Investigation on air-conditioning energy usage in educational office buildings based on sensing and clustering

Yue Yuan [1], Yixing Chen [1,2,*], Chengcheng Song [1], Yaling He [1],
[1] College of Civil Engineering, Hunan University, Changsha 410082, China
[2] Key Laboratory of Building Safety and Energy Efficiency of Ministry of Education, Hunan University, Changsha 410082, China
*Corresponding Author: yixingchen@hnu.edu.cn

Abstract

To establish a model that can reflect the characteristics and behavior patterns of air conditioning electricity consumption in personnel changing seasons and cooling seasons, and guide the simulation modeling of office buildings. This paper collected occupant behavior and environmental data for three office buildings in Changsha, such as environmental data, air conditioning behavior-related data, electricity consumption data, etc. Firstly, each room is analyzed, and then the daily air conditioning load curve of each type of room is analyzed, and finally, the correlation analysis is carried out.

Highlights

- 21 rooms on the campus have been investigated
- Impact factors are analyzed via random forest
- Clustering method and evaluation method has been adopted to analyze the behavior pattern
- Each type has three different HVAC usage patternsand patterns have diversity

Introduction

Building energy consumption accounts for about 40% of the total energy consumption of the terminal (Zhang et al. 2023). Therefore, to effectively manage the building load, optimizing its energy use structure is of significant significance to help energy conservation and emission reduction and achieve the goal of "zero carbon" (Hong 2012). Building Energy Simulation (BES) has gradually become an important tool for optimizing building structures and evaluating building energy saving potential. However, in the current building simulation software, the human behavior model is often treated with extreme simplification (Yan et al. 2017), which becomes one of the main causes of the difference between simulation and real energy consumption.

Therefore, the depiction of human behavior closer to the real situation helps improve the performance of building simulation. Ding et al. (2022) reviewed the advantages and disadvantages of different sensing and modeling methods in previous Occupancy modeling research. They developed an OB-based energy prediction model (Ding et al. 2019). Li et al. (2020) proposed a cost-sensitivity analysis model for establishing personalized occupant behavior. The indoor environment could be more energy-saving andcomfortable if a more accommodating Occupancy model could be based (Day et al. 2020). O'Brien eta Gunay (2015) introduced the uncertainty Occupancy model into the simulation to estimate energy and use caused by occupancy.

Undoubtedly, the HVAC system is the largest contributor to the energy-saving potential for the application of buildings. Thus, the accuracy OB model for the HVAC system can benefit in achieving an energy efficient building. Many types of research have been modeling occupant behavior on the HVAC system for different building types, such as residential buildings (Yao 2018), office buildings (Obrien et al. 2019), high-rise apartments (Ryu eta Kim 2021) (Stopps eta Touchie 2020), facilities (Mawson eta Hughes 2020), and so on. Stopps et al. (Stopps eta Touchie 2020) investigated the opportunities to improve thermal comfort and energy efficiency by surveying the participants and modeling the relationship. Obrien et al. (2019) developed a data-driven stochastic tenant model to describe individual HVAC occupant behavior in an office building. Jung eta Jazizadeh (2019)- demonstrated that thermal comfort of Occupancy has a statistical significance to the HVAC system. The campus building is a kind of special building type, the campus contains many kinds of building models, but the modeling is complex, and there is no prototype campus model. In order to assist researchers in building a typical campus office building simulation model.

To enhance our understanding of OB diversity and differences in campus office buildings located in hot summer and cold winter areas in China, and to facilitate the simulation modeling of office buildings, it is essential to develop a model that accurately reflects the air conditioning electricity consumption characteristics and staff behavior patterns during the cooling season.

Methodology

Workflow of this study

This study aimed to investigate the air conditioning energy usage of the educational office building in hot summer and cold winter climate zone. A field measurement study is conducted in 21 college office rooms of educational buildings in Changsha, China. The study selected three office buildings in a typical

https://doi.org/10.26868/25222708.2023.1140

comprehensive university in Changsha as research objects, where the energy use behavior of users in the office buildings, which accounted for a significant proportion of the total energy consumption, was investigated. The research subjects were mainly postgraduate students who occupied the buildings. The monitoring content mainly included the time of air conditioning use, frequency of use, and usage habits of each research subject in the office buildings.

As Figure 1 shown, first, this paper investigated the indoor thermal comfort environment and analyzed the distribution of temperature and humidity in 23 rooms, and then, this paper investigated the use of air conditioning. The energy consumption per unit area and unit person were counted separately monthly. Then the influencing factors of energy consumption were analyzed using the random forest method. Random Forest (RF) is adopted to discover influencing key factors of Air-Conditioning energy consumption from outdoor and indoor environment parameters within the obtained period. Finally, a clustering model based on k-means was developed to extract the Air-Conditioning energy use patterns of the 24-hourly daily profile for each room. And The effectiveness of the cluster analysis was evaluated using silhouette width (S_i).

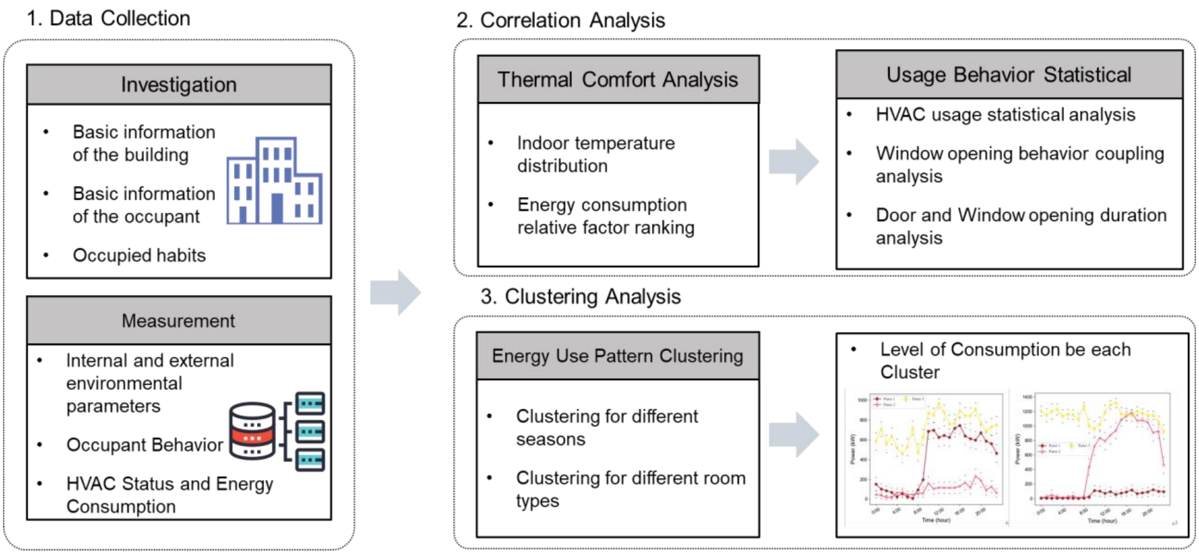


Figure 1: The workflow of this paper

Overview description of the dataset

In this study, sensors were arranged in 21 rooms in four functionally similar office buildings used for research at a university in a hot summer and cold winter region of China. As shown in Table 1, the analysis period covers the whole cooling season from Mar 22nd to Oct 18th, 2022. The rooms range from 8m2 to 43m2 and include public and private offices. This study will analyze the effects of outdoor environmental parameters and occupant behavior on air conditioning energy consumption in rooms of similar size and the same function. The experiment was conducted in three office buildings on the university campus used by graduate students. Therefore, the 85 subjects involved were all graduate students, including 37 males and 48 females involved. There are four single offices and 17 public offices. The per capita use area is between 2.67m2 and 22.75m2, and all offices adopt sash windows.

Through field measurement, investigation and interview, this study collected information about different room sizes, window opening methods, window sizes, window orientation and so on. Through investigation and interview, the basic information about each office staff, including gender and window opening habits, is also collected.

The data collected are divided into four main types: 1. basic building information and personnel information, 2. outdoor weather data, 3. OB-related indoor environmental data, and 4. HVAC-related Occupant behaviour. Sensors collected indoor environmental data, including indoor temperature, relative humidity, indoor light intensity and atmospheric pressure. Personnel data include door and window switch condition, air conditioning operation condition and power consumption, socket power consumption, and personnel movement at the door K- means clustering is popular because it can cluster large amounts of data quickly and efficiently.

https://doi.org/10.26868/25222708.2023.1140

Table 1: Basic Information on investigated rooms

Room#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Building	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B	B	C
Occupancy	5	3	4	7	1	6	1	1	3	1	2	2	3	3	4	4	5	5	3	4	8
Area	42	8	12	54	22.75	54	13.5	13.5	13.5	18	18	18	18	18	18	14	18	31.5	14	18	42.56
Height	4	4	4	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Level	1	1	1	2	2	3	3	3	2	2	2	2	2	2	2	3	3	4	4	4	2
Window orientation	S	N	N	N	N	N	N	N	N	N	N	S	S	S	S	S	S	N	S	S	E
Numbers of Window	2	1	1	1	1	4	2	2	1	2	2	3	2	2	3	1	3	2	1	2	4

Energy use distribution

In this paper, the energy consumption per unit area and per person are calculated respectively. Initially, the energy consumption of air-conditioning per unit area ranged from 11.7 kWh/m2 to 152.2 kWh/m2, with an average value of 68.9 kWh/m2 (SD=37.5 kWh/m2). Per capita, energy consumption for a single room ranged from 46.84 kWh/m2 to 1883.7 kWh/m2, with an average value of 464.7 kWh/m2 (SD = 385.4 kWh/m2).

Figure 2 shows the energy usage distribution with unit area for each room, and Figure 3 shows the energy usage per person for each room. Room 10, which consumes 1883.70 kWh/per person, is a single office. The room's area, orientation and structure are the same as Room 11, and the work schedule and energy consumption pattern are similar. However, due to the small number of people in the room, the per capita energy consumption value is higher than the average level. The highest energy consumption month occurs in July and August due to the summer season significantly reducing energy consumption levels.

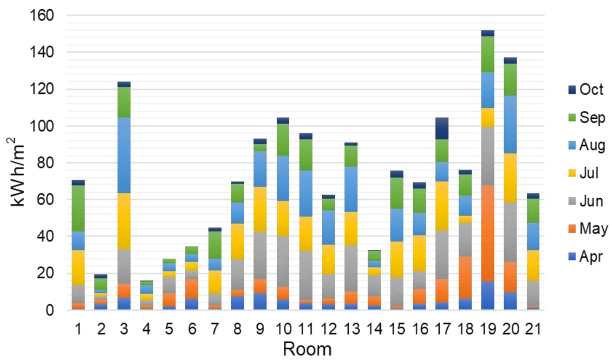


Figure 2: Comparison of average energy consumption of each room capital area

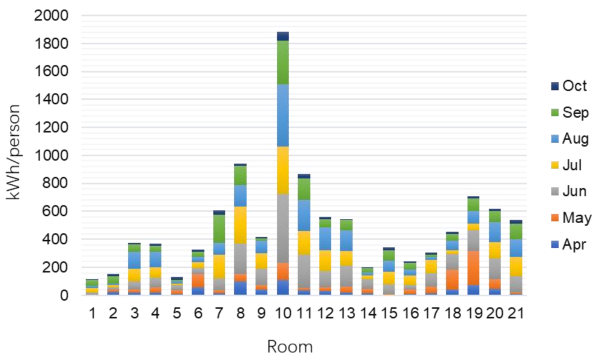


Figure 3: Comparison of average energy consumption of each room per person

The indoor environment analysis

This study collected six types of meteorological data: outdoor temperature, outdoor relative humidity, wind speed, wind direction, wind level and PM 2.5. Among them, the sources of outdoor temperature, outdoor relative humidity, wind speed, wind direction and wind level were meteorological stations set up in the laboratory. The PM2.5 data came from the China Meteorological Data Network. Moreover, occupant-related data is collected by smart sensors, such as indoor temperature and indoor relative humidity.

The indoor and outdoor environments of different types of rooms under long-term monitoring were plotted based on the monitored indoor and outdoor temperature and humidity. During monitoring, the outdoor temperature ranged from 8.5 to 41.5 °C with an average of 25.07 °C (SD = 7.1 °C), and the outdoor relative humidity ranged from 18 to 99.3% with an average of 69.1% (SD = 18.3 %). The indoor air temperature ranged from 21.17 to 29.25 °C with an average of 26.6 °C (SD = 3.4 °C), and the indoor relative humidity ranged from 50.8 to 72.2% with an average of 60.8 % (SD = 11.4 %). Based on the room division and data quality, we show the indoor and outdoor temperature and humidity in Figure 4.

https://doi.org/10.26868/25222708.2023.1140

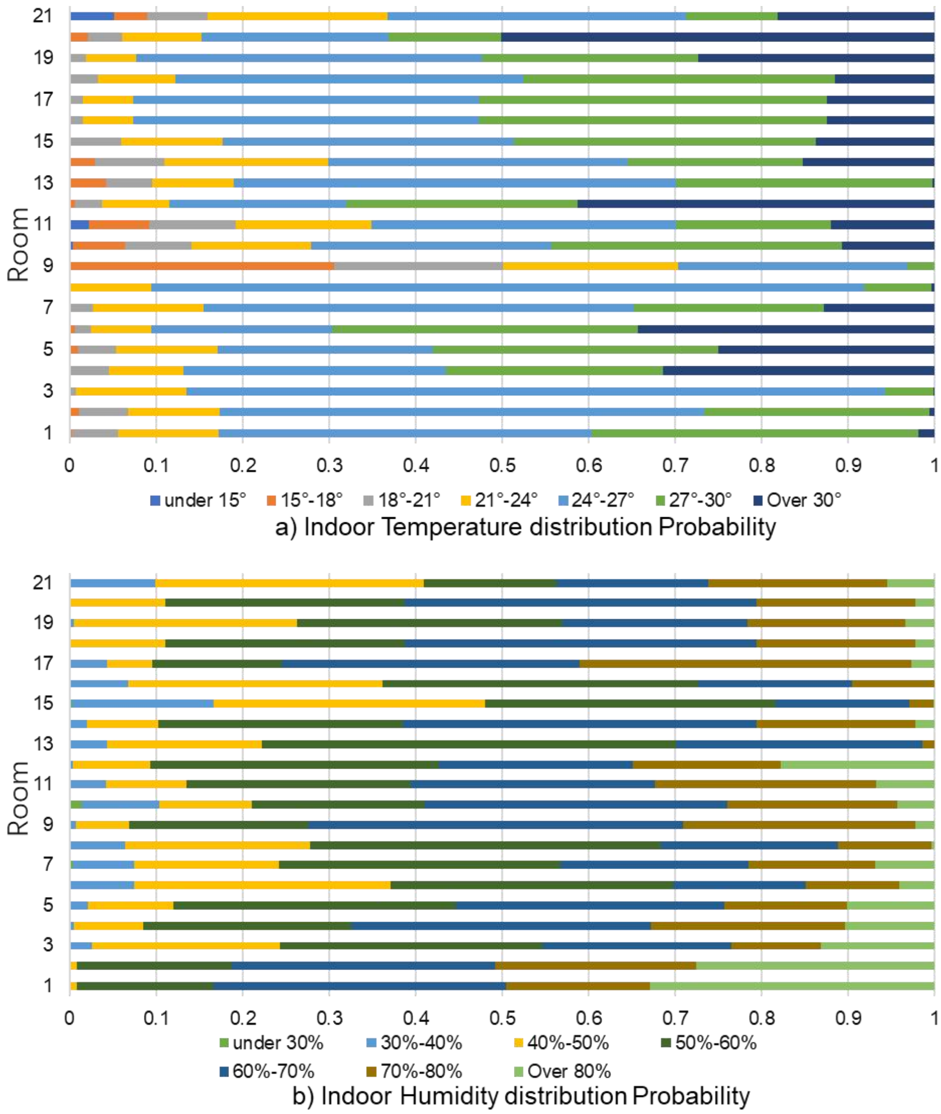


Figure 4: Indoor and outdoor environmental conditions of different types of rooms during the monitoring period

Firstly, a data-driven method is proposed to discover influencing key factors of Air-Conditioning energy consumption within the obtained period. In addition, clustering models were developed to extract the Air-Conditioning OB energy use patterns based on physical observational and long-term monitoring data. The results show that the distribution of OB on air conditioning energy usage patterns varies among offices and determine the regular high- and low-energy consumption periods. It is further recognized that days with similar area and occupancy patterns can have nearly similar energy use patterns. Moreover, this study can provide a capital energy use reference for the educational office building in China’s hot summer and cold winter climate zone for simulation modeling.

Impact factors analysis via Random Forest

In this paper, the random forest algorithm is chosen to rank the indoor and outdoor environmental factors and occupant behavior factors that affect HVAC energy consumption.

Random Forest is a popular artificial intelligence algorithm besides ANN (Wang et al. 2018). Like a forest, its basic principle is an algorithm that integrates multiple trees through the Bagging idea of ensemble learning: its basic unit is the decision tree. Predictions are made by averaging the predictions for each decision tree.

This paper selects the CART regression tree as the weak classifier, and the Bagging algorithm is used for integration. CART regression tree adopts the minimum mean square error (MSE) for error correction; that is, for the data sets D1 and D2 divided into both sides of the corresponding arbitrary partition point s for any partition feature A, the corresponding feature and eigenvalue partition points that minimize the mean square error of D1 and D2 respectively and the sum of the mean square error of D1 and D2 are obtained(Zekić-Sušac et al. 2021). The expression is shown as:

$$\underbrace{min}_{A,s}\left[\underbrace{min}_{c_1}\sum_{x_i\in D_1(A,s)}(y_i-c_1)^2+\underbrace{min}_{c_2}\sum_{x_i\in D_2(A,s)}(y_i-c_2)^2\right] \quad (1)$$

https://doi.org/10.26868/25222708.2023.1140

Clustering method of HVAC energy usage profile

The clustering method of the HVAC energy usage profile is K-means. K-means is popular in energy usage pattern clustering because it can achieve better results with lower algorithmic complexity. Nepal et al. (Nepal et al. 2019) proved that the k-means algorithm could improve accuracy. Yang et al. (Yang et al. 2017) analyzed ten institutional buildings in 3 different typologies with K- means. Accordingly, K-means were adopted in this study.

In cluster analysis, it is generally necessary to determine the optimal number of clusters, namely the k value, through evaluation methods. This paper selects two internal assessment methods to determine the k value: the Silhouette Coefficient (Si) and Calinski-Harbasz Score.

The Silhouette Coefficient method works as follows steps:

1. Calculate the average distance ai between sample i and the cluster's other samples. The smaller ai, the more sample i should be clustered into the cluster. Let ai be the intra-cluster dissimilarity of sample i. The average ai of all samples in a certain cluster C is called the cluster dissimilarity of cluster C.

2. Calculate the average distance bij of all samples from sample i to some other cluster Cj, called the dissimilarity between sample i and cluster Cj. Defined as the inter-cluster dissimilarity of sample i: bi =min{bi1, bi2,..., bik}, the dissimilarity between clusters of a sample is the lowest among the average distance between this sample and all samples of other clusters. The larger bi is, the less sample i belongs to any other cluster.

As the function (2) and (3) are shown, the contour coefficient (Si) is jointly determined by ai and bi. If Si is close to 1, the clustering of sample i is reasonable; if Si is close to -1, the sample should be clustered into another cluster.

$$S(i)=\frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \tag{2}$$

$$S(i)=\begin{cases}1-\dfrac{a(i)}{b(i)},\ a(i)<b(i)\\ 0,a(i)=b(i)\\ \dfrac{b(i)}{a(i)}-1,a(i)>b(i)\end{cases} \tag{3}$$

The Calinski-Harbasz Score is calculated by assessing the inter-class and intra-class variance. Score S is the ratio of inter-cluster dispersion to intra-cluster dispersion and is calculated by evaluating inter-class and intra-class variance. The larger the score, the better the clustering effect.

$$S(K)=\frac{T_r(B_K)}{T_r(W_K)}\times\frac{N-K}{K-1} \tag{4}$$

Where is represent the discrete matrix between groups, W_K is represents the discrete matrix in the group.

RESULTS

Impact factors analysis

The results shown in Figure 5 indicated that the most important factors affecting energy consumption are time sequence, indoor and outdoor temperature and humidity. The next most significant factors are window status, outdoor humidity and wind level, and the next is wind direction.

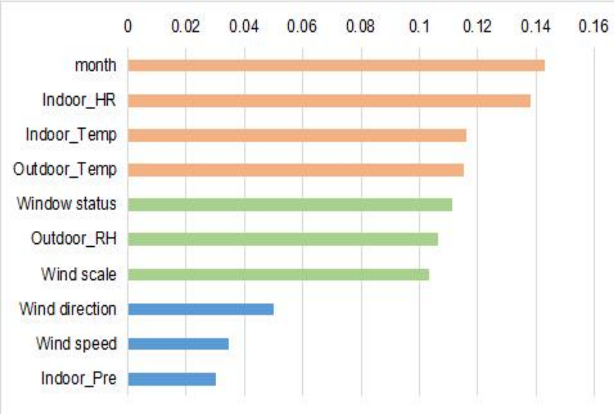


Figure 5: Impact factors rankings

Clustering analysis

Due to the great individual variability of data quality in each room and the different working hours of the personnel. Therefore, we need to cluster the rooms one by one and compare the energy use curves obtained from the clusters one by one.

Firstly, we preliminarily determined the optimal classification cluster as k=3 by hierarchical clustering method, as Figure 6 shown, then conducted k-means for each room one by one, and evaluated the Silhouette Coefficient (Si) and Calinski-Harbasz Score when k=3, 4, 5, respectively.

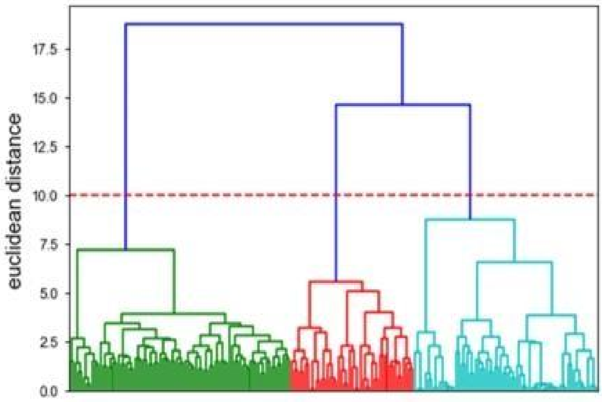


Figure 6: The Euclidean clustering results

As shown in Table 2, the Silhouette Coefficient and Calinski-Harbasz Score are used to compare the clustering result of k-means. When k = 3, the clustering performance is the best. Moreover, Calinski-Harbasz has

https://doi.org/10.26868/25222708.2023.1140

better discrimination for the evaluation.

Due to space limitations, we selected Room #3, Room #8, Room #10 and Room #15 as representatives for presentation according to building, room size, orientation, floor height and other factors. Conversely, Room #3, Room #8, Room #10 and Room #15, with good data quality were selected for comparative analysis. It can be found that the energy consumption of Type 3 is higher than the other three groups through comparison. First, it is more reasonable to cluster into three categories by calculating the Euclidean distance, drawing a clustering dendrogram, and adding auxiliary line markers.

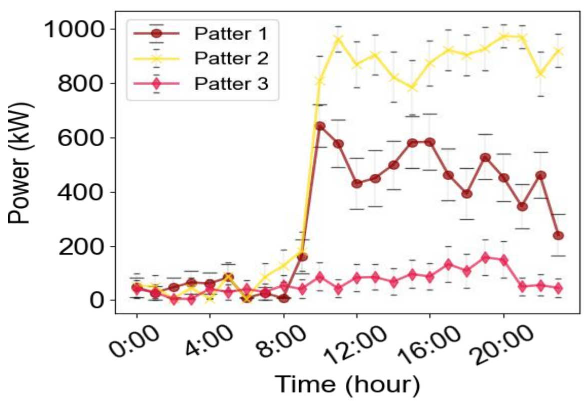
Due to the data quality of each room, there are great individual differences, and personnel work and rest time are different. Therefore, cluster analysis should be done individually to compare the energy use curves obtained by clustering. In the first stage, Room #3, Room #8, Room #10 and Room #15, with good data quality were selected for comparative analysis. By comparison, the energy consumption of Type 3 is higher than that of the other three groups.

Experimental data should be presented with uncertainty/error bounds wherever possible, and a statement of how these bounds were determined should be presented. K-means clustering of the collected energy consumption data shows obvious differences in the energy use patterns of each type of room. But all of them can be classified into high energy consumption mode, medium energy consumption mode and low energy consumption mode.

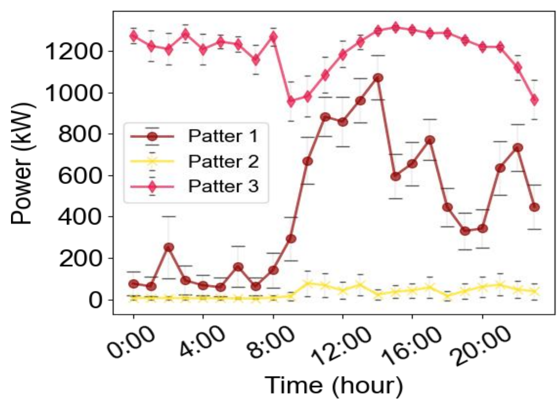
Clustering result evaluation.

By comparing Room #3, Room #8, Room #10 and Room #15, it can be found that rooms with similar areas have similar energy consumption levels. The reason for the difference in August is that the Type1 room has occupied the state during the summer vacation. Hence, its energy consumption in August differs from other occupied rooms. The result has shown that air conditioning energy usage patterns vary among offices and determine the regular three patterns as high-, median- and low-energy consumption periods.

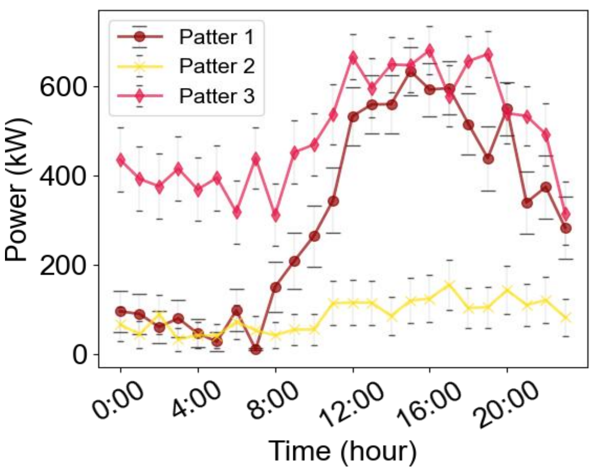
The clustering algorithm is then applied to the energy consumption dataset of 3 types for daily-base pattern identification of each room. It is observed that the clustering result shows a significant diversity. These observation results could be utilized as a reference for building energy simulation



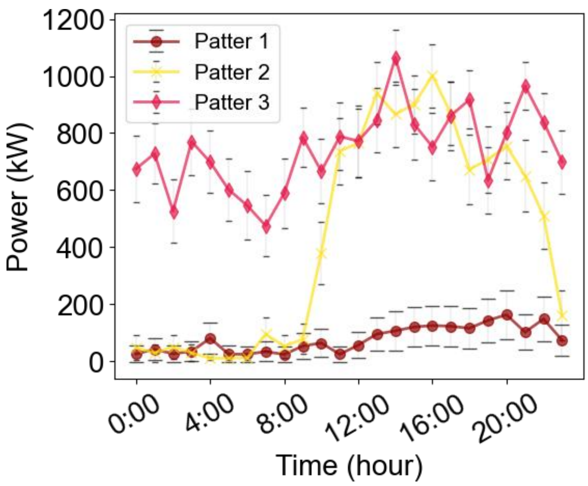
a) Room#3



b) Room #8



c) Room #10

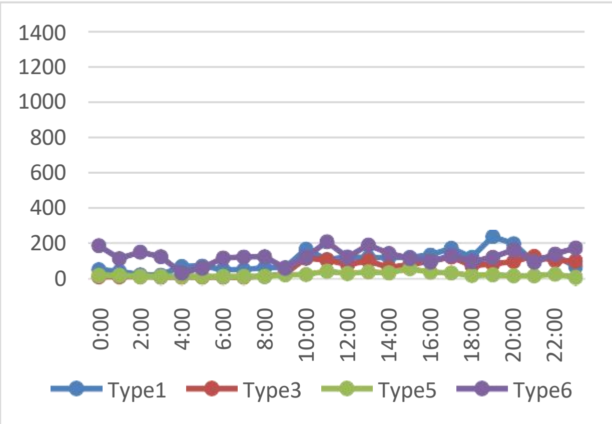


d) Room #15

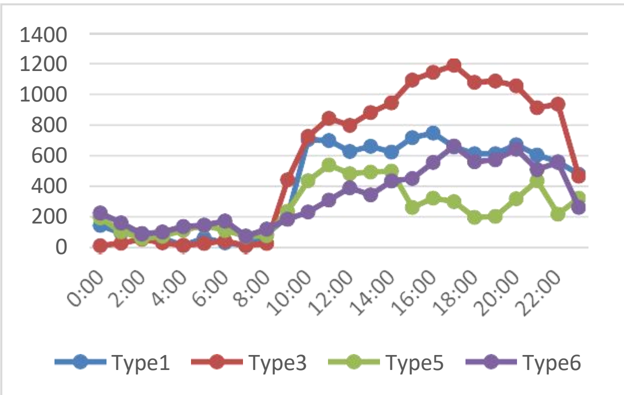
Figure 7: Clustering for different types

https://doi.org/10.26868/25222708.2023.1140

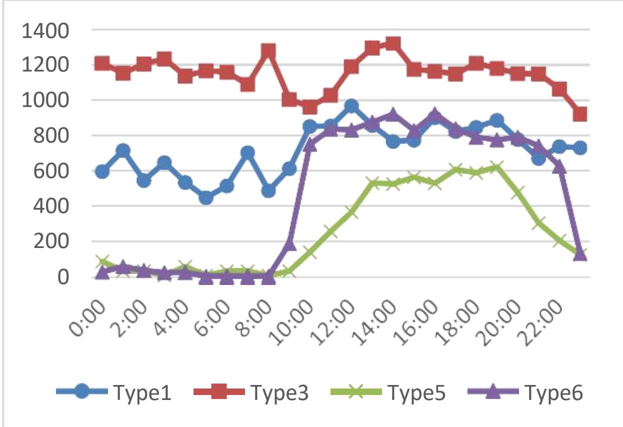
After clustering, the three energy consumption patterns of high, median- and low-low energy were extracted for analysis. In Figure 8, all three patterns for each room are compared. It was found that large differences emerged among the three different types of rooms. Comparing Room #3, Room #8, Room #10, and Room #15 shows that rooms with similar areas have similar energy consumption levels. The reason for the difference in August is that Room #3 has occupied the state during the summer vacation. Hence, its energy consumption in August differs from other occupied rooms. The average energy consumption in June and July is relatively high among them. This is because many rooms are idle during the summer vacation in August.



a) Energy Use Pattern 1



b) Energy Use Pattern 2



c) Energy Use Pattern 3

Figure 8: Average energy consumption of each room

DISCUSSIONS

This study aimed to investigate the HVAC energy usage patterns of four university office buildings. Initially, statistical analysis was conducted to examine the indoor thermal comfort and overall energy usage conditions of each room. Subsequently, a random forest approach was employed to rank the impact factors of HVAC energy usage. Furthermore, a K-means clustering method was utilized to analyze the electricity usage patterns of the university buildings. To validate the clustering result, 2 Cluster Validity Indices are calculated.

The conclusions are as follows:

1. This paper investigates the energy consumption level of the Campus office Building, providing a certain simulation reference. The energy consumption of air-conditioning per unit area ranged from 11.7 kWh/m2 to 152.2 kWh/m2, with an average value of 68.9 kWh/m2 (SD=37.5 kWh/m2). Per capita, energy consumption for a single room ranged from 46.84 kWh/m2 to 1883.7 kWh/m2, with an average value of 464.7 kWh/m2 (SD = 385.4 kWh/m2). This provides a practical reference for the later simulation and the establishment of typical buildings.

2. Although the energy use patterns of each room can be classified into three categories, there exists a significant variation in the energy use patterns within the same category for each room.

Acknowledgment

This paper is supported by the National Natural Science Foundation of China (NSFC) through Grant No. 51908204 and the Natural Science Foundation of Hunan Province of China through Grant No. 2020JJ3008.

References

Day JK, McIlvennie C, Brackley C, et al (2020) A review of select human-building interfaces and their relationship to human behavior, energy use and occupant comfort. Build Environ 178:106920. https://doi.org/10.1016/j.buildenv.2020.106920

Ding Y, Han S, Tian Z, et al (2022) Review on occupancy detection and prediction in building simulation. Build Simul 15:333–356. https://doi.org/10.1007/s12273-021-0813-8

Ding Y, Wang Q, Wang Z, et al (2019) An occupancy-based model for building electricity consumption prediction: A case study of three campus buildings in Tianjin. Energy Build 202:109412. https://doi.org/10.1016/j.enbuild.2019.109412

Hong T (2012) Occupant Behavior: impact on energy use of private offices. ASim 2012 - 1st Asia Conf Int Build Perform Simul Assoc

Jung W, Jazizadeh F (2019) Comparative assessment of HVAC control strategies using personal thermal comfort and sensitivity models. Build Environ 158:104–119. https://doi.org/10.1016/j.buildenv.2019.04.043

Li Z, Zhu H, Ding Y, et al (2020) Establishment of a personalized occupant behavior identification model for occupant-centric buildings by considering cost sensitivity. Energy Build 225:110300. https://doi.org/10.1016/j.enbuild.2020.110300

Mawson VJ, Hughes BR (2020) Thermal modelling of manufacturing processes and HVAC systems. Energy 204:117984. https://doi.org/10.1016/j.energy.2020.117984

Nepal B, Yamaha M, Sahashi H, Yokoe A (2019) Analysis of building electricity use pattern using k-means clustering algorithm by determination of better initial centroids and number of clusters. Energies 12:. https://doi.org/10.3390/en12122451

O'Brien W, Gunay HB (2015) Mitigating office performance uncertainty of occupant use of window blinds and lighting using robust design. Build Simul 8:621–636. https://doi.org/10.1007/s12273-015-0239-2

Obrien W, Abdelalim A, Gunay HB (2019) Development of an office tenant electricity use model and its application for right-sizing HVAC equipment. J Build Perform Simul 12:37–55. https://doi.org/10.1080/19401493.2018.1463394

Ryu J, Kim J (2021) Effect of different hvac control strategies on thermal comfort and adaptive behavior in high-rise apartments. Sustain 13:1–20. https://doi.org/10.3390/su132111767

Stopps H, Touchie MF (2020) Managing thermal comfort in contemporary high-rise residential buildings: Using smart thermostats and surveys to identify energy efficiency and comfort opportunities. Build Environ 173:106748. https://doi.org/10.1016/j.buildenv.2020.106748

Wang Z, Wang Y, Zeng R, et al (2018) Random Forest based hourly building energy prediction. Energy Build 171:11–25. https://doi.org/10.1016/j.enbuild.2018.04.008

Yan D, Hong T, Dong B, et al (2017) IEA EBC Annex 66: Definition and simulation of occupant behavior in buildings. Energy Build 156:258–270. https://doi.org/10.1016/j.enbuild.2017.09.084

Yang J, Ning C, Deb C, et al (2017) k-Shape clustering algorithm for building energy usage patterns analysis and forecasting model accuracy improvement. Energy Build 146:27–37. https://doi.org/10.1016/j.enbuild.2017.03.071

Yao J (2018) Modelling and simulating occupant behaviour on air conditioning in residential buildings. Energy Build 175:1–10. https://doi.org/10.1016/j.enbuild.2018.07.013

Zekić-Sušac M, Has A, Knežević M (2021) Predicting energy cost of public buildings by artificial neural networks, CART, and random forest. Neurocomputing 439:223–233.

https://doi.org/10.1016/j.neucom.2020.01.124

Zhang Y, Teoh BK, Wu M, et al (2023) Data-driven estimation of building energy consumption and GHG emissions using explainable artificial intelligence. Energy 262:125468. https://doi.org/10.1016/j.energy.2022.125468