Analyzing Building Energy Consumption Patterns for Green Smart City Development Using a Data-driven Method

Jian Zhang [1], Xin Guo [1,a*], Hao Fu [1,b*], Tingxu Chen [2], Yue Zhang [2]

[1]School of Systems Science, Beijing Jiaotong University, Beijing, 100044, China
[2]School of Traffic and Transportation, Beijing Jiaotong University, Beijing, 100044, China

[a]goxin@bjtu.edu.cn; [b]haofl@bjtu.edu.cn.

Abstract. In the current era of dual-carbon construction, analyzing the energy consumption patterns of buildings is crucial for the development of green and low-carbon smart cities. This study analyzes the energy consumption of buildings with various functions, revealing the spatial and temporal patterns and their probability distributions. For the same type of buildings, the study further examines the correlations among main energy consumptions—water, electricity, and natural gas. The analysis is achieved by adopting the principal component analysis to reduce the dimensionality of high-dimensional building data, and employing the K-means clustering algorithm to categorize the energy consumption of buildings for various purposes. This study find that the energy consumption patterns of different functional buildings are different, and there is a positive correlation between all kinds of energy consumption, and the time variation is relatively strong. The conclusions can help to lay a solid foundation of building operations and maintenance in terms of the energy consumption and manage energy utilization for different regions and time periods for the development of green smart city.

Keywords: Building energy; Smart city; Principal component analysis; Clustering; Energy mapping

1 INTRODUCTION

The study of building energy consumption patterns within the context of smart city is of great significance for fostering sustainable development [1]. It uncover the distinct characteristics of energy usage across different building types and operational scenarios, paving the way for the development of more efficient energy solutions tailored to each unique situation [2]. The construction sector, a primary source of global energy use and carbon emissions [3], can dramatically reduce its carbon footprint and contribute to climate change mitigation by enhancing energy efficiency and decreasing dependence on fossil fuels. Neural network approaches are used to forecast energy consumption in residential settings [4]. Regarding research methodologies, Principal Component Analysis (PCA) has traditionally been applied to extract significant

B. Yuan et al. (eds.), *Proceedings of the 2024 8th International Conference on Civil Architecture and Structural Engineering (ICCASE 2024)*, Atlantis Highlights in Engineering 33,
https://doi.org/10.2991/978-94-6463-449-5_72

characteristics from various building groups [5], and an enhanced K-means clustering algorithm has been extensively utilized in forecasting various types of energy consumption with other algorithms [6], though the integration of these two methods remains underexplored.

This study utilizes a data-driven method, specifically PCA, to process complex, high-dimensional building data. Moreover, the K-means clustering algorithm is adeptly applied to precisely categorize building energy consumption, enhancing the data-driven, scientific refinement in the analysis of consumption patterns. The strengths of our approach lie in its capability to detect unusual energy consumption patterns.

2 METHODOLOGY

2.1 Data Descriptions

The open data set used in this study was released by the American Society of Heating, Refrigeration, and Air-Conditioning Engineers, with 3,053 meters across 1,636 buildings during the years 2016 and 2017.

2.2 Methods

A. Data dimensionality reduction algorithm

The specific steps are given in Algorithm 1. Consider the linear transformation as equation (1),

$$\begin{cases} Z_1 = \boldsymbol{a}_1'\boldsymbol{X} = a_{11}X_1 + a_{21}X_2 + \cdots + a_{p1}X_p \\ Z_2 = \boldsymbol{a}_2'\boldsymbol{X} = a_{12}X_1 + a_{22}X_2 + \cdots + a_{p2}X_p \\ \vdots \\ Z_p = \boldsymbol{a}_p'\boldsymbol{X} = a_{1p}X_1 + a_{2p}X_2 + \cdots + a_{pp}X_p \end{cases} \tag{1}$$

$$\mathrm{var}(Z_i) = \boldsymbol{a}_i'\Sigma\boldsymbol{a}_i,\ i = 1,2,\cdots,p \tag{2}$$

$$\mathrm{cov}(Z_i, Z_j) = \boldsymbol{a}_i'\boldsymbol{\Sigma}\boldsymbol{a}_j,\ i = 1,2,\cdots,p \tag{3}$$

Equation (2) and (3) calculates the variance and covariance matrix of the distribution. For $\mu_i = E(X_i)$, var $(X_i) = \sigma_i^2$,

$$X_i^* = \frac{X_i - E(X_i)}{\sqrt{\mathrm{var}(X_i)}} = \frac{X_i - \mu_i}{\sigma_i} \tag{4}$$

Equation (4) is used to achieve standardization.

Algorithm 1: PCA

Input: dataset $(X = x_1, x_2, \ldots, x_n)$

Output: the reduced dataset (Y=y1, y2, ... ,yn)

(1) Standardize each data point (x_i)

(2) Calculate the covariance matrix

(3) Find the eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_d)$and eigenvectors $(v_1, v_2, \ldots, v_d)$

(4) Arrange the eigenvalues and corresponding eigenvectors in descending order

(5) Select the eigenvectors $(v_1, v_2, \ldots, v_p)$

(6) Form a $d \times p$ projection matrix (W)
(7) Transform the original dataset to a new subspace $(Y = X'W)$
(8) Output the reduced dataset Y

B. K-means clustering algorithm

The K-means clustering algorithm used in this study contains the following points, and the specific steps are given in Algorithm 2:

(1) K-means algorithm needs to specify the number of clusters k in advance.

(2) μ_i is the mean vector of cluster C_i. $\mu_i = \frac{1}{m_i}\sum_{x \in C_i} x, m_i$ is the data in cluster C_i.

(3) The measure of the distance μ_i :

The formula of Euclidean distance is as equation (5).

$$d = \sqrt{(X-Y)^{\mathrm{T}}(X-Y)} = \sqrt{\sum_{i=1}^{n} (X_i - Y_\mathrm{i})^2} \quad (5)$$

X_i and Y_i denote the components of the sample vectors, and n denotes the dimension of the sample vectors.

(4) Our objective function is to minimize the sum of squares of errors, which is calculated as equation (6).

$\min \sum_{i=1}^{k} \sum_{x \in C_i} \text{dist}\,(\mu_i, x)^2$ (6)

Where μ_i is the center of mass of cluster c_i, dist (μ_i, x) denotes the distance from sample x to the center of mass of cluster μ_i.

Algorithm 2: K-means clustering algorithm

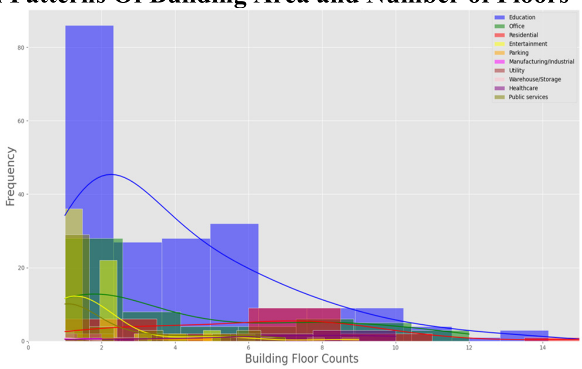
Input: Sample set $D = \{x_1, x_2, \cdots, x_n\}$, number of clusters k

Output: Cluster division $\{C_1, C_2 \cdots, C_k\}$

(1) Select k samples from D as the center of the initial cluster: $\{\mu_1, \mu_2 \cdots, \mu_k\}$
(2) Calculate the distance, and assign them to form k clusters
(3) Recalculate the centers of mass of the k clusters
(4) Repeat (2) and (3) until the centers of mass of the k clusters do not change
(5) Output the cluster division $\{C_1, C_2 \cdots, C_k\}$

3 RESULTS AND DISCUSSIONS

A. Distribution Patterns Of Building Area and Number of Floors



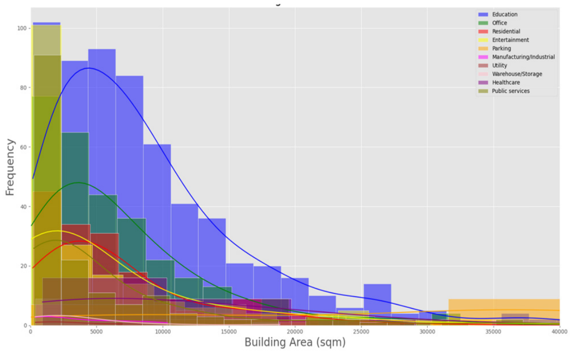


Fig. 1. Probability distributions of floors and areas

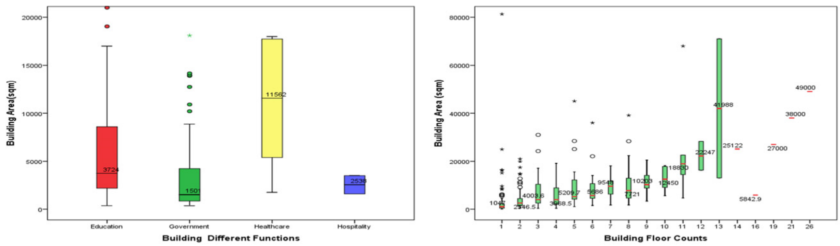


Fig. 2. The relationship between building area and functions, floor counts

Fig. 1. shows the differences in the distribution of area and number of floors among buildings with different functions. Fig. 2. presents the differences in the distribution of area and number of floors for typical buildings. There are substantial differences in the use of buildings based on their intended purposes, hence, when conducting energy consumption analysis, more attention should be given to the functional functions of the buildings.

B. Distribution Patterns of Energy Consumption for Buildings

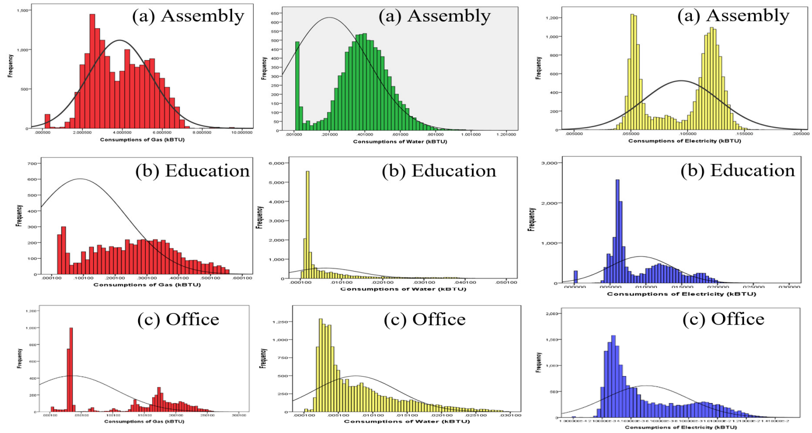


Fig. 3. Distribution of gas, water and electricity consumption in typical buildings

Generally, Fig. 3. shows that the consumption of natural gas and water follows a unimodal distribution, which is more concentrated; whereas electricity consumption exhibits a bimodal distribution, indicating a more dispersed pattern. However, the distribution of natural gas and water consumption in Assembly-type buildings is more scattered, and natural gas consumption in Office-type buildings displays a bimodal distribution. Electricity consumption in Assembly-type, Education-type, and Office-type buildings all show a multimodal distribution.

C. Analysis of Downscaling and Clustering Results

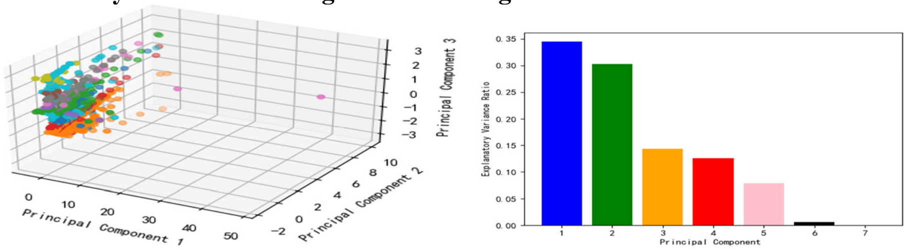


Fig. 4. Explained variance ratio of principal component analysis

From Fig. 4., after dimension reduction using the PCA method, the explained variance ratio is 0.90, meaning that 90% of the data information is retained after removing redundant information.

Table 1. The K-means clustering results of energy consumptions

Centroids					
		Consumptions of Natural Gas		**Building Area(sqm)**	
		Mean	Std. Deviation	Mean	Std. Devia-
Cluster Qual-ity: Fair	1	253921.9411	786011.7235	4705.893	4580.9023
	2	53912.4493	123395.6486	14785.87	16611.08326
	3	1096389.418	2240142.994	3315.344	3226.93889
	4	148655.7569	331345.105	6140.758	4145.06382
	Com-	358661.5133	1171167.065	6577.403	8614.01155
		Consumptions of Water		**Building Area(sqm)**	
Cluster Qual-ity: Good	1	7307757000	10388956557	29305.75	22679.04273
	2	595356161.3	1353893207	4103.937	6313.44295
	3	857798667.6	1839872016	2807.386	2581.49484
	4	253075088.1	895176601.2	8337.142	6248.56101
	5	74084292.06	374355622.4	5218.637	5461.46517
	6	25122256.92	66233682.03	4736.268	4639.44214
	Com-	774751122.1	3257190849	7346.542	9854.09001
		Consumptions of Electricity		**Building Area(sqm)**	
Cluster Qual-ity: Fair	1	485304.1964	629503.6722	9135.602	7900.2196
	2	2708458.832	3024412.684	24554.3	23424.42019
	3	531160.9492	605110.9517	6708.315	6289.50597
	4	482941.2131	655889.5976	4678.708	5244.48306
	5	485145.9491	792169.2076	5907.156	8265.95509
	6	304355.2699	354358.1584	4490.978	5729.23232

	7	483379.6542	552276.5482	6556.283	6408.6428
	Com-	619887.8631	1120177.654	8297.788	10195.70171

Table 1. presents the relationship between building area and energy consumption, highlighting the differences in energy usage among different buildings. Through cluster analysis and by identifying groups of buildings with higher energy consumption, a foundation is provided for energy management and optimization.

D. Correlation Analysis

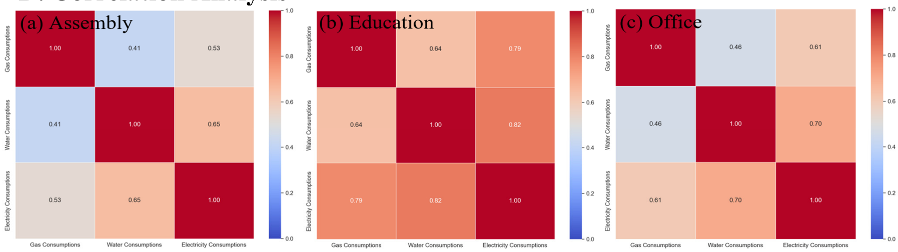


Fig. 5. The correlation relationship between consumptions for buildings with the same function

Fig. 6. The correlation of consumptions between buildings with different functions

Fig. 5. indicates that within buildings of the same type, there is a positive correlation between natural gas consumption and water consumption, natural gas consumption. In terms of the correlation of consumptions between buildings with different functions, Fig. 6. shows that there is a strong association between some types of buildings with specific functions. Overall, grasping the energy usage in buildings necessitates a detailed and the significance of employing data-driven approaches that considers the various building categories and their distinctive energy consumption scenarios.

4 CONCLUSION

The study finds that the function of a building is a significant indicator for evaluating energy consumption. In this dataset, the number of floors for most buildings is concentrated at the lower levels, while the distribution of floors varies for buildings with residential, office, commercial, and healthcare functions. Although building area tends to increase with the number of floors, there is a noticeable variability in the area of high-rise buildings. Overall, there is a positive correlation between building area and energy consumption, meaning the larger the area, the higher the consumption. However, K-means clustering results also indicate a certain degree of dispersion in energy consumption within each building type, with significant differences in energy efficiency and demand characteristics across different types of buildings. Correlation mining also reveals that within the same type of buildings, there is a certain positive correlation between the consumption of natural gas, water, and electricity. Additionally, targeted management measures can be adopted based on the consumption characteristics of different energy types.

Through the analysis of building energy consumption patterns, urban planners can more accurately predict energy demands. Future research may focus on employing data-driven models to process real-time and dynamic data.

ACKNOWLEDGEMENTS

This work is partially supported by the National Key R&D Program of China [grant number 2022YFC3801300].

REFERENCES

1. Chastas, P., Theodosiou, T., & Bikas, D. (2016) Embodied energy in residential buildings-towards the nearly zero energy building: a literature review. *Building and Environment*, *105*, 267-282. https:/doi.org/https://doi.org/10.1016/j.buildenv.2016.05.040
2. Wang, P., Yang, Y., Ji, C., & Huang, L. (2023) Influence of built environment on building energy consumption: a case study in nanjing, china. *Environment Development and Sustainability*, 1-24. https:/doi.org/10.1007/s10668-023-02930-w
3. Robinson, C., Dilkina, B., Hubbs, J., Zhang, W., Guhathakurta, S., Brown, M. A., & Pendyala, R. M. (2017). Machine learning approaches for estimating commercial building energy consumption. *Applied Energy*, *208*, 889-904. https://doi.org/10.1016/j.apenergy.2017.09.060
4. Biswas, M. A. R., Robinson, M. D., & Fumo, N. (2016) Prediction of residential building energy consumption: a neural network approach. *Energy*, *117*, 84-92. https://doi.org/10.1016/j.energy.2016.10.066
5. Ruch, D., Chen, L., Haberl, J. S., & Claridge, D. E. (1993) "A Change-Point Principal Component Analysis (CP/PCA) Method for Predicting Energy Usage in Commercial Buildings: The PCA Model." ASME. J. Sol. Energy Eng, 115(2), 77–84. https://doi.org/10.1115/1.2930035
6. Troccoli, E. B., Cerqueira, A. G., Lemos, J. B., & Holz, M. (2022) "K-means clustering using principal component analysis to automate label organization in multi-attribute seismic facies analysis." Journal of Applied Geophysics, 198, 104555. https://doi.org/10.1016/j.jappgeo.2022.104555