Sustainable Cities and Society 88 (2023) 104250

Contents lists available at ScienceDirect

Sustainable Cities and Society

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

Analysis of single- and multi-family residential electricity consumption in a large urban environment: Evidence from Chicago, IL

Jorge E. Pesantez [a,b,*], Grace E. Wackerman [c], Ashlynn S. Stillwell [a]

[a] Civil and Environmental Engineering, University of Illinois Urbana-Champaign, 205 North Mathews Avenue, MC-250, Urbana, IL, 61801-2350, USA
[b] Department of Civil and Geomatics Engineering, California State University, Fresno, 2320 E. San Ramon Ave. M/S EE94, Fresno, CA, 93740-0001, USA
[c] Electrical and Computer Engineering, University of Illinois Urbana-Champaign, 306 North Wright St, MC-702, Urbana, IL, 61801-2918, USA

ARTICLE INFO

Dataset link: https://stillwell.cee.illinois.edu/data/

Keywords:
Demand management
Socioeconomic predictors
Electricity usage
Urban environment

ABSTRACT

Natural and human-caused extreme events can alter residential electricity demand in urban areas and stress the electricity grid, with different types of residential electricity consumers exhibiting different consumption patterns. Residential electricity demands have been widely analyzed considering single-family consumers; however, multi-family consumption patterns remain comparatively understudied. The deployment of smart electricity meters enables the identification of single- and multi-family residential electricity consumption patterns at high temporal resolution. Using smart electricity meter data for the greater Chicago area, we compare electricity demand profiles reported by smart meters from single- and multi-family consumers in a large and diverse urban environment to understand residential electricity patterns better. Our study comprehensively analyzes the daily electricity demand profiles of these two types of residential consumers to identify peak electricity consumption times and magnitudes. Results show that the electricity demand of both residential end-users follows similar time of use patterns, and single-family users approximately double the demand of multi-family users on a per household basis. We also present predictive models of the electricity demand with socioeconomic data at the zip code level. Predictive model results show that multiple linear regression models explain up to 62% and 41% of the mean daily electricity (MDE) demand of single- and multi-family users, respectively. The median age of occupants, percent age 65 and older, mean commute time, and percent high school or higher education are statistically significant predictors of the MDE demand of single-family users, with percent high school or higher education having the highest relative importance. Similarly, median building age, percent multi-family, percent female, median age of occupants, and mean commute time are statistically significant predictors of multi-family electricity consumption, with median age of occupants having the highest relative importance. Modeling electricity demand to uncover differences between single- and multi-family residential electricity demands can assist city planners and utility managers to develop tailored demand management strategies.

1. Introduction

Electricity demand has increased over time in most urban areas and this trend is likely to continue as cities are projected to be home to 70% of the worldwide population by 2050 (Chen, Ban-Weiss, & Sanders, 2020). Residential customers are an important part of the total electricity demand in the United States, representing nearly 40% of total consumption (US EPA, 2021). The demand exerted by the residential sector is expected to increase due to several factors, including the need for more space conditioning (e.g., air conditioning, heating) and the adoption of electrically-powered technologies (e.g., electric vehicles). In most cities, the residential sector is comprised of single- and multi-family electricity users. Single-family refers to one account from one building and a single consumption meter device. Multi-family refers to multiple accounts from one building (e.g., apartments, flats, lofts, etc.) and their corresponding meter devices. Multi-family users are crucial residential end-users, especially in areas of high population density, where they account for up to 90% of electricity consumption (Jain, Smith, Culligan, & Taylor, 2014). While extensive research has been conducted on the effects of weather and time of consumption for single-family electricity demand (Alberini, Prettico, Shen, & Torriti, 2019; Beccali, Cellura, Brano, & Marvuglia, 2008; Liddle & Huntington, 2021; Sandels, Widén, Nordström, & Andersson, 2015), there is a limited number of studies on how social and economic factors affect the electricity usage of both the single- and multi-family residential end-uses.

* Corresponding author.
E-mail address: jpesantez@mail.fresnostate.edu (J.E. Pesantez).

https://doi.org/10.1016/j.scs.2022.104250
Received 8 June 2022; Received in revised form 9 October 2022; Accepted 10 October 2022
Available online 11 November 2022

Acronyms

ACS:	American Community Survey
CAR:	Companion to Applied Regression
ComEd:	Commonwealth Edison Electric Company
E.U.:	European Union
GHG:	Greenhouse gas
HS:	High school degree
kWh:	Kilowatt-hour
MDE:	Mean Daily Electricity
MLR:	Multiple Linear Regression
OLS:	Ordinary Least Squares
Pr:	Probability
USA:	United States of America
VIF:	Variance Inflation Factor

While single-family users can constitute the predominant residential end-use in the United States and in many parts of the world, the population living in apartment buildings, condominiums, and other multi-family residences represents an important component in highly dense urban environments. Reports on housing characteristics show that 47% of the U.S. population lives in apartments in buildings with 5 or more units (NMHC, 2019). In Europe, 46% of the E.U. population lived in apartments in 2018 (Eurostat, 2020). Throughout the residential sector, the widespread use of smart meters in the energy sector has provided essential tools to understand electricity consumption at high temporal and spatial resolution (Avancini, et al., 2019), including labeling differences between single- and multi-family residences.

Data analysis of smart metered electricity demand presents different challenges and opportunities. For example, as more utilities use smart meters, the need for implementing data analytics increases for electricity companies (vom Scheidt, et al., 2020). Also, using big data analysis can uncover insights about electricity consumption (Yildiz, Bilbao, Dore, & Sproul, 2017). Aggregating data to hourly and daily time steps enables the graphical representation of electricity profiles to analyze trends and peaks of electricity demand during the day, week, and across different areas of a city (Fan, MacGill, & Sproul, 2017). Similarly, aggregating electricity demand data from the user level to the zip code level enables the use of publicly available socioeconomic factors from Census data to understand how these demands vary across different geographical areas over time (Elnakat, Gomez, & Booth, 2016).

Modeling electricity demand at the user level provides valuable insights to utility managers to predict future consumption, develop demand response programs, and support operational planning. However, data collection for these high spatial resolution models is challenging due to security, privacy, and lack of availability (Zechman Berglund, et al., 2021). Hence, developing models with readily available predictors aggregated at spatial resolution levels that enhance privacy and security remains an open research question. Furthermore, modeling electricity demand with socioeconomic predictors at the zip code level can assist city planners and utility managers in developing tailored demand management strategies to provide safe and affordable basic electricity service. From a customer perspective, descriptive and predictive models of electricity demand may guide how to use energy more efficiently to decrease bills and work towards one of the Sustainable Development Goals (SDGs) to ensure access to affordable, reliable, and modern energy services (Clark & Wu, 2016).

Electricity demand varies across different geographical areas in cities around the world (Huang, 2015; Kavousian, Rajagopal, & Fischer, 2013; Wyatt, 2013). However, this variation has been studied using billing records with aggregated temporal resolution, such as monthly and even annual time steps. Analysis of the effects of socioeconomic variables and daily electricity demand measured by smart meters can provide new strategic tools to understand how electricity is consumed across a city. Socioeconomic variables, including income, percentage of occupancy, and building age, among others, affect the daily electricity demand of both single- and multi-family users; however, these links between socioeconomic conditions and electricity demand remain underquantified. By analyzing the effect of these variables on the electricity demand of single- and multi-family residential users, a better understanding of how electricity is consumed (e.g., daily peak magnitude and time) across different geographical areas is achievable.

The objectives of this work are to describe graphically and quantitatively the electricity demand exerted by the two types of residential users, single- and multi-family accounts, in a large, diverse urban environment. This work also quantifies how much of the mean daily electricity (MDE) demand over a year reported at the zip code level can be explained by publicly available socioeconomic variables retrieved from the American Community Survey (ACS). Finally, this study builds predictive models and identifies the most important predictors of the MDE demand.

This study analyzes the electricity demand reported by residential users to improve the understanding of how electricity is used across a large, diverse urban environment such as the City of Chicago. Among others, the novelty of this work relies on the study of electricity demand divided into two types of residential users: single- and multi-family accounts. This study also reports the per capita residential electricity demand at the zip code level, which differs from previous works that reported these values at the city or census block levels. Finally, the model uses publicly available socioeconomic variables from the ACS and investigates the importance of each predictor in the predictive capabilities of the mean daily electricity demand during 2016. In this work, we examine smart electricity meter data from January 1st to December 31st of 2016 reported at a 30-minute resolution at the account level for Chicago, Illinois, USA. Chicago was selected to apply our analysis to a large, diverse urban environment with a comparable percentage of single- and multi-family accounts across the city. Furthermore, the selection of Chicago as the project's test bed was mainly driven by data availability. We use smart meter data, in conjunction with socioeconomic data (U.S. Census Bureau, 2016), to develop daily electricity profiles for single- and multi-family residential users, revealing electricity consumption trends. We leverage the high temporal resolution of smart meters (e.g., 30 min) to determine electricity demand from both types of residential users under different circumstances like typical work days, weekends, and daily peak and non-peak times. From the demand profiles, we analyze differences in electricity consumption across zip codes and predict electricity consumption for single- and multi-family accounts as functions of socioeconomic data. We aggregated electricity demand reported at the anonymized customer level to the zip code level to analyze electricity demand while maintaining anonymity across the study area. The results present trends and correlations for the Chicago area while demonstrating the utility of smart electricity meter data collection, sharing, and analysis beyond the study's geographic location.

The remaining parts of the manuscript are organized as follows. Previous research implementations based on electricity demand are part of Section 2. Then, the applied descriptive and predictive methodology is explained in Section 3. Section 4 summarizes our graphical and numerical results for describing profiles and correlation and regression analyses, respectively. The discussion of our results and findings is highlighted in Section 5. Finally, Section 6 summarizes results and describes a path forward.

2. Background

The analysis of residential electricity demand has been an increasing topic of interest due in part to the widespread installation of smart meters connected at the user level and the implications of understanding electricity consumption to develop demand management programs (Avancini, et al., 2019). Researchers have developed descriptive and forecasting models using electricity demand records at different temporal scales with a wide range of predictors (Amasyali & El-Gohary, 2018; Chou & Tran, 2018; Jain et al., 2014). These models reflect the growing knowledge base of factors affecting electricity demand in different contexts. In the next paragraphs, we present a systematic review of relevant works that used smart metered electricity demand, factors affecting residential electricity demand, and a wide range of descriptive and predictive models.

2.1. Smart meter data analysis

The ongoing roll-out of smart meters collecting high temporal electricity demand has been the backbone of recent demand-side energy management programs (Corbett, Wardle, & Chen, 2018; Meyabadi & Deihimi, 2017). Previous studies have leveraged smart electricity meters to better understand predictors of customer consumption and acceptance of smart meter infrastructure. When using smart meter data to examine climate, occupant, and housing impacts on residential electricity consumption, Kavousian et al. (2013) found that the number of occupants, high consumption appliances, weather, location, floor area, number of refrigerators, entertainment devices, and pet ownership had significant correlation with the electricity consumption profiles of the households. Income, home ownership, and building age did not show significant correlation; however, zip code variation explained up to 46% of electricity consumption variability (Kavousian et al., 2013). Classification and regression analyses of electricity demand data from households at 30-minute time steps have revealed household characteristics such as occupancy status, floor area, and employment status from smart metered data (Beckel, Sadamori, Staake, & Santini, 2014). The study also highlights the importance of linking electricity demand with household characteristics for utilities to improve their information to users. A study of the willingness-to-pay for smart meters among household customers in Germany found that intention to change usage behaviors and usefulness of consumption feedback were the most important factors for users to pay for smart metering installation (Gerpott & Paukert, 2013). Conversely, a study that used a representative sample of Great Britain energy users concluded that most users were concerned about sharing their data at a high temporal resolution (Spence, et al., 2015).

Several studies have focused on the analysis of daily electricity consumption to understand peak consumption magnitude, its corresponding time of use, and the presence of space cooling and heating (Chen, Sanders, & Ban-Weiss, 2019; Gouveia & Seixas, 2016). Electricity profiles have been partially validated with door-to-door surveys reporting space cooling as the most representative end-use for households in Evora, Portugal (Gouveia, Seixas, & Mestre, 2017). The analysis of daily profiles has also provided insights for demand-side management systems. Using smart metering data and surveys, Elma and Selamoğullar (2017) identified electricity end-uses being used during weekdays and weekends. For peak electricity demand, a wide range of research shows that variables like weather and socioeconomic conditions, including household demographics and appliance ownership, are the key drivers of peak electricity consumption analyzed mainly with daily demand profiles (Fan et al., 2017). Seasonal effects on peak electricity consumption show how peak values are highly correlated with air temperature during the summer and winter seasons, with the time of use of these peak values between 6:00 and 9:00 PM for both seasons (Chen, Ban-Weiss, & Sanders, 2018; Fan et al., 2017; Hu, Yoshino, & Jiang, 2013). As the penetration rate of smart meters has increased steadily over time (Kochański, Korczak, & Skoczkowski, 2020), customers have expressed concern about sharing information that could compromise security. Hence, some utilities have opted to increase the level of protection (encrypted communications) or to develop additional demand-side management programs at the city or zone levels (Frayssinet, et al., 2018; Noor, Yang, Guo, van Dam, & Wang, 2018).

2.2. Factors related to residential electricity consumption

In the residential environment, the multi-family housing sector has reported less electricity consumption than the single-family sector. However, multi-family accounts are the predominant residential users in large urban areas, and comparing single- and multi-family electricity consumption across different geographic areas of a city is a relatively understudied aspect of building energy consumption (Tran, Gao, Novianto, Ushifusa, & Fukuda, 2021). Previous works show that the main factors affecting residential electricity consumption are related to household characteristics including family size, age of occupants, education level, family income, and the type of house (Guo, et al., 2018; Van den Brom, Hansen, Gram-Hanssen, Meijer, & Visscher, 2019). For demand management purposes, Guo, et al. (2018) highlight the importance of differentiating daily demand profiles between the types of houses. When reviewing methods that combined end-use identification and questionnaires, Sakah, du Can, Diawuo, Sedzro, and Kuhn (2019) report that factors such as income, household size, and floor space explain up to 57% of the total household electricity consumption in an urban city in Ghana. In terms of exogenous factors, the relationships between residential electricity consumption and weather variables have been also considered at a different temporal resolution ranging from monthly to sub-hourly (Alberini et al., 2019; Son & Kim, 2017). The most repeated weather variables used by predictive models include temperature, humidity, rainfall, and cooling degree days (Son & Kim, 2017). While descriptive approaches have extensively studied the factors affecting the total residential electricity demand, Amasyali and El-Gohary (2018) found that only 19% of the reviewed research efforts focused on residential buildings and the remaining works were related to prediction models for commercial and/or educational buildings. In terms of conservation attitudes, energy-efficient features are adopted much less regularly in residential rented units than in owned housing units, with rates of adoption ranging from 5.3% to 21.6% in a study of 10 U.S. cities (Im, Seo, Cetin, & Singh, 2017). Lastly, studies have found that residential electricity consumption varies depending on the geographical location of the buildings across a city or a region (Liddle & Huntington, 2021; Liddle & Lung, 2010; Ota, Kakinaka, & Kotani, 2018).

2.3. Socioeconomic variables and electricity consumption

Various socioeconomic factors affect residential electricity consumption, with statistical significance depending on context (Huebner, Hamilton, Chalabi, Shipworth, & Oreszczyn, 2015). A study of 189 Dutch households found that overall energy use was related to sociodemographic variables, while changes in energy use were often related to psychological variables (Abrahamse & Steg, 2009). A later study of Dutch households found that households with children or elderly tended to consume more energy than other households (Brounen, Kok, & Quigley, 2012). Similarly, electricity load profiles in Europe showed strong dependence on household size, net income, age of reference person, and employment status (Hayn, Bertsch, & Fichtner, 2014). Electricity consumption has also been studied on a U.S. zip code level to determine the impact of large-scale electric vehicle adoption (Ahkamiraad & Wang, 2018). Elnakat and Gomez (2015) found that there was an 80% higher per capita energy consumption in female-dominated households compared to male-dominated households, with twice the natural gas consumption in the former. When studying the socioeconomic, demographic, and gendered influences on a household's energy consumption at the zip code level in San Antonio, Texas, Elnakat et al. (2016) found higher energy use to be associated with zip codes that were female-dominant, with a median age over 40, and with higher levels of income and education. The study also found that renters tended to use less energy than homeowners. Similarly, Karatasou, Laskari, and Santamouris (2018) found that households with more occupants, living in older and less insulated buildings, with greater floor area and electric water heating were more likely to be high energy consumers. A more recent study (Berrill, Gillingham, & Hertwich, 2021) found that household income, size, and total area from urban homes in the United States were statistically significant predictors of energy end-uses, with single-family homes using more energy than multi-family homes.

2.4. Electricity consumption at city, sub-city, and building levels

Descriptive models that provide information to city planners at different spatial resolutions are part of long-term electricity demand programs. The use of Geographic Information Systems (GIS) tools was applied by Pincetl, Graham, Murphy, and Sivaraman (2016) to investigate electricity demand variation at the city council, neighborhood, and census block levels in Los Angeles, California. The authors found that building age and size are primary variables of residential electricity demand. Similarly, at the city level, researchers have clustered electricity demand to group accounts into different categories based on their demand. Similar works analyzed spatially-aggregated demand in Australia at daily, weekly, and annual time steps identifying consumers with similar consumption patterns and geographical locations as focus areas to assist utilities in designing tailored conservation policies. Describing electricity demand with multi-annual data has proved to be an effective way of estimating residential electricity expenditure. The study found that housing and household size, energy source, and type of building are the essential features to model total electricity demand in Spain (Sanchez-Sellero & Sanchez-Sellero, 2019). When analyzing energy consumption across cities, Facchini, Kennedy, Stewart, and Mele (2017) and Bettignies, et al. (2019) compared the energy generation and usage in 27 megacities across the globe. The studies showed that total energy per capita is related to urban population density based on a power law relationship. More recently, a study based on energy analysis at sub-city levels or districts in Lima, Peru, showed that the energy use and greenhouse gas (GHG) emissions have steadily increased between 2007 and 2015 (Cárdenas-Mamani, Kahhat, & Vázquez-Rowe, 2022).

Researchers have compared the energy consumption of conventional and LEED-certified buildings using energy reports like the City of Chicago Energy Benchmarking. Their findings reflected that LEED-certified buildings do not use less energy than the non-LEED-certified buildings. Interestingly, LEED-certified schools use more energy than conventional school buildings (Scofield & Doane, 2018). Studies have also focused on shifting and reducing buildings' main end-use: Heating, Ventilation, and Air Conditioning (HVAC). The model suitability was investigated in Austin, TX, and Chicago, IL using dynamic controls and a price threshold related to peak demand. The findings revealed that energy savings were up to 6% on an annual scale (Yoon, Baldick, & Novoselac, 2016). Building material and energy consumption have also been analyzed to inform building designers about energy efficiency. Other work surveyed building materials and prices to report that buildings with hollow concrete blocks for masonry material, combined with fluorescent lights and windows with double pane absorbing heat, were the best for energy savings (Shehadi, 2018). Moreover, a study investigating the climate effects on the water-electricity demand nexus of six Midwestern U.S. cities showed that climate fluctuations alone could explain 23 to 71% of the water-electricity nexus demand. The authors compared the monthly water (gal) and electricity demand (MWh) at the city level with non-seasonal variables, such as maximum dry bulb temperature, average dew point temperature, average relative humidity, average wind speed, and the El Niño Southern Oscillation Index. These non-seasonal weather variables can explain up to 42% and 71% of electricity and water use, respectively (Obringer, Kumar, & Nateghi, 2019).

2.5. Electricity consumption prediction models

Several quantitative approaches have been used to describe electricity consumption variability with location and time. Cluster analysis, multivariate linear regression, or support vector machines combined with classification based on surveys can provide a more detailed analysis of household load profiles (Do & Cetin, 2018; Gouveia & Seixas, 2016; Howard, et al., 2012; McLoughlin, Duffy, & Conlon, 2015; Yildiz et al., 2017). Relationships between income and energy consumption have been studied using regression (Abreu, Silva, Amaro, & Magalhães, 2016) and clustering techniques have been used to group customers for electrical load pattern analysis (Chicco, 2012; Wang, et al., 2015). In one study, a hierarchical clustering method was used to detect resident profiles, finding statistical significance for months of the year, working versus weekend days, hours of the day, temperature, and baseline energy consumption (Abreu et al., 2016). A study of Danish households' hourly electricity consumption predicted that the peak electricity consumption for households will likely increase significantly for workdays by January 2030 (Andersen, Larsen, & Boomsma, 2013). A comparison study of neural networks, conditional demand analysis, and engineering model approaches determined that while all three methods can be used, a neural network model was best for end-use energy consumption modeling (Aydinalp-Koksal & Ugursal, 2008). Another end-use simulation/forecasting model combined load data with survey results to estimate the total residential load curve in New South Wales, Australia, including considerations for monthly and daily variations as well as weather dependencies (Bartels, Fiebig, Garben, & Lumsdaine, 1992). Parti and Parti (1980) similarly obtained detailed household level data for 5,286 households in San Diego County via a mail questionnaire, using the results to determine a conditional demand framework to disaggregate household demand into 16 appliance categories. As electricity prediction models are abundant (Hong, et al., 2020; Yildiz et al., 2017), it is important to consider the model's application, simplifications, and how to apply it to diverse urban populations.

Traditional methods for energy demand forecasting for demand side management include time series, regression, and econometric modeling, while soft techniques such as fuzzy logic, genetic algorithms, and neural networks have also been used (Suganthi & Samuel, 2012). Upgrade options for reducing residential energy consumption in Canada were evaluated using an end-use electricity consumption model, which found that upgrading appliances would lead to significant savings (Farahbakhsh, Ugursal, & Fung, 1998). Different statistical analysis methods have been used to estimate residential end-use load curves, such as conditional demand analysis, seemingly unrelated regressions, and the random coefficient model (Fiebig, Bartels, & Aigner, 1991), while others use artificial intelligence for load forecast models (Raza & Khosravi, 2015).

Despite these advances in modeling residential electricity demands and load patterns, differences between the single- and multi-family residential housing sectors across different geographic zones of a city remain understudied in the context of residential electricity consumption. We aim to fill this knowledge gap for the residential housing sector in Chicago, Illinois, analyzing both electricity load profiles and socioeconomic factors related to electricity consumption. We present a multiple linear regression analysis to discern the most important socioeconomic characteristics to model the mean daily electricity demand of single- and multi-family users. Also, we compare the consumption of these two types of residential customers and describe the main differences in magnitude and time of use. The models developed in this research can assist utility managers and city planners to identify the variables of interest for developing demand-side management strategies for sustainable energy consumption in highly diverse urban environments.

3. Methodology

This section describes how we model the daily profile of electricity demand reported by single- and multi-family accounts using

Table 1
Socioeconomic data of the city of Chicago at the zip code level reported by the American Community Survey 5-Year Estimates.

Demographics	Housing and location	Income	Education
Population (*Pop*)	Median building age (*MedBuildAge*)	Median household income (*MedHhInc*)	Percent with a high school degree (*%HS*)
Median age of occupants (*MedAgeOcc*)	Percent of multi-family housing (*%Multi*)	Mean household income (*MeanHhInc*)	Percent with a bachelor's degree (*%BS*)
Percent 65 and over (*%65+*)	Percent occupancy (*%Occ*)	Percent under poverty line (*%Pov*)	
Percent female (*%Fem*)	Total housing units (*TotHouses*)	Unemployment percent (*%Unemp*)	
Household with children under 6 (*Children<6*)	Mean commute time (*MeanCommute*)		
Households with children 6 to 17 (*6<Children<17*)			

half-hourly data reported by smart meters. The first subsection describes the electricity demand data and the aggregation process from electricity demand at the account level to the zip code level and their exogenous socioeconomic predictor variables. The second subsection describes the approach applied to group data based on the time of day and day of the week. The third subsection describes the socioeconomic data variables used in the predictive analysis. Finally, the statistical predictive models are explained in detail, with the evaluation metric reported as the Coefficient of Determination.

3.1. Electricity consumption across zip codes

Commonwealth Edison (ComEd) has installed smart electricity meters throughout its service area in northern Illinois. These data are anonymized smart meter readings reported from 00:00 of January 1st to 23:30 of December 31st of 2016 with 30-minute resolution (17,568 observations/meter) within the ComEd service area. The spatial resolution of the electricity demand data is at the account level. Each smart meter and its corresponding data were reported with its zip code, a unique account identifier, and the delivery service type, designating single- or multi-family residential and electric or non-electric space heating.

With over 500 zip codes served by ComEd, we focused on the 56 zip codes covering the City of Chicago to represent a large, diverse city with a range of housing options and distinct neighborhoods with reported demographic information (U.S. Census Bureau, 2016). ComEd has installed over 4 million smart meters in northern Illinois and there are 10,000 smart meters on average across the 56 zip codes classified as either "Single-Family" or "Multi-Family" users, as displayed in Fig. 1.

The smart meters have an accuracy within +/- 0.2% with 6 W starting Watts at a voltage of 120 V. The meters operate over a temperature range between - 40 C through +85 C. Smart meter-Utility consumption communication uses Radio Frequency, Power Line Communication, and a cellular network (Zheng, Gao, & Lin, 2013). Real-time transmission between users and utility allows two-way communication for demand management strategies.

The analysis of ComEd smart meter data for single- and multi-family residential electricity consumption has two components as shown in Fig. 2: (1) Profiles Modeling, which represents the mean daily electricity (MDE) demand over the year for each zip code, and (2) Regression Models to build statistical models using the MDE from step 1 as the dependent variable and socioeconomic data at the zip code level as the predictors. Step 1 was implemented in the open-source Python programming language version 3.7 (Van Rossum & Drake, 2009) and step 2 was modeled in the open-source statistical programming language R (R Core Team, 2017).

3.2. Profiles modeling

The data set was first filtered using the 'Delivery Service Name' variable to include only electricity demand from labeled single- and multi-family residential consumers in the analysis. As shown in Fig. 2, we aggregated the data by calculating the average consumption of each meter for each month and for each zip code. To decrease the impact of extreme values but still preserve unusual events, we retained electricity demand values within the 5th and 95th percentiles, expanding the interquartile range approach mentioned in the review by Walfish (2006). Then, we determine the daily consumption for both types of residential end-users and report differences on electricity consumption during weekdays and weekends, and hourly peak and non-peak time and magnitude. We also calculate the average daily profiles for each zip code and month of the year to analyze the electricity demand of single- and multi-family residential users in Chicago. This information was passed to the regression analysis to build regression models for the mean daily electricity demand data using socioeconomic explanatory variables.

3.3. Zip code socioeconomic data

To study the impact of demographic and socioeconomic variables on electricity consumption, it is necessary to have robust data about both electricity consumption and demographic variables for the same geographic location (Liu, et al., 2021). With anonymized smart meters, it is infeasible to determine the characteristics of the specific households represented by each smart meter. However, the American Community Survey (ACS) 5-Year Estimate reports annual demographic information on a zip code level (U.S. Census Bureau, 2016). Although the inclusion of environmental data might help understand electricity demand (Jovanović, Savić, Bojić, Djordjević, & Nikolić, 2015; Zhang, Chen, Hu, Deng, & He, 2021; Zhang, Liao, & Mi, 2019), the purpose of the presented method is to compare the mean daily electricity demand across zip codes; hence, environmental variables are approximated as equal across the analyzed geographic area. With the smart meter data labeled by zip code, we analyzed the electricity consumption data in conjunction with collected demographic and socioeconomic data from 2016 to estimate correlations and create a predictive model for both single- and multi-family residential users. Table 1 summarizes the socioeconomic data from the ACS 5-Year Estimates that were collected for each zip code in Chicago. These socioeconomic data shown in Table 1 were selected as possible explanatory variables of residential electricity demand based on previous results reported in literature (Aydinalp-Koksal & Ugursal, 2008; Brounen et al., 2012; Elnakat & Gomez, 2015; Gouveia & Seixas, 2016; Hayn et al., 2014; Kavousian et al., 2013; McLoughlin, Duffy, & Conlon, 2012). The socioeconomic variables reported at the zip code level can be classified into four groups depending on their characteristics. Demographics include the number of inhabitants per zip code *(Pop)*, the median age of occupants *(MedAgeOcc)*, percentage of people 65 and over *(% 65+)*, number of households with children under six years old *(Children<6)*, and households with children between 6 and 17 years old *(6<Children<17)*. Housing and Location include the median building age as the difference between 2016 and its construction year *(MedBuildAge)*, the percentage of multi-family housing at the zip code level *(% Multi)*, percentage of occupancy *(% Occ)*, the total number of housing units *(TotHouses)*, and the reported mean commute time in minutes *(MeanCommute)*. Two measures of central tendency, the annual median *(MedHhInc)* and mean *(MeanHhInc)* household income, the percentage of households under the poverty line *(% Pov)*, and the percentage of unemployment *(% Unemp)*, are part of the Income group. Finally, Education includes the percentage of households reporting a high school *(% HS)* and a bachelor's *(% BS)* degree, respectively.

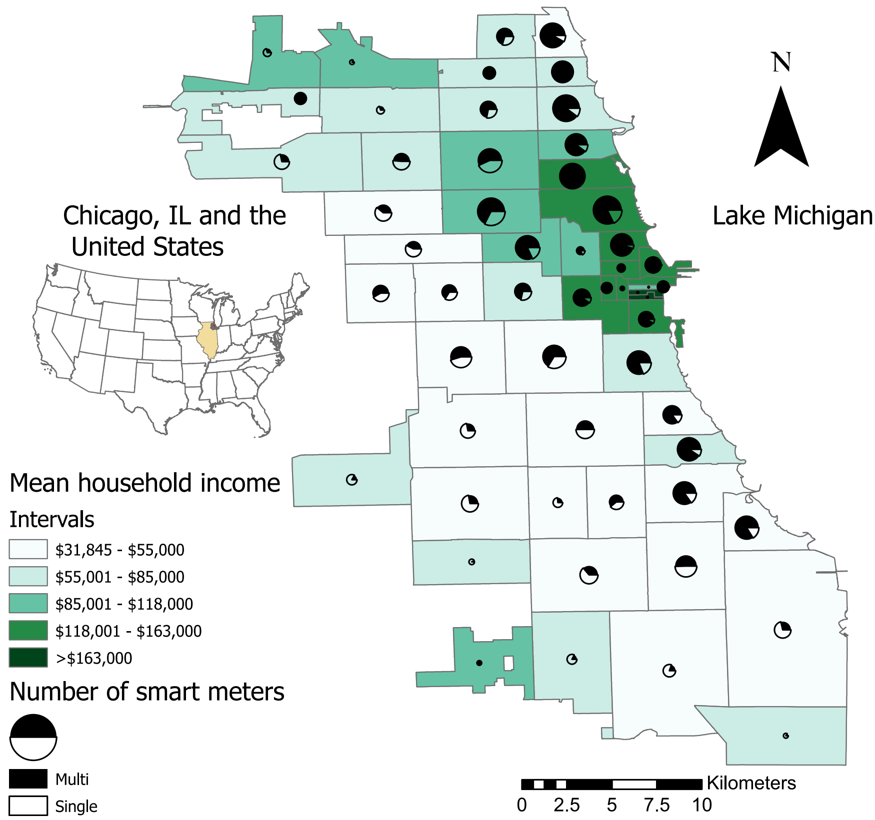


Fig. 1. The number of smart meters for each zip code within Chicago is shown with different-size markers. The mean household income shows low values with light green and high values with dark green for each zip code, respectfully.

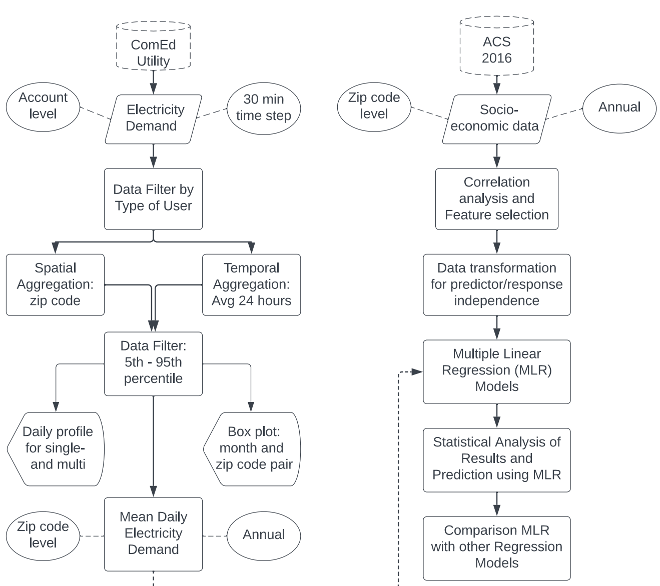


Fig. 2. The two sequential approaches presented in the Methodology to analyze socioeconomic at the zip code level of the city of Chicago and the mean daily electricity demand for both single- and multi-family residential users.

3.4. Correlation analysis of predictors and response

Based on previous implementations (Kim, Kim, & Srebric, 2020a; Rhodes, Cole, Upshaw, Edgar, & Webber, 2014), a pairwise correlation analysis was performed for two reasons. First, we visualize how the Mean Daily Electricity (MDE) demand varies with respect to all the socioeconomic variables. Second, a pairwise correlation of the variables may identify highly correlated predictors that can help the model's performance but compromise its broader applications. Hence, the Pearson correlation coefficient was calculated for all the predictors and the response variable to identify correlation at the 5% significance level. We performed the mentioned pairwise correlation analysis between the predictors (socioeconomic data at the zip code level) and the response variable (mean daily electricity demand of single- and multi-family residential users). This step determines the strength of the relationship between each of the predictors and the response variable and to select the most significant predictors as the input of the regression model (dos Santos & Balestieri, 2018). As shown in Eq. (1), we calculated the pairwise correlation as follows:

$$r_{X,Y} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sigma_X \sigma_Y} \tag{1}$$

where $r_{X,Y}$ is the Pearson product moment correlation coefficient; $\overline{x}$ and $\overline{y}$ are the sample means of variables X (shown in Table 1) and variable Y (mean daily electricity demand) calculated for each i value of the total n data points; σ_X and σ_Y are the sample standard deviation of variables X and Y, respectively. The Pearson correlation coefficient of two variables indicates high positive or negative correlation when r is +1 or -1, respectively, whereas no correlation exists when r is 0 (Taylor, 1990).

3.5. Annual electricity consumption prediction model and statistical analysis

With the developed single- and multi-family residential electricity profiles and the corresponding identified socioeconomic variables across zip codes, we created a regression model to evaluate socioeconomic indicators as possible predictors of the aggregated mean daily electricity demand over a year for single- and multi-family residential

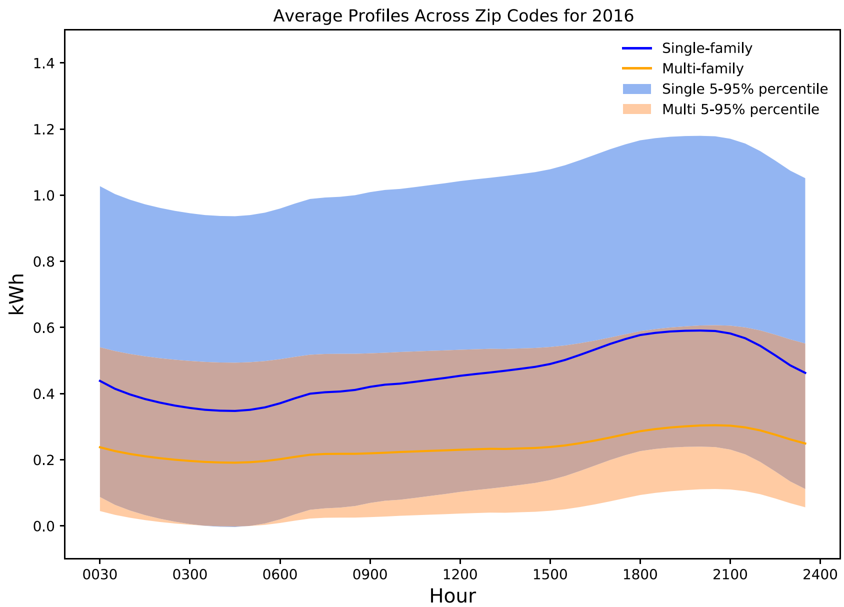


Fig. 3. Mean daily electricity demand profiles for single- and multi-family residential users across Chicago zip codes (n = 56) for 2016.

electricity consumption (Afaifia, Djiar, Bich-Ngoc, & Teller, 2021; Kim, Kim, & Srebric, 2020b). Since socioeconomic data were available on an annual timestep, a single- and multi-family residential electricity profile for each zip code for an average day in the year was used with linear regression to reveal the best socioeconomic predictors of electricity consumption in these two residential end-users. The best-fit models, one for single-family users and one for multi-family users, were determined through ordinary least squares (OLS) regression using the form listed in Eq. (2) through linear modeling in the open-source statistical programming language R and the Companion to Applied Regression (CAR) package (Fox & Weisberg, 2019).

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \cdots + \beta_{nsp} X_{nsp} \tag{2}$$

where Y represents the residential single- or multi-family mean daily electricity demand for each observation i, $\beta_{0,1,\ldots,nsp}$ are the coefficients of their corresponding socioeconomic predictor $X_{0,1,\ldots,nsp}$; and, nsp is the number of selected predictors.

We evaluated the OLS assumptions of normality and constant variance of residuals, and independence of predictor variables and residuals to verify the model's inference validity. The normality and constant variance of the residuals were quantified using the Shapiro–Wilk normality test (Shapiro & Wilk, 1965) and Tukey test (Tukey, 1953), respectively, using the CAR package in R. Independence of predictor variables was determined by evaluating multicollinearity and the Durbin–Watson test for autocorrelation (Durbin & Watson, 1950) in R with the Diagnostic Checking in Regression Relationships (lmtest) package (Zeileis & Hothorn, 2002). Diagnostics were performed on the statistical significance at the 95% confidence level of the coefficient estimates and the model form, including estimation of the goodness-of-fit R^2 statistic (Eq. (3)). We supplemented OLS multiple linear regression with relative importance analysis, using a customized version of R code available from Tonidandel and LeBreton (2011). Relative importance analysis partitions the explained variance among the predictor variables (i.e., $X_{1,\ldots,nsp}$) to estimate the contribution of each predictor such that the sum of each predictor's relative importance equals the R^2 value (Tonidandel & LeBreton, 2011).

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n} (Y_i - \bar{Y}_i)^2} \tag{3}$$

where Y represents the residential single- or multi-family mean daily electricity (MDE) demand (kWh) for each observation i from 1 to n zip codes, $\hat{Y}$ is the predicted MDE demand (kWh); and, $\bar{Y}$ is the arithmetical mean of the observed MDE demand values (kWh).

We also built a regression tree model to explore the capabilities of regression trees in comparison to the multiple linear regression model. The regression tree hyperparameters used in this analysis correspond to a tree with the mean squared error as the splitting criterion and a depth controlled by the 'minLeafSize' value. To compare the R^2 values from multiple linear regression and regression trees, we split the data set into 70% training and 30% testing as performed in previous literature (Pesantez, Berglund, & Kaza, 2020). Finally, the importance of each predictor is reported by both the MLR and regression tree models.

4. Results

4.1. Electricity profile graphical analysis

Through visual analysis of monthly average daily electricity consumption for each zip code, high variances were observed between months within each zip code. High variances also exist between zip codes within each month. The daily electricity profiles show a similar trend where the minimum consumption occurs at 4:00 AM and the maximum consumption occurs between 6:00 and 9:00 PM with magnitudes of 0.6 and 0.2 kWh for single- and multi-family users, respectively (Fig. 3). Single-family users consistently report more electricity demand on average than multi-family users, and this difference in electricity demand of the two residential end-users is in accordance with studies reported in literature (Berrill et al., 2021).

The MDE demand at the zip code level reported by the single- and multi-family accounts shows that the average of multi-family electricity demand generally spreads evenly across the greater Chicago area. However, the MDE demand reported by single-family accounts is considerably high in the northeast part of the city. Furthermore, more populated zip codes do not report the highest MDE demand values (Fig. 4). MDE demand values range between 17 and 33 kWh and 8 and 15 kWh for single- and multi-family accounts, respectively. These results are similar to previous studies of other large U.S. cities (Pincetl et al., 2016) where the median annual residential electricity demand is reported at the neighborhood and census block group levels to protect customers' privacy.

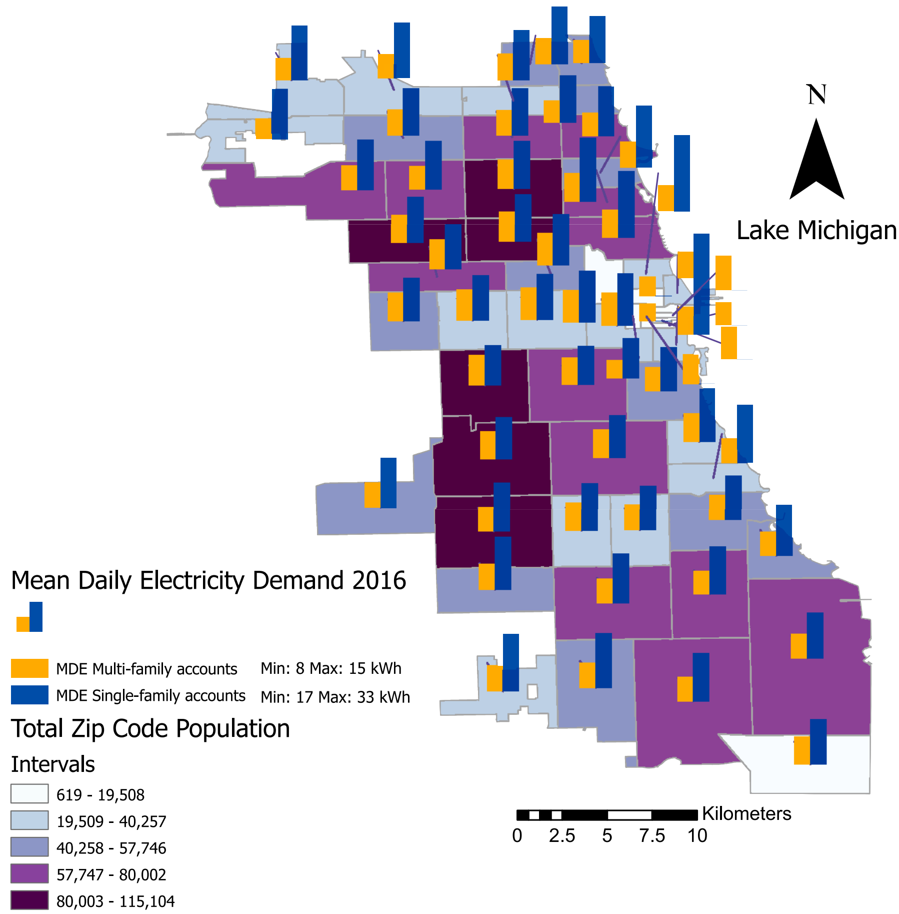


Fig. 4. Population and MDE demand reported by single- and multi family accounts at zip code level for the city of Chicago.

Previous works report population density as the primary driver of energy consumption in big cities (Facchini et al., 2017). While Bettignies, et al. (2019) compared total electricity demand, our study focuses on the two types of residential demand, and the predictors for these types of customers are analyzed in the regression analysis. Similar to the analysis of Bettignies, et al. (2019), our study did not find a significant correlation between the MDE demand per capita of single- and multi-family accounts and the zip code's population density for the city of Chicago. Also, as the MDE demand per capita differs at "intra-city" or zip code levels (Fig. 5), it is significantly correlated to the mean income of these city sublevel areas. When comparing single- and multi-family MDE demand per capita, it can be seen that northwest Chicago exerts more MDE demand from single- than multi-family accounts. The highest MDE demand from multi-family accounts concentrates in downtown Chicago due to the significant presence of multi-family buildings.

4.2. Daily and hourly electricity demand

The daily electricity demand of single-family accounts roughly duplicates the demand of multi-family accounts across the days of the week in the city of Chicago during the analysis period (Fig. 6). This result was expected considering different factors like the built form, relative size of the dwelling spaces, and the number of occupants likely to be living in them (Huebner, et al., 2015; Van den Brom, et al., 2019). From a city perspective, peak electricity demand from both types of accounts decreases about 10% on Thursday, Friday, and Saturday compared to the rest of the days. The vast majority of multi-family accounts are in a specific area of the City of Chicago (see Fig. 1) where residential electricity consumption is not the predominant type of consumption when compared to commercial. Visibility of electricity consumption from single- and multi-family accounts at high temporal resolution across different Chicago areas can support the city plan to unlock energy and cost savings for residents (City of Chicago, 2016; Won, No, & Alhadidi, 2019).

The electricity demand of both types of residential end-users varies monthly in response to air temperature. During the coldest winter months, February and March, single-family customers show nearly double the electricity consumption of multi-family users (Figs. 7(a)–7(d)). This situation also occurs during the warmest summer months, July and August, where single-family users' minimum consumption remains higher than multi-family users' maximum consumption. Also, Figs. 7 and 8 illustrate the strong temporal and spatial variability in single-family electricity consumption during the summer. During winter and summer months, there was higher electricity consumption compared to the rest of the year, likely due to space heating and air conditioning (Chen et al., 2019; Gouveia & Seixas, 2016). During the winter months, there are two consumption peaks that occur at the same time of the day for both types of residential end-users. The first peak occurs at approximately 7:00 AM and the second and more pronounced peak occurs at 7:00 PM, which could be associated with cooking and entertainment activities. During the summer months, one pronounced and sustained peak is present for both single- and multi-family users. The peak consumption time during summer occurs at 6:00 PM and remains until 10:00 PM. There are also electricity values reporting unusual high demand at different times of the day from both types of residential customers that could be attributed to data inaccuracies or specific electricity uses that utility managers may be interested in identifying. Similar visualization for electricity consumption during spring and fall months, which had the lowest electricity consumption variability, are presented in the Supporting Information.

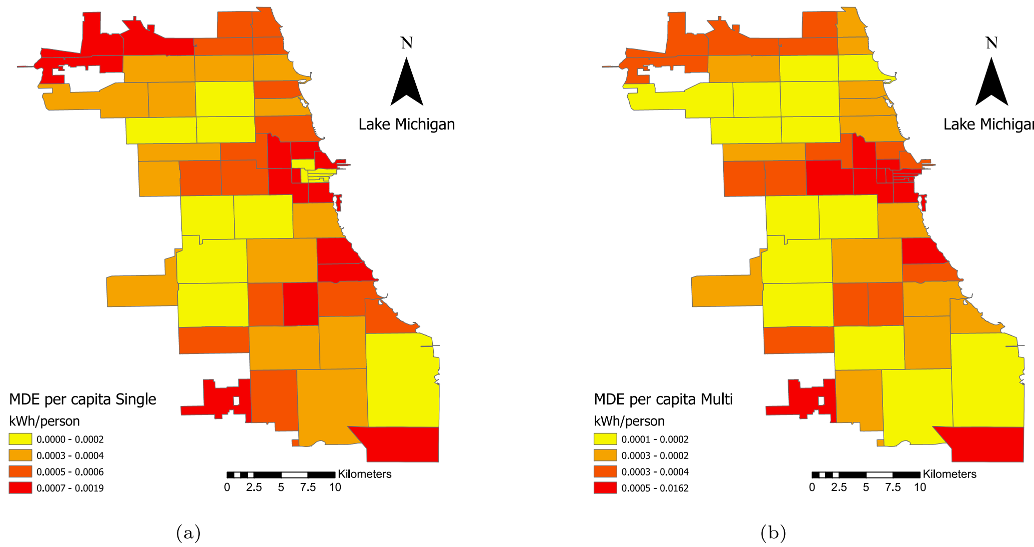


Fig. 5. Mean Daily Electricity (MDE) demand per capita reported at the zip code level from (a) single- and (b) multi-family residential accounts in Chicago during 2016.

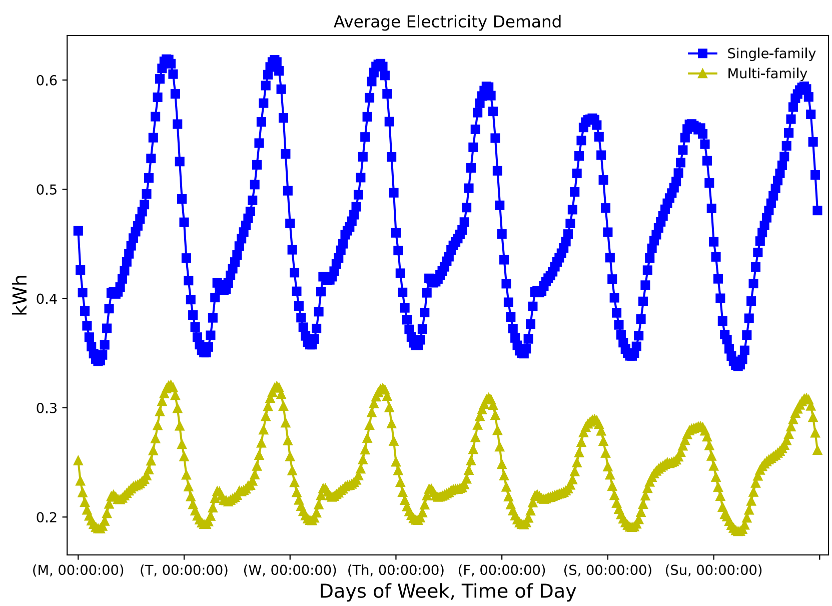


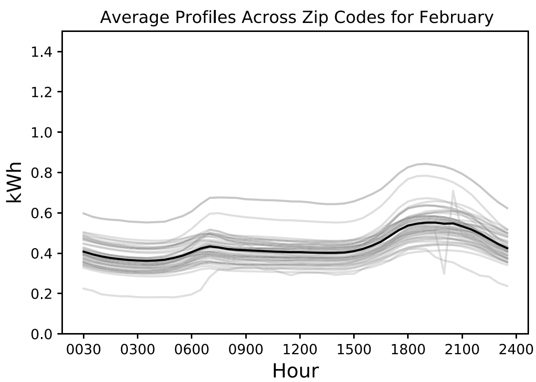
Fig. 6. Mean hourly electricity demand over the days of the week for single- and multi-family residential accounts across Chicago zip codes for 2016.

4.3. Spatial variation in electricity consumption

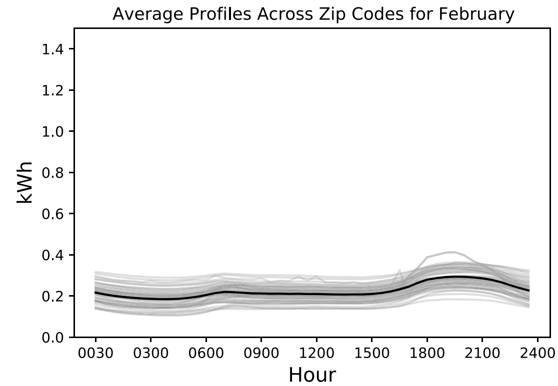
Due to socioeconomic factors, single- and multi-family residential electricity consumption varies across specific zip codes. Fig. 1 shows zip codes with different mean annual income ranges. Darker polygons represent zip codes with higher mean household incomes, based on 2016 data (U.S. Census Bureau, 2016). Similarly, Chicago zip codes also contain different proportions of single- and multi-family households, indicated by the pie charts within each zip code (Fig. 1). For comparison purposes, we calculated the average consumption of five zip codes with the highest and lowest mean income values for the multi-family users (shaded areas shown in Figure S1 of the Supporting Information section). The average mean daily electricity demand of these two income classes of customers (i.e., low- and high-mean household income values) is presented in Fig. 9, shown only for multi-family users as certain Chicago zip codes have no single-family users (see Fig. 1). In February of 2016, a cold temperature month in Chicago, the electricity demand of multi-family accounts presented a similar trend and magnitude on an hourly basis for both low- and high-mean-annual income (Figs. 9(a) and 9(b)). In August 2016, the daily electricity consumption of multi-family users from low-income zip codes shows higher variation than the consumption of users from high-income zip codes (Figs. 9(c) and 9(d)). Despite showing a similar trend during warm weather, the mean electricity demand in multi-family residential households from low-income zip codes is generally higher than the mean electricity demand in multi-family residential households from high-income zip codes. One possible explanation for this result is that the high electricity demand exerted by low-income zip codes could be due to less efficient appliances installed in aged buildings and variability in home size. We performed a t-test to examine the statistical relevance of the difference between the multi-family residential electricity demand for low- and high-income zip codes reported in August of 2016 and shown in Figs. 9(c) and 9(d). The following hypothesis was evaluated:

H_0: The difference between the mean electricity demand of low-income and high-income zip codes is zero.

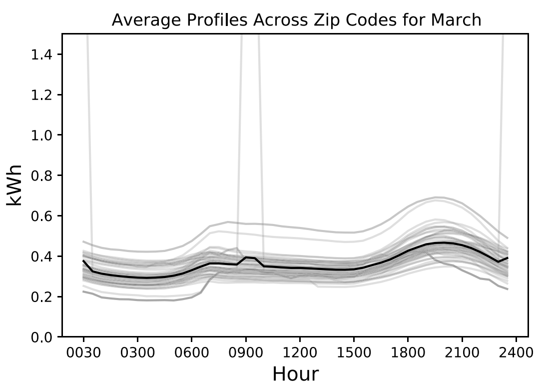
$$\mu_{\text{low-income}} - \mu_{\text{high-income}} = 0$$



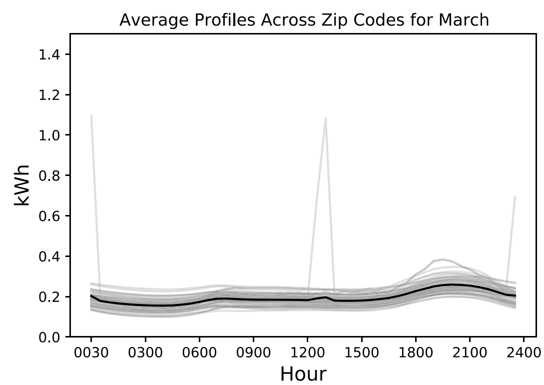
(a) Single-family for all zip codes in February



(b) Multi-family for all zip codes in February



(c) Single-family for all zip codes in March



(d) Multi-family for all zip codes in March

Fig. 7. Mean daily electricity demand profiles of single-family ((a) and (c)) and multi-family ((b) and (d)) residential users. Electricity load varies with time of day across the analyzed zip codes for the two coldest months of winter (February and March) during 2016. The average profiles are represented as gray lines for each zip code, with the black line reflecting the average across Chicago zip codes.

H_A: The difference between the mean electricity demand of low-income and high-income zip codes is not zero.

$$\mu_{\text{low-income}} - \mu_{\text{high-income}} \neq 0$$

The two-tailed t-test resulted in a p-value close to zero. Therefore, we reject the null hypothesis (H_0) of having identical expected values in favor of the alternate hypothesis (H_A) that the mean electricity demand reported in August of 2016 in low-income zip codes is significantly different from that in high-income zip codes at a 5% confidence level.

We explored the electricity demand of two geographic areas characterized by differences in the median age of inhabitants. Similar to mean household income, age of occupants can determine residential electricity consumption (Elnakat et al., 2016; Hayn et al., 2014). Fig. 10 illustrates the differences in electricity demand during a cold temperature month (February) and a warm temperature month (August) between zip codes with differing occupant median age. Garfield Park with a much lower median age (28.7 years) than The Loop (44.6 years) (U.S. Census Bureau, 2016). The difference in electricity demand across zip codes with different median age highlights a potential impact of socioeconomic indicators on multi-family residential electricity consumption, shown in Fig. 10 for these zip codes. We performed a t-test for February and one for August to examine the statistical relevance of the difference between the electricity demand for low- and high-age values, including the zip codes of Garfield Park and The Loop areas. The following hypothesis was evaluated:

H_0: The difference between the mean electricity demand of zip codes with low-age values and high-age values is zero.

$$\mu_{\text{low-age}} - \mu_{\text{high-age}} = 0$$

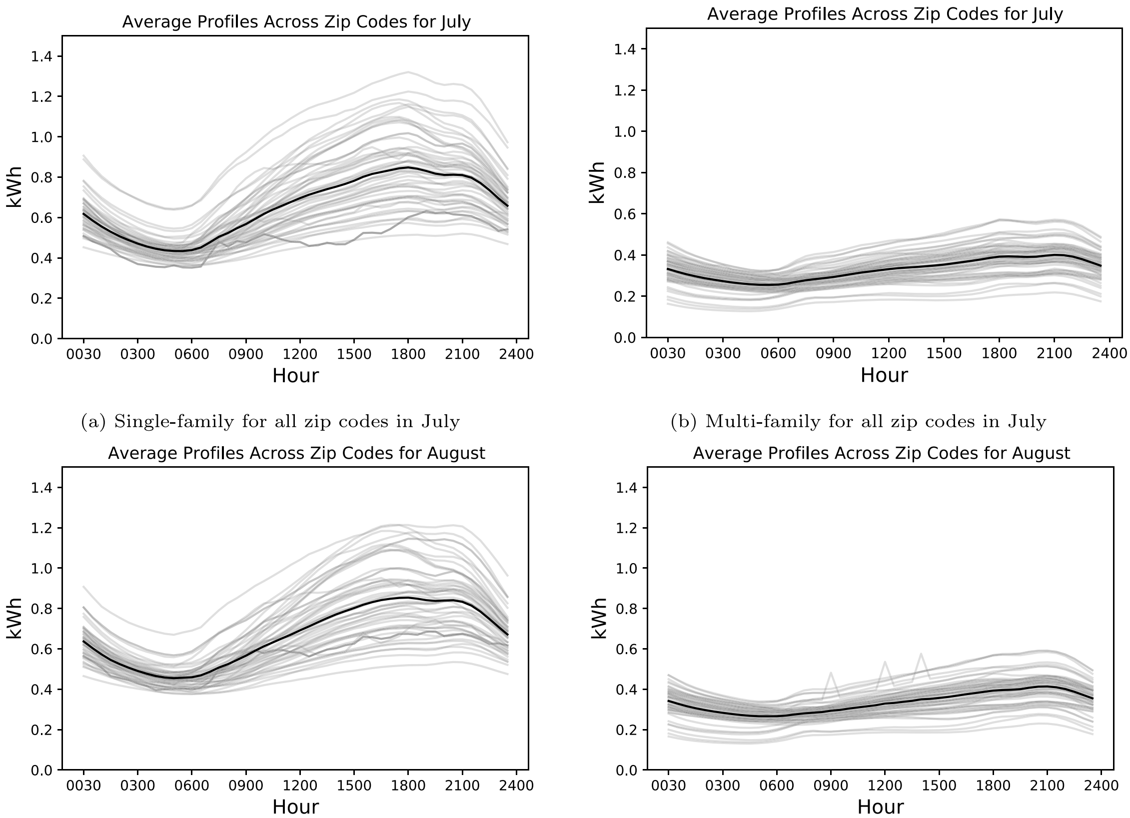
H_A: The difference between the mean electricity demand of zip codes with low-age values and high-age values is not zero.

$$\mu_{\text{low-age}} - \mu_{\text{high-age}} \neq 0$$

The two-tailed t-test resulted in a p-value close to zero for both months. Therefore, we reject the null hypothesis (H_0) of having identical expected values, in favor of the alternate hypothesis (H_A) that the mean electricity demand in zip codes with low-age occupants is significantly different than that in zip codes with high-age occupants at a 5% confidence level. The results of electricity demand were the same across the eleven remaining months of 2016 and the box plots are presented in the Supporting Information section.

4.4. Correlation

Correlation analysis was performed to reveal interdependencies between predictors and the mean daily electricity demand of single- and multi-family residential users and to analyze the feasibility of regression analysis. As shown in Fig. 11, the pairwise correlation of socioeconomic variables and mean daily electricity demand shows that for single-family users, percentage under the poverty line, percentage of unemployment, number of children under 6 and between 6 and 17 years old, and the median age of building are negatively correlated to electricity demand. Similarly, predictors including the percentage of people with bachelor and postgraduate education, mean income, percentage of occupancy, and percentage of females are positively correlated to electricity demand (Fig. 11(a)). For the multi-family users, the mean daily electricity demand is negatively correlated to the median age of inhabitants, percentage of elderly, percentage of occupancy, and the percentage of people that at least finished high



(a) Single-family for all zip codes in July

(b) Multi-family for all zip codes in July

(c) Single-family for all zip codes in August

(d) Multi-family for all zip codes in August

Fig. 8. Mean daily electricity demand profiles of single-family ((a) and (c)) and multi-family ((b) and (d)) residential users. Electricity load varies with time of day across the analyzed zip codes for the two warmest months of summer (July and August) during 2016. The average profiles are represented as gray lines for each zip code, with the black line reflecting the average across Chicago zip codes.

school. Lastly, there is a positive correlation between the mean daily electricity demand at the zip code level and the percentage of people under the poverty line, the percentage of multi-family users, and the percentage of children under 6 years old (Fig. 11(b)).

As part of the correlation analysis, we calculated the Pearson's correlation and *p*-values for testing the null hypothesis of no correlation against the right-tailed alternative that the correlations are more significant than zero (Benesty, Chen, Huang, & Cohen, 2009). Figs. 12(a) and 12(b) show an extract of this correlation analysis for single- and multi-family accounts, respectively, between the response variable (Mean daily electricity demand (kWh)) and predictors Percent with a high school degree, Mean annual income, and Building age. This analysis shows that electricity usage increases as households have occupants holding a high school degree and increasing income, decreasing in old buildings. For multi-family accounts, occupants with a high school degree consume less electricity, and there is no correlation between electricity usage, income, and building age. The Supporting Information illustrates the remaining correlation figures.

4.5. Multiple linear regression models and statistical analysis

We created a multiple linear regression model to further investigate the ability of socioeconomic indicators to explain variation in both single- and multi-family residential electricity consumption in Chicago. Using an ordinary least squares modeling approach, a best-fit multiple linear regression model was created following the form of Eq. (2), using the socioeconomic data for Chicago zip codes as indicators. The OLS approach used backwards stepwise regression to create a best-fit model of statistically-significant predictor variables. The coefficient estimates shown in Tables 2 and 3 reflect the multiple linear regression model, rounded to two significant figures, explaining approximately 62% and 41% of the variability in single- and multi-family residential electricity consumption in Chicago, respectively. Tables 2 and 3 also include measures of the statistical significance of each coefficient and the model form, along with the relative importance analysis results.

OLS assumptions of normality and constant variance of residuals, and independence of residuals and predictor variables were evaluated statistically. The normality of the residuals was measured using the Shapiro–Wilk normality test, providing enough evidence to fail to reject the null hypothesis of normally distributed residuals. We tested the OLS assumption of constant variance of the residuals using the Tukey test. For the multi-family model, we fail to reject the null hypothesis of constant variance across residuals, shown graphically in Figure S4. In the case of the single-family model, we reject the null hypothesis in favor of the alternate for the Tukey test, indicating non-constant variance in the residuals. To cope with this heteroskedasticity, we used a Box–Cox transformation (Box & Cox, 1964) for the single-family multiple linear regression model shown in Table 2. The non-transformed single-family multiple linear regression model, exhibiting non-constant variance in the residuals, is shown in the Supporting Information (Table S1). With this transformation, we ensure that the dependent variable follows a normal distribution. For the Shapiro–Wilk test, the transformed model also complies with failing to reject the null hypothesis as with the non-transformed model. For the transformed single-family model, the Tukey test provides sufficient information to fail to reject the null hypothesis such that we can approximate constant variance across residuals (Figure S3) as obtained with the multi-family model. To assess

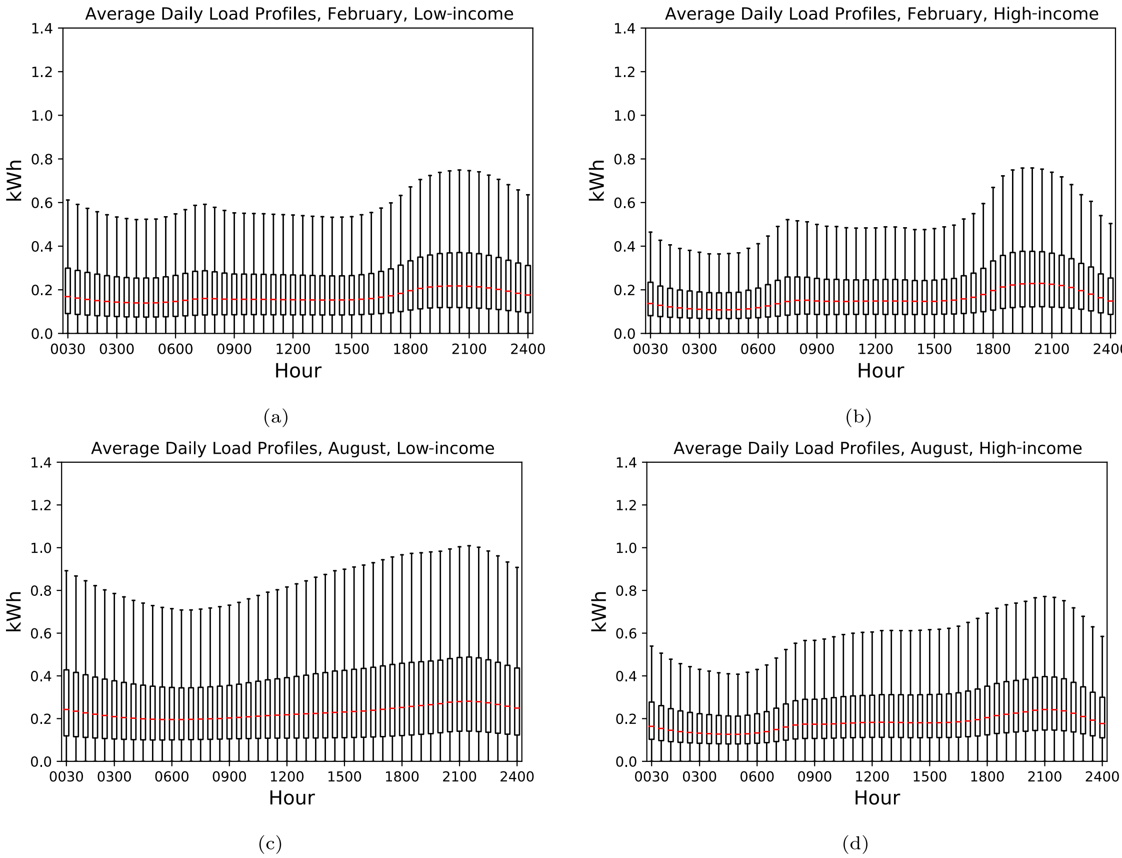


Fig. 9. Daily load profiles of multi-family residential users in February and August 2016 corresponding to low-mean-income ((a) and (c)) and high-mean-income ((b) and (d)) zip codes.

Table 2
Socioeconomic indicators were statistically significant predictors of the transformed single-family residential electricity consumption for Chicago zip codes, explaining approximately 62% of the variability.

Factor	Coefficient	Estimate	Std. error	t-Value	Pr(>|t|)	Relative importance
Constant	β_0	6.2e−01	2.2e−04	280	0.00	–
Percent occupancy	β_1	1.3e−02	2.4e−03	5.3	3.7e−06	0.14
Median age of occupants	β_2	−3.1e−04	6.3e−05	−4.9	1.4e−05	0.055
Percent 65 and over	β_3	2.7e−04	5.5e−05	4.8	1.8e−05	0.085
Mean commute time	β_4	1.1e−04	2.9e−05	3.8	4.1e−04	0.028
Percent high school or higher	β_5	8.1e−05	1.1e−05	7.1	9.6e−09	0.35
Multiple R^2 = 0.65; Adjusted R^2 = 0.62; F-statistic = 16 (p-value = 4.0e−09)						

Results are rounded to 2 significant digits.

Table 3
Socioeconomic indicators were statistically significant predictors of multi-family residential electricity consumption for Chicago zip codes, explaining approximately 41% of the variability.

Factor	Coefficient	Estimate	Std. Error	t-Value	Pr(>|t|)	Relative importance
Constant	β_0	19	3.6	5.2	4.4e−06	–
Median building age	β_1	0.029	0.012	2.3	0.024	0.022
Percent multi-family	β_2	−1.9	0.86	−2.2	0.031	0.030
Percent female	β_3	0.14	0.071	1.9	0.059	0.0086
Median age of occupants	β_4	−0.28	0.046	−6.1	1.4e−07	0.36
Mean commute time	β_5	−0.14	0.051	−2.8	0.008	0.046
Multiple R^2 = 0.47; Adjusted R^2 = 0.41; F-statistic = 8.8 (p-value = 4.2e−06)						

Results are rounded to 2 significant digits.

independence of the residuals of the transformed single-family and the multi-family models (Tables 2 and 3), shown in Figures S3 and S4, the Durbin–Watson statistic was calculated to quantify autocorrelation, with p-values of 0.31 and 0.75, respectively. The hypothesis test decision is to fail to reject the null hypothesis, confirming the assumption of independence of the residuals of both models. Independence of the socioeconomic predictor variables was determined by estimating the variance inflation factors (VIF) as a measure of multicollinearity. VIF > 10 is an indication of correlation between predictor variables, violating assumptions of independence. Table 4 summarizes the VIF values for the socioeconomic indicators, confirming independence among the predictor variables. OLS assumption hypothesis tests and values are reported in further detail in the Supporting Information.

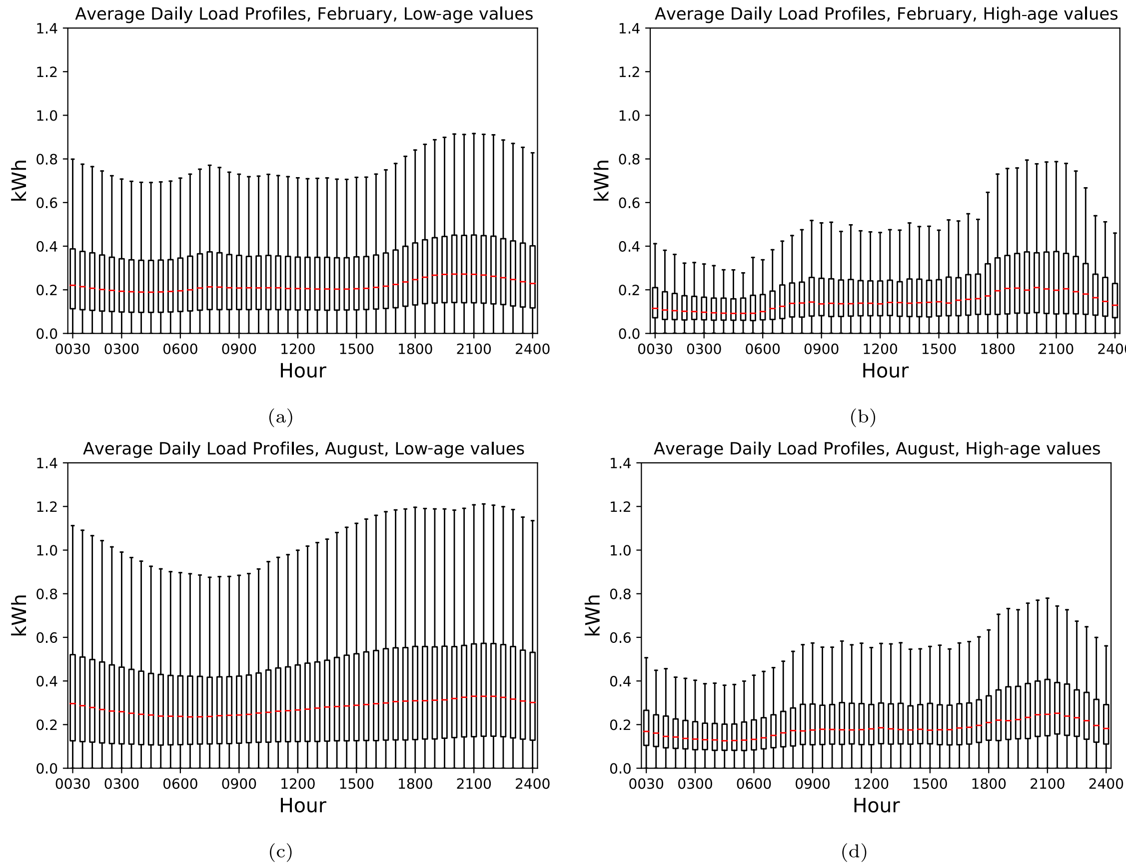


Fig. 10. Daily load profiles vary for multi-family residential housing across zip codes, shown for February and August 2016 in Garfield Park (median age: 28.7 years) ((a) and (c)) and The Loop (median age: 44.6 years) ((b) and (d)).

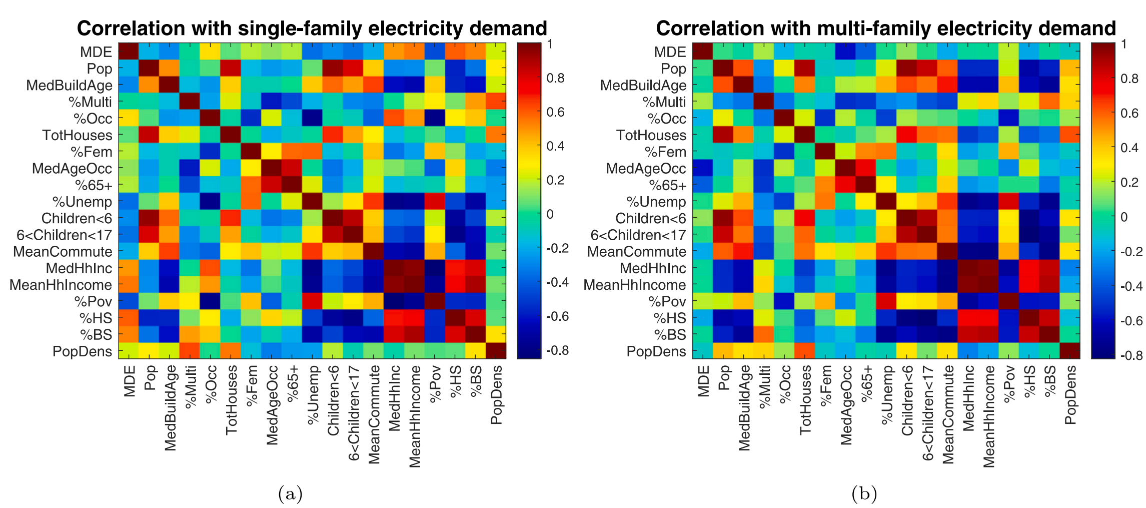


Fig. 11. Pairwise correlation map of the analyzed socioeconomic variables and the mean daily electricity demand of (a) single- and (b) multi-family residential users in Chicago.

Table 4
The socioeconomic indicators did not exhibit multicollinearity as all Variance Inflation Factors (VIF) values were less than 10 for both models.

Predictor variable	Single-family model VIF	Multi-family model VIF
Percent of occupancy	2.22	–
Median age of occupants	6.91	1.43
Percent 65 and over	5.66	–
Mean commute time	1.76	2.97
Percent high school or higher	1.55	–
Median building age	–	2.32
Percent multi-family	–	1.69
Percent female	–	1.40

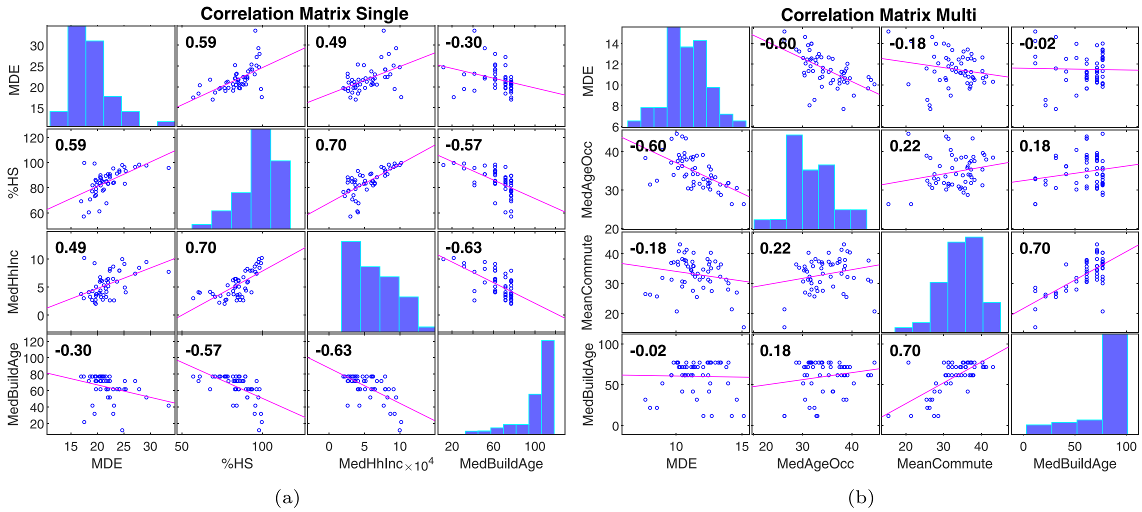


Fig. 12. Pearson's correlation between the mean daily electricity (MDE) demand (kWh) and zip code-level socioeconomic predictors for (a) single- and (b) multi-family residential users in Chicago.

With OLS assumptions verified, the best-fit models of annual single- and multi-family residential electricity consumption (Y_i), in kilowatt-hours, are the models shown in Eqs. (4) and (5):

Single-family residential average annual electricity consumption model:

$$\begin{aligned}\frac{Y_{i,single}^{\lambda-1}}{\lambda} = {} & 0.62 + 0.013(\text{Percent occupancy}) \\ & - 0.00031(\text{Median age of occupants}) \\ & + 0.00027(\text{Percent 65 and over}) \\ & + 0.00011(\text{Mean commute time}) \\ & + 0.000081(\text{Percent high school or higher}) \end{aligned} \tag{4}$$

where λ is -1.56 calculated with the Box Cox transformation.

Multi-family residential average annual electricity consumption model:

$$\begin{aligned} Y_{i,multi} = {} & 19 + 0.029(\text{Median building age}) - 1.90(\text{Percent multi-family}) \\ & + 0.14(\text{Percent female}) - 0.28(\text{Median age of occupants}) \\ & - 0.14(\text{Mean commute time}) \end{aligned} \tag{5}$$

Higher electricity demand of single-family residential users is associated with higher percentage occupancy, younger age of occupants, higher percent elderly, longer mean commute times, and higher education beyond high school. Similarly, higher multi-family residential electricity consumption is associated with older buildings, lower percent multi-family housing, higher percent female populations, younger median age, and shorter mean commute time, based on data at a zip code level. Eqs. (4) and (5), however, only explain about 62% and 41% of the variability in single- and multi-family residential electricity consumption, respectively, across Chicago zip codes. Many other factors, including personal behaviors and seasonal effects, influence residential electricity consumption but were not part of the analyzed data sets. Future applications could include the development of surveys to collect and aggregate data for electricity predictive models.

The relative importance analysis revealed different socioeconomic predictors that explain a large portion of the variation in single- versus multi-family residential electricity consumption. In the transformed single-family residential multiple linear regression model, the percent with high school or higher education accounted for over half of the explained variation in average electricity consumption, shown in Table 2 with a relative importance of 0.35 (of the multiple R^2 value of 0.65). The next most important socioeconomic predictor, in terms of relative importance, was the percent occupancy at 0.14. In the multi-family residential multiple linear regression model, the median age of occupants accounted for over three-fourths of the explained variation in average electricity consumption, shown in Table 3 with a relative importance of 0.36 (of the multiple R^2 value of 0.47). The remaining predictors in the multi-family residential model have similar relative importance, constituting the remaining one-fourth of the variation in average electricity consumption.

Table 5
Goodness-of-fit R^2 values for the training and testing data sets of the Multiple Linear Regression model (MLR) and Regression Trees (RT) models.

Single-family models		
R^2	Multiple linear regression	Regression tree
Train	0.67	0.49
Test	0.33	0.14
Multi-family models		
R^2	Multiple linear regression	Regression tree
Train	0.49	0.31
Test	0.22	0.08

4.6. Regression trees

We trained two regression trees using the mean daily electricity demand and all the socioeconomic variables to compare regression trees' performance with the single- and multi-family multiple linear regression models. We partitioned the data into training (70%) and testing (30%) sets, resulting in training and testing sets with 35 and 15 records, respectively. A random sampling algorithm ensuring no overlapping was used to partition the data. Based on previous studies (Pesantez et al., 2020) to evaluate the effects of random sampling, the data partition was re-initialized for each of the 100 trials. The goodness-of-fit R^2 values for training and testing sets reported by regression trees and multiple linear regression models are shown in Table 5.

According to the goodness-of-fit R^2 calculated for the training and testing data sets (Table 5), the multiple linear regression models outperform regression trees for both single- and multi-family residential electricity models. However, these results come with the nuance of a limited size data set. Regression tree models require more data to learn and generalize better than mechanistic models. The most important

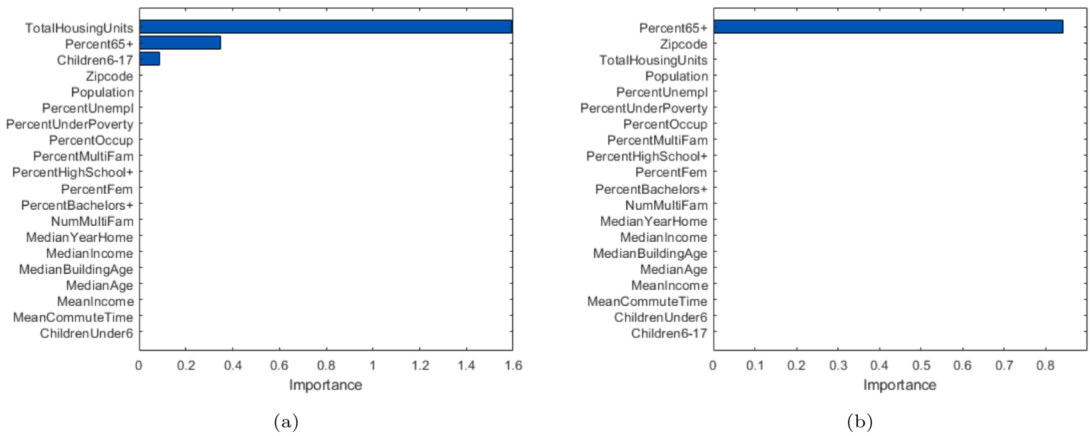


Fig. 13. Feature importance reported by the regression tree models for the mean daily electricity demand of (a) single- and (b) multi-family residential users.

predictors identified by the regression tree models for single-family users were the total number of housing units followed by the percent of people of ages 65 and over, and the number of children between 6 and 17 years old (Fig. 13(a)). Similarly, for the multi-family users, the most important predictor was reported as only the percent of people of ages 65 and over (Fig. 13(b)).

5. Discussion: Implications and data limitations

We used smart electricity meter data to first model the daily profiles of electricity consumption and then aggregate those profiles to the mean daily electricity (MDE) demand over a year at the zip code level to analyze electricity with socioeconomic data retrieved from the American Community Survey (ACS). The daily electricity profiles across zip codes can improve the understanding of when high and low electricity consumption occurs and how these values change across seasons over a year. These results can also be used to compare how extreme events, such as changes in population dynamics and consumption behaviors across cities (e.g., in response to the COVID-19 pandemic), affected the electricity demand exerted by single- and multi-family residential accounts, similar to findings in the transportation, water, and grid-scale power generation sectors (Goenaga, Matini, Karanam, & Underwood, 2021; Pesantez, Alghamdi, Sabu, Mahinthakumar, & Berglund, 2022; Roidt, Chini, Stillwell, & Cominola, 2020; Zechman Berglund, et al., 2021). For example, in studying the COVID-19 pandemic, Spearing, et al. (2021) found that building-scale water use was tightly coupled to occupancy, while energy was less tied to occupancy and presented unique challenges in uncertain conditions of occupant behaviors. Using the results from our analysis of single- and multi-family residential electricity consumption, our models could be used to compare to current electricity demands to understand impacts of such changes in consumption patterns and demands across different geographic areas of a city with different socioeconomic contexts.

Modeling electricity demand with data from single- and multi-family residential users provides insights about differences in consumption time and magnitude. Although several predictors are statistically significant in the models and explain an important percentage of the mean daily electricity demand, these models have limitations. While using zip code-level socioeconomic data can give an estimate of the demographics of the data sample, these data do not reflect individual households' characteristics to comply with data security protocols. Socioeconomic information is also estimated on an annual scale, such that there are limitations in using U.S. Census data to understand sub-annual variation in single- and multi-family residential electricity consumption. Therefore, models' validation using door-to-door surveys is out of the scope of this work due mainly to anonymity requirements stipulated in the data–sharing agreement.

Our study describes and quantifies the 24-hour mean electricity profiles of the two types of residential users. Then, the model aggregates these profiles by taking their average to get the mean daily electricity (MDE) demand over a year. MDE is analyzed with socioeconomic predictors that are reported annually and vary across zip codes. Including other exogenous predictors, such as weather variables, is limited by a lack of sufficient zip code-level weather data across the study area. Furthermore, since our data set includes only one year of electricity demand data, it is infeasible to perform a multi-annual analysis with weather variables.

While the regression analyses include significant variables that explain the mean daily electricity demand of single- and multi-family accounts, this study presents additional limitations. Past studies have found that electricity demand is heavily correlated to building area (Cárdenas-Mamani et al., 2022; Kennedy, et al., 2015; Pincetl et al., 2016). Our data set does not include this predictor as it was unavailable from the ACS database. Furthermore, our model includes data aggregation, which has two effects on the analysis. First, temporal aggregation of electricity demand from 30-minute to mean daily values hides consumption patterns that utilities could use to identify peak consumption times during the day. Second, spatial aggregation from account to zip code level prevents spatial analysis that could inform electricity providers of specific areas with irregular consumption.

Data available on a zip code level can be useful for analysis; however, aggregate data can present limitations. Zip codes can include very different geographic ranges depending on whether the area is rural or urban, and may also consist of two discontinuous areas. Information grouped by Census block can be better for analysis of demographic and socioeconomic data (Grubesic, 2008). However, these different aggregated scales can lead to different conclusions. For example, Harris and Liu (1993) found that income did not have a statistically significant effect on U.S. residential electricity consumption for the period 1969–1990, while others have demonstrated statistically-significant relationships between income and electricity consumption (Berrill et al., 2021; Elnakat et al., 2016; Hayn et al., 2014). While a reasonable percentage of daily electricity demand is explained by our multiple linear regression models with zip code-scale predictors, model results might improve by exploring socioeconomic data at the household level; however, this approach could raise privacy concerns about data sharing from the household occupants.

Profiles description and predictive modeling have direct links to at least two of the Sustainable Development Goals (SDGs). Understanding the time of consumption and electricity demand trends and magnitude gives the user a valuable tool to tailor their consumption during the day. Hence, this work assists utility managers and city planners towards SDG 7 "By 2030, ensure universal access to affordable, reliable and modern energy services". Furthermore, with a predictive model, users can make informed decisions about future electricity demand usage across different geographical areas of the City of Chicago. This predictive model can also help users and providers achieve SDG 11 "By 2030, ensure access to adequate, safe, and affordable basic services" (Clark & Wu, 2016).

6. Conclusions

We analyzed data from Commonwealth Edison (ComEd) smart electricity meters in Chicago in conjunction with data collected from the U.S. Census ACS for 2016 to compare differences in single- and multi-family residential electricity consumption profiles across zip codes. Using these data, we created statistically significant multiple linear regression models for both single- and multi-family residential customers to predict the mean daily annual electricity consumption using socioeconomic characteristics as explanatory variables. Even for the restricted scope of the City of Chicago, sufficient social variation existed across zip codes and single- and multi-family electricity consumption to visualize differences in daily load profiles and create statistically relevant models. Our results show that both types of residential end-users present the same consumption trend. However, single-family electricity consumption exceeds multi-family consumption throughout the day and across the week, based on measured 2016 consumption. From a spatial perspective, the results show differences in electricity load profiles across zip codes for both types of residential users, and those differences are present concerning time-of-day and season of the year. Electricity consumption from multi-family residential customers located in zip codes reporting high income is generally lower than electricity reported from multi-family residential customers in lower income zip codes, which could reflect the presence of new and energy-saving appliances, weatherization, or other efficiency measures, raising important energy justice questions (Cong, Nock, Qiu, & Xing, 2022; Reames, 2016). Finally, in areas with primarily multi-family users, demand profiles show that zip codes reporting low median age consume more electricity than zip codes with higher median age values. Urban planners can use these analyses to develop tailored demand management strategies in a diverse and densely populated urban environment.

The findings show that more populated areas do not necessarily represent the highest demand for electricity values. These results agree with previous works about electricity demand at sub-city levels (Pincetl et al., 2016). Furthermore, when analyzing the demand per capita exerted by different Chicago zip codes, the highest residential demand occurs in the downtown area for both the single- and multi-family accounts. It is worth noting that the most important predictor for single-family accounts was the percentage of people with a high school degree or higher. For the multi-family accounts, the most important predictor was the median age of occupants. Information at the census block or zip code level was the key to finding these relationships between electricity demand and socioeconomic characteristics and has the potential to become even more insightful when gathering data at multi-annual resolution and weather variables can be included.

This research demonstrates that socioeconomic characteristics of zip codes can be used to explain a significant amount of variation in electricity consumption for single- and multi-family homes in Chicago using a multiple linear regression model. We used a linear regression model to represent the statistical relationship between socioeconomic explanatory variables and the mean daily electricity demand over the year of 2016. The characteristics and size of the data set predictors were not large enough to train a suitable regression tree model. The single-family multiple linear regression model showed a goodness-of-fit (adjusted R^2) of 0.62 with statistically significant predictors including the percent occupancy, median age of occupants, percent age 65 and over, mean commute time, and percent of individuals with at least a high school degree, with percent with high school or higher education showing the highest relative importance in the model. The multi-family multiple linear regression model showed an adjusted R^2 of 0.41 and its significant predictors included the median building age, percent of multi-family units, percent female, median age of occupants, and mean commute time, with the median age of occupants showing the highest relative importance in the model. Socioeconomic variables reported at the zip code level explain an important portion of the mean daily residential electricity demand of the City of Chicago, demonstrating the utility in quantifying behavior, infrastructure, and socioeconomic factors regarding electricity consumption and management with smart meter data.

The results reported at the zip code level for both types of residential users level can be used to elaborate more descriptive statistics focusing on, for example, the zip code reporting the highest electricity consumption. Prioritizing specific zip codes could lead to demand management strategies using predictors at a finer spatial resolution (e.g., specific building characteristics) to foster electricity conservation policies and obtain more insightful models. Cities like Chicago have already implemented publicly available electricity consumption benchmarking programs. The models developed in this work may be applied to those data sets with specific building characteristics to improve the understanding of significant users and create awareness of the factors affecting single- and multi-family residential electricity demand.

CRediT authorship contribution statement

Jorge E. Pesantez: Created the single-family residential model, Regression tree analysis, Validation, Formal analysis, Writing. **Grace E. Wackerman:** Data curation, Created the multi-family residential model, Validation, Writing. **Ashlynn S. Stillwell:** Conducted the relative importance analysis, Formal analysis, Supervision, Writing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Models and code that support the findings of this study are available in the Supporting Information. Data representing daily average electricity consumption profiles are available in the Supporting Information and at https://stillwell.cee.illinois.edu/data/.

Acknowledgments

This work was supported by the Department of Civil and Environmental Engineering, USA and the Department of Electrical and Computer Engineering at the University of Illinois Urbana-Champaign, USA. Additional support was provided by the Center for Infrastructure Resilience in Cities as Livable Environments (CIRCLE) through the Zhejiang University - University of Illinois Urbana-Champaign Joint Research Center, USA Project No. DREMES202001, funded by Zhejiang University. Commonwealth Edison electricity data were obtained through a partnership with the Environmental Defense Fund, under a data sharing agreement with the University of Illinois.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.scs.2022.104250.

References

Abrahamse, W., & Steg, L. (2009). How do socio-demographic and psychological factors relate to households' direct and indirect energy use and savings? *Journal of Economic Psychology*, *30*(5), 711–720. http://dx.doi.org/10.1016/J.JOEP.2009.05.006.

Abreu, P. H., Silva, D. C., Amaro, H., & Magalhães, R. (2016). Identification of Residential Energy Consumption Behaviors. *Journal of Energy Engineering*, *142*(4), Article 04016005. http://dx.doi.org/10.1061/(ASCE)EY.1943-7897.0000340.

Afaifia, M., Djiar, K. A., Bich-Ngoc, N., & Teller, J. (2021). An energy consumption model for the Algerian residential building's stock, based on a triangular approach: Geographic information system (GIS), regression analysis and hierarchical cluster analysis. *Sustainable Cities and Society*, *74*, Article 103191. http://dx.doi.org/10.1016/j.scs.2021.103191, URL https://www.sciencedirect.com/science/article/pii/S2210670721004698.

Ahkamiraad, A., & Wang, Y. (2018). An Agent-Based Model for Zip-Code Level Diffusion of Electric Vehicles and Electricity Consumption in New York City. *Energies*, *11*(3), 640. http://dx.doi.org/10.3390/EN11030640, URL https://doaj.org/article/55df1df41dd045a197442afc9bf3b2b6.

Alberini, A., Prettico, G., Shen, C., & Torriti, J. (2019). Hot weather and residential hourly electricity demand in Italy. *Energy*, *177*, 44–56.

Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, *81*, 1192–1205. http://dx.doi.org/10.1016/J.RSER.2017.04.095.

Andersen, F. M., Larsen, H. V., & Boomsma, T. K. (2013). Long-term forecasting of hourly electricity load: Identification of consumption profiles and segmentation of customers. *Energy Conversion and Management*, *68*, 244–252.

Avancini, D. B., Rodrigues, J. J., Martins, S. G., Rabêlo, R. A., Al-Muhtadi, J., & Solic, P. (2019). Energy meters evolution in smart grids: A review. *Journal of Cleaner Production*, *217*, 702–715.

Aydinalp-Koksal, M., & Ugursal, V. I. (2008). Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector. *Applied Energy*, *85*(4), 271–296. http://dx.doi.org/10.1016/J.APENERGY.2006.09.012.

Bartels, R., Fiebig, D. G., Garben, M., & Lumsdaine, R. (1992). An end-use electricity load simulation model: Delmod. *Utilities Policy*, *2*(1), 71–82.

Beccali, M., Cellura, M., Brano, V. L., & Marvuglia, A. (2008). Short-term prediction of household electricity consumption: Assessing weather sensitivity in a Mediterranean area. *Renewable and Sustainable Energy Reviews*, *12*(8), 2040–2065.

Beckel, C., Sadamori, L., Staake, T., & Santini, S. (2014). Revealing household characteristics from smart meter data. *Energy*, *78*, 397–410.

Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson correlation coefficient. In *Noise reduction in speech processing* (pp. 1–4). Springer.

Berrill, P., Gillingham, K. T., & Hertwich, E. G. (2021). Linking housing policy, housing typology, and residential energy demand in the United States. *Environmental Science and Technology*, *55*(4), 2224–2233.

Bettignies, Y., Meirelles, J., Fernandez, G., Meinherz, F., Hoekman, P., Bouillard, P., et al. (2019). The scale-dependent behaviour of cities: A cross-cities multiscale driver analysis of Urban energy use. *Sustainability*, *11*(12), 3246.

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *26*(2), 211–243.

Brounen, D., Kok, N., & Quigley, J. M. (2012). Residential energy use and conservation: Economics and demographics. *European Economic Review*, *56*(5), 931–945. http://dx.doi.org/10.1016/J.EUROECOREV.2012.02.007.

Cárdenas-Mamani, Ú., Kahhat, R., & Vázquez-Rowe, I. (2022). District-level analysis for household-related energy consumption and greenhouse gas emissions: A case study in Lima, Peru. *Sustainable Cities and Society*, *77*, Article 103572.

Chen, M., Ban-Weiss, G. A., & Sanders, K. T. (2018). The role of household level electricity data in improving estimates of the impacts of climate on building electricity use. *Energy and Buildings*, *180*, 146–158.

Chen, M., Ban-Weiss, G. A., & Sanders, K. T. (2020). Utilizing smart-meter data to project impacts of urban warming on residential electricity use for vulnerable populations in Southern California. *Environmental Research Letters*, *15*(6), Article 064001.

Chen, M., Sanders, K. T., & Ban-Weiss, G. A. (2019). A new method utilizing smart meter data for identifying the existence of air conditioning in residential homes. *Environmental Research Letters*, *14*(9), Article 094004.

Chicco, G. (2012). Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy*, *42*(1), 68–80. http://dx.doi.org/10.1016/J.ENERGY.2011.12.031.

Chou, J.-S., & Tran, D.-S. (2018). Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. *Energy*, *165*, 709–726.

City of Chicago (2016). Energy Benchmarking Report. https://www.chicago.gov/city/en/progs/env/building-energy-benchmarking-transparency.html. (Accessed 19 Aguest 2021).

Clark, H., & Wu, H. (2016). The sustainable development goals: 17 goals to transform our world. In *Furthering the work of the United Nations; UN: New York, NY, USA* (pp. 36–54).

Cong, S., Nock, D., Qiu, Y. L., & Xing, B. (2022). Unveiling hidden energy poverty using the energy equity gap. *Nature Communications*, *13*, 1–12.

Corbett, J., Wardle, K., & Chen, C. (2018). Toward a sustainable modern electricity grid: The effects of smart metering and program investments on demand-side management performance in the US electricity sector 2009–2012. *IEEE Transactions on Engineering Management*, *65*(2), 252–263.

Do, H., & Cetin, K. S. (2018). Residential Building Energy Consumption: A Review of Energy Data Availability, Characteristics, and Energy Performance Prediction Methods. *Current Sustainable/Renewable Energy Reports 2018 5:1*, *5*(1), 76–85. http://dx.doi.org/10.1007/S40518-018-0099-3, URL https://link.springer.com/article/10.1007/s40518-018-0099-3.

dos Santos, H. T. M., & Balestieri, J. A. P. (2018). Spatial analysis of sustainable development goals: a correlation between socioeconomic variables and electricity use. *Renewable and Sustainable Energy Reviews*, *97*, 367–376.

Durbin, J., & Watson, G. S. (1950). Testing for serial correlation in least squares regression: I. *Biometrika*, *37*(3/4), 409–428.

Elma, O., & Selamoğullar, U. S. (2017). A survey of a residential load profile for demand side management systems. In *2017 IEEE international conference on smart energy grid engineering* (pp. 85–89). IEEE.

Elnakat, A., & Gomez, J. D. (2015). Energy engenderment: An industrialized perspective assessing the importance of engaging women in residential energy consumption management. *Energy Policy*, *82*(1), 166–177. http://dx.doi.org/10.1016/J.ENPOL.2015.03.014.

Elnakat, A., Gomez, J. D., & Booth, N. (2016). A zip code study of socioeconomic, demographic, and household gendered influence on the residential energy sector. *Energy Reports*, *2*, 21–27. http://dx.doi.org/10.1016/J.EGYR.2016.01.003.

Eurostat (2020). House or flat: where do you live? - Products Eurostat News - Eurostat. URL https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20200513-1.

Facchini, A., Kennedy, C., Stewart, I., & Mele, R. (2017). The energy metabolism of megacities. *Applied Energy*, *186*, 86–95.

Fan, H., MacGill, I., & Sproul, A. (2017). Statistical analysis of drivers of residential peak electricity demand. *Energy and Buildings*, *141*, 205–217.

Farahbakhsh, H., Ugursal, V., & Fung, A. (1998). A residential end-use energy consumption model for Canada. *International Journal of Energy Research*, *22*(13), 1133–1143.

Fiebig, D. G., Bartels, R., & Aigner, D. J. (1991). A random coefficient approach to the estimation of residential end-use load profiles. *Journal of Econometrics*, *50*(3), 297–327.

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Thousand Oaks CA: Sage, URL https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Frayssinet, L., Merlier, L., Kuznik, F., Hubert, J.-L., Milliez, M., & Roux, J.-J. (2018). Modeling the heating and cooling energy demand of Urban buildings at city scale. *Renewable and Sustainable Energy Reviews*, *81*, 2318–2327.

Gerpott, T. J., & Paukert, M. (2013). Determinants of willingness to pay for smart meters: An empirical analysis of household customers in Germany. *Energy Policy*, *61*, 483–495. http://dx.doi.org/10.1016/J.ENPOL.2013.06.012.

Goenaga, B., Matini, N., Karanam, D., & Underwood, B. (2021). Disruption and recovery: Initial assessment of COVID-19 traffic impacts in North Carolina and Virginia. *Journal of Transportation Engineering Part A*, *147*(4), Article 06021001.

Gouveia, J. P., & Seixas, J. (2016). Unraveling electricity consumption profiles in households through clusters: Combining smart meters and door-to-door surveys. *Energy and Buildings*, *116*, 666–676. http://dx.doi.org/10.1016/J.ENBUILD.2016.01.043.

Gouveia, J. P., Seixas, J., & Mestre, A. (2017). Daily electricity consumption profiles from smart meters-proxies of behavior for space heating and cooling. *Energy*, *141*, 108–122.

Grubesic, T. H. (2008). Zip codes and spatial analysis: Problems and prospects. *Socio-Economic Planning Sciences*, *42*(2), 129–149. http://dx.doi.org/10.1016/J.SEPS.2006.09.001.

Guo, Z., Zhou, K., Zhang, C., Lu, X., Chen, W., & Yang, S. (2018). Residential electricity consumption behavior: Influencing factors, related theories and intervention strategies. *Renewable and Sustainable Energy Reviews*, *81*, 399–412.

Harris, J. L., & Liu, L. M. (1993). Dynamic structural analysis and forecasting of residential electricity consumption. *International Journal of Forecasting*, *9*(4), 437–455. http://dx.doi.org/10.1016/0169-2070(93)90072-U.

Hayn, M., Bertsch, V., & Fichtner, W. (2014). Electricity load profiles in Europe: The importance of household segmentation. *Energy Research & Social Science*, *3*(C), 30–45. http://dx.doi.org/10.1016/J.ERSS.2014.07.002.

Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D., & Zareipour, H. (2020). Energy forecasting: A review and outlook. *IEEE Open Access Journal of Power and Energy*, *7*, 376–388.

Howard, B., Parshall, L., Thompson, J., Hammer, S., Dickinson, J., & Modi, V. (2012). Spatial distribution of urban building energy consumption by end use. *Energy and Buildings*, *45*, 141–151. http://dx.doi.org/10.1016/J.ENBUILD.2011.10.061.

Hu, T., Yoshino, H., & Jiang, Z. (2013). Analysis on urban residential energy consumption of hot summer & cold winter zone in China. *Sustainable Cities and Society*, *6*, 85–91.

Huang, W.-H. (2015). The determinants of household electricity consumption in Taiwan: Evidence from quantile regression. *Energy*, *87*, 120–133.

Huebner, G. M., Hamilton, I., Chalabi, Z., Shipworth, D., & Oreszczyn, T. (2015). Explaining domestic energy consumption–the comparative contribution of building factors, socio-demographics, behaviours and attitudes. *Applied Energy*, *159*, 589–600.

Im, J., Seo, Y., Cetin, K. S., & Singh, J. (2017). Energy efficiency in U.S. residential rental housing: Adoption rates and impact on rent. *Applied Energy*, *205*, 1021–1033. http://dx.doi.org/10.1016/J.APENERGY.2017.08.047.

Jain, R. K., Smith, K. M., Culligan, P. J., & Taylor, J. E. (2014). Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. *Applied Energy*, *123*, 168–178.

Jovanović, S., Savić, S., Bojić, M., Djordjević, Z., & Nikolić, D. (2015). The impact of the mean daily air temperature change on electricity consumption. *Energy*, *88*, 604–609.

Karatasou, S., Laskari, M., & Santamouris, M. (2018). Determinants of high electricity use and high energy consumption for space and water heating in European social housing: Socio-demographic and building characteristics. *Energy and Buildings*, *170*, 107–114. http://dx.doi.org/10.1016/J.ENBUILD.2018.04.019.

Kavousian, A., Rajagopal, R., & Fischer, M. (2013). Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior. *Energy*, *55*, 184–194. http://dx.doi.org/10.1016/J.ENERGY.2013.03.086.

Kennedy, C. A., Stewart, I., Facchini, A., Cersosimo, I., Mele, R., Chen, B., et al. (2015). Energy and material flows of megacities. *Proceedings of the National Academy of Sciences*, *112*(19), 5985–5990.

Kim, M. K., Kim, Y.-S., & Srebric, J. (2020a). Impact of correlation of plug load data, occupancy rates and local weather conditions on electricity consumption in a building using four back-propagation neural network models. *Sustainable Cities and Society*, *62*, Article 102321.

Kim, M. K., Kim, Y.-S., & Srebric, J. (2020b). Predictions of electricity consumption in a campus building using occupant rates and weather elements with sensitivity analysis: Artificial neural network vs. linear regression. *Sustainable Cities and Society*, *62*, Article 102385. http://dx.doi.org/10.1016/j.scs.2020.102385, URL https://www.sciencedirect.com/science/article/pii/S2210670720306065.

Kochański, M., Korczak, K., & Skoczkowski, T. (2020). Technology innovation system analysis of electricity smart metering in the European Union. *Energies*, *13*(4), 916.

Liddle, B., & Huntington, H. (2021). How prices, income, and weather shape household electricity demand in high-income and middle-income countries. *Energy Economics*, *95*, Article 104995.

Liddle, B., & Lung, S. (2010). Age-structure, urbanization, and climate change in developed countries: revisiting STIRPAT for disaggregated population and consumption-related environmental impacts. *Population and Environment*, *31*(5), 317–343.

Liu, S., Zeng, A., Lau, K., Ren, C., wai Chan, P., & Ng, E. (2021). Predicting long-term monthly electricity demand under future climatic and socioeconomic changes using data-driven methods: A case study of Hong Kong. *Sustainable Cities and Society*, *70*, Article 102936. http://dx.doi.org/10.1016/j.scs.2021.102936, URL https://www.sciencedirect.com/science/article/pii/S2210670721002225.

McLoughlin, F., Duffy, A., & Conlon, M. (2012). Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study. *Energy and Buildings*, *48*, 240–248.

McLoughlin, F., Duffy, A., & Conlon, M. (2015). A clustering approach to domestic electricity load profile characterisation using smart metering data. *Applied Energy*, *141*, 190–199. http://dx.doi.org/10.1016/J.APENERGY.2014.12.039.

Meyabadi, A. F., & Deihimi, M. H. (2017). A review of demand-side management: Reconsidering theoretical framework. *Renewable and Sustainable Energy Reviews*, *80*, 367–379.

NMHC (2019). NMHC | Household Characteristics. URL https://www.nmhc.org/research-insight/quick-facts-figures/quick-facts-resident-demographics/household-characteristics/.

Noor, S., Yang, W., Guo, M., van Dam, K. H., & Wang, X. (2018). Energy demand side management within micro-grid networks enhanced by blockchain. *Applied Energy*, *228*, 1385–1398.

Obringer, R., Kumar, R., & Nateghi, R. (2019). Analyzing the climate sensitivity of the coupled water-electricity demand nexus in the Midwestern United States. *Applied Energy*, *252*, Article 113466.

Ota, T., Kakinaka, M., & Kotani, K. (2018). Demographic effects on residential electricity and city gas consumption in the aging society of Japan. *Energy Policy*, *115*, 503–513.

Parti, M., & Parti, C. (1980). The total and appliance-specific conditional demand for electricity in the household sector. *Bell Journal of Economics*, 309–321.

Pesantez, J. E., Alghamdi, F., Sabu, S., Mahinthakumar, G., & Berglund, E. Z. (2022). Using a digital twin to explore water infrastructure impacts during the covid-19 pandemic. *Sustainable Cities and Society*, *77*, 103520.

Pesantez, J. E., Berglund, E. Z., & Kaza, N. (2020). Smart meters data for modeling and forecasting water demand at the user-level. *Environmental Modelling & Software*, *125*, Article 104633.

Pincetl, S., Graham, R., Murphy, S., & Sivaraman, D. (2016). Analysis of high-resolution utility data for understanding energy use in urban systems: The case of Los Angeles, California. *Journal of Industrial Ecology*, *20*(1), 166–178.

R Core Team (2017). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, URL https://www.R-project.org/.

Raza, M. Q., & Khosravi, A. (2015). A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. *Renewable and Sustainable Energy Reviews*, *50*, 1352–1372.

Reames, T. G. (2016). Targeting energy justice: Exploring spatial, racial/ethnic and socioeconomic disparities in Urban residential heating energy efficiency. *Energy Policy*, *97*, 549–558.

Rhodes, J. D., Cole, W. J., Upshaw, C. R., Edgar, T. F., & Webber, M. E. (2014). Clustering analysis of residential electricity demand profiles. *Applied Energy*, *135*, 461–471.

Roidt, M., Chini, C. M., Stillwell, A. S., & Cominola, A. (2020). Unlocking the impacts of COVID-19 lockdowns: Changes in thermal electricity generation water footprint and virtual water trade in Europe. *Environmental Science & Technology Letters*, *7*(9), 683–689.

Sakah, M., du Can, S. d. l. R., Diawuo, F. A., Sedzro, M. D., & Kuhn, C. (2019). A study of appliance ownership and electricity consumption determinants in urban Ghanaian households. *Sustainable Cities and Society*, *44*, 559–581.

Sanchez-Sellero, M.-C., & Sanchez-Sellero, P. (2019). Variables determining total and electrical expenditure in Spanish households. *Sustainable Cities and Society*, *48*, Article 101535.

Sandels, C., Widén, J., Nordström, L., & Andersson, E. (2015). Day-ahead predictions of electricity consumption in a Swedish office building from weather, occupancy, and temporal data. *Energy and Buildings*, *108*, 279–290.

Scofield, J. H., & Doane, J. (2018). Energy performance of LEED-certified buildings from 2015 Chicago benchmarking data. *Energy and Buildings*, *174*, 402–413.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3–4), 591–611. http://dx.doi.org/10.1093/biomet/52.3-4.591.

Shehadi, M. (2018). Energy consumption optimization measures for buildings in the Midwest regions of USA. *Buildings*, *8*(12), 170.

Son, H., & Kim, C. (2017). Short-term forecasting of electricity demand for the residential sector using weather and social variables. *Resources, Conservation and Recycling*, *123*, 200–207.

Spearing, L. A., Tiedmann, H. R., Sela, L., Nagy, Z., Kaminsky, J. A., Katz, L. E., et al. (2021). Human–Infrastructure interactions during the COVID-19 pandemic: Understanding water and electricity demand profiles at the building level. *ACS ES&T Water*, Article ASAP.

Spence, A., Demski, C., Butler, C., Parkhill, K., & Pidgeon, N. (2015). Public perceptions of demand-side management and a smarter energy future. *Nature Climate Change*, *5*(6), 550–554.

Suganthi, L., & Samuel, A. A. (2012). Energy models for demand forecasting—A review. *Renewable and Sustainable Energy Reviews*, *16*(2), 1223–1240.

Taylor, R. (1990). Interpretation of the correlation coefficient: a basic review. *Journal of Diagnostic Medical Sonography*, *6*(1), 35–39.

Tonidandel, S., & LeBreton, J. M. (2011). Relative importance analysis: A useful supplement to regression analysis. *Journal of Business and Psychology*, *26*(1), 1–9.

Tran, L. N., Gao, W., Novianto, D., Ushifusa, Y., & Fukuda, H. (2021). Relationships between household characteristics and electricity end-use in Japanese residential apartments. *Sustainable Cities and Society*, *64*, Article 102534.

Tukey, J. W. (1953). *The problem of multiple comparisons*. Princeton University, Chapman and Hall.

U. S. Census Bureau (2016). U.S. Census Bureau QuickFacts: Chicago city, Illinois. URL https://www.census.gov/quickfacts/chicagocityillinois.

US EPA (2021). Electricity Customers | US EPA. URL https://www.epa.gov/energy/electricity-customers.

Van den Brom, P., Hansen, A. R., Gram-Hanssen, K., Meijer, A., & Visscher, H. (2019). Variances in residential heating consumption–Importance of building characteristics and occupants analysed by movers and stayers. *Applied Energy*, *250*, 713–728.

Van Rossum, G., & Drake, F. L. (2009). *Python 3 reference manual*. Scotts Valley, CA: CreateSpace.

vom Scheidt, F., Medinová, H., Ludwig, N., Richter, B., Staudt, P., & Weinhardt, C. (2020). Data analytics in the electricity sector–A quantitative and qualitative literature review. *Energy and AI*, *1*, Article 100009.

Walfish, S. (2006). A review of statistical outlier methods. *Pharmaceutical Technology*, *30*(11), 82.

Wang, Y., Chen, Q., Kang, C., Zhang, M., Wang, K., & Zhao, Y. (2015). Load profiling and its application to demand response: A review. *Tsinghua Science and Technology*, *20*(2), 117–129. http://dx.doi.org/10.1109/TST.2015.7085625.

Won, C., No, S., & Alhadidi, Q. (2019). Factors affecting energy performance of large-scale office buildings: Analysis of benchmarking data from New York City and Chicago. *Energies*, *12*(24), 4783.

Wyatt, P. (2013). A dwelling-level investigation into the physical and socio-economic drivers of domestic energy consumption in England. *Energy Policy*, *60*, 540–549.

Yildiz, B., Bilbao, J. I., Dore, J., & Sproul, A. B. (2017). Recent advances in the analysis of residential electricity consumption and applications of smart meter data. *Applied Energy*, *208*, 402–427. http://dx.doi.org/10.1016/J.APENERGY.2017.10.014.

Yoon, J. H., Baldick, R., & Novoselac, A. (2016). Demand response control of residential HVAC loads based on dynamic electricity prices and economic analysis. *Science and Technology for the Built Environment*, *22*(6), 705–719.

Zechman Berglund, E., Thelemaque, N., Spearing, L., Faust, K. M., Kaminsky, J., Sela, L., et al. (2021). Water and wastewater systems and utilities: Challenges and opportunities during the COVID-19 pandemic. *Journal of Water Resources Planning and Management*, *147*(5), Article 02521001.

Zeileis, A., & Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News*, *2*(3), 7–10, URL https://CRAN.R-project.org/doc/Rnews/.

Zhang, M., Chen, Y., Hu, W., Deng, N., & He, W. (2021). Exploring the impact of temperature change on residential electricity consumption in China: The 'crowding-out'effect of income growth. *Energy and Buildings*, *245*, Article 111040.

Zhang, C., Liao, H., & Mi, Z. (2019). Climate impacts: temperature and electricity consumption. *Natural Hazards*, *99*(3), 1259–1275.

Zheng, J., Gao, D. W., & Lin, L. (2013). Smart meters in smart grid: An overview. In *2013 IEEE green technologies conference* (pp. 57–64). IEEE.