Applied Energy 197 (2017) 303–317

Contents lists available at ScienceDirect

Applied Energy

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

A data-driven predictive model of city-scale energy use in buildings

Constantine E. Kontokosta [a,*], Christopher Tull [b]

[a] Center for Urban Science and Progress & Tandon School of Engineering, New York University, 1 Metrotech Center, 19th Floor, Brooklyn, NY 11201, United States
[b] Center for Urban Science and Progress, New York University, 1 Metrotech Center, 19th Floor, Brooklyn, NY 11201, United States

HIGHLIGHTS

- This paper provides insight into urban energy dynamics.
- Machine learning is used to predict building energy use at the city scale.
- Actual energy use data for more than 20,000 buildings is used.
- Energy use intensity is predicted for all 1.1 million buildings in New York City.
- Predictions are validated using actual building and zip code level energy data.

ARTICLE INFO

Article history:
Received 21 November 2016
Received in revised form 25 March 2017
Accepted 1 April 2017

Keywords:
Machine learning
Building energy
Energy efficiency
Urban dynamics
Energy prediction

ABSTRACT

Many cities across the United States have turned to building energy disclosure (or benchmarking) laws to encourage transparency in energy efficiency markets and to support sustainability and carbon reduction plans. In addition to direct peer-to-peer comparisons, the benchmarking data published under these laws have been used as a tool by researchers and policy-makers to study the distribution and determinants of energy use in large buildings. However, these policies only cover a small subset of the building stock in a given city, and thus capture only a fraction of energy use at the urban scale. To overcome this limitation, we develop a predictive model of energy use at the building, district, and city scales using training data from energy disclosure policies and predictors from widely-available property and zoning information. We use statistical models to predict the energy use of 1.1 million buildings in New York City using the physical, spatial, and energy use attributes of a subset derived from 23,000 buildings required to report energy use data each year. Linear regression (OLS), random forest, and support vector regression (SVM) algorithms are fit to the city's energy benchmarking data and then used to predict electricity and natural gas use for every property in the city. Model accuracy is assessed and validated at the building level and zip code level using actual consumption data from calendar year 2014. We find the OLS model performs best when generalizing to the City as a whole, and SVM results in the lowest mean absolute error for predicting energy use within the LL84 sample. Our median predicted electric energy use intensity for office buildings is 71.2 kbtu/sf and for residential buildings is 31.2 kbtu/sf with mean absolute log accuracy ratio of 0.17. Building age is found to be a significant predictor of energy use, with newer buildings (particularly those built since 1991) found to have higher consumption levels than those constructed before 1930. We also find higher electric consumption in office and retail buildings, although the sign is reversed for natural gas. In general, larger buildings use less energy per square foot, while taller buildings with more stories, controlling for floor area, use more energy per square foot. Attached buildings – those with adjacent buildings and a shared party wall – are found to have lower natural gas use intensity. The results demonstrate that electricity consumption can be reliably predicted using actual data from a relatively small subset of buildings, while natural gas use presents a more complicated problem given the bimodal distribution of consumption and infrastructure availability.

1. Introduction

Cities are increasingly adopting long-term sustainability plans designed to increase the efficiency of energy infrastructure, reduce operating costs, and mitigate the negative effects of climate change

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

http://dx.doi.org/10.1016/j.apenergy.2017.04.005
0306-2619/

[1–4]. As buildings account for a majority of primary energy use and carbon emissions in dense urban areas, these plans tend to focus on the "greening" of existing buildings as a path to greater resource efficiency. One of the most notable urban policy innovations for buildings has been energy disclosure. These laws require annual energy use reporting for a subset of buildings in a city's inventory, typically those larger than a particular size threshold. In New York City (NYC), Local Law 84 covers approximately 20,000 buildings larger than 50,000 square feet each year. While this is a significant sample that accounts for approximately 45 percent of the City's total energy consumption, it represents only 2 percent of the NYC building stock ([5–8]). Beginning in 2018, this mandate will expand to cover all buildings over 25,000 square feet, similar to disclosure policies in other U.S. cities. There are legitimate political, financial, and privacy concerns that constrain the expansion of these laws to smaller buildings, particularly driven by the potential cost to building owners. Given this reality, it is imperative for policy-makers tasked with reducing city-wide energy use and carbon emissions to have alternative, but reliable, methods to understand energy consumption across multiple spatial scales.

This study develops a predictive model of energy use at the building, district, and city scales using training data from energy disclosure policies and predictors from widely-available property and zoning information. For city planners and energy policymakers, understanding energy use dynamics is critical to (1) knowing where and how energy is being consumed across the morphologic and socioeconomic contours of the city, (2) providing situational awareness of energy use to better allocate resources and target policy interventions to reduce consumption, and (3) identifying cost-efficient savings opportunities across the city. In addition to developing a time-invariant snapshot of consumption patterns, reliable energy use predictions can inform forecasting and policy scenario modeling over time. From an energy management perspective, a building-specific, city-wide energy profile provides opportunities for efficiencies through tracking, benchmarking, and impact evaluation of new programs. For planners, the ability to understand future energy demand – given expectations for land use changes, urban development, and other technological, architectural, and behavioral factors that could alter future energy use patterns – provides an important yardstick by which to evaluate policy alternatives, set goals, and measure progress.

In this paper, we evaluate several prediction algorithms, including ordinary least squares regression, support vector regression, and random forest, and two feature selection approaches to predict building-specific annual energy use and energy use intensity (EUI) from existing property and land use administrative records. We combine actual energy use and building attribute data from NYC's Local Law 84 (LL84) for 20,652 buildings with the NYC Department of City Planning's Primary Land Use Tax Output (PLUTO) database, covering all 1,082,437 properties in NYC, to predict in-sample and out-of-sample building energy use, and validate our prediction model output against actual zip code-level energy use data provided by New York City's energy utilities, ConEdison and National Grid. We begin by presenting the relevant literature and describing our data and data cleaning process. We then discuss our methodology and model results, and conclude with an exploration of the implications and significance of our findings for urban energy analytics and energy policy.

2. Literature review

Despite the importance of cities and buildings to the reduction of global energy use and carbon emissions, relatively little attention has been paid to the data-driven study of urban energy dynamics. The traditional focus of demand-side energy studies has been on building simulation and systems-level physical models of building technologies and components, rather than city-scale empirical models informed by data on actual consumption patterns. For instance, Dhakal [9] examines urban energy consumption in Chinese cities using a coarse estimate of energy use in urban areas derived from the energy intensity of economic activities (measured by Gross Regional Product). The study found that cities account for 84% of China's total commercial energy use, and that the 35 largest cities account for approximately 40% of national consumption and carbon emissions. Bennett and Newborough [10] present a framework for auditing citywide energy use. They highlight the importance of the city scale to identifying and promulgating carbon emission reduction strategies. The authors point out the need for various types of data, including surveys and direct monitoring, but do not provide any empirical examples of the citywide audit in practice. The paper highlights the need for such city-scale modeling to predict energy efficiency potential across urban sectors.

In a paper by Lin et al. [11], the authors utilize a LEAP energy simulation model to estimate future energy demand under various policy scenarios. The model applies demographic, socioeconomic, and macroeconomic variables for the city of Xiamen to forecast energy demand across the city over time. The authors rely on coarse consumption data that do not differentiate among building-specific characteristics beyond type of use. Although distinguished from the work of Bennett and Newborough [10] by the recommended data collection and by the fact that Lin et al. attempt to estimate energy use in Xiamen, both approaches rely on relatively low spatial resolution estimates of consumption predicated on numerous assumptions about the patterns and drivers of energy use. In another attempt to develop an urban energy demand model, Brownsword et al. [12] utilize a linear programming approach to estimate current and future energy use. The authors apply a broad typology to urban energy "consumers", separating the city into domestic, commercial, and industrial uses across size bins of small, medium, and large. While the authors stated goal is the replicability of the model, this approach significantly dismisses the variation in urban land use and building typologies and the impact of such characteristics on energy use (see [8,13]). Such assumptions limit the usefulness of an energy model to identify and predict future energy savings measures at scale.

Heiple and Sailor [14] estimate daily property-level energy consumption levels using annual building simulation results for prototypical buildings in the city of Houston, Texas. The authors apply energy simulation outputs from building prototypes and match these prototypes to existing buildings in the city using GIS-based tax lot data. The study attempts to scale up building-level estimations to the entire city. However, the use of prototypical buildings introduces significant error in the prediction, coupled with the uncertainty in matching prototype buildings to existing buildings based on sparse tax lot data. The limited data on actual energy use at the building level and building characteristics constrain the predictive power of the citywide model. Facing similar data constraints, Touchie et al. [15] attempt to identify building characteristics that influence energy consumption in residential buildings. Their findings are significantly limited by a sample size of only approximately 60 buildings, and they rely on multiple data providers to collect building-level characteristics. The limitations of the Touchie et al. study highlight the value of building energy disclosure data given the relatively large sample of buildings covered (particularly in New York City) and the richness of the building features included in the datasets.

Kavgic et al. [16] model the energy use in residential buildings in the context of Belgrade, Serbia. The authors highlight the limitations of city-scale energy models derived from the extrapolation of archetypical building profiles to an entire urban building stock.