Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 158 (2019) 6478–6483

Energy Procedia

www.elsevier.com/locate/procedia

10th International Conference on Applied Energy (ICAE2018), 22-25 August 2018, Hong Kong, China

DUE-A: Data-driven Urban Energy Analytics for understanding relationships between building energy use and urban systems

Zheng Yang[a], Karan Gupta[a], Rishee K. Jain[a]*

[a] *Urban Informatics Lab, Stanford University, 473 Via Ortega, Room 269B, Stanford, CA 94305, United States*

Abstract

Cities account for over 75% of all primary energy usage in the world with buildings making up the bulk of this usage. It is well acknowledged that the building energy usage is greatly impacted by urban context and thus understanding the relationships between building energy use and surrounding urban systems is critical for more energy efficient and holistic planning. This paper proposes a Data-driven Urban Energy Analytics (DUE-A) workflow to investigate and quantify the relationships between building energy usage and the spatial proximity of other urban systems. A case study of 530 buildings in a mid-size city in the Unites States is conducted to validate the performance of the workflow and demonstrate the statistical significance of relationships between building energy use and spatial proximity of other systems. Results show that spatial proximity of other buildings, roads and trees can have both positive and negative impacts on the mean, variability and distribution of building energy usage, and indicate that more holistic planning and design of cities could unlock urban energy efficiency and low-carbon municipal pathways.

Peer-review under responsibility of the scientific committee of ICAE2018 – The 10th International Conference on Applied Energy.

Keywords: Data analytics; Building energy; Energy efficiency; Spatial proximity; Urban systems; Relationship learning

1. Introduction

Cities account for more than 75% global energy use with the majority of this consumption coming from buildings [1]. As a result, enhancing the energy efficiency of the urban built environment is critical for meeting the world's climate change and sustainability goals. Due to the development of new information technologies, massive amounts

* Corresponding author. Tel.: 1-650-721-8559; fax: 1-650-723-4806.
E-mail address: rishee.jain@stanford.edu

1876-6102
Peer-review under responsibility of the scientific committee of ICAE2018 – The 10th International Conference on Applied Energy.
10.1016/j.egypro.2019.01.114

of data are now being collected on an array of urban systems and thus opening up opportunities to analyze and optimize urban energy use that was previously not possible. Urban building energy efficiency is impacted by surrounding urban systems, including adjacent buildings (e.g, due to occupant dynamics), roads (e.g., due to vehicles), and trees (e.g., due to shading and winds) [2-4]. While previous work has acknowledged the impact surrounding systems can have on urban building energy use, a systematic study and quantification of the impact spatial proximity of surrounding urban systems has on energy usage is lacking. This paper proposes a Data-driven Urban Energy Analytics (DUE-A) workflow to investigate and quantify the relationships between building energy usage and the spatial proximity of other urban systems. The DUE-A workflow contains two steps: (1) extract features on spatial proximity for each building through a series of physical relationship learning algorithms; (2) model the relationships between building spatial proximity features and energy use features through multivariate multiple regression.

2. Literature Review

Urban building energy use is significantly impacted by urban context [5]. One important factor of urban context is the spatial proximity of a building to other urban systems like other buildings, vegetation, and roads. For example, a building's outside temperature could be abnormally high due to urban heat island effects caused by close proximity to adjacent buildings [6] or due to shadows and shading [2]. Previous work [7] has also demonstrated an "inter-building effect" (IBE) for lighting related energy usage. Air flow is another key element that drives building heat transfer (e.g., convection), humidity, cooling, and ventilation loads [3] and can be influenced by nearby buildings and trees [4].

Previous research has analyzed how spatial proximity of a building in respect to other urban systems impacts building energy usage at the micro and macro level. At the micro level, Howard et al. analyzed the spatial distributions of building energy consumption by end use and found use patterns across zip codes in New York city [8]. Keirstead and Calderon developed an optimization based method to capture the spatial variations in energy use at both the whole city and the district scale [9]. Zhu et al. studied spatial-temporal patterns of fuel and electricity consumptions in the residential sector of China [10]. However, these studies all focused on spatial impacts on energy use of large scale building stock, no insights for individual buildings can be extracted and it is still unclear how a building's spatial proximity to other systems impact its energy use. At the macro level, Han et al. designed an urban building network model to investigate complex mutual impacts within spatially proximal buildings [11]. Bouyer et al. analyzed differences of building energy use by the presence of trees in the surrounding configuration [12]. Perini and Magliocco investigated the effects of surrounding vegetation on building energy use and found effects are more significant with higher temperatures and lower relative humidity [13]. While these studies provided insights into how other urban systems specifically impact individual building energy usage, they are limited in their ability to holistically analyze and quantify the impact spatial proximity to multiple urban systems has on building energy use. As a result, we propose a Data-driven Urban Energy Analytics (DUE-A) workflow capable of systematically modeling and quantifying the relationships between building energy usage and the spatial proximity of other urban systems.

3. Methodology

3.1. Feature extraction of building spatial proximity properties

First, the DUE-A workflows extracts features of building spatial proximity properties through a series of physical relationship learning algorithms elucidated in [14]. Based on the geometric properties, urban systems are divided into three categories: polygonal (e.g., buildings), linear (e.g., roads) and point (e.g., trees). Based on this categorization, the relationship learning algorithm utilize an iterative process to determine whether a spatial proximity exists between two urban elements. For example, the relationship between two buildings is learned by calculating the projection of a building (building B) on a side of a target building (building A) as presented in the following vector equation:

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

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:

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

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, … p_n of the building A, the projections i_2, i_3, … i_n , could be calculated. Let max_p=max(i_1,i_2, i_3, … i_n) and min_p=min(i_1,i_2, i_3, … 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 the building A, then we pick the side which is 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 the threshold. Similarly, trees and roads which are within the threshold distance of the building A can be identified along with the side of building B upon which they are located. After the algorithm iterates through all the buildings, the results are aggregated to construct a set of spatial proximity features for each building in respect to other buildings, trees and roads (see Table 1).

Table 1. Spatial proximity features extracted for each building

Spatial proximity features for buildings	Spatial proximity features for trees	Spatial proximity features for roads
Number of proximate buildings (B_No)	Number of proximate trees (T_No)	Number of proximate roads (R_No)
Maximum distance of proximate buildings (B_Max)	Maximum distance of proximate trees (T_Max)	Maximum distance of proximate roads (R_Max)
Minimum distance of proximate buildings (B_Min)	Minimum distance of proximate trees (T_Min)	Minimum distance of proximate roads (R_Min)
Mean distance of proximate buildings (B_Mean)	Mean distance of proximate trees (T_Mean)	Mean distance of proximate roads (R_Mean)
Standard deviation of distance of proximate buildings (B_SD)	Standard deviation of distance of proximate tress (T_SD)	Standard deviation of distance of proximate roads (R_SD)
Median distance of proximate buildings (B_Median)	Median distance of proximate trees (T_Median)	Median distance of proximate roads (R_Median)
Interquartile range of distance of proximate buildings (B_IQR)	Interquartile range of distance of proximate trees (T_IQR)	Interquartile range of distance of proximate roads (R_IQR)
Skewness of distance of proximate buildings (B_Skew)	Skewness of distance of proximate trees (T_Skew)	Skewness of distance of proximate roads (R_Skew)
Kurtosis of distance of proximate buildings (B_Kur)	Kurtosis of distance of proximate trees (T_Kur)	Kurtosis of distance of proximate roads (R_Kur)

3.2. Multivariate Multiple Regression

Second, the DUE-A workflow models the relationships between building spatial proximity features (Section 3.1) and energy use through a multivariate multiple regression. Features are then extracted to represent the characteristics of building energy use, including median energy use (E_Median), mean energy use (E_Mean), standard deviation of energy use (E_Sd), interquartile range of energy use (E_IQR), mean absolute deviation of energy use (E_Meanad), median absolute deviation from median of energy use (E_Mad), skewness of energy use (E_Skew), and kurtosis of energy use (E_Kur). As a result, the training data$\{(\mathbf{x_1},\mathbf{y_1}),(\mathbf{x_2},\mathbf{y_2}),\dots,(\mathbf{x_n},\mathbf{y_n})\}$, where $\mathbf{y}=(\mathrm{y}_1,\mathrm{y}_2,\dots,y_m)$ is the vector of eight energy use features for a building (e.g., mean of monthly energy use), and $\mathbf{x}=(\mathrm{x}_1,\mathrm{x}_2,\dots,\mathrm{x}_p)$ is the vector of spatial proximity features for the same building calculated in Section 3.1. The objective is to model statistical mapping from building spatial proximity features to energy use features:

$$\mathbf{x}=(\mathrm{x}_1,\mathrm{x}_2,\dots,\mathrm{x}_p)\xrightarrow{\boldsymbol{h(x)}}\mathbf{y}=(\mathrm{y}_1,\mathrm{y}_2,\dots,\mathrm{y}_m) \tag{3}$$

In order to maintain the interpretability of the relationships, it is assumed $\boldsymbol{h(x)}$ is linear $\boldsymbol{h(x)}=\boldsymbol{A}^T\boldsymbol{x}$. Since there might exist dependencies among the features of building energy use, for example the kurtosis of one building's monthly energy use could be highly associated with its monthly mean and standard deviation of energy use:

$$\mathrm{P}(\mathbf{y}|\mathbf{x})\neq\prod_1^m\mathrm{P}(\mathrm{y_i}|\mathbf{x}) \tag{4}$$

The errors of $\boldsymbol{h(x)}$ associated with $\mathbf{y}$ of same building may have different variances and may be correlated. The multivariate least-square loss is defined as:

$$L(h,P)=\int\sum_{j=1}^{m}\left(y_j-h_j(\boldsymbol{x})\right)^2dP(\boldsymbol{x},\boldsymbol{y}) \tag{5}$$

The process for multivariate least squares is implemented with Frobenius norm, L_1 norm, and L_2 norm being added as regularization to the loss function to control overfitting, the overall sparsity of the model, the total number of impactful building spatial proximity features, respectively.

The estimated coefficient matrix represents the impacts of different building spatial proximity features on energy use features. Hyperparameters in the regularization terms are determined by 10-fold cross validation. Construction of the spatial proximity features and multivariate multiple regression analysis are conducted in Python.

4. Case study and Results

4.1. Data Collection

To validate the performance of the DUE-A workflow described above, a real-world case study is conducted for the downtown area of a mid-size city in California, USA. This area contains 530 buildings of varying sizes and mixed functions, including offices, residential, stores, restaurants, etc. (Fig. 1a for a map of the test area). Spatial proximity properties for each building in respect to other buildings, trees and roads in the test area are computed using the method described in section 3.1 We note that as part of the data cleansing process, we discarded any building with more than 5 features missing and imputed other missing values based on the mean value across the test area sample. Two-year's worth of monthly energy use data for each building (normalized by building area) was collected. If more than 3 months of data for a building is missing, that building was removed. Other missing values were imputed using energy use for the same month in a different year. Features representing energy use characteristics were calculated (Section 3.2) and then standardized to remove the mean and scale to unit variance. It can be seen from Fig. 1b that building energy use features are correlated to each other and corroborates assumptions made in the proposed methodology.

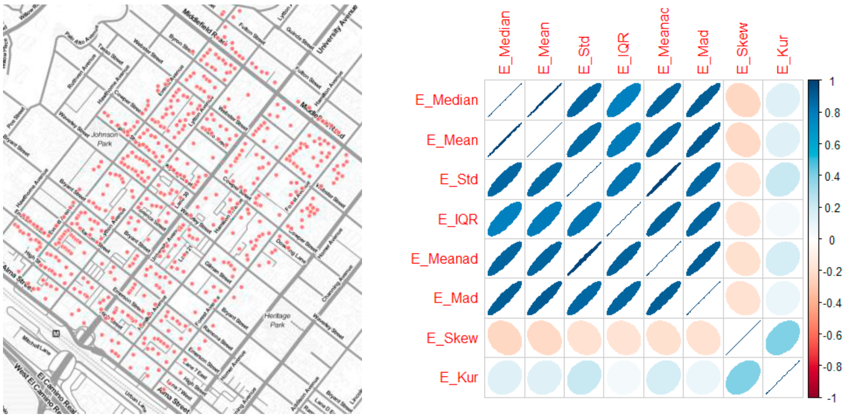

Fig. 1a. Test area in a mid-size city of California, United States with buildings marked (Map credit: OpenStreetMap). Fig. 1b. Correlation between building energy use features.

4.2. Results and Discussion

The multivariate multiple regression analysis described in Section 3.2 is implemented to understand the relationships between building spatial proximity and energy use. The absolute values of coefficients of spatial property features are between 0.1 and 2.5 (Fig. 2). It can be seen that the spatial connections/proximity to other buildings, roads

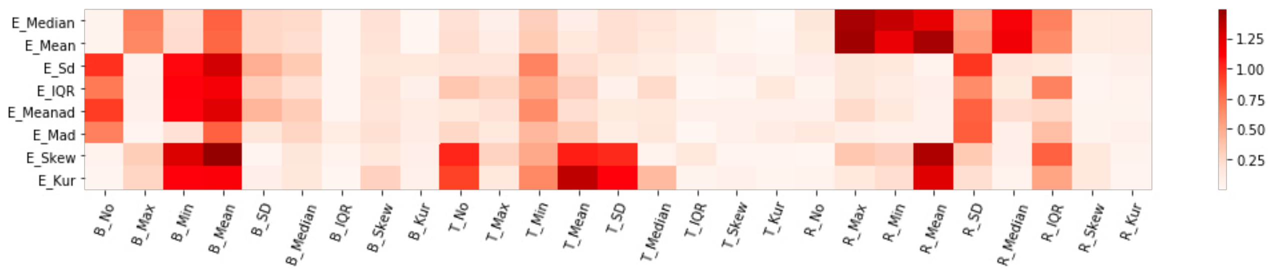

Fig. 2. Coefficients of building spatial properties regarding energy use features.

and trees have different patterns of impacts on building energy use. The relationships between building spatial proximity features and different energy use features are all statistically significant with all p-values less than 0.05 and have an adjusted R-square ranging between 0.4 and 0.6.

For more detailed analysis, first the relationships between building spatial proximity features and the centering of energy use (e.g. mean - *E_Mean*) are investigated (blue bars in Fig. 3). It can be seen that the distance between a building and roads has the largest negative impacts on the mean energy use of the building (*R_Max, R_Min, R_Mean, R_Median*). In other words, the closer the buildings are to roads, the more energy they consume. We postulate that this could be due to the extra heat produced from vehicle engine combustion or street noise that coerces building occupants to close windows and reduce natural ventilation both of which impact energy loads for heating/cooling.

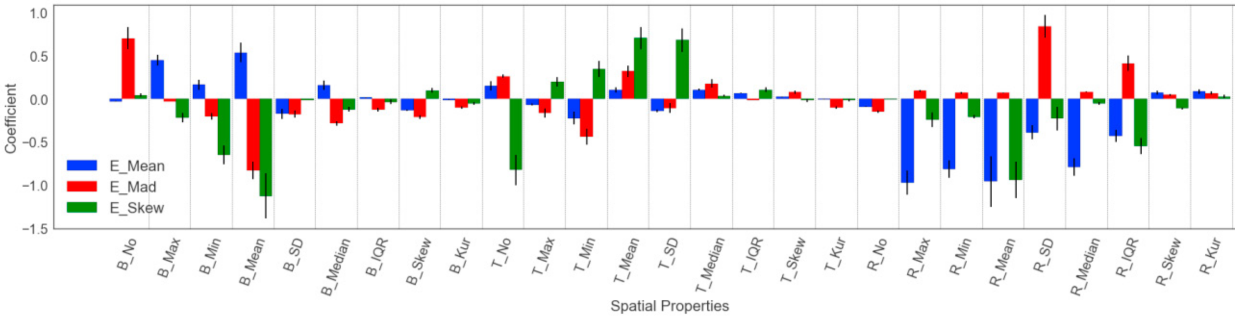

Fig. 3. Coefficients of building spatial proximity features in respect to: central tendency of energy use (*E_Mean*, blue bar); variability of energy use (*E_Mad*, red bar); distribution of energy use (*E_Skew*, green bar)

The relationships between building spatial proximity features and the variability of energy use (e.g. median absolute deviation from mean- *E_Mad*) are indicated by the red bars in Fig. 3. Results indicate that the variability of energy use is most impacted (positively) by the standard deviation of distance to proximate roads (*R_SD*). This could be attributed to the fact that roads close to each other might have similar traffic patterns and thus have more consistent impacts onto microclimate and external loads of buildings. Analysis also revealed that the distance between a building and adjacent buildings (*B_Mean*) decreases the variability of building energy use, but a higher number of adjacent buildings (*B_No*) increases the variability of building energy use. This observation could be due to seasonal urban heat island effects and/or underlying joint community human dynamics previously shown to impact energy use [15].

Lastly, the relationships between building spatial proximity features and the distribution of energy use (e.g. skewness - *E_Skew*) are analyzed (green bars in Fig. 3). It can be seen that the distance to roads (*R_Mean*) and other buildings (*B_Mean*) have the largest negative impacts on the skewness of a building's energy use. This indicates that being close to other buildings and roads could result in higher energy usage as the impacts of urban context can be more significant. On the contrary, if a building is closer to more trees (*T_No*), its energy use pattern is less skewed, demonstrating vegetation could be a good buffer for the variations of microclimate and resultant heating/cooling loads.

5. Conclusions and Future Work

It has been widely acknowledged that urban building energy efficiency is largely impacted by urban context. Extensive work has been conducted to analyze how the spatial connections and proximity to other urban systems impact a building's energy use. However, previous studies were limited in their ability to systematically and explicitly quantify relationships between building energy use and proximity to other urban systems (i.e., roads, trees). This paper proposes a Data-driven Urban Energy Analytics (DUE-A) workflow to investigate and quantify such relationships. First, spatial proximity features of each building are extracted using a series of physical relationship learning algorithms. Second, multivariate multiple regression is utilized to model how building spatial proximity features are associated with variations in building energy use. We applied the DUE-A model to a case study of 530 buildings and found the relationships between building spatial proximity features and energy use features to be statistically significant. Results also revealed that spatial proximity of other buildings, roads and trees can have both positive and negative impact on the mean, variability and distribution of building energy usage in a city.

This paper represents a first-step in understanding and quantifying the relationship between building energy use and proximity to urban systems and several limitations exist. First, only a few features are extracted to represent both building spatial proximity and building energy use, and our case study sample was limited to mid-size city in the United States. Future work should explore more informative features to create generalizable relationships between building spatial proximity properties and energy use characteristics and extend the analysis to cities of varying size around the world. Second, it is assumed the impacts of building spatial proximity to other urban systems on energy use are linear and therefore multivariate multiple regression is utilized. Future work aims to employ more advanced algorithms (e.g., conditional random fields) to examine whether non-linear relationships exist. Third, the application for such analysis in urban systems planning and decision-support are currently limited. Future work could mitigate this issue by developing concrete reasoning (e.g., what-if) rules to formulate the relationships between building spatial proximity and energy use. A deeper understanding of how building energy use and other urban systems relate could provide insights into more holistic urban systems planning and provide viable pathways to energy efficient and low-carbon cities.

Acknowledgements

The material presented in this manuscript is based in part upon work supported by the U.S. National Science Foundation (NSF) under Grants No. 1642315. 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

[1] United Nations. For a better urban future. (2016). Available at: http://unhabitat.org/urban-themes/energy/.
[2] Chen, Y., Hong, T., Piette, M.A. Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis, Applied Energy. 205 (2017) 323-335.
[3] Liu, J., Heidarinejad, M., Gracik, S., Srebric, J. 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 (2015) 449-463.
[4] Toparlar, Y., Blocken, B, Vos, P., Van Heijst, G., Janssen, W., van Hooff, T., et al. CFD simulation and validation of urban microclimate: A case study for Bergpolder Zuid, Rotterdam, Building and Environment. 83 (2015) 79-90.
[5] Nutkiewicz, A. Yang, Z., Jain, R.K. Data-driven Urban Energy Simulation (DUE-S): A framework for integrating engineering simulation and machine learning methods in a multi-scale urban energy modeling workflow, Applied Energy. (2017) In press.
[6] Samuelson, H., Claussnitzer, S., Goyal, A., Chen, Y., Romo-Castillo, A. Parametric energy simulation in early design: High-rise residential buildings in urban contexts, Building and Environment. 101 (2016) 19-31.
[7] Pisello, A.L., Castaldo, V.L, Taylor, J.E., Cotana, F. Expanding Inter-Building Effect modeling to examine primary energy for lighting, Energy and Buildings. 76 (2014) 513-523.
[8] Howard, B., Parshall, L., Thompson, J., Hammer, S., Dickinson, J., Modi, V. Spatial distribution of urban building energy consumption by end use, Energy and Buildings. 45 (2012) 141-151.
[9] Keirstead, J., Calderon, C. Capturing spatial effects, technology interactions, and uncertainty in urban energy and carbon models: Retrofitting newcastle as a case-study, Energy Policy. 46 (2012) 253-267.
[10] Zhu, D., Tao, S., Wang, R., Shen, H., Huang, Y., Shen, G. et al. Temporal and spatial trends of residential energy consumption and air pollutant emissions in China, Applied Energy. 106 (2013) 17-24.
[11] Han, Y., Taylor, J.E., Pisello, A.L. Exploring mutual shading and mutual reflection inter-building effects on building energy performance, Applied Energy. 185 (2017) 1556-1564.
[12] Bouyer, J., Inard, C., Musy, M. Microclimatic coupling as a solution to improve building energy simulation in an urban context, Energy and Buildings. 43 (2011) 1549-1559.
[13] Perini, K., Magliocco, A. Effects of vegetation, urban density, building height, and atmospheric conditions on local temperatures and thermal comfort, Urban Forestry & Urban Greening. 13 (2014) 495-506.
[14] Gupta, K., Yang, Z., Jain, R.K. Urban Data Integration (UDI) using proximity relationship learning for design, management and operations of sustainable urban systems, Journal of Computing in Civil Engineering. (2018).
[15] Chen, J., Jain, R.K., Taylor, J.E. Block Configuration Modeling: A novel simulation model to emulate building occupant peer networks and their impact on building energy consumption, Applied Energy. 105 (2013) 358-368.