OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

# Modeling Energy Use Intensity and Carbon Emissions in Seattle Urban Buildings Using Public Benchmarking Data

David Fernandez-Yoon
davidresearch12@gmail.com

## ABSTRACT

Urban buildings play an important role in energy consumption and greenhouse gas emissions, making building performance an essential target for sustainable engineering and city planning. In this study, energy use intensity, building size, and property type were examined to see if they could predict greenhouse gas emissions among large benchmarked buildings in Seattle. Using publicly available 2024 data from the Seattle Building Energy Benchmarking Program, a total of 3,616 buildings were included in the final sample after cleaning datasets and analyzing them. To evaluate relationships between building characteristics and total greenhouse gas emissions, descriptive statistics, Pearson correlations, property-type comparisons, and multiple linear regression were used. Due to the nature of strongly right-skewed distribution of emissions, Site Energy Use Intensity (EUI), and gross floor area, natural-log transformations were applied in the regression model. Total emissions were positively associated with Site EUI ($r = 0.514$) and also with gross floor area ($r = 0.459$). Approximately 79.1% of the variation in log-transformed greenhouse gas emissions was explained by the multivariable regression model. Both log Site EUI and log gross floor area were significant positive predictors ($p < 0.001$). Emission levels also varied across property types. Substantially higher median emissions were reported in Health/Lab and Hotel buildings than Multifamily buildings. These findings indicate that practical, data-driven evaluation of urban building energy and carbon performance can be supported by public benchmarking data.

**Keywords:** *Building energy benchmarking; energy use intensity; greenhouse gas emissions; urban buildings; sustainable architecture; multiple linear regression*

## INTRODUCTION

Buildings play an important role for urban energy demand and also for climate-change mitigation. Commercial and multifamily buildings consume energy for heating, cooling, lighting, ventilation, and equipment. This creates substantial direct and indirect greenhouse gas emissions. According to earlier reviews, buildings were identified as a main share of energy consumption in developed economies. Therefore, efficiency has become a central component to sustainable urban development (Perez-Lombard et al., 2008). In more recent research, the need for reliable building-level information has been emphasized to evaluate energy use, identify major drivers, and support decarbonization strategies

(Gonzales-Torres et al., 2022). This creates a need for quantitative methods to identify inefficient building performance for civil and industrial engineering.

## LITERATURE REVIEW

Measured building performance is compared with past performance or with similar properties by building energy benchmarking. Transparency is improved by benchmarking and rating systems, helping users interpret energy performance across buildings (Perez-Lombard et al., 2009). Statistical benchmarking methods, such as regression, can explain differences among properties when assessing efficiency (Chung, 2011). When evaluating efficiency across buildings of different sizes, data-driven frontier methods have also been used (Kavousian and Rajagopal, 2014). Furthermore, according to studies of disclosure datasets, relatively simple benchmarking information can provide useful predictive information about actual building energy performance (Hsu, 2014).

Municipal benchmarking laws have created useful datasets for urban energy research. Kontokosta developed a market-specific performance index by incorporating occupancy and physical building characteristics through New York City data (Kontokosta, 2015). Information barriers may be reduced by benchmarking ordinances, encouraging efficiency investments (Palmer and Walls, 2017). According to city-scale modeling, publicly reported energy data and building characteristics may predict energy use across large urban building stocks (Kontokosta and Tull, 2017). Furthermore, reductions in energy use intensity over time were associated with benchmarking and disclosure policies in New York City (Meng et al., 2017).

These approaches have been expanded by recent research. Particularly, energy drivers across different levels of building performance have been identified by quantile regression (Roth and Rajagopal, 2018). In addition, longitudinal research has examined performance under urban disclosure requirements (Yang et al., 2018). Neighborhood urban form has been associated with multifamily building energy use in Seattle (Ahn and Sohn, 2019). Benchmarking policy effects in Chicago have been evaluated by other studies (Shang et al., 2020), developing multi-city models of energy use and carbon emissions (Li et al., 2024) and proposing improved benchmarking indices as a way of addressing limitations of conventional floor-area-based EUI measures (Lee et al., 2025).

Despite this literature, only a few studies provide a simple and interpretable analysis of how greenhouse gas emissions have been related to energy use intensity, gross floor area, and property type jointly within Seattle's benchmarked building stock. This study seeks to address the literature gap using benchmarking data in Seattle and the multiple linear regression to assess relationships between readily available building characteristics and operational carbon emissions. Detailed energy-performance data are provided by Seattle's benchmarking program for its large nonresidential and multifamily building stock, making it well suited for this type of analysis. This study aims to answer the research question about how building size, property type, and energy use intensity are associated with greenhouse gas emissions among benchmarked urban buildings in Seattle. This study hypothesizes that buildings with higher energy use

intensity and larger gross floor area will have higher greenhouse gas emissions, and that adjusted greenhouse gas emission levels will differ among property types.

## METHODS AND MATERIALS

### Dataset and Study Design

In this study, publicly available data from the Seattle Building Energy Benchmarking Program administered by the City of Seattle Office of Sustainability and Environment were used. In the program, large nonresidential and multifamily buildings in Seattle were required to report annual information on energy consumption and building characteristics. In the complete dataset, a total of 34,699 building-year observations were included from 2015 to 2024. Due to the nature of repeated appearance of individual buildings across multiple years, only observations from 2024 were chosen for this study, creating a cross-sectional dataset, while avoiding treating repeated measurements of the same building as independent observations.

The 2024 subset initially contained a total of 3,712 building records. Inclusion in the analytical sample required usable values for Site EUI, total GHG emissions, gross floor area, EPA property type, year built, and number of floors. Eighty-seven records had at least one missing required value, and nine additional records reported a non-positive number of floors. This resulted in a final analytical sample of 3,616 buildings. Site EUI (86 records), total GHG emissions (62 records), and EPA property type (6 records) were the most common missing variables. These counts overlap as some observations were missing more than variable. All retained values for Site EUI, total GHG emissions, and gross floor area were positive. Therefore, they were suitable for natural-log transformation.

### Variables

Total Greenhouse Gas Emissions, measured in metric tons of carbon dioxide equivalent (metric tons $CO_2e$) was the dependent variable. Site Energy Use Intensity, measured in thousand British thermal units per square foot (kBtu/sf) was the primary independent variable. Site EUI shows the amount of energy consumed by a building relative to its floor area. Since Site EUI shows energy use normalized by floor area, while total GHG emissions are divided by total building energy consumption and the emissions associated energy sources used, Site EUI and gross floor area are structurally related to total emissions.

As a measure of building size, gross floor area, measured in square feet, was included. Building age was calculated as 2024 minus year built, while including the number of floors as an additional structural characteristic. Detailed EPA property classifications were consolidated into nine categories of Multifamily, Office, Education, Hotel, Warehouse/Storage, Retail/Food, Mixed Use, Health/Lab, and Other. Among them, multifamily buildings were regarded as the reference category in the regression analysis.

**Mathematical Model**

To estimate the relationship between building characteristics and greenhouse gas emissions, a multiple linear regression model was generated. Due to the strongly right-skewed distribution of GHG emissions, Site EUI, and gross floor area, natural-log transformations were applied. The model was specified as follows:

$$\ln ln\left(GHG_i\right) = \beta_0 + \beta_1 \ln ln\left(EUI_i\right) + \beta_2 \ln ln\left(GFA_i\right) + \beta_3\left(\frac{Age_i}{10}\right) + \beta_4\left(\frac{Floors_i}{10}\right) + \sum_{k=1}^{8} \gamma_k PropertyType_{ik} + \varepsilon_i$$

Where $GHG_i$ represents total greenhouse gas emissions for building i, $EUI_i$ shows Site Energy Use Intensity, and $GFA_i$ represents gross floor area. $\beta_0$ is the intercept, $\beta_1$-$\beta_4$ are coefficients for the continuous building characteristics, $\gamma_k$ shows the coefficient for each property-type indicator relative to the Multifamily reference category, and $\varepsilon_i$ is the error term

**Statistical Analysis**

Python-based statistical procedures were used to process and analyze data. Including mean, standard deviation, median, interquartile range, minimum, and maximum, descriptive statistics were calculated for major variables. Due to strongly right-skewed distribution of both variables, median Site EUI and GHG emissions were also compared across property groups.

To examine unadjusted relationships among emissions, Site EUI, gross floor area, building age, and number of floors, Pearson coefficients were initially calculated using the original values. Since GHG emissions, Site EUI, and gross floor area were strongly right-skewed, Pearson correlations were also calculate using their natural-log transformed values as a robustness check. To determine whether Site EUI and building size remained associated with GHG emissions after adjustment for property type, age, and floors, the multiple linear regression model was estimated. To reduce sensitivity to heteroscedasticity, robust standard errors were used. Using $R^2$ and adjusted $R^2$, model performance was evaluated, while defining statistical significance at $p < 0.05$.

**Results**

In the final analysis, a total of 3,616 Seattle buildings from the 2024 benchmarking year were included after excluding the records with missing or incomplete values for the primary variables. In both building energy use and greenhouse gas (GHG) emissions, a significant amount of variation was observed. Mean Site Energy Use Intensity (EUI) was 52.75 kBtu/sf (SD = 143.82), while the median was substantially lower value at 35.20 kBtu/sf. This indicates a right skewed distribution. Mean total GHG emissions was 124.98 metric tons $CO_2e$ (SD = 568.31), and the median of it was 34.00 metric ton $CO_2e$. The mean gross floor area of buildings was 110,268 $ft^2$, along with a mean age of 48.60 years and an average of 5.33 floors (Table 1).

| Variable | N | Mean | SD | Median | IQR | Minimum | Maximum |
|---|---|---|---|---|---|---|---|
| Site EUI (kBtu/sf) | 3,616 | 52.75 | 143.82 | 35.20 | 25.70 | 0.60 | 5,666.40 |
| Total GHG emissions (metric tons $CO_2e$) | 3,616 | 124.98 | 568.31 | 34.00 | 85.90 | 0.10 | 18,907.00 |
| Gross floor area ($ft^2$) | 3,616 | 110,268.20 | 185,219.19 | 49,285.50 | 75,470.50 | 20,008 | 2,099,379 |
| Building age (years) | 3,616 | 48.60 | 35.17 | 41.00 | 49.00 | 1 | 124 |
| Number of floors | 3,616 | 5.33 | 6.04 | 4.00 | 3.00 | 1 | 76 |

Table 1. Descriptive Statistics for the 2024 Seattle Building Sample

According to the bivariate analysis, total GHG emissions were positively correlated with both energy use intensity and building size. Total GHG emissions were moderately positively correlated with Site EUI ($r = 0.514$), and emissions were positively correlated with gross floor area ($r = 0.459$). The relationships remained positive after log transformation: ln(GHG emissions) was correlated with ln(Site EUI) at $r = 0.674$ and with ln(gross floor area) at $r = 0.624$ (both $p < 0.001$). This indicated that the observed relationships were not attributable solely to extreme observations. Figure 1 shows the relationship between Site EUI and GHG emissions. Due to strongly right-skewed distribution of both variables, logarithmic scales were used in the figure. As shown in the distribution, buildings with greater energy consumption per square foot generally showed higher total emissions even though much variation still remained among buildings with smaller EUI values.

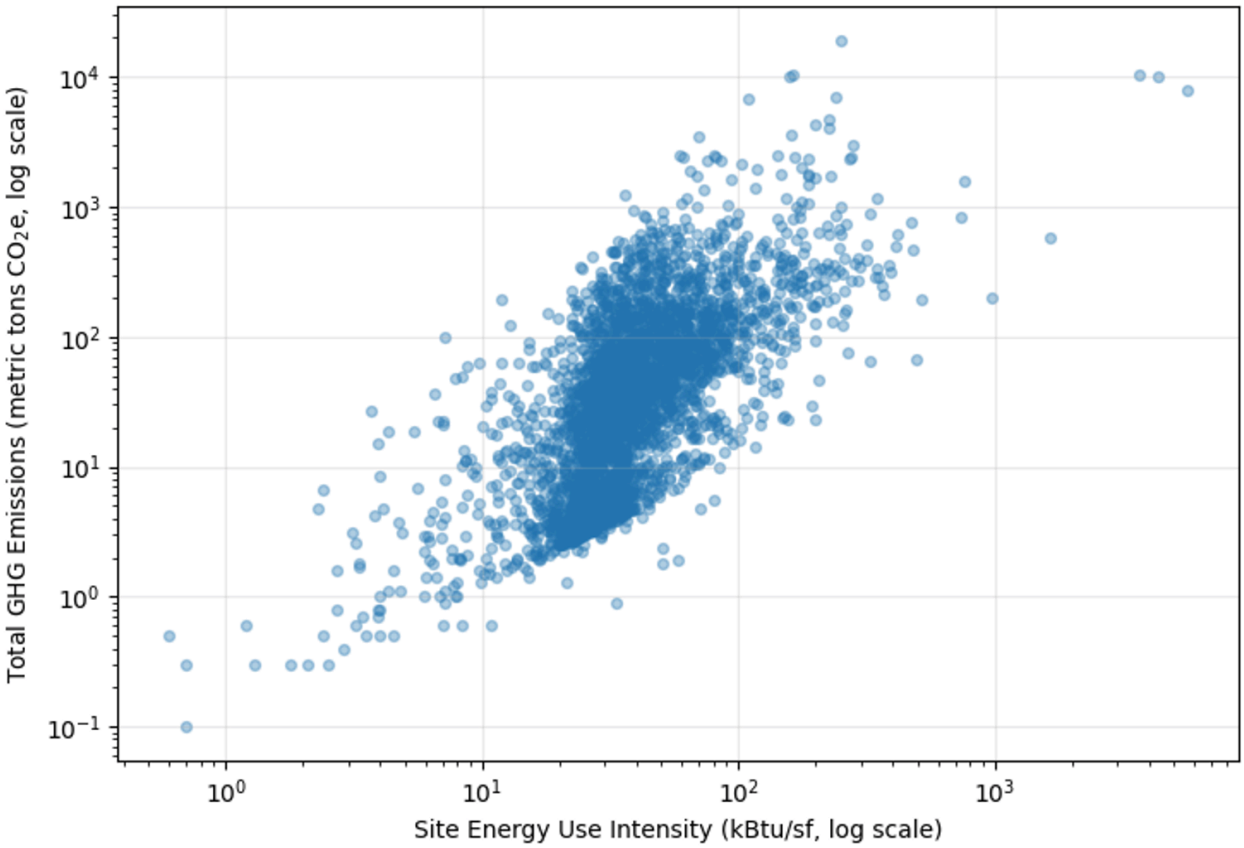



Figure 1. Relationship between Site Energy Use Intensity and total greenhouse emissions among a total of 3,616 Seattle buildings in 2024. Both axes are displayed on logarithmic scales.

Large differences were also observed across property categories. The groups sizes were Multifamily (n = 1,951), Office (n = 514), Other (n = 383), Warehouse/Storage (n = 240), Education (n = 172), Mixed Use (n = 140), Retail/Food (n = 92), Hotel (n = 82), and Health/Lab (n = 42). The highest median total GHG emissions were shown in Health/Lab buildings at 691.85 metric Tons $CO^2e$, followed by Hotels at 183.1, Mixed Use Buildings at 86.15, and Education buildings at 73.10 metric tons $CO^2e$. As the largest group, multifamily buildings had a median emission level of only 21.90 metric ton $CO_2e$. A similarly low median of 22.20 metric tons $CO_2e$ was shown in Warehouse/Storage buildings. The highest median Site EUI was also shown in Health/Lab properties at 187.60 kBtu/sf, compared with 31.70 kBtu/sf for multifamily buildings (Figure 2).

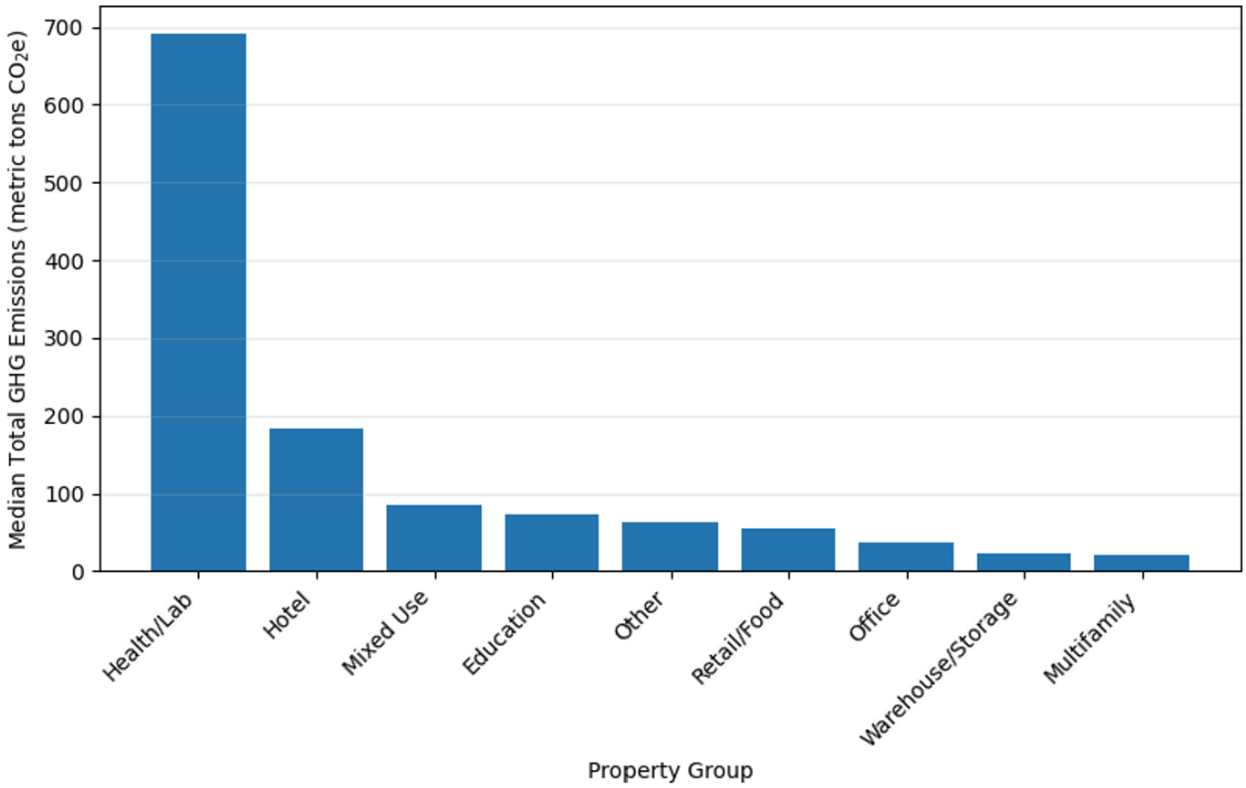



Figure 2. Median total greenhouse gas emissions across the nine consolidated property groups in the 2024 Seattle benchmarking dataset.

To determine whether Site EUI and building size remained associated with emissions after considering property type, building age, and number of floors, multiple linear regression was used. Around 79.1% of the variation in the log-transformed GHG emissions was explained by the model ($R^2 = 0.791$; adjusted $R^2 = 0.790$). Log-transformed Site EUI reported to be a strong positive predictor of emission ($\beta = 1.418$, robust SE = 0.025, $p < 0.001$, 95% CI [1.368, 1.468]). Gross floor area was also strongly associated with emissions ($\beta = 1.017$, SE = 0.017, $p < 0.001$, 95% CI [0.983, 1.051]) (Table 2).

| **Predictor** | **β** | **Robust SE** | **95% CI** | **p-value** |
|---|---|---|---|---|
| Intercept | -13.271 | 0.204 | [-13.670, -12.872] | <0.001 |
| ln(Site EUI) | 1.418 | 0.025 | [1.368, 1.467] | <0.001 |
| ln(Gross floor area) | 1.017 | 0.017 | [0.983, 1.051] | <0.001 |
| Building age / 10 years | 0.048 | 0.004 | [0.040, 0.055] | <0.001 |
| Floors / 10 floors | 0.095 | 0.023 | [0.049, 0.141] | <0.001 |
| Education | 0.617 | 0.049 | [0.522, 0.713] | <0.001 |
| Health/Lab | -0.105 | 0.093 | [-0.287, 0.077] | 0.258 |
| Hotel | 0.339 | 0.062 | [0.218, 0.460] | <0.001 |

| Predictor | β | Robust SE | 95% CI | p-value |
|---|---|---|---|---|
| Mixed Use | -0.071 | 0.065 | [-0.199, 0.056] | 0.273 |
| Office | -0.438 | 0.039 | [-0.514, -0.363] | <0.001 |
| Other | 0.252 | 0.050 | [0.153, 0.350] | <0.001 |
| Retail/Food | 0.016 | 0.081 | [-0.143, 0.176] | 0.841 |
| Warehouse/Storage | 0.608 | 0.054 | [0.502, 0.714] | <0.001 |

Table 2. Multiple Linear Regression Results Predicting Log-Transformed Total GHG Emissions. Outcome variable is ln(total GHG emissions), N = 3,616; $R^2$ = 0.791; adjusted $R^2$ = 0.790. Multifamily is the reference property category. HC3 is the robust standard error. EUI = Energy Use Intensity; CI = confidence interval.

After adjustment, both building age and number of floors remained statistically significant. Each additional 10 years of building age was associated with a coefficient of $\beta = 0.048$ ($p < 0.001$), while additional 10 floors had a coefficient of $\beta = 0.094$ ($p < 0.001$). Compared to multifamily buildings, significantly higher adjusted log emissions were observed in Education ($\beta = 0.617$), Warehouse/Storage ($\beta = 0.608$), Hotel ($\beta = 0.339$), and Other properties ($\beta = 0.245$). However, significantly lower adjusted emissions were observed in Office properties ($\beta = -0.435$; all $p < 0.001$). Overall, the hypothesis in this study was supported by the findings that higher greenhouse gas emissions were associated with higher energy use intensity and greater building size, but emission patterns varied among property types.

## DISCUSSION

In this study, energy use intensity, building size, and property type were examined for their association with greenhouse gas emissions among benchmarked Seattle buildings. The hypothesis in this study was supported by the findings that greater total GHG emissions were produced by buildings with higher Site EUI and larger gross floor area. Approximately 79% of the variation in log-transformed emissions was explained by the regression model, suggesting that a large portion of the differences in operational carbon emissions may be captured by a relatively simple set of building characteristics. Both Site EUI and gross floor area were strongly associated with total GHG emissions in the regression model, highlighting the importance of considering both energy efficiency and building scale when assessing urban building performance. The relatively high model fit should not be interpreted as evidence of independent causality of Site EUI and floor area to greenhouse gas emissions. Part of the strong associations and high $R^2$ is expected from the construction of these measures as total emissions are divided from total energy use that is related to both energy intensity and floor area. Therefore, the regression should be primarily interpreted as describing associations within the benchmarking dataset.

Differences across property types were also substantial. In particular, much higher median emissions were observed in Health/Lab and Hotel buildings than Multifamily and Warehouse/Storage buildings. These differences may reflect variation in occupancy patterns, equipment loads, ventilation requirements, and operating schedules across different building functions. Therefore, the findings in this study suggest that

building type should be considered when identifying high-emission properties or prioritizing energy-efficiency improvements with benchmarking programs. For all building categories, a single performance threshold may not be equally appropriate.

Despite all aforementioned findings, there are several limitations to be considered in this study. First, cross-sectional data from 2024 were used in this study. Therefore, the results identify associations rather than causal relationships. Second, since benchmarking data are self-reported, there may be reporting or measurement errors contained in the dataset. Third, the model may not include potentially important factors, including occupancy density, renovation history, HVAC system type, local microclimate, and detailed operating schedules. Finally, large benchmarked buildings in Seattle were mainly used in this study that may limit generalizability to smaller buildings or other cities.

Overall, the findings in this study show that practical, data-driven approaches can be supported by public benchmarking data for urban energy planning and building decarbonization.

## CONCLUSION

In this study, 2024 Seattle Building Energy Benchmarking data were used to examine how greenhouse gas emissions were associated with energy use intensity, building size, and property type in large urban buildings. According to the results, both Site EUI and gross floor area were strongly associated with total GHG emissions. However, emission levels turned out to vary across building categories. Approximately 79% of the variation in log-transformed emissions was explained by the regression model, showing that useful information about building-level carbon performance can be provided by a relatively simple set of building characteristics. These findings emphasize the value of public benchmarking datasets when identifying high-emission buildings and supporting energy-efficiency planning. In spite of the limitations of this study to one city and one reporting year, a practical framework is provided to apply statistical modeling to urban sustainability questions. It is recommended for future studies to apply similar approaches to extend the scope of research with additional cities or multiple years to evaluate broader patterns in building energy use and decarbonization.

## REFERENCES

Ahn, Y. J., & Sohn, D. W. (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 and Buildings, 196, 124–133.

Chung, W. (2011). Review of building energy-use performance benchmarking methodologies. Applied Energy, 88(5), 1470–1479.

González-Torres, M., Pérez-Lombard, L., Coronel, J. F., Maestre, I. R., & Yan, D. (2022). A review on buildings energy information: Trends, end-uses, fuels and drivers. Energy Reports, 8, 626–637.

Hsu, D. (2014). How much information disclosure of building energy performance is necessary? Energy Policy, 64, 263–272.

Kavousian, A., & Rajagopal, R. (2014). Data-driven benchmarking of building energy efficiency utilizing statistical frontier models. Journal of Computing in Civil Engineering, 28(1), 79–88.

Kontokosta, C. E. (2015). A market-specific methodology for a commercial building energy performance index. The Journal of Real Estate Finance and Economics, 51, 288–316.

Kontokosta, C. E., & Tull, C. (2017). A data-driven predictive model of city-scale energy use in buildings. Applied Energy, 197, 303–317.

Lee, K., Lim, H., Hwang, J., & Lee, D. (2025). Development of building benchmarking index for improving gross-floor-area-based energy use intensity. Energy and Buildings, 328, 115103.

Li, T., Liu, T., Sawyer, A. O., Tang, P., Loftness, V., Lu, Y., & Xie, J. (2024). Generalized building energy and carbon emissions benchmarking with post-prediction analysis. Developments in the Built Environment, 17, 100320.

Meng, T., Hsu, D., & Han, A. (2017). Estimating energy savings from benchmarking policies in New York City. Energy, 133, 415–423.

Palmer, K., & Walls, M. (2017). Using information to close the energy efficiency gap: A review of benchmarking and disclosure ordinances. Energy Efficiency, 10, 673–691.

Pérez-Lombard, L., Ortiz, J., & Pout, C. (2008). A review on buildings energy consumption information. Energy and Buildings, 40(3), 394–398.

Pérez-Lombard, L., Ortiz, J., González, R., & Maestre, I. R. (2009). A review of benchmarking, rating and labelling concepts within the framework of building energy certification schemes. Energy and Buildings, 41(3), 272–278.

Roth, J., & Rajagopal, R. (2018). Benchmarking building energy efficiency using quantile regression. Energy, 152, 866–876.

Shang, L., Lee, H. W., Dermisi, S., & Choe, Y. (2020). Impact of energy benchmarking and disclosure policy on office buildings. Journal of Cleaner Production, 250, 119500.

Yang, E., Lee, Y., & Li, Q. (2018). Energy disclosure law in New York City: Building energy performance benchmarking progress. Facilities, 36(11/12), 571–583.