

#### **PAPER • OPEN ACCESS**

## Cluster analysis-based energy performance assessment for office building stock

To cite this article: Ji Hyun Oh et al 2023 J. Phys.: Conf. Ser. 2600 082025

View the article online for updates and enhancements.

#### You may also like

- Harvesting big data from residential building energy performance certificates: retrofitting and climate change mitigation insights at a regional scale João Pedro Gouveia and Pedro Palma
- Local accommodation energy efficiency in Lisbon: a red flag for tourism, indoor thermal comfort, and energy renovation targets Inês Valente, Bárbara Fernandes, André César et al.
- <u>Science for Policy: Insights from</u> <u>Supporting an EU Roadmap for the</u> <u>Reduction of Whole Life Carbon of</u> <u>Buildings</u> M Röck, G Pristerá, D Ramon et al.

# Join the Society Led by Scientists, for Scientists Like You!



This content was downloaded from IP address 202.30.9.62 on 23/05/2025 at 09:00

### **Cluster analysis-based energy performance assessment for** office building stock

#### Ji Hyun Oh<sup>1</sup>, Hye Gi Kim<sup>2</sup> and Sun Sook Kim<sup>3,\*</sup>

<sup>1</sup> Department of Smart Convergence Architecture, Ajou University, 206 wordcupro, Suwon, Gyeonggi-do 16499, South Korea

<sup>2</sup> Department of Building Energy Research, Korea Institute of Civil Engineering and Building Technology, Goyangdaero, Goyang-Si, Gyeonggi-Do 10223, South Korea <sup>3</sup> Department of Architecture, Ajou University, 206 wordcupro, Suwon, Gyeonggi-do 16499, South Korea

\* Corresponding author : kss@ajou.ac.kr

Abstract. To achieve carbon neutrality at the national or city level, the energy performance and conservation measures of large buildings should be evaluated. However, the assessment of the energy performance of existing building stock is often based on annual energy use intensity calculated from energy bill data due to data acquisition limitations. This approach has limitations in analyzing seasonal effects and establishing effective energy conservation strategies. In this paper, we propose a novel energy performance assessment method for existing office building stock. Our method classifies monthly electricity, gas, and heat energy use patterns using clustering algorithms without requiring additional database construction beyond the National Building Energy Database in Korea. We discuss the clustering results and provide an application method to assist policymakers through a case study.

#### 1. Introduction

To achieve carbon neutrality at the national or city level, the energy performance of building stock should be evaluated and energy conservation measures implemented. The execution of energy performance assessments for building stock often differs from those for individual buildings because the former can be time-consuming. Many countries have therefore implemented benchmarking systems for assessing existing building stock, such as Energy Star in the USA, Display Energy Certificates in the UK, and Commercial Building Disclosure in Australia. These systems evaluate buildings based on their annual energy use intensity (EUI, kWh/m<sup>2</sup>yr), which considers factors such as building usage type and region. This approach is useful for existing building stock, as it minimizes the need for additional data collection and allows for evaluation using only monthly bill data. This approach has also been adopted in Korea. The disclosure of building energy performance information, including annual EUI by energy source, annual carbon emissions, and energy efficiency rating, is mandatory for apartment buildings (over 100 units) and office buildings (over 2000 m<sup>2</sup> total floor area). While this annual value is intuitive, it has limitations in analyzing seasonal effects and establishing effective energy conservation strategies. It only provides information on whether the energy use of a building is higher or lower than that of its peer group or itself.

In contrast, analyzing energy use patterns over time can provide insights into the energy use characteristics of buildings, and unsupervised clustering methods have been widely employed for this

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI. Published under licence by IOP Publishing Ltd 1

purpose [1,2]. However, most studies have collected data using metering devices for individual or a small number of buildings, which can be costly and difficult to apply to building stock. Some studies [3,4] have proposed energy performance analysis methods for building stock using monthly energy bill data, but these studies only considered electricity and presented results in terms of annual EUI and energy signature values.

Using a 12-month EUI pattern, both energy use quantity and characteristics can be evaluated, including heating, cooling, and base energy. In a previous study [5], we employed k-means clustering analysis to derive five monthly energy use patterns for office buildings in Korea and suggested energy conservation strategies for each pattern. However, there were limitations to estimating equipment information using only monthly total EUI data. Therefore, the objective of this study is to propose an energy performance assessment method for building stock that can support policymakers' decision-making using monthly EUI data according to energy source. This method utilizes existing national databases, eliminating the need for additional data collection costs. We discuss the feasibility of this method through a case study.

#### 2. Method

#### 2.1. Data acquisition and preprocessing

The scope of this study is limited to office building stock, for which we collected basic building information and energy use data from the National Building Energy Database in Korea. The energy data were obtained from 2019 monthly energy bill information and include electricity, gas, and heat (including both district heating and cooling) energy data.

The following cases were then filtered from the raw 12,707 data points.

- Total floor area (TFA) or energy use data were missing.
- The energy use value was negative or the electric energy use was zero.
- Annual energy use exceeded 1.5 times the interquartile range.
- The energy use of a specific month was excessively high.

After applying all filters, a total of 11,947 building datasets were identified, and no missing values were present in the final dataset. Before conducting the analysis, we normalized the datasets using the Min-Max normalization method with a range of 0-1.

Because the monthly EUI data represent sequential observations with weak correlations between them, they can be considered both a time-series attribute and a feature attribute. Therefore, to better capture the energy use characteristics of the buildings, we incorporated additional features, such as heating, cooling, and mid-season energy ratios, as well as TFA, in addition to the monthly EUI. Table 1 lists the features and example values of the data before normalization. The seasonal energy use ratio represents the ratio of the sum of EUI for specific months within each season to the annual EUI.

Features	Unit	Example	Note
12-month energy use	kWh/m <sup>2</sup>	22.9	Jan 2019–Dec 2019
Heating season energy use ratio	%	49.1	Jan, Feb, Mar, Nov, and Dec
Cooling season energy use ratio	%	23.1	Jun, Jul, and Aug
Mid-season energy use ratio	%	27.8	Apr, May, Sep, and Oct
TFA	m <sup>2</sup>	5425.9	

#### 2.2. k-means clustering

In this study, the k-means algorithm was implemented with a Euclidean distance metric. The algorithm, which is an unsupervised learning technique, aims to partition n observations into a k cluster. Many studies have used this algorithm for analyzing energy use patterns. The number of k was determined via the Elbow method, Silhouette (Si) scores, and cluster centroid pattern. Si score is also a

representative validity indicator for quality of clustering results. The clustering analysis was performed in Python.

#### 3. Results and Discussion

#### 3.1. Data overview

Table 2 shows the analysis results for the energy source-type distribution. The energy source type was classified into E (electricity only), EG (electricity and Gas), EH (electricity and heat), and EGH (electricity, gas, and heat). The Building Energy Conservation Code in Korea recommends adopting nonelectricity cooling systems (e.g., district heating & cooling and renewable energy) for electricity peak management. These systems require specialized maintenance personnel and are adopted for energy mix in large buildings, whereas individual systems such as electricity and gas heat pump (EHP and GHP) are preferable for small buildings. The average TFA results in Table 2 reveal this tendency. A total of 97% of the office building stock in Korea utilized E or EG as their energy source type.

Type	Source			Number of observations	Average $TFA(m^2)$	
Type _	Electricity	Gas	Heat	(Ratio)	menuge mm(m)	
Е	•			6,678 (56%)	2,171.8	
EG	•	•		4,960 (41%)	6,798.7	
EH	•		•	114 (1%)	7,644.2	
EGH	•	•	•	195 (2%)	25,707.3	

#### 3.2. Clustering results

K-means clustering was conducted for electricity (obs. 11,947), gas (obs. 5,156), and heat (obs. 309). We determined the number of k for each of the aforementioned considering the Elbow method and average Si score results, as shown in Figure 1. First, we checked the elbow chart and narrowed the range of k to 2–5. The average Si score can represent the overall performance of the clustering results and is calculated for a range from -1 to 1. The higher the score the less ambiguously data are classified [2]. However, All average Si scores were less than 0.5. Because our dataset covered the entire office building stock and it was not segmented based on region or TFA, consequently the average Si scores were low.

We determined the final number of k based on cluster centroids results that demonstrated a significant difference in energy use pattern. As shown in Table 3, each number of k was determined as 2 for electricity, 4 for gas, and 3 for heat. Table 4 shows the characteristics and centroid results by cluster. The characteristics were labeled as per the centroid analysis results.



Figure 1. Elbow chart (left) and average Si score results (right) by energy source.

**IOP** Publishing

Energy source	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Electricity (k=2)	a ho her cyc hay so on her be:	normal sector se		
	70	»	70	70
Gas (k=4)	an the force for twy and july and the contract of the section of t	<sup>10</sup> <sup>20</sup> <sup>20</sup> <sup>20</sup> <sup>20</sup> <sup>20</sup> <sup>20</sup> <sup>20</sup> <sup>2</sup>	as hits Her, Apr. Shr. Jan. Base 405. Nrs. Dec. <b>g3</b>	an interview rays may any any any any any any any any any a
Heat (k=3)	a nite ner opr før jan jal nag beg och her der h	in the ther Apr Nav An Ang Ste Oc Nav. Dec. h2	an the late up first an Al Arg Ste Cle Nin Che. h3	

 Table 3. k-means clustering results by energy source.

Table 4. Analysis results of clustering characteristics and centroids.

	Cluster	Datio	Characteristic	TFA (m <sup>2</sup> )	Annual	Seasonal EUI ratio (%)		
Source		(%)			EUI <sup>a</sup> (kWh/m <sup>2</sup> )	Heating	Cooling	Mid
Elec.	e1	69.8	Low	3,935.2	60.9	46.9	23.8	29.4
(N=11,947)	e2	30.2	High	5,901.2	143.7	44.3	25.5	30.2
Gas (N=5,156)	g1	28.2	Base	8,342,8	16.7	46.7	23.0	30.3
	g2	9.9	H↑&C	4,971.1	120.7	63.5	12.9	23.6
	g3	27.3	H↓&C	9,644.6	72.5	50.1	25.1	24.9
	g4	34.6	Н	5,878.1	28.8	83.7	3.1	13.2
Heat (N=309)	h1	50.5	H↓	2119.8	23.6	88.1	3.3	8.6
	h2	38.5	H&C	17,811.4	57.6	42.4	40.0	17.6
	h3	11.0	$\mathrm{H}\uparrow$	11533.2	77.1	81.4	5.1	13.5

<sup>a</sup> Calculated by energy source respectively

For electricity, the energy use patterns were classified according to the EUI value in all cases of k (2–5). This indicates that the clustering was solely based on the amount of EUI and was not influenced by seasonal pattern changes. Because in buildings, electricity is required for various purposes, for example, lighting, appliance, and HVAC equipment regardless of season, we selected k = 2, which had the highest Avg.Si value of 0.448.

On the other hand, the gas and heat results demonstrated differences in energy use patterns according to the season. Gas with k = 4 (Avg.Si = 0.306) was classified into a pattern of steady usage levels within a year (g1), increased usage during heating and cooling seasons (g2 and g3), and increased usage during the heating season only (g4). Accordingly, we estimated the gas energy use for each cluster. All gas clusters have a baseline energy usage for cooking and hot water supply. Cluster g1 uses gas energy as a boiler heat source for hot water, whereas g2 and g3 are used for heating and cooling systems, such as GH, and g4 is used for a heating system only.

The heat energy use pattern with k = 3 (Avg.Si = 0.446) was classified into patterns with increased usage only during the heating season (h1 and h3) and increased usage during the heating and cooling season (h2). In some large buildings, hot water absorption chiller utilize district heating and cooling as

the heat source and are used to reduce electricity peak throughout all season. These cases correspond to h2, where the heating and cooling seasonal EUI ratio was 42.4% and 40.0%, respectively.

#### 3.3. Application

To verify the feasibility of the clustering analysis results, we analyzed three cases by considering HVAC equipment information. Table 5 shows the basic building and energy usage information of the selected cases. All three cases with different construction years, TFA, and energy source type were selected.

Bld. Appro date of	Approval	TFA	Annual EUI (kWh/m <sup>2</sup> )	Seasonal EUI ratio (%) <sup>a</sup>			EUI ratio by source (%) <sup>a</sup>		
	date of use	(m <sup>2</sup> )		Heating	Cooling	Mid	Elec.	Gas	Heat
А	Sep 2003	11,119.5	147.1	41.8	32.3	25.8	75.1	0.2	24.7
В	Oct 1996	49,888.4	135.7	44.3	30.1	25.5	65.0	2.5	32.5
С	Apr 2013	6,387.7	168.8	41.9	26.8	31.3	60.2	39.8	-

Table 5. Building and energy information of case buildings.

<sup>a</sup> Calculated based on total annual energy use data.

The HVAC system information and clustering results are shown in Table 6. The cluster names correspond to the characteristics describe in Table 4. The clustering results for total monthly energy use were derived from a previous study, which classified the data into five clusters; [Low use], [Typical use], [High use], [Over use], and [Heating sensitive] [5] and all were classified as the same [High use] building. However, energy source-based clusters were assessed differently depending on the conditions of each building. In cases A and B, gas and heat were classified as [Base] and [H&C], respectively, which matched the information on using a hot water absorption chiller and a gas-fired steam boiler for humidification. On the other hand, electricity was classified as [High] for case A and [Low] for case B. This was attributed to two reasons. First, the annual EUI of case A (147.1 kWh/m<sup>2</sup>) is higher than that of case B (135.7 kWh/m<sup>2</sup>), indicating high usage of appliances. Second, the electricity EUI of case A was relatively overestimated due to the small TFA. Case C was classified into the [high] and [H↓&C] clusters for electricity and gas, which was consistent with the actual HVAC system condition using EHP and GHP. Because the gas energy use of Case C was 67.2 kWh/m<sup>2</sup> (39.8% of the total), it was classified as [H↓&C] not [H↑&C].

Table 6. HVAC system information and cluster results of case buildings.

D14	Uppting and popling system	Cluster				
Ыû.	Heating and cooling system	Total	Elec.	Gas	Heat	
A	<ul><li> 2 Hot water absorption chiller</li><li> 4 AHU</li><li> 1 Gas-fired steam boiler for humidifying AHU</li></ul>	High use	High	Base	H&C	
В	<ul><li> 4 Hot water absorption chiller</li><li> 14 AHU</li><li> 1 Gas-fired steam boiler for humidifying AHU</li></ul>	High use	Low	Base	H&C	
С	<ul> <li>5 EHP outdoor units (Clg: 300 kW / Htg: 342.4 kW)<sup>a</sup></li> <li>10 GHP outdoor units (Clg: 494 kW / Htg: 552 kW)<sup>a</sup></li> <li>1 Gas-fired boiler for hot water</li> </ul>	High use	High	H↓&C	-	

<sup>a</sup> Sum of capacity of all outdoor units

The case study results reveal that the HVAC system types and energy usage characteristics can be estimated based on the energy source-based clustering results. From the perspective of policy makers,

even if the total energy use is classified as [High use], different strategies can be developed depending on the energy source results. For instance, in the case of an office building stock with similar results as Case A, despite energy mix, energy conservation measures can be proposed by policymakers due to address the high total and electricity usage. The followings are examples of energy conservation measures:

- Replace existing lighting and equipment with high-efficiency lighting and equipment (e.g., fan, pump)
- Inspect fan and pump operation.
- Inspect appliance and lighting operation schedule.
- Install renewable energy system.
- Install building energy monitoring and HVAC control system.

#### 4. Conclusion

In this study, we proposed a simple energy performance assessment method for existing building stock with sparse building information and monthly energy bill data. The suggested method has the advantage of utilizing existing national databases without the need for additional database construction. We discussed the feasibility of the method through a case study, which revealed that even though all the total energy use clusters were classified as [High use] buildings, the HVAC systems could be estimated, and appropriate energy conservation measures could be recommended from the energy source-based results. However, for electricity, only the usage amount could be verified through the clusters, and additional building information was necessary to explain the high or low usage.

The results of this study can contribute to the assessment of building energy performance at the city or country level and inform policymakers in developing specific carbon neutrality implementation plans and building energy efficiency policies, such as setting targets and determining scopes (e.g. priority of energy conservation measures). In future work, we will categorize the HVAC systems used in office buildings and develop energy conservation strategies based on matches between the HVAC system type and clusters.

#### Acknowledgments

This study was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Science, ICT & Future Planning (grant number NRF2020R1A2C1103033).

#### References

- Bourdeau, M., Basset, P., Beauchêne, S., Da Silva, D., Guiot, T., Werner, D., & Nefzaoui, E. (2021). Classification of daily electric load profiles of non-residential buildings. Energy and Buildings, 233, 110670.
- [2] Westermann, P., Deb, C., Schlueter, A., & Evins, R. (2020). Unsupervised learning of energy signatures to identify the heating system and building type using smart meter data. *Applied Energy*, 264, 114715.
- [3] Choi, S., & Yoon, S. (2023). Energy signature-based clustering using open data for urban building energy analysis toward carbon neutrality: A case study on electricity change under COVID-19. Sustainable Cities and Society, 92, 104471.
- [4] Zhou, H., Tian, X., Yu, J., Zhao, Y., Lin, B., & Chang, C. (2022). Identifying buildings with rising electricity-consumption and those with high energy-saving potential for government's management by data mining approaches. Energy for Sustainable Development, 66, 54–68.
- [5] Oh, Ji-Hyun, Kim, Hye-Gi, Choi, Byung-Ju, Kim and Sun-Sook. (2022). Development of Energy Conservation Measures for Office Buildings by Analyzing Monthly Energy Use Patterns. *Journal of the Architectural Institute of Korea*, 38(5), 139-146.