2016.06.011

Howard, L. (1833). *The climate of London*. First.

Huang, X., Zhang, L., & Li, P. (2007). Classification and extraction of spatial features in urban areas using high-resolution multispectral imagery. *IEEE Geoscience and Remote Sensing Letters, 4*, 260–264. https://doi.org/10.1109/LGRS.2006.890540

Ihara, T., Kikegawa, Y., Asahi, K., Genchi, Y., & Kondo, H. (2008). Changes in year-round air temperature and annual energy consumption in office building areas by urban heat-island countermeasures and energy-saving measures. *Applied Energy, 85*, 12–25. https://doi.org/10.1016/j.apenergy.2007.06.012

IPMVP Committee and others. (2002). *International Performance Measurement and Verification Protocol: Concepts and options for determining energy and water savings*. IPMVP. In this issue.

Karan, E., Mohammadpour, A., & Asadi, S. (2016). Integrating building and transportation energy use to design a comprehensive greenhouse gas mitigation strategy. *Applied Energy, 165*, 234–243. https://doi.org/10.1016/j.apenergy.2015.11.035

Gupta, K., Yang, Z., & Jain, R. K. (2019). Urban data integration using proximity relationship learning for design, management, and operations of sustainable urban systems. *Journal of Computing in Civil Engineering, 33*, Article 04018063. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000806

Kim, H., Gu, D., & Kim, H. Y. (2018). Effects of Urban Heat Island mitigation in various climate zones in the United States. *Sustainable Cities and Society, 41*, 841–852. https://doi.org/10.1016/j.scs.2018.06.021

Kontokosta, Constantine (2012). *Local Law 84 Energy Benchmarking Data: Report to the New York City Mayor's Office of Long-Term Planning and Sustainability*. New York: New York City Mayor's Office of Long-Term Planning and Sustainability.

Lauzet, N., Rodler, A., Musy, M., Azam, M.-H., Guernouti, S., Mauree, D., et al. (2019). How building energy models take the local climate into account in an urban context – A review. *Renewable and Sustainable Energy Reviews, 116*, Article 109390. https://doi.org/10.1016/j.rser.2019.109390

Lee, S.-H., & Kim, S.-T. (2015). Estimation of anthropogenic heat emission in the Gyeong-In region of Korea. *Asia-Pacific Journal of Atmospheric Sciences, 96*, 291–303. https://doi.org/10.1007/s00704-008-0040-6

Li, D. H. W., & Wong, S. L. (2007). Daylighting and energy implications due to shading effects from nearby buildings. *Applied Energy, 84*, 1199–1209. https://doi.org/10.1016/j.apenergy.2007.04.005

Liu, J., Heidarinejad, M., Gracik, S., & Srebric, J. (2015). The impact of exterior surface convective heat transfer coefficients on the building energy consumption in urban neighborhoods with different plan area densities. *Energy and Buildings, 86*, 449–463. https://doi.org/10.1016/j.enbuild.2014.10.062

Mohammadi, N., & Taylor, J. E. (2017). Urban energy flux: Spatiotemporal fluctuations of building energy consumption and human mobility-driven prediction. *Applied Energy, 195*, 810–818.

Morakinyo, T. E., Dahanayake, K. W. D., Kalani, C., Adegun, O. B., & Balogun, A. A. (2016). Modelling the effect of tree-shading on summer indoor and outdoor thermal condition of two similar buildings in a Nigerian university. *Energy and Buildings, 130*, 721–732. https://doi.org/10.1016/j.enbuild.2016.08.087

New York City Mayor's Office of Sustainability. (2016). *New York City's Energy and Water Use 2013 Report*. New York City Mayor's Office of Sustainability.

Oke, T. R. (1982). The energetic basis of the urban heat island. *Quarterly Journal of the Royal Meteorological Society, 108*(55), 1–24.

Open Data · City of Palo Alto (2020) https://data.cityofpaloalto.org/home (accessed September 18, 2020).

Pacheco, R., Ordóñez, J., & Martínez, G. (2012). Energy efficient design of building: A review. *Renewable and Sustainable Energy Reviews, 16*, 3559–3573. https://doi.org/10.1016/j.rser.2012.03.045

Pearlmutter, D., Berliner, P., & Shaviv, E. (2007). Integrated modeling of pedestrian energy exchange and thermal comfort in urban street canyons. *Building and Environment, 42*, 2396–2409. https://doi.org/10.1016/j.buildenv.2006.06.006

Perini, K., & Magliocco, A. (2014). Effects of vegetation, urban density, building height, and atmospheric conditions on local temperatures and thermal comfort. *Urban Forestry & Urban Greening, 13*, 495–506. https://doi.org/10.1016/j.ufug.2014.03.003

Pisello, A. L., Taylor, J. E., Xu, X., & Cotana, F. (2012). Inter-building effect: Simulating the impact of a network of buildings on the accuracy of building energy performance predictions. *Building and Environment, 58*, 37–45. https://doi.org/10.1016/j.buildenv.2012.06.017

Pisello, A. L., Castaldo, V. L., Poli, T., & Cotana, F. (2014). Simulating the thermal-energy performance of buildings at the urban scale: Evaluation of inter-building effects in different urban configurations. *Journal of Urban Technology, 21*, 3–20. https://doi.org/10.1080/10630732.2014.884386

Pisello, A. L., Castaldo, V. L., Taylor, J. E., & Cotana, F. (2014). Expanding Inter-Building Effect modeling to examine primary energy for lighting. *Energy and Buildings, 76*, 513–523. https://doi.org/10.1016/j.enbuild.2014.02.081

Porse, E., Derenski, J., Gustafson, H., Elizabeth, Z., & Pincetl, S. (2016). Structural, geographic, and social factors in urban building energy use: Analysis of aggregated account-level consumption data in a megacity. *Energy Policy, 96*, 179–192. https://doi.org/10.1016/j.enpol.2016.06.002

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

Samuelson, H. W., Claussnitzer, S., Goyal, A., Chen, Y., & Romo-Castillo, A. (2016). *Parametric energy simulation in early design: High-rise residential buildings in urban contexts*. https://doi.org/10.1016/j.buildenv.2016.02.018

M. Santamouris. (2001). Energy and Climate in the Urban Built Environment. On the built environment - the urban influence, n.d., p. 3–18.

Santamouris, M., Papanikolaou, N., Livada, I., Koronakis, I., Georgakis, C., Argiriou, A., et al. (2001). On the impact of urban climate on the energy consumption of buildings. *Solar Energy, 70*, 201–216. https://doi.org/10.1016/S0038-092X(00)00095-5

Scofield, John H. (2013). Efficacy of LEED-certification in reducing energy consumption and greenhouse gas emission for large New York City office buildings. *Energy and Buildings, 67*, 517-24 https://doi.org/10.1016/j.enbuild.2013.08.032. In this issue.

Tong, Z., Chen, Y., Malkawi, A., Adamkiewicz, G., & Spengler, J. D. (2016). Quantifying the impact of traffic-related air pollution on the indoor air quality of a naturally ventilated building. *Environment International, 89–90*, 138–146. https://doi.org/10.1016/j.envint.2016.01.016

Toparlar, Y., Blocken, B., Vos, P., van Heijst, G. J. F., Janssen, W. D., van Hooff, T., et al. (2015). CFD simulation and validation of urban microclimate: A case study for Bergpolder Zuid, Rotterdam. *Building and Environment, 83*, 79–90. https://doi.org/10.1016/j.buildenv.2014.08.004

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. https://doi.org/10.1016/j.rser.2017.05.248

Tuhus-Dubrow, D., & Krarti, M. (2010). Genetic-algorithm based approach to optimize building envelope design for residential buildings. *Building and Environment, 45*, 1574–1581. https://doi.org/10.1016/j.buildenv.2010.01.005

United Nations. (2015). *A better urban future*. United Nations.

US DOE. (2008). *M&V guidelines: Measurement and verification for federal energy projects. Version 3.0*. US DOE.

US EPA. (2017). *EPA's ENERGY STAR Resources: Portfolio Manager and More*. US EPA.

Wang, Z.-H. (2014). Monte Carlo simulations of radiative heat exchange in a street canyon with trees. *Solar Energy, 110*, 704–713. https://doi.org/10.1016/j.solener.2014.10.012

Yamaguchi, Y., Shimoda, Y., & Mizuno, M. (2007). Proposal of a modeling approach considering urban form for evaluation of city level energy management. *Energy and Buildings, 39*, 580–592. https://doi.org/10.1016/j.enbuild.2006.09.011

Yi, Y. K., & Malkawi, A. M. (2009). Optimizing building form for energy performance based on hierarchical geometry relation. *Automation in Construction, 18*, 825–833. https://doi.org/10.1016/j.autcon.2009.03.006