KIEAE Journal, Vol. 26, No. 5, Oct. 2026, pp.45-51.

Check for updates

KIEAE Journal
Korea Institute of Ecological Architecture and Environment
141

An Empirical Study on Energy Consumption Factors and Benchmark Models Based on a Residential Energy Consumption Survey Dataset

Jonghoon Ahn*

* Professor, School of Architecture and Design Convergence, Hankyong National Univ., South Korea (architectism@hknu.ac.kr)

ABSTRACT

Purpose: This study empirically investigates the principal physical, spatial, and climatic drivers that determine the total energy consumption of residential buildings. The ultimate objective is to provide a highly interpretable, heuristic predictive model that architects, urban planners, and policymakers can utilize for residential energy benchmarking and sustainable architectural design. **Method:** To construct a robust framework, the research leverages a large-scale, nationally representative microdata sample from the Residential Energy Consumption Survey spanning the 2009, 2015, and 2020 survey waves. The study employs Pearson correlation analysis to measure linear associations and filter variables, followed by the application of statistical models to quantify the isolated impact of each variable on total energy consumption. **Results:** The findings indicate that a building's physical scale acts as the primary driver of residential energy consumption. The analysis also reveals an asymmetric behavior in climate variables. The Heating Degree Days exhibit a robust positive correlation, whereas Cooling Degree Days demonstrate a negative correlation, reflecting the fundamental efficiency gap between fossil-fuel space heating and electric space cooling. Additionally, the regression model highlights nuanced spatial dynamics, such as a negative energy penalty associated with the number of bathrooms when the total square footage is held constant. In conclusion, the study successfully establishes a mathematically robust multiple regression predictive model that serves as a highly accessible forecasting tool. This model can empower researchers and practitioners to rapidly estimate the baseline annual energy consumption of a residential unit based purely on basic architectural parameters and local climate conditions.

ⓒ 2026. KIEAE all rights reserved.

KEYWORD

Residential Building
Energy Consumption
RECS
Energy Benchmark
Regression Model

ACCEPTANCE INFO

Received Sep. 3, 2026
Final revision received Sep. 14, 2026
Accepted Sep. 18, 2026

1. Introduction

1.1. Background and Necessity of the Study

As global initiatives toward carbon neutrality intensify, mitigating the energy footprint of the residential building sector has emerged as a paramount challenge. Unlike commercial buildings, which are predominantly driven by internal heat gains, residential buildings exhibit unique thermodynamic behaviors highly sensitive to exterior climate variations, envelope integrity, and household operational patterns. However, much of the existing research relies on localized datasets or isolated building energy simulations which often fail to capture the macro-scale statistical realities of diverse housing stocks.

1.2. Purpose of the Study

This study empirically investigates the principal physical, spatial, and climatic drivers that determine the total energy consumption (TOTALBTU) of residential buildings. By leveraging the Residential Energy Consumption Survey (RECS), a large-scale, nationally representative microdata sample provided by the U.S. Energy Information Administration (USEIA), this research employs Pearson correlation analysis and Multiple Linear Regression (MLR). The ultimate objective is to provide a highly interpretable, heuristic predictive model that architects, urban planners, and policymakers can utilize for residential energy benchmarking and sustainable architectural design.

2. Literature Review

2.1. Energy Use Intensity and Benchmarking

The transition toward a sustainable built environment necessitates rigorous methodologies to quantify, evaluate, and compare building energy performance. At the core of this evaluation framework is the concept of Energy Use Intensity (EUI). EUI is a universally accepted metric defined as the annual energy consumption of a building normalized by its gross floor area, typically expressed in kWh/m^2·yr or kBTU/sqft·yr [1]. By normalizing absolute energy consumption, EUI facilitates equitable comparisons across buildings of disparate sizes and geometric configurations, making it a foundational indicator for

pISSN 2288-968X, eISSN 2288-9698
http://dx.doi.org/10.12813/kieae.2026.26.5.045

both macroscopic energy policies and microscopic building diagnostics [2]. Building energy benchmarking leverages EUI to compare a specific building's performance against a peer group of similar buildings or established regulatory standards. Chung categorized building energy benchmarking into two primary approaches: statistical methods based on empirical data and simulation-based methods relying on physical models [3]. Empirical benchmarking programs, such as the U.S. Environmental Protection Agency's (EPA) ENERGY STAR portfolio manager, rely heavily on historical EUI data to assign percentile rankings to buildings, thereby incentivizing owners to pursue energy-efficient retrofits [4]. The literature consistently demonstrates that continuous benchmarking and the public disclosure of EUI metrics directly correlate with long-term energy savings. For instance, Ke et al. highlighted that mandatory energy benchmarking policies can lead to a 3% to 8% reduction in EUI over a multi-year period due to increased operational awareness and the identification of baseline inefficiencies [5]. However, traditional EUI metrics often face criticism for oversimplifying the complex thermodynamic realities of buildings. EUI treats the entire floor area as a homogenous zone, frequently ignoring the disproportionate energy intensities of specific spaces, such as data centers or intensive medical equipment zones within commercial facilities, or diverse spatial partitioning in residential sectors [6]. Consequently, recent literature emphasizes the need to decompose standard EUI by integrating multi-dimensional variables, such as operational hours, occupant density, and climate severity, into the benchmarking process to establish more accurate, context-specific baselines.

2.2. Residential Energy Consumption Survey

While physics-based building energy modeling (BEM) tools, such as EnergyPlus and TRNSYS, are widely utilized for predicting energy performance during the architectural design phase, they inherently suffer from the "energy performance gap" a well-documented discrepancy between simulated predictions and actual operational energy consumption [7]. This gap is primarily driven by the unpredictability of occupant behavior, stochastic weather variations, and degradation of mechanical systems over time [8]. To bridge this gap, researchers increasingly rely on large-scale, empirical microdata collected directly from the operational building stock. In the United States, the Residential Energy Consumption Survey (RECS), administered by the U.S. Energy Information Administration (EIA), serves as the most comprehensive and statistically representative dataset for residential energy analysis [9]. Conducted periodically since 1978, RECS captures detailed multidimensional data, including physical housing characteristics, household demographics, installed HVAC equipment, local climatic indicators (Heating and Cooling Degree Days, HDD/CDD), and actual billing data for total energy consumption. Kaza utilized RECS data to demonstrate that housing typology and size are the most dominant determinants of residential energy consumption, often superseding the impact of household income or behavioral conservation efforts [10]. The unparalleled advantage of the RECS dataset lies in its macro-scale empirical validity. Ewing and Rong leveraged RECS to investigate the relationship between urban sprawl and residential energy use, proving that households in sprawling, detached single-family contexts consume significantly more energy than those in dense, compact urban environments due to greater exposed surface areas and transmission losses [11]. By providing a statistically weighted cross-section of the national housing stock, RECS allows researchers to bypass the theoretical assumptions of BEM tools and construct data-driven models that accurately reflect the macroscopic reality of residential energy consumption patterns.

2.3. Statistical Method

The proliferation of large-scale building energy datasets has necessitated the application of robust statistical methodologies to extract actionable insights. Zhao and Magoules classified building energy prediction models into engineering methods, statistical methods, and artificial intelligence methods [12]. While AI and machine learning algorithms (e.g., Artificial Neural Networks, Support Vector Machines) have gained immense popularity due to their high predictive accuracy in handling non-linear data, they are frequently criticized for their "black-box" nature. The inability to transparently trace how specific variables interact limits their utility for architects and policymakers who require explicitly interpretable guidelines [13]. Consequently, classical statistical approaches, specifically Pearson Correlation Analysis and Multiple Linear Regression (MLR), remain highly prevalent and essential in building energy research. The Pearson correlation coefficient (γ) is utilized as a rigorous preliminary tool for feature selection and dimensionality reduction. By quantifying the degree of linear association between a continuous independent variable (e.g., floor area, HDD) and the dependent variable (EUI or total energy consumption), researchers can eliminate statistically insignificant features and identify critical drivers of energy use [14]. For example, a high positive correlation between the number of occupants and total energy consumption instantly validates the necessity of integrating occupant-centric controls into the HVAC system.Building upon correlation analysis, Multiple Linear Regression (MLR) models are employed to predict energy

consumption and quantify the isolated impact of individual variables under the assumption of ceteris paribus (all other things being equal) [15]. Fumo and Biswas successfully applied MLR to residential energy data, demonstrating that simple linear combinations of floor area, operating hours, and climate data could predict building energy consumption with an accuracy closely rivaling complex dynamic simulations [16]. The primary mathematical advantage of MLR lies in its partial regression coefficients (β). These coefficients provide explicit, quantifiable metrics such as estimating the exact annual BTU penalty incurred by adding an additional room or a specific mechanical equipment unit. This level of interpretability provides architectural planners with a highly accessible heuristic tool to forecast energy profiles and optimize spatial configurations during the early stages of design, without the computational overhead of advanced simulations [17].

2.4. Summary

A review of the existing literature reveals that while the concepts of EUI and energy benchmarking are well-established, many benchmarking frameworks still rely on deterministic physical simulations that fail to capture the stochastic nature of actual building operations. Furthermore, while numerous studies have utilized the EIA's RECS dataset, a significant portion of this research has either focused purely on socioeconomic analyses or employed opaque machine learning models that offer limited practical interpretability for architectural design. There remains a critical need for studies that bridge the gap between large-scale empirical datasets and highly interpretable statistical modeling. By rigorously applying Pearson correlation and Multiple Linear Regression to the RECS microdata, this study seeks to extract transparent, mathematically quantifiable coefficients that define the exact energetic cost of specific residential architectural features and climatic variables, thereby providing a robust and intuitive heuristic model for sustainable building design.

3. Research Methodology

3.1. Energy Consumption Benchmark

In the domain of building energy analysis, benchmarking models have routinely employed regression techniques to systematically assess the statistical significance of performance-driving parameters. This quantitative approach allows for the rigorous evaluation of diverse variables, ranging from environmental climatic conditions and installed thermal system capacities to occupant density profiles and facility operational schedules. Through this methodology, the relative impact of both structural and behavioral factors can be mathematically isolated and measured. The underlying foundational equation that constructs this multi-variable linear regression framework is typically represented as follows [18,19]:

$$EUI = a + b_1 x_1^* + b_2 x_2^* + \cdots + b_k x_k^* + \varepsilon \quad \text{(Eq. 1)}$$

where "a" is an intercept; "$b_1, \cdots, b_k$" are regression coefficients; "$x_1^*, \cdots, x_k^*$" are significant standardized factors; "ε" is random error.

Prior to statistical modeling, all raw data points were transformed using Z-score standardization to neutralize unit variance and align heterogeneous variables on a non-dimensional scale. This preprocessing step mitigates scale bias by converting absolute measurements into relative standard deviations from their respective sample means, expressed as follows:

$$Z = (X - \mu)/\sigma \quad \text{(Eq. 2)}$$

where "X" is the raw variable value, "μ" is the population mean, and "σ" represents the standard deviation.

These analytical models assess building energy performance by systematically isolating the primary physical and operational drivers, subsequently estimating their corresponding statistical weights. However, to ensure that these predictive frameworks are both accurate and generalizable, they must be subjected to rigorous statistical validation [20,21]. While fundamental descriptive metrics, such as standard deviation, are limited to quantifying the general spread of the raw data, advanced inferential techniques, namely Ordinary Least Squares (OLS) regression and Analysis of Variance (ANOVA), are employed to explicitly parse out and verify specific parameter relationships. To quantitatively benchmark this predictive reliability, two primary diagnostic metrics are utilized: the Root Mean Square Error (RMSE) assesses the absolute residual variance between predicted and observed energy values, whereas the coefficient of determination (R^2 or adjusted R^2) delineates the exact proportion of total energy variance that is successfully explained by the model's selected predictors [22].

$$\begin{aligned} &R\text{-}squared\,(R^2) \\ &\equiv 1 - (The\ sum\ of\ squared\ of\ residuals)/ \\ &\quad (The\ total\ sum\ of\ squares) \end{aligned} \quad \text{(Eq. 3)}$$

$$RMSE = squared\ root\ \left(\left(\sum_{t=1}^{n} (\hat{y_t} - y_t)^2\right)/n\right) \quad \text{(Eq. 4)}$$

ⓒ 2026. Korea Institute of Ecological Architecture and Environment all rights reserved.

where "n" is the sample size, "$\hat{y}_i$" represents the predicted values, "y_i" is the observed dependent variable, and bar is its mean.

In order to facilitate equitable comparisons across datasets with disparate numerical scales and physical magnitudes, the Coefficient of Variation of the Root Mean Square Error (CvRMSE) is employed. This diagnostic metric normalizes the absolute prediction error by dividing it by the empirical mean of the observed data, thereby providing a scale-independent, dimensionless percentage that accurately reflects model performance [22].

$$CvRMSE = RMSE / \bar{y} \quad \text{(Eq. 5)}$$

Minimizing the sum of squared residuals simultaneously optimizes both metrics—maximizing R^2 while lowering the CvRMSE. Thus, a high R^2 coupled with a low CvRMSE signifies high precision and superior explanatory capability.

The mathematical objective of the regression algorithm is to minimize the sum of squared residuals, a process that inherently optimizes both diagnostic metrics simultaneously. By reducing the unexplained variance, this minimization mathematically maximizes the coefficient of determination (R^2) while concurrently driving down the Coefficient of Variation of the Root Mean Square Error (CVRMSE). Consequently, achieving a high R^2 in tandem with a low CVRMSE provides quantitative confirmation that the model possesses high predictive precision and superior explanatory capability.Building upon this statistical foundation, this study details the rigorous construction of a multiple linear regression framework from an integrated engineering and mathematical perspective. By harmonizing longitudinal data from the Residential Energy Consumption Survey (RECS) across three distinct waves (2009, 2015, and 2020), this framework aims to accurately forecast future residential energy consumption patterns. Crucially, before advancing to complex, non-linear control algorithms for HVAC systems or deploying "black-box" artificial neural network (ANN) predictive models, establishing this multivariate linear model is an essential prerequisite. It serves as a transparent, mathematically robust baseline, fundamentally characterizing the macroscopic linear dynamics of the overall building system and explicitly determining the statistical weights of its underlying features.

3.2. Data Preprocessing

In order to secure a robust, large-scale sample, this study pooled historical and contemporary datasets from the RECS microdata. The total residential energy consumption (TOTALBTU) was established as the dependent variable. To ensure statistical integrity, the following preprocessing protocols were applied:

1) Prevention of Tautological Bias: Sub-metered or fuel-specific energy metrics (e.g., electricity kWh, natural gas usage, specific space heating BTUs) that mathematically sum to TOTALBTU were strictly excluded from the independent variable pool to prevent severe multicollinearity and tautological errors.
2) Missing Value Refinement: Samples with missing values in critical continuous variables (e.g., total square footage, degree days, room counts) were processed using listwise deletion, yielding a highly reliable pooled dataset of over 36,000 households for the final regression modeling.
3) Variable Selection: Variables that exhibited inconsistent binning or categorical coding across the three survey years were excluded, prioritizing robust continuous and ordinal metrics (e.g., Area, Rooms, Climatic variables) for the linear models.

3.3. Data Analysis

Pearson Correlation Analysis was adopted as the primary feature selection tool to measure the strength and direction of linear associations between architectural/climatic variables and TOTALBTU.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad \text{(Eq. 6)}$$

This methodology offers paramount interpretability. It immediately quantifies the directional impact (positive or negative) of a specific spatial or climatic trait on overall energy demand, effectively filtering out noise before the construction of the predictive regression model. Following the correlation analysis, verified independent variables were simultaneously fed into an Ordinary Least Squares (OLS) Multiple Linear Regression model to predict TOTALBTU (Y). This mathematical formulation quantifies the pure, isolated impact of a single physical or climatic variable.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n + \epsilon \quad \text{(Eq. 7)}$$

4. Results and Discussion

Using the preprocessed pooled RECS dataset, the Pearson correlation coefficients between major architectural/climatic characteristics and total energy consumption (TOTALBTU) were calculated. The findings are summarized in Table 1.

Consistent with the laws of thermodynamics, the building's physical scale-represented by Total Area (TOTSQFT_EN, γ =0.529) and Number of Stories (STORIES, γ =0.568)-acts as the primary driver of residential energy consumption. Furthermore, variables indicating spatial partitioning, such as Total Rooms (TOTROOMS, γ =0.365), also showed significant positive correlations. This suggests that highly partitioned homes require more energy to overcome thermal zoning inefficiencies and air distribution friction.

A highly compelling statistical discovery is the asymmetric behavior of climate variables. Heating Degree Days (HDD65) exhibited a robust positive correlation (γ =0.338), whereas Cooling Degree Days (CDD65) demonstrated a negative correlation (γ =-0.283). This perfectly encapsulates the fundamental energy dynamics of the residential sector. Space heating inherently relies on high-BTU fossil fuels (natural gas, heating oil, propane), whereas space cooling utilizes highly efficient, electricity-driven vapor-compression refrigeration cycles (air conditioners/heat pumps) with high Coefficients of Performance (COP). Consequently, homes situated in severely cold climates consume massively larger quantities of raw source BTUs than homes in severely hot climates. In order to build a practical forecasting tool, five continuous core variables were integrated into an MLR model (Model Adj. R^2=0.369, p<0.001). The derived optimal estimated regression equation is as follows:

$$\begin{aligned} TOTALBTU = {} & 2{,}319.6 + (22.73 * TOTSQFT) \\ & + (1{,}004.37 * TOTROOMS) \\ & - (6{,}775.71 * NCOMBATH) + (5.52 * HDD65) \\ & + (1.43 * CDD65) \end{aligned} \quad \text{(Eq. 8)}$$

Physical Meaning and Interpretation of Regression Coefficients (β) can be summarized:

1) Total Area (TOTSQFT_EN): Assuming all else is constant, every 1 sqft increase in residential floor area dictates a fixed annual increase of 22.73 BTUs.
2) Total Rooms (TOTROOMS): The addition of a partitioned room introduces an energy penalty of approximately 1,004.37 BTUs annually, underscoring the thermodynamic cost of complex interior layouts.
3) The Bathroom Paradox (NCOMBATH): Statistically, the coefficient for bathrooms emerged as highly negative (-6,775.71). This implies that under strictly controlled conditions-where the total square footage and total number of rooms are exactly identical-a house with more bathrooms actually consumes less energy. Thermodynamically, bathrooms are often transiently occupied, heavily occupied by plumbing fixtures rather than heat-generating electronics, and frequently maintained as unconditioned or semi-conditioned thermal zones. As a result, the replacement of living spaces with bathroom spaces structurally reduces the home's baseline thermal and plug loads.
4) Climatic Impact (HDD65 vs CDD65): The regression confirms the correlation anomaly: every single Heating Degree Day adds 5.52 BTUs to the annual load, while a Cooling Degree Day decreases the comparative total source energy by 1.43 BTUs, reflecting the profound systemic efficiency gap between heating and cooling technologies.

Table 2. describes the performance and accuracy of the developed regression model by comparing its predicted outputs against actual data. In order to conduct this comparative analysis, a subset of 5,000 samples was extracted from the comprehensive dataset. The evaluation focuses on two primary building energy metrics: Average Total BTU and Energy Use Intensity (EUI),

Table 1. Rank of the statistical significance for the independent variables

Rank	Category	Variable name (code)	Description	Coefficient (r)	Statistical significance
1	Form/Scale	Number of Stories (STORIES)	Total above-ground stories	0.568	Strong positive (+)
2	Scale/Structure	Total Area (TOTSQFT_EN)	Total conditioned square footage	0.529	Strong positive (+)
3	Envelope	Number of Windows (WINDOWS)	Binned/total window count	0.508	Strong positive (+)
4	Spatial	Total Rooms (TOTROOMS)	Total number of partitioned rooms	0.365	Moderate positive (+)
5	Spatial	Bedrooms (BEDROOMS)	Number of bedrooms	0.351	Moderate positive (+)
6	Climate	Heating Degree Days (HDD65)	Climate indicator for heating demand	0.338	Moderate positive (+)
7	Spatial	Bathrooms (NCOMBATH)	Number of full/half bathrooms	0.209	Weak positive (+)
8	Env/Equip	Equipment Types (EQUIPM)	Types of mechanical equipment	-0.054	Negligible
9	Envelope	Insulation (ADQINSUL)	Adequacy of building insulation	-0.058	Negligible
10	Climate	Cooling Degree Days (CDD65)	Climate indicator for cooling demand	-0.283	Weak/mod negative (-)

ⓒ 2026. Korea Institute of Ecological Architecture and Environment all rights reserved.

Table 2. Prediction effectiveness

Category		Type of control model		
		Actual data	Predicted result	Difference
Average total BTU		73,668.26	71,175.33	3.38 %
EUI (kWh/m^2·yr)		140.62	143.35	1.94 %
CvRMSE	Average total BTU	-	27,466.68	-
	EUI (kWh/m^2·yr)	-	62.99	-

alongside their respective error variances.

The regression model demonstrates a strong capability to estimate central tendencies, showing minimal deviation from the actual aggregate data. For the Average Total BTU, the actual data records an average of 73,668.26, whereas the regression model predicts 71,175.33. This results in a marginal difference of only 3.38%, indicating a slight underprediction by the model. Similarly, the actual EUI stands at 140.62kWh/m^2·yr compared to the predicted EUI of 143.35kWh/m^2·yr. This yields an even smaller difference of 1.94%, with the model slightly overpredicting the intensity. Despite the high accuracy in predicting overall averages, the table reports notable values in the CvRMSE category – 27,466.68 for Average Total BTU and 62.99 for EUI. Given the absolute magnitude of these figures relative to their respective averages, they indicate a significant degree of variance at the individual sample level across the 5,000 items, even though the aggregated macro-level means align very closely.

The discrepancy between the highly accurate mean predictions and the notable individual-level variance can be attributed to several operational realities inherent in building energy performance. Regression models inherently smooth out high-frequency dynamic changes and extreme outliers to establish a generalized predictive trend. In reality, actual building energy data is subject to highly unpredictable stochastic variables, most notably occupant behavior. Variations in occupancy schedules, manual overrides of localized thermal controls, and unpredictable equipment plug loads introduce sudden fluctuations that a standard regression model cannot perfectly capture on a case-by-case basis. And the actual operational data naturally encompasses real-world systemic inefficiencies. Factors such as localized HVAC component degradation, sensor drift, or the mechanical lag inherent in conventional thermostatic control responses create "noise" in the data. The regression model, which likely assumes a more idealized or mathematically consistent operational state, accurately captures the baseline energy demand but struggles to replicate the granular friction of physical system operations. Additionally, localized environmental factors and microclimates may cause short-term spikes or drops in thermal demand. While the regression model is robust enough to absorb these variations over a large aggregate, keeping the overall mean difference well under 4%, the individual data points still reflect these localized thermal shocks. Ultimately, the model proves highly reliable for macro-level energy estimation and baseline planning, though individual-level predictions remain sensitive to real-world operational and environmental noise.

5. Conclusion

This study successfully derived the core architectural and climatic variables dictating residential energy consumption based on a massive, nationally representative dataset and proposed a robust multiple regression predictive model. The primary conclusions and implications are as follows:

First, unlike commercial buildings dominated by internal loads, residential energy demand is profoundly shaped by climatic geography and spatial scale. The statistical evidence proves that mitigating winter heat loss is drastically more critical for raw energy consumption reduction than addressing summer heat gain, largely due to the systemic inefficiencies of fossil-fuel-based heating versus electric cooling.

Second, the regression modeling reveals nuanced spatial dynamics such as the negative energy penalty of bathrooms when total area is controlled. This provides architectural planners with crucial heuristic insights into how interior space allocation inherently alters a home's thermal load profile, even before applying complex building information modeling simulations.

Third, the proposed regression equation serves as a highly accessible forecasting tool. It empowers researchers and policymakers to rapidly and reasonably estimate the baseline annual energy consumption of a residential unit based purely on basic architectural parameters and local climate data.

A follow-up study should focus on segmenting the residential energy consumption survey data by specific housing typologies (e.g., single-family detached and multi-family apartments) and integrating advanced non-linear machine learning algorithms to capture the complex, compounding interactions between structural age, envelope retrofits, and smart thermostat adoptions.

References

[1] L. Perez-Lombard, J. Ortiz, C. Pout, Review on buildings energy consumption information, Energy and Buildings, 40(3), 2008, pp.394-398.
[2] S. Borgeson, G. Brager, Benchmarking building energy performance, Center for the Built Environment, UC Berkeley, 2011.
[3] W. Chung, Review of building energy-use performance benchmarking methodologies, Applied Energy, 88(5), 2011, pp.1470-1479.

[4] Energy Star, Energy Star portfolio manager methodology for greenhouse gas inventory and tracking calculations, U.S. Environmental Protection Agency, 2020.
[5] M.-T. Ke, C.-H. Yeh, J.-T. Jian, Analysis of building energy consumption parameters and energy savings measurement and verification by applying eQUEST software, Energy and Buildings, 61, 2013, pp.100-107.
[6] T. Sharp, Energy benchmarking in commercial office buildings, In Proceedings of the ACEEE Summer Study on Energy Efficiency in Buildings, 4, 1996, pp.321-329.
[7] P. De Wilde, The gap between predicted and measured energy performance of buildings: A framework for investigation, Automation in Construction, 41, 2014, pp.40-49.
[8] J. Ahn, A Comparative analysis of Influencing Factors in the residential energy consumption survey 2009, 2015, and 2020, KIEAE Journal, 26(4), 2026, pp.5-12, http://dx.doi.org/10.12813/kieae.2026.26.4.005
[9] U.S. Energy Information Administration, Residential Energy Consumption Survey (RECS) Methodology, U.S. Department of Energy, 2020.
[10] N. Kaza, Understanding the spectrum of residential energy consumption: A quantile regression approach, Energy Policy, 38(11), 2010, pp.6574-6585.
[11] R. Ewing, F. Rong, The impact of urban form on US residential energy use, Housing Policy Debate, 19(1), 2008, pp.1-30.
[12] H. Zhao, F. Magoules, A review on the prediction of building energy consumption, Renewable and Sustainable Energy Reviews, 16(6), 2012, pp.3586-3592.
[13] K. Amasyali, N. El-Gohary, A review of data-driven building energy consumption prediction studies, Renewable and Sustainable Energy Reviews, 81, 2018, pp.1192-1205.
[14] T. Catalina, V. Iordache, B. Caracaleanu, Multiple regression model for fast prediction of the heating energy demand, Energy and Buildings, 57, 2013, pp.302-312.
[15] Z. O'Neill, B. Dong, Data-driven modeling of building energy consumption, In: J.L.M. Hensen, R. Lamberts, editors, Building performance simulation for design and operation, Routledge, 2017.
[16] N. Fumo, M. Biswas, Regression analysis for prediction of residential energy consumption, Renewable and Sustainable Energy Reviews, 47, 2015, pp.332-343.
[17] J. Hygh et al., Multivariate regression as an energy assessment tool in early building design, Building and Environment, 57, 2012, pp.165-175.
[18] J. Ahn, Development of energy performance metrics for airport terminal buildings using multivariate regression modeling (Doctoral disseratation), Raleigh, NC: North Carolina State University, 2016.
[19] J. Ahn, S. Cho, D. Chung, Development of a statistical analysis model to benchmark the energy use intensity of subway stations, Applied Energy, 179, 2016, pp.488-496. https://doi.org/10.1016/j.apenergy.2016.06.065
[20] C. Kim, H. Shin, J. Ahn, Energy performance analysis of airport terminal buildings by use of architectural, operational information and benchmark metrics, Journal of Air Transport Management, 83, 2020, 101762. https://doi.org/10.1016/j.jairtraman.2020.101762
[21] J. Ahn, A benchmark methodology to assess the energy performance of train station complexes, KIEAE Journal, 19(6), 2019, pp.5-12. http://dx.doi.org/10.12813/kieae.2019.19.6.005
[22] G. Casella, G. Berger, Statistical inference, Pacific Grove, CA: Duxbury, 2002.

ⓒ 2026. Korea Institute of Ecological Architecture and Environment all rights reserved.