

Data-Driven Analysis of Influential Factors on Residential Energy End-use in the US

Biao Kuang ^a, Chelsea Schelly ^b, Gaby Ou ^c, Mostafa Sahraei-Ardakani ^d, Shardul Tiwari ^b, Jianli Chen ^{a*}

^a Department of Civil and Environmental Engineering, The University of Utah, Salt Lake City, UT 84102, USA

^b Department of Social Sciences, Michigan Technological University, Houghton, MI 49931, USA

^c Department of Civil and Coastal Engineering, University of Florida, Gainesville, FL 32603, USA

^d Department of Electrical and Computer Engineering, The University of Utah, Salt Lake City, UT 84112, USA

Abstract

Decarbonizing the residential buildings sector is important to realize a net-zero future. However, little research comprehensively explores how residential energy by end-uses, i.e., space heating, space cooling, water heating, and appliances, vary across different climate, building, and occupant characteristics. Therefore, based on Residential Energy Consumption Survey 2015, this study investigates how these energy end-uses correlated with different influential factors, i.e., climate, demographics, housing and appliance features. Recursively random forest regression and partial dependency plots were employed to analyze and quantify the impact of influencing factors on end-use energy usage. The developed models illustrate nonlinear and varying impacts of influential factors on different end-uses, with R-square from 0.481 (appliance) to 0.885 (space cooling). Specifically, housing size and climate are determinant factors of space heating/cooling, while family size and total appliance number determine water heating and appliance usage, respectively. Use frequency, size, and vintage of appliances (e.g., refrigerator, laundry) affect appliance energy consumption. Furthermore, one-way ANOVA confirms the statistically significant differences in energy consumption across varying household income groups, i.e., households with higher incomes consume more energy. Houses with older vintages use more total energy, especially in space heating, while less in cooling due to lower penetration of air conditioners. Finally, the analysis proves that residential end-use electrification, mainly space heating and water heating, is beneficial for energy efficiency. The findings contribute to understanding influential factors on different end-uses in households, which from a holistic level informs the potential pathway for building decarbonization, especially targeting at specific end-uses or occupant groups.

Keywords

Residential Energy Consumption; Influential Factors; Residential End-Use; Data-driven Analysis; Machine Learning

* Corresponding author.

E-mail address: jianli.chen@utah.edu (J. Chen)

1 Introduction

Multiple initiatives (e.g., zero-carbon buildings) have been implemented to reduce buildings' carbon emissions in the past few years (Aydin & Brounen, 2019; Duzgun et al., 2022), although the building sector still contributes to around 40% of total global energy consumption (Khalil et al., 2022). In the United States (US), the share of energy consumed by residential buildings has increased gradually, reaching 21% of total energy usage in 2021 (US EIA, 2022). US residential energy consumption accounts for 17% of global residential energy consumption (Fumo and Rafe Biswas, 2015). Consequently, US residential buildings show great energy saving and carbon reduction potential.

Residential buildings consume electricity, natural gas, and other end-use energy sources to provide various services, i.e., space heating, air conditioning (space cooling), water heating, and appliances (Fell, 2017; Swan and Ugursal, 2009). Analyzing these energy end-uses is crucial for the design of specific energy-conservation programs and system resiliency for the security of energy services (Tiwari et al., 2022). Compared with the research on total energy consumption, the main obstacle to analyzing energy end-use, which corresponds to the range of energy services utilized in residential homes, is the disaggregation of total energy usage at the end-use level due to measuring difficulty of utility meters (González-Torres et al., 2022).

Generally, several techniques can be used to break down total energy consumption for different energy service use. The first is the sensor-based disaggregation method. Although this method has high accuracy based on distributed direct sensors, the installation and maintenance cost is non-negligible (Larsen and Nesbakken, 2004), which impedes its application on a large scale. Another approach is statistical modeling (i.e., data-based method), e.g., conditional demand analysis (CDA) and regression methods. However, CDA has limited capability in modeling end-use energy consumption, in that only a few variables are allowed to adjust in the model (Aydinalp-Koksal and Ugursal, 2008). Regression, especially linear regression, is widely applied due to its straightforward interpretation but has poor performance in explaining the nonlinear impacts (Kim et al., 2020). Additionally, engineering models (i.e., model-based analyses) are commonly used to decompose total energy consumption. This method does not rely on historical consumption data but is based on occupant behaviors, appliance ratings, building characteristics, etc. (Swan and Ugursal, 2009). The energy consumption simulated by engineering models can precisely match the actual energy by obtaining the details of the building and occupants (Zhao, 2012). Since 2015, Residential Energy Consumption Survey (RECS), a national survey administered by US Environment Information Administration (EIA), started adopting engineering models to disaggregate total energy use. Furthermore, RECS 2015 was conducted and validated through a cross-sectional household survey and an energy supplier survey to ensure the quality of energy disaggregation (US EIA, 2015a). Hence, RECS is reliable for a detailed analysis of residential energy end-uses.

Numerous studies have been conducted on building total energy use and corresponding influencing factors based on various datasets, including RECS, but few studies focus on energy services differentiated by energy end-use. Estiri (2014) examined the direct and indirect impacts of building and household characteristics on residential total energy consumption based on RECS 2015. Likewise, Karatasou & Santamouris (2019) used RECS 2015 to model the relationship between socioeconomic status (e.g., income, housing size, and the number of rooms) and total energy consumption. Debs & Metzinger (2022) also analyzed RECS 2015 but compared the energy consumption and the influencing factors in different climate zones. Furthermore, a few studies are based on particular energy fuels, mainly electricity.

Sanquist et al. (2012) analyzed how household demographics and energy consumption behavior affect residential electricity consumption using multiple regression modeling based on RECS 2005. Iraganaboina & Eluru (2021) compared influencing factors on residential energy use from the perspective of different energy sources, including electricity, natural gas, and fuel oil. In contrast, fewer studies analyzed end-use energy consumption. Matsumoto (2016) only explored the impacts of household characteristics on electrical appliance energy consumption in Japan. Similarly, in a recent study, Zhu et al. (2023) focused on the contributors to space cooling energy use in China. RECS can be applied to analyze end-use consumption as well. Min et al. (2010) and Li (2014) adopted RECS 2005 and RECS 2009 to model end-use energy consumption using linear regression; however, the influencing factors considered are incomprehensive, mainly household and housing characteristics.

Understanding the differences in energy end-uses (i.e., space heating/cooling, water heating, and appliances, which provide different energy services needed at home) and their driving factors is the prerequisite to developing energy efficiency programs for residential buildings and, ultimately, conserving energy consumption. However, the research on energy end-uses is limited. Although it has been revealed that household demographics, housing characteristics, home appliances, and climate affect residential energy consumption (Estiri, 2014), most former studies are from the perspective of total household energy consumption, and the incorporated factors are incomprehensive. How these factors affect each end-use energy consumption remains unclear. Hence, comprehensive investigations and comparisons of the influencing factors across various residential energy end-uses are required.

Therefore, this study aims to explore and compare the driving factors of residential energy consumption from the perspective of end-uses based on RECS 2015. The remainder of this paper is organized as follows: section 2 reviews existing studies on the influencing factors of energy consumption. Section 3 describes the data screening and processing and introduces the used methods. In section 4, the results of descriptive analysis and fitted regression models are presented. Then, section 5 discusses and compares the driving factors and their impacts across different end-uses. Finally, section 6 draws up conclusions.

2 Literature Review

2.1 Influencing Factors of Total Residential Energy Use

Various factors contribute to residential energy consumption, including climate and geography, household demographics, housing and appliance characteristics. One of the most studied areas is the influence of occupant demographics on energy use. Household income and family size are generally identified as the determining factors that drive energy consumption, although the findings on the impacts of demographics are inconsistent among existing studies (Frederiks et al., 2015; Mohr, 2018; Wang et al., 2020). Age is considered another contributor to energy consumption due to elderly residents' frugal energy consumption behaviors (Chen et al., 2013). Besides these, household members' education levels and employment status are recognized as potential factors (Jones and Lomas, 2015; Karatasou and Santamouris, 2019; Kumar et al., 2022; Tiwari et al., 2022). In particular, these demographics reflect the socio-economic status of households (Chen et al., 2013). A few studies demonstrated that socio-economic status indirectly affects energy consumption by influencing housing choice, e.g., housing type, number of appliances, and housing size (Estiri, 2014; Tso and Guan, 2014).

Another key driving factor to energy consumption is housing, which can be classified into housing and appliance characteristics (Jones et al., 2015). Housing size, measured by total square footage or heated/cooled area, significantly contributes to energy consumption (Karatasou and Santamouris, 2019). Generally, larger housing sizes are associated with higher energy usage. The total number of rooms is related to housing size and positively affects energy consumption (Iraganaboina and Eluru, 2021; Wang et al., 2021). Besides, housing vintage (i.e., the built year of the house) is another influencing factor; however, its impact is inconsistent across existing studies. The housing of old vintage has less electricity usage in Bedir et al. (2013), while Bartusch et al. (2012) found that newly constructed houses consume less electricity. Also, the impacts of housing vintage on different energy fuels (e.g., electricity and natural gas) are different (Brounen et al., 2012). Additionally, energy consumption varies by housing type (Brounen et al., 2012; Kaza, 2010).

Compared with other factors, fewer studies have included appliance characteristics in their investigation (Jones et al., 2015). Bartusch et al. (2012) proposed that energy consumption cannot be fully explained by household and housing characteristics alone. For appliance characteristics, the number of appliances effectively predicts energy consumption (Mo and Zhao, 2021). Iraganaboina & Eluru (2021) analyzed the number of some specific appliances, e.g., desktops, clothes dryers, refrigerators, etc. Similarly, Karatasou & Santamouris, (2019) found that the number of appliances, measured as the sum of the number of refrigerators, freezers, dishwashers, clothes washers, dryers, TVs, personal computers, and ACs, has a positive correlation with energy consumption. Use frequency and size of common appliances, e.g., AC, laundry, personal computer, and TV, also greatly impact energy consumption (Sanquist et al., 2012; Genjo et al., 2005). Similarly, Steemers and Yun (2009) and Ruan et al. (2017) concluded that appliance usage behavior, mainly AC, contributes to energy consumption. Additionally, appliance energy consumption is increased with aging (Paul et al., 2022); however, the impacts of appliances' vintage are not fully explored (Wang, 2017). Đurišić et al. (2020) only considered the average age of kitchen appliances, laundry, and HVAC. Chen et al. (2022) incorporated the age of appliances but only focused on the AC when analyzing the influencing factors among different income groups.

In addition, the climate and location of houses are recognized to affect residential energy consumption. It is proven that outdoor temperature affects residential energy consumption (Fumo and Rafe Biswas, 2015). Some studies adopted heating or cooling degree days (HDD/ CDD) to measure the climate of the areas. Strong correlations between energy consumption and HDD exist in the European Union (Tsemekidi et al., 2019) and the US (Li et al., 2022); that is, the colder the region, the higher the energy consumption. Besides HDD and CDD, climate zones are also used to classify regional climates. Differences in climate zones contribute to energy consumption gaps in total energy consumption (Mo and Zhao, 2021). Debs and Metzinger (2022) asserted that energy consumption's driving factors have varying effects in different climate zones. Furthermore, the effects of the housing location, usually classified into rural and urban, are identified in a few studies. There are some differences in energy consumption between urban and rural households (Chun-sheng et al., 2012). Zheng et al. (2014) found the urban-rural variance in energy consumption as well.

2.2 Residential End-use Consumption and Influencing Factors

The residential sector consumes energy for various end-uses or services, e.g., space heating, space cooling, domestic hot water, and appliances, for occupant living (Swan and Ugursal, 2009). These energy end-uses can be referred as energy services (Fell, 2017; González-Torres et al., 2022). Some studies

reported end-use energy consumption; however, they still studied the total energy consumption. For instance, Aydinalp et al. (2002) studied the total energy usage of appliances, lighting, and space cooling, which are major sectors of building energy use. Hence, this research still focuses on aggregated energy consumption. Moreover, although Wang et al. (2021) used “residential energy end-use” as their research topic, unfortunately, they mixed the concepts of energy end-use and total energy consumption. In this case, they have not disaggregated the total energy consumed to end-use level nor analyzed the influencing factors of each energy end-use.

Generally, total energy consumption can be obtained directly by energy meters or billings (Swan and Ugursal, 2009), while it is challenging to obtain the energy consumed by specific categories of end uses (González-Torres et al., 2022). Small-scale research always uses direct measurement to obtain end-use data. Lee et al. (2019) adopted gas or electricity meters to measure the annual end-use energy consumption data, i.e., heating, cooling, water heating, ventilation, electric appliances, and cooking. This study covered 71 households with less than 40 valid samples for most end-uses. In a similar manner, Xie & Noor (2022) collected 66 households’ annual residential end-use energy consumption data using an installed measurement system. Nevertheless, direct measurement is not feasible on large scales, and some statistical or engineering models are needed to disaggregate energy consumption (González-Torres et al., 2022). Among earlier studies, large-scale end-use energy consumption data were roughly estimated. Farahbakhsh et al. (1998) divided the total energy consumption into end uses by constant percentages of each end-use. Similarly, the cooling energy consumption in Aydinalp et al. (2002) was approximately estimated by a reported general frequency of AC usage. Afterward, more information, including housing characteristics, household demographics, and occupant behaviors, has been collected to disaggregate energy consumption, e.g., RECS 2015; hence, data quality is improved correspondingly (Jin et al., 2023).

Compared with the total energy consumption, less research is conducted on residential energy end-uses, and the factors considered when analyzing end-use energy consumption are incomplete. The end-use models developed by Ren et al. (2013) are based on occupancy patterns, requiring many details about occupants and their behaviors to predict energy consumption; however, this information is difficult to obtain in most cases (Wang, 2017). Min et al. (2010) and Li (2014) modeled the residential energy end-uses using RECS 2005 and RECS 2009, respectively. However, both studies’ influencing factors are incomplete, e.g., lacking housing appliances. Karatasou et al. (2018) only compared the impacts of housing and family size on the energy consumption of electric appliances, space heating, and water heater of 244 dwellings. Lee & Song (2022) analyzed the determinants of housing, socio-demographics, and appliance on end-use energy consumption based on 71 apartment households in South Korea. However, these studies are small-scale with insufficient variables under consideration.

Furthermore, a few studies focus on certain specific end-uses. Matsumoto (2016) examined the effects of households’ socioeconomic characteristics on the electric usage of 19 residential appliances in Japan. Huo et al. (2021) only studied space heating in Chongqing, China, and viewed electricity as the sole energy source. Other fuel types providing space heating and appliance characteristics are ignored in these two studies. Guerra Santin et al. (2009) investigated the influence of occupancy and building on the total energy consumption of space and water heating. Liu et al. (2021) considered multiple fuels to explore the main driving factors of space heating in China, including types of energy consumed, heating area, and climate, but used a small dataset (only around 200). Feng et al. (2016) pointed out the determinant impact of occupant behavior on space cooling usage based on survey and energy bills. However, the behavior involved is incomplete, with only a general cooling setpoint. Yun and Steemers

(2011) adopted RECS 2001 to analyze the impacts of behavior, physical, and socio-economic factors on residential energy consumption, but only space cooling usage was considered. Hence, the comprehensive analysis and comparisons of the influencing factors of each end-use energy consumption are limited and need further exploration.

Overall, climate, household demographics, housing characteristics, and appliance characteristics contribute to residential energy consumption. Existing studies on the influencing factors of energy consumption are mainly from the perspective of total energy consumption. Although a few studies have been conducted at the end-use level, they are not nationwide, and the influencing factors considered are incomplete.

3 Data and Methods

3.1 Data and Data Screening

RECS is a national survey conducted every five years by US EIA to collect detailed information on household energy-related data. RECS 2015, the most recently available energy consumption survey data, is adopted for this study (US EIA, 2015b). The housing characteristics (including physical housing characteristics, household demographics, energy-related behavior, and home appliance) of RECS 2015 were collected between August 2015 and April 2016, with 5686 households involved representing 118.2 million US households. Besides this, to collect the energy consumption and expenditures of the surveyed houses, EIA uses the energy supplier survey to gather household energy billing data from energy suppliers directly, including electricity, gas, and fuel oil (US EIA, 2015a). The energy billings cover the energy usage from September 2014 to April 2016.

Due to the cost and difficulty of large scale submetering of end-uses, the end-use energy consumption is decomposed from metering data of total household energy usage with calibration and validation based on the engineering model-based method, as commonly adopted method in large-scale building surveys (González-Torres et al., 2022; Swan and Ugursal, 2009). Specifically, building characteristics, demographics, climate, owned appliances, and occupant behavior were first collected in RECS 2015 (US EIA, 2015a). Based on this information, physics-based engineering models use published values of building and appliance parameters (e.g., Unit Energy Consumption [UEC], space heating/cooling load) and consider occupant behaviors, building characteristics, and equipment efficiency to derive energy usage for each category of end-uses. For example, coffee maker energy consumption is based on the average UEC value from published sources, the presence of coffee makers, and the use frequency of surveyed occupants. Space heating energy usage is modeled based on the equipment characteristics, reported usage behaviors, and required space heating load. After model-based decomposition, the end-use energy consumption is further calibrated and validated using the collected total energy bills of each fuel for the surveyed household, considering the uncertainties of end-use estimate and correlations between energy end-uses (e.g., clothes washers and cloth dryers). The detailed information can be found in (US EIA, 2015a).

The housing unit in the RECS was classified into 5 types, including mobile home, single-family detached house, and single-family attached house, among which single-family detached houses occupy the most part, with around 62%. To reduce the impact of different housing units (Estiri, 2014; Kaza, 2010), this study focuses on the single-family detached house. Also, considering the impact of bill

payment responsibilities (Best et al., 2021), this study further delimits the research samples only to consider the households responsible for paying all used energy bills, i.e., electricity, natural gas, propane, and fuel oil. Therefore, the final sample size is 3702.

3.2 Factors Selection Overview

Figure 1 shows the flowchart of factor selection. First, based on the literature review and the available factors in the RECS dataset, we preselect explainable variables for each end-use. These factors are processed and recoded for further analysis. Then, Spearman rank correlation analysis is employed to find the correlated factors with each end-use and total energy consumption, where the threshold is set as 0.2 (Akoglu, 2018). Next, these correlated factors are fed into random forest regression models, from which the variable importance (VI) scores are calculated, followed by a recursive feature elimination process (Darst et al., 2018). In this iterative step, we keep tuning the model and eliminating the variables with a VI score less than 0.01 until the VI scores of all factors in the regression model are larger than or equal to 0.01. Based on the final model, we obtain the determinants and corresponding VI scores.

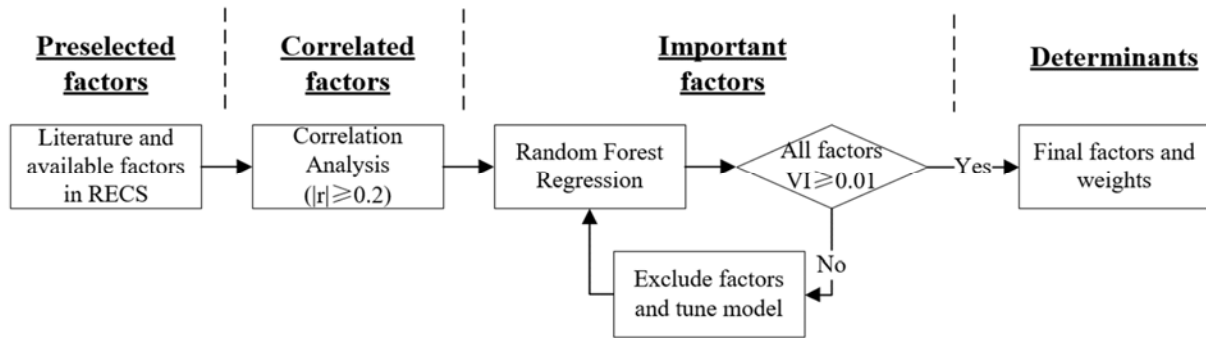


Figure 1 Variables Selection Process

3.3 Factors and Measurements

The preselected variables in the RECS can be classified into four types: climate and location, occupant demographics, building characteristics, and appliance characteristics (Rafsanjani, 2016), including categorical and numerical variables. The factors and descriptions are shown in Appendix Table 1. Some categorical variables are processed into ordinal or dummy variables.

Climate and Location We select the urban type of housing (Zhang et al., 2016), climate zones (Debs and Metzinger, 2022), and heating/cooling degree days (Nie et al., 2018) to describe the climate and location of the surveyed houses. Urban types and climate zones are converted into dummy variables. Specifically, UATYP10 is converted into UATYP10_R and UATYP10_U, among which each value of the two dummy variables equal to 1 means the house is in rural or urban areas, respectively; if both values equal 0, the house is in the urban cluster. Moreover, we use the Building American Climate Zone to classify the climate zones. The original variable (i.e., CLIMATE_REGION) is one-hot coded using “Cold/Very Cold”, “Hot-Dry/Mixed-Dry”, “Hot-Humid”, “Marine”, and “Mixed-Humid” to represent the five climate zones. Additionally, HDD65 and CDD65 are continuous numerical variables, measuring how cold/hot the temperature was in the area during 2015 based on 65F.

Household Demographics RECS 2015 did not collect the age, education levels, and employment status of all occupants in the household; therefore, the three variables are represented by the respondents' age (HHAGE), education (EDUCATION), and employment status (EMPLOYHH) (Estiri, 2014; Karatasou and Santamouris, 2019). Family size (NHSLDMEM), household income (MONEYPY), and the number of weekdays someone is at home (ATHOME) are driving factors commonly identified in existing studies as well (Debs and Metzinger, 2022; Tso and Guan, 2014). We recode the employment status with 0 representing not employed/retired, 0.5 representing employed part-time, and 1 representing employed full-time. In this way, household income, education, and employment status are ordinal variables. Family size, age, and the number of weekdays someone is at home are continuous variables.

Housing Characteristics Various housing characteristics, e.g., vintage, housing area, and the number of total rooms, influence residential energy consumption (Li et al., 2022; Wang, 2017; Wang et al., 2021). RECS uses square footage to measure the housing size, which is converted to square meters. In this study, three types of housing size variables, i.e., total cooled area in square meters (TOTCM2), total heated area in square meters (TOTHM2), and total housing area in square meters (TOTHOUSING), are used to model the space cooling, space heating, and other end-use energy consumption. Total number of rooms (TOTROOMS) in the housing unit is a numerical variable. Vintage (YEARMADERANGE), the number of stories (STORIES) and windows (WINDOWS), level of insulation (ADQINSUL), and frequency of drafty (DRAFTY) are ordinal variables, which we viewed as continual variables. ATTIC and CELLAR are dummy variables to show if the house has a basement or attic.

Appliance Characteristics Some characteristics of essential appliances, e.g., equipment type (Debs and Metzinger, 2022), vintage (Jones et al., 2015), usage frequency and behavior (Iraganaboina and Eluru, 2021; Tso and Guan, 2014), and fuel type (Mo and Zhao, 2021), are selected and encoded. Specifically, the ages of vital appliances (e.g., refrigerator, freezer, space heating, and air conditioner) are recoded by using 0-6 to indicate not applicable, less than 2 years old, 2-4 years old, 5-9 years old, 10-14 years old, 15-19 years old, and over 20 years old, respectively. Then, we use "0" to replace "-2" in the RECS's original codes of size and use frequency of appliance to represent not applicable. In this way, the size of refrigerators, freezers, TVs, and water heaters are ordinal variables. Larger values of these variables mean a larger size of the appliance. The use frequency of the cooktop and oven part of the stove, separate cooktop and oven, microwave, and TV are numerical variables. Furthermore, HVAC usage behavior, type of fuels used by appliances, and equipment type of appliances (e.g., stove, clothes dryer, main and secondary space heating, and water heating) are also dummy coded. Finally, we add another variable, "NUMAP", to represent the total number of appliances owned (Karatasou and Santamouris, 2019). It is measured by the total number of refrigerators, separate freezers, stoves, separate cooktops, separate ovens, microwaves, outdoor grills, toasters, toaster ovens, coffee makers, crockpots, food processors, rice cookers, blenders or juicers, dishwashers, laundry (i.e., clothes washers and dryers), televisions, computers (desktop, laptop, and tablet), printers, fans, humidifiers, and light bulbs.

End-use Energy Consumption This study uses kilowatt hours (kWh) as the energy unit to combine different fuels or energy sources on an equal basis. Except for electricity, RECS uses British thermal units (Btu) to quantify other consumed fuels, e.g., natural gas and propane, which are correspondingly converted to kWh[†]. The calculations for the end-use energy consumption are shown in equations (1)-(4).

[†] 1 kWh = 3412.14 Btu

Space cooling (SPCKWH) only consumes electricity, consisting of air conditioning, air handlers used for cooling, and evaporative coolers. In contrast, space heating (SPHKWH) can be provided by electricity, natural gas, propane, and fuel oil/kerosene. The electricity consumed by air handlers and boiler pumps for heating is incorporated into space heating. Water heating (WTHKWH) also consumes electricity, natural gas, and fuel oil. The remaining energy consumption is by appliances (APKWH). Some appliances only consume electricity, e.g., refrigerators, freezers, microwaves, clothes washers, and dishwashers, while other applications support multiple energy sources, e.g., cooking, clothes dryers, and swimming pool pumps and heaters. The total energy consumption is summed by RECS and called “TOTALBTU”, converted to kWh and named “TOTALKWH”. Finally, after testing, we transform the energy consumption of space heating, space cooling, appliance, and total energy consumption to log and use Btu for better fitting the model (Debs and Metzinger, 2022), while the results are reported in kWh for easier interpretations.

$$\text{SPCKWH} = \text{KWHCOL} + \text{KWHAHUCOL} + \text{KWHEVAPCOL} \quad (1)$$

SPCKWH: Total energy consumption by space cooling, in kWh

KWHCOL: Electricity usage for air conditioning (central systems and individual units), in kWh

KWHELAHUCOL: Electricity usage for air handlers used for cooling, in kWh

KWHELEVAPCOL: Electricity usage for evaporative coolers, in kWh

$$\text{SPHKWH} = \text{KWHSPH} + \text{KWHELAHUHEAT} + \text{KWHNGSPH} + \text{KWHLPSPH} + \text{KWHFOSPH} \quad (2)$$

SPHKWH: Total energy consumption by space heating, in kWh

KWHELSPH: Electricity usage for main and secondary space heating, in kWh

KWHELAHUHEAT: Electricity usage for air handlers and boiler pumps used for heating, in kWh

KWHNGSPH: Natural gas usage for main and secondary space heating, in kWh

KWHLPSPH: Propane usage for main and secondary space heating, in kWh

KWHFOSPH: Fuel oil/kerosene usage for main and secondary space heating, in kWh

$$\text{WTHKWH} = \text{KWHWTH} + \text{KWHNGWTH} + \text{KWHFOWTH} \quad (3)$$

WTHKWH: Total energy consumption by water heating, in kWh

KWHELWTH: Electricity usage for main and secondary water heating, in kWh

KWHNGWTH: Natural gas usage for main and secondary water heating, in kWh

KWHFOWTH: Fuel oil/kerosene usage for main and secondary water heating, in kWh

TOTALKWHWTH: Total usage for main and secondary water heating, in kWh

$$\begin{aligned} \text{APKWH} = & \text{KWHELRFG} + \text{KWHELFRZ} + \text{KWHELMICRO} + \text{KWHELWCW} + \text{KWHELDWH} + \text{KWHELLGT} + \text{KWHELTVREL} + \\ & \text{KWHELFCAN} + \text{KWHELDHUM} + \text{KWHELHUM} + \text{TOTALKWHCOK} + \text{TOTALKWHCDR} + \text{TOTALKWHPL} + \\ & \text{TOTALKWHHTB} + \text{TOTALKWHNEC} \end{aligned} \quad (4)$$

APKWH: Total energy consumption by appliances, in kWh

KWHELRFG: Electricity usage for all refrigerators, in kWh

KWHELFRZ: Electricity usage for freezers, in kWh

KWHELMICRO: Electricity usage for microwaves, in kWh

KWHELWCW: Electricity usage for clothes washers, in kWh

KWHELDWH: Electricity usage for dishwashers, in kWh
 KWHELLGT: Electricity usage for indoor and outdoor lighting, in kWh
 KWHELTVREL: Electricity usage for all televisions and related peripherals, in kWh
 KWHELCFAN: Electricity usage for ceiling fans, in kWh
 KWHELDHUM: Electricity usage for dehumidifiers, in kWh
 KWHELHUM: Electricity usage for humidifiers, in kWh
 TOTALKWHCOK: Total usage for cooking (stoves, cooktops, and ovens), in kWh
 TOTALKWHCDR: Total usage for clothes dryers, in kWh
 TOTALKWHPL: Total usage for swimming pool pumps and heaters, in kWh
 TOTALKWHNEC: Total usage for other devices and purposes not elsewhere classified, in kWh

3.4 Spearman Rank Correlation

Spearman rank correlation measures the degree of correlation between two variables by calculating the Pearson correlation coefficient between the ranked variables (see equation (5)). The Spearman correlation coefficient is denoted by r , whose value is between -1 and +1. Spearman correlation works well with nonlinear monotonic and ordinal data, as well as data with outliers (Schober et al., 2018). Correlation analysis can be used to select features before developing models by selecting correlated variables to reduce the risk of model retrofitting (Marques Ramos et al., 2020; Zhang et al., 2018). The larger the absolute coefficient, the stronger the correlation between the two variables. Generally, the absolute value of r greater than 0.2, 0.4, and 0.6 means that the correlation between the two variables is weak, moderate, and strong, respectively (Akoglu, 2018). To incorporate more variables into our models, we selected 0.2 as the threshold of correlation analysis. Variables with an absolute correlation coefficient value with end-use or total energy consumption larger than 0.2 are selected.

$$r(X, Y) = \frac{Cov(R(X), R(Y))}{\sqrt{Var(R(X))Var(R(Y))}} \quad (5)$$

Where $R(X)$ and $R(Y)$ are ranked variables X and Y ; $Cov(X, Y)$ is the covariance of ranked X and Y ; $Var(X, Y)$ is the variance between the ranked variables.

3.5 Radom Forest Regression

Random forest (RF) is an ensemble learning method based on the decision tree (Breiman, 2001). Ensemble learning is a technique to combine predictions from multiple machine learning algorithms for higher predictive performance than a single model (Sagi and Rokach, 2018). Random forest uses the Bagging (short for Bootstrap aggregating) sampling method to randomly extract k different sample datasets, as the sub-training sets of each decision tree (i.e., the single weak learner), from the original dataset with replacement. The size of each sample set is the same as that of the original dataset (denoted as m). With the increase of dataset size (m), around 37% of samples will not be selected for training due to sampling with replacement, see equation (6). The unselected sample of each Bagging sampling is called Out-of-Bag (OOB), which can be used to evaluate the performance of RF models, i.e., OOB score, measuring the model's R-squared in the out-of-bag samples (37% of the training dataset). Then, k sub-training sets are used to generate k decision trees correspondingly based on partial features randomly selected from total features. The training process is repeated until the random forest is generated by trees.

For regression task, the predicted result of the algorithm is the averaging predictions of individual trees. Therefore, four parameters are generally selected to tune the model using grid search method to obtain an optimal model:

- `n_estimators`: the number of trees in the forest.
- `max_depth`: the maximum depth of the tree.
- `min_samples_split`: the minimum number of samples required to split an internal node.
- `min_samples_leaf`: the minimum number of samples required at a leaf node.

$$\lim_{m \rightarrow \infty} \left(1 - \frac{1}{m}\right)^m \rightarrow \frac{1}{e} \approx 0.368 \quad (6)$$

Where m is the size of the original training dataset.

RF can solve classification and regression and allow for nonlinear relationships between explanatory and response variables, especially working well with high-dimensional problems (Darst et al., 2018). Also, random forest reduces the risk of overfitting due to the Bagging sampling method and only a small set of features selected to grow each tree (Ziegler and König, 2014). Besides, the random forest has the ability to assess variable importance (VI) scores. The VI score is calculated by comparing the prediction accuracy between a tree with and without the presence of this predictor variable (Chehreh Chelgani et al., 2016). In particular, some variables with small VI scores can be eliminated to improve the model's interpretability and reduce the risk of overfitting (Gregorutti et al., 2017). To ensure only important variables are incorporated in the random forest model, a recursive feature elimination is commonly employed in existing studies by tuning the model recursively and excluding variables with VI scores less than a threshold (Darst et al., 2018; Zhou et al., 2014). Random forest, like other machine learning methods, can also employ partial dependence (PD) plots to interpret the effects of independent variables in the fitted model on the predicted variable (Molnar et al., 2021). PD plots show the functional relationship between partial independent variables and predictions. The PD function is the marginal effect of the variable by calculating the average of the predictions for all combinations of other variables in the dataset (Jiang et al., 2022).

Based on these advantages, we selected RFR to build the energy consumption models for each end-use and total energy consumption. The screened dataset is randomly split into training and test datasets using the "train_test_split" Python package, where the training-test ratio is 0.3. Then, the "RandomForestRegressor" package is used to develop random forest regression models and calculate the variable importance (VI) scores. During the data split and training process, the random state is specified as 0. To improve the generalization ability of the fitted model, 10-fold cross-validation is used to train and tune the regression model for each energy consumption model (Shi et al., 2021). The cross-validation score is measured as the average of the R-squared of 10-fold training. The variables are selected recursively until the VI scores of all variables in the final model are larger than 0.01. Finally, we adopt the "pdpbox" package to plot the partial dependence of important variables in the fitted model to interpret their impact on energy consumption.

In addition to the Cross-validation score and OOB score, three additional metrics are adopted to evaluate the performance of the model: R-squared (R^2), root mean square error (RMSE), and mean average error (MAE). Their calculation formulas are shown in equations (7)-(9). R^2 score shows how well the developed model is fitted to the data by comparing it to the average line of the dependent variable. R^2

is upper bound by 1, representing perfect fit; a higher R^2 is considered desirable (Chicco et al., 2021). Similarly, the higher value of cross-validation score and OOB means better model fitness. MAE and RMSE measure the difference between predicted and true values; therefore, a lower value of MAE and RMSE implies higher model accuracy.

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (9)$$

Where $\hat{y}_i$, y_i , and $\bar{y}$ represent the predicted values, true values, and mean values of the true values of energy consumption.

4 Results

4.1 Descriptive Analysis

4.1.1 Household Demographics

The distributions of household demographic indicators are shown in Figure 2. Among 3702 surveyed households, around 40% of single-family households have 2 persons, followed by 3 persons and 1 person with around 17% and 16%. In terms of annual gross income, about half of the households have an annual gross household income of less than \$60,000, among which 20% of respondents earned \$20,000 - \$39,999 (the largest part). The youngest respondent is 18 years old, and the oldest is 85 years old, with an average age of 54.5 years old. Over 60% of respondents are between 40-70 years old.

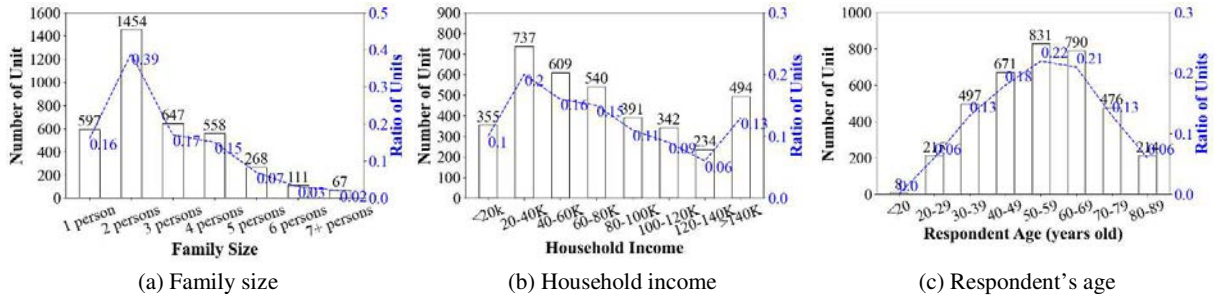


Figure 2 Distribution of Household Demographics

4.1.2 Housing Characteristics

Figure 3 shows the distributions of housing characteristics of all surveyed households. The numbers of houses built before 1950 and between 2000 and 2009 are almost the same, at about 610 (17%), followed by 1970-1979, 1990-1999, and 1980-1989. Over 72% of houses have 5-8 rooms, excluding bathrooms in the unit. The total housing area of the housing unit is a continuous variable. The minimum and maximum areas for single-family houses are 29.45 square meters (m^2) and 789.77 m^2 , respectively, with a mean value of around 238.77 m^2 and a standard deviation of 118.28 m^2 . The distribution of housing area shows that 21% of houses' areas are between 150-200 m^2 , followed by 100-150 m^2 and 200-250 m^2 . 61% of houses have only 1 story, and only 4% have 3 or more stories. 40% of houses have 10-15

windows, two times larger than those with 6-9 windows. Also, 86% of houses are well or adequately insulated. There are 56% of houses with an attic, 42% of houses with a basement, and only 1% of households use solar.

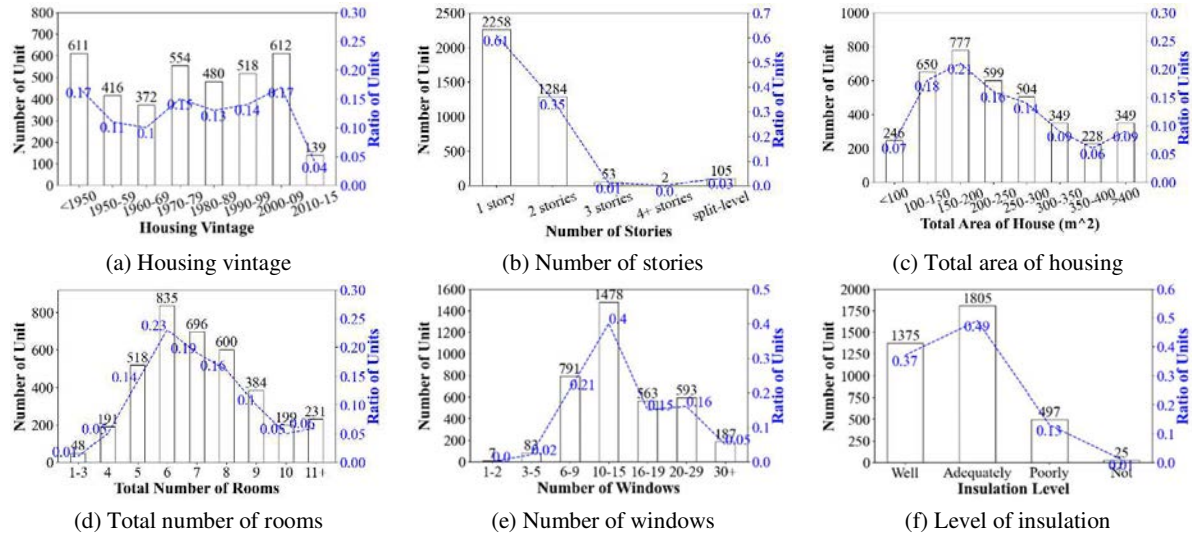


Figure 3 Distribution of Housing Characteristics

4.1.3 Appliance Characteristics

The total number of appliances owned is 25.8 on average, with a minimum of 3 and a maximum of 107. We analyzed the ownership of common appliances (see Figure 4), including freezers, dishwashers, and televisions. Among surveyed single-family households, almost all households have one or more refrigerators and televisions. 87% and 77% of households have at least 1 stove and 1 dishwasher, respectively. The proportion of households with clothes washers, clothes dryers, or microwaves at home is over 95%. Around 70% of households have outdoor grills, toasters, coffee makers, or laptops.

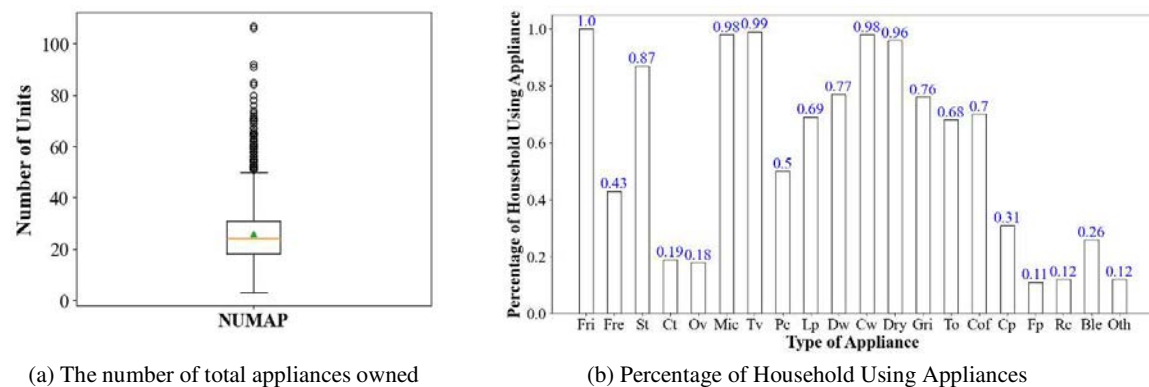


Figure 4 Distribution of Housing Appliances

(Fri: refrigerators; Fre: separate freezers; St: stoves; Ct: separate cooktops; Ov: separate ovens; Mic: microwave; Tv: television; Pc: desktop computers; Lp: laptop computers; Dw: dishwasher; Cw: clothes washer; Dry: clothes dryer; Gri: outdoor grill; To: toaster; Cof: coffee maker; Cp: crockpot or slow cooker; Fp: food processor; Rc: rice cooker; Ble: blender or juicer; Oth: other small appliance.)

In terms of HVAC (see Figure 5), most houses (68%+5%=73%) have a central conditioner, and 20% (15%+5%) have an individual wall/window or portable unit, while 12% of houses do not have an air

conditioner at home. In contrast, 97% of houses have space heating equipment, among which a central furnace is the most common (66%), four times larger than houses with a heat pump (15%). Also, around 46% of houses have secondary space heating, with 18% having a portable electric heater.

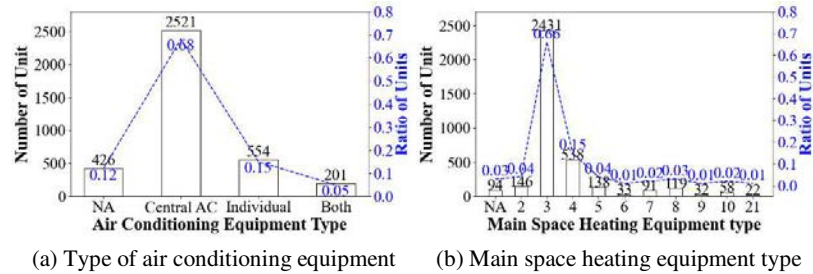


Figure 5 Type of Air Condition and Space Heating

(Figure (b-c): 2 Steam/hot water system with radiators or pipes; 3 Central furnace; 4 Heat pump; 5 Built-in electric units installed in walls, ceilings, baseboards, or floors; 6 Built-in floor/wall pipeless furnace; 7 Built-in room heater burning gas, oil, or kerosene; 8 Wood-burning stove (cordwood or pellets); 9 Fireplace; 10 Portable electric heaters; 21 Some other equipment.)

4.1.4 Climate and Location

The distribution of climate factors (Figure 6) shows that 36% of houses are in cold/very cold zones, which is three times more than the hot-dry/mixed-dry zone and twice larger than the hot-humid zone. The mean and median of CDD65 are 1686 and 1343, respectively, with 60% of houses' CDD65 between 500 and 2000. On the other hand, the mean and median of HDD65 are 3775 and 3922, respectively, with over 60% of houses' HDD65 above 3000. Additionally, we analyzed the urban type of houses: 64% in urban areas, compared with 25% in rural areas. The remaining 11% of houses are in the urban cluster.

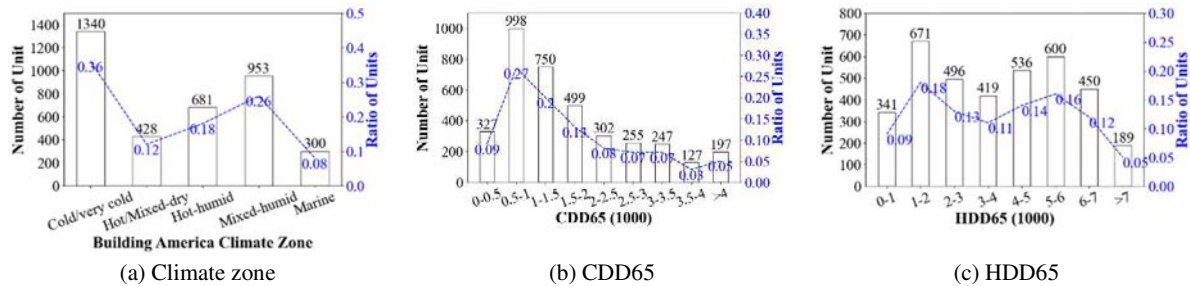
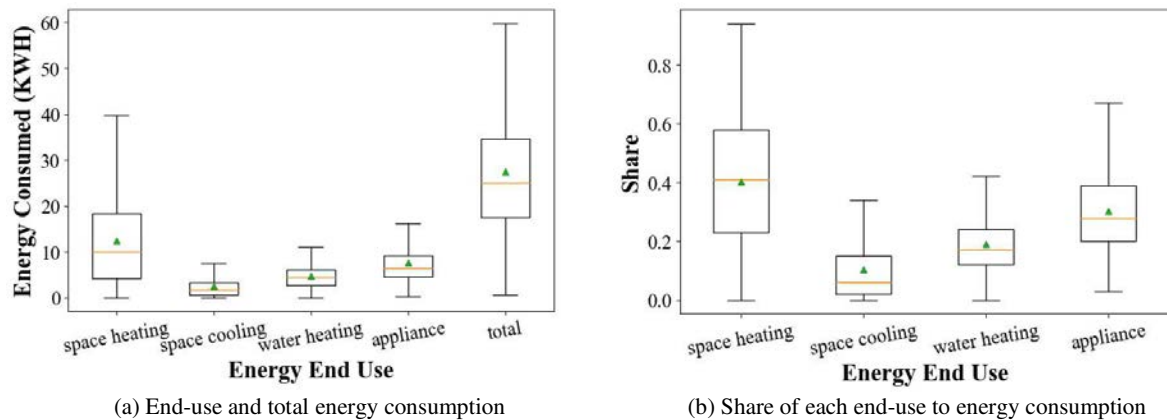


Figure 6 Distribution of Climate Factors

4.1.5 Energy Consumption

The distributions of end-use energy usage and corresponding share to total energy consumption are shown in Figure 7 and Table 1. The total energy consumption has the highest of 143.66 kWh and the lowest of 0.64 kWh, with an average of 27.38 kWh and a median of 25.09 kWh. There are 134, 374, and 7 households not consuming energy for space heating, space cooling, and water heating, respectively, while appliance usage is necessary for each household, with a minimum of 0.35 kWh. Generally, space heating consumes the largest part of the energy, with around 40% of total energy consumption, followed by appliance and water heating, while space cooling only consumes 10% of total energy.



(a) End-use and total energy consumption (b) Share of each end-use to energy consumption
Figure 7 Distribution of End-Use and Total Energy Consumption with Outlier Excluded

Table 1 Descriptive Analysis of Energy Usage (KWH) and Percentage of Each End-Use

End Use	Min.		Quartile		Median		3 Quartiles		Max.		Mean		STD.	
	Usage	Share	Usage	Share	Usage	Share	Usage	Share	Usage	Share	Usage	Share	Usage	Share
Space heating	0	0	4.20	23%	10.04	41%	18.49	58%	80.04	94%	12.48	40%	10.56	0.22
Space cooling	0	0	0.65	2%	1.61	6%	3.39	15%	21.78	82%	2.46	10%	2.63	0.12
Water heating	0	0	2.77	12%	4.34	17%	6.12	24%	21.65	80%	4.74	19%	2.70	0.10
Appliance	0.35	3%	4.60	20%	6.39	28%	9.22	39%	105.35	91%	7.71	30%	5.36	0.14
Total	0.64		17.53		25.09		34.63		143.66		27.38		13.67	

4.2 Correlation Analysis

Correlation analysis is conducted to select the variables for further regression modeling. The factors with an absolute value of correlation coefficient with end-use energy consumption or total energy consumption over 0.2 are selected, summarized in Table 2. We only consider the explainable factors. For example, although FUELH2O_1 and FUELH2O_5 (using natural gas and electricity for the main water heater) have coefficients of correlation with space heating over 0.2, they are not able to explain the energy consumption of space heating. Hence, both factors are excluded from the regression model for space heating. In a similar manner, TOTHM2, TOTCM2, and TOTHOUSING are used to model space heating, space cooling, and other end-uses, respectively. Space heating and cooling models are established with HDD65 and CDD65, respectively.

Correlation results demonstrate that the energy consumption of space heating positively correlates with the Cold/Very Cold climate zone and HDD65, while space cooling has positive correlations with the Hot-Humid climate zone and CDD65. The colder the area where the house is located, the higher the energy consumption of space heating. The hotter the area where the house is located, the higher the energy consumption of space cooling. In contrast, the correlations between appliance energy consumption and climate factors are poor, with coefficients less than 0.2. Also, the correlation coefficients between housing urban type and end-use and total energy consumption are small, with an absolute value of less than 0.2. Regarding household demographics, water heating energy consumption strongly correlates with family size (NHSLDMEM, $r=0.684$). Household income weakly correlates with water heating, appliance, and total energy consumption ($r \approx 0.3$). However, the correlations of space heating/ space cooling with demographics are negligible. Various housing characteristics, e.g., housing size and the number of total rooms and windows, have different degrees of correlation with energy consumption. In terms of the

appliance, although the appliances vary by end-use, in general, the age, fuel type, equipment type, and size of appliances correlate with the corresponding energy consumption.

Table 2 Correlated Factors with Each End-Use and Total Energy Consumption*

Use	Climate	Household	Housing	Appliance
SPH (15)	Cold/Very Cold (0.496), Hot/Mixed-Dry (-0.296), Hot-Humid (-0.402), HDD65 (0.649)	None	CELLAR (0.524), STORIES (0.292), TOTROOMS (0.257), WINDOWS (0.304), TOTHM2 (0.364)	EQUIPM_2 (0.201), EQUIPM_3 (0.364), EQUIPM_4 (-0.297), FUELHEAT_1 (0.436), FUELHEAT_3 (0.253), FUELHEAT_5 (-0.450)
SPC (9)	Cold/Very Cold (-0.449), Hot-Humid (0.488), Mixed-Humid (0.220), CDD65 (0.742)	None	CELLAR (-0.259), TOTCM2 (0.506)	AGECENAC (0.418), COOLTYPE_1 (0.531), USECENAC_1 (0.259)
WTH (12)	Hot-Humid (-0.243), CDD65 (-0.211), HDD65 (0.223)	MONEYPY (0.244), NHSLDMEM (0.684), HHAGE (-0.324)	TOTROOMS (0.211)	DWASHUSE (0.230), WASHLOAD (0.332), DRYRUSE (0.321), FUELH2O_1 (0.405), FUELH2O_5 (-0.445)
AP (22)	None	MONEYPY (0.356), NHSLDMEM (0.326)	YEARMADERANGE (0.213), TOTROOMS (0.367), WINDOWS (0.291), TOTHOUSING (0.296)	NUMAP (0.623), AGERFRI2 (0.381), AGEFRZR (0.224), DWASHUSE (0.414), WASHLOAD (0.420), DRYRUSE (0.436), SEPOVENUSE (0.229), AMTMICRO (0.223), SIZRFRI1 (0.300), SIZRFRI2 (0.397), TVSIZE1 (0.247), TVSIZE2 (0.323), TVONWD2 (0.264), TVONWE2 (0.276), DWCYCLE_2 (0.258), SIZFREEZ (0.249)
Total (26)	Cold/Very Cold (0.353), Hot-Dry/Mixed-Dry (-0.200), Hot-Humid (-0.242), HDD65 (0.434), CDD65 (-0.237)	MONEYPY (0.266), NHSLDMEM (0.230)	CELLAR (0.413), STORIES (0.332), TOTROOMS (0.391), WINDOWS (0.401), TOTHOUSING (0.454)	NUMAP (0.372), AGERFRI2 (0.261), DWASHUSE (0.272), WASHLOAD (0.227), DRYRUSE (0.248), SIZRFRI2 (0.266), TVSIZE2 (0.205), STOVENFUEL_5 (-0.203), EQUIPM_3 (0.303), EQUIPM_4 (-0.215), FUELHEAT_1 (0.391), FUELHEAT_5 (-0.412), FUELH2O_1 (0.296), FUELH2O_5 (-0.359)

Note: *All correlations are significant at the 0.01 level (2-tailed).

Abbreviation: SPH: space heating, SPC, space cooling, WTH: water heating, AP: appliance (the same below).

MONEYPY: Annual gross household income, NHSLDMEM: number of household members, HHAGE: household age.

CELLAR: with cellar, STORIES: number of stories, TOTROOMS: total number of rooms, WINDOWS: number of windows, TOTHM2: total heated area, TOTCM2: total cooled area, TOTHOUSING: total housing area.

EQUIPM_2: steam/hot water system with radiators or pipes for main space heating, EQUIPM_3: central furnace for main space heating, EQUIPM_4: heat pump for main space heating, FUELHEAT_1: natural gas from underground pipes as main space heating fuel, FUELHEAT_3: fuel oil as main space heating fuel, FUELHEAT_5: electricity as main space heating fuel, AGECEAC: age of central air conditioner, COOLTYPE_1: central air conditioning system, USECENAC_1: set one temperature and leave it there most of the time, DWASHUSE: frequency of dishwasher use, WASHLOAD: frequency of clothes washer use, FUELH2O_1: natural gas for water heater, FUELH2O_1: electricity for water heater, NUMAP: total number of owned appliances, AGERFRI2: age of second most-used refrigerator, AGEFRZR: age of most-used freezer, SEPOVENUSE: frequency of separate oven use, AMTMICRO: frequency of microwave use, SIZRFRI1: size of most-used refrigerator, SIZRFRI2: size of second most-used refrigerator, TVSIZE1: size of most-used TV, TVSIZE2: size of second most-used TV, TVONWD2: second most-used TV usage on weekdays, TVONWE2: second most-used TV usage on weekends, DWCYCLE_2: normal or default cycle with heated dry using dishwasher, SIZFREEZ: size of most-used freezer, STOVENFUEL_5: electricity for stove.

Furthermore, the correlations between demographics and housing characteristics are analyzed, as shown in Table 3. Household annual income positively correlates with housing size and the number of appliances, with r around 0.4. Housing vintage has slight correlations with housing size ($r \approx 0.2$). Also, a correlation between family size and the number of total appliances owned can be found. The number of total rooms strongly correlates with housing size as well, with a coefficient over 0.5.

Table 3 Correlations between Housing Characteristics and Household Demographics**

Variables	1	2	3	4	5	6	7	8	9
1. Housing income (MONEYPY)	-								
2. Family size (NHSLDMEM)	0.262	-							
3. Education level (EDUCATION)	0.494	0.051	-						
4. Housing vintage (YEARMADERANGE)	0.221	0.131	0.176	-					
5. Number of rooms (TOTROOMS)	0.398	0.201	0.285	0.174	-				
6. Total cooled area-square meters (TOTCM2)	0.379	0.130	0.267	0.350	0.481	-			
7. Total heated area-square meters (TOTHM2)	0.384	0.125	0.263	0.231	0.602	0.668	-		
8. Total housing area-square meters (TOTHOUSING)	0.366	0.131	0.252	0.178	0.592	0.608	0.863	-	
9. Number of appliances (NUMAP)	0.468	0.379	0.330	0.224	0.450	0.372	0.402	0.387	-

** All correlations are significant at the 0.01 level (2-tailed).

4.3 Determinants of End-Use and Total Energy Consumption

Based on correlation results, correlated variables are selected to train preliminary models using default parameters for each energy consumption model. Then, the models are tuned recursively using grid search methods and eliminating unimportant variables. The parameters and final model performance (e.g., obb score and R squared) for the end-use and total energy consumption are shown in Table 4.

Table 4 Parameter Setting and Model Performance for Energy Consumption Models

	Model	Log SPH	Log SPC	WTH	Log AP	Log Total
	Number of factors included	7	2	4	21	19
Parameter	n_estimators	92	40	257	273	93
	max_depth	7	5	5	11	12
	min_samples_split	6	7	19	19	14
	min_samples_leaf	1	3	19	3	5
Model Performance	Cross-validation score	0.760	0.876	0.650	0.478	0.538
	Obb Score	0.741	0.862	0.655	0.471	0.520
	R squared	0.761	0.885	0.643	0.481	0.542
	RMSE	6.577	2.782	1.625	3.924	8.832
	MAE	4.472	0.953	1.019	2.284	6.589

4.3.1 Space Heating

7 important variables are incorporated in the energy consumption model of space heating. The R-square is 0.761, demonstrating that the developed regression model can explain around 76% of the variability in space heating energy consumption. Combined with other metrics, the developed model for predicting space heating energy consumption is reliable.

The VI scores of these variables are shown in Figure 8. We plot the partial dependence of the important variables to visualize their marginal effects on space heating energy consumption, see Figure 9. Since the model is the logarithm of space heating energy consumption in Btu, the partial dependence plots are converted to the true energy consumption value in kWh by exponent arithmetic. Among these influencing factors, total heated area (TOTHM2) is the most important for predicting the energy consumption of space heating, followed by heating degree days based on 65F (HDD65), with 66.45% and

13.60% of VI scores, respectively. Both factors positively affect the energy consumption of space heating, and their corresponding impact ranges on energy consumption are around 9 kWh. The energy consumed for space heating grows sharply when TOTHM2 changes from 50 to 150 m². Then, the increasing trend gradually slows down and reaches the minimum when the TOTHM2 is larger than 400 m², after which the energy consumption almost keeps unchanged. Likewise, as HDD65 increases from 1000 to 5000, the energy consumption goes up dramatically from 0.5 to 8.5 kWh. After that, HDD65 has little effect on energy consumption. Furthermore, some characteristics related to space heating equipment affect energy consumption. If the main space heating equipment is EQUIPM_3 (central furnace), the energy consumption of space heating will increase by 2.6 KWH. The fuel type of main space heating also determines the space heating energy consumption. The households using FUELHEAT_1 (natural gas) or FUELHEAT_3 (fuel oil) as the space heating fuel type consumes more energy than those using FUELHEAT_5 (electricity). Additionally, more windows contribute to higher energy consumption for space heating, especially when the number of windows increases from 10-15 to 16-19.

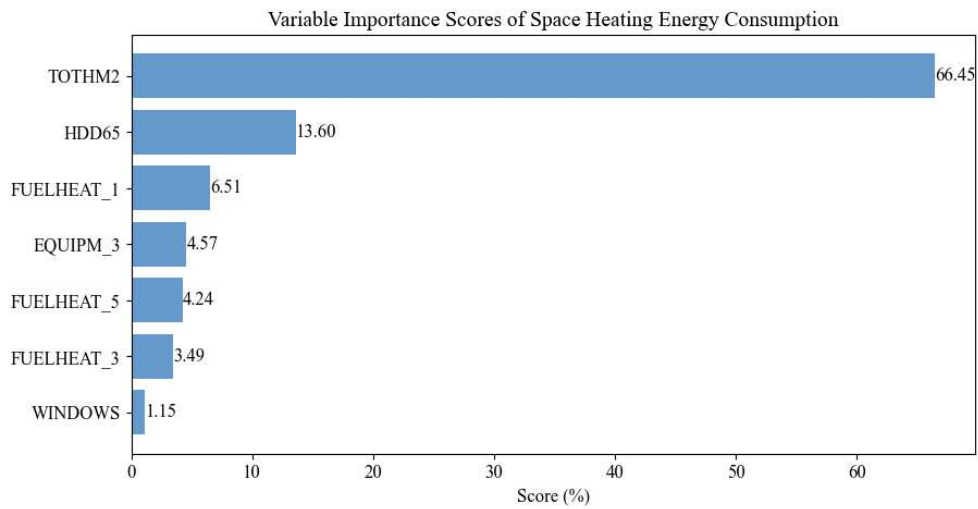
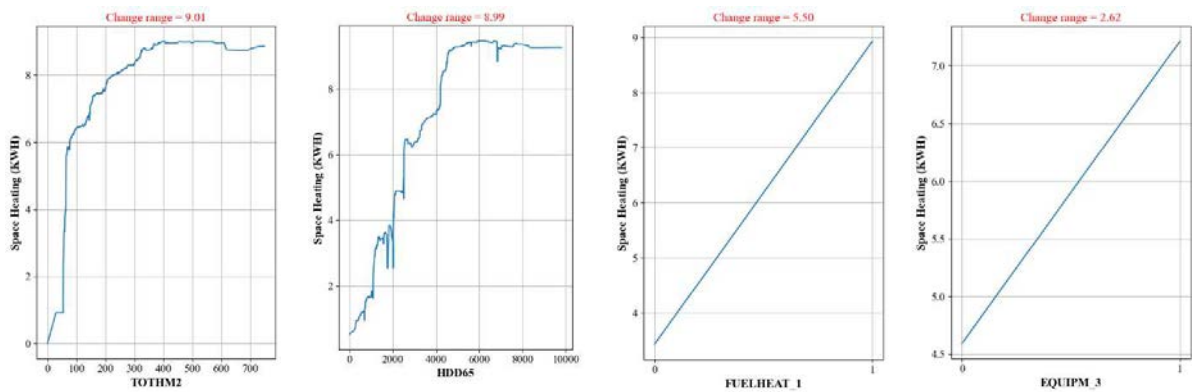


Figure 8 Rank of Variable Importance of Space Heating



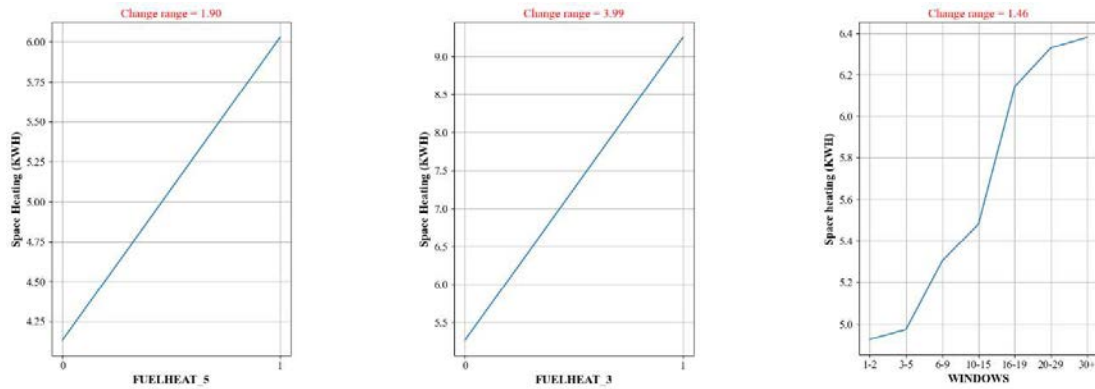


Figure 9 Marginal effects of important factors on space heating energy consumption

4.3.2 Space Cooling

Total cooled area (TOTCM2) and cooling degree days based on 65F (CDD65) are determinants for space cooling energy consumption. The percentage of variance explained by this model on energy consumption for space cooling is 88.5%, which is the highest among the four end-uses.

Figure 10 and Figure 11 show the VI scores and partial dependency of the two variables. TOTCM2 determines the energy consumption of space cooling, with a VI score of 87.10%. The energy consumed increases dramatically from 0 to 2 kWh with the changes of TOTCM2 from 0 to 200 m². For residential buildings with TOTCM2 larger than 250 m², the impact of TOTCM2 on space cooling becomes insignificant. CDD65 has a positive impact on space cooling energy consumption as well. When the CDD65 is less than 3000, the energy consumption of space cooling rises considerably, although there are some fluctuations. The impact range on the energy consumption of space cooling caused by CDD65 is 1.83 kWh.

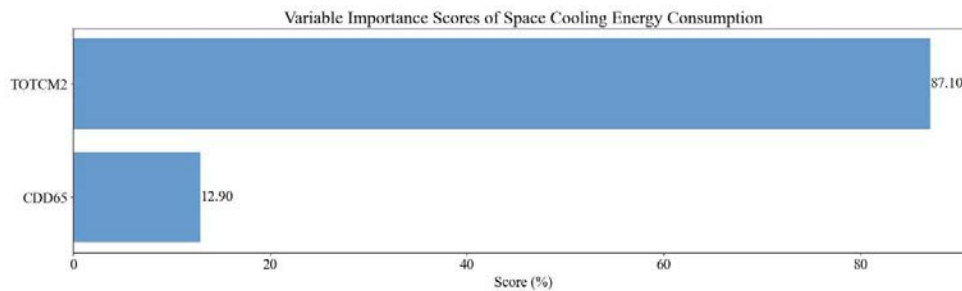


Figure 10 Variable importance scores of space cooling energy consumption

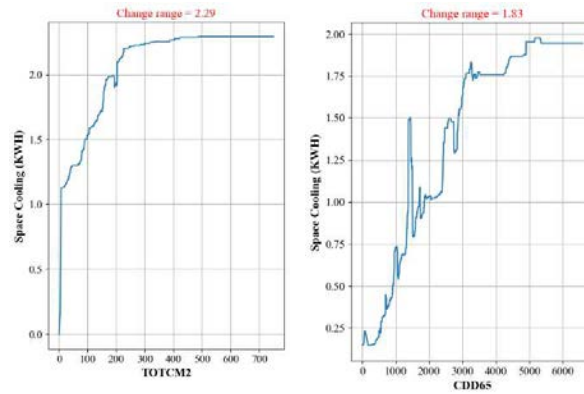


Figure 11 Marginal Effects of Important Factors on Space Cooling Energy Consumption

4.3.3 Water Heating

There are 4 important variables included in the final regression model of water heating energy consumption, i.e., family size (NHSLDMEM), electricity used for the main water heater (FUElh2O_5), and heating and cooling degree day (HDD65 and CDD65). Around 64% of the variability in water heating energy consumption can be explained by the fitted model.

The VI scores and partial dependency plots of three variables (see Figure 12 and Figure 13) illustrate that NHSLDMEM is a determinant of water heating energy consumption, with a VI score of 66.45% and an impact range of 5.94 kWh. The energy consumed for water heating increases gradually with the family size between 1 and 4. Then, the upward trend of energy consumption slows down until the difference in energy consumed between households with 7 persons or more becomes not obvious. The increase in HDD65 is also associated with rising water heating energy use, especially when HDD65 changes from 500 to 4000. Contrarily, FUElh2O_5 and CDD65 have negative impacts on water heating energy consumption. Households using electricity for water heating consume 2.17 kWh less than those using other energy sources, e.g., natural gas. The higher CDD65 is associated with less energy consumed for water heating, which decreases by around 0.44 kWh, with CDD65 increasing from 500 to 3500.

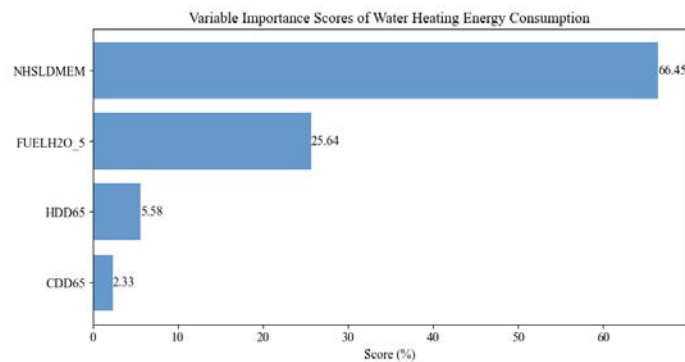


Figure 12 Variable importance scores of Water Heating energy consumption

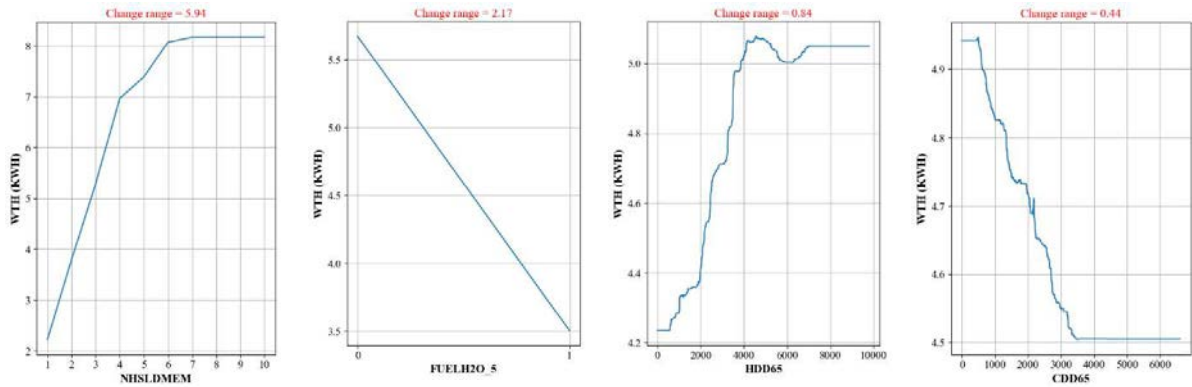


Figure 13 Marginal effects of important factors on water heating energy consumption

4.3.4 Appliance

The RFR model for appliance energy consumption has 21 factors, with an R squared of 0.481. These variables are sorted based on the VI scores (see Figure 14), and corresponding partial dependency plots are shown in Figure 15.

Among the 21 variables, the number of total appliances (NUMAP) is the most determinant, with a VI score of over 50%. The energy consumed by appliances rises significantly from 3 to 13 kWh, with the number of total appliances changing from 5 to 60. Some other characteristics of appliances affect energy consumption as well, such as the frequency of using appliances (e.g., clothes dryer [DRYRUSE], clothes washer [WASHLOAD], dishwasher [DWASHUSE], microwave [AMTMICRO], and second most-used television [TVONWD2 and TVONWE2]), the vintage of appliances (e.g., second most-used refrigerator [AGERFRI2] and freezer [AGEFRZR]), and the size of appliances (e.g., the most and second most-used refrigerator [SIZRFRI1 and SIZRFRI2] and TV [TVSIZE1 and TVSIZE2]). Generally, the more frequently these appliances are used, or the older the appliances are, or the larger the size of the appliances is, the more energy these appliances consume. Among these appliances' characteristics, DRYRUSE is the most important, with a VI score of 5.62% and an impact range of around 2 kWh. Additionally, some housing characteristics, e.g., total housing area (TOTHOUSING), vintage (YEARMADERANGE), and the number of total rooms and windows (TOTROOMS and WINDOWS), positively affect appliance energy consumption. The impact of TOTHOUSING on appliance energy consumption is small when TOTHOUSING is less than 600 m², while the energy consumption goes up sharply when TOTHOUSING increases to 700 m². Furthermore, in terms of the impacts of demographics, annual gross household income (MONEYPY) contributes to appliance energy consumption. Households with an annual gross income of \$140,000 or more consume the most energy for appliances. Also, a household with a larger family size consumes more energy, but when the family members are greater than or equal to 7, there is no significant difference in the appliance energy consumption.

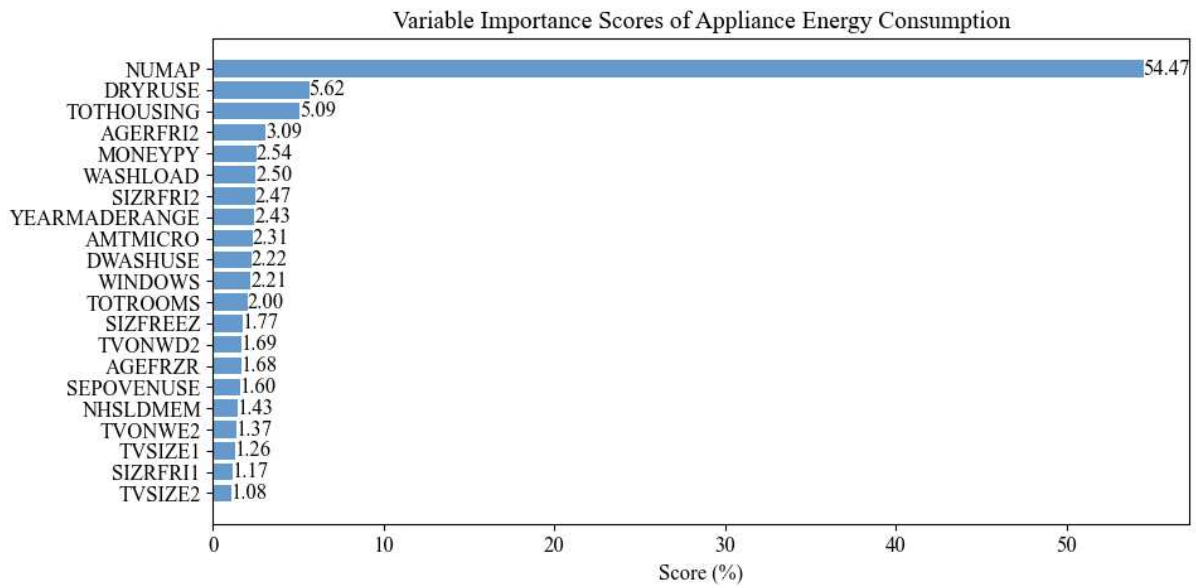
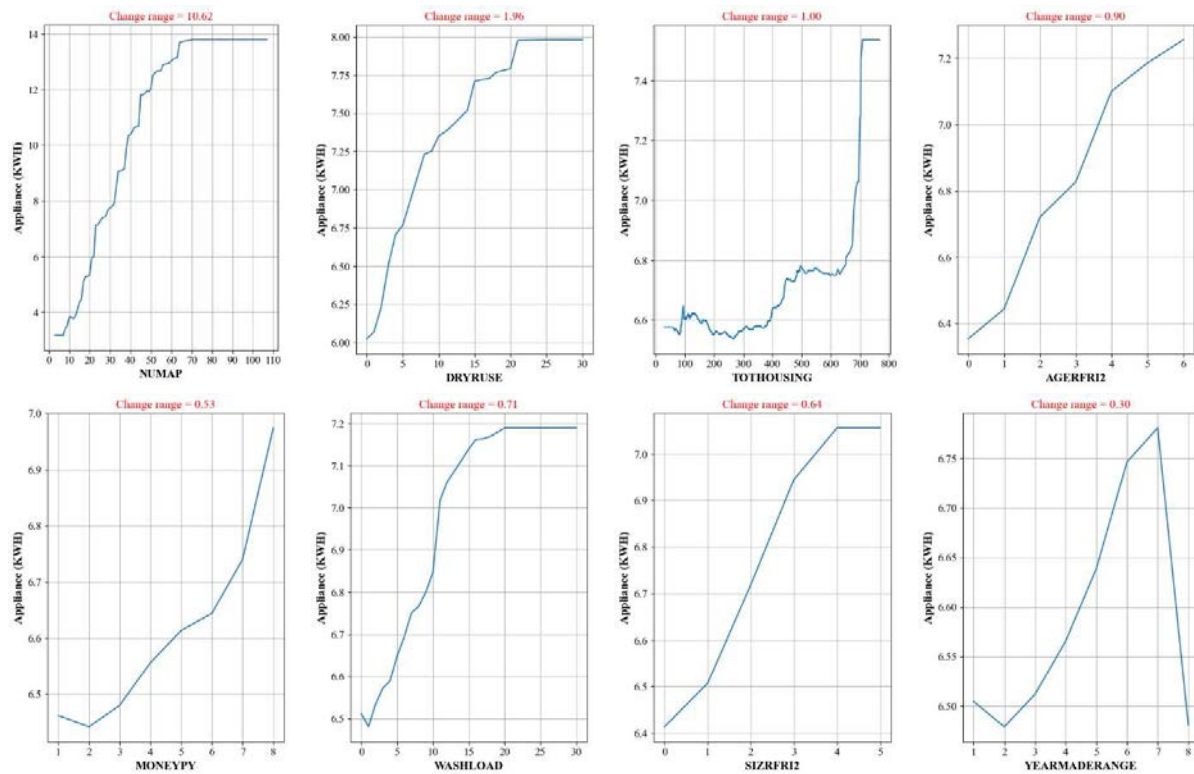


Figure 14 Variable Importance Scores of Appliance Energy Consumption



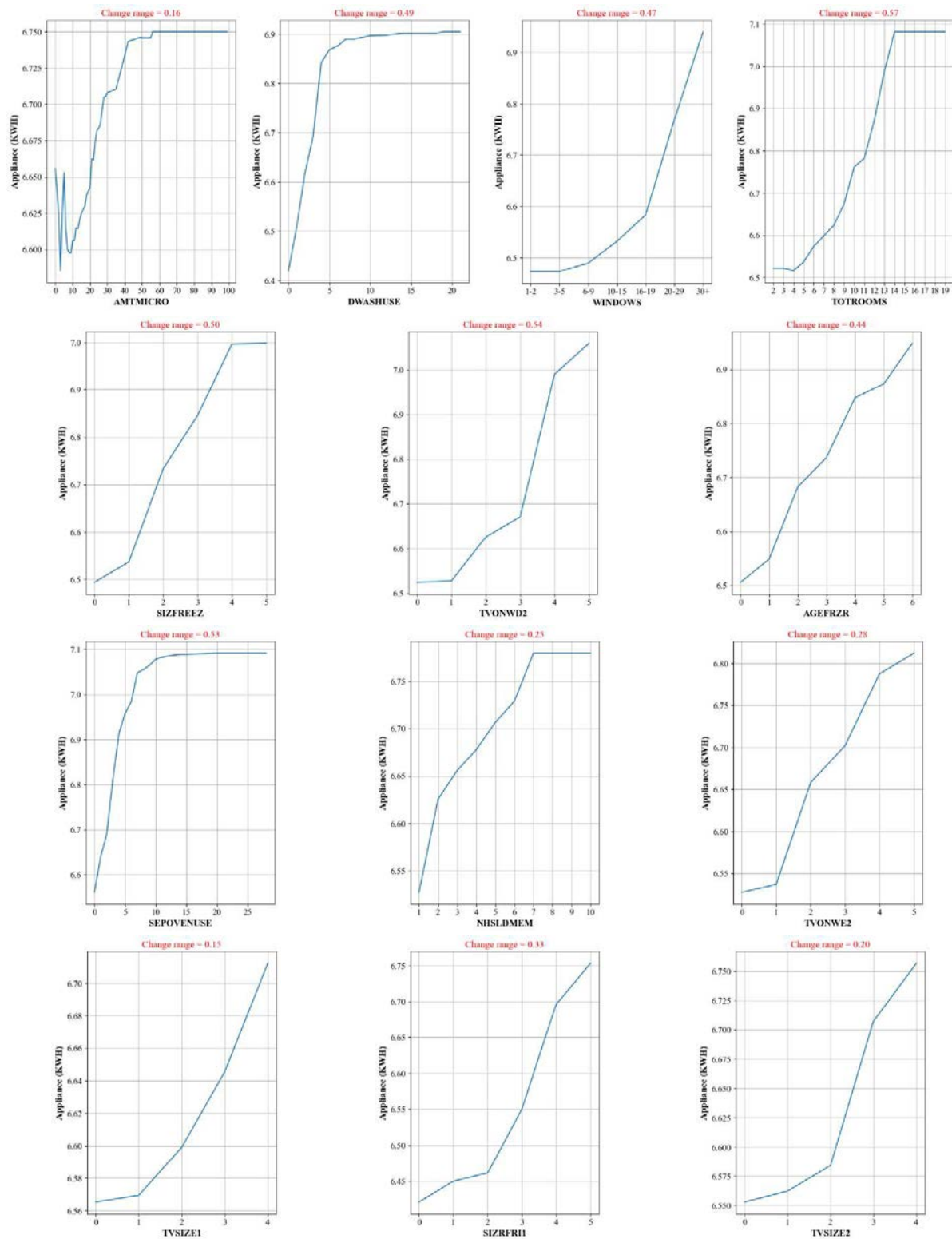


Figure 15 Marginal Effects of Important Factors on Appliance Energy Consumption

4.3.5 Total Energy Consumption

There are 19 important variables included in the fitted model of total energy consumption, which has good interpretability in total energy consumption, with an R-square of 0.542. Figure 16 and Figure 17 show these important variables' VI scores and marginal effects.

Climate factors (i.e., HDD65 and CDD65) are dominant for total energy consumption, with VI scores of 20.6% and 6.2%, respectively. The two climate factors positively correlate with total energy consumption. The higher the CDD65 or HDD65 of the area where the house is located, the more energy is consumed by the households. Total housing area (TOTHOUSING) is the second most important driving factor for total energy consumption. There is an upward trend in total energy consumption with the increase of TOTHOUSING. In particular, when TOTHOUSING is larger than 200 m², the total energy consumption rockets. The numbers of windows and total rooms are two other additional housing characteristics driving the total energy consumption. The more windows or rooms a residential building has, the more energy the household consumes. Moreover, the number of appliances (NUMAP) and their conditions, including fuel type, size, vintage, and use frequency, highly affect total energy consumption. Specifically, the VI score of NUMAP is 11.72%, with an impact change range of over 9 kWh. With the increase of NUMAP, total energy consumption goes up accordingly, while the energy consumption remains stable after NUMAP is larger than 50. For fuel types of end-uses, households using electricity for space heating (FUELHEAT_5) or water heating (FUELH2O_5) consume less energy than those using natural gas (FUELHEAT_1). The size of the second most-used TV (TVSIZE2) and refrigerator (SIZERFRI2), as well as the vintage of the second most-used refrigerator (AGERFRI2) contribute to total energy consumption. The higher frequency of using clothes dryer (DRYRUSE), dishwasher (DWASHUSE), and clothes washer (WASHLOAD) increase energy consumption. Additionally, household demographics are another aspect driving energy consumption. The total energy consumption increases with the family size (NHSLDMEM) and household income (MONEYPY).

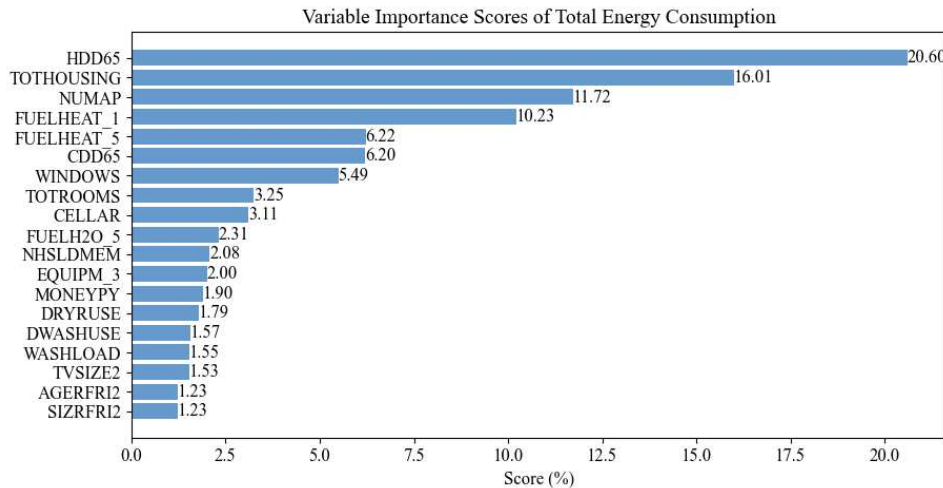
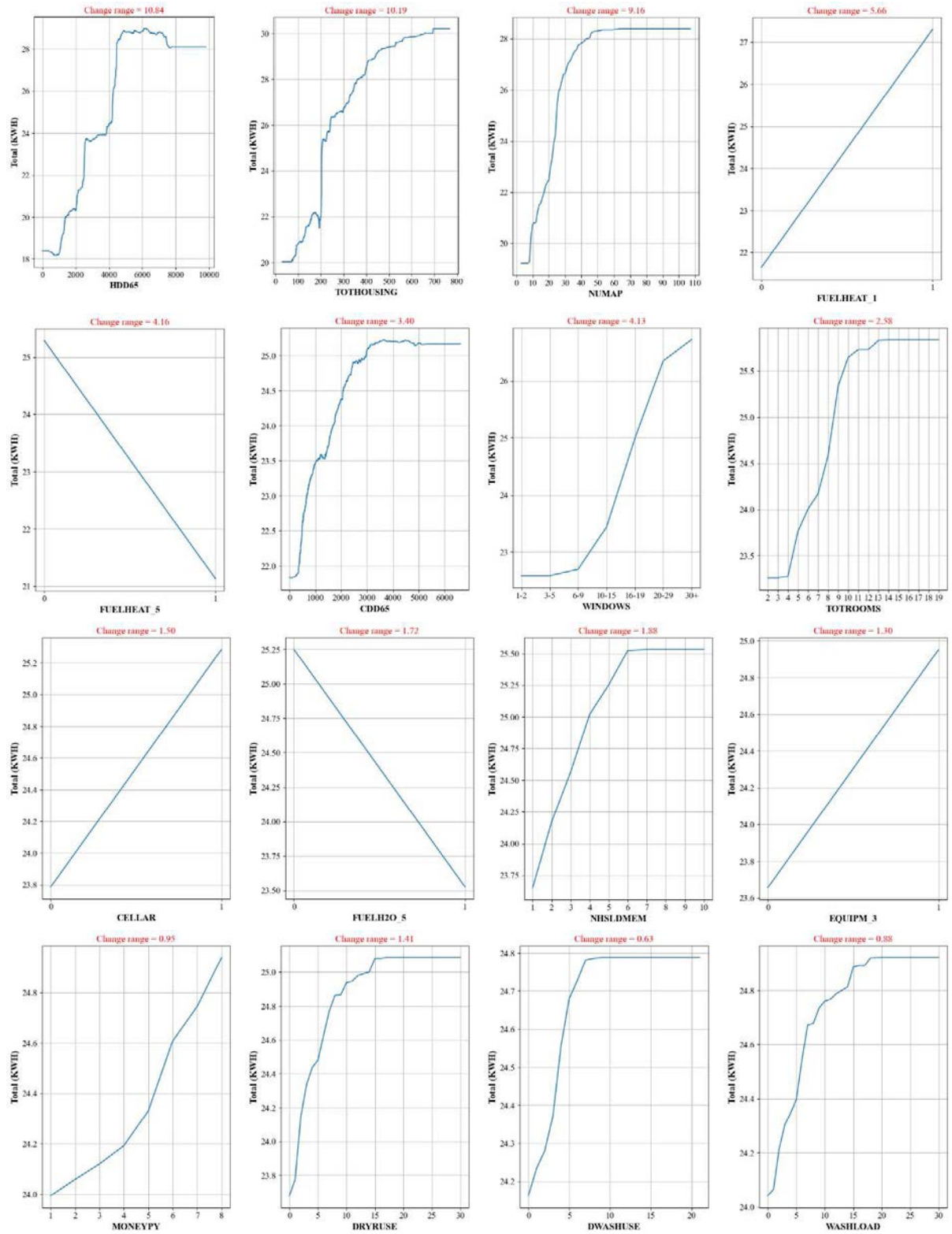


Figure 16 Variable Importance Scores of Total Energy Consumption



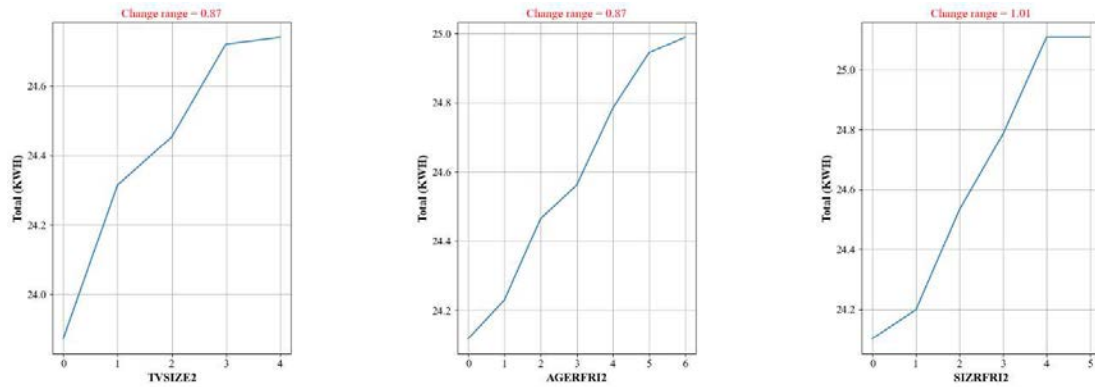


Figure 17 Marginal Effects of Important Factors on Total Energy Consumption

5 Discussions

The developed RFR models for energy consumption perform well in explaining the variability of residential end-use energy consumption, with R-square ranging from 0.481 (appliance) to 0.885 (space cooling). These fitted models reveal not only the importance of the driving factors using VI scores, but also the influences of corresponding variables on energy consumption by partial dependency plots.

5.1 Comparison of Influencing Factors across Different Energy End-Uses

The driving factors of residential energy consumption are multidimensional, including housing characteristics, household demographics, climate, and appliance characteristics, and vary by end-use. These end-uses correspond to the diversity in household energy service needs, which vary across households based on environmental, structural, as well as demographic factors. Specifically, the impacts of climate and housing size on HVAC energy consumption are more important than demographic and HVAC equipment characteristics. This finding is consistent with existing studies (Huo et al., 2021; Iraganaboina and Eluru, 2021; Zhu et al., 2023). In contrast, water heating energy consumption is less affected by climate factors and housing size. Instead, family size is the most important predictor of the energy use of water heating, which is relevant to the finding of Lee and Song (2022), which were based on the metering methods to collect water heating energy usage data and concluded that households with larger family size consume more energy for water heating services.

Additionally, we found that the fuel type affects the energy consumption of space heating and water heating. Using electricity consumes less energy than other energy sources, mainly natural gas, which is neglected by most studies related to residential energy consumption. This finding verifies that end-use electrification is more energy-efficient (González-Torres et al., 2022). Overall, end-use electrification is beneficial to decarbonization, especially using the electricity generated by clean energy (Ebrahimi et al., 2018). The efficiency of using heat pumps for space heating and water heating is 2-3 times higher than using fossil fuels (Dennis, 2015). RECS 2015 shows that space heating and water heating are more commonly accomplished with natural gas than electricity, with the corresponding gas-to-electricity ratios of approximately 54 to 28 and 50 to 42. Also, only 14.5% of studied houses use heat pumps (n=538), and 65.7% use central furnaces (n=2431) for space heating, among which 89% of heat pumps and 16% of central furnaces consume electricity. Hence, there is still a gap in residential end-use electrification. However, in the residential sector, the use of heat pumps for space heating and electric water heater is

increasing (Schwartz et al., 2017), which indicates a further improvement trend of residential energy efficiency in the future.

Of the four end-uses studied here, which correspond to particular energy service needs within a home, every house has appliance energy usage, which accounts for around 30% of total energy use. On average, there are around 26 appliances in each single-family house. The total number of appliances is the most influential factor affecting appliance energy consumption, followed by appliance characteristics, including use frequency, vintage, and size, of common appliances (e.g., washers and driers, dishwashers, TVs, and refrigerators, existing in over 96% of single-family homes). Similarly, our analysis found that the total number of appliances and these appliance characteristics contribute to total energy consumption, as identified by former studies (Jones et al., 2015; Xie and Noor, 2022). However, fewer studies involve the appliance characteristics comprehensively, especially the vintage of appliances. We found that the ages of the second most-used refrigerator and freezer are vital for energy consumption. Appliances' vintage affects their performance, which further affects energy consumption (Wang, 2017). Besides, climate factors (the determinants of space heating and cooling) and some characteristics of space heating (e.g., fuel types) drive total energy use as well, since space heating/cooling are major end-uses in households.

Figure 18 compares the influence of each major category of factors on end uses. In general, the energy consumption of space heating and space cooling is affected most by housing characteristics, whereas demographics and appliance characteristics determine the energy consumption of water heating and appliance, respectively. In terms of total energy consumption, the impact of appliances is the most obvious, followed by structural housing characteristics and climate. These differences in influences across factors could be implicit if only total energy use is analyzed in existing studies. Also, we verified that the impacts of most driving factors on energy consumption are nonlinear through partial dependency plots, e.g., the impact of housing size, climate factors, and family size. However, currently, linear regression is still commonly used to predict energy consumption (Debs and Metzinger, 2022; Fumo and Rafe Biswas, 2015).

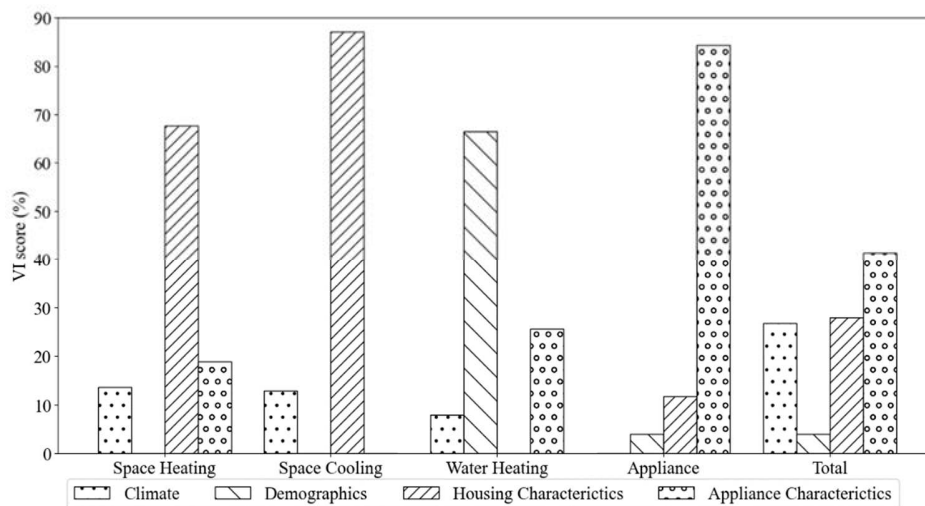


Figure 18 Comparison Influencing Factors of End-Use and Total Energy Consumption

5.2 Impacts of Housing Characteristics and Household Demographics

Compared with the above important variables, some housing characteristics and household demographics, especially housing vintage and household income, have smaller VI scores and are not reflected in most RFR models. However, numerous studies have revealed that household income positively affects residential energy consumption (Huo et al., 2021). One of the potential reasons for the differences is that the major impact of household income on energy consumption is indirect, for example, by affecting the ownership of appliances (Matsumoto, 2016) or determining the housing size (Debs and Metzinger, 2022; Karatasou and Santamouris, 2019). Our correlation analysis (see Table 3) shows that household income positively correlates (r around 0.4) with housing size and the number of total rooms and appliances, as determinant factors on energy consumption of space heating/cooling and appliances. However, considering the importance of housing and appliance characteristics in regression and the limited capacity of the regression model to capture these indirect impacts (Estiri, 2014; Kim et al., 2020; Sanquist et al., 2012), the influences of household income on energy end-use could be moderated in the derived regression models. In the former analysis, the structural equation model of total energy consumption developed by Estiri (2014) confirms that the indirect impact of household demographics intermediated by building characteristics (e.g., housing size, number of total rooms) is four times larger than the direct impact. Similarly, the indirect impact of socioeconomic status on energy consumption is also larger than its direct impact (Karatasou and Santamouris, 2019). Also, it is likely that the property of variables, i.e., household income and housing vintage are categorical variables of 8 categories while housing size and climate factors (i.e., HDD65 and CDD65) are continuous variables, will affect the variable importance in regression. As categorical variables are less informative than continuous variables due to the insufficient divergences and details within the same category (Lazic, 2008), the implications may be implicit in regression models.

To further evaluate the implications of housing characteristics and household demographics on energy end-use, we performed a one-way ANOVA analysis and found that there are significant differences in energy consumption for all end-use and total energy consumption across different income groups ($p < 0.001$, $\alpha = 0.95$) and among different housing vintage groups ($p < 0.001$, $\alpha = 0.95$). The F values of each energy consumption and mean differences between different groups are shown in Figure 19. By comparing end-use energy consumption across different household income groups, energy consumption increases steadily with the rise of household income. The results are similar to (Chen et al., 2022; Debs and Metzinger, 2022). For housing vintage, the newer houses generally consume less energy in total than older houses. Brounen et al. (2012) also found that total energy consumption decreases in newly built houses. However, the impacts of housing vintage vary with different end-uses. In newer houses, the energy consumed for space heating decreases, while households of newer vintage consume more energy in other end-uses, i.e., space cooling, water heating, and appliances, but with varying impacts. The increase of space cooling energy consumption in newer houses is most likely due to higher penetration of central air conditioners in newer buildings, with around 87% in the houses built between 2010 and 2015 and 47% in houses before 1950. Similarly, new buildings have more appliances than old buildings (correlation coefficient between vintage and number of appliances = 0.224), which contributes to the higher appliance energy usage in newer houses. Therefore, future work could further explore and compare the impact of housing vintage and household income on the energy consumption of different end-uses. Further, demographics beyond income and number of household members, such as age composition of residents, occupation and structure of work life (i.e., working from home), political party affiliation of

adult residents, and resident values and beliefs, may all provide additional insight into consumption of end-use energy services at the residential scale, although these datapoints are not collected within the RECS 2015 survey.

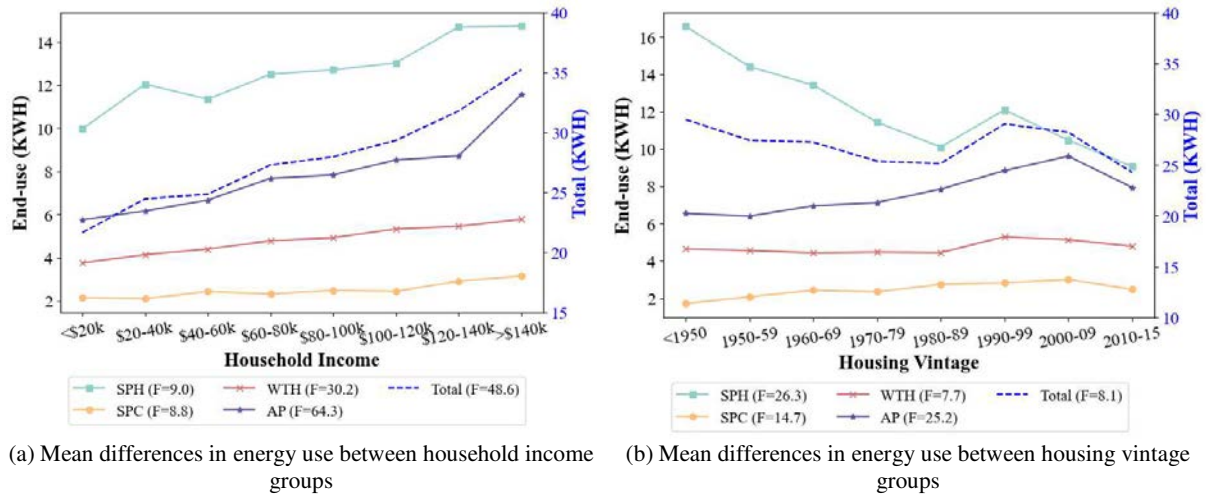


Figure 19 Energy Consumption Differences Between Household Income/ Housing Vintage Groups

Finally, in disagreement with previous studies, it is noticeable that we did not find a correlation between the urban type of houses and energy consumption ($-0.2 < r < 0.2$). Although this finding is inconsistent with the studies of Zheng et al. (2014) and Zhang et al. (2016), it concurs with the explanation that the impacts of urban type (urban vs. rural) are weakened with regional economic development (Zhang and Bai, 2018) and are not significant in developed countries (Wang et al., 2020). Additionally, the influence of age on energy consumption is not reflected in our RFR models, which contradicts Brounen et al. (2012) and Zhu et al. (2023), which found that older occupants consume less energy. Considering that RECS 2015 only recorded the age of respondents instead of all household members, the analysis results could be biased. It is worth mentioning that RECS 2020 is more detailed in collecting the number of occupants below 18, 18-64, and over 65 separately. This would lead to more insights for future analysis on how occupants' age will influence end-use.

6 Conclusions

This nationwide study comprehensively analyzes the influencing factors on residential end-use energy consumption in US single-family detached houses and quantifies their impacts by zooming into different end-uses, corresponding to particular energy services, i.e., space heating/cooling, water heating, and appliances. Multiple driving factors contribute to energy consumption, including housing characteristics, household demographics, climate, and appliance characteristics. Also, we compare the influencing factors of energy consumption across different end-uses, thereby elucidating differences and the potential driving factors that were left out in previous studies when only total energy consumption was analyzed.

First, the influencing factors vary by end-use. The housing size and climate factors (i.e., HDD65 and CDD65) are the key contributors to space heating and space cooling energy usage, while the family size and the number of total appliances determine the energy consumption of water heating and appliances, respectively. Specifically, climate factors (HDD65 and CDD65) and home size positively influence the

energy use of space heating and cooling. On the contrary, water heating energy consumption decreases with the increase of CDD while positively correlating with the family size. Also, this study confirms that the application of electrified end-uses could lead to reduced energy use. As with space heating, using electricity for water heating consumes less energy. In terms of appliances, besides the total number of appliances, certain characteristics (e.g., use frequency, size, and vintage) of common appliances (e.g., refrigerators, televisions, and cloth washers and driers) also demonstrate important implications on the energy use of appliances.

Second, these impacts on energy consumption are nonlinear and are further compounded by the correlations among independent variables. The fitted RFR models verify these nonlinear effects, e.g., the housing size, total number of appliances, and family size, on energy services. Moreover, the correlations between demographics and characteristics of houses and appliances are non-negligible, especially considering the strong correlations between household income with housing size and the total number of appliances.

However, there are some limitations in this study. First, due to these correlations between independent variables, the indirect impacts of demographics may exist but not be captured sufficiently in our models. Therefore, future work can further examine the indirect effect of demographics on each type of end-use energy service. Also, this study only focuses on single-family detached houses. Other housing types could be further explored. Besides, the end-use energy consumption data in RECS 2015 was decomposed based on engineering models. Although these energy end-use data were calibrated and validated, the data quality is still subject to the validity of engineering models and detailed calibration and validation methods. Even if RECS 2015 is already a trustworthy data source, additional national-scale efforts with direct sub-metering data collection are expected to further improve the data quality and analysis as resources allow. Nevertheless, this study has important implications for the conservation of residential energy. The identified influencing factors on different end-uses can provide references for developing energy efficiency programs, especially targeting specific residential end-use services or occupant groups based on consumption across different end-use services.

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.

Acknowledgments

The research is funded by National Science Foundation (NSF). Award title: EAGER: SAI: Socio-Technological Guided Enhancement of Power Infrastructure Resilience (#2121875).

References

- Akoglu, H., 2018. User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine* 18, 91–93. <https://doi.org/10.1016/j.tjem.2018.08.001>
- Aydin, E., Brounen, D., 2019. The impact of policy on residential energy consumption. *Energy* 169, 115–129. <https://doi.org/10.1016/j.energy.2018.12.030>

834 Aydinalp, M., Ismet Ugursal, V., Fung, A.S., 2002. Modeling of the appliance, lighting, and space-
835 cooling energy consumptions in the residential sector using neural networks. *Applied Energy* 71,
836 87–110. [https://doi.org/10.1016/S0306-2619\(01\)00049-6](https://doi.org/10.1016/S0306-2619(01)00049-6)

837 Aydinalp-Koksal, M., Ugursal, V.I., 2008. Comparison of neural network, conditional demand analysis,
838 and engineering approaches for modeling end-use energy consumption in the residential sector.
839 *Applied Energy* 85, 271–296. <https://doi.org/10.1016/j.apenergy.2006.09.012>

840 Bartusch, C., Odlare, M., Wallin, F., Wester, L., 2012. Exploring variance in residential electricity
841 consumption: Household features and building properties. *Applied Energy* 92, 637–643.
842 <https://doi.org/10.1016/j.apenergy.2011.04.034>

843 Bedir, M., Hasselaar, E., Itard, L., 2013. Determinants of electricity consumption in Dutch dwellings.
844 *Energy and Buildings* 58, 194–207. <https://doi.org/10.1016/j.enbuild.2012.10.016>

845 Best, R., Burke, P.J., Nishitaten, S., 2021. Factors Affecting Renters' Electricity Use: More Than Split
846 Incentives. *EJ* 42. <https://doi.org/10.5547/01956574.42.5.rbes>

847 Breiman, L., 2001. Random Forests. *Machine Learning* 45, 5–32.
848 <https://doi.org/10.1023/A:1010933404324>

849 Brounen, D., Kok, N., Quigley, J.M., 2012. Residential energy use and conservation: Economics and
850 demographics. *European Economic Review, Green Building, the Economy, and Public Policy* 56,
851 931–945. <https://doi.org/10.1016/j.euroecorev.2012.02.007>

852 Chehreh Chelgani, S., Matin, S.S., Hower, J.C., 2016. Explaining relationships between coke quality
853 index and coal properties by Random Forest method. *Fuel* 182, 754–760.
854 <https://doi.org/10.1016/j.fuel.2016.06.034>

855 Chen, C., Xu, X., Adua, L., Briggs, M., Nelson, H., 2022. Exploring the factors that influence energy use
856 intensity across low-, middle-, and high-income households in the United States. *Energy Policy*
857 168, 113071. <https://doi.org/10.1016/j.enpol.2022.113071>

858 Chen, J., Wang, X., Steemers, K., 2013. A statistical analysis of a residential energy consumption survey
859 study in Hangzhou, China. *Energy and Buildings* 66, 193–202.
860 <https://doi.org/10.1016/j.enbuild.2013.07.045>

861 Chicco, D., Warrens, M.J., Jurman, G., 2021. The coefficient of determination R-squared is more
862 informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation.
863 *PeerJ Comput. Sci.* 7, e623. <https://doi.org/10.7717/peerj-cs.623>

864 Chun-sheng, Z., Shu-wen, N., Xin, Z., 2012. Effects of household energy consumption on environment
865 and its influence factors in rural and urban areas. *Energy Procedia, 2011 2nd International*
866 *Conference on Advances in Energy Engineering (ICAEE)* 14, 805–811.
867 <https://doi.org/10.1016/j.egypro.2011.12.1015>

868 Darst, B.F., Malecki, K.C., Engelman, C.D., 2018. Using recursive feature elimination in random forest to
869 account for correlated variables in high dimensional data. *BMC Genetics* 19, 65.
870 <https://doi.org/10.1186/s12863-018-0633-8>

871 Debs, L., Metzinger, J., 2022. A Comparison of Energy Consumption in American Homes by Climate
872 Region. *Buildings* 12, 82. <https://doi.org/10.3390/buildings12010082>

873 Dennis, K., 2015. Environmentally Beneficial Electrification: Electricity as the End-Use Option. *The*
874 *Electricity Journal* 28, 100–112. <https://doi.org/10.1016/j.tej.2015.09.019>

875 Đurišić, V., Rogić, S., Smolović, J.C., Radonjić, M., 2020. Determinants of household electrical energy
876 consumption: Evidences and suggestions with application to Montenegro. *Energy Reports,*
877 *Technologies and Materials for Renewable Energy, Environment and Sustainability* 6, 209–217.
878 <https://doi.org/10.1016/j.egy.2019.10.039>

879 Duzgun, B., Koksal, M.A., Bayindir, R., 2022. Assessing drivers of residential energy consumption in
880 Turkey: 2000–2018. *Energy for Sustainable Development* 70, 371–386.
881 <https://doi.org/10.1016/j.esd.2022.08.016>

882 Ebrahimi, S., Mac Kinnon, M., Brouwer, J., 2018. California end-use electrification impacts on carbon
883 neutrality and clean air. *Applied Energy* 213, 435–449.
884 <https://doi.org/10.1016/j.apenergy.2018.01.050>

- Estiri, H., 2014. Building and household X-factors and energy consumption at the residential sector: A structural equation analysis of the effects of household and building characteristics on the annual energy consumption of US residential buildings. *Energy Economics* 43, 178–184. <https://doi.org/10.1016/j.eneco.2014.02.013>
- Farahbakhsh, H., Ugursal, V.I., Fung, A.S., 1998. A residential end-use energy consumption model for Canada. *International Journal of Energy Research* 22, 1133–1143. [https://doi.org/10.1002/\(SICI\)1099-114X\(19981025\)22:13<1133::AID-ER434>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1099-114X(19981025)22:13<1133::AID-ER434>3.0.CO;2-E)
- Fell, M.J., 2017. Energy services: A conceptual review. *Energy Research & Social Science* 27, 129–140. <https://doi.org/10.1016/j.erss.2017.02.010>
- Feng, F., Li, Z., Ruan, Y., Xu, P., 2016. An Empirical Study of Influencing Factors on Residential Building Energy Consumption in Qingdao City, China. *Energy Procedia, Clean Energy for Clean City: CUE 2016--Applied Energy Symposium and Forum: Low-Carbon Cities and Urban Energy Systems* 104, 245–250. <https://doi.org/10.1016/j.egypro.2016.12.042>
- Frederiks, E.R., Stenner, K., Hobman, E.V., 2015. The Socio-Demographic and Psychological Predictors of Residential Energy Consumption: A Comprehensive Review. *Energies* 8, 573–609. <https://doi.org/10.3390/en8010573>
- Fumo, N., Rafe Biswas, M.A., 2015. Regression analysis for prediction of residential energy consumption. *Renewable and Sustainable Energy Reviews* 47, 332–343. <https://doi.org/10.1016/j.rser.2015.03.035>
- Genjo, K., Tanabe, S., Matsumoto, S., Hasegawa, K., Yoshino, H., 2005. Relationship between possession of electric appliances and electricity for lighting and others in Japanese households. *Energy and Buildings* 37, 259–272. <https://doi.org/10.1016/j.enbuild.2004.06.025>
- González-Torres, M., Pérez-Lombard, L., Coronel, J.F., Maestre, I.R., Yan, D., 2022. A review on buildings energy information: Trends, end-uses, fuels and drivers. *Energy Reports* 8, 626–637. <https://doi.org/10.1016/j.egyr.2021.11.280>
- Gregorutti, B., Michel, B., Saint-Pierre, P., 2017. Correlation and variable importance in random forests. *Stat Comput* 27, 659–678. <https://doi.org/10.1007/s11222-016-9646-1>
- Guerra Santin, O., Itard, L., Visscher, H., 2009. The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock. *Energy and Buildings* 41, 1223–1232. <https://doi.org/10.1016/j.enbuild.2009.07.002>
- Huo, T., Cai, W., Zhang, W., Wang, J., Zhao, Y., Zhu, X., 2021. How does income level impact residential-building heating energy consumption? Micro-level evidence from household surveys. *Environmental Impact Assessment Review* 91, 106659. <https://doi.org/10.1016/j.eiar.2021.106659>
- Iraganaboina, N.C., Eluru, N., 2021. An examination of factors affecting residential energy consumption using a multiple discrete continuous approach. *Energy and Buildings* 240, 110934. <https://doi.org/10.1016/j.enbuild.2021.110934>
- Jiang, H., Mei, L., Wei, Y., Zheng, R., Guo, Y., 2022. The influence of the neighbourhood environment on peer-to-peer accommodations: A random forest regression analysis. *Journal of Hospitality and Tourism Management* 51, 105–118. <https://doi.org/10.1016/j.jhtm.2022.02.028>
- Jin, X., Zhang, C., Xiao, F., Li, A., Miller, C., 2023. A review and reflection on open datasets of city-level building energy use and their applications. *Energy and Buildings* 285, 112911. <https://doi.org/10.1016/j.enbuild.2023.112911>
- Jones, R.V., Fuertes, A., Lomas, K.J., 2015. The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings. *Renewable and Sustainable Energy Reviews* 43, 901–917. <https://doi.org/10.1016/j.rser.2014.11.084>
- Jones, R.V., Lomas, K.J., 2015. Determinants of high electrical energy demand in UK homes: Socio-economic and dwelling characteristics. *Energy and Buildings* 101, 24–34. <https://doi.org/10.1016/j.enbuild.2015.04.052>
- 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.
<https://doi.org/10.1016/j.enbuild.2018.04.019>
- Karatasou, S., Santamouris, M., 2019. Socio-economic status and residential energy consumption: A latent variable approach. *Energy and Buildings* 198, 100–105.
<https://doi.org/10.1016/j.enbuild.2019.06.013>
- Kaza, N., 2010. Understanding the spectrum of residential energy consumption: A quantile regression approach. *Energy Policy, Energy Efficiency Policies and Strategies with regular papers*. 38, 6574–6585. <https://doi.org/10.1016/j.enpol.2010.06.028>
- Khalil, M., McGough, A.S., Pourmirza, Z., Pazhoohesh, M., Walker, S., 2022. Machine Learning, Deep Learning and Statistical Analysis for forecasting building energy consumption — A systematic review. *Engineering Applications of Artificial Intelligence* 115, 105287.
<https://doi.org/10.1016/j.engappai.2022.105287>
- Kim, Y.-J., Lee, S.-J., Jin, H.-S., Suh, I.-A., Song, S.-Y., 2020. Comparison of linear and nonlinear statistical models for analyzing determinants of residential energy consumption. *Energy and Buildings* 223, 110226. <https://doi.org/10.1016/j.enbuild.2020.110226>
- Kumar, P., Caggiano, H., Cuite, C., Andrews, C.J., Felder, F.A., Shwom, R., Floress, K., Ahamed, S., Schelly, C., 2022. Behaving or not? Explaining energy conservation via identity, values, and awareness in U.S. suburban homes. *Energy Research & Social Science* 92, 102805.
<https://doi.org/10.1016/j.erss.2022.102805>
- Larsen, B.M., Nesbakken, R., 2004. Household electricity end-use consumption: results from econometric and engineering models. *Energy Economics* 26, 179–200.
<https://doi.org/10.1016/j.eneco.2004.02.001>
- Lazic, S.E., 2008. Why we should use simpler models if the data allow this: relevance for ANOVA designs in experimental biology. *BMC Physiol* 8, 16. <https://doi.org/10.1186/1472-6793-8-16>
- Lee, S.-J., Kim, Y.-J., Jin, H.-S., Kim, S.-I., Ha, S.-Y., Song, S.-Y., 2019. Residential End-Use Energy Estimation Models in Korean Apartment Units through Multiple Regression Analysis. *Energies* 12, 2327. <https://doi.org/10.3390/en12122327>
- Lee, S.-J., Song, S.-Y., 2022. Determinants of residential end-use energy: Effects of buildings, sociodemographics, and household appliances. *Energy and Buildings* 257, 111782.
<https://doi.org/10.1016/j.enbuild.2021.111782>
- Li, C., 2014. Home energy consumption estimation by end use and energy efficiency upgrade recommendations. Duke University.
- Li, J., Shui, C., Li, R., Zhang, L., 2022. Driving factors analysis of residential electricity expenditure using a multi-scale spatial regression analysis: A case study. *Energy Reports* 8, 7127–7142.
<https://doi.org/10.1016/j.egyr.2022.05.026>
- Liu, G., Kong, Z., Dong, J., Dong, X., Jiang, Q., Wang, K., Li, J., Li, C., Wan, X., 2021. Influencing Factors, Energy Consumption, and Carbon Emission of Central Heating in China: A Supply Chain Perspective. *Frontiers in Energy Research* 9.
- Marques Ramos, A.P., Prado Osco, L., Elis Garcia Furuya, D., Nunes Gonçalves, W., Cordeiro Santana, D., Pereira Ribeiro Teodoro, L., Antonio da Silva Junior, C., Fernando Capristo-Silva, G., Li, J., Henrique Rojo Baio, F., Marcato Junior, J., Eduardo Teodoro, P., Pistori, H., 2020. A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. *Computers and Electronics in Agriculture* 178, 105791.
<https://doi.org/10.1016/j.compag.2020.105791>
- Matsumoto, S., 2016. How do household characteristics affect appliance usage? Application of conditional demand analysis to Japanese household data. *Energy Policy* 94, 214–223.
<https://doi.org/10.1016/j.enpol.2016.03.048>
- Min, J., Hausfather, Z., Lin, Q.F., 2010. A High-Resolution Statistical Model of Residential Energy End Use Characteristics for the United States. *Journal of Industrial Ecology* 14, 791–807.
<https://doi.org/10.1111/j.1530-9290.2010.00279.x>

- Mo, Y., Zhao, D., 2021. Effective factors for residential building energy modeling using feature engineering. *Journal of Building Engineering* 44, 102891. <https://doi.org/10.1016/j.jobbe.2021.102891>
- Mohr, T.M., 2018. Fuel poverty in the US: Evidence using the 2009 Residential Energy Consumption Survey. *Energy Economics* 74, 360–369. <https://doi.org/10.1016/j.eneco.2018.06.007>
- Molnar, C., Freiesleben, T., König, G., Casalicchio, G., Wright, M.N., Bischl, B., 2021. Relating the Partial Dependence Plot and Permutation Feature Importance to the Data Generating Process. <https://doi.org/10.48550/arXiv.2109.01433>
- Nie, H., Kemp, R., Xu, J., Vasseur, V., Fan, Y., 2018. Drivers of urban and rural residential energy consumption in China from the perspectives of climate and economic effects. *Journal of Cleaner Production* 172, 2954–2963. <https://doi.org/10.1016/j.jclepro.2017.11.117>
- Paul, A., Baumhögger, E., Elsner, A., Reineke, M., Hueppe, C., Stamminger, R., Hoelscher, H., Wagner, H., Gries, U., Becker, W., Vrabec, J., 2022. Impact of aging on the energy efficiency of household refrigerating appliances. *Applied Thermal Engineering* 205, 117992. <https://doi.org/10.1016/j.applthermaleng.2021.117992>
- Rafsanjani, H.N., 2016. Factors Influencing the Energy Consumption of Residential Buildings: A Review, in: *Construction Research Congress 2016*. Presented at the Construction Research Congress 2016, American Society of Civil Engineers, San Juan, Puerto Rico, pp. 1133–1142. <https://doi.org/10.1061/9780784479827.114>
- Ren, Z., Foliente, G., Chan, W.-Y., Chen, D., Ambrose, M., Paevere, P., 2013. A model for predicting household end-use energy consumption and greenhouse gas emissions in Australia. *International Journal of Sustainable Building Technology and Urban Development* 4, 210–228. <https://doi.org/10.1080/2093761X.2013.801801>
- Ruan, Y., Cao, J., Feng, F., Li, Z., 2017. The role of occupant behavior in low carbon oriented residential community planning: A case study in Qingdao. *Energy and Buildings* 139, 385–394. <https://doi.org/10.1016/j.enbuild.2017.01.049>
- Sagi, O., Rokach, L., 2018. Ensemble learning: A survey. *WIREs Data Mining and Knowledge Discovery* 8, e1249. <https://doi.org/10.1002/widm.1249>
- Sanquist, T.F., Orr, H., Shui, B., Bittner, A.C., 2012. Lifestyle factors in U.S. residential electricity consumption. *Energy Policy* 42, 354–364. <https://doi.org/10.1016/j.enpol.2011.11.092>
- Schober, P., Boer, C., Schwarte, L.A., 2018. Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia* 126, 1763–1768. <https://doi.org/10.1213/ANE.0000000000002864>
- Schwartz, L., Wei, M., Morrow, W., Deason, J., Schiller, S.R., Leventis, G., Smith, S., Leow, W.L., Levin, T., Plotkin, S., Zhou, Y., 2017. Electricity End Uses, Energy Efficiency, and Distributed Energy Resources Baseline.
- Shi, G.-Y., Zhou, Y., Sang, Y.-Q., Huang, H., Zhang, J.-S., Meng, P., Cai, L.-L., 2021. Modeling the response of negative air ions to environmental factors using multiple linear regression and random forest. *Ecological Informatics* 66, 101464. <https://doi.org/10.1016/j.ecoinf.2021.101464>
- Steemers, K., Yun, G.Y., 2009. Household energy consumption: a study of the role of occupants. *Building Research & Information* 37, 625–637. <https://doi.org/10.1080/09613210903186661>
- Swan, L.G., Ugursal, V.I., 2009. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews* 13, 1819–1835. <https://doi.org/10.1016/j.rser.2008.09.033>
- Tiwari, S., Schelly, C., Ou, G., Sahraei-Ardakani, M., Chen, J., Jafarishiadeh, F., 2022. Conceptualizing resilience: An energy services approach. *Energy Research & Social Science* 94, 102878. <https://doi.org/10.1016/j.erss.2022.102878>
- Tsemekidi, S., Bertoldi, P., Diluiso, F., Castellazzi, L., Economidou, M., Labanca, N., Ribeiro Serrenho, T., Zangheri, P., 2019. Analysis of the EU Residential Energy Consumption: Trends and Determinants. *Energies* 12, 1065. <https://doi.org/10.3390/en12061065>

- Tso, G.K.F., Guan, J., 2014. A multilevel regression approach to understand effects of environment indicators and household features on residential energy consumption. *Energy* 66, 722–731. <https://doi.org/10.1016/j.energy.2014.01.056>
- US EIA, 2022. Monthly Energy Review – November 2022 282.
- US EIA, 2015a. Residential Energy Consumption Survey (RECS) 2015 Consumption and Expenditures Technical Documentation Summary.
- US EIA, 2015b. 2015 RECS Survey Data [WWW Document]. URL <https://www.eia.gov/consumption/residential/data/2015> (accessed 10.2.22).
- Wang, E., 2017. Decomposing core energy factor structure of U.S. residential buildings through principal component analysis with variable clustering on high-dimensional mixed data. *Applied Energy* 203, 858–873. <https://doi.org/10.1016/j.apenergy.2017.06.105>
- Wang, L., Lee, E.W.M., Hussian, S.A., Yuen, A.C.Y., Feng, W., 2021. Quantitative impact analysis of driving factors on annual residential building energy end-use combining machine learning and stochastic methods. *Applied Energy* 299, 117303. <https://doi.org/10.1016/j.apenergy.2021.117303>
- Wang, Q., Lin, J., Zhou, K., Fan, J., Kwan, M.-P., 2020. Does urbanization lead to less residential energy consumption? A comparative study of 136 countries. *Energy* 202, 117765. <https://doi.org/10.1016/j.energy.2020.117765>
- Xie, Y., Noor, A.I.M., 2022. Factors Affecting Residential End-Use Energy: Multiple Regression Analysis Based on Buildings, Households, Lifestyles, and Equipment. *Buildings* 12, 538. <https://doi.org/10.3390/buildings12050538>
- Yun, G.Y., Steemers, K., 2011. Behavioural, physical and socio-economic factors in household cooling energy consumption. *Applied Energy* 88, 2191–2200. <https://doi.org/10.1016/j.apenergy.2011.01.010>
- Zhang, M., Bai, C., 2018. Exploring the influencing factors and decoupling state of residential energy consumption in Shandong. *Journal of Cleaner Production* 194, 253–262. <https://doi.org/10.1016/j.jclepro.2018.05.122>
- Zhang, M., Song, Y., Li, P., Li, H., 2016. Study on affecting factors of residential energy consumption in urban and rural Jiangsu. *Renewable and Sustainable Energy Reviews* 53, 330–337. <https://doi.org/10.1016/j.rser.2015.08.043>
- Zhang, X.M., Grolinger, K., Capretz, M.A.M., Seewald, L., 2018. Forecasting Residential Energy Consumption: Single Household Perspective, in: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). Presented at the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 110–117. <https://doi.org/10.1109/ICMLA.2018.00024>
- Zhao, H., 2012. A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews* 7.
- Zheng, X., Wei, C., Qin, P., Guo, J., Yu, Y., Song, F., Chen, Z., 2014. Characteristics of residential energy consumption in China: Findings from a household survey. *Energy Policy* 75, 126–135. <https://doi.org/10.1016/j.enpol.2014.07.016>
- Zhou, Qifeng, Zhou, H., Zhou, Qingqing, Yang, F., Luo, L., 2014. Structure damage detection based on random forest recursive feature elimination. *Mechanical Systems and Signal Processing* 46, 82–90. <https://doi.org/10.1016/j.ymssp.2013.12.013>
- Zhu, M., Huang, Y., Wang, S.-N., Zheng, X., Wei, C., 2023. Characteristics and patterns of residential energy consumption for space cooling in China: Evidence from appliance-level data. *Energy* 265, 126395. <https://doi.org/10.1016/j.energy.2022.126395>
- Ziegler, A., König, I.R., 2014. Mining data with random forests: current options for real-world applications. *WIREs Data Mining and Knowledge Discovery* 4, 55–63. <https://doi.org/10.1002/widm.1114>