Energy and Buildings 163 (2018) 58–69

Contents lists available at ScienceDirect

Energy and Buildings

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

DUE-B: Data-driven urban energy benchmarking of buildings using recursive partitioning and stochastic frontier analysis

Zheng Yang, Jonathan Roth, Rishee K. Jain*

Urban Informatics Lab, Department of Civil and Environmental Engineering, Stanford University, United States

ARTICLE INFO

Article history:
Received 19 June 2017
Received in revised form
14 November 2017
Accepted 17 December 2017
Available online 18 December 2017

Keywords:
Benchmarking
City scale
Data-driven
Energy efficiency
Energy performance
Recursive partitioning
Stochastic frontier
Urban energy

ABSTRACT

With the world rapidly urbanizing, addressing the energy intensive urban built environment is becoming increasingly important. Cities across the United States and the world are turning to energy benchmarking as a means of understanding the relative energy efficiency of their building stock and identifying potential opportunities to reduce energy usage. Benchmarking utilizes building characteristics and energy use data to measure a building's energy consumption against a performance baseline and derive a level of energy efficiency. Over twenty cities in the United States and many others across the world have passed laws mandating the collection and disclosure of energy use data to enable benchmarking and pinpoint potential energy saving opportunities. However, municipalities are struggling to convert this data into actionable insights and identify which buildings are prime candidates for energy efficiency interventions. Although an extensive body of work exists on benchmarking building energy performance, previous works are limited in their ability to leverage such emerging data streams and conduct analysis at the city scale. Moreover, previous methods are largely based on black-box models that limit the interpretability of results and in turn hinder the ability of policy-makers to employ such models in their policy design and decision-making processes. In this paper, we propose DUE-B, a new Data-driven Urban Energy Benchmarking methodology based on recursive partitioning and stochastic frontier analysis. To test DUE-B, we evaluate its performance using real energy and building data from over 10,000 buildings in New York City, and we compare the results to other common benchmarking models using the Kendall tau-b correlation coefficient. Results indicate that DUE-B is more robust than conventional benchmarking methods in respect to identifying subsets of efficient and inefficient buildings. Furthermore, we highlight how results from DUE-B can be utilized by municipal officials and other policy-makers to target inefficient buildings for energy efficiency interventions, incentives, and programs. Specifically, we indicate how DUE-B can be utilized by municipalities to target the most inefficient buildings for subsidized onsite energy audits and less inefficient buildings for less capital-intensive energy efficiency strategies such as incentives and educational programs. In the end, more robust benchmarking methods like DUE-B have the potential to enhance the efficacy of municipal energy efficiency programs and help transition cities to a more sustainable energy future.

1. Introduction

Cities account for over 75% of all primary energy use and over 80% of direct and indirect greenhouse gas (GHG) emissions worldwide, with the bulk of such consumption and emissions coming from buildings [1]. With the world rapidly urbanizing, addressing the energy intensive urban built environment is integral to our transition to a more sustainable world. Given that the majority of existing buildings are believed to be inefficient in some capacity, enhancing the energy efficiency of our existing urban building stock through retrofits (e.g., upgrades to buildings systems, changes to building envelope) and/or adoption of energy conservation measures would yield significant energy, environmental, and economic benefits [2–5].

In order to accelerate the adoption of energy efficiency measures in cities, municipal officials and policy-makers have turned

* Corresponding author at: Civil & Environmental Engineering, Stanford University; 473 Via Ortega Way, Rm 269A, Stanford, CA 94305, USA.
E-mail address: rishee.jain@stanford.edu (R.K. Jain).

https://doi.org/10.1016/j.enbuild.2017.12.040
0378-7788/

to energy benchmarking[1] as means of understanding the energy efficiency of their building stock and identifying opportunities to reduce energy usage. As such, over twenty cities across the United States (e.g., New York, Philadelphia, Seattle, and San Francisco) have passed legislation mandating the disclosure of energy use data for a majority of their building stock [6]. Municipalities are struggling to turn this data into insights and actionable policy that can lead to effective and targeted energy efficiency programs. Current benchmarking methods are limited in their ability to handle this new type of building and energy data. Traditional engineering benchmarking methods (e.g., EnergyPlus) require the development of highly-detailed simulation models and/or an extensive on-site engineering surveys rendering them infeasible for benchmarking thousands of buildings across a city [7,8]. Other methods such as EPA's EnergyStar ranking system [9] are applicable for city scale benchmarking, but are based on national level databases and therefore fail to leverage local emerging data streams from disclosure legislation in their analysis. New data-centric methods and models [10–12] are emerging as part of a more concentrated effort on harnessing local energy usage data for benchmarking purposes. While such models have shown promise in city scale applications, those that rely on complex machine learning methods [13] reduce interpretability and thus limit their applicability to inform the design of energy efficiency policies and programs.

In this paper, we propose DUE-B, a new Data-driven Urban Energy Benchmarking method for buildings using recursive partitioning and stochastic frontier analysis. A major challenge in conducting energy benchmarking at the urban scale is the heterogeneity of building stock and its associated energy usage across an entire city. To address this challenge, we utilize an integrated approach that first partitions buildings based on both energy usage and building characteristics, with a Classification and Regression Tree (CART), and then applies Stochastic Frontier Analysis (SFA) on each partitioned group to separate random errors from sources of inefficiency in each building. Our integrated approach was designed to strike a balance between generalizability and interpretability. We aimed to create a model that could be utilized by any city despite differences in data collected as part of disclosure legislation, yet still provides robust and highly interpretable results that can inform the design and development of targeted energy efficiency policies and programs. We demonstrate the merits of our proposed DUE-B method by applying it to building data from over 10,000 buildings in New York City and compare results of our method to two other common benchmarking methods (EUI, EnergyStar) using a correlation analysis based on the Kendall tau-b metric.

2. Related work

Numerous benchmarking models exist in the literature ranging from simple to complex methods. The simplest method is the use of a key performance indicator (KPI), such as the European Energy Performance Indicator (EPI) or Energy Use Intensity (EUI). These methods are easy to compute and have been utilized in numerous previous works [14,15]. Additionally, the output of such simple methods is easy to interpret and can be used to rank buildings in order to estimate levels of energy efficiency. However, such simple indicators can only provide a rough estimate of the numerical differences between the energy efficiency of buildings and do not account for other influential factors like physical characteristics, occupancy, and building systems. Specifically, EUI values do not normalize for other factors (e.g., HVAC systems, age, number of floors) and make the erroneous underlying assumption that energy and floor area scale linearly [16]. As a result, simple indicators can be unreliable in assessing the relative energy efficiency of buildings at the city scale where there is a large variation in types, sizes and uses of buildings.

Point-based rating systems use a predefined scoring system to grade the performance of buildings and have been embedded in previous benchmarking systems [17]. Numerous point-based systems exist but for the sake of brevity we highlight two main types: credit- based and rank-based. Credit-based methods, such as Leadership in Energy and Environmental Design (LEED) in the U.S. and Building Research Establishment Environmental Assessment Method (BREEAM) in the U.K. evaluate the overall sustainability and environmental footprint of a building. However, previous work [18,19] has pointed to the limitations of credit-based systems as they do not compare the energy efficiency among buildings after they are constructed (i.e., during the occupancy phase) and cannot recognize buildings that are more operationally energy inefficient than peers. Ranking-based systems, such as EnergyStar [20], use national survey data (e.g., CBECS-Commercial Building Energy Consumption Survey) as the basis to quantify how efficient a building is compared to the best-practice peers with the same primary use type. While rank-based systems provide a quick and intuitive numeric evaluation of building energy performance, they are limited by their use of static national level data. As a result, ranking-based systems like EnergyStar are limited in their ability to provide locally contextualized results for a city and do not quantify the extent and source of energy inefficiencies within a building.

Simulation-based methods [21,22] for benchmarking apply whole building energy simulation programs, such as EnergyPlus and ESP-r, to compare current energy use of a building to a simulated building's energy use by virtually representing and reproducing periodical load dynamics, heat transfer and building-system interactions. Simulation-based methods can quantitatively account for numerous factors that influence the energy performance of a building but previous work [23] has demonstrated that simulations require time intensive expert calibration in order to yield accurate results. As a result, simulation benchmarking methods main drawback is the significant time, effort and expertise required to ensure consistent performance thereby limiting their applicability to city scale energy benchmarking.

Linear statistical models are also widely used for benchmarking building energy efficiency [24,25] by establishing a mathematically linear fit between building characteristics and energy use. Recent work [26] has improved regression models to fit a fuzzy environment by capturing the fuzzy structures in the dependent and independent variables. Despite the ease of implementation and interpretability of regression, there are several key limitations. First, the residuals measure the difference between energy use of a building with a fitted average, which cannot be directly interpreted as the actual level of energy efficiency. Second, the residuals capture relative inefficiency, statistical noise, unexplained factors and any measurement error in the dependent variables. This makes it difficult to tease out true levels of inefficiency from statistical noise and measurement error. Third, regression is highly susceptible to outliers which can skew the regression fit and is likely to occur when being applied to a diverse set of buildings across an entire city [4]. This is due to the fact that ordinary least squares regression finds the mean value of the data where outliers hold more weight. Thus, a few poor performing buildings could lead to an incorrect conclusion that the majority of buildings are efficient and thereby significantly skew energy benchmarking results.

More recently, Artificial Neural Networks (ANN) methods have been applied to building energy performance benchmarking [27,28]. ANN methods work by learning the statistical relationships between building characteristics and energy use through

[1] In this paper, we define energy benchmarking as the process of collecting building energy data and measuring a building's consumption against a performance baseline to indicate its relative energy efficiency.

a feed-forward back-propagation network. While ANN methods have demonstrated the ability to model non-linear and non-normal relationships [29–31], the complexity and black box nature of these methods make interpreting the results difficult. Moreover, because knowledge gained in the training process does not have any explanatory power and the model derived network does not have a direct physical meaning, insights into what is causing energy inefficiencies are limited. As a result, ANN based methods have limited applicability to city scale energy benchmarking where interpretability is paramount for augmenting and informing decision-making regarding energy efficiency policies and programs.

Clustering systems have also been utilized to allocate buildings into groups, where the best and worst performers within each group can be identified as baselines for energy performance benchmarking [32]. Both parametric methods (e.g., k-means) and nonparametric methods (e.g., decision trees) have been utilized in such systems. Specifically, the k-means algorithm has been utilized in previous work [33] to divide buildings into a fixed number of groups by minimizing the differences in energy use or building characteristics among different clusters. While k-means clustering is computationally efficient and has been applied to cluster a large number of buildings [34]. It requires a preset number of groups which in general is difficult to determine *a priori* by city officials and other energy efficiency decision-makers. Additionally, k-means is also unable to account for correlations among building characteristics (e.g., building area) that are common in city scale datasets and thus could result in conflicting results in benchmarking applications. Decision tree methods have been utilized in previous work [35,36] to form sub-groups by dividing buildings based on the relationship between characteristics and energy use. Recent work [37] has utilized such methods to determine the energy efficiency of a particular building by comparing its energy usage to the mean and median values within a clustered sub-group. Decision tree methods can equally handle both numeric and categorical characteristics, and are robust to missing data [38]. This ability is valuable for city scale building benchmarking given the nature and collection process of municipal benchmarking data. However, a key limitation of decision tree methods is that the statistics used for benchmarking only represent the typical or average energy performance of a group of buildings and thus cannot be solely utilized to quantify the potential level of efficiency attainable for an individual building.

Frontier analysis based benchmarking systems are used to measure the divergence of energy use between a constructed frontier, of energy-efficient buildings and any other building in the dataset. Depending on whether stochastic variations of the frontier are incorporated, there are two main types of frontier analysis: Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA) [4]. DEA considers buildings as individual decision-making units and applies fractional programming to measure the efficiency level of each building [39,40]. Although DEA can quantitatively compare relative energy performance, it cannot evaluate buildings outside the training set, and the calculated level of energy efficiency from DEA is the combination of random errors and relative inefficiency, reducing the reliability of benchmarking results. DEA's deterministic nature assumes that efficient buildings consistently exist for comparison and that variations in energy use are strictly due to technical inefficiency sources and not influenced by other random factors such as data input error, system malfunctions, etc. In order to overcome such limitations of DEA, SFA was introduced to differentiate random errors in energy use from inefficiency sources in several recent studies [41–43]. SFA is a parametric method for benchmarking that builds a regression in which it is assumed that random errors and inefficiency sources are independently distributed. An "energy efficient" frontier is established by combining the deterministic kernel and random errors to represent the maximum potential efficiency level of building energy use. The energy efficiency level of a specific building can then be precisely quantified by comparing its actual energy use with the frontier. However, previous work has only applied SFA to buildings comprised of a single type (e.g., commercial) without considering variances in the magnitude of energy usage. This poses a challenge when extending SFA to the city scale because a city's building stock is heterogeneous in nature and comprised of various types of buildings (e.g., commercial office space, retail, multi-family) that have significant variances in the magnitude of their energy usage. Such large variations can make it impossible for SFA to create a single frontier for benchmarking. As a result, applying SFA to city scale benchmarking requires pre-processing of the data to minimize the variance of both the dependent and independent variables (energy use and other building characteristics, respectively).

This paper proposes a novel integrated Data-driven Urban Energy Benchmarking (DUE-B) method that quantitatively estimates a building's relative energy efficiency (or inefficiency) and can be applied to a diverse set of buildings across an entire city. Additionally, DUE-B aims to preserve interpretability of benchmarking results in order to facilitate the design of effective and targeted energy efficiency policies and programs.

3. Data-driven urban energy benchmarking (DUE-B) method

Our proposed integrated data-driven and highly interpretable, DUE-B method contains four main steps (visually depicted in Fig. 1): 1) collect and clean building related data; 2) recursively partition buildings into different subgroups based on building characteristics and energy consumption to reduce variance in both; 3) define a stochastic frontier for each group (the maximum potential level of efficiency) and calculate energy efficiency estimates for buildings to benchmark energy performance; 4) identify the energy inefficient/efficient buildings for targeted energy policy and programs.

3.1. Recursive partitioning: CART

After data is collected and cleaned, a classification and regression tree (CART) is applied to recursively partition buildings into similar groups based on building characteristics and total energy usage or energy use intensity (EUI). The choice of dependent variable as total energy usage or EUI is determined by the modeler for a specific city and dataset. We discuss our choice for the dependent variable in our experimental test of the method later in the paper as part of Section 4.1. This clustering technique was chosen over other popular methods, like k-means or random forest, due to its high interpretability, its compatibility with categorical data, its resistance to irrelevant characteristics and its ability to easily handle missing data. Results from other clustering methods can be more difficult to interpret since distance, or density, between two buildings holds little real world meaning outside of the realm of data analysis (distance in clustering models refers to an abstract metric, measuring similarity between data points across the entire feature space, and not physical distance). These same methods often have difficulty handling categorical data, and often require tuning parameters that can be somewhat arbitrary when dealing with a problem where prediction is not the goal. CART can accept categorical data, increasing the retention of the already limited information in city datasets, and automatically chooses the most important data when constructing a tree, eliminating the use of irrelevant information. Finally, CART easily handles missing data and is robust to outliers, which is especially important when dealing with heterogeneous municipal data that often contains missing data fields. Rather than impute data for those missing entries, which

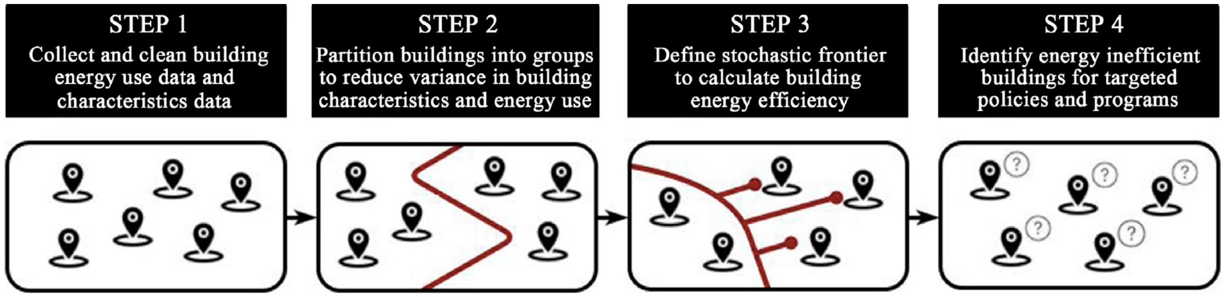

Fig. 1. Process of proposed Data-driven Urban Energy Benchmarking (DUE-B) method.

creates strong biases in building data, CART uses surrogate variables to partition buildings based on available data, which results in higher clustering accuracy. The use of CART is well suited for this application since city building data is often highly variable, heteroskedastic, and contains categorical and missing data. Moreover, CART can classify two buildings as comparable based on several characteristics, therefore avoiding potentially overly simplistic categorizations based on only one variable. This is desirable given the numerous building characteristics that can influence energy usage in a building.

First, CART performs an analysis of variance (ANOVA) process at the root node to begin building a regression tree for subsequent hierarchical binary partitions. Then at each node of the tree, building characteristics are examined to select the characteristic that results in the maximum reduction of variance in energy use (by maximizing the inter-cluster sum of squares in the analysis of variance) and then partitions the buildings into two nodes based on the selected characteristic. This process inherently reduces the variance of the independent variables as well, thereby clustering comparable buildings together. The left node holds the values that satisfy the calculated cut-off condition for the selected characteristic, while the right node contains the remaining buildings. All buildings are partitioned into smaller, lower variance groups using this binary tree process and thereby form a hierarchical structure of nodes and edges.

There are two main phases when building a CART: *growing* and *pruning*. *Growing* is when the tree is constructed and is dependent on two user-set parameters: the minimum number of buildings in a node for an attempted partition (*minsplit*) and a prior complexity (*cp*). The *minsplit* ensures that the tree's final leaves have a minimum number of data points and is set to avoid over partitioning. The *cp* is used to set the minimum reduction of mean square error (MSE) that must be attained for a node to be partitioned. This value ensures that the tree does not grow too large and that each split has a meaningful reduction in energy use variance. A low *cp* value will result in a large tree with many nodes that each contain a small number of buildings (set by *minsplit*). The inherent goal of each partition is to reduce the MSE of total building energy use and is evaluated using Eqs. (1) and (2) for node *j* which contains *N* buildings

$$\mathrm{e}_j = \frac{1}{N}\sum_{x_i \in R_j} (y_i - \bar{y}_i)^2 \tag{1}$$

$$\mathrm{e}_s = e_j - e_{jl} - e_{jr} \tag{2}$$

where y_i is the energy use of the *i*th building, and e_j is the MSE of energy use for all *N* buildings. e_{jl} and e_{jr} are the corresponding MSE for the left (*jl*) and right (*jr*) subgroup of node *j*, respectively. When e_s is greater than or equal to the user set *cp* value, the partitioning adequately satisfies the predefined threshold and is thus accepted. Otherwise *j* is set as a final cluster that will not be further partitioned. The performance of the tree is evaluated based on two metrics derived from the jackknifed error: *relative error* (equivalent to 1-R^2) and *xerror*. The *relative error* is the ratio of the MSE after running CART to the MSE before running cart, while *xerror* is the average MSE from ten-fold cross validation. The second phase for CART, which happens after the tree is grown, is *pruning*. This process eliminates branches that are formed by partitions that did not significantly reduce the *xerror*, resulting in a tree with fewer groups. To do this, the one-standard error rule (1-SE) rule is applied which determines the final number of partitions, and therefore the final *cp* value for the tree. First, the absolute minimum *xerror* found for the grown tree is summed with its corresponding standard error. Then, the *xerror* with the fewest number of partitions below this summed value is selected and its corresponding *cp* value is used to construct the final tree.

3.2. Stochastic frontier analysis

The main advantage of SFA is primarily based on its ability to differentiate a random error term from an inefficiency term [44,45]. Building energy use can be thought of similarly, where the random error term depicts unpredictable situations like system malfunctions, unexpected occupancy dynamics, weather fluctuations and data input omissions. This random error term may have either positive or negative influence on building energy use and can be generalized as stochastic variations around potential levels of energy efficiency. The inefficiency term, conversely, is a solely non-negative influence on building energy use and can be categorized into three categories: occupant behavior, building envelope, and building control systems. These types of inefficiencies are often addressed with upgrades to insulation, HVAC systems, automation technologies, demand-driven controls, lighting systems, etc. Buildings also contain many fixed characteristics that are seldom changed, like "year built" or "number of floors", that are better used to classify buildings into similar groups, and are more integral in the partitioning phase of our method (i.e., CART). Since the random factors and inefficiency factors have significantly different impacts on building energy performance and are related to different actions for improvement, we utilize SFA to quantitatively separate variations of building energy use from each other.

Within each group defined by CART, SFA is applied to determine the efficiency frontier, which signifies the maximum potential level of energy efficiency attainable, stripped from the influences of random error. For this paper, SFA is used with the cost function since we are interested in evaluating the relation between energy use intensity (EUI) and provided building services given the same *technology set*, or set of input features. By calculating the lower-bound frontier (cost function), we are establishing the minimum level of energy consumption possible given certain building characteristics, and thus the relative energy efficiency for any building can be measured. Therefore, the output of SFA is the energy use intensity (EUI) and the inputs are building characteristics. Given our goal to benchmark buildings, EUI is used as the output in SFA, as opposed to raw energy consumption, as used in CART, because relative energy efficiency is primarily an approximation of a building's energy "effectiveness" and unit energy use is more representative of this effectiveness. In this paper, we selected the commonly used Cobb-Douglas function (linear in logarithms) for the cost function [46] as shown in Eq. (3).

$$y = f(x;\beta) = \beta_0 x_1^{\beta_1} x_2^{\beta_2} \cdots x_k^{\beta_k} \tag{3}$$

As a result, the base SFA model that combines the random error term and inefficiency term can be formulated as Eq. (4),

$$y_i = f\left(x_i;\beta\right) + v_i + u_i,\ i = 1,\ldots,N \tag{4}$$

where y_i is the log of building EUI, the x_i is a vector of building characteristics log transformed, and $f()$ is the Cobb-Douglas function (Eq. (3)). β is the vector of coefficients defining the contributions of building characteristics on EUI, including an intercept value. The composite error term is separated into two independent components where $v_i \sim iidN\left(0,\sigma_v^2\right)$ accounts for the impacts of random errors and $u_i \sim iidN^+\left(0,\sigma_u^2\right)$ accounts for the sources of inefficiency. u_i is non-negative and is assumed to have a truncated half normal distribution. More importantly, u_i represents how much a building exceeds the relatively maximum potential level of efficiency. If $u_i = 0$, the building is 100% efficient. Therefore, the SFA model from Eq. (4), is found by determining the unknown values, $\beta, \sigma_v^2, \sigma_u^2$ for the Cobb-Douglas function, which can then be applied to gauge the energy performance of individual buildings. In this parametric process, the maximum likelihood principle is applied, by which actual observations from different buildings are used to estimate the parameter values that make the observations have the highest probability to happen. To do so, first the density function for $\varepsilon_i = v_i + u_i = y_i - f(x_i;\beta)$ can be formulated by the marginal density function of u_i and v_i as shown in Eq. (5),

$$\mathrm{f}_\epsilon(\varepsilon_\mathrm{i}) = \int_{\varepsilon_\mathrm{i}}^{\infty} \mathrm{f}_\mathrm{u}(\varepsilon_\mathrm{i} - \mathrm{v}_\mathrm{i})\mathrm{f}_\mathrm{v}(\mathrm{v}_\mathrm{i})\mathrm{d}\mathrm{v}_\mathrm{i} \tag{5}$$

where f_u is the density for u_i and f_v is the density function for v_i. Assuming $\varepsilon_i = u_i + v_i$ and $\sigma^2 = \sigma_u^2 + \sigma_v^2$, Eq. (5) can be transformed to Eq. (6),

$$\mathrm{f}_{\varepsilon_\mathrm{i}}(\varepsilon_\mathrm{i}) = 2\left[1 - \Phi\left(\frac{\varepsilon_\mathrm{i}\sigma_\mathrm{u}}{\sigma_\mathrm{v}\sigma}\right)\right]\left(2\pi\sigma^2\right)^{-\frac{1}{2}}\exp\left(-\frac{\frac{1}{2}{\varepsilon_\mathrm{i}}^2}{\sigma^2}\right) \tag{6}$$

where Φ is the standard normal cumulative distribution function. The density function for y_i can be defined to form the log-likelihood function for a group of buildings, as shown in Eqs. (7) and Eq. (8),

$$\mathrm{L}(\theta;\mathrm{y}) \equiv -\frac{1}{2}\mathrm{N}\ln(\frac{1}{2}\pi) - \frac{1}{2}\mathrm{N}\ln\sigma^2 + \sum_{\mathrm{t}=1}^{\mathrm{N}}\ln\left[1 - \Phi(\mathrm{z}_\mathrm{i})\right] - \frac{1}{2}\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{y}_\mathrm{i} - \mathrm{x}_\mathrm{i}\beta)^2/\sigma^2 \tag{7}$$

$$z_i = \left[\frac{y_i - x_i\beta}{\sigma}\right]\left[\frac{\gamma}{1-\gamma}\right]^{\frac{1}{2}} \tag{8}$$

where θ represents the parameter vector for $(\beta,\ \sigma^2,\ \gamma)$ and $\gamma = \frac{\sigma_u^2}{\sigma_u^2+\sigma_v^2}$, which ranges from 0 to 1. If γ is close to 0 when either $\sigma_u^2 \to 0$ or $\sigma_v^2 \to \infty$, the model is an "average frontier model" with no inefficiency among the buildings. If γ is close to 1, the frontier is deterministic and called a "full frontier model" with no stochastic noise. In cases where only one data period is available, an error component frontier is utilized without a time effect [47]. In the parameterization, σ_u^2 and σ_v^2 are replaced by $\sigma^2 = \sigma_u^2 + \sigma_v^2$, and $\gamma = \frac{\sigma_u^2}{\sigma_u^2+\sigma_v^2}$ for the derivation of the log-likelihood function. The maximum likelihood estimates of parameters for the SFA model are obtained by equating the first order Taylor expansion of the logarithm of likelihood function (with respect to β, σ^2, γ) to zero. There is no direct closed-form solution thus a three-step numerical search is implemented: 1) attain the unbiased β vector from the ordinary least square (OLS) estimates; 2) perform a grid search and calculate γ, where all β values are the OLS estimates while the intercept and σ are updated according to the corrected ordinary least squares algorithm [48]; 3) implement an iterative optimization method to replace β^k with the newly calculated value β^{k+1} by Eq. (9) until β^{k+1} does not change from β^k(given by: $|\beta^{k+1} - \beta^k| < 10^{-4}$) to obtain the final maximum likelihood estimates of the parameters [49]. If the optimization does not converge or return parameters for the global maximization of the likelihood function, the input of building characteristics are not sufficient or appropriate to construct the SFA model and require revision.

$$\beta^{k+1} = \beta^k - \left(\frac{\partial^2 L\left(\beta^k\right)}{\partial\beta^2}\right)^{-1}\frac{\partial L\left(\beta^k\right)}{\partial\beta} \tag{9}$$

After the functional form is estimated, the deviations of individual-building energy use intensity can be decomposed into noise and inefficiency. The energy efficiency score for an individual building, denoted as *efficiency estimate* (units of kBtu/ft^2), is defined as the unit energy saving potential by comparing the actual site EUI with the fitted frontier site EUI ($f\left(x_i;\beta\right) + v_i$) to represent best building practices (*actual site EUI – fitted frontier site EUI*). The *efficiency estimate* can range between zero and infinity, where a score of one indicates a fully energy *efficient* building, while a score of infinity indicates a completely energy *inefficient* building. A high *efficiency estimate* indicates a high potential for unit energy savings for a building, and poor building energy performance. Buildings that are prime candidates for interventions or retrofits can then be identified based upon the *efficiency estimates*. One main advantage of stochastic frontier analysis is its ability to handle uncertainty, which prevents random error impacting the *efficiency estimates* and can accurately form the lower-bound frontier for buildings with certain characteristics. To keep our proposed method generalizable, this paper does not allocate energy inefficiency to specific sources (e.g., elements vs. systems).

4. Data, results, and discussion

We utilize a dataset from New York City (NYC) to validate and demonstrate the merits of our proposed DUE-B method. NYC passed Local Law 84 (LL84) in 2009, which was the first major mandate requiring the disclosure of building energy consumption in the United States [50]. Under LL84, buildings larger than 50,000 square feet and tax lots with combined building area of more than 100,000 square feet were required to disclose their annual energy consumption to the city [51]. Energy data collected through LL84 has been utilized in several previous works [52–54]. The energy data in this study is for the 2014 calendar year and is supplemented and merged with a comprehensive dataset regarding basic building characteristics that was acquired from New York City's Primary Land Use Tax Lot Output (PLUTO) database. The PLUTO dataset includes fields about building features for every tax lot in NYC [55]. It is maintained by city governmental agencies including the Department of City Planning, the Department of Finance, and the Department of City-wide Administrative Service. Initially, the merged dataset contained information for 13,912 buildings in the city. An extensive data cleansing process was performed to eliminate buildings with missing energy use data, combine repeated cells, and correct for contradictory or erroneous building characteristics (e.g., year altered is earlier than year built). The descriptions of the sixteen building characteristics used in this paper are summarized in the Appendix A. The description in the Appendix includes all variables that are of potential interest and only eliminates variables such as address, code numbers, and other information not pertinent to the study. Several buildings with reported EUI numbers higher than 1,000 kBtu per square feet and were assumed to be outliers based on heuristics in previous work [56] and removed. Finally, buildings with a use type that appeared less than one percent of the time were eliminated as these buildings are considered to be highly specialized, making their comparison to the remaining building stock unsuitable. In total 10,153 buildings were retained after the data cleansing process and entered into the CART analysis.

A preliminary test was conducted using SFA on the whole dataset to test its efficacy on the non-partitioned, highly variable data. The log likelihood test result ($\gamma \to 0$) gave a value of zero, indicating no inefficiency among buildings could be determined (i.e., all of the error was contained in the random error term v_i). The results of this preliminary test indicate that the variance in building characteristics and energy usage is too large on the entire dataset to provide a basis for comparison in the SFA benchmarking model and underscores the need to run CART to form subgroups within the large city wide dataset. Inherently, the CART partitioning process reduces the variance in both energy consumption and building characteristics and thereby provides a better peer group of buildings for comparison.

4.1. CART results

We implemented CART using the *rpart* package in the R programming language [57] to recursively partition the 10,153 buildings in the dataset. Total building energy use (in kBtu) was log-transformed to account for the wide range and positively skewed energy consumption values. For this experimental test-case, we utilized total energy consumption instead of EUI as the dependent variable as we did not want to make the strong assumption that energy usage scales linearly with floor area as this assumption has been demonstrated to not hold across datasets that are heterogeneous in nature [4]. We acknowledge that this choice of dependent variable introduces its own set of challenges as it could be possible that two buildings with similar energy usage but differing characteristics could be partitioned into the same group. However, the choice of CART as our partitioning method provides some robustness against this issue as it aims to reduce the variance in the target variable by partitioning buildings on selected key features (i.e., building characteristics), thus also reducing the variance of the feature space. Additionally, we would also like to note that the DUE-B method is flexible in nature and the modeler has freedom to set the dependent variable and independent variables based on observed trends within their own dataset.

Using *cp* and *minsplit* (set to 10) the sample size of buildings was partitioned to reduce the variance in building characteristics and energy usage. To keep the tree as simple as possible while maximizing the reduction of variance, the determination of *cp* was advised by the 1-SE rule. We applied this rule to prune the tree and obtain a simpler structure with fewer partitions than the original tree. The resulting *cp* was then selected as the final threshold capacity (*cp* = 0.00197 and *xerror* = 0.2093). After pruning, our final tree (shown in Fig. 2) consisted of 15 groups partitioned on the basis of three of the sixteen building characteristics our method analyzed: Property Floor Area, Building Class, and Assessed Total Value. The building class characteristic encompasses information about the regulations, technologies, architectural trends, construction quality, and design work for different types of buildings and indicate the importance of these aspects to energy usage. The presence of property floor area in our final CART model signifies the importance of size in relation to building energy consumption. Intuitively, larger buildings will consume more energy than smaller buildings as internal systems will have to condition larger spaces and must meet the needs of a greater number of occupants. Assessed total value is a significant variable as it encompasses several key attributes of a building including quality of build, implemented technology, and general opulence; these factors are known to affect energy performance and can provide some precursory insights into building operations.

The CART algorithm successfully handled buildings with missing feature values. These buildings were processed using the surrogates determined by the CART algorithm, which are constructed by finding an alternative characteristic that most closely matches the partitioning results of the original characteristic. If data was missing for all surrogates, buildings were partitioned to be included in the group with the most data. As shown in Fig. 2, the maximum number of buildings in a group is 2732 (Group 1) while the minimum number is 40 (Group 15), resulting in a mean of 477 and a median of 543. To visually depict the number of buildings partitioned at each step, the branch widths in Fig. 2 are sized proportionally to the number of buildings in the ensuing node. The different shades of color in the figure represent the mean energy use (log-transformed) for each group; the darker shade represents high mean in energy use while the light shade represents low mean energy use. As shown in Fig. 3, each group contains buildings with different values and variances of energy use, further demonstrating the need to partition buildings into groups before proceeding with benchmarking using SFA. These boxplots show that there is still a sizeable spread in the data for each cluster (i.e., exhibiting potential inefficiencies), but the variance is much smaller than that of the whole dataset. This eliminates the issues SFA has with highly variable data where it is unable to find any inefficiencies within the dataset. By using this tree, a building owner can quickly determine which cluster their building belongs to. For example, a building that is less than 181,612 square feet would go down the first left branch. If the building was more than 97,488 square feet, it would then proceed down the next right branch. Then if the building had an assessed total value greater than $10,421,550 it would go down the right branch. Lastly if the building belonged to class C (walk-up apartments), D (elevator apartments), L (loft buildings), R (condominiums) or W (educational structures) it would proceed down the left branch and be classified in group 7 that consists of 583 buildings.

Another interesting note, is the exclusion of building class for several clusters. Typically, building class is a key attribute when comparing energy use of buildings, yet examining buildings solely on this attribute has its drawbacks. For example, it is very difficult to compare an apartment building with four units to one with seven-hundred units. Instead, CART can partition these buildings so that floor area is also considered since large buildings are more likely to have similar systems in place even if the buildings are of a different type. Therefore, CART would deem two very large buildings as comparable, and separate the small and large apartment structures into different clusters. Also consider the situation where two medium sized apartment buildings have very different consumption levels. At first glance, it is easy to consider them comparable buildings, but they may be considered dissimilar due to other building characteristics; the high energy consuming building may be equipped with lavish apartments and includes a gym, pool, and spa. CART may classify this building as being more like a hotel, which often includes well-conditioned rooms, a gym, and a pool. The high consuming apartment building and the hotel may be clustered into the same node based on assessed value of the property, as the low consuming apartment building is not as luxurious as the other two buildings. Given the added level of services offered by both the hotel and high consuming apartment building, it makes more sense to compare these two structures due to their similar amenities, size, and perceived value, rather than solely on their building type.

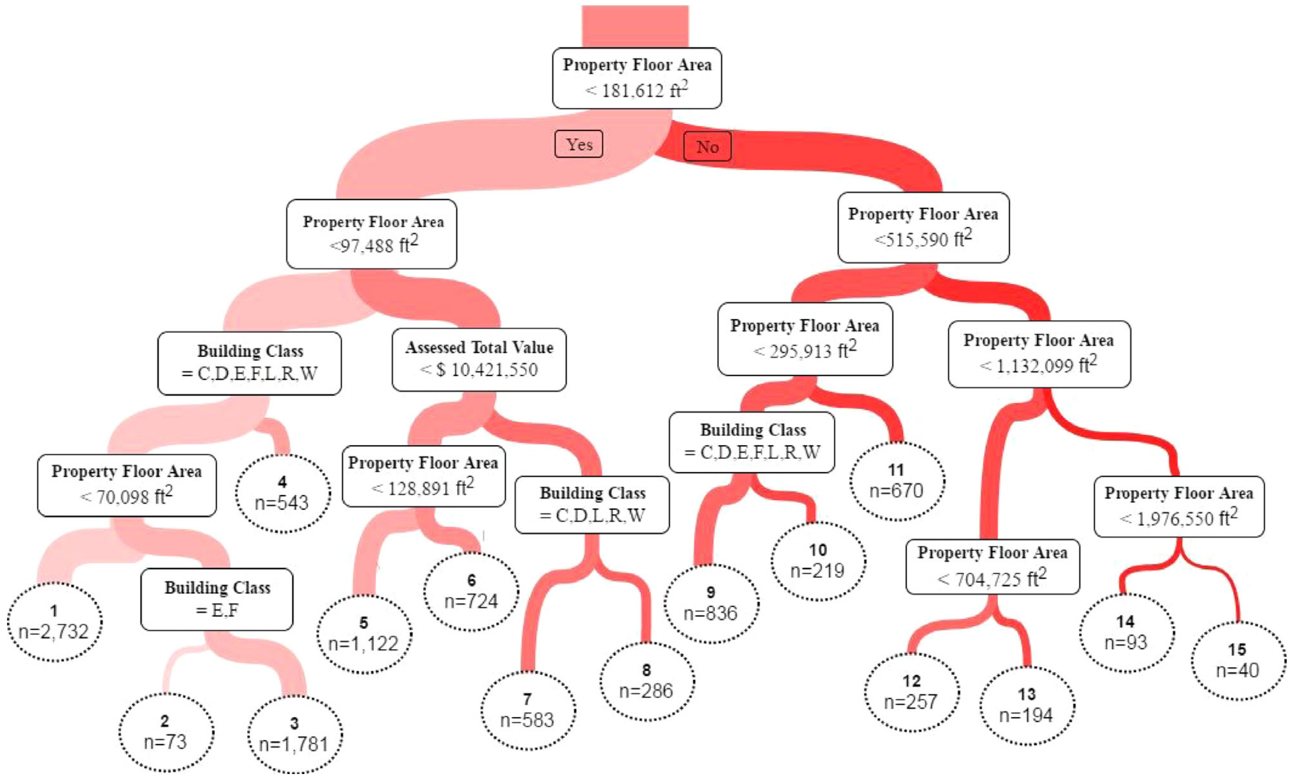

Fig. 2. Classification and regression tree for each group (different tones represent different levels of mean log-transformed total energy use for different groups. Lighter color indicates lower mean energy usage, darker indicates higher mean energy usage).

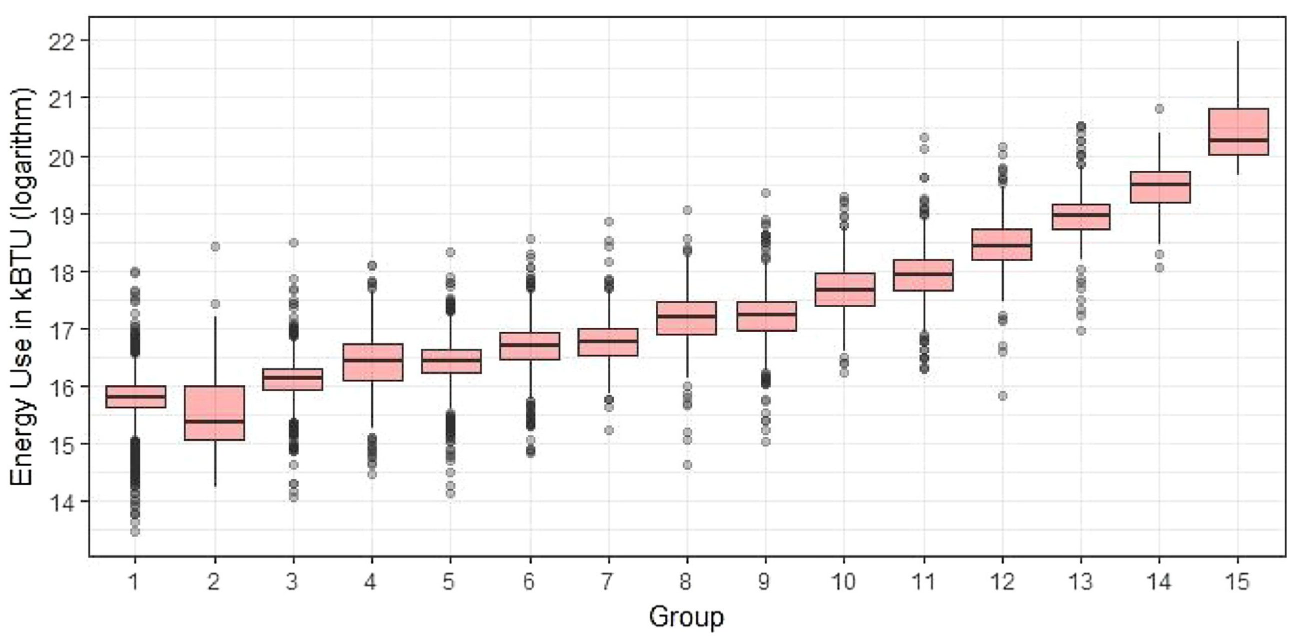

Fig. 3. Boxplot of energy use for all partitioned groups.

4.2. SFA results

SFA was implemented for each partitioned cluster of buildings using the *sfa* package in the R programming language [58]. All available building characteristics were used to perform an exhaustive search to determine the optimal subset of features based on goodness of fit (R^2 value) to use for the final SFA model for each cluster. Since the Cobb-Douglas cost function is used in this paper, both the output (building site EUI) and numeric explanatory inputs (buildings characteristics) were log-transformed. In the *frontier* package, *Error Components Frontier (ECF)* was selected because only panel data without time effects are analyzed in this paper. *ineffDecrease* was set to FALSE because inefficiency increases the composite error term ε over a lower-bound frontier for building energy use intensity. *minusU* was set to FALSE to calculate the Shepard-type efficiencies which represent the $E\left[exp\left(\hat{u}\right)\right]$. The maximum likelihood principle, explained in Section 3.2, was then applied to estimate the SFA functional form. To be clear, the characteristics

Table 1
Maximum Likelihood Estimates (MLEs) of coefficients for SFA model of cluster 2.

Building Characteristics	Coefficient Estimate
Building floor area	−12.742
Building front	−0.040
Assessed value	0.010
Commercial floor area	11.856
Building depth	−0.006
Year of built	−16.420
Number of floors	−0.189
Lot front	0.089
Built FAR	−0.013

Table 2
Maximum Likelihood Estimates (MLEs) of parameters for SFA model of cluster 2.

Parameter	Estimate
Intercept	137.079
γ	0.856
Log-likelihood (ECF)	−77.691
σ^2	1.180
λ	2.441
Log-likelihood (OLS)	−79.424
σ_u^2	1.011
σ_v^2	0.169
Ratio Test	0.031

to construct SFA for different clusters are not necessarily the same between different clusters, since there are different relationships between site EUI and building characteristics for different types of buildings. As an example, the estimates of the SFA model for cluster 2, which contains 73 buildings (warehouses, factory and industrial buildings), are shown in Tables 1–2. The parameter $\gamma = 0.856$ is close to the upper bound of 1, signifying that most variations in site EUI stem from sources of inefficiency, with only a fraction originating from other random errors. The likelihood ratio test was also performed to show that the inefficiency term was statistically significant, demonstrating that differences regarding energy efficiency among buildings were identified. Our estimated coefficients were also consistent with findings from previous studies; for example, larger buildings consumed energy less intensely than their smaller counterparts [43].

The boxplots of *efficiency estimates* for different clusters of buildings are presented in Fig. 4, along with histograms of each cluster's distribution to provide a more holistic summary of results. Due to the random error term, several buildings were found to have actual EUI values that were lower than the frontier. Rather than assign them with a negative difference and deem them more efficient than the frontier, their *efficiency estimates* were set to zero, indicating maximum efficiency. Although some buildings perform relatively well, the boxplots show that the distributions have a long right tail, demonstrating that there are a substantial number of highly inefficient buildings with a wide range of magnitudes. Inefficient buildings with *efficiency estimates* much larger than the upper hinge of the boxplot (1.5% interquartile) could be classified as the worst performing buildings, in their respective partitioned clusters (indicated by the circle dots in Fig. 4). These inefficient buildings contained would be prime candidates for targeted energy efficiency policies or programs. Examining the histograms in Fig. 4, the majority of clusters have distributions with long right tails. This indicates a high variability among the most inefficient buildings (*the higher the inefficiency estimate, the more inefficient the building*). Significant differences exist in *efficiency estimates* among different clusters, which confirms the important contributions of different building characteristics to the partitioning of buildings into sub-clusters. These histograms can facilitate the analysis of building energy performance among the partitioned clusters by city officials and others. A kernel density estimation (KDE) was also applied by using a Gaussian function as the smoothing kernel in order to provide an estimate of the underlying distribution (density) of energy *efficiency estimates* for all clusters (solid red lines in Fig. 4)

5. Validation and implications

5.1. Validation of proposed DUE-B method

In order to validate the performance of DUE-B, results were compared to two widely used benchmarking methods: site EUI and Portfolio Manager EnergyStar. We chose to compare our results with both EnergyStar scores and EUI due to the fact that many cities (e.g., New York, San Francisco, and Philadelphia) are utilizing both metrics in their current benchmarking reports [59–61]. We applied the Kendall tau-b correlation coefficient as a non-parametric test to evaluate the association between two benchmarking methods without prior assumptions about the frequency distribution of the results. Since different benchmarking methods have different ways of representing the level of building energy efficiency, comparing the rank orders of energy efficiency among different methods is more robust than taking the absolute values. If more than one building had the same benchmarking results (e.g., the same EnergyStar scores), they were assigned to the same rank. Kendall's tau-b rank correlation coefficient was then calculated to represent the level of correlation between each pair of methods. The equation for calculating the Kendall tau-b correlation coefficient is provided in Eq. (10):

$$\tau_b = \frac{(\#\,concordant\ pairs) - (\#\,discordant\ pairs)}{\sqrt{N1} \times \sqrt{N2}} \tag{10}$$

When building A is evaluated as more energy efficient than building B by the two methods, A–B is called a concordant pair; when the relative rank of building A and building B is evaluated as opposite by the two methods, A–B is called a discordant pair. In Eq. (10), if the number of concordant pairs is much larger or smaller than that of discordant pairs, the two methods are positively or negatively correlated. The correlation coefficient ranges from −1 to 1. In addition, in Eq. (10) N_1 and N_2 are the numbers of pairs with different levels of energy efficiency by the two methods to be compared.

Before the analysis was performed, buildings with missing EnergyStar scores were assigned the average score of buildings in the same CART cluster. The results and correlation coefficients for all 15 clusters are shown in Fig. 5. The Kendall tau-b correlation analysis shows that the *efficiency estimates* consistently correlate higher with EUI and EnergyStar scores than they do with each other. The displayed robustness of DUE-B may indicate that the two existing models are overlooking, or oversimplifying, intra-building effects that are causing the building to be inefficient in its use of energy. We postulate that our proposed DUE-B method is more robust than site EUI and EnergyStar because neither of these methods differentiate random factors from inefficiency factors and because EnergyStar scores are not computed with respect to local peer buildings. Perhaps more importantly, cities that have either switched from EUI to EnergyStar or are simultaneously using both methods may be experiencing inconsistent and variable results that could undermine the effectiveness of ensuing policies and programs. It is important to note that the results presented in Fig. 5 are contingent on the use of both CART and SFA as our preliminary test results (see Section 4) demonstrated the need for both methods to be utilized in conjunction. The results are also contingent of our choice of dependent variable (total energy usage) in the CART partitioning process of our method. We underscore that this choice yielded more robust results for the New York City dataset but the choice for the dependent variable can be modified by the modeler based on observed trends in their benchmarking dataset.

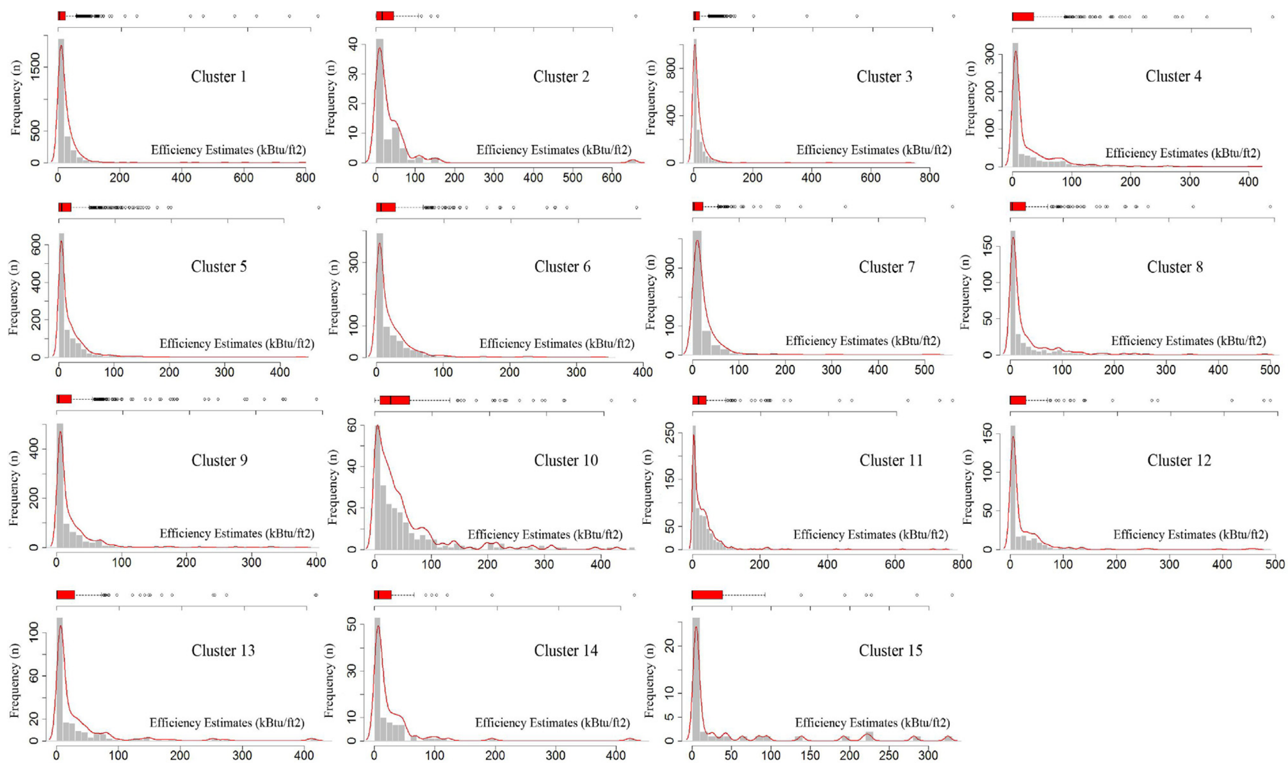

Fig. 4. Boxplot, histogram, and density of *efficiency estimates* for the CART partitioned clusters of New York City buildings.

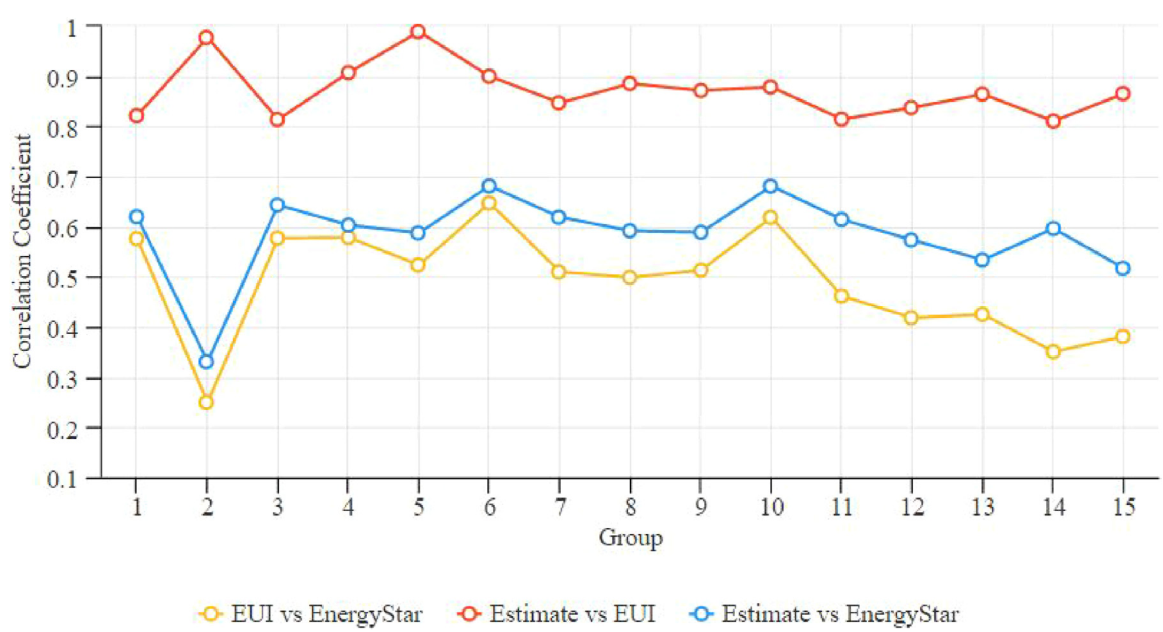

Fig. 5. Kendall's tab-b correlation test of SFA inefficiency term $E[\exp(u_i)]$ by the DUE-B proposed method (short as Estimate), site EUI (short as EUI), and Energy Star score (short as EnergyStar) for all fifteen clusters.

5.2. Implications for urban energy efficiency

Overall, this paper contributes a new methodological approach to the growing body of research in the area of data-driven energy benchmarking [9,34,37]. This work aims to build on previous work [13,62] that conducts analysis of building energy performance at a city-scale by introducing a new data-driven benchmarking method that utilizes local publicly available data, maintains interpretability, and produces more robust results than currently used benchmarking methods (EUI, EnergyStar). Moreover, this research extends previous work on stochastic frontier based energy analysis [43] to city-scale energy benchmarking by integrating a stochastic frontier model with a highly-interpretable recursive partitioning method.

It is generally difficult for municipal officials to use data from buildings across a city to develop and implement energy efficiency policies and programs. Resources to implement and interpret data-driven analytical tools are limited, as is the ability to translate benchmarked data into effective policies and targeted programs. Our primary focus in this paper was to introduce a new benchmarking method that leverages new disclosure data streams and

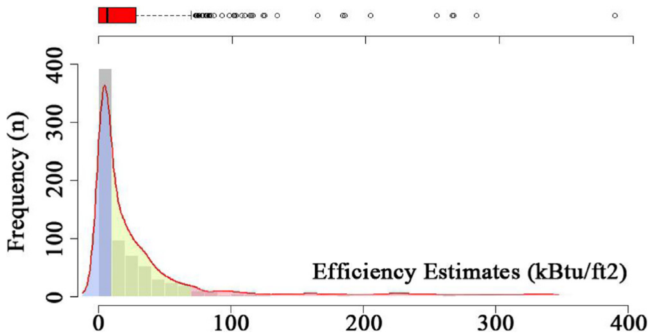

Fig. 6. *Efficiency Estimates* for Group 6. The red shaded region (right tail) is the 5% "most inefficient" buildings (n = 38), the green shaded region is the 40% "relatively inefficient" buildings (n = 292), and the blue shaded region is the 55% "relatively well performing" buildings (n = 394). A total of 724 buildings are in this group. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

provides municipal officials with robust and interpretable estimates of the energy efficiency of their entire urban building stock. This paper provides a thorough investigation on how computational methods can be used effectively to identify significant energy savings while maintaining interpretability for a wide audience.

Specifically, results from our model could facilitate the design and implementation of targeted building energy efficiency policies and programs. For example, Fig. 6 shows the distribution of *efficiency estimates* for Group 6 from our results. Different regions are shaded on the graph to indicate subsets of buildings that could be targeted with different interventions. The red region represents the 5% most inefficient buildings (shown on the far right tail) and could be deemed most suitable for subsidized on-site energy audits, as they are mostly likely to have significant potential for energy savings. The larger green region encompases the next 40% of inefficient buildings that could be targeted through less costly incentives, mandates, or educational outreach programs. Finally, the blue region holds the remaining 55% of buildings where less focus is needed because they are shown to be relatively efficient compared to their peers. In the end, DUE-B aims to help municipal officials allocate the limited resources they have for energy efficiency programs more intelligently and enable them to reach their energy and sustainability goals.

6. Limitations and future work

While our proposed method addresses many of the shortcomings (e.g., limited leveraging of local benchmarking data, not scalable for city wide analysis, difficult to interpret, unable to differentiate between random factors and inefficiency factors) in existing benchmarking models, its introduction also precipitates several limitations. First, the data we analyzed did not allow our proposed method to quantitatively allocate inefficiency to individual sources within a building. With the current state of the model, its best use case is to determine which buildings are considered inefficient within a city, where effort could then be further spent determining the drivers of consumption within those poor performers with subsidized on-site energy audits and/or in-depth energy simulations. Once more detailed data is collected and made available, our model is designed to be flexible and could leverage such data to better explore the drivers of inefficiencies.

Like any data-driven, the methodology proposed and the resulting conclusions would be more profound if more granular or complete data (e.g., HVAC systems, building orientation, exterior enclosure, occupancy behavior) was acquired. Yet even with current limited data, the DUE-B model demonstrated its robustness when compared to conventional benchmarking methods (EUI, EnergyStar). In future work, we aim to test the use and implementation of DUE-B with city officials, buildings managers, and other parties to assess the interpretability of our methodology and determine if the results can be appropriately used to inform the design and development of energy efficiency policies and programs.

7. Conclusions

This paper proposes DUE-B, a new data-driven methodology for benchmarking building energy consumption at the urban scale that integrates recursive partitioning and stochastic frontier analysis. We propose DUE-B as an alternative to existing benchmarking models due to its numerous advantages. Specifically, DUE-B eliminates human data entry, exclusively uses publicly available datasets, is not susceptible to biased scores as a result of outliers or heterogeneous data and maintains a high-level of interpretability to facilitate decision-making. As with other benchmarking methodologies, quantitatively establishing its superiority is difficult due to the fact that ground truth data of energy efficiency are infeasible to collect for thousands of buildings across a city. Nevertheless, we established some quantitative advantages of this model through the novel application of the Kendall tau-b correlation analysis and demonstrate that our model is more robust than EUI and EnergyStar. This robustness is critical as unreliable and irregular results can compromise potential energy efficiency policies and programs that are informed by benchmarking models.

Specifically, a major challenge in conducting energy benchmarking at the urban scale is the heterogeneity of building stock and associated energy usage across an entire city. To address this challenge, our proposed DUE-B model first uses a classification and regression tree (CART) to recursively partition buildings into different subgroups with reduced variance in both building characteristics and energy usage. Stochastic frontier analysis is then applied to construct the efficiency frontier for each individual subgroup, which represents the theoretical level of efficient energy consumption that buildings are benchmarked against. Unlike traditional benchmarking methods, DUE-B leverages local energy disclosure data, enables city scale analysis and can successfully separate the impact of random factors from that of inefficiency factors while maintaining interpretability. We applied our proposed DUE-B method to energy and building data from over 10,000 buildings in New York City and results indicate that our proposed method is more robust than two other commonly used benchmarking methods (EUI, EnergyStar).

Overall, this research represents a key first-step towards addressing the energy intensive urban built environment. By leveraging emerging data streams being collected by cities around the world, our proposed model aims to enhance our understanding of how our urban buildings consume energy and identify opportunities for improving energy efficiency. Utilizing these data-driven insights, municipal officials and policy-makers can develop effective energy efficiency policies and programs that tackle the most inefficient buildings in a city and in turn realize substantial energy, environmental and economic savings. Amidst growing urbanization, cities have become a focal point for energy efficiency and new data-driven methods will undoubtedly be crucial to helping transition our cities to a more energy efficient and sustainable future.

Acknowledgements

This material is based upon the work supported in part by the Stanford Precourt Institute for Energy, Stanford School of Engineering under a Terman Faculty Fellowship and National Science Foundation under Grants No. 1461549 and 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 the National Science Foundation. The authors also acknowledge the City of New York for granting access to the LL84 and PLUTO datasets through their Open Data portal.

Appendix A.

Field Name	Characteristic and Explanation	
Property floor area	property area: the gross square footage of the property, per the Department of Finance records (ft^2)	
Building class	building class: the major use of structures of the property	
	A	one family dwellings
	B	two family dwellings
	C	walk-up apartments
	D	elevator apartments
	E	warehouses
	F	factory and industrial buildings
	G	garages and gasoline stations
	H	hotels
	I	hospital and health
	J	theatres
	K	storage buildings (taxpayers included)
	L	loft buildings
	M	churches, synagogues, etc.
	N	asylums and homes
	O	office buildings
	P	places of public assembly (indoor) and cultural
	Q	outdoor recreation facilities
	R	condominiums
	S	residential – multiple use
	T	transportation facilities (assessed in ore)
	U	utility bureau properties
	V	vacant land
	W	educational structures
	Y	selected government installations
	Z	miscellaneous
Assessed total value	assessed total value: the tentative assessed total value of the property ($)	
Assessed land value	assessed land value: the tentative assessed land value ($)	
Area unit price	price per square feet: the unit price per building area ($/ft^2)	
Building floor area	total building floor area: the total gross area in square feet (ft^2)	
Lot area	lot area: the total area of the tax lot (ft^2)	
Number of buildings	number of buildings: the number of buildings on the lot	
Number of floors	number of floors: the number of full and partial stories	
Number of total units	number of total units: the units in all buildings on the lot	
Number of residence units	number of residence units: the residential units in all buildings on the lot	
Commercial floor area	commercial floor area: exterior dimensions of the structure for commercial use (ft^2)	
Residential floor area	residential floor area: exterior dimensions of the structure for residential use (ft^2)	
Year built	year built: the year building construction completed	
Building frontage	building frontage: the building frontage along the street (ft)	
Building depth	building depth: the effective perpendicular distance (ft)	

References

[1] United Nations-Habitat, Energy, 2017 (Accessed on 5 September 2017) http://unhabitat.org/urban-themes/energy/.

[2] C.A. Balaras, A.G. Gaglia, E. Georgopoulou, S. Mirasgedis, Y. Sarafidis, D.P. Lalas, European residential buildings and empirical assessment of the Hellenic building stock, energy consumption, emissions and potential energy savings, Build. Environ. 42 (2007) 1298–1314, http://dx.doi.org/10.1016/j.buildenv.2005.11.001.

[3] E. Hoşgör, P.S. Fischbeck, Virtual home energy auditing at scale: predicting residential energy efficiency using publicly available data, Energy Build. 92 (2015) 67–80, http://dx.doi.org/10.1016/j.enbuild.2015.01.037.

[4] W. Chung, Review of building energy-use performance benchmarking methodologies, Appl. Energy 88 (2011) 1470–1479, http://dx.doi.org/10.1016/j.apenergy.2010.11.022.

[5] R.K. Jain, J.E. Taylor, G. Peschiera, Assessing eco-feedback interface usage and design to drive energy efficiency in buildings, Energy Build. 48 (2012) 8–17, http://dx.doi.org/10.1016/j.enbuild.2011.12.033.

[6] Map: U.S. Building Benchmarking and Transparency Policies, Institute for Market Transformation, 2017 (Accessed on 5 September 2017) http://www.imt.org/resources/detail/map-u.s.-building-benchmarking-policies.

[7] P. Hernandez, K. Burke, J.O. Lewis, Development of energy performance benchmarks and building energy ratings for non-domestic buildings: an example for Irish primary schools, Energy Build. 40 (2008) 249–254, http://dx.doi.org/10.1016/j.enbuild.2007.02.020.

[8] Z. Yang, B. Becerik-Gerber, A model calibration framework for simultaneous multi-level building energy simulation, Appl. Energy 149 (2015) 415–431, http://dx.doi.org/10.1016/j.apenergy.2015.03.048.

[9] P.A. Mathew, L.N. Dunn, M.D. Sohn, A. Mercado, C. Custudio, T. Walter, Big-data for building energy performance: lessons from assembling a very large national database of building energy use, Appl. Energy 140 (2015) 85–93, http://dx.doi.org/10.1016/j.apenergy.2014.11.042.

[10] E.H. Borgstein, R. Lamberts, J.L.M. Hensen, Evaluating energy performance in non-domestic buildings: a review, Energy Build. 128 (2016), http://dx.doi.org/10.1016/j.enbuild.2016.07.018.

[11] E.H. Borgstein, R. Lamberts, Developing energy consumption benchmarks for buildings: bank branches in Brazil, Energy Build. 82 (2014), http://dx.doi.org/10.1016/j.enbuild.2014.07.028.

[12] L. Pérez-Lombard, J. Ortiz, R. González, I.R. Maestre, A review of benchmarking, rating and labelling concepts within the framework of building energy certification schemes, Energy Build. 41 (2009) 272–278, http://dx.doi.org/10.1016/j.enbuild.2008.10.004.

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

[14] C. Filippín, Benchmarking the energy efficiency and greenhouse gases emissions of school buildings in central Argentina, Build. Environ. 35 (2000) 407–414, http://dx.doi.org/10.1016/S0360-1323(99)00035-9.

[15] S.-M. Deng, J. Burnett, A study of energy performance of hotel buildings in Hong Kong, Energy Build. 31 (2000) 7–12, http://dx.doi.org/10.1016/S0378-7788(98)00067-X.

[16] W. Chung, Y.V. Hui, Y.M. Lam, Benchmarking the energy efficiency of commercial buildings, Appl. Energy 83 (2006) 1–14, http://dx.doi.org/10.1016/j.apenergy.2004.11.003.

[17] T. Hong, M.A. Piette, Y. Chen, S.H. Lee, S.C. Taylor-Lange, R. Zhang, K. Sun, P. Price, Commercial building energy saver: an energy retrofit analysis toolkit, Appl. Energy 159 (2015) 298–309, http://dx.doi.org/10.1016/j.apenergy.2015.09.002.

[18] J.H. Scofield, Efficacy of LEED-certification in reducing energy consumption and greenhouse gas emission for large New York City office buildings, Energy Build. 67 (2013) 517–524, http://dx.doi.org/10.1016/j.enbuild.2013.08.032.

[19] G.R. Newsham, Do LEED-certified buildings save energy? Not really, Energy Build. 41 (2009) 1386–1390, http://dx.doi.org/10.1016/j.enbuild.2009.08.006.

[20] U.S. Environmental Protection Agency, EnergyStar, Environmental Protection Agency, U.S, 2017 (Accessed on 5 September 2017) https://www.energystar.gov/.

[21] W.L. Lee, J. Burnett, Benchmarking energy use assessment of HK-BEAM, BREEAM and LEED, Build. Environ. 43 (2008) 1882–1891, http://dx.doi.org/10.1016/j.buildenv.2007.11.007.
[22] A.-T. Nguyen, S. Reiter, P. Rigo, A review on simulation-based optimization methods applied to building performance analysis, Appl. Energy 113 (2014) 1043–1058, http://dx.doi.org/10.1016/j.apenergy.2013.08.061.
[23] M. Ahmad, C.H. Culp, Uncalibrated building energy simulation modeling results, HVAC&R Res. 12 (2006).
[24] W. Xuchao, R. Priyadarsini, L. Siew Eang, Benchmarking energy use and greenhouse gas emissions in Singapore's hotel industry, Energy Policy 38 (2010) 4520–4527, http://dx.doi.org/10.1016/j.enpol.2010.04.006.
[25] P. Bohdanowicz, I. Martinac, Determinants and benchmarking of resource consumption in hotels-case study of Hilton International and Scandic in Europe, Energy Build. 39 (2007) 82–95, http://dx.doi.org/10.1016/j.enbuild.2006.05.005.
[26] W. Chung, Using the fuzzy linear regression method to benchmark the energy efficiency of commercial buildings, Appl. Energy 95 (2012) 45–49, http://dx.doi.org/10.1016/j.apenergy.2012.01.061.
[27] M. Yalcintas, U. Aytun Ozturk, An energy benchmarking model based on artificial neural network method utilizing US Commercial Buildings Energy Consumption Survey (CBECS) database, Int. J. Energy Res. 31 (2007) 412–421, http://dx.doi.org/10.1002/er.1232.
[28] S.-M. Hong, G. Paterson, D. Mumovic, P. Steadman, Improved benchmarking comparability for energy consumption in schools, Build. Res. Inf. 42 (2014) 47–61, http://dx.doi.org/10.1080/09613218.2013.814746.
[29] F. Khayatian, L. Sarto, G. Dall, Application of neural networks for evaluating energy performance certificates of residential buildings, Energy Build. 125 (2016) 45–54, http://dx.doi.org/10.1016/j.enbuild.2016.04.067.
[30] S.L. Wong, K.K.W. Wan, T.N.T. Lam, Artificial neural networks for energy analysis of office buildings with daylighting, Appl. Energy 87 (2010) 551–557, http://dx.doi.org/10.1016/j.apenergy.2009.06.028.
[31] H.-X. Zhao, F. Magoulès, A review on the prediction of building energy consumption, Renew. Sustain. Energy Rev. 16 (2012) 3586–3592, http://dx.doi.org/10.1016/j.rser.2012.02.049.
[32] X. Gao, A. Malkawi, A new methodology for building energy performance benchmarking: an approach based on intelligent clustering algorithm, Energy Build. 84 (2014), http://dx.doi.org/10.1016/j.enbuild.2014.08.030.
[33] C.E. Kontokosta, A market-specific methodology for a commercial building energy performance Index, J. Real Estate Finance Econ. 51 (2015) 288–316, http://dx.doi.org/10.1007/s11146-014-9481-0.
[34] I. Farrou, M. Kolokotroni, M. Santamouris, A method for energy classification of hotels: a case-study of Greece, Energy Build. 55 (2012) 553–562, http://dx.doi.org/10.1016/j.enbuild.2012.08.010.
[35] C. Koo, T. Hong, M. Lee, H.S. Park, Development of a new energy efficiency rating system for existing residential buildings, Energy Policy 68 (2014) 218–231, http://dx.doi.org/10.1016/j.enpol.2013.12.068.
[36] Z. Li, Y. Han, P. Xu, Methods for Benchmarking Building Energy Consumption Against Its Past or Intended Performance: An Overview, 2014, http://dx.doi.org/10.1016/j.apenergy.2014.03.020.
[37] H.S. Park, M. Lee, H. Kang, T. Hong, J. Jeong, Development of a new energy benchmark for improving the operational rating system of office buildings using various data-mining techniques, Appl. Energy 173 (2016) 225–237, http://dx.doi.org/10.1016/j.apenergy.2016.04.035.
[38] H. Dahan, S. Cohen, L. Rokach, O. Maimon, Proactive Data Mining With Decision Trees Springer Briefs in Electrical and Computer Engineering, 2014, http://dx.doi.org/10.1007/s13398-014-0173-7.2.
[39] W.S. Lee, Benchmarking the energy efficiency of government buildings with data envelopment analysis, Energy Build. 40 (2008) 891–895, http://dx.doi.org/10.1016/j.enbuild.2007.07.001.
[40] P. Zhou, B.W. Ang, K.L. Poh, A survey of data envelopment analysis in energy and environmental studies, Eur. J. Oper. Res. 189 (2008) 1–18, http://dx.doi.org/10.1016/j.ejor.2007.04.042.
[41] A. Kavousian, R. Rajagopal, Data-driven benchmarking of building energy efficiency utilizing statistical frontier models, J. Comput. Civil Eng. 28 (2014) 79–88, http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000327.
[42] M. Filippini, L.C. Hunt, US residential energy demand and energy efficiency: a stochastic demand frontier approach, Energy Econ. 34 (2012) 1484–1491, http://dx.doi.org/10.1016/j.eneco.2012.06.013.
[43] J. Buck, D. Young, The potential for energy efficiency gains in the Canadian commercial building sector: a stochastic frontier study, Energy 32 (2007) 1769–1780, http://dx.doi.org/10.1016/j.energy.2006.11.008.
[44] D. Aigner, C.A.K. Lovell, P. Schmidt, Formulation and estimation of stochastic frontier production function models, J. Econom. 6 (1977) 21–37, http://dx.doi.org/10.1016/0304-4076(77)90052-5.
[45] W. Meeusen, J. van Den Broeck, Efficiency estimation from Cobb-Douglas production functions with composed error, Int. Econ. Rev. 18 (1977) 435, http://dx.doi.org/10.2307/2525757.
[46] G.E. Battese, G.S. Corra, Estimation of a production frontier model: with application to the pastoral zone of eastern Australia, Aust. J. Agric. Econ. 21 (1977) 169–179, http://dx.doi.org/10.1111/j.1467-8489.1977.tb00204.x.
[47] G.E. Battese, T.J. Coelli, Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India, J. Product. Anal. 3 (1992) 153–169, http://dx.doi.org/10.1007/BF00158774.
[48] T. Coelli, Estimators and hypothesis tests for a stochastic frontier function: a Monte Carlo analysis, J. Product. Anal. 6 (1995) 247–268, http://dx.doi.org/10.1007/BF01076978.
[49] M.M. Pitt, L.-F. Lee, The measurement and sources of technical inefficiency in the Indonesian weaving industry, J. Dev. Econ. 9 (1981) 43–64, http://dx.doi.org/10.1016/0304-3878(81)90004-3.
[50] A.C. Burr, C. Keicher, D. Leipziger, Building Energy Tranparency: A Framework for Implementing U.S. Commercial Energy Rating and Disclosure Policy, 2017 (Accessed on 5 September 2017) http://www.imt.org/uploads/resources/files/IMT-Building_Energy_Transparency_Report.pdf.
[51] City of New York, The Local Laws of City of New York, 2017 (Accessed on 5 September 2017) http://www.nyc.gov/html/gbee/downloads/pdf/nycbenchmarkinglaw.pdf.
[52] R.K. Jain, J.M. Moura, C.E. Kontokosta, Big data + big cities: graph signals of urban air pollution, IEEE Signal Process. Mag. 5 (2014) 130–136, http://dx.doi.org/10.1109/MSP.2014.2330357.
[53] D. Hsu, How much information disclosure of building energy performance is necessary? Energy Policy 64 (2014) 263–272, http://dx.doi.org/10.1016/j.enpol.2013.08.094.
[54] C.E. Kontokosta, R.K. Jain, Modeling the determinants of large-scale building water use: implications for data-driven urban sustainability policy, Sustain. Cities Soc. 18 (2015) 44–55, http://dx.doi.org/10.1016/j.scs.2015.05.007.
[55] City of New York. PLUTO, 2017 (Accessed on 5 September 2017) https://www1. nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page.
[56] C.M. Saldanha, S.M. O'Brien, a study of energy use in new York city and leed-Certified buildings, IBPSA-USA J. 6 (2016) (Accessed on 11 May 2017) http://ibpsa-usa.org/index.php/ibpusa/article/view/382.
[57] T. Therneau, B. Atkinson, B. Ripley, M.B. Ripley, R Package 'rpart', 2017 (Accessed on 5 September 2016) http://cran.ma.ic.ac.uk/web/packages/rpart/rpart.pdf.
[58] A. Straub, R Package 'sfa, 2017 (Accessed on 5 September 2016) https://cran.r-project.org/web/packages/sfa/sfa.pdf.
[59] City of New York, New York City Local Law 88 Benchmarking Report, 2017 (Accessed on 5 September 2017) http://www1. nyc.gov/site/planning/data-maps/open-data.page.
[60] San Francisco Environment, San Francisco Existing Commercial Buildings Performance Report, 2017 (Accessed on 5 September 2017) https://sfenvironment.org/download/san-francisco-existing-commercial-buildings-performance-report-20102014.
[61] City of Philadelphia, City of Philadelphia 2016 Energy Benchmarking Report, 2016 (Accessed on 5 September 2017) http://www.phillybuildingbenchmarking.com/wp-content/uploads/2016/04/2016-Benchmarking-Report.pdf.
[62] D. Hsu, Improving energy benchmarking with self-reported data, Build. Res. Inf. 42 (2014) 641–656, http://dx.doi.org/10.1080/09613218.2014.887612.