Applied Energy 221 (2018) 576–586

Contents lists available at ScienceDirect

Applied Energy

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

Pattern recognition in building energy performance over time using energy benchmarking data

Sokratis Papadopoulos, Bartosz Bonczak, Constantine E. Kontokosta*

Department of Civil and Urban Engineering and Center for Urban Science and Progress, New York University, 370 Jay Street, 12th Floor, Brooklyn, NY 11201, United States

HIGHLIGHTS

- We analyze building energy time series data to identify patterns over time.
- We use a large-scale, time series cross-sectional dataset of energy disclosure data.
- Machine learning methods are used to define under- and over-performing clusters.
- Our results show a differential response to energy disclosure.

ARTICLE INFO

Keywords:
Energy benchmarking
Energy disclosure
Time series clustering
Machine learning
Building energy performance
New York City

ABSTRACT

In recent years, many cities have adopted energy disclosure policies to better understand how energy is consumed in the urban built environment and how energy use and carbon emissions can be reduced. The diffusion of such policies has generated large-scale streams of building energy data, creating new opportunities to develop the fundamental science of urban energy dynamics. Nevertheless, there is limited research that rigorously analyzes building energy performance patterns over time. This paper provides a comprehensive framework to analyze building energy time series data and identify buildings with similar temporal energy performance patterns. We use data from approximately 15,000 properties in New York City, covering a six-year reporting period from 2011 to 2016. After pre-processing and merging the data for each constituent year, we use an unsupervised learning algorithm to optimally cluster the energy time series and statistical tests and supervised learning methods to infer how building characteristics vary between clusters. Our results show that energy reductions in New York City are mainly driven by its commercial building stock, with larger, newer, and higher-value buildings demonstrating the largest improvements in energy intensity over the study period. Moreover, voluntary energy conservation schemes are found to be more effective in boosting energy performance of commercial properties, compared to residential buildings. Our results suggest two distinct temporal patterns of energy performance for commercial and residential buildings, characterized by energy use reductions and increases. This finding highlights the differential response to energy reporting and disclosure, and presents a more complex picture of energy use dynamics over time when compared to previous studies. In order to realize significant energy use improvements over time and reach energy and carbon reduction goals, cities need to design and implement comprehensive energy policy frameworks, bringing together information transparency and reporting with targeted mandates and incentives.

1. Introduction

1.1. Background and motivation

It is now largely acknowledged that climate change is a global threat and immediate action needs to be taken to mitigate its most significant effects [1]. Existing buildings are responsible for approximately 40% of primary energy consumption worldwide, drawing the attention of energy policy and carbon emission reduction efforts [2,3]. Given the density and scale of the urban built environment, cities are leading the way in climate action [4], with many setting aggressive long-term carbon reduction goals (e.g. New York City (NYC) aims to reduce its carbon footprint by 80% by 2050 [5]). These climate action plans inevitably focus on the building sector as a source for improved energy

* Corresponding author.
E-mail address: ckontokosta@nyu.edu (C.E. Kontokosta).

https://doi.org/10.1016/j.apenergy.2018.03.079
Received 8 December 2017; Received in revised form 22 February 2018; Accepted 25 March 2018

efficiency and reduced emissions, a fact driven by the large proportion of energy consumed by buildings in urbanized areas and the relatively high return on investment for energy efficiency improvements [6].

To support energy use and carbon emission reductions, energy disclosure ordinances constitute a significant policy tool to accelerate building energy efficiency market transformation [7]. There is a growing body of cities and local governments that have adopted such energy disclosure mandates in the United States (U.S.), with Austin, Texas and Washington, DC being among the first in 2008, and New York City (NYC) following in 2009 [8,9]. Recently, smaller municipalities, such as Cambridge (2014), Boulder (2015) and Berkeley (2015), among others, have enacted similar reporting policies [10]. These ordinances mandate that building owners report their property's energy consumption and, by extension, allow decision makers (DMs) to benchmark building performance and more completely assess a city's energy use profile. The rationale behind energy benchmarking can be encapsulated in Michael R. Bloomberg's statement in 2010, then Mayor of NYC: *"You cannot manage what you do not measure, and benchmarking the City's buildings lets us determine where energy costs can be reduced"* [11]. According to Perez-Lombard [12], governments should consider energy disclosure and benchmarking as the basis of any energy efficiency policy pertaining to the building sector, prior to additional actions, as there are multiple benefits of having such energy data available. From the end users' perspective, the simple act of reporting consumption might increase tenants' awareness of energy issues and lead to end use reductions through behavior changes or impacts on locational decisions [7,13]. For DMs, monitoring and reporting energy data allows them to track progress towards energy reduction goals, understand how energy is consumed at the urban scale [14,15], or develop market-specific energy performance metrics [16–18], to name a few examples.

Early adopters of energy disclosure policies, such as NYC, have already collected as much as seven years of data [19]. Aggregating these data presents an unprecedented opportunity to fill the gaps in existing research by analyzing the temporal energy performance patterns in individual buildings. Given the novelty of these data, and their relative sparsity to date, there are few, if any, studies that have attempted to examine such relationships. The main purpose of this research is to detect and analyze buildings with similar energy performance patterns over time and identify common characteristics they might share. Specifically, we seek to: (a) develop an optimal method to cluster building energy performance time series data using unsupervised learning, (b) statistically test the difference in various building characteristics within the identified clusters, and (c) assess the likelihood of a building belonging to a certain performance cluster given its characteristics. This knowledge will enable stakeholders and policy makers to study subgroups of buildings with similar energy behavior, and to understand the factors that promote, or hinder, energy efficiency adoption and improvements. Given the significant energy and carbon reduction goals established by cities around the world, this is a critical element of understanding the potential for energy savings in buildings and to target policy interventions to improve performance over time across different sub-groups of buildings. In the remainder of this section, we provide an overview of current research on building energy data and energy performance, highlighting gaps in the literature that our research attempts to address. In Section 2, we describe in detail our data and methods, followed by a presentation of the results (Section 3). The paper concludes with a discussion of the findings and their relevance for energy decision-making and energy efficiency policy and regulations.

1.2. Limitations in previous research

Although energy disclosure is a relatively recent policy innovation, the rapid diffusion of such policies across cities and states has resulted in new, large-scale data streams, which have catalyzed a growing body of research on city-wide building energy consumption and performance [20,8,21,22,16,23,13,24]. While these studies contribute to an understanding of how buildings consume energy in urban areas, many have been constrained by the nature and volume of data available at the time of the research. This has resulted in both limitations to previous work and, given the growing adoption of disclosure policies and availability of data, new opportunities to address existing research gaps.

A majority of previous research is focused on analyzing static snapshots of buildings' energy performance, rather than dynamic performance trends over time [21,16,23]. Since the earliest data available, released by NYC, only dates to 2010, much of the previous literature is limited to quantifying energy performance as a time-invariant peer comparison, or to understand the drivers of energy use, such as building age or size. In one of the earliest studies of energy disclosure data, Kontokosta [16] used energy and correlative data from approximately 20,000 buildings in NYC, obtained through the City's Local Law 84 (LL84) energy benchmarking ordinance. Specifically, the author analyzed the relationship between commercial buildings' energy use intensity and design, system, occupancy, and spatial land use characteristics using a multivariate robust regression model. The study found that characteristics such as age, size, construction type, and occupancy significantly influence a building's energy intensity. Focusing on NYC's residential building stock, Ma and Cheng [21] employed a random forests algorithm to analyze the influence of 171 different features on energy intensity. In addition to building-related attributes, the authors include socioeconomic and demographics features in their analysis. Their results showed that areas with lower educational attainment and higher percentage of fuel oil-heated buildings tend to be more energy intense, likely a function of the quality of housing for different income groups. A paper by Reina and Kontokosta [20] focuses on the issue of social equity in energy demand, specifically analyzing the relative energy efficiency of subsidized (low-income) housing. The authors find non-trivial differences between subsidy type and program, such as Public Housing, as well as the importance of sub-metering to building energy efficiency. In another recent study [14], the authors used reported energy data in an attempt to predict city-scale energy consumption, using utility data as a validation set. They trained several machine learning models on 23,000 buildings reported under LL84, and predicted the energy use of 1.1 million buildings in NYC. Their results demonstrated that city-wide electricity consumption can be predicted accurately from a relatively small sample of buildings, whereas natural gas consumption prediction is a more complex problem given utility infrastructure constraints and the bimodal distribution of use between heating and cooking. Energy benchmarking data have also been used in web-based visualization tools [25,26], as a means to provide transparency to the public and create a feedback loop of disclosed information back to building owners to support data-driven decision-making.

As energy efficiency improvements and long-term carbon reductions are the ultimate goal of climate action, understanding changes over time is critical to effective policy design, evaluation, and implementation. Recent studies that do include more than one year of disclosed energy data [8,24,27] try to capture the overall effect of the adoption of disclosure policies on the existing building stock's energy efficiency, rather than focusing on sub-groups of buildings and their energy behavior across multiple years. There are only two existing studies utilizing panel energy disclosure data [8,24]. In both cases, the authors proposed a difference-in-difference regression model to capture the overall effect of the adoption of a disclosure law on energy efficiency. Based on their model, Meng et al. [8] suggest that NYC's disclosure policy reduced energy intensity by 6%, on average, in the three years after its implementation, and by 14% after the fourth year. Although the results of this study are limited by selection bias in the control group and the absence of pre-intervention data, it provides a useful exploratory policy evaluation of disclosure ordinances. Palmer and Walls [24] provide an insightful overview of the issues in evaluating disclosure policies as a driver of energy use reductions; however, they

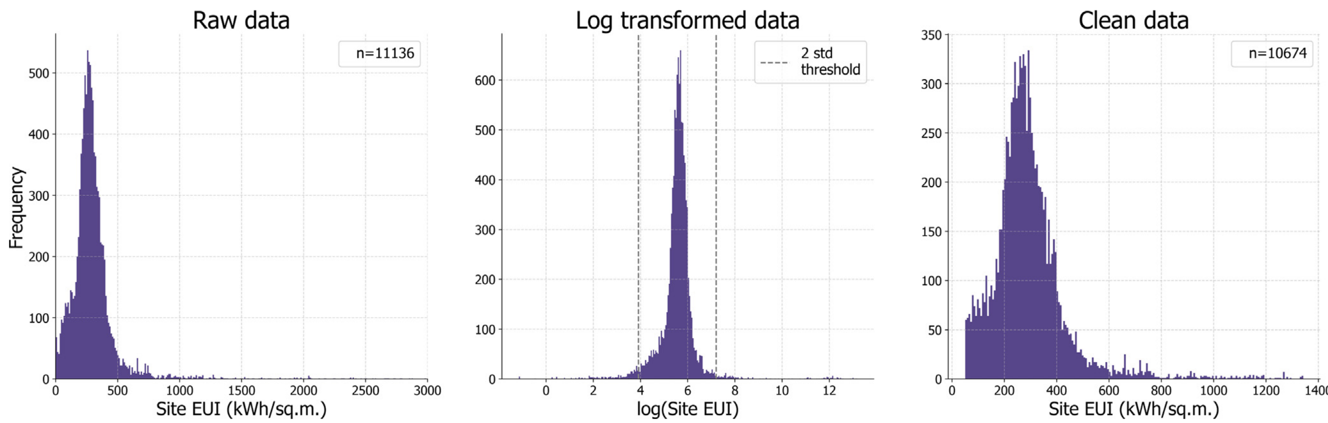

Fig. 1. Raw data distribution for multifamily housing stock (left), logarithmic transformed data (center), and cleaned distribution (right). Dashed lines indicate the two standard deviation outlier detection thresholds.

perform only a descriptive analysis of changes over time given significant data limitations. From the perspective of driving energy and carbon reductions in cities, these studies focus only on quantifying the overall effect of the adoption of the policy on energy use, rather than the differential response of buildings that lead them to reduce (or increase) their consumption and the characteristics of buildings that can explain such an effect. Our hypothesis is that energy awareness and the adoption of energy conservation measures varies across building typologies, and thus there are fundamental differences in the likelihood of any particular building improving its energy performance over time. Discovering similarities in buildings that reduce their energy use, in absence of regulations or requirements to do so, will help cities evaluate potential energy efficiency and carbon reduction strategies, ranging from simple reporting polices to mandates and codes stipulating specific energy performance standards.

While predictive modeling using energy data has been studied extensively, research on identifying buildings with similar characteristics for targeted energy policy is limited. Kontokosta [16] used K-means clustering to establish building performance grades, following the calculation of a novel energy performance index for NYC's commercial building stock. Hsu [23] compared various clustering algorithms on building energy consumption data to assess the tradeoff between cluster stability and predictive accuracy. The results showed that K-means yields more stable clusters when the correct number of clusters is chosen. Both studies highlight the strengths of clustering as a tool for building sub-grouping and targeted energy policy.

Whether predictive or descriptive, the analyses discussed in this section come with a major limitation: they study static snapshots of energy performance, typically focusing on one year of data, and thus omit a temporal analysis of building energy use over multiple years. An immediate benefit of energy disclosure has been found to be that measurement of a building's energy use can lead, by itself, to energy use reductions [28]. Consequently, there is a need to analyze energy performance trends in buildings subject to energy disclosure laws to identify patterns and shared characteristics between buildings that reduce their consumption, and those that do not.

2. Data and methods

In this section, we discuss the data sets used in this study and our methods to: (a) cluster buildings with similar energy performance patterns and (b) extract insights from these patterns. We begin by describing our data pre-processing and cleaning prior to the analysis. Then, we detail our clustering approach and statistical tests for whether buildings in the same performance clusters share similar characteristics. Finally, we discuss the methods we use to quantify the impacts of individual building and management attributes on the likelihood of a building belonging to a particular cluster.

2.1. Data sources

NYC's LL84 energy benchmarking is the main data set used in this study, covering seven years of reporting (2010–2016) and consisting of more than 100 features for each of the approximately 15,000 properties (accounting for 21,000 buildings) included. In effect since 2010, the LL84 ordinance mandates all properties larger than 4645 square meter (sq.m.) (equivalent to 50,000 square feet (sq.ft.)) to report their annual energy and water consumption [29]. Along with this information, the data set includes features such as fuel type contribution to energy end-use, occupant density, and building physical characteristics. Energy use is reported in three ways: (i) absolute values (annual kBtu), (ii) normalized by building area (annual kBtu/sq.ft.) as energy use intensity (EUI), and (iii) weather normalized EUI. In this paper, we consider weather normalized site EUI as our primary variable of interest, hereinafter referred to as *EUI*. The LL84 data set covers properties in all five boroughs of NYC, accounting for approximately 45% of the City's building energy consumption and almost 280 million sq.m. of space [16]. For the purpose of this study we focus on the two main building typologies encountered in the data: (a) Office and (b) Multifamily housing.

LL84 data are self-reported; therefore, we need to remove any misreported or anomalous (*outlier*) entries prior to any analysis [30]. First, we remove observations with duplicate or missing Borough, Block, and Lot (BBL) number. The BBL number is a unique spatial identifier for tax lots in NYC. We then remove entries with zero or missing values in their reported energy use or weather normalized site EUI. Misreported values in either the energy use or gross floor areas fields (e.g. accidental addition/omission of zeros, misreported units) can dramatically distort EUI values [14]. For each property type, we apply a logarithmic transformation to the EUI values in order to approximate the normal distribution given the log-normal distribution observed in the raw data. Then, we identify observations falling outside the threshold of two standard deviations from the logged sample mean as outliers and filter them out of our analysis. Fig. 1 shows an illustration of the outlier detection process.

We repeat the cleaning process for each year of available data (2010–2016) and use the BBL identifier as a matching key to merge the individual data sets into a panel. Starting from the first reporting year, we append the following year's data only if the property consistently reports its energy consumption each year. Due to lower compliance rates in the first year of LL84 implementation (2010), we omit 2010 data from the analysis and use 2011 as the base year for the time series analysis (Fig. 2).

In order to test various hypotheses on building energy performance

Fig. 2. Compliance/noncompliance rates for the years LL84 has been active (2010–2016). Numbers in the bar plots indicate the number of properties reported each year.

time trends, we incorporate additional data sources into the analysis. The NYC Primary Land Use Tax Lot (PLUTO) database is a detailed property and land use data set, containing the geographic coordinates of each building, as well as physical, zoning, and financial attributes [31]. We use PLUTO data to extract the geolocation, height, age, and asset value for each building in the LL84 data set. Local Law 87 (LL87) is NYC's energy audit and retro-commissioning law. Each year, a portion (approximately 10%) of LL84 buildings is required to undergo an energy audit and report the results [32]. We use LL87 data to flag properties that conducted an energy audit in years 2013 or 2014, the third and fourth year of our panel, respectively. LL87 data are confidential, and provided to the authors by the NYC Mayor's Office of Sustainability. Another data source we use includes all NYC buildings with fuel oil boilers. The oil boiler data are publicly available through the City's open data portal [33]. We identify properties with heavy oil boilers (i.e. those that use fuel oils #5 or #6) and match them with LL84 data using their BBL identifier. Finally, we identify all commercial and residential buildings participating in the NYC Carbon Challenge [34], a public–private partnership between real estate owners and the City of New York aiming to reduce greenhouse gas emissions by at least 30% over the course of ten years. This voluntary program is used as a proxy for buildings whose owners or managers have explicitly stated their intention to improve energy efficiency over time.

2.2. Clustering the energy performance time series

For each building, we extract the weather normalized site EUI value time series from 2011 to 2016, and scale the range of each time series in [0, 1]. Time series scaling allows the clustering algorithm to exclusively account for changes in the energy performance trend, rather than its absolute value. Additionally, feature scaling is an essential pre-processing step for distance-based clustering algorithms that boosts their convergence and accuracy [35,36]. We apply K-means clustering to identify groups of buildings with similar energy performance over time. The K-means algorithm is a partitioning unsupervised learning algorithm that separates the data into K equal variance groups, by minimizing the within-sample sum of squares (i.e. inertia) (Eq. (1)).

$$argmin \sum_{i=1}^{k} \sum_{j=1}^{n} (||x_j - \mu_i||)^2 \tag{1}$$

where, k is the number of clusters, n the number of data points, $||x_j - \mu_i||$ is the Euclidian distance between data point x_j and cluster center μ_i.

K-means starts by randomly assigning cluster centers and calculates the distance between the centers and each data point. In step two, it assigns each data point to the cluster with the minimum distance from its center. In step three, the algorithm recalculates cluster centers as the mean of the data points assigned in the cluster (Eq. (2)).

$$\mu_i = \frac{1}{n} \sum_{j=1}^{n} x_j \tag{2}$$

Steps two and three are repeated until the algorithm converges [37].

Although simple in nature, the K-means algorithm is a powerful, widely used machine learning tool [38], that has been successfully applied in various applications [39–42]. Examples of K-means clustering applications in energy-related research can be found in the works of Hsu [23], Yu et al. [43], Rasanen et al. [44], Gaitani et al. [45].

2.2.1. Selecting the number of clusters

However fast and robust, K-means comes with a major drawback: the number of clusters k needs to be defined *a priori*. To overcome this limitation, we run K-means with several k values in an attempt to identify its optimal value. This equates to the value that minimizes the variance within the clusters, while maximizing the variance among different clusters. We employ two performance metrics, the Silhouette score [46] and the Dunn index [47] to assess this trade-off. For a given assignment of clusters, higher metric values indicate better clustering.

2.3. Statistical tests on clusters' differences

Once we successfully form clusters of buildings with similar energy performance patterns over time, we test various hypotheses on the characteristics of buildings belonging to different clusters. We are concerned with identifying statistically significant differences in building attributes between the clusters to determine if certain features are associated with increasing (or decreasing) energy performance over time. Moreover, we use a spatial correlation index to test whether

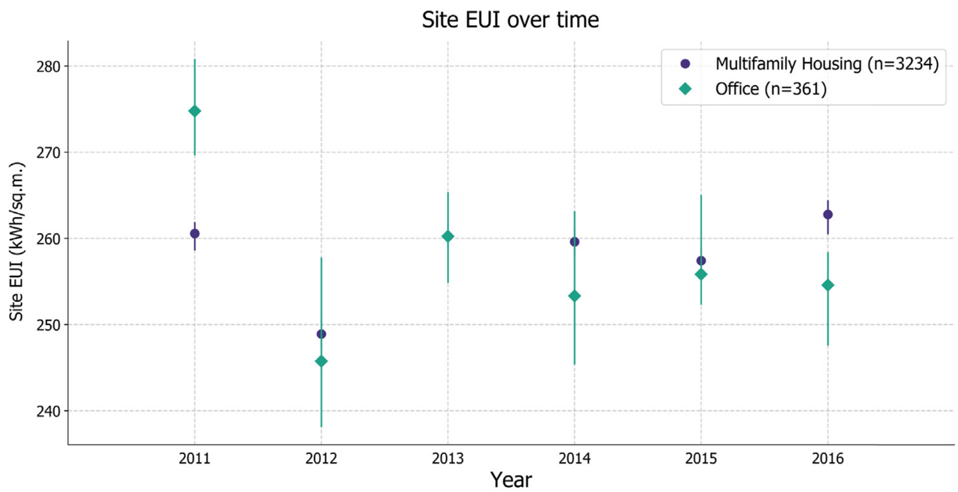

Fig. 3. Median site EUI values for multifamily housing and office typologies (2011–2016).

spatial relationships exist in the data. It is conceivable that buildings in the same cluster, and therefore those that behave similarly over time, may be spatially proximate. This could result from local competition between buildings for tenants, the geographic focus of certain larger management or real estate companies, or spatially-correlated building attributes (such as year of construction) [48].

2.3.1. Mann–Whitney U test

We use the Mann–Whitney *U* test [49] to determine if continuous features between two independent groups are homogeneous and follow the same distribution. For instance, we want to examine whether buildings belonging to different clusters vary in age, size, or assessed value. The Mann-Whitney *U* test is a non-parametric test, and unlike the two sample *t*-test does not rely on distribution assumptions. It tests the null hypothesis (*Ho*): two samples come from the same population, against the alternative hypothesis (*Ha*): the distribution of the one sample differs from the other's.

2.3.2. Fisher's exact test

Fisher's exact test of independence is used to test differences in proportions in nominal features [50]. In the case of energy performance time series clustering, this is used to answer questions about the impact of discrete activities or interventions, such as whether the proportion of buildings that have conducted an energy audit is different between two clusters. *Ho* suggests that the proportions of one group are the same as another, whereas *Ha* suggests that there is difference in these proportions.

2.3.3. Moran's I index

Moran's I index is a measure of global spatial autocorrelation. Unlike one-dimensional autocorrelation, spatial autocorrelation is multi-dimensional and characterizes the relationship between data points that are located nearby in space. Moran's I index tests the null hypothesis of spatial randomness [51]. We use Moran's I to determine if there is spatial correlation between buildings assigned to different clusters.

2.4. Association between building characteristics and cluster assignment

In the last part of our methodology, and building on the previous steps, we identify the factors explaining a building's energy performance pattern (expressed as its cluster assignment). We specify a logistic regression model to predict cluster assignment, and based on the model's coefficients we quantify the degree of association between a building's characteristics and its assigned cluster.

Logistic regression is a well-established linear classification approach for both binary and multinomial problems [52]. The logistic regression model predicts the logit transformation of the probability of an outcome given a set of independent variables (Eq. (3)).

$$logit(P) = log\frac{P(X)}{1-P(X)} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_\nu X_\nu \quad (3)$$

where *P(X)* is the probability of an outcome given X, *X* are the model's independent variables, and β are the regression coefficients.

To remove non-significant variables and obtain the optimal logistic regression model, we run a backward feature elimination algorithm. Backward elimination starts with fitting a model to all candidate variables. Then the least significant variable is omitted and the new model's performance is evaluated based on the goodness of fit (i.e. likelihood-ratio statistic). The process is repeated until removal of additional variables compromises the likelihood-ratio [53].

2.4.1. Odds ratio and model interpretation

By taking the exponential of Eq. (3) for both sides of the equation, it can be rewritten as:

$$odds = \frac{P(X)}{1-P(X)} = e^{\beta_0} + e^{\beta_1 X_1} + e^{\beta_2 X_2} + \cdots + +e^{\beta_\nu X_\nu} \quad (4)$$

From Eq. (4), we observe that a unit change in X_ν results in the odds changing by a factor $e^{(\beta_\nu)}$. Factor $e^{(\beta_\nu)}$ is a measure of association between dependent and independent variables and defined as the odds ratio. The odds ratio explains the change in the likelihood of the outcome when an independent variable changes, adjusted for the influence of confounding variables.

Finally, for each coefficient β we compute the Wald statistic as:

$$Wald = \left[\frac{\beta}{s.\ e.\ (\beta)}\right]^2 \quad (5)$$

where $s.\ e.\ (\beta)$ is the standard error of the coefficient. Wald statistic is the square of the t-statistic and we use it as a measure of variable importance in the logistic regression model.

3. Results

Basic descriptive statistics are shown in Fig. 3 for the aggregated LL84 data. In the base year of 2011, it is clear that office buildings are more energy intense compared to residential. However, offices have reduced their median energy use intensity by 8%, on average, during the six-year study period, a finding that does not hold for multifamily housing, which exhibited a modest increase in intensity over time. By the end year of our analysis, multifamily housing properties show

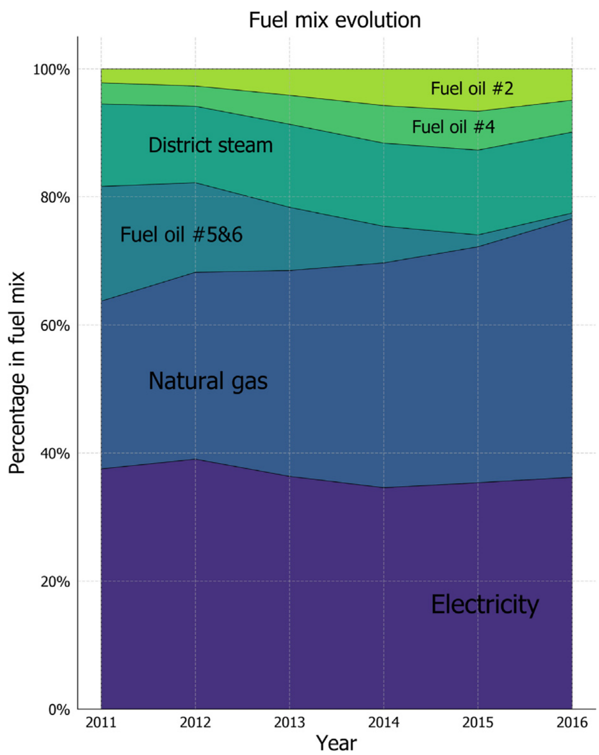

Fig. 4. Fuel mix for office and multifamily housing buildings.

higher median site EUIs compared to offices. Finally, we observe a steep drop in EUI in 2012 for both building typologies. This anomaly is largely attributed to hurricane Sandy, which hit NYC during October 2012 and caused widespread power outages [8].

In Fig. 4, we show NYC's fuel mix for both Office and Multifamily housing buildings. Interestingly, heavy fuel oils, including types #5 and #6, accounted for almost 20% of the fuel mix in 2011, but have been almost eliminated by 2016. This confirms NYC's success in phasing out the use of these fuel oils [54] and their replacement by cleaner energy sources, particularly natural gas. On the other hand, the share of electricity in the energy mix remains the same in the 2011–2016 period, approximately 40%.

3.1. Clustering results

Table 1 shows the Silhouette scores and Dunn indices for the number of clusters ranging between two and ten. Interestingly, for the two building typologies, both scores suggest the existence of two primary clusters in the energy performance time series data.

After fitting the K-Means model to the EUI time series, we obtain clusters with different temporal trends (Figs. 5 and 6). In both Office (Fig. 5) and Multifamily housing (Fig. 6), we observe two similar patterns: (a) a cluster of buildings increasing their median EUI levels (dashed purple colored time series), and (b) a second cluster showing a decreasing EUI trend (green colored time series).

Table 1
Silhouette score and Dunn index for various cluster values.

Number of clusters	Office		Multifamily housing	
	Silhouette score	Dunn index	Silhouette score	Dunn index
2	0.29	1.31	0.23	1.31
3	0.24	1.23	0.19	1.26
4	0.21	1.15	0.17	1.17
5	0.21	1.03	0.18	1.07
6	0.22	1.03	0.18	1.05
7	0.21	0.94	0.2	1.02
8	0.21	0.94	0.1	0.99
9	0.21	0.92	0.22	0.96
10	0.22	0.84	0.22	0.95

The two clusters are more balanced in the case of Multifamily housing, whereas in the Office case, the cluster with lower EUIs over time ($n = 210$) is larger than those that increased EUI ($n = 151$). To study the homogeneity in the data we plot the 68% (one standard deviation) and 95% (two standard deviations) confidence intervals around the time series. We observe that Office buildings show higher variance in EUI when compared to the Multifamily housing buildings. This finding can be attributed to the underlying factors driving energy consumption in commercial buildings, such as equipment intensity, space utilization, or occupancy patterns, that might fluctuate significantly across different properties. A closer look at the clustered time series plots shows variations in the gradient of changes in energy performance, between both building typologies and different clusters. In Fig. 5, we show that Office buildings in cluster 0 increase their median EUI by 13% in 2013 and up to 18% in 2015 and 2016, where the trend begins to plateau. On the other hand, buildings in cluster 1 reduce their median EUI by approximately 11% in 2013 compared to the base year of 2011, reaching 18% reduction in 2015. This result indicates a much more nuanced picture of building energy performance over time than what is presented by Meng et al. [8] or Palmer and Walls [24]. Here, we find significant divergence between certain buildings in their energy use trends, highlighting a non-uniform response to energy reporting and disclosure.

Looking at the absolute values of energy consumption within each cluster (Table 2), we see that Office buildings that reduced their energy use (cluster 1) accounts for a disproportionately larger share of energy consumption when compared to buildings in cluster 0. We can infer that Cluster 1 consists of larger properties in general, a fact that we statistically confirm in the following section. In the Multifamily housing building stock, the consumption between the two clusters is more balanced, similar to the EUI trends shown in Fig. 6. Overall, the aggregate energy use of buildings in the sample dropped by 0.51 billion kW hs over the study period, equivalent to a 3.5% reduction.

3.2. Cluster characteristics

After identifying clusters of buildings with similar energy performance behavior over time, we statistically test for differences in building characteristics between the two clusters. Table 3 summarizes the statistical tests' results for Office and Multifamily housing buildings. The statistical tests examine differences between ratios for categorical variables (Fisher's test) and distributions for continuous variables (Mann-Whitney's test).

The ratio of buildings with a heavy fuel oil boiler is found to be lower in the second (performance improving) cluster for both Multifamily housing and Office buildings (although very few office buildings use fuel oil for heat), showing correlations between compliance with laws requiring conversion from heavy fuel oils, as well as poor overall energy performance associated with this energy source. Also, the energy use metrics EUI and ENERGY STAR scores vary significantly between the clusters, since EUI values were used to perform the clustering. We find, though, that less-efficient buildings in 2011 were more likely to improve their performance over time. In the case of Office buildings, larger, newer, higher-value, and more heavily occupied buildings demonstrate improved performance over time. Buildings that ended up improving their energy performance had an approximately 30% higher EUI in 2011 than buildings that consumed more energy over the study period. Participation in NYC's Carbon Challenge program appears to boost energy performance in Office buildings as well, with cluster 1 having an approximately three times higher ratio of properties participating in the program. Conversely, larger and taller

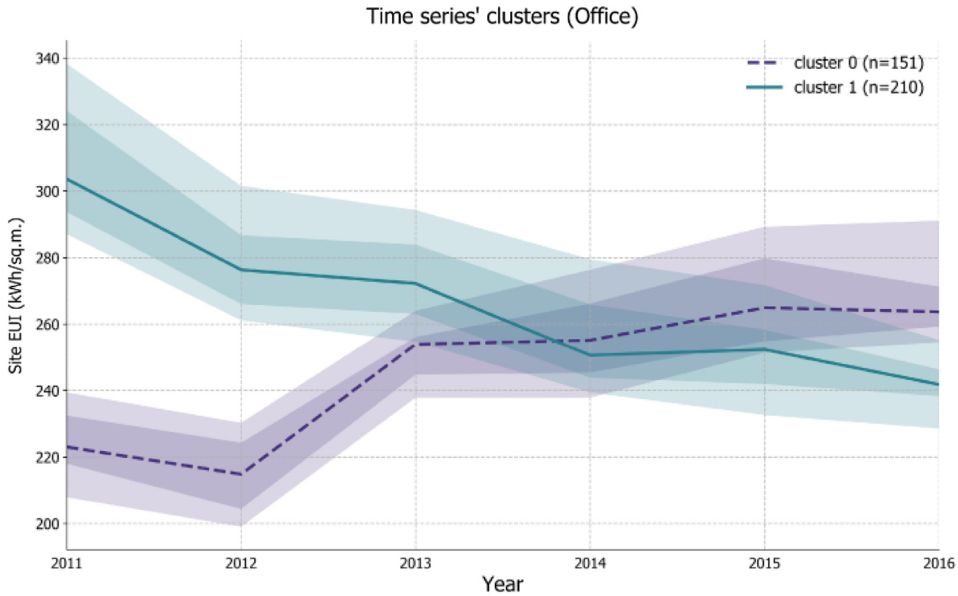

Fig. 5. Clustered time series (office buildings).

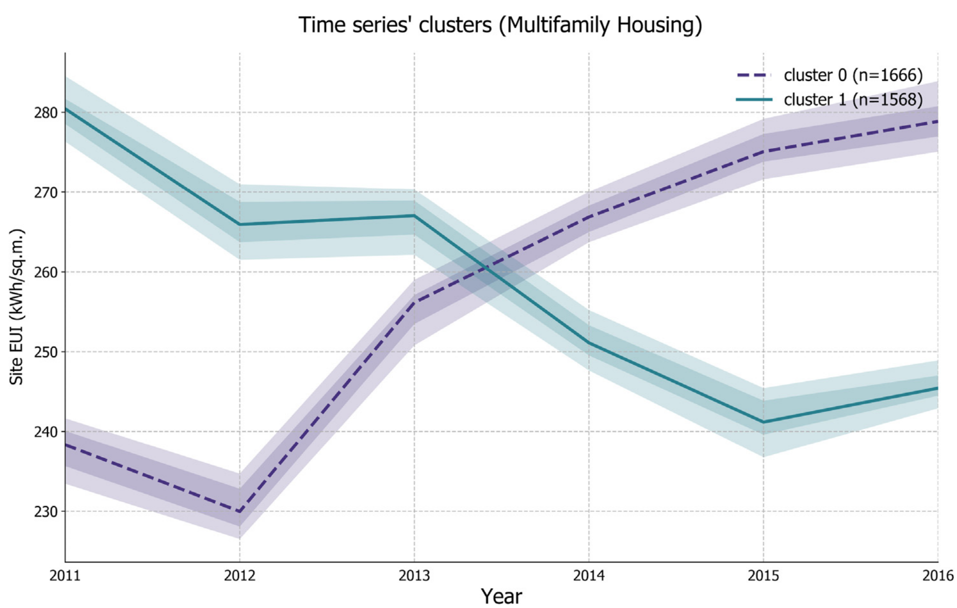

Fig. 6. Clustered time series (multifamily housing buildings).

Table 2
Inter-cluster difference in absolute energy consumption.

	Office		Multifamily housing	
	Cluster 0	Cluster 1	Cluster 0	Cluster 1
Total energy consumption 2011 (billion kW h)	1.12	3.27	4.81	5.62
Total energy consumption 2016 (billion kW h)	1.25	2.67	5.68	4.71
Difference 2011–2016 (billion kW h)	0.13	−0.60	0.87	−0.91

Multifamily housing buildings tend to perform worse over the six-year period. Multifamily housing buildings with higher unit density, on the other hand, show improved performance over time, indicating the effect of space utilization when managing a building's energy loads and the potential influence of shifts in tenant energy behavior. Additionally, if a residential building's data are reported by a large energy service provider (defined as top-5 market share holders), it is more likely to be assigned to cluster 0. Contrary to the expected finding, major energy service companies are not associated with energy use reductions over time in residential buildings. Finally, we do not observe the positive effect of Carbon Challenge participation for residential buildings, although multifamily properties are under-represented in the program.

Factors such as whether the building has been through an energy audit, is managed by top-5 real estate organization, or of the building is located outside Manhattan, do not vary significantly between clusters, for both building typologies, suggesting that these factors are not significant determinants of energy performance over time. These are important findings that will be discussed further in the next section.

3.3. Cluster assignment and influential variables

Table 4 shows the logistic regression results for Office and Multifamily housing buildings, using the backward feature elimination

Table 3
Statistical tests on cluster differences (office and multifamily housing buildings).

Binary features	Office			Multifamily housing		
	Cluster 0 ratio	Cluster 1 ratio	Fisher's test statistic	Cluster 0 ratio	Cluster 1 ratio	Fisher's test statistic
Audited under LL87	0.23	0.26	0.84	0.16	0.18	0.87
Has heavy oil boiler	0.32	0.18	2.11[b]	0.42	0.38	1.17[a]
Owner participates in NYC Carbon Challenge	0.08	0.21	0.32[b]	0.05	0.04	1.14
Not located in Manhattan	0.09	0.10	0.87	0.53	0.54	0.96
Managed by top-5 organizations	0.42	0.36	1.30	0.51	0.53	0.93
Reported by top-5 energy service providers	0.40	0.39	1.01	0.38	0.33	1.22[b]
Has data center space	0.03	0.06	0.45	–	–	–
Continuous/categorical features	**Cluster 0 median**	**Cluster 1 median**	**Mann-Whitney's test statistic**	**Cluster 0 median**	**Cluster 1 median**	**Mann-Whitney's test statistic**
Site EUI 2011 (kW h/sq.m.)	223	304	8647[b]	238	280	818,046[b]
Site EUI 2016 (kW h/sq.m.)	264	242	12479.5[b]	279	245	933,632[b]
ENERGYSTAR score 2011	75	63	12194.5[b]	–	–	–
ENERGYSTAR score 2016	72	79	12985.5[b]	54	70	1052606.5[b]
Total building area (sq.m.)	13,285	27,179	11460.5[b]	9002	8235	1237485.5[b]
Computer density 2011 (computers/100 sq.m.)	2.4	2.9	11,899[a]	–	–	–
Computer density 2016 (computers/100 sq.m.)	2.2	2.9	14633.5	–	–	–
Weekly operating hours 2011 (hr)	61.0	62.0	14786.5	–	–	–
Weekly operating hours 2016 (hr)	62.5	60.0	15,828	–	–	–
Worker density 2011 (workers/100 sq.m.)	2.5	2.8	12474.5[b]	–	–	–
Worker density 2016 (workers/100 sq.m.)	2.5	2.6	15,582	–	–	–
Laundry machine density 2011 (machines/100 sq.m.)	–	–	–	0.087	0.088	1,283,321
Laundry machine density 2016 (machines/100 sq.m.)	–	–	–	0.090	0.089	1233026.5[b]
Total number of residential units	–	–	–	81	80	1,286,651
Unit density 2011 (units/1000 sq.m.)	–	–	–	9.3	9.7	1192789.5[b]
Unit density 2016 (units/1000 sq.m.)	–	–	–	9.6	10.0	1,200,030[b]
Number of floors	16	20	12991.5[b]	7	6	1,261,674[a]
Year built	1926	1937	13152.5[b]	1940	1942	1,266,532
Asset value ($/sq.m.)	1061	1307	13,329[b]	399	371	1,271,667

[a] Significant at 95% confidence level (p-value < 0.05).
[b] Significant at 99% confidence level (p-value < 0.01).

algorithm described in the methods section.

As expected, variables that are found to be significant in the Fisher's and the Mann-Whitney's tests are included in the models for both building typologies. Based on the Wald statistic, the number of floors for Offices and the unit density for Multifamily housing are the variables contributing the most to the explanatory power of the respective regression models. Specifically, for each additional floor in an Office building, it is 3% more likely to belong in cluster 1 (i.e. improve its performance over time), whereas each unit increase in a Multifamily housing building's unit density (measured as units per 1000 sq.m.) increases its likelihood of being assigned to the energy-reducing cluster by 4.7%. Moreover, if an Office building participates in the NYC Carbon Challenge, it is 138% more likely to reduce its energy consumption over time. Similarly, an Office building located outside of the borough of Manhattan is 155% more likely to reduce its consumption. We approach this latter finding with caution, since the majority of commercial buildings are located in Manhattan and thus the class of buildings outside of the borough is underrepresented in the logistic regression model. For Multifamily housing properties, the existence of a heavy oil boiler or the use of a leading energy service provider for data reporting both reduce the probability of improving performance by approximately 16%. The association between carbon-intense energy systems,

Table 4
Logistic regression model (features obtained from backward elimination algorithm).

	Feature	β	Standard error (s.e.)	Wald statistic	e^{β}
Office	Intercept	−0.787[b]	0.294	7.14	0.455
	Owner participates in NYC Carbon Challenge	0.870[a]	0.360	5.823	2.387
	Not located in Manhattan	0.938[a]	0.407	5.300	2.555
	Number of floors	0.029[b]	0.010	8.544	1.029
	Asset value ($/sq.m.)	0.002	0.001	2.784	1.002
Multifamily housing	Intercept	−0.369[b]	0.113	10.732	0.691
	Reported by top-5 energy service provider	−0.172[a]	0.076	5.218	0.842
	Has heavy oil boiler	−0.188[a]	0.075	5.568	0.838
	Unit density (units/1000 sq.m.)	0.046[b]	0.010	20.936	1.047

[a] Significant at 95% confidence level (p-value < 0.05).
[b] Significant at 99% confidence level (p-value < 0.01).

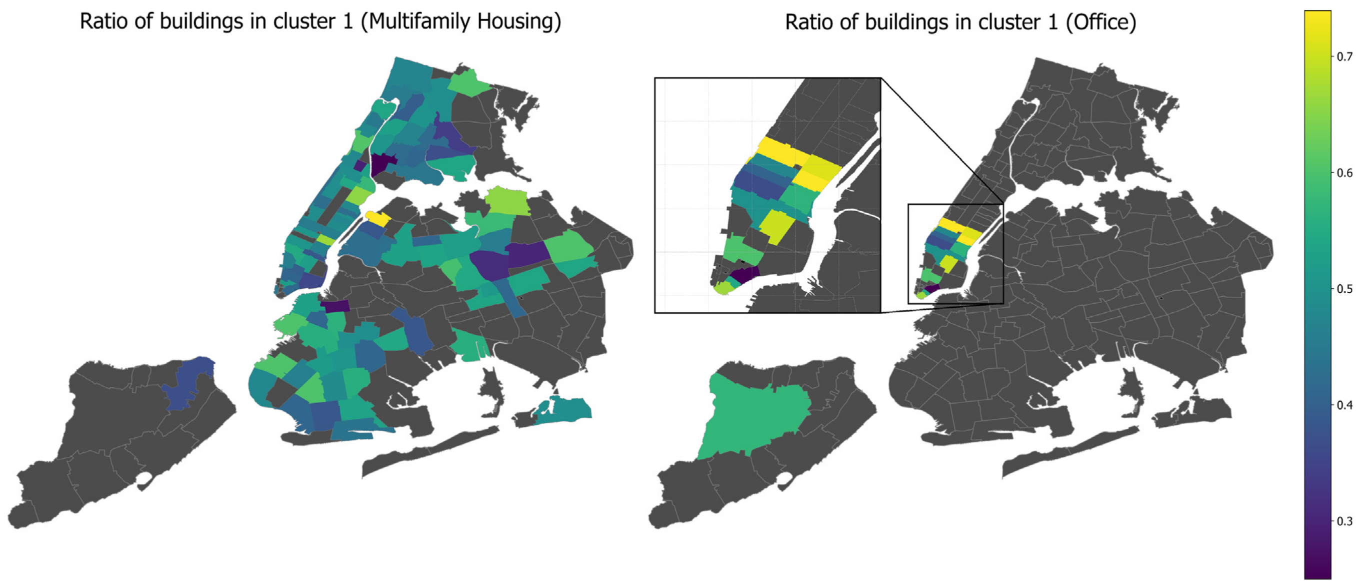

Fig. 7. Ratio of multifamily housing (left) and office buildings assigned to cluster 1 over total number of properties; grouped by zip code. Grey areas represent zip codes with less than ten properties.

such as heavy fuel oil boilers, and poor energy performance over time highlights the importance of environmental awareness on energy and carbon reduction efforts.

3.4. Spatial analysis

In this last part of the Results section, we explore the spatial dimension of energy performance over time. We spatially aggregate the clustered buildings to the zip code level, and calculate the ratio of buildings assigned to cluster 1 (i.e. "good performers") over the total number of clustered properties for each zip code. We decide to study only zip codes including at least five buildings from our model, to avoid potential bias (Fig. 7).

Visual inspection of the choropleths in Fig. 7 reveals a rather weak relationship between nearby zip codes and their buildings' energy performance over time. Moran's I index values are: 0 (*p-value: 0.90*) for Multifamily housing, and -0.03 (*p-value: 0.86*) for Office buildings. The high p-values for both building typologies suggest spatial randomness in the distribution of buildings' energy performance over time.

4. Discussion

Our results suggest two distinct temporal patterns of energy performance for commercial and residential buildings, characterized by energy use reductions and increases. This finding highlights the differential response to energy reporting and disclosure, and presents a more complex picture of energy use dynamics over time when compared to previous studies. While a sub-group of buildings do improve their performance, these tend to be poorly-performing buildings at the outset, with EUIs well above the means of the sample. This sub-group, then, may be motivated by (1) the avoidance of being viewed as an energy inefficient building when compared with peers or (2) excessive risk exposure to future possible energy standards or mandates based on building performance. It is also quite possible that this sub-group of buildings was simply unaware of their energy performance, or the importance of tracking energy use, and responded to the disclosure of its energy data by reducing consumption. However, for Office buildings that demonstrated the most substantial energy performance improvement, this assumption is weakened by the fact that both clusters had similar proportions of buildings managed by a major energy services firm. One would have assumed that larger and more capable management firms would lead to more efficient building operations, or the ability to identify and implement energy conservation measures. The finding of two distinct clusters for each property type also suggest that energy reporting is useful for spurring behavioral changes that address the "low-hanging fruit" of energy efficiency investments. Since the buildings that improved their energy performance over time had higher initial EUIs, and thus were under-performing in the early years of the reporting mandate, it is conceivable that tracking their energy use initiated more modest changes that resulted in reverting their EUI values toward the mean.

It is important to emphasize that energy disclosure policies do not mandate change; rather, they are a mechanism to shift market behavior and decision-making through the reduction of information asymmetries [7,10]. As evidenced here, cities relying solely on market-driven change will face non-trivial barriers to large-scale energy efficiency gains. While voluntary schemes, such as NYC Carbon Challenge, may boost energy performance in commercial buildings, their effect is not found to be significant for residential buildings, which hold the largest share of the City's building stock. On the other hand, the heavy fuel oil phase-out regulation had substantial impacts on energy use. The suggestion here is that policies predicated on energy reporting, without a mandate to improve performance, will incentivize certain buildings to modestly reduce their energy consumption, while having no effect on others.

What is a particularly surprising finding is the sub-group of buildings that actually increased their energy consumption over time. Despite a policy landscape increasingly focused on energy efficiency and carbon reductions, these buildings – generally older, smaller, and lower quality, but with better initial energy performance – used 18% more energy, on average, when comparing 2016 to 2011 data. One possible explanation is that these buildings were identified as "good" performers in 2011, and thus experienced less pressure to improve over time, or assumed that no more action was needed. Future work should focus on analyzing which part of this change is attributed to actual energy inefficiency and which to changes in underlying factors driving energy use (e.g. occupancy density and operating hours).

The implications of our study are of direct relevance to urban energy policy. While market-driven strategies, such as disclosure, may be effective in reducing consumption in some buildings, they do not have a consistent impact across the entire building stock. In fact, we see that disclosure may have a counter-intuitive impact for a sub-group of buildings, particularly those that are initially deemed to be performing well. This suggests that a comprehensive urban energy policy must couple market-driven, information-based tools with regulations requiring improvements over time. Our methodology also provides a tool for targeted policy interventions that can be applied more efficiently and equitably than universal policies that may have unexpected consequences, or create unnecessary hardships on building owners that would have improved their performance regardless of the regulatory intervention.

Before concluding our analysis, we present recommendations for the disclosed energy data collection process that could support further research. Due to the self-reported nature of the energy data, we often encountered entry errors in both energy consumption and building characteristics (e.g floor area), raising data quality and reliability concerns. We argue that more clear documentation in the energy reporting tools or data quality checks in the users' inputs would minimize the misreported data entries. Also, more granular inputs, such as tenant-level energy or occupancy data as well as monthly energy consumption reporting could help us better understand energy consumption patterns in individual buildings. Our last suggestion would be to track energy conservation measures implemented in buildings that are reporting energy use data each year. This would allow us to go further in exploring the causes of observed increases/decreases in building energy performance.

5. Conclusion

Cities across the globe are turning to information disclosure as a means to understand and reduce energy use and carbon emissions. In first-mover cities, such as NYC, disclosure policies have created cross-sectional time series energy data streams, presenting an unprecedented opportunity to study the evolution of building energy performance over time.

The analysis presented in this work identifies buildings with similar temporal energy performance patterns, as well as shared characteristics among them. Using a K-means clustering algorithm, we reveal two distinct clusters of "improving" and "declining" energy performance over time, for both commercial and residential buildings in NYC. We employ Fisher's and Mann-Whitney's statistical tests to test for significantly different characteristics between the clusters. We find that energy reductions are mostly driven by Office buildings, with larger, newer, and higher-value buildings showing significant improvement in terms of energy use intensity between 2011 and 2016. From our logistic regression model, office buildings that participate in the NYC Carbon Challenge program are 138% more likely to have improved their performance over the study period. Similarly, residential buildings with heavy oil boilers are 16% more likely to have increasing EUIs over time.

Overall, we demonstrate that although disclosure might lead to better energy performance in some buildings, its effect is not consistent across the entire NYC building stock. In order to realize significant energy use improvements over time and reach energy and carbon reduction goals, cities need to design and implement comprehensive policy frameworks, bringing together information transparency and reporting with targeted mandates and incentives.

Our work provides an important foundational analysis of the patterns of building energy performance over time. Future research should explore more deeply the factors that drive energy use changes and determine the response to energy disclosure, whether they are social-behavioral (e.g. management quality, occupant demographics), physical (e.g. individual building systems) or economic (e.g. firm size and revenue for Office buildings or household income for residential tenants). Moreover, our methodology can be used as the basis of targeted and more equitable energy policy, allowing cities to allocate limited resources more efficiently in their attempt to reduce energy use and carbon emissions. Finally, with energy disclosure policies being widely adopted, our analytical framework can be applied on other cities' data sets, as well. Although our findings are representative of NYC specifically, it would be interesting for future studies to analyze similar energy performance patterns in cities with different characteristics. These characteristics could vary from urban morphology and location to existing energy policy frameworks and carbon reduction goals. This would enable policy makers assess drivers of energy performance not only in local, but in national or global level.

Acknowledgment

The authors would like to thank the New York City Mayor's Office of Sustainability for their contribution of data for this research, as well as important feedback and expertise. This material is based upon work supported by the National Science Foundation under Grant No. 1653772. All errors remain the responsibility of the authors.

References

[1] Pachauri RK, Allen MR, Barros VR, Broome J, Cramer W, Christ R, et al. Climate change 2014: synthesis report. Contribution of working groups I, II and III to the fifth assessment report of the intergovernmental panel on climate change, IPCC; 2014.
[2] Zhao H-X, Magoulès F. A review on the prediction of building energy consumption. Renew Sustain Energy Rev 2012;16:3586–92.
[3] Pérez-Lombard L, Ortiz J, Pout C. A review on buildings energy consumption information. Energy Build 2008;40:394–8.
[4] Rosenzweig C, Solecki W, Hammer SA, Mehrotra S. Cities lead the way in climate-change action. Nature 2010;467:909–11.
[5] Wright D, Leigh R, Kleinberg J, Abbott K, et al. New York city can eliminate the carbon footprint of its buildings by 2050. Energy Sustain Dev 2014;23:46–58.
[6] Ma Z, Cooper P, Daly D, Ledo L. Existing building retrofits: methodology and state-of-the-art. Energy Build 2012;55:889–902.
[7] Kontokosta CE. Energy disclosure, market behavior, and the building data ecosystem. Ann New York Acad Sci 2013;1295:34–43.
[8] Meng T, Hsu D, Han A. Estimating energy savings from benchmarking policies in New York city. Energy 2017;133:415–23.
[9] Burr AC, Keicher C, Leipziger D. Building energy transparency: a framework for implementing us commercial energy rating and disclosure policy. Washington (DC): Institute for Market Transformation; 2011.
[10] Palmer K, Walls M. Using information to close the energy efficiency gap: a review of benchmarking and disclosure ordinances. Energy Efficiency 2017;10:673–91.
[11] City of New York. Mayor Bloomberg and Commissioner Hirst announce that every major City-owned building has been benchmarked for energy use. < http://www1.nyc.gov/office-of-the-mayor/news/192-10/mayor-bloomberg-commissioner-hirst-that-every-major-city-owned-building-has-been > ; 2010.
[12] Pérez-Lombard L, Ortiz J, González R, Maestre IR. A review of benchmarking, rating and labelling concepts within the framework of building energy certification schemes. Energy Build 2009;41:272–8.
[13] Palmer K, Walls M. Can benchmarking and disclosure laws provide incentives for energy efficiency improvements in buildings? Resour Fut Disc Pap 2015;1:5–9.
[14] Kontokosta CE, Tull C. A data-driven predictive model of city-scale energy use in buildings. Appl Energy 2017;197:303–17.
[15] Robinson C, Dilkina B, Hubbs J, Zhang W, Guhathakurta S, Brown MA, et al. Machine learning approaches for estimating commercial building energy consumption. Appl Energy 2017;208:889–904.
[16] Kontokosta CE. A market-specific methodology for a commercial building energy performance index. J Real Estate Financ Econ 2015;51:288–316.
[17] Santamouris M, Mihalakakou G, Patargias P, Gaitani N, Sfakianaki K, Papaglastra M, et al. Using intelligent clustering techniques to classify the energy performance of school buildings. Energy Build 2007;39:45–51.
[18] Nikolaou T, Kolokotsa D, Stavrakakis G. Review on methodologies for energy benchmarking, rating and classification of buildings. Adv Build Energy Res 2011;5:53–70.
[19] Urban Green Council. The New York City energy and water use 2017 report. < https://urbangreencouncil.org/content/projects/new-york-city-energy-and-water-use-2017-report?_ga=2.205398292.548664148.1511817700-1719214256.1509306203 > ; 2017.
[20] Reina VJ, Kontokosta C. Low hanging fruit? Regulations and energy efficiency in subsidized multifamily housing. Energy Policy 2017;106:505–13.
[21] Ma J, Cheng JC. Identifying the influential features on the regional energy use intensity of residential buildings based on random forests. Appl Energy 2016;183:193–201.
[22] Kontokosta CE, Jain RK. Modeling the determinants of large-scale building water use: implications for data-driven urban sustainability policy. Sustain Cities Soc 2015;18:44–55.
[23] Hsu D. Comparison of integrated clustering methods for accurate and stable prediction of building energy consumption data. Appl Energy 2015;160:153–63.
[24] Palmer K, Walls M. Does information provision shrink the energy efficiency gap? A cross-city comparison of energy benchmarking and disclosure laws. Technical report. Discussion paper 15–12. Washington (DC): Resources for the Future; 2015.
[25] Kontokosta CE, Tull C. Energyviz: web-based eco-visualization of urban energy use

from building benchmarking data. In: Proceedings of the international conference on computing in civil and building engineering.
[26] City of Philadelphia. Building energy benchmarking. < http://www.phillybuildingbenchmarking.com/ > ; 2017.
[27] Papadopoulos S, Kontokosta CE. Big buildings and big data: do energy disclosure policies impact energy use over time? In: ASCE international workshop on computing in civil engineering. p. 248–55.
[28] Hart Z. The benefits of benchmarking building performance. Washington: Institute for Market Transformation and Pacific Coast Collaborative; 2015.
[29] NYC Mayor's Office of Sustainability. NYC benchmarking law. < http://www.nyc.gov/html/gbee/html/plan/ll84.shtml > ; 2017.
[30] Kontokosta C, Bonczak B, Duer-Balkind M. DataIQ – a machine learning approach to anomaly detection for energy performance data quality and reliability. In: Proceedings of the ACEEE.
[31] NYC Department of City Planning. Property land use tax lot output (PLUTO) data. < https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page > ; 2017.
[32] NYC Mayor's Office of Sustainability. LL87: energy audits & retro-commissioning. < http://www.nyc.gov/html/gbee/html/plan/ll87.shtml > ; 2017.
[33] NYC Open Data. Oil boilers – detailed fuel consumption and building data. < https://data.cityofnewyork.us/Housing-Development/Oil-Boilers-Detailed-Fuel-Consumption-and-Building/jfzu-yy6n > ; 2017.
[34] NYC Mayor's Office of Sustainability. The New York City carbon challenge. < http://www.nyc.gov/html/gbee/html/challenge/nyc-carbon-challenge.shtml > ; 2017.
[35] Al Shalabi L, Shaaban Z, Kasasbeh B. Data mining: a preprocessing engine. J Comput Sci 2006;2:735–9.
[36] Aksoy S, Haralick RM. Feature normalization and likelihood-based similarity measures for image retrieval. Pattern Recogn Lett 2001;22:563–82.
[37] Bishop CM. Pattern recognition and machine learning. Springer; 2006.
[38] Jain AK. Data clustering: 50 years beyond k-means. Pattern Recogn Lett 2010;31:651–66.
[39] Aghabozorgi S, Shirkhorshidi AS, Wah TY. Time-series clustering – a decade review. Inform Syst 2015;53:16–38.
[40] Liao TW. Clustering of time series data – a survey. Pattern Recogn 2005;38:1857–74.
[41] Bagnall A, Janacek G. Clustering time series with clipped data. Mach Learn 2005;58:151–78.
[42] Vlachos M, Lin J, Keogh E. A wavelet-based anytime algorithm for k-means clustering of time series. In: Proc workshop on clustering; 2003. p. 23–30.
[43] Yu Z, Fung BCM, Haghighat F, Yoshino H, Morofsky E. A systematic procedure to study the influence of occupant behavior on building energy consumption. Energy Build 2011;43:1409–17.
[44] Räsänen T, Voukantsis D, Niska H, Karatzas K, Kolehmainen M. Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. Appl Energy 2010;87:3538–45.
[45] Gaitani N, Lehmann C, Santamouris M, Mihalakakou G, Patargias P. Using principal component and cluster analysis in the heating evaluation of the school building sector. Appl Energy 2010;87:2079–86.
[46] Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987;20:53–65.
[47] Dunn JC. Well-separated clusters and optimal fuzzy partitions. J Cybernet 1974;4:95–104.
[48] Kontokosta CE. Modeling the energy retrofit decision in commercial office buildings. Energy Build 2016;131:1–20.
[49] Nachar N, et al. The Mann-Whitney U: a test for assessing whether two independent samples come from the same distribution. Tutorials Quant Methods Psychol 2008;4:13–20.
[50] Routledge R. Fisher's exact test. Encyclopedia Biostat 2005.
[51] Goodchild MF. Spatial autocorrelation, vol. 47. Geo Books; 1986.
[52] Friedman J, Hastie T, Tibshirani R. The elements of statistical learning, vol. 1. Springer series in statistics New York; 2001.
[53] Harrell Jr. FE. Binary logistic regression. Regression modeling strategies. Springer; 2015. p. 219–74.
[54] NYC Clean Heat. NYC Clean Heat program regulations. < https://www.nyccleanheat.org/content/regulations > ; 2017.