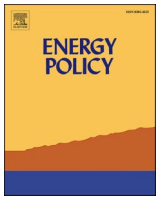
Energy Policy 202 (2025) 114594

Contents lists available at ScienceDirect

Energy Policy

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



Determinants of U.S. residential energy consumption at national and state levels: Policy implications

Sepideh Sadat Korsavi [a,*], Rahman Azari [a], Lisa D. Iulo [a], Mehrdad Mahdavi [b]

[a] Department of Architecture, The Pennsylvania State University, 105 Stuckeman Family Building, University Park, PA, 16802, United States
[b] Department of Computer Science & Engineering, The Pennsylvania State University, W365 Westgate Building, University Park, PA, 16802, United States

ARTICLE INFO

Keywords:
Energy consumption
Residential buildings
Machine learning modeling
Electrified heating
Sensitivity analysis
Policy implications

ABSTRACT

Literature has extensively studied the effects of physical and environmental parameters on building energy consumption through physics-based simulations. However, energy use of the building sector at state and national levels is influenced by more complex factors that vary spatially and temporally and include sociodemographic, socioeconomic, physical, climatic, and microclimatic parameters.

This study identifies key determinants of U.S. residential energy consumption at national and state levels using the 2020 Residential Energy Consumption Survey (RECS) data and examines whether they increase or decrease energy use. This research uses Machine learning algorithms for energy modeling and SHAP (SHapley Additive exPlanations) sensitivity analysis to explain the contribution of each feature to the model. Top determinants influencing national energy include using electricity for space and water heating, Heating Degree Days, and energy-consuming areas. Using electricity for heating can significantly reduce on-site residential energy consumption. Given that electricity is the second most common heating source in American homes after natural gas, these findings highlight the potential benefits of transitioning to electric heating systems like heat pumps. Further policies derived from key state-level determinants promote attached housing, efficient setpoint temperature behaviors, energy-efficient appliances and lighting, and on-site electricity generation.

1. Introduction

The residential building sector, with its significant energy use, is an important factor of sustainability initiatives in the United States and worldwide. In 2022, American households used 12.3 quadrillion British Thermal Units (BTUs) — equivalent to 12.98 EJ or 16 % of the country's delivered energy — and accounted for 21.7 quadrillion BTUs (22.9 EJ or 21.6 % of the total) of the primary energy consumption in the U.S (U.S. EIA, 2023a). The diverse determinants of housing energy use including climatic, physical, and occupant-related factors have been studied. For example, in a comprehensive meta-analytical review, Tran et al. (2023) showed that occupant-related determinants are the most frequently studied parameters of energy use, followed by building features and outdoor environmental factors.

When compared to individual home energy use, however, the energy consumption of the residential sector at the larger regional and national scales is influenced by a broader set of factors, including population and demographic variations, household structures, economic changes, and urban parameters. For example, the number of single detached homes has grown by about 10.7 % between 2011 and 2021 (U.S. Census Bureau, 2021) and the U.S. population has become older in the past two decades, with American households having fewer young children now (Blakeslee et al., 2023).

Lifestyle changes and preferences for larger housing sizes as well as an expected rise in older populations all affect building energy needs at a macroscale (Estiri and Zagheni, 2019). As a result, although individual homes are becoming more energy-efficient, factors like population growth, demographic shifts, and lifestyle changes can still lead to increased energy consumption across the U.S. (Lima et al. 2013). Understanding the determinants of household energy consumption should be an integral part of efforts to promote energy efficiency in the residential sector (Belaïd, 2016). It is important to note that the drivers of residential energy use vary spatially and temporally and there is a need for the research community to document, examine, and observe the changes in type and significance of these drivers and their effects across scales of space and time.

This study uses the most recent Residential Energy Consumption Survey (RECS) dataset (i.e., 2020 data published in 2023) to identify the

* Corresponding author.
E-mail addresses: ssk5573@psu.edu (S.S. Korsavi), razari@psu.edu (R. Azari), ldi1@psu.edu (L.D. Iulo), mzm616@psu.edu (M. Mahdavi).

https://doi.org/10.1016/j.enpol.2025.114594
Received 26 June 2024; Received in revised form 7 February 2025; Accepted 4 March 2025
Available online 27 March 2025

drivers of U.S. residential energy use. More specifically, we a) model national and state-level residential building energy consumption using machine learning methods, b) identify the key determinants of residential energy consumption to determine whether they increase or decrease energy use, and quantify their effects using SHAP (SHapley Additive exPlanations) sensitivity analysis, and c) provide energy policy recommendations by studying the strongest explanatory features.

The 2020 RECS dataset is particularly valuable for this study and is distinctive compared with previous RECS datasets as it includes a greater number of variables and a larger sample size. This allows for a fresher perspective on residential energy consumption in the U.S. with the potential to create a more reliable picture of its drivers. By positioning the results in the context of literature, this study will therefore help us to understand how drivers of energy use have changed across time horizons.

2. Literature review

We provide an overview of studies that have applied machine learning or statistical analysis to identify the determinants of building energy consumption in the U.S. at national and regional scales.

2.1. Determinants of national residential energy consumption

Several studies have used the RECS dataset to investigate drivers of the U.S. residential energy sector; therefore, we categorized these findings by dataset to account for variations in feature sets and sample sizes.

2015 RECS: A study by X. Cui et al. (2024) used the 2015 RECS dataset and identified total square footage, space heating with natural gas, climate conditions, and building age as the most important features influencing Energy Use Intensity (EUI) in U.S. apartments and single-family houses. Wang et al. (2021) employed artificial neural networks and the Monte Carlo method on the 2015 RECS dataset and indicated a positive relationship between energy end-use and determinants such as 'total building area', 'number of rooms, 'number of windows', 'winter temperature with heating', 'level of insulation', and 'respondent's age', and a negative relationship between energy end-use and 'summer temperature with cooling'. Goldstein et al. (2022) used the same 2015 RECS dataset and demonstrated that housing quality, floor area, and home ownership status were the main drivers of energy use.

Combined RECS Datasets: Burnett and Lynne Kiesling (2022) applied machine-learning models on the 2001–2015 RECS dataset and showed that the most important predictors of residential energy use in the U.S. were the total square footage of the heated or cooled spaces in a home, followed by residential natural gas prices, number of bedrooms, living in large apartment units, and electricity prices. Using 1987, 1990, 2005, and 2009 RECS data, Estiri and Zagheni (2019) examined the age-energy interrelations in the U.S. housing sector and showed an overall increasing trend in energy consumption as the household head ages, controlling for other factors.

2001, 2005, and 2009 RECS: Mostafavi, Farzinmoghadam, and Hoque (2017) employed Quantile Regression (QR) on the 2009 RECS database and predicted the effects of physical and socioeconomic variables on space heating, cooling, water heating, lighting, and appliance energy consumption. Sanquist et al. (2012) used a multivariate approach to analyze U.S. residential electricity patterns from the 2005 RECS and identified key lifestyle factors such as air conditioning, laundry, personal computer usage, and TV usage affecting energy consumption. Yun and Steemers (2011) used the 2001 RECS dataset and revealed that an increase in the use of air conditioning (AC), number of rooms with AC, cooling degree days (CDD), size of the dwelling, number of household members, and total household income increased the cooling energy consumption.

2.2. Determinants of urban and state-level energy consumption in the U.S

To complement the residential energy trends at the national scale, we provide a literature review of urban- and state-level energy consumption in the United States. Bednar, Reames, and Keoleian (2017) examined residential annual heating energy consumption of census block groups in Detroit, Michigan, and showed that areas with a higher median household income and more homeowners experienced greater heating energy consumption. In a study of New York City, Kontokosta and Tull (2017) used machine learning and predicted the energy use of 1.1 million mixed-use buildings. The outcomes revealed that larger buildings consumed less energy per square foot, whereas taller buildings consumed more energy per square foot. Moreover, attached buildings exhibited a decrease in natural gas EUI due to the thermal properties of shared walls. In another study of large buildings in New York City, Movahedi and Derrible (2021) showed that the choice of technology for space and water heating, building use, and building density were the main drivers of energy and water usage. Shams Amiri et al. (2023) estimated building energy use for residential and commercial buildings for the years 2015 (base case) and 2045 (scenario) in Philadelphia and showed that the most influential features, after monthly electricity cost, included the presence of single-family attached units, number of rooms per housing unit, property value, monthly natural gas cost, household income, and number of bedrooms. In a study on Seattle, Ahn and Sohn (2019) suggested that larger dwelling unit sizes had lower EUI, while higher EUI values were observed in older buildings, households with higher income, and buildings with more units. In a study by Pesantez et al. (2023) in Chicago key factors influencing single-family daily electricity demand were median occupant age, the proportion of seniors, average commute time, and education level, with education showing the greatest impact. On a state level, Shen and Yang (2020a) explored the effects of climate change and urban growth on energy consumption in Texas, and showed that rising population, growth in urbanization, and fluctuating temperatures contributed to increased energy usage in the region. Tables 2 and 3 organize and summarize key findings from the literature review, categorizing studies based on household and occupant attributes (Table 2) and building, climatic, urban, and socioeconomic factors (Table 3) affecting U.S. residential energy consumption.

2.3. Research gaps and contributions

Previous studies have utilized RECS datasets up to 2015, but to the best of the authors' knowledge, there has yet to be a study on the RECS 2020 dataset with the outlined aim and objectives. This dataset is particularly distinctive for its increased number of features, including emerging technologies (e.g., EVs) and additional variables (e.g., DBT1 and DBT99), as well as its broader sample size and improved data granularity. While previous research has identified various determinants of energy consumption, a gap remains in understanding whether some of these features increase or decrease energy consumption. Moreover, fewer studies have focused on predicting state-level energy consumption and related determinants, a gap this study aims to address. By using a comprehensive set of features and analyzing their influence on energy patterns at both national and state levels, our study provides a holistic understanding of the key determinants of energy consumption, the magnitude and direction of their impact, and the underlying factors driving these relationships. We highlight the growing role of electrification, which has become more influential in shaping energy demand. The findings of this study offer new insights into policy design at both levels.

3. Methodology

This section outlines our research methodology, beginning with selection of the 2020 RECS dataset, followed by feature selection, data preprocessing, machine learning model application, and sensitivity

analysis. Fig. 1 provides an overview of the methodology framework.

3.1. 2020 RECS dataset

This research utilizes the 2020 RECS microdata provided by the U.S. Energy Information Administration (EIA). Drawn from the responses of 18,496 houses, this dataset is representative of the energy profiles of around 123.5 million individual homes and their occupants (U.S. EIA, 2023c). The dataset comprises over 300 attributes, encompassing topics such as household features, appliance usage, electronic devices, heating

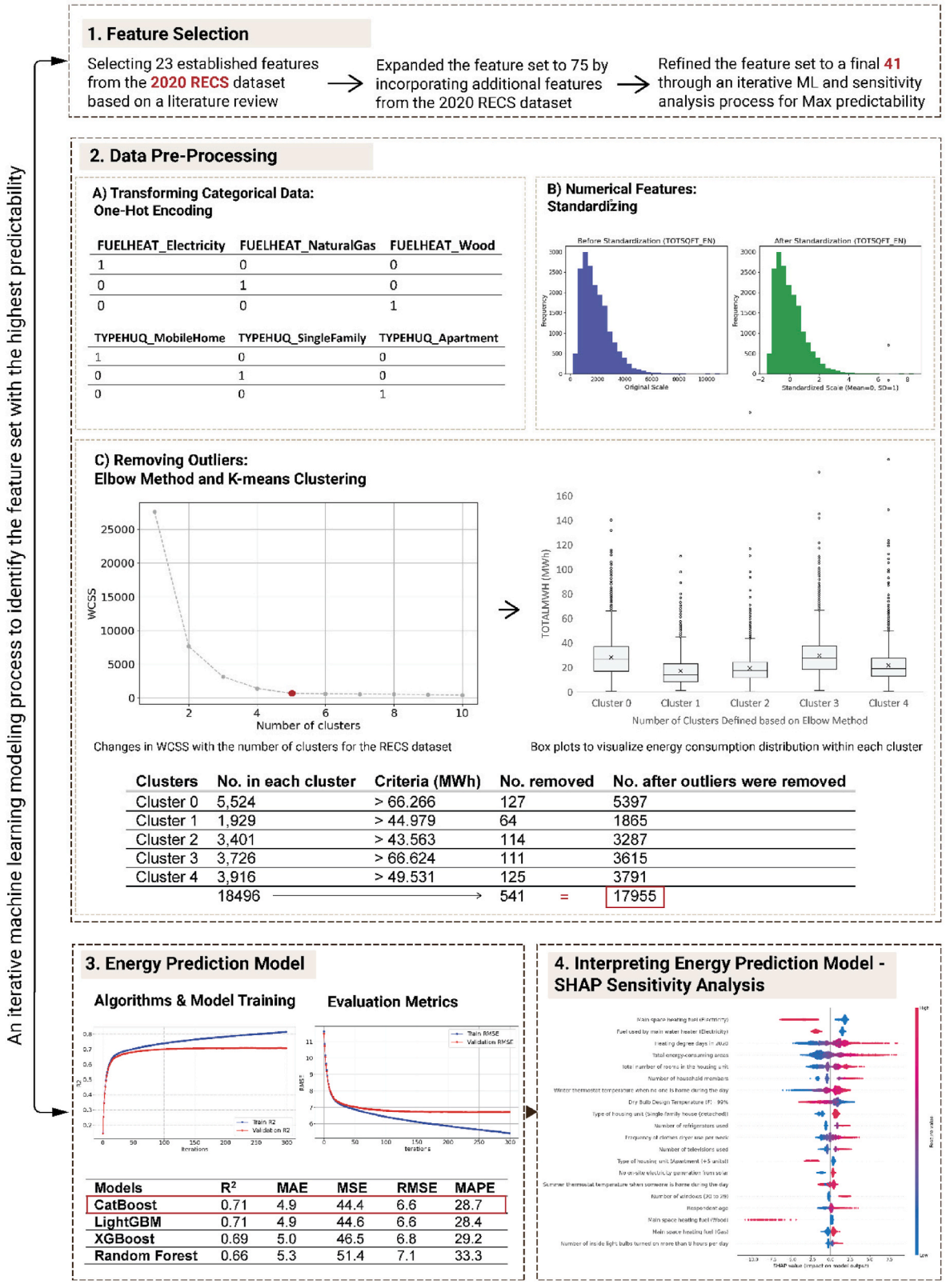


Fig. 1. An overview of the research methodology.

Table 1
Household attributes affecting the U.S. residential energy consumption (Estiri and Zagheni, 2019; Yun and Steemers, 2011; Sanquist et al., 2012a; Mostafavi, Farzinmoghadam, and Hoque, 2017a; Burnett and Lynne Kiesling, 2022; Wang et al., 2021; Goldstein et al., 2022; Bednar, Reames, and Keoleian, 2017a; Kontokosta and Tull, 2017a; Movahedi and Sybil, 2021; Shapley, 1953; Ahn and Sohn, 2019; Shen and Yang, 2020b; Zhang et al., 2023; Jiang et al., 2022; Abbasabadi et al., 2019; Pan and Zhang, 2020; Pesantez et al., 2023).

		Outputs (Type or unit of energy studied)											Households' Characteristics (sociodemographic and socioeconomic)								Households' Characteristics (Adaptive behaviors, occupancy, and comfort)					
References	Location	Total Energy Consumption	Energy intensity	EUI	Cooling Energy	Heating Energy	Lighting or Appliances	Water Consumption	Electricity use per capita	Electricity Consumption	Gas Consumption	Annual delivered energy per household	Age	Gender	Persons per household	Type of households (Single or multi-family)	Race	Income	Education	Employment situation	Home ownership	Number of Appliances and Equipment	Number of (air-conditioned) rooms	Occupied spaces / number of days at home	Heating setpoint temperature	Cooling setpoint temperature	The strongest explanatory feature(s)
(Estiri and Zagheni 2019)	U.S. National*	•											• (+)					• (+)									NA
(Yun and Steemers 2011)	U.S. National*				•								• (-)		• (+)			•(+)					•(+)				The frequency of AC use and CDD
(Sanquist et al. 2012a)	U.S. National*									•												•					
(Mostafavi, Farzinmoghadam, and Hoque 2017a)	U.S. National*				•	•	•	•					• (NC)		• (+)			• (NC)	• (NC)		• (+)						Housing Type and building size/Area
(Burnett and Kiesling 2022)	U.S. National*	•											• (+)	•	• (+)		• (NF)	• (NC)			•	• (+)		• (+)	• (+)		Total square feet of heated or cooled space
(Wang et al. 2021)	U.S. National*								•				• (+)												• (+)	• (-)	HDD, CDD, building area, number of rooms, and number of windows
(Goldstein, Reames, and Newell 2022)	U.S. National*			•													• (NF)				• (-)						Floor Area, Building age, Homeownership
(Bednar, Reames, and Keoleian 2017a)	Det, Urban			•		•							•		• (-)		• (NF)	•(NC)	•(-)		• (+)						NA
(Kontokosta and Tull 2017a)	NYC, Urban		•						•	•	•	•															NA
(Movahedi and Derrible 2021)	NYC, Urban							•		•	•								• (+)								Building Type, Technology used for Energy and building density
(Shams Amiri, Mueller, and Hoque 2023)	Phila, Urban	•																• (NC)									Electricity costs
(Ahn and Sohn 2019)	SEA, Urban			•														• (+)									Variations in building heights
(Shen and Yang 2020b)	TX, State	•																									Population
(Zhang et al. 2023)	SEA	•																									NA
(Jiang et al. 2022)	MTN, Urban			•									•	•		•	•	•	•	•				•			NA
(Abbasabadi et al. 2019)	CHI, Urban			•											•	•		•	•					• (+)			Gross Floor Area
(Pan and Zhang 2020)	SEA, Urban			•																							Gross Floor Area
(Pesantez, Wackerman, and Stillwell 2023)	CHI, Urban									•			•	•		•			•								Education and Occupant Age
This study	State, National	•											• (+)		• (+)			• (+)				• (+)			• (+)	• (-)	Electricity for space heating

• : a link between the feature and energy.
• (+): a positive link between the feature and energy.
• (-): a negative link between the feature and energy.
• (NC): Not Consistent – Different categories of the feature show varied effects.
• (NF): Nominal Features – Each category may relate differently to energy consumption.
* Using RECS Dataset

and cooling systems, lighting, and demographic information (U.S. EIA, 2023b). Table 3 shows that this dataset is distinguished from previous RECS datasets by including a higher number of features and a broader sample set. Larger sample sizes typically lead to reduced standard errors and narrower confidence intervals, particularly when estimating smaller subpopulations (U.S. Energy Information Administration, 2023). The final end-use consumption estimates were generated by calibrating a home's engineered end-use model outputs with 2020 billing data. This ensures that total estimates align with actual usage. Table 3 compares RECS datasets from 2001 to 2020. The data in this Table is organized based on the methodology section of each respective RECS.

3.2. Feature selection

Due to the large number of features in the 2020 RECS dataset, we pursued a systematic feature selection procedure to enhance model performance and reduce overfitting by retaining the most influential predictors. Step 1) We shortlisted an initial set of 23 influential features as suggested by previous studies (Estiri and Zagheni, 2019; Wang et al., 2021; Yun and Steemers, 2011; Mostafavi, Farzinmoghadam, and Hoque, 2017a; Burnett and Lynne Kiesling, 2022; Goldstein et al., 2022). The complete list can be found in Tables 1 and 2. The objective of this step is to ensure alignment with established research in the field. Step 2) Assuming that previous literature might not be comprehensive in their approach to the selection of features and because the 2020 RECS dataset incorporates more features compared with previous versions, we used our team's expert judgment to include additional features that could potentially affect residential building energy consumption. This led to an aggregate list of 75 potential features included in this work. Step 3) We trained the model using all 75 features, tested its performance, and applied SHAP sensitivity analysis to identify the most influential ones. We removed features with negligible impact based on SHAP values and re-evaluated the model to measure performance improvements. We repeated this iterative process, progressively refining the feature set by retaining the most influential features and removing the least impactful ones. This continued until we arrived at a final selection of 41 features (as shown in Table 4), which resulted in the highest predictive performance (R^2) and the lowest error metrics (RMSE, MSE, and MAE). The selection of features in this paper offers a thorough representation of various influencing categories. Note that we converted the unit of energy output in the original 2020 RECS dataset from thousand British Thermal Units (Btu) to Megawatt-hours (MWh) by applying the appropriate conversion rate, labeled 'TOTALMWH' in Table 4.

Adjusting for Household Variability: Our study explores the determinants of residential energy consumption at the state and national levels using household-level data from the RECS dataset. The 2020 RECS dataset applies weighting adjustments to ensure that the responding sample represents housing units at the national, census region, census division, and state levels. To align our analysis with this methodology and ensure that findings at broader geographic scales reflect population-level energy consumption rather than just an upward aggregation of household trends, we incorporated the NWEIGHT variable in our modeling. These statistical weights, as specified in the RECS methodology (U.S. EIA, 2023b), account for sampling probabilities, nonresponse adjustments, and post-stratification, making the dataset representative of the U.S. housing population. By incorporating NWEIGHT, we adjust for household-level variability, allowing our state- and national-level estimates to more accurately reflect broader residential energy consumption trends.

3.3. Data pre-processing

Following feature selection, we pursued a data pre-processing approach to prepare data for future machine learning steps of this

Table 2
Building, climatic, urban, and regional socio-economic attributes affecting the U.S. residential energy consumption (Estiri and Zagheni, 2019; Yun and Steemers, 2011; Mostafavi, Farzinmoghadam, and Hoque, 2017a; Burnett and Lynne Kiesling, 2022; Wang et al., 2021; Goldstein et al., 2022; Bednar, Reames, and Keoleian, 2017a; Kontokosta and Tull, 2017a; Movahedi and Sybil, 2021; Shapley, 1953; Ahn and Sohn, 2019; Shen and Yang, 2020b; Zhang et al., 2023; Jiang et al., 2022; Abbasabadi et al., 2019; Pan and Zhang, 2020).

	Building Characteristics													Climatic Characteristics						Urban Characteristics					Regional Socio-economic Characteristics				
Reference	Building Age	Housing Type (Ref: Mobile, detached)	Building Use (sector)	Rooms per dwelling, (total number of rooms	Building height / no. of floors	Size / Floor Area	Energy class (Ref: A/B)	Number of windows	Building insulation	Shape form factor	Vertical to horizontal Ratio	Buildings Width-length	Type of HVAC (Ref: Centrally heated or not)	Technology used for energy (Ref: electricity, gas)	HDD	CDD	Region (north, south, ...)	Mean temperature	SD of temperature	Urbanization level	Density	Lot coverage area	Vegetative fraction/ tree canopy area	Boroughs/ Regions	Height diversity	Energy price	Poverty rate	population	Housing Price and rent	Unemployment rate
(Estiri and Zagheni 2019)	• (+)	• (NF)				• (+)									• (+)														
(Yun and Steemers 2011)	• (NS)	• (NF)				• (+)								(NF)		• (+)													
(Mostafavi, Farzinmoghadam, and Hoque 2017a)	• (+)	• (NF)				• (+)														• (-)						•			
(Burnett and Kiesling 2022)		• (NF)		• (+)		• (+)									• (+)	• (+)	• (NC)									• (-)			
(Wang et al. 2021)			•		• (+)		• (+)	• (+)						•	• (-)														
(Goldstein, Reames, and Newell 2022)	• (+)					•																							
(Bednar, Reames, and Keoleian 2017a)	•	• (NF)		• (-)										• (NF)													• (-)		
(Kontokosta and Tull 2017a)	• (-)	• (NF)	• (NF)		• (+)	• (-)																		• (NF)					
(Movahedi and Dembie 2021)	• (-)		• (NF)											• (NF)							• (-)								
(Shams Amiri, Mueller, and Hoque 2023)		•	• (NF)	• (+)	• (+)	•															• (+)					• (+)		•	
(Ahn and Sohn 2019)	• (+)			• (+)		• (-)	• (-)														• (-)	• (-)	• (+)		• (+)				
(Shen and Yang 2020b)		• (NF)													• (+)	• (-)		• (+)	• (+)	• (+)								• (+)	
(Zhang et al. 2023)	•				• (+)	• (+)	• (-)			• (+)	• (+)	• (+)		• (NF)							• (+)		•						
(Jiang et al. 2022)	• (+)		• (NF)	•	•	•																		• (NF)			•	•	
(Abbasabadi et al. 2019)	•				•	•																					•		•
(Pan and Zhang 2020)	• (+)	•			• (+)	• (+)	• (-)										•				•								
This study		• (NF)		• (+)		• (+)		• (+)						• (NF)	• (+)														

research. This step involved the treatment of categorical features, standardization of numerical features, and outlier removal.

Treatment of categorical features: We applied one-hot encoding to convert each categorical feature into multiple binary numerical features to ensure that each category is distinctly represented. Each category is encoded as a vector, where 1 ("hot") indicates the presence of a category, and 0 ("cold") indicates its absence (Aurélien, 2017). One-hot encoding was specifically chosen to enhance interpretability and capture the directional impact of categorical variables on energy consumption. For instance, rather than using a single categorical feature like 'fuel for heating,' one-hot encoding transforms it into multiple binary features such as 'electricity for heating,' which allows us to analyze the specific impact of each heating type on energy consumption separately. Additionally, our machine learning models are well-equipped to handle high-dimensional data and mitigate potential overfitting concerns.

Standardization: We applied the standardization process to all the numerical features to ensure uniform scaling and compatibility with various modeling techniques. This process involves rescaling the numerical features so that they would have a mean of zero and a standard deviation of 1 (Scikit-learn Developers, 2024).

Removing Outliers: To systematically handle outliers, we first applied a clustering method to group observations (data points) of similar quality into distinct categories. This clustering process helped ensure that data points with similar characteristics were analyzed together. Within each category, we then identified and removed extreme values in energy consumption, as these were considered outliers that could distort the analysis. For clustering, we used K-means, an unsupervised learning algorithm that divides a dataset into K clusters based on similarities (Jain, 2010). It starts by initializing K centroids and iteratively refines them by grouping data points closest to each centroid and recalculating the averages (Lloyd, 1982). The goal is to minimize the average squared distance within clusters (Arthur and Vassilvitskii, 2007), with the optimal number of clusters, K, often determined by the Elbow method, which uses the Within-Cluster Sum-of-Squares (WCSS) to measure variance (M. Cui, 2020). The method identifies the point where increasing K yields diminishing returns. Fig. 1 (Data pre-processing, part C) displays how WCSS changes with the number of clusters for the RECS dataset, leading to the selection of 5 clusters. To identify outliers, we used box plots to visualize the distribution of total energy consumption within each cluster, Fig. 1 (Data pre-processing, part C). We identified data points exceeding the upper limits of each box plot as outliers and removed them. Clusters Table in Fig. 1 (Data pre-processing, part C) summarizes the count of data items in each cluster, the criteria used for outlier removal (data points exceeding the upper limits of each box plot), and the number of data items excluded from each cluster. This approach led to the removal of a total of 541 data items from the dataset, hence, 17,955 data points were kept in the analysis.

3.4. Machine learning

In this study, we employed tree-based machine learning algorithms, including CatBoost, XGBoost, LightGBM, and Random Forest, due to their ability to capture complex nonlinear relationships, robustness to feature scaling, and effectiveness in handling categorical and missing data. These characteristics make them particularly well-suited for structured datasets, such as those used in energy consumption analysis. All four algorithms belong to the ensemble learning family, with Random Forest employing bagging (bootstrap aggregation) to enhance accuracy and robustness by aggregating outputs from multiple decision trees (Kontokosta and Tull, 2017a). In contrast, CatBoost, LightGBM, and XGBoost use boosting techniques to iteratively improve predictive performance by combining multiple weak learners and focusing on correcting errors from previous iterations (Prokhorenkova et al., 2017). CatBoost stands out for its ability to handle categorical features natively and not need extensive hyperparameter tuning (Pan and Zhang, 2020; Prokhorenkova et al.). Additionally, it effectively manages multicollinearity through its ordered boosting algorithm and built-in regularization techniques, which mitigate overfitting and reduce the influence of highly correlated features on model predictions (Hancock and Khoshgoftaar, 2020). LightGBM uses leaf-wise tree growth instead of the widely used level-wise tree growth to speed up the training by reducing the number of tree nodes (Gan et al., 2021), which makes it suitable for handling large datasets. XGBoost overcomes the weaknesses of decision trees and effectively controls the bias-variance trade-off (Shapley, 1953).

In our study, 80 % of the data was allocated to the training set and the remaining 20 % was assigned to the test set. We also used several evaluation metrics to assess the performance of the models including R^2, adjusted R^2, MAE, MSE, RMSE, and MAPE (James et al., 2013). R^2 (Equation (1)) measures how much of the variance in TOTALMWH is explained by the model, while Adjusted R^2 (Euquation 2) accounts for the number of predictors (Chicco et al., 2021), penalizing for overfitting to provide a more accurate assessment of model performance. For both metrics, values closer to 1 indicate a better fit.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Predicted\ TOTALMWH_i - Actual\ TOTALMWH_i)^2}{\sum_{i=1}^{n}(Actual\ TOTALMWH_i - Mean\ of\ TOTALMWH_i)^2}$$

Equation (1)

$$Adjusted\ R^2 = 1 - (1 - R^2) * \frac{n-1}{n-p-1}$$

Equation (2)

Where:

- n = Total number of observations
- p = Number of predictors

MAE measures the average absolute error, MSE calculates the average squared error, RMSE represents the square root of MSE (Willmott and Matsuura, 2005) (in the same units as TOTALMWH), and MAPE expresses the average percentage error between actual and predicted TOTALMWH values; for all, lower values indicate better model performance.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Actual\ TOTALMWH_i - Predicted\ TOTALMWH_i|$$

Equation (3)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Actual\ TOTALMWH_i - Predicted\ TOTALMWH_i)^2$$

Equation (4)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Actual\ TOTALMWH_i - Predicted\ TOTALMWH_i)^2}$$

Equation (5)

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Actual\ TOTALMWH_i - Predicted\ TOTALMWH_i}{Actual\ TOTALMWH_i}\right| * 100$$

Equation (6)

3.5. Sensitivity analysis and SHAP

Interpreting machine learning models is a critical challenge in the field of energy prediction, as these models often function as opaque systems with limited transparency. This lack of interpretability not only creates challenges for researchers trying to understand the factors influencing the predictions but also hinders adoption by stakeholders, such as building owners, who require actionable insights rather than abstract outputs. Addressing this limitation requires methods that can demystify the model's behavior and attribute the influence of input features on its predictions.

One effective solution to this challenge is Shapley Additive Explanations (SHAP), which is derived from the Shapley value, a concept rooted in cooperative game theory (Shapley, 1953). Mathematically, the Shapley value (φ_i) is expressed as:

$$\varphi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!}\left[f(S \cup \{i\}) - f(S)\right]$$

Equation (7)

Where:

- *F:* The set of all input features.
- *S:* A subset of features excluding the feature iii.
- $f(S \cup \{i\}) - f(S)$*:* The change in the model's prediction for TOTALMWH when feature i is included versus excluded (Shapley, 1953; Lundberg and Lee, 2017).

Table 3
Comparison of RECS datasets.

RECS Dataset Year	Sample Size (Households)	Number of Features (Approx.)	Estimated U.S. Households Represented (millions) (Approx.)	Data Collection Method	Key Survey/Questionnaire Changes
2020	18,496	~300	~123.5	Self-administered web & paper questionnaires	- First RECS with state-level estimates for all 50 states & DC - Updating and adding more questions (e.g., on electric vehicles) and features (e.g., DBT1 and DBT99) - New square footage estimation method (self-reported instead of interviewer-measured) - Billing-level energy modeling calibration
2015	5686	~192	~118.2	Self-administered web, paper questionnaires and computer-assisted personal interviews (CAPI)	- Introduction of engineering-based end-use modeling instead of statistical regression - Improved questionnaire with open-ended appliance use responses instead of categorical choices - First RECS to use minimum variance estimation for calibration
2009	12,083	~120	~113.6	Computer-assisted personal interviews and mails	- Last RECS to use statistical regression models for end-use estimation - Household square footage measured by interviewers - Higher sample size than 2015 RECS but less advanced calibration methods
2005	4382	~112	~111.1	Computer-assisted personal interviews and mails	- Basic statistical modeling for energy use estimation - Simpler questionnaire design (fewer features collected compared to later years)
2001	4822	~98	~107	Computer-assisted personal interviews and mails	- Basic statistical modeling for energy use estimation

Table 4
The final dataset, including 41 features and one output.

Variables	Description and Labels	Response Codes
State_name	State Name	1. Hawaii 2. Florida …. 50. North Dakota 51. Alaska
BA_CLIMATE	Building America Climate Zone	1. Subarctic 2. Very-Cold 3. Cold 4. Mixed-Dry 5. Mixed-Humid 6. Marine 7. Hot-Dry 8. Hot-humid
UATYP10	2010 Census Urban Type Code	1. Rural area 2. Urban Cluster 3. Urban area
HDD65	Heating degree days in 2020, base temperature 65F	0–17383
CDD65	Cooling degree days in 2020, base temperature 65F	0–5534
DBT1	Dry Bulb Design Temperature (F) - temp expected to be exceeded 1 % of the time	57.7–111.4
DBT99	Dry Bulb Design Temperature (F) - temp expected to be exceeded 99 % of the time	−44.9–67.5
TYPEHUQ	Type of housing unit	1 Mobile home 2 Single-family house (detached) 3 Single-family house (attached) 4 Apartment (2–4 units) 5 Apartment (+5 units)
STORIES	Number of stories in a single-family home	1 One story 2 Two stories 3 Three stories 4 Four or more stories 5 Split-level −2 Not applicable
KOWNRENT	Own or rent	1 Own 2 Rent 3 Occupy without payment of rent
YEARMADERANGE	Range when housing unit was built	1 Before 1950 2 1950 to 1959 3 1960 to 1969 4 1970 to 1979 5 1980 to 1989 6 1990 to 1999 7 2000 to 2009 8 2010 to 2015 9 2016 to 2020
TOTROOMS	Total number of rooms in the housing unit, excluding bathrooms	1–15
ROOFTYPE	Major roofing material	1 Ceramic or clay tiles 2 Wood shingles/ shakes 3 Metal 4 Slate or synthetic slate 5 Shingles 6 Concrete tiles 99 Other −2 Not applicable
WINDOWS	Number of windows	1 1 or 2 windows 2 3 to 5 windows 3 6 to 9 windows 4 10 to 15 windows 5 16 to 19 windows 6 20 to 29 windows 7 30 or more windows
TYPEGLASS	Type of glass in most windows	1 Single-pane glass 2 Double-pane glass 3 Triple-pane glass
FUELPOOL	Fuel used for heating swimming pool	0 None, not heated 5 Electricity 1 Natural gas from underground pipes 2 Propane (bottled gas) 3 Fuel oil 99 Other −2 Not applicable
NUMFRIG	Number of refrigerators used	0–9
DWASHUSE	Frequency of dishwasher use per week	0 - 21 −2 Not applicable
DRYRUSE	Frequency of clothes dryer use per week	0 - 30 −2 Not applicable
TVCOLOR	Number of televisions used	0–14
DESKTOP	Number of desktop computers used	0–8
EQUIPM	Main space heating equipment type	3 Central furnace 2 Steam or hot water system with radiators or pipes 4 Central heat pump 13 Ductless heat pump 5 Built-in electric units 7 Built-in room heater burning gas or oil 8 Wood or pellet stove 10 Portable electric heaters 99 Other −2 Not applicable
FUELHEAT	Main space heating fuel	5 Electricity 1 Natural gas from underground pipes 2 Propane (bottled gas) 3 Fuel oil 7 Wood or pellets 99 Other −2 Not applicable
NUMCFAN	Number of ceiling fans used	0–15
TYPETHERM	Type of thermostat used	1 Yes, a manual or non-programmable thermostat 2 Yes, a programmable thermostat 3 Yes, a “smart” or Internet-connected thermostat 0 Does not have thermostat −2 Not applicable
TEMPGONE	Winter thermostat temperature in home when no one is home during the day	50 - 90 −2 Not applicable
TEMPHOMEAC	Summer thermostat temperature in home when someone is home during the day	50 - 90 −2 Not applicable
FUELH2O	Fuel used by main water heater	5 Electricity 1 Natural gas from underground pipes 2 Propane (bottled gas) 3 Fuel oil 7 Wood 8 Solar thermal 99 Other
LGTIN1TO4	Number of inside light bulbs turned on 1–4 h per day	0–90
LGTIN4TO8	Number of inside light bulbs turned on 4–8 h per day	0–84
LGTINMORE8	Number of inside light bulbs turned on more than 8 h per day	0–99

(continued on next page)

Table 4 (*continued*)

Variables	Description and Labels	Response Codes
LGTOUTNITE	Number of outside light bulbs left on all night	1 - 65 −2 Not applicable
SOLAR	On-site electricity generation from solar	1 Yes 0 No −2 Not applicable
HHAGE	Respondent age (top-coded)	18–90
EMPLOYHH	Respondent employment status	1 Employed full-time 2 Employed part-time 3 Retired 4 Not employed
EDUCATION	Highest level of education completed by respondent	1 Less than high school diploma or GED 2 High school diploma or GED 3 Some college or Associate's degree 4 Bachelor's degree 5 Master's, Professional, or Doctoral degree
HOUSEHOLDER_RACE	Householder (respondent) race	1 White Alone 2 Black or African/ American Alone 3 American Indian or Alaska Native Alone 4 Asian Alone 5 Native Hawaiian or Other Pacific Islander Alone 6 2 or More Races Selected
NHSLDMEM	Number of household members (top-coded)	1–7
MONEYPY	Annual gross household income for the past year	1 Less than $5000 2 $5000 - $7499 3 $7500 - $9999 4 $10,000 - $12,499 5 $12,500 - $14,999 6 $15,000 - $19,999 7 $20,000 - $24,999 8 $25,000 - $29,999 9 $30,000 - $34,999 10 $35,000 - $39,999 11 $40,000 - $49,999 12 $50,000 - $59,999 13 $60,000 - $74,999 14 $75,000 - $99,999 15 $100,000 - $149,999 16 $150,000 or more
TOTSQFT_EN	Total energy-consuming area (square footage) of the housing unit.	200–15000
NWEIGHT	Final Analysis Weight	437.9–29279.1
TOTALMWH	Total usage including electricity, natural gas, propane, and fuel oil, in Megawatt-hours, 2020	0.34639046-400.6916

The SHAP value for each feature explains how the presence (or absence) of that feature changes the models' prediction from the baseline (Lundberg and Lee, 2017). In the context of SHAP analysis, the "baseline" is often referred to as the average prediction of the model over a reference dataset, typically the entire training dataset.

In this study, we used SHAP to identify and interpret the key features driving the predictions for TOTALMWH to ensure transparency and understand the model's decision-making process. The SHAP evaluation was carried out on the test set to ensure its findings are generalizable to the average house in the U.S.

We focused on the application of existing algorithms without aiming for their enhancement or modification, therefore, we did not provide the in-depth theoretical and mathematical foundations of these methods. We used Google Colab for cloud-based execution in an interactive environment.

4. Results

We conducted energy modeling at both the national and state levels to provide a comprehensive understanding of residential energy consumption. The national-level analysis offers insights into aggregate energy usage that can inform the development of unified, large-scale energy policies. Conversely, state-level modeling is essential for capturing regional variations in building practices.

4.1. National energy consumption modeling

We used various tree-based machine learning models to predict national energy consumption and compared their performance, as shown in Table 5. CatBoost emerges as the best-performing model with the highest R^2 value of 0.71. This indicates that the model explains 71 % of the variance in national residential energy consumption. Additionally, CatBoost's MAE, MSE, and RMSE scores are among the lowest, which suggests smaller prediction errors compared to other models. The robust performance of CatBoost highlights its effectiveness in capturing the complex relationships between various features and energy consumption.

4.2. Determinants of national residential energy consumption

Based on the most reliable model for national energy consumption (CatBoost), we conducted a SHAP sensitivity analysis to identify the key determinants influencing energy consumption. Fig. 2 displays a beeswarm plot of the most influential features and showcases the distribution of their SHAP values across individual samples. Features are ranked by the aggregate impact they have on model output. The horizontal axis represents the SHAP value, and the color gradient (from blue to red) indicates the feature value, with red representing higher values. Features on the right side of the graph with higher SHAP values, represented in red, increase energy consumption. This means that an increase in their values or presence is linked to higher energy use. These features include HDDs, energy-consuming areas, total number of rooms, number of household members, winter thermostat setting when no one is home, single-family detached homes, number of refrigerators, frequency of clothes dryer use, number of televisions, no on-site electricity generation from solar, number of windows (20–29), respondent age, use of gas for space heating, and the number of inside light bulbs that are on for more than 8 h a day. Features on the left side of the graph with higher SHAP values reduce energy consumption, meaning their increased values or presence are linked to lower energy use. These features include the use of electricity for space and water heating, Dry Bulb Design Temperature (F) −99 %, apartments with more than five units, and the use of wood for space heating. DBT99, a measure of extreme cold temperature that is exceeded only 1 % of the time in a given year, was not previously analyzed in the context of U.S. national energy consumption and was introduced as a significant determinant of energy use in this study.

Fig. 3 ranks the key features in a model based on their average absolute SHAP values, which measure the overall impact of each feature on the model's output, regardless of the direction of their effect. Higher SHAP values indicate a greater influence. The model predicts an average energy consumption of 23.173 MWh for a typical house in the dataset, assuming no specific attributes are considered. Key features that predominantly influence the energy model alongside their average absolute contributions are electricity for main space heating (2.35 MWh), electricity for main water heater (1.67 MWh), HDD65 (1.63 MWh), energy-consuming areas (1.58 MWh), total number of rooms (1.27 MWh), number of household members (1.21 MWh), winter temperature setting when no one is at home (1.14 MWh), Dry Bulb Design Temperature (1.04 MWh), single-family detached homes (0.87 MWh) and number of refrigerators (0.73). For context, these contributions need to be

evaluated in relation to a base value of 23.173 MWh. The most influential features affecting energy consumption in this study are presented in Tables 1 and 2 to provide a comparison with existing literature.

4.3. State-level energy models

We modeled each state's energy consumption separately using CatBoost to analyze the data at a more granular level for key determinants that may not be visible at a national level. Fig. 4 shows the coefficient of determination (R^2) values for estimating energy consumption in each state. Around 30 % of the states showed a strong fit ($R^2 > 0.65$), 57 % showed a moderate fit ($0.5 \leq R^2 \leq 0.65$), and 13 % showed a weak fit ($R^2 < 0.5$). This indicates that for approximately 87 % of the states, the model's performance is moderate to strong, suggesting reliable predictive capability in most cases. In states with lower model performance (e.g., Louisiana, Mississippi, Iowa, and Montana) the features presented may not fully account for most energy predictions. This suggests that additional determinants not presented in the RECS dataset could be driving energy consumption in those regions.

4.4. Determinants of states' residential energy consumption

We conducted a SHAP sensitivity analysis to identify the key determinants influencing energy consumption in each state. Fig. 5 shows a bubble graph of the top five factors influencing residential energy consumption across states. The size of the bubbles represents the importance of each feature's impact on the energy consumption model. A red color indicates that an increase in the feature value or its presence leads to higher energy consumption, whereas a blue color suggests that an increase or presence of the feature reduces energy consumption. The bubble graph shows that the top five factors affecting residential energy consumption at the state level are energy-consuming area (98 % of states), heating fuel type (90 %), total number of rooms (65 %), housing type (41 %), number of appliances and their frequency of operation (39 %), household size (35 %), thermostat settings (12 %) and on-site solar generation (10 %). Higher energy consumption in states is linked to larger energy-consuming areas, more rooms, single-family detached homes, gas and oil heating, a higher number and greater use of appliances, larger households, higher winter thermostat settings, lack of on-site solar power, and annual incomes over $150,000.

5. Discussion

5.1. Key determinants: a RECS-Based analysis

Studies on previous RECS datasets have consistently identified building size, housing type (e.g., detached homes), building age, HDD, CDD, number of household members, occupant age, income, and the frequency of appliance use as key determinants of residential energy consumption (Tables 1 and 2) (X. Cui et al., 2024; Goldstein et al., 2022; Burnett and Lynne Kiesling, 2022; Estiri and Zagheni, 2019; Mostafavi, Farzinmoghadam, and Hoque, 2017b; Sanquist et al., 2012b; Yun and Steemers, 2011). While our study confirms the significance of building size, housing type, and HDD, we observe notable shifts in the relative importance of certain factors. Electrification, particularly the use of electricity for space and water heating, has emerged as a dominant driver of energy use, which reflects changes in energy infrastructure. Additionally, climate-related factors such as Dry Bulb Design Temperature, occupant-related factors like winter thermostat settings and on-site electricity generation have become more important. Interestingly, while income and occupant age were historically among the strongest predictors, our findings suggest that electrification trends, occupant behaviors, and technology adoption are now playing a greater role in reshaping residential energy consumption patterns.

5.2. The dominance of heating-related factors

An overview of our findings shows that heating-related factors such as fuel type for space and water heating, HDD65, Dry Bulb Design Temperature – 99 %, and winter thermostat settings have a greater impact on energy use than cooling-related factors like summer thermostat settings, CDD65, and DBT1. This finding can be supported by the fact that heating demands generally surpass cooling demands, particularly in colder regions such as the Midwest and Northeast, where extended winters drive higher energy consumption. Additionally, heating systems rely on a mix of fuel sources, including natural gas, electricity, and oil, each with different efficiencies, whereas cooling is primarily electricity-based. The importance of HDD65 and DBT99 in our analysis further highlights the dominance of heating-related factors, as extreme cold conditions lead to greater energy use for space heating compared to cooling needs in warmer climates.

5.3. Key determinants and their underlying drivers

5.3.1. Heating fuels and systems

Heating Fuel: Our results indicate that at both state and national levels, using electricity as the primary fuel for space and water heating significantly reduces on-site energy consumption, whereas gas use for the same purposes increases it. This finding aligns with existing research in the U.S., which shows that greater reliance on electricity is associated with lower heating energy use and EUI (Bednar, Reames, and Keoleian, 2017b), while higher natural gas use tends to increase both energy consumption and EUI (Zhang et al., 2023; X. Cui et al., 2024). The main reason is that natural gas heating is more energy-intensive on-site compared to electric heating, due to fuel conversion losses occurring within the furnace (Hojjati and Wade, 2012). Based on the RECS dataset, houses utilizing electricity for primary space heating reported a mean and median heating EUI of 0.0294 and 0.0206 MWh/m^2, respectively. These values are significantly lower than 0.0951 and 0.080 MWh/m^2 when gas dominates as the main heating fuel in houses. In this dataset, gas and electricity represent 52 % and 30 % of the heating system fuels respectively, therefore, we study the mean and median heating EUI for houses using these two fuels across different heating systems. For our analysis, we use EUI to allow for a standardized comparison of energy performance across different heating systems regardless of house size.

Heating Systems: Fig. 6 shows that houses equipped with central heat pumps have the lowest heating EUI, with mean and median of 0.0237 and 0.0175 MWh/m^2 compared to residences with other electric and gas-based heating systems. In contrast, houses with the gas-operated central furnace, the most common system in this dataset, exhibit a heating EUI roughly 3.9 times higher than those with central heat pumps, with a mean and median of 0.0918 and 0.0776 MWh/m^2. This is mainly because modern electric heating systems, such as heat pumps, are often more efficient than traditional combustion-based heating methods. According to the U.S. Department of Energy, heat pumps transfer heat instead of generating it, therefore, they can reduce electricity use for heating by around 50 % relative to electric resistance heating such as furnaces and baseboard heaters (U.S. Department of Energy). Also, combustion processes in furnaces, boilers, and other equipment are not 100 % efficient. In practical terms, not every joule of energy contained in the fuel gets converted into useful heat. A portion of this energy is lost as waste heat, or in incomplete combustion, and is

Table 5
Performance comparison of tested ML models.

Models	R^2	Adjusted R^2	MAE	MSE	RMSE	MAPE
CatBoost	0.712	0.700	4.958	44.477	6.669	28.712
LightGBM	0.711	0.696	4.937	44.670	6.683	28.432
XGBoost	0.699	0.683	5.074	46.574	6.824	29.268
Random Forest	0.667	0.650	5.373	51.427	7.171	33.360

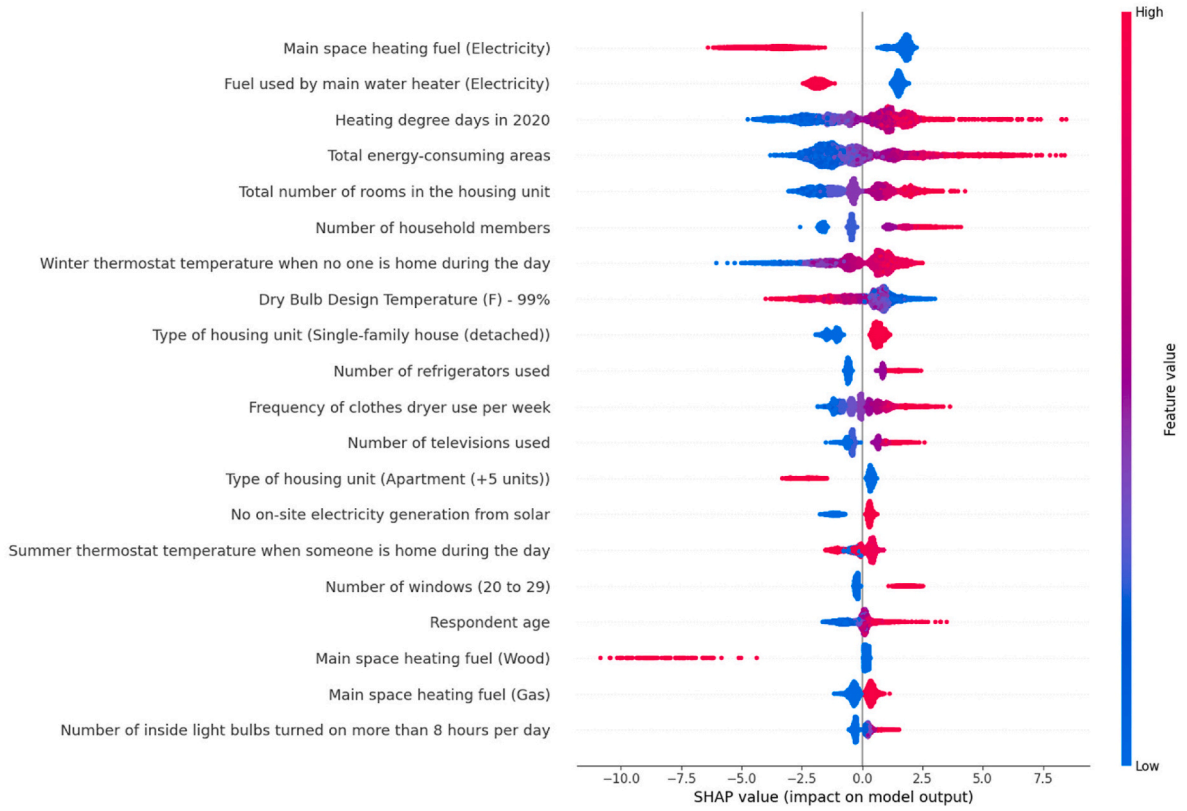


Fig. 2. The plot of SHAP values for each feature across all samples.

vented away (Moran et al., 2010).

Water Heating Systems: We also studied the distribution of water heating EUI in houses with various fuels (Fig. 6), excluding wood and solar thermal due to their low frequency. Houses with electric water heaters demonstrated lower water heating EUI, with mean and median values of 0.0224 and 0.0170 MWh/m^2, in comparison to propane, gas, and oil water heaters. In contrast, houses with oil-operated water heaters showed the highest heating EUI, with average and median values at 0.0429 and 0.0362 MWh/m^2, respectively, indicating approximately double consumption compared to houses with electric water heaters. This is because electric resistance water heaters are highly efficient and convert most of the electricity they use into heat. Also, electric on-demand or tankless water heaters heat water directly without needing a storage tank, which eliminates the standby heat losses that storage water heaters experience when stored water gradually releases heat to the environment (U.S. DOE).

5.3.2. *Housing type*

The RECS dataset shows that the average and median EUI for single-family (detached) homes are 0.169 and 0.151 MWh/m^2 (Fig. 6). In contrast, apartments (+5 units) have a mean and median EUI of 0.126 and 0.111 MWh/m^2. This is because apartments in multi-unit buildings tend to have reduced heat loss in colder climates and reduced heat gain in warmer climates due to shared walls, floors, and ceilings (Kontokosta and Tull, 2017b; Namazkhan et al., 2020). In contrast, detached single-family homes have more surface area exposed to outdoor conditions (Navamuel et al., 2018), resulting in greater heat loss in winter and heat gain in summer. There is substantial evidence suggesting that detached single-family homes consume more energy compared to more compact housing configurations like apartments (Estiri and Zagheni, 2019; Mostafavi, Farzinmoghadam, and Hoque, 2017a; Bednar, Reames, and Keoleian, 2017a). Burnett and Lynne Kiesling (2022) investigation of the 2001–2015 RECS dataset found that apartment occupants use 12–40 % less energy than those in mobile homes. Shen and Yang (2020) study on Texas residential energy consumption showed that a 1 % increase in apartments with more than five units could save 2216 GWh in total primary energy use.

5.3.3. *Occupant characteristics*

Our national- and state-level analysis shows that occupant behavior, demographic characteristics, and economic status influence residential energy consumption.

Thermostat temperature: Winter thermostat temperature is a key factor in national-scale energy consumption and ranks among the top five contributors in Colorado, Connecticut, Maine, Massachusetts, New York, and South Dakota (Fig. 5). Higher setpoint temperatures are associated with increased energy consumption, supported by existing literature (Belaïd et al., 2019; Staffell et al., 2023; Mohammadiziazi et al., 2021). This means that even small adjustments to thermostat settings can lead to significant variations in energy consumption. As a result, occupant behavior in managing indoor temperatures plays a crucial role in overall energy demand.

Respondent's Age: National-level analysis shows that energy consumption increases with occupant age, consistent with previous studies (Estiri and Zagheni, 2019; Froemelt et al., 2020). This trend is largely driven by older adults' preference for higher indoor temperatures, as their lower metabolic rate and reduced thermoregulatory response affect heat retention and perception (van Hoof et al., 2017). However, the relationship can reverse for cooling demand, as older occupants tend to use less air conditioning (Yun and Steemers, 2011), which has implications for energy planning in warmer climates.

Home Ownership: Homeownership was not a strong determinant of national energy consumption; however, in North Dakota and Wyoming, it ranked among the top five factors, with rented homes consuming less energy than owned homes (Fig. 5). Fig. 7 shows that the mean total energy consumption for owned homes is 25.95 MWh, whereas rented homes consume significantly less at 14.84 MWh. This pattern aligns with existing research, which has consistently shown that an increase in home

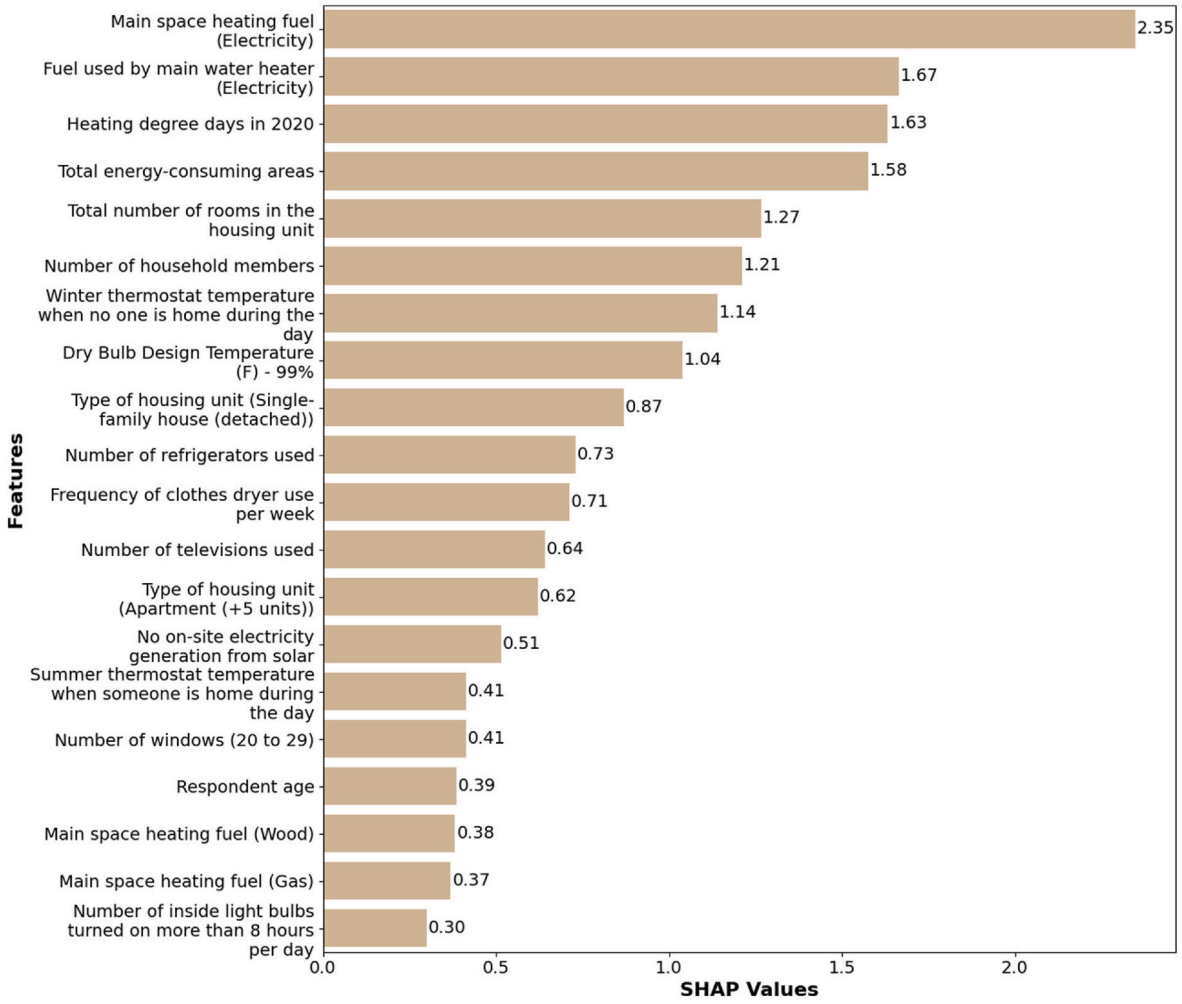


Fig. 3. Features ranked based on their average absolute SHAP values.

ownership is associated with higher total energy consumption (Belaïd, 2016; Bednar, Reames, and Keoleian, 2017b; Porse et al., 2016). However, when we normalized energy consumption by home size, the difference was less pronounced. EUI values are 0.15 MWh/m^2 for owned homes and 0.14 MWh/m^2 for rented homes (Fig. 7). This suggests that homeowners' higher energy use may be attributed to the larger size of their homes. It is also shown that homeowners generally use more electricity due to their ability to purchase more appliances (Huang, 2015) (see Fig. 7).

5.4. Policies at the national and state level

To suggest policies based on key determinants, we focus on the top five, as they have the greatest impact on energy use. The top five determinants of energy consumption at the national level are primarily related to electricity for heating, climatic conditions, and building size, which reflects broader trends in residential energy use across the country. However, at the state level, alongside these dominant determinants, the top five also incorporate a wider range of factors such as housing type, appliance number and usage, on-site electricity generation, socioeconomic and demographic characteristics (including homeownership, income, and occupant age), and type of space heating equipment. This variation suggests the importance of studying energy consumption at a more granular level, as national-scale analyses may overlook state-specific patterns and localized drivers of energy use. Fig. 8 presents a heatmap of key features and the policies inferred from them at both levels.

Electric Heating Systems: One of the most important factors at both scales is the type of fuel used for heating. We showed that using electricity for space and water heating is associated with lower on-site residential energy consumption. Looking deeper into the reasons behind this, we linked it to the type of heating and water heating systems. Our analysis demonstrated that electric heat pumps have the lowest heating EUI, making them a strong candidate for policy consideration. 'Electrification Futures Study' by the National Renewable Energy Laboratory (Mai et al., 2018) along with several other studies (Nyangon and Byrne, 2021; Goldstein et al., 2022; Andreou et al., 2020), have emphasized the benefits of adopting high-efficient heat pumps and water heaters for substantial energy savings. However, as more households transition to electric heating, this shift will place additional demand on the power grid. To fully realize the benefits of electrification, grid infrastructure should be strengthened to handle higher demand, ensure reliable power supply, and minimize transmission losses.

Efficient Occupant Behavior: Thermostat temperature setting ranks among the top five features in 12 % of states. This emphasizes that even small adjustments in occupant behavior, such as lowering the thermostat setpoint during the heating season and raising it during the non-heating season, directly influence energy consumption, leading to significant energy savings. Developing energy-conscious habits, such as mindful thermostat adjustments, is crucial for long-term efficiency, and studies show that educating consumers on such habits is an effective strategy for reducing energy use (Hu et al., 2019; Nsangou et al., 2022; Estiri, 2014). For example, Sakah et al. (2019) advocate for large-scale energy conservation campaigns on major television networks and

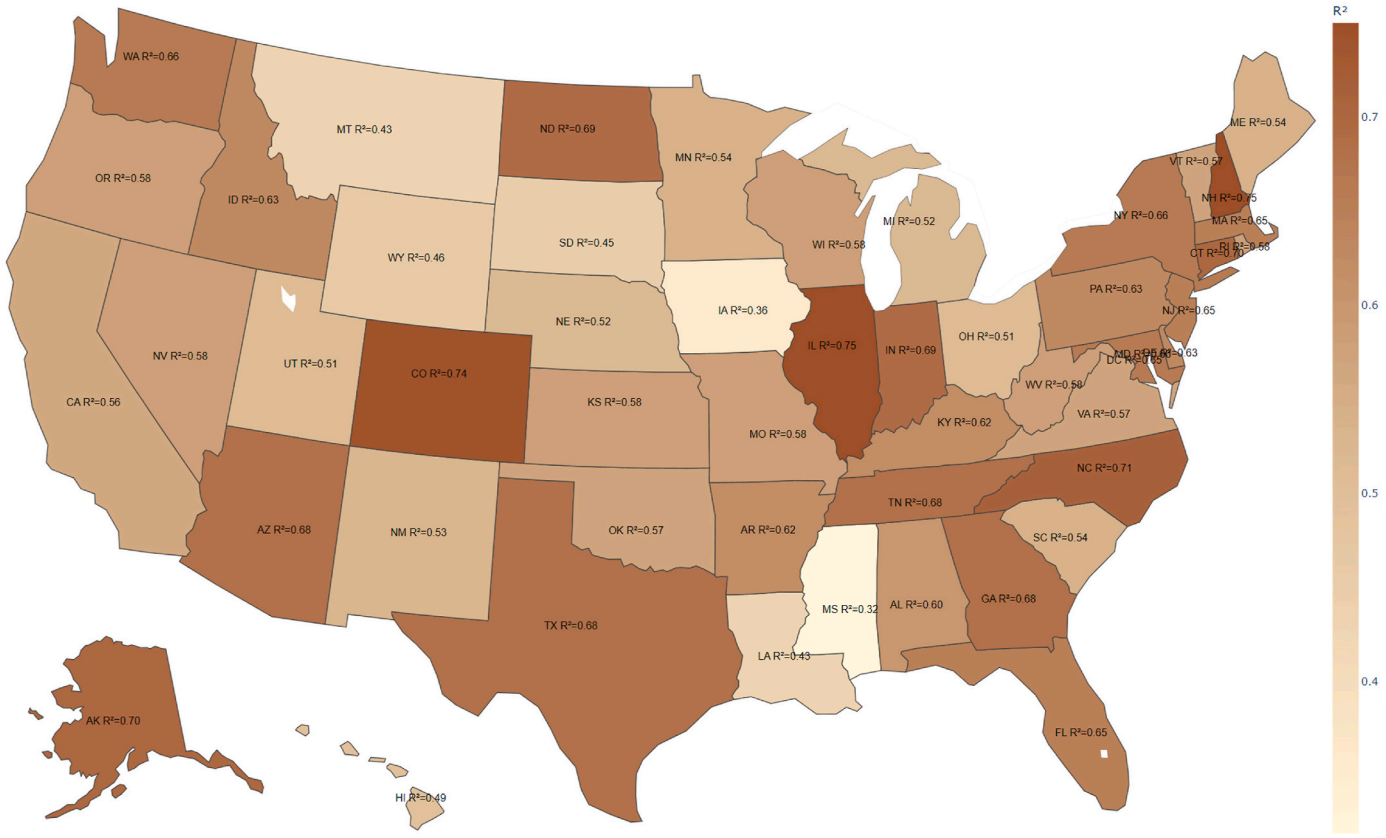


Fig. 4. R^2 values for energy consumption models in each state.

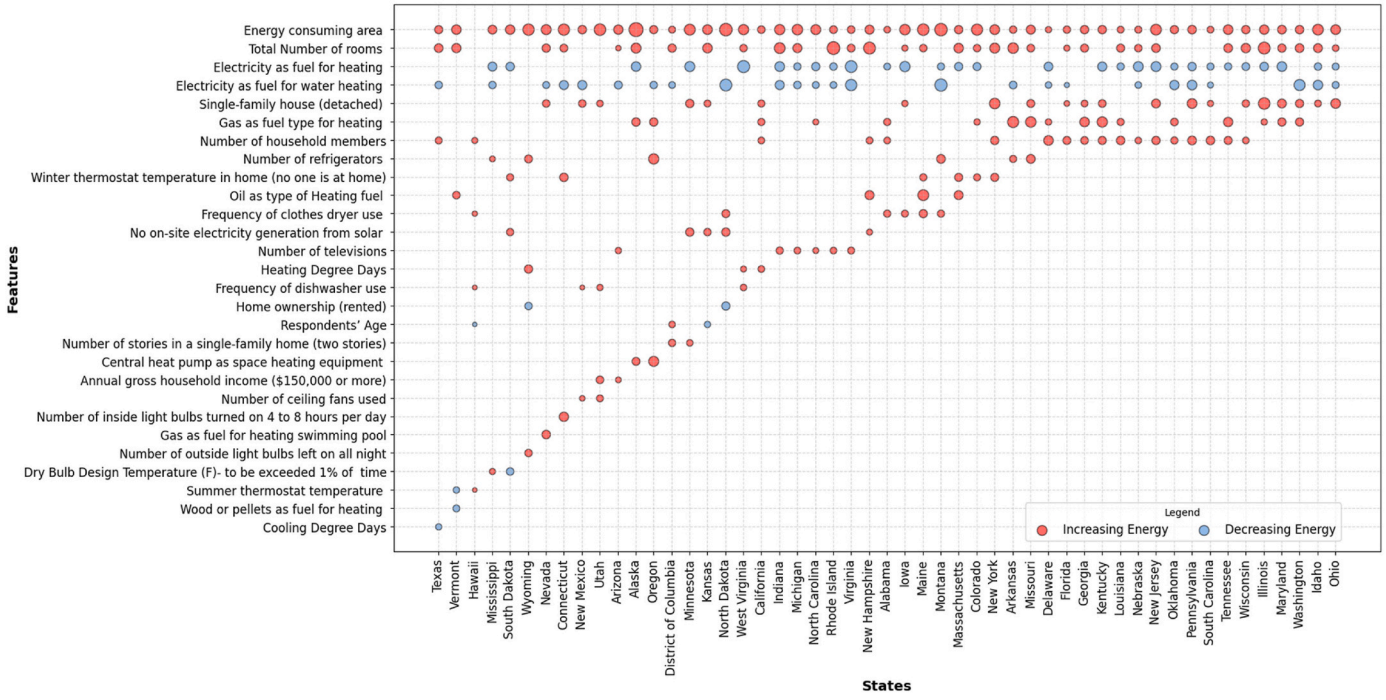


Fig. 5. Bubble graph illustrating top five features affecting residential energy consumption across various states.

integrating energy-saving education into school curriculums. Alongside top-down methods, Belaïd (2016) suggests that bottom-up approaches, such as community-driven initiatives, can play a crucial role in long-term behavioral change and energy efficiency.

Energy-Efficient Appliances: Our analysis shows that a higher number of refrigerators and televisions, along with the frequent use of clothes dryers and dishwashers, are major drivers of increased residential energy consumption in 39 % of states. While the number and

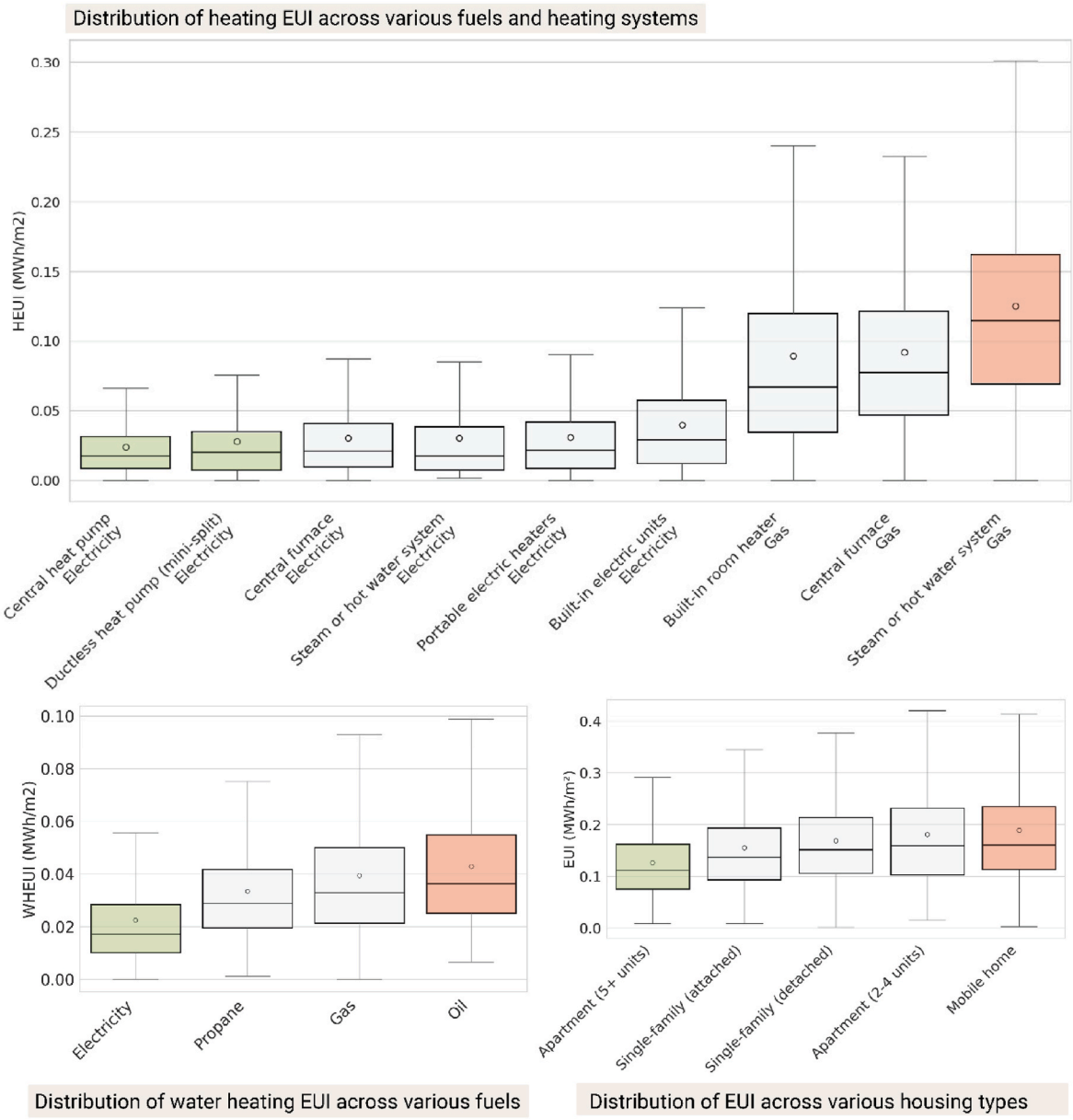


Fig. 6. Distribution of heating EUI across different fuel types and water heating systems, as well as EUI variations across different building types.

frequency of appliance use are influenced by occupant behavior and lifestyle, improving the efficiency of these appliances can significantly reduce overall energy demand. Existing research supports adopting strategies like Minimum Energy Performance Standards (MEPS) and energy-efficiency labeling for improving appliance efficiency (Sakah et al., 2019; Pan and Zhang, 2020; Nyangon and Byrne, 2021; Hojjati and Wade, 2012). These measures also phase out low-cost, inefficient appliances.

On-site Electricity Generation: We showed that on-site electricity generation can reduce residential energy consumption. According to RECS microdata, only 3.48 % of US residences use on-site solar electricity. This low adoption rate highlights the need for greater support and incentives to expand decentralized renewable technologies, which can help households cope with high energy costs while reducing transmission losses and large-scale building energy consumption (Bednar, Reames, and Keoleian, 2017b; Pan and Zhang, 2020; Andreou et al., 2020). Goldstein et al. (2022) suggest that government-led initiatives, like 'Solarize' campaigns, can speed up clean energy adoption by helping groups of buyers negotiate better contracts with solar companies.

The success of the mentioned policies in reducing overall energy consumption depends on considering socio-economic and demographic factors while mitigating unintended consequences, such as the rebound effect (also known as Jevons' Paradox). This phenomenon occurs when improvements in efficiency lead to increased use, ultimately offsetting energy savings (Burnett and Madariaga, 2018; Newton and Meyer, 2012; Froemelt et al., 2021). This is due to changes in user behavior and consumption patterns, where greater efficiency may encourage increased use. Additionally, ensuring policy effectiveness requires addressing socio-economic disparities. A study by Elmallah et al. (2024) analyzed residential heating and cooling access in Northern California and found uneven distribution across socioeconomic lines, influenced by income and housing tenure. It highlighted that equitable electrification policies should not only promote technology adoption, like heat pumps, but also ensure fair access to the essential services these technologies provide. By integrating socio-economic and behavioral considerations, policies can achieve meaningful energy savings without unintended consequences.

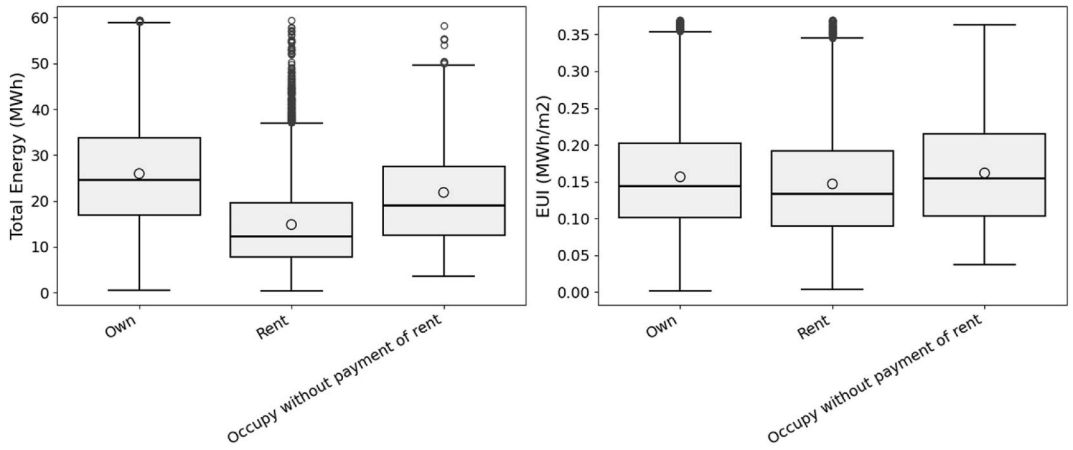


Fig. 7. Distribution of total energy consumption and EUI by home ownership status.

5.5. Limitations and future studies

Site vs. Source Energy Consumption: The RECS dataset measures site energy consumption, which includes electricity, natural gas, propane, and fuel oil used directly at the location. This does not account for energy losses during generation, transmission, and distribution. Future studies should include both site and source energy metrics for a complete analysis of energy use and efficiency.

Model Performance: Weak model performance ($R^2 < 0.50$) in 13 % of states indicates that additional features are needed to better capture energy consumption patterns and improve prediction accuracy.

Electricity for Heating: Using electricity as the primary fuel for space and water heating, often through energy-efficient systems like heat pumps, reduces on-site energy consumption. Therefore, further exploration of various electrified heating systems is recommended.

Challenges and Potentials of Electrification: Future studies should address the challenges of extensive home electrification, such as new peak loads, and assess its potential to reduce carbon emissions, particulate matter, and air pollution.

Temporal Variations: Future research could adopt a multi-year approach by integrating multiple RECS datasets to provide a more comprehensive analysis of evolving residential energy consumption patterns and better capture temporal variations.

Towards Policy Validation: The Database of State Incentives for Renewables & Efficiency (DSIRE, 2024) provides a comprehensive list of state-level policies and incentives, including appliance and equipment efficiency standards, building energy codes, green building incentives, energy efficiency resource standards, and solar/wind access policies. We recommend that future studies compare the key determinants identified in our analysis with the policies listed in DSIRE to evaluate how well existing policies align with observed energy consumption trends and to identify opportunities for policy improvement.

6. Conclusion and policy implications

We developed bottom-up data-driven models (CatBoost + SHAP sensitivity analysis) based on the 2020 RECS micro dataset to pinpoint the primary determinants of energy consumption in U.S. residential buildings at national and state levels. The key conclusions of our study are summarized as follows.

- The most important features positively (higher use) correlated with national energy consumption include HDDs, energy-consuming areas, total number of rooms, number of household members,

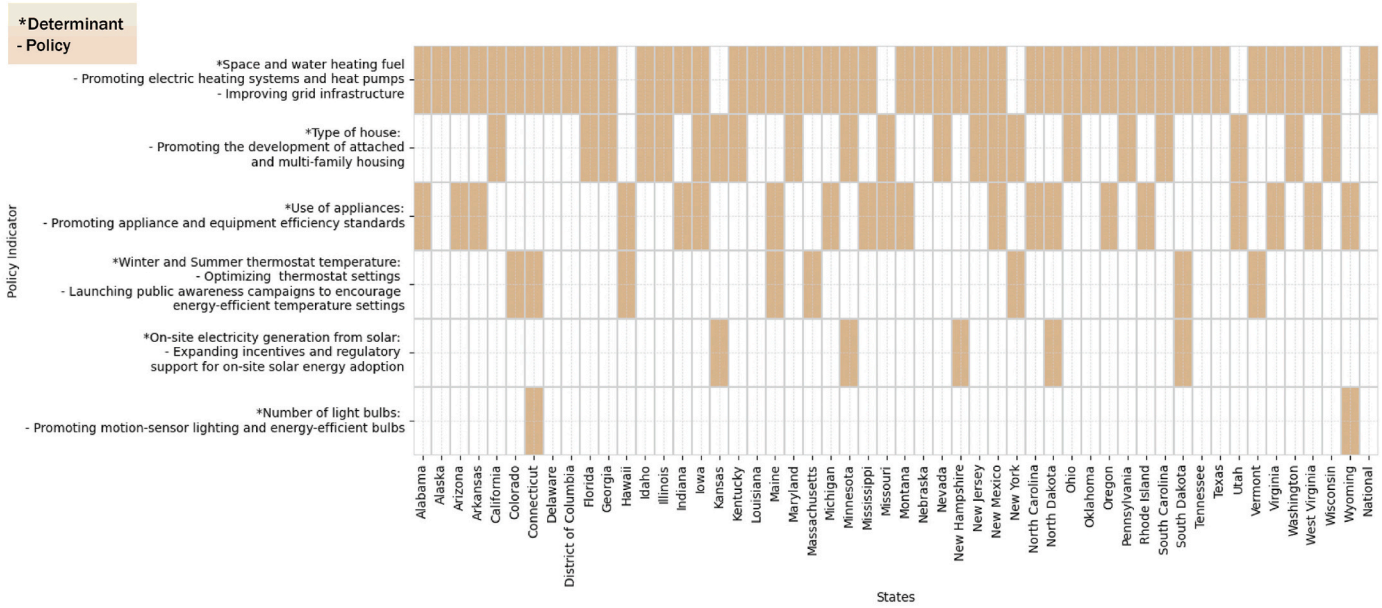


Fig. 8. The list of important features along with policies that can be inferred from each feature.

winter thermostat settings, and single-family detached homes. On the other hand, features whose increase or presence is linked to lower energy use include the use of electricity for space and water heating, Dry Bulb Design Temperature (F) −99 %, and apartments with more than five units.

- The five most important features affecting energy consumption, alongside their average absolute contributions, are electricity for main space heating (2.35 MWh), electricity for main water heater (1.67 MWh), HDD65 (1.63 MWh), energy-consuming areas (1.58 MWh) and total number of rooms (1.27 MWh) compared to the baseline of 23.173 MWh. This indicates that using electricity for space and water heating is the most significant factor in lowering national energy consumption.
- Key factors affecting residential energy consumption at the state level include energy-consuming area (98 % of states), heating fuel type (90 %), total number of rooms (65 %), housing type (41 %), appliance number and usage (39 %), household size (35 %), thermostat settings (12 %), and on-site solar generation (10 %).
- Our findings align with previous RECS-based studies, which have consistently identified building size, housing type, and HDD/CDD as key determinants of residential energy consumption. However, electrification has emerged as a dominant driver. Additionally, Dry Bulb Design Temperature, winter thermostat settings, and on-site electricity generation have become more influential. These shifts indicate that electrification trends, occupant behaviors, and technology adoption are increasingly shaping residential energy consumption patterns.
- The dominance of heating fuel as the strongest explanatory feature of national energy consumption is largely due to the significant efficiency differences between electric and gas-based heating systems. Our findings confirm that using electricity, particularly through energy-efficient heat pumps, leads to lower heating energy use intensity (EUI) compared to natural gas systems, which suffer from fuel conversion losses. Heating accounts for a substantial share of residential energy consumption and given that natural gas remains the most common heating fuel in the dataset, its higher energy intensity makes it a major driver of overall energy use.
- Our findings support the integration of electrified heating systems such as heat pumps into both state and national energy policies. Policies based on key state-level determinants include promoting attached housing, optimizing setpoint temperatures, encouraging energy-efficient appliances and lighting, and supporting on-site electricity generation. Incorporating socio-economic and behavioral factors can help policies achieve meaningful energy savings.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used Chat GPT to improve the readability and language of the manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

CRediT authorship contribution statement

Sepideh Sadat Korsavi: Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization, Validation. **Rahman Azari:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Lisa D. Iulo:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Mehrdad Mahdavi:** Writing – review & editing, Methodology, Validation.

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

We gratefully acknowledge the support provided by Penn State's Hamer Center for Community Design and its Resource and Energy Efficiency (RE2) Lab, the Stuckeman Center for Design Computing (SCDC), and the Institute of Energy and the Environment (IEE).

Data availability

Data will be made available on request.

References

Abbasabadi, Narjes, Ashayeri, M., Azari, Rahman, Stephens, B., Heidarinejad, Mohammad, 2019. An integrated data-driven framework for urban energy use modeling (UEUM). Appl. Energy 253 (November). https://doi.org/10.1016/j.apenergy.2019.113550.

Ahn, Yong Jin, Sohn, Dong Wook, 2019. The effect of neighbourhood-level urban form on residential building energy use: a GIS-based model using building energy benchmarking data in Seattle. Energy Build. 196 (August), 124–133. https://doi.org/10.1016/j.enbuild.2019.05.018.

Amiri, Shams, Shideh, Mueller, Maya, Hoque, Simi, 2023. Investigating the application of a commercial and residential energy consumption prediction model for urban planning scenarios with machine learning and Shapley additive explanation methods. Energy Build. 287 (May). https://doi.org/10.1016/j.enbuild.2023.112965.

Andreou, Andreas, Barrett, John, Taylor, Peter G., Brockway, Paul E., Wadud, Zia, 2020. Decomposing the drivers of residential space cooling energy consumption in EU-28 countries using a panel data approach. Energy and Built Environment 1 (4), 432–442. https://doi.org/10.1016/j.enbenv.2020.03.005.

Arthur, David, Vassilvitskii, Sergei, 2007. K-Means++: the advantages of careful seeding. Soda 7, 1027–1035.

Aurélien, G., 2017. Hands-on Machine Learning with Scikit-Learn & Tensorflow. O'Reilly.

Bednar, Dominic J., Reames, Tony Gerard, Keoleian, Gregory A., 2017a. The intersection of energy and justice: modeling the spatial, racial/ethnic and socioeconomic patterns of urban residential heating consumption and efficiency in Detroit, Michigan. Energy Build. 143 (May), 25–34. https://doi.org/10.1016/j.enbuild.2017.03.028.

Bednar, Dominic J., Reames, Tony Gerard, Keoleian, Gregory A., 2017b. The intersection of energy and justice: modeling the spatial, racial/ethnic and socioeconomic patterns of urban residential heating consumption and efficiency in Detroit, Michigan. Energy Build. 143 (May), 25–34. https://doi.org/10.1016/j.enbuild.2017.03.028.

Belaïd, Fateh, 2016. Understanding the spectrum of domestic energy consumption: empirical evidence from France. Energy Policy 92 (May), 220–233. https://doi.org/10.1016/j.enpol.2016.02.015.

Belaïd, Fateh, Roubaud, David, Galariotis, Emilios, 2019. Features of residential energy consumption: evidence from France using an innovative multilevel modelling approach. Energy Policy 125 (February), 277–285. https://doi.org/10.1016/j.enpol.2018.11.007.

Blakeslee, L., Rabe, M., Caplan, Z., Roberts, A., 2023. An Aging U.S. Population with Fewer Children in 2020. U.S. Census Bureau. https://www.census.gov/library/stories/2023/05/aging-united-states-population-fewer-children-in-2020.html. (Accessed 25 May 2023).

Burnett, J. Wesley, Lynne Kiesling, L., 2022. How do machines predict energy use? Comparing machine learning approaches for modeling household energy demand in the United States. Energy Res. Social Sci. 91 (September). https://doi.org/10.1016/j.erss.2022.102715.

Burnett, J. Wesley, Madariaga, Jessica, 2018. A top-down economic efficiency analysis of U.S. HouseholdEnergy consumption. Energy J. 39 (4), 1–30. https://doi.org/10.5547/01956574.39.4.jbur.

Chicco, Davide, Warrens, Matthijs J., Jurman, Giuseppe, 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and

RMSE in regression analysis evaluation. PeerJ Computer Science 7, 1–24. https://doi.org/10.7717/PEERJ-CS.623.
Cui, Mengyao, 2020. Introduction to the K-means clustering algorithm based on the Elbow method. Accounting, Auditing and Finance 1 (1), 5–8. https://doi.org/10.23977/accaf.2020.010102.
Cui, Xue, Lee, Minhyun, Koo, Choongwan, Hong, Taehoon, 2024. Energy consumption prediction and household feature analysis for different residential building types using machine learning and SHAP: toward energy-efficient buildings. Energy Build. 309 (April). https://doi.org/10.1016/j.enbuild.2024.113997.
DSIRE, 2024. Policies & incentives by type, 2024. https://programs.dsireusa.org/system/program/tables.
Elmallah, Salma, Crespo Montañés, Cristina, Duncan, Callaway, 2024. Who heats and cools? Access to residential heating and cooling in Northern California and implications for energy transitions. Energy Policy 191 (August). https://doi.org/10.1016/j.enpol.2024.114169.
Estiri, Hossein, 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 Econ. 43, 178–184. https://doi.org/10.1016/j.eneco.2014.02.013.
Estiri, Hossein, Zagheni, Emilio, 2019. Age matters: ageing and household energy demand in the United States. Energy Res. Social Sci. 55 (September), 62–70. https://doi.org/10.1016/j.erss.2019.05.006.
Froemelt, Andreas, Buffat, René, Hellweg, Stefanie, 2020. Machine learning based modeling of households: a regionalized bottom-up approach to investigate consumption-induced environmental impacts. J. Ind. Ecol. 24 (3), 639–652. https://doi.org/10.1111/jiec.12969.
Froemelt, Andreas, Geschke, Arne, Wiedmann, Thomas, 2021. Quantifying carbon flows in Switzerland: top-down meets bottom-up modelling. Environ. Res. Lett. 16 (1). https://doi.org/10.1088/1748-9326/abcdd5.
Gan, Min, Pan, Shunqi, Chen, Yongping, Chen, Cheng, Pan, Haidong, Zhu, Xian, 2021. Application of the machine learning LightGBM model to the prediction of the water levels of the lower columbia river. J. Mar. Sci. Eng. 9 (5), 496. https://doi.org/10.3390/JMSE9050496, 2021, 496 9.
Goldstein, Benjamin, Reames, Tony G., Newell, Joshua P., 2022. Racial inequity in household energy efficiency and carbon emissions in the United States: an emissions paradox. Energy Res. Social Sci. 84 (February). https://doi.org/10.1016/j.erss.2021.102365.
Hancock, John T., Khoshgoftaar, Taghi M., 2020. CatBoost for big data: an interdisciplinary review. J. Big Data 7 (1). https://doi.org/10.1186/s40537-020-00369-8.
Hojjati, Behjat, Wade, Steven H., 2012. U.S. Household energy consumption and intensity trends: a decomposition approach. Energy Policy 48 (September), 304–314. https://doi.org/10.1016/j.enpol.2012.05.024.
Hoof, J. van, Schellen, L., Soebarto, V., Wong, J.K.W., Kazak, J.K., 2017. Ten questions concerning thermal comfort and ageing. Build. Environ. 120 (August), 123–133. https://doi.org/10.1016/j.buildenv.2017.05.008.
Hu, Shan, Yan, Da, Qian, Mingyang, 2019. Using bottom-up model to analyze cooling energy consumption in China's urban residential building. Energy Build. 202 (November). https://doi.org/10.1016/j.enbuild.2019.109352.
Huang, Wen Hsiu, 2015. The determinants of household electricity consumption in Taiwan: evidence from quantile regression. Energy 87 (July), 120–133. https://doi.org/10.1016/j.energy.2015.04.101.
Jain, Anil K., 2010. Data clustering: 50 Years beyond K-means. Pattern Recognit. Lett. 31 (8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011.
James, Gareth, Witten, Daniela, Hastie, Trevor, Tibshirani, Robert, 2013. An Introduction to Statistical Learning, vol. 112. Springer, New York. https://doi.org/10.1007/978-1-4614-7138-7.
Jiang, Feifeng, Jun, Ma, Li, Zheng, Ding, Yuexiong, 2022. Prediction of energy use intensity of urban buildings using the semi-supervised deep learning model. Energy 249 (June). https://doi.org/10.1016/j.energy.2022.123631.
Kontokosta, Constantine E., Tull, Christopher, 2017a. A data-driven predictive model of city-scale energy use in buildings. Appl. Energy 197, 303–317. https://doi.org/10.1016/j.apenergy.2017.04.005.
Kontokosta, Constantine E., Tull, Christopher, 2017b. A data-driven predictive model of city-scale energy use in buildings. Appl. Energy 197, 303–317. https://doi.org/10.1016/j.apenergy.2017.04.005.
Lima, Azevedo, Inês, Granger Morgan, M., Palmer, Karen, Lave, Lester B., 2013. Reducing U.S. Residential energy use and CO2 emissions: how much, how soon, and at what cost? Environ. Sci. Technol. 47 (6), 2502–2511. https://doi.org/10.1021/es303688k.
Lloyd, Stuart P., 1982. Least squares quantization in PCM. IEEE Trans. Inf. Theor. 28 (2), 129–137.
Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems, vol. 30.
Mai, Trieu, Jadun, Paige, Logan, Jeffrey, Mcmillan, Colin, Muratori, Matteo, Steinberg, Daniel, Vimmerstedt, Laura, Ryan, Jones, Benjamin, Haley, Nelson, Brent, 2018. Electrification futures study: scenarios of electric technology adoption and power consumption for the United States. Golden, CO (United States). https://www.nrel.gov/docs/fy18osti/71500.pdf.
Mohammadiziazi, Rezvan, Copeland, Samuel, Bilec, Melissa M., 2021. Urban building energy model: database development, validation, and application for commercial building stock. Energy Build. 248 (October). https://doi.org/10.1016/j.enbuild.2021.111175.
Moran, M.J., Shapiro, H.N., Boettner, D.D., Bailey, M.B., 2010. Fundamentals of Engineering Thermodynamics. John Wiley & Sons.
Mostafavi, Nariman, Farzinmoghadam, Mohamad, Hoque, Simi, 2017a. Urban residential energy consumption modeling in the integrated urban metabolism analysis tool (IUMAT). Build. Environ. 114 (March), 429–444. https://doi.org/10.1016/j.buildenv.2016.12.035.
Mostafavi, Nariman, Farzinmoghadam, Mohamad, Hoque, Simi, 2017b. Urban residential energy consumption modeling in the integrated urban metabolism analysis tool (IUMAT). Build. Environ. 114 (March), 429–444. https://doi.org/10.1016/j.buildenv.2016.12.035.
Movahedi, Ali, Sybil, Derrible, 2021. Interrelationships between electricity, gas, and water consumption in large-scale buildings. J. Ind. Ecol. 25 (4), 932–947. https://doi.org/10.1111/jiec.13097.
Namazkhan, Maliheh, Albers, Casper, Steg, Linda, 2020. A decision tree method for explaining household gas consumption: the role of building characteristics, socio-demographic variables, psychological factors and household behaviour. Renew. Sustain. Energy Rev. 119 (March). https://doi.org/10.1016/j.rser.2019.109542.
Navamuel, Elena Lasarte, Morollón, Fernando Rubiera, Cuartas, Blanca Moreno, 2018. Energy consumption and urban sprawl: evidence for the Spanish case. J. Clean. Prod. 172 (January), 3479–3486. https://doi.org/10.1016/j.jclepro.2017.08.110.
Newton, Peter, Meyer, Denny, 2012. The determinants of urban resource consumption. Environ. Behav. 44 (1), 107–135. https://doi.org/10.1177/0013916510390494.
Nsangou, Jean Calvin, Kenfack, Joseph, Nzotcha, Urbain, Paul, Salomon Ngohe Ekam, Voufo, Joseph, Tamo, Thomas T., 2022. Explaining household electricity consumption using quantile regression, decision tree and artificial neural network. Energy 250 (July). https://doi.org/10.1016/j.energy.2022.123856.
Nyangon, Joseph, Byrne, John, 2021. Spatial energy efficiency patterns in New York and implications for energy demand and the rebound effect. Energy Sources B Energy Econ. Plann. 16 (2), 135–161. https://doi.org/10.1080/15567249.2020.1868619.
Pan, Yue, Zhang, Limao, 2020. Data-driven estimation of building energy consumption with multi-source heterogeneous data. Appl. Energy 268 (June). https://doi.org/10.1016/j.apenergy.2020.114965.
Pesantez, Jorge E., Wackerman, Grace E., Stillwell, Ashlynn S., 2023. Analysis of single- and multi-family residential electricity consumption in a large urban environment: evidence from Chicago, IL. Sustain. Cities Soc. 88 (January), 104250. https://doi.org/10.1016/j.scs.2022.104250.
Porse, Erik, Derenski, Joshua, Gustafson, Hannah, Elizabeth, Zoe, Pincetl, Stephanie, 2016. Structural, geographic, and social factors in urban building energy use: analysis of aggregated account-level consumption data in a megacity. Energy Policy 96 (September), 179–192. https://doi.org/10.1016/j.enpol.2016.06.002.
Prokhorenkova, Liudmila, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. . "CatBoost: Unbiased Boosting with Categorical Features." https://github.com/catboost/catboost.
Prokhorenkova, Liudmila, Gusev, Gleb, Vorobev, Aleksandr, Dorogush, Anna Veronika, Gulin, Andrey, 2017. CatBoost: unbiased boosting with categorical features. Adv. Neural Inf. Process. Syst. 6638–6648, 2018-December (June.
Sakah, Marriette, Rue du Can, Stephane de la, Diawuo, Felix Amankwah, Sedzro, Morkporkpor Delight, Kuhn, Christoph, 2019. A study of appliance ownership and electricity consumption determinants in urban Ghanaian households. Sustain. Cities Soc. 44 (January), 559–581. https://doi.org/10.1016/j.scs.2018.10.019.
Sanquist, Thomas F., Orr, Heather, Bin, Shui, Bittner, Alvah C., 2012a. Lifestyle factors in U.S. Residential electricity consumption. Energy Policy 42 (March), 354–364. https://doi.org/10.1016/j.enpol.2011.11.092.
Sanquist, Thomas F., Orr, Heather, Bin, Shui, Bittner, Alvah C., 2012b. Lifestyle factors in U.S. Residential electricity consumption. Energy Policy 42 (March), 354–364. https://doi.org/10.1016/j.enpol.2011.11.092.
Scikit-learn Developers, 2024. "StandardScaler." scikit-learn 1.5.0 documentation, 2024. https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html.
Shapley, Lloyd S., 1953. "A value for N-person games. Contribution to the Theory of Games 2.".
Shen, Pengyuan, Yang, Biao, 2020a. Projecting Texas energy use for residential sector under future climate and urbanization scenarios: a bottom-up method based on twenty-year regional energy use data. Energy 193 (February). https://doi.org/10.1016/j.energy.2019.116694.

Shen, Pengyuan, Yang, Biao, 2020b. Projecting Texas energy use for residential sector under future climate and urbanization scenarios: a bottom-up method based on twenty-year regional energy use data. Energy 193 (February). https://doi.org/10.1016/j.energy.2019.116694.

Staffell, Iain, Pfenninger, Stefan, Johnson, Nathan, 2023. A global model of hourly space heating and cooling demand at multiple spatial scales. Nat. Energy. https://doi.org/10.1038/s41560-023-01341-5.

Tran, Le Na, Cai, Gangwei, Gao, Weijun, 2023. Determinants and approaches of household energy consumption: a review. Energy Reports. Elsevier Ltd. https://doi.org/10.1016/j.egyr.2023.08.026.

U.S. Census Bureau, 2021. American housing Survey (AHS), 2021. https://www.census.gov/programs-surveys/ahs/data/interactive/ahstablecreator.html?s_areas=00000&s_year=2021&s_tablename=TABLE1&s_bygroup1=1&s_bygroup2=1&s_filtergroup1=1&s_filtergroup2=1.

U.S. Department of Energy (DOE). . Heat Pump Systems. Energy Saver. Accessed June 18, 2024. https://www.energy.gov/energysaver/heat-pump-systems.

U.S. DOE. n.d. "Tankless or Demand-Type Water Heaters." Energy Saver. Accessed June 18, 2024. https://www.energy.gov/energysaver/tankless-or-demand-type-water-heaters.

U.S. EIA, 2023. Annual energy review. July 2023. https://www.eia.gov/totalenergy/data/annual/.

U.S. EIA, 2023b. Methodology: 2020 residential energy consumption Survey. Household characteristics technical documentation summary, 2023. https://www.eia.gov/consumption/residential/data/2020/index.php?view=methodology.

U.S. EIA, 2023c. Microdata: 2020 residential energy consumption Survey, 2023. https://www.eia.gov/consumption/residential/data/2020/index.php?view=microdata.

U.S. Energy Information Administration, 2023. Comparing the 2020 RECS with previous RECS and other studies. www.eia.gov.

Wang, Lan, Lee, Eric W.M., Hussian, Syed Asad, Yuen, Anthony Chun Yin, Feng, Wei, 2021. Quantitative impact analysis of driving factors on annual residential building energy end-use combining machine learning and stochastic methods. Appl. Energy 299 (October). https://doi.org/10.1016/j.apenergy.2021.117303.

Willmott, Cort J., Matsuura, Kenji, 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Research 30 (1), 79–82. https://doi.org/10.2307/24869236.

Yun, Geun Young, Steemers, Koen, 2011. Behavioural, physical and socio-economic factors in household cooling energy consumption. Appl. Energy 88 (6), 2191–2200. https://doi.org/10.1016/j.apenergy.2011.01.010.

Zhang, Yan, Teoh, Bak Koon, Wu, Maozhi, Chen, Jiayu, Zhang, Limao, 2023. Data-driven estimation of building energy consumption and GHG emissions using explainable artificial intelligence. Energy 262 (January). https://doi.org/10.1016/j.energy.2022.125468.