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
0306-2619/

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