Energy Engineering

Tech Science Press

DOI: 10.32604/ee.2022.019654

ARTICLE

Research on Electricity Consumption Model of Library Building Based on Data Mining

Jiaming Dou[1], Hongyan Ma[1,2,3,*] and Rong Guo[1]

[1]School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing, 102616, China

[2]Institute of Distributed Energy Storage Safety Big Data, Beijing University of Civil Engineering and Architecture, Beijing, 102616, China

[3]Beijing Key Laboratory of Intelligent Processing for Building Big Data, Beijing, 102616, China

[*]Corresponding Author: Hongyan Ma. Email: mahongyan@bucea.edu.cn

Received: 06 October 2021 Accepted: 21 December 2021

ABSTRACT

With the exponential development of Chinese population, the massive energy consumption of buildings has recently become an interest subject. Although much research has been conducted on residential buildings, heating ventilation and air conditioning (HVAC), little research has been conducted on the relationship between student's behavior, campus buildings, and their subsystems. Using classical seasonal decomposition, hierarchical clustering, and apriori algorithm, this paper aims to provide an empirical model for consumption data in campus library. Smart meter data from a library in Beijing, China, is adopted in this paper. Building electricity consumption patterns are investigated on an hourly/daily/monthly basis. According to the monthly analysis, electricity consumption peaks each year around June and December due to teaching programs, social exams, and outdoor temperatures. Hourly data analysis revealed a relatively stable consumption pattern. It shows three different types of daily load profiles. Daily data analysis demonstrated a high relationship between HVAC consumption and building total consumption, with a lift value of 5.9. Furthermore, links between temperature and subsystems were also discovered. Through a case study of library, this study provides a unique insight into campus electricity use. The results could help to develop operational strategies for campus facilities.

KEYWORDS

Electricity consumption; data mining; load profile; campus building

1 Introduction

Energy consumption is becoming recognized as a critical environmental issue with serious social and economic consequences [1]. The buildings consume the most energy and emits the most greenhouse gases [2]. According to International Energy Agency (IEA) statistics [3], the buildings and their construction sectors account for more than one-third of global energy consumption and approximately 40% of total CO_2 emissions. Statistics from the United Nations Environment Programme (UNEP) also indicate that residential and commercial buildings consume more than 40% of energy [4,5], with HVAC systems contributing more than 40% of building energy [6,7]. As a result, governments have adopted a number of measures to monitor the rise of primary energy demand and CO_2 emissions in response to the trend of energy consumption [8]. China has placed steps in place to peak its CO_2 emissions around 2030, and it is pushing hard to reach this target faster [9,10]. Building energy conservation is both necessary and urgent. The most promising approaches to construction technology research concentrate on improving building energy efficiency [11,12].

The exponential growth in building consumption puts energy shortages and environmental sustainability at risk. Balance of supply and demand grid has been a more complicated problem [13]. To overcome this issue, there is a lot of interest in grid demand side control [14,15], and renewable energy generation [16,17]. On the one hand, renewable energy development is conducive to energy saving and emission reduction. Renewable energy has a robust stochastic character, which differs from conventional power systems. In order to be able to integrate renewable energy with the grid, it is necessary to align the energy supply side and the consumption side in terms of load characteristic patterns. In that case, it will be very beneficial to develop the power grid. On the other hand, the advancement of sensing and measuring technologies, network connectivity technologies, and intelligent control technologies has created new possibilities for power grid demand management [18].

Demand side management is regarded as a less capital-intensive system balance approach [13,19,20]. Among it, the idea of building energy management systems has piqued the interest of researchers for a long time [14]. The data collection layer, analytics layer, and application layer comprise the building energy management system. The data analytics layer is at the heart of the whole structure [21]. Nowadays, smart meters are also commonly used in residential buildings. These meters have measuring and coordination features, allowing them to monitor consumers' fine-grained energy consumption in vast amounts [22]. Data of various types/steps is itemized and accumulated in vast amounts. This opens up the possibility of implementing data mining and interpretation at the data analytics layer. Advanced data analysis-based decision support is becoming increasingly relevant in the growth, application, and management of smart energy systems [23,24].

Different data scales have different analysis methods and application potential. Analyzing hourly data can help to understand daily load variation patterns and manage staggered electricity consumption. Data of daily steps can be used to discover correlations between energy-consuming systems and provide guidance for fuel supply and even equipment maintenance. Monthly data can be used for annual load judgment and long-term planning. There are various approaches to meet the needs of various scales and volumes. Building energy management systems may use this data to assess the energy usage of each subsystem, consider load characteristics [25,26], and minimize excessive energy loss [27].

However, because user behavior is complex and stochastic, it will take a long time to fully understand how end-users apply energy on the demand side [2]. Traditional analysis methods [28], such as the degree-day [29], are inefficient and difficult to uncover hidden information behind the data. Their analysis requires a high level of professional knowledge and is mostly only suitable for professionals. With the advent of big data in recent years, scholars have proposed a variety of data mining models that can be applied to energy management systems. These models are capable of identifying relevant characteristics of consumer behavior [22]. For example, some studies have proposed methods for predicting the future energy consumption generated by buildings [30,31].

Clustering is the most widely used method for analyzing building energy now. It divides customers into groups based on their actual electricity usage. Objects from one cluster are more like one another than objects from other clusters [22]. Many clustering methods have been proposed in previous studies [1,32], including k-means [32,33], fuzzy k-means [34,35], hierarchical algorithms [22,36], SOM [37,38], probabilistic and generative models [39,40], DBSCAN [41,42], and k-shapes [43]. k-means and hierarchical clustering are the two most common clustering approaches used to construct consumption data [44]. Nonetheless, k-means have some inherent drawbacks that may have an impact on the performance of electricity usage pattern mining. For example, clustering centers are typically randomly initialized prior to clustering, causing clustering results to easily fall into local optima and affecting the algorithm's efficiency [45]. As a result, hierarchical clustering is used as the clustering method in this paper.

Furthermore, association rule mining is a data-driven method for analyzing energy consumption. It can discover interesting relationships in large amounts of data. In recent years, association rule mining has been widely used in building equipment parameter control [46] and occupant profiling [47]. Existing research has concentrated on the relationship between HVAC equipment parameters and less on the relationship between the building and subsystems, as well as between different subsystems. The apriori algorithm is a straightforward and easy-to-understand association rule mining algorithm. It is the algorithm of choice for rule mining in this paper since it has low data requirements.

For buildings with different functions, most studies have been done on residential and commercial buildings. Few scholars have studied libraries in campus. Some studies have shown that changes in occupant behavior during the building life cycle have a great potential for energy savings [2]. Different occupant needs and behaviors require specific technological solutions, which may induce or change behavioral patterns, and occupant behavior can influence technology adaptation and implementation [10,48,49]. The occupants of library are mainly students. Students, in general, are an important group of electricity users who should not be overlooked. High frequency, high volume, and dispersion characterize student electricity consumption. The analysis of students' electricity usage behavior can help with demand-side control, grouping of electricity customers, and tariff setting. In the future, student electricity use will be included in the demand-side regulation of the smart grid to understand the intelligence and low carbon of campus electricity use.

By investigating the electricity usage of a campus library building, this paper aims to fill a research void in library building energy consumption study. Via a case study in Beijing, China, this paper employs seasonal decomposition, hierarchical clustering, and apriori algorithm to explain building electricity consumption patterns. This thesis describes the dynamic electricity usage behavior of a library in a meaningful way. Based on the results, it makes recommendations for various stakeholders. The remainder of the paper is organized as follows. The study's methodology is described in Chapter 2. In Chapter 3, a case study of a campus building is conducted. Chapter 4 summarizes and concludes the research.

2 Methodology

Technical roadmap in this paper is depicted in Fig. 1. First, the gathered data were filtered and categorized according to the students' living habits. Second, the monthly/daily/hourly data is analyzed using seasonal decomposition, the apriori algorithm, and hierarchical clustering.

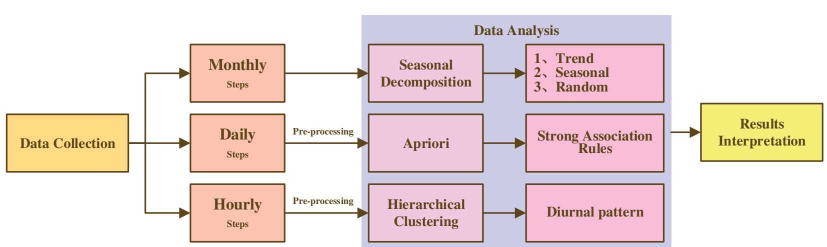

Figure 1: Technical roadmap

2.1 Seasonal Decomposition

Since seasonal decomposition requires a breakdown of seasonal time series. The collected monthly data shows seasonal patterns. So, we use seasonal decomposition for analyzing monthly data. In time series, three categories of variables may be statistically decomposed as seasonal factors exist:

- Trend factor: a long-term upward or downward trend in the research subject.
- Seasonal factor: variations in cycles over the course of a year, typically pertaining to seasons.
- Stochastic factor: fluctuations caused by random events that cannot be explained by trend or seasonal impacts.

Seasonal decomposition is divided into additive and multiplicative models for the decomposition of time series. If the components are independent of each other, it is appropriate to use the additive model [50]. In this paper, three components are independent of each other, so the additive model is used to decompose the time series of electricity consumption. The sum of each factor in the additive model corresponds to the electricity consumption of the corresponding period, which can be described in Eq. (1).

$$Y_t = T_t + S_t + R_t \tag{1}$$

In Eq. (1), Y_t denotes the observed value at the t time; T_t denotes the value of trend factor at that moment; S_t denotes the seasonal factor value at that moment; R_t denotes the random factor value.

Locally weighted regression smoothing (LOESS) is a common method for decomposing series into trend, seasonal, and stochastic terms. In regression analysis, LOESS smoothing is a common tool for creating a smoothed line in a time plot or scatter plot. It can beneficial to understand the relationship between variables and the predicted trend.

2.2 Hierarchical Clustering

Representative daily load profiles will facilitate effective building energy system scheduling, fault detection and diagnosis [51]. Clustering is a commonly used method to obtain load curves, which can analyze data from a given sample or individual by classifying them based on the similarity or distance of their features. The data is clustered into different classes, so that the objects in the same class are very similar, and the objects in different classes are quite different. Hierarchical clustering assumes a hierarchical structure between classes and that samples are clustered into hierarchical classes. Since each sample belongs to only one class, hierarchical clustering is a type of hard clustering. There are two different methods, namely aggregated clustering (bottom-up) and split clustering (top-down). The method used in this paper is aggregate clustering, which is a bottom-up clustering method, also known as systematic clustering. The main flow of the hierarchical clustering algorithm is shown in Fig. 2.

The general process is as follows:

1. Treat each sample as a category;
2. Merge the smallest pair of samples into a new category;
3. Repeat the operation until the stop condition is satisfied.

It consists of three main points: calculation of the distance or similarity, merge rule, stopping condition. For example, the distance or similarity can be an absolute distance, a Euclidean distance, a Mink's distance, or a Chebyshev's distance, etc. The stopping condition can reach a threshold, reach a depth per node, or reach a set maximum number of clusters. When constructing daily load curves, data at scales of hours or smaller needs to be clustered. Otherwise, the curve is not well characterized. Therefore, data at hourly scales is used for clustering.

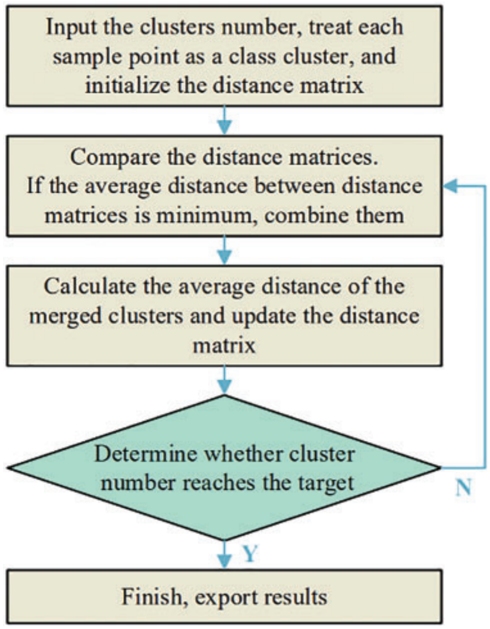

Figure 2: Cluster analysis flow chart

2.3 Apriori Algorithm

In this paper, apriori algorithm will be used to analyze the relationship between weather temperature and energy consumption of building subsystems from daily data. The reason for not using the hourly step data is that the hourly data is too random and unstable.

Apriori is a type of association rule algorithm, the basic principle of which is to iterate through the relationships of items by using an iterative method. The apriori algorithm is used to study the relationship between weather temperature and energy consumption of subsystems in the building. The steps are to use an algorithm to scan the database multiple times, generate frequent sets using candidate frequent sets each time, and then investigate the association rules between the frequent itemsets by calculating the support, confidence, and lift between the frequent items. The support parameter represents the likelihood of the association rule being present in all itemsets. The confidence is the probability that the correlated result will occur under the condition that the precondition is present. The lift is an indicator of whether the prerequisites and the associated results are independent of each other. The higher the lift, the more correlated the two are.

The flow chart to process daily scale data is shown in Fig. 3. First, divide the electricity data and temperature into groups to reduce the number of continuous attribute values. Then, hierarchical clustering is used to discretize the continuous data and the actual data values are replaced by the labels of the categories. The generalized dataset is obtained. In the preceding example, a generalized dataset is fed into an apriori program to generate strong association rules. Finally, energy consumption is analyzed based on these strong association rules.

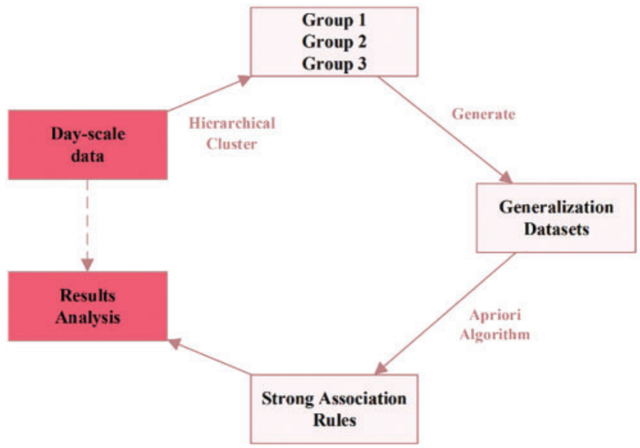

Figure 3: Apriori algorithm flow chart

3 Case Study

Building consumption generally covers lighting & socket consumption, HVAC consumption, special electricity consumption and other consumption. The data in this paper was gathered from a library in Beijing. The library is located at 39.75 degrees north latitude and 116.29 degrees east longitude, a typical temperate continental climate. The library is a square-shaped building, 32 meters high, 70 meters long and wide, 7 floors above ground and 1 floor underground, with a total construction area of 38,580 square meters. Its structure is reinforced concrete frame-steel support system, central patio, inner and outer annular space, glass curtain wall, external hanging diamond-shaped glass fiber reinforced concrete panel exterior wall decoration.

The data covers two periods. The first period data is the monthly data from January 2017 to August 2019. It is made up of three parts: overall library consumption, lighting and socket consumption, and HVAC consumption. The second is the hourly usage from August 01 to October 28, 2020, and is divided into four parts: overall building consumption, lighting and socket consumption, HVAC consumption, and special electricity consumption. The whole building is centrally air-conditioned and centrally heated by electricity. The HVAC power consumption includes both fans and heating and cooling. The heat source of the heating is the power plant. And the heat source of the cooling is chilled water. The HVAC power consumption includes both fans and heating and cooling.

The data are filtered and sorted manually. The collection, transmission, and recording of energy data can be subject to disturbances or faults in the power system, resulting in anomalies or missing data. Also, the regularity of load curves can change, affecting data mining's potential due to vacation and weekends' specific nature. Therefore, it is essential to pre-process the raw load data before cluster and correlation. Data pre-processing's main task is to complete the missing data and clean up insufficient data, such as deleting the electricity data for unknown reasons without a reliable modification basis [52].

Monthly electricity consumption data from January 2017 to August 2019 is relatively neat and clean and does not require cleaning. Hourly data from 01 August to 28 October 2020, including more special cases (Table 1), such as library opening and closing, needs to be filtered and sorted. The data are checked to exclude holiday, transition period and night retreats. The abnormal data for the above dates have been excluded from this paper as they do not reflect the library's real electricity consumption patterns. After excluding the data mentioned above, we selected two periods of time for the analysis: from 17 September to 29 September and 01 October to 28 October.

Table 1: Building operations affair

Whether to open?	Opening times	Event
Closed (From 01 August to 30 August)	None	Summer Vacation
Opening (From 31 August to 28 October)	From 31 August to 13 September 8:00∼21:00	Students Not Yet Fully Back to Campus
	From 14 September to 16 September 8:00∼16:00	Library Maintenance
	From 17 September to 29 September 8:00∼21:00	Normal
	From 30 September to 08 October 8:00∼17:00	Library Closed on National Day
	From 09 October to 28 October 8:00∼22:00	Normal

Fig. 4 shows the total consumption of the library for the 32 months from January 2017 to August 2019. A visual determination of the library's monthly electricity consumption data shows an overall cyclical trend, but with substantial variations from month to month. The summation of the 32 months consumption was 5,005,591.13 kWh, with an average monthly consumption of 156,424.72 kWh. The peak periods were May, June, September in 2017, and May, June, July, August in 2018. The months of February 2017, August 2017, February 2018, February 2019 are the summer and winter vacation when most students will be leaving campus and home, so electricity consumption during these periods is significantly lower.

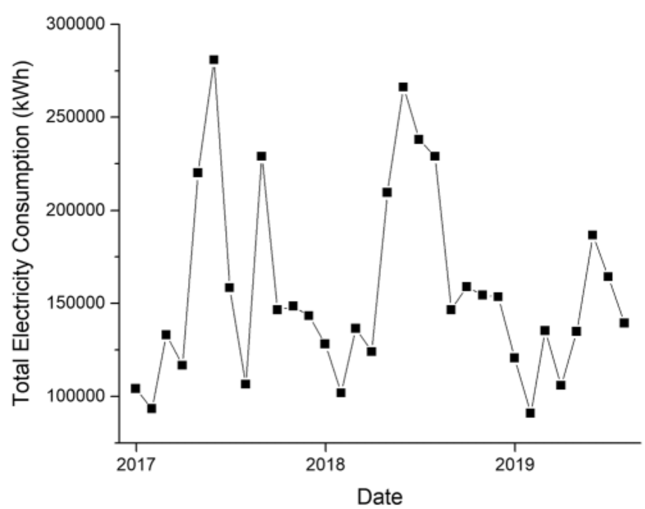

Figure 4: Monthly library electricity consumption

Figs. 5 and 6 show the monthly consumption of lighting & sockets and HVAC. Average monthly power consumption of the lighting & sockets is 86277.78 kWh, accounting for 36.08%–69.28%, with an average of 55.30%. HVAC consumption is 20443.38 kWh, accounting for 1.60%–31.11%, with an average of 11.13%. From the changing trend, lighting, sockets, and HVAC electricity consumption have the following characteristics.

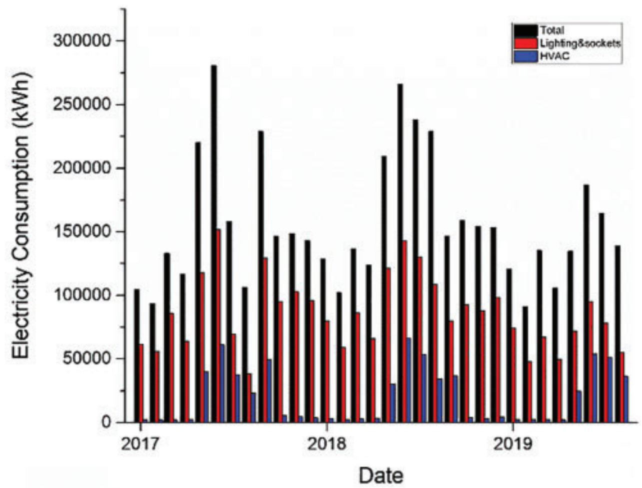

Figure 5: Proportion of electricity used for lighting & sockets and HVAC

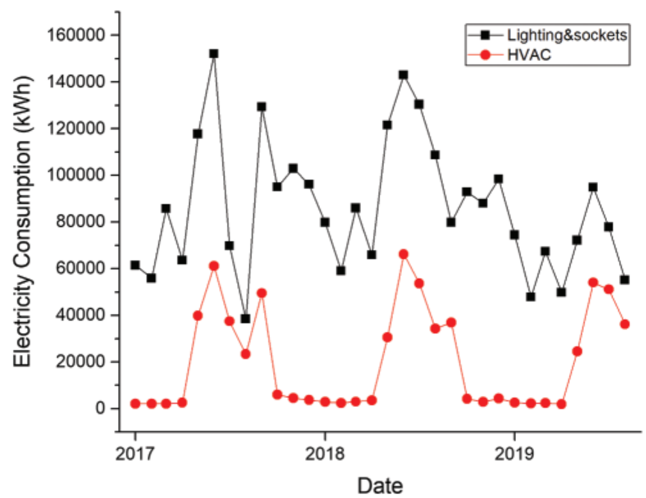

Figure 6: Comparison of electricity consumption for lighting & sockets and HVAC

Electricity consumption from lighting and sockets varies considerably but is not significantly affected by the season. There is a certain regularity in HVAC consumption, greatly influenced by seasonality, with electricity consumption mainly concentrated in the five summer months of May, June, July, August, and September. It seems that the teaching schedule and social examinations influence the total energy consumption, such as the postgraduate entrance examination.

3.1 Seasonal Decomposition

We use the LOESS smoothing to do seasonal decomposition of lighting & sockets consumption (Fig. 7). There is an apparent increase in the series and then a decrease in the trend series, but this is apparent in magnitude. Based on the investigation, it was found that the library was under renovation during summer 2018. Almost all lamps are open during this period, making the power consumption of lighting and outlets slightly higher than the same period in 2017. Seasonal factor sequence presents a noticeable annual variation trend. It reaches a local maximum in March, June, September, and December. It can be explained that March and September are the first half of the semester and the second half of the semester by students' time, respectively. This time may be mainly to return books to the library and prepare for the exam. June and December are the final exams of the next semester and the review time of the postgraduate entrance examination every year. During this period, the sockets & lighting consume more electricity.

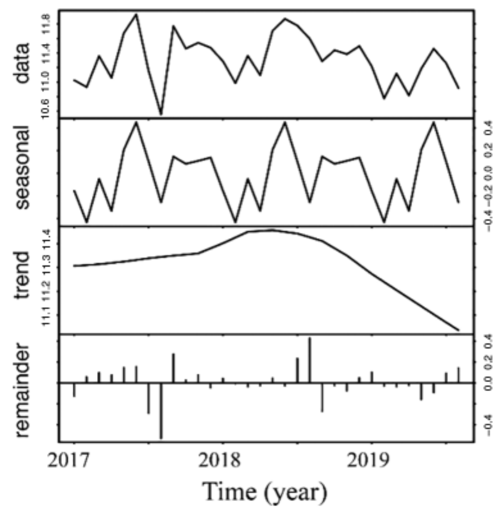

Figure 7: Seasonal decomposition of monthly lighting & sockets consumption

We use the same method to analyze HVAC consumption (Fig. 8). It can be found in the decomposition results of the HVAC consumption series. The trend series has a small process of first increasing and then decreasing, but always remains in the same order of magnitude 9–9.20. The seasonal factor sequence showed a noticeable trend in the annual cycle. HVAC consumption is high from May to September every year. Fig. 9 is the comparison of normalized outdoor average maximum temperature and normalized average maximum and minimum temperature with HVAC power consumption. It can be found that the changing trend of temperature is very similar to that of HVAC consumption. Due to air-condition opening in May and June, the building's power consumption has significantly increased. The random fluctuation series showed a significant random decline from July to August 2017, but there was no significant random decline from July to August 2018, indicating certain special events at that time.

Based on the above analysis, it can be seen that:

Winter vacation and summer vacation are the two lowest consumption periods of the year. The monthly consumption during winter vacation is lower than that during summer vacation. There are still some students studying in campus during summer vacation, while most students go home for Chinese New Year during winter vacation.

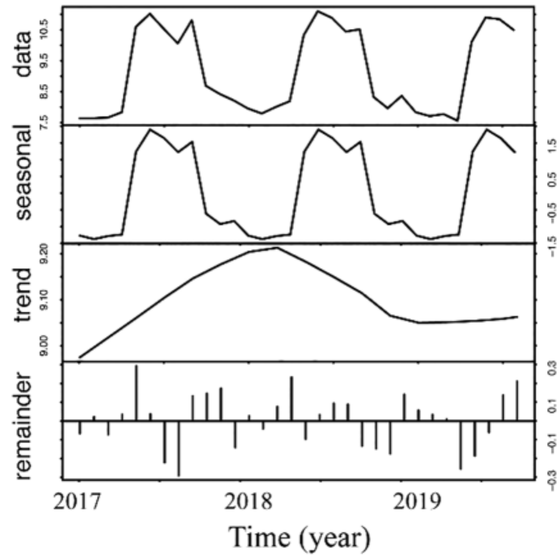

Figure 8: Seasonal decomposition of monthly lighting & sockets consumption

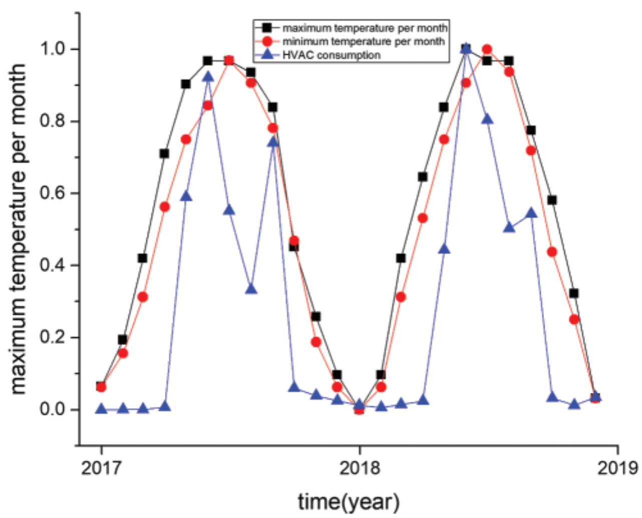

Figure 9: Normalized maximum temperature, minimum temperature, and HVAC consumption

For the total electricity consumption of buildings, the maximum occurs in May, June, and July each year. The electricity consumption in June was more than 60% of the average monthly consumption. It can be explained by teaching schedule. The end of the semester is May and June each year, so the students often begin to review their lessons at this time, and the amount of electricity used for lighting & sockets has increased significantly. Also, according to climate change, the weather gradually turns hot in June and July, and the library turns on air-conditioning and refrigeration, and the amount of electricity used for air-conditioning has increased significantly. Therefore, under the double influence of the two factors, the library consumption peak appeared in May and June.

The above analysis will assist in the development of a long-term campus electricity consumption plan, which should focus on increasing the reserve capacity for the library in May, June, and July each year to cope with the higher electricity load during those months.

3.2 Typical Daily Load Curves Based on Hierarchical Clustering

This part describes the typical characteristics of consumption on working days. Preliminary treatment is required because of the variables involved, such as student opening, including: removal of rest days, retention of working days; removal of select dates with varying opening times.

After deducting some random factors, this paper selected the working days from September 17 to September 29 and October 09 to October 28 as the analysis objects. Table 2 shows partial electricity consumption data. The amount of data is 24, and each data has 24 dimensions, corresponding to 24 h of electricity consumption. This paper only considered the load variation during the library working day. Fig. 10 is the working day load curve for the entire library. We have used silhouette coefficient method [53] to filter the optimal number of clusters. The silhouette coefficients were calculated separately for the number of clusters from 2 to 10. The trend of the silhouette coefficient with the number of clusters is shown in Fig. 11. The closer the silhouette coefficient is to 1, the more reasonable and rational the sample clustering result is. Therefore, the optimal number of clusters is four.

Table 2: Partial electricity consumption data

No.	Dates	Day	Hourly electricity consumption (kWh)				
			00:00∼01:00	01:00∼02:00	......	22:00∼23:00	23:00∼00:00
1	09/17	Thursday	121.5625	111.25	......	125.5937	136.8125
2	09/18	Friday	123.2188	135.2187	......	126.4375	117.5625
3	09/21	Monday	90.375	100.8125	......	114.375	103.25
4	09/22	Tuesday	91.1875	103.1875	......	115.25	114.375
5	09/23	Wednesday	102.4375	111.9688	......	54.4063	45.5937
6	09/24	Thursday	52	51.1875	......	100.0313	99.1875
7	09/25	Friday	89.625	98.375	......	103.2188	90.3437
8	09/28	Monday	99.2188	91.1562	......	103.1562	100.0313
9	09/29	Tuesday	99.2187	88.7813	......	104	99.9688
10	10/09	Friday	88.8438	79.1562	......	105.625	80
11	10/10	Saturday	91.9375	92.0625	......	84.0625	82.375
12	10/12	Monday	85.625	76	......	93.5625	88.8125
13	10/13	Tuesday	87.1875	78.375	......	100.0625	79.9375
14	10/14	Wednesday	85.625	78.375	......	88	85.625
15	10/15	Thursday	85.625	77.5625	......	89.6875	86.375
16	10/16	Friday	84	76.75	......	95.1875	76
17	10/19	Monday	84.8125	84	......	100.75	84
18	10/20	Tuesday	96	92	......	92.8125	74.4375
19	10/21	Wednesday	76	92.75	......	92	77.625
20	10/22	Thursday	87.1875	85.625	......	84.8125	87.1875

(Continued)

Table 2 (continued)

No.	Dates	Day	Hourly electricity consumption (kWh)				
			00:00∼01:00	01:00∼02:00	. . . . . .	22:00∼23:00	23:00∼00:00
21	10/23	Friday	76.75	84.0625	. . . . . .	100.8125	76
22	10/26	Monday	84	75.1875	. . . . . .	94.4375	84.8125
23	10/27	Tuesday	83.1875	84.8125	. . . . . .	111.25	84.75
24	10/28	Wednesday	94.4375	83.9375	. . . . . .	103.1875	77.625

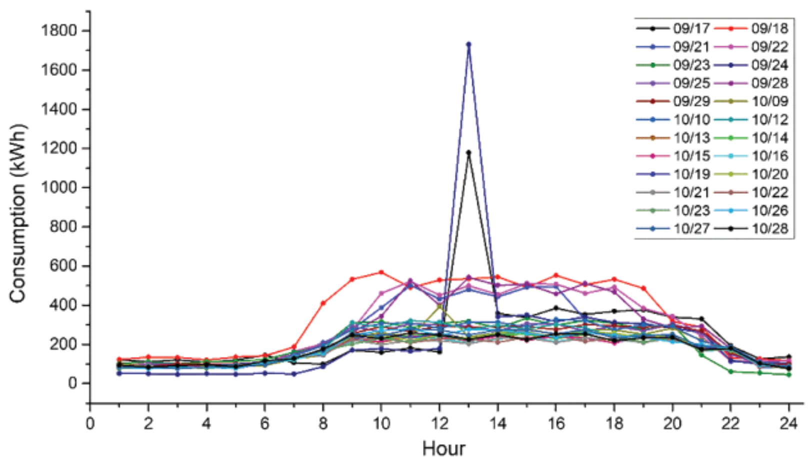

Figure 10: Daily electricity consumption

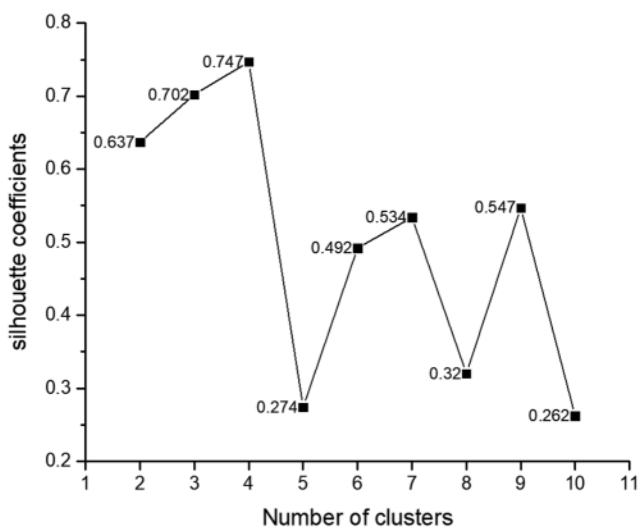

Figure 11: Trend of silhouette coefficient

We clustered the data using a hierarchical approach when the number of clusters is four. The results of hierarchical clustering are as Fig. 12. In this figure, three dates (09/17, 09/18, and 09/24) were clustered into one cluster individually, and the remaining 21 days were clustered into one cluster. Through observations, Cluster 1 and Cluster 3 can be grouped and are characterized by high consumption around 13:00, which is not an order of magnitude higher than at other times. It is assumed that at 12:00 and 13:00, the building has a high consumption of equipment. However, by observing the data of three subsystems (lighting & sockets, HVAC and special) consumption, it was found that they consumed different amounts of power, but no abnormal values appeared. There was no discernible increase, which is thought to be the result of other electrical equipment systems or meter damage. This was also demonstrated by using the k-means approach. The k-means results are shown in Fig. 13, which can be divided into the following four clusters. It is obvious from the Fig. 13 that k-means approach put Cluster 1 and Cluster 3 into one class.

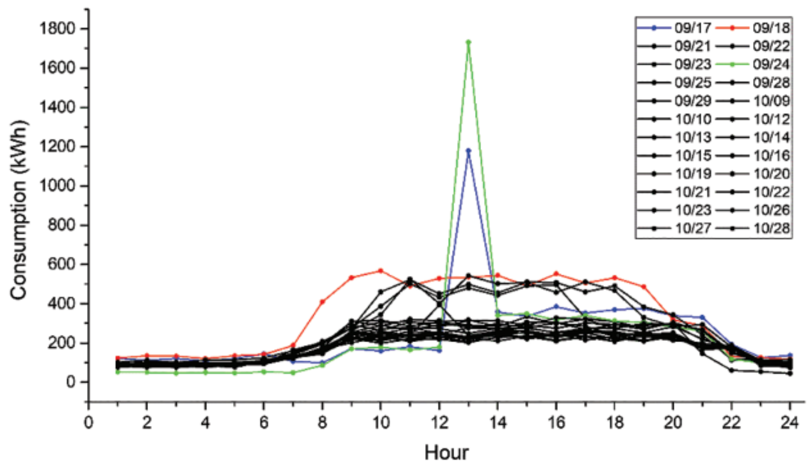

Figure 12: Results of hierarchical clustering

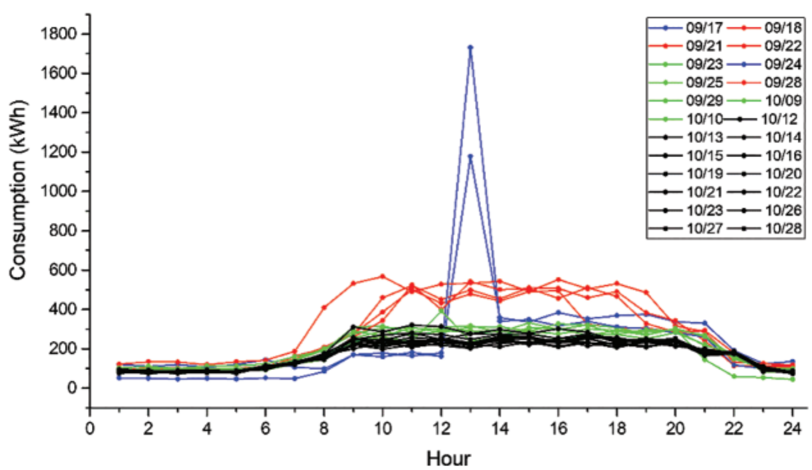

Figure 13: Results of k-means approach

In addition, the building maintained an overall high-power consumption on 09/18. Hierarchical approach clustered it into a separate class, while the k- means approach clustered it into a class with the dates 09/21, 09/22, and 09/28. This is what we do not want to see. Because we have found a distinctive peculiarity of 09/18 through our research. The temperature on 09/18 was not much more specialized to the other dates, but the energy consumption was significantly different. We defined it as a separate special pattern. Because in this pattern, the library conducts activities such as holding exhibitions and visitor visits. At that time, the library's lighting sockets, HVAC, etc., will be at a higher load. Therefore, hierarchical clustering continues to be adopted for this analysis. In Fig. 14, the four types of clustered curves are drawn, respectively. The Cluster 4 in Fig. 14 is the average of all the fourth type data in Fig. 12. The characteristics of the four clusters are different, showing different energy use behaviors.

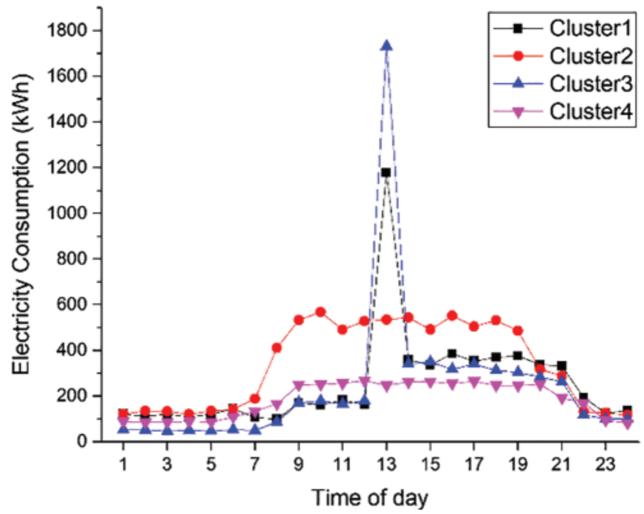

Figure 14: Diurnal pattern of hourly electricity consumption (the 1st cluster)

Cluster 1: A peak value of around 150 kWh before 8:00 am, a significant peak at noon, a sudden increase to around 1200 kWh between 12:00 am–2:00 pm, a return to a level of around 300 kWh after 14:00, and a gradual decline to around 300 kWdh after 9:00 at night, around 100 kWh.

Cluster 2: No significant peaks, mainly characterized by a small consumption, less than 200 kWh until 7:00 am, an accelerating decreasing consumption increment from 7:00–10:00 am, a stable consumption from 10–16:00 am, and a significant decreasing process after 19:00 am.

Cluster 3: This cluster is very similar to Cluster 1 in terms of electricity consumption characteristics, with a significant peak at midday, which is even higher than the Cluster 1 peak at around 1700 kWh higher energy consumption.

Cluster 4: No significant peaks, an overall stable electricity consumption with no rapid fluctuations, and consumption of around 270 kWh from 9:00 to 20:00, which is significantly higher than the consumption from 23:00 to 7:00 the following day (90 kW/h).

After our investigation, we found that the Cluster 1 and Cluster 3 are not representative, as they are all special patterns of electricity use for special reasons. The Cluster 2 and Cluster 4 can be analyzed separately as an electricity consumption model in everyday life. These data is now presented in Fig. 15. Two different types of electricity consumption patterns can be clearly observed in Fig. 15, so we set the cluster number to 2, and do a new round of systematic clustering. The result is shown in Fig. 16. The first mode is shown by the red line. The building consumption is a low electricity level from 11:00 pm–5:00 am, about 95 kW per hour. Two transitional periods are at 6:00 am–9:00 am and

8:00 pm–11:00 pm. After the students entered the building, the first transition time can be understood as the equipment started up and the consumption of sockets increased. The second transition is just the opposite, for the socket electricity declines and equipment is off. 9:00 am–8:00 pm is a relatively smooth period of electricity, about 260 kW/h. The second model is shown as a black line. It can also be divided into four sections: 11:00 pm–5:00 am; 6:00 am–9:00 am; 10:00 am–8:00 pm; 9:00 pm–11:00 pm.

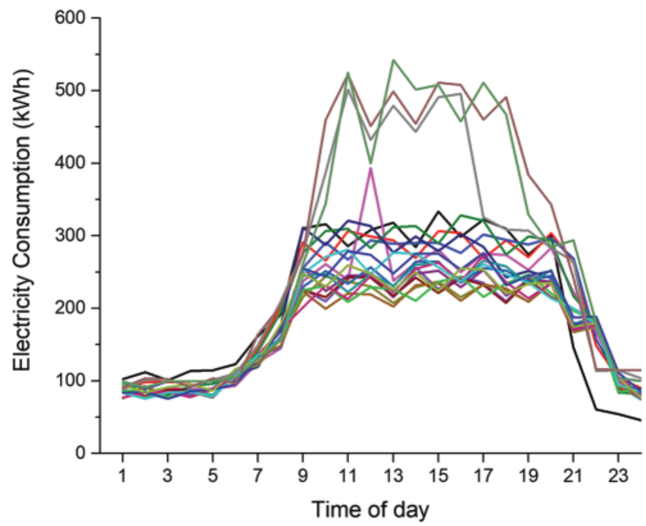

Figure 15: Hourly electricity consumption (the 2nd cluster)

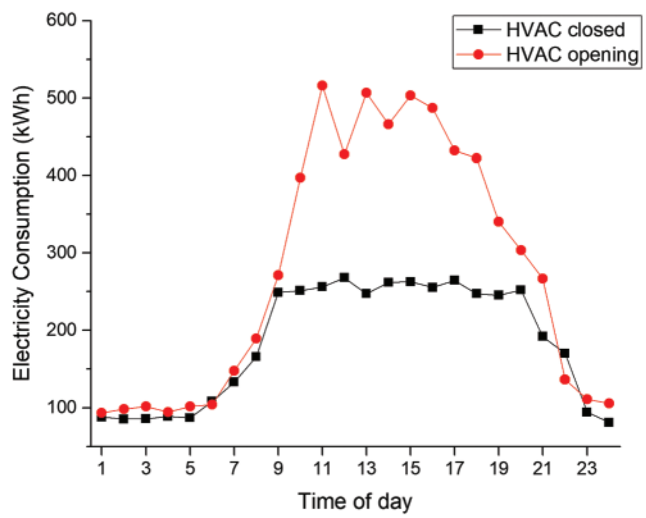

Figure 16: Diurnal pattern of hourly electricity consumption (the 2nd cluster)

Through the comparison of the two modes, it can be found that the two modes before 9:00 am and after 10:00 pm are not much different. The biggest difference is that during the day, the energy consumption of the air conditioning system is higher than in the first mode. Fig. 17 summarizes the daily electrical pattern of this library. To balance the seasonal fluctuations in power supply caused by HVAC, the campus's energy supplies should improve the reliability of the campus supply grid and prevent mandatory peak clipping steps. Energy awareness education should be broadly introduced in order to increase building energy efficiency while not interfering with students' daily learning.

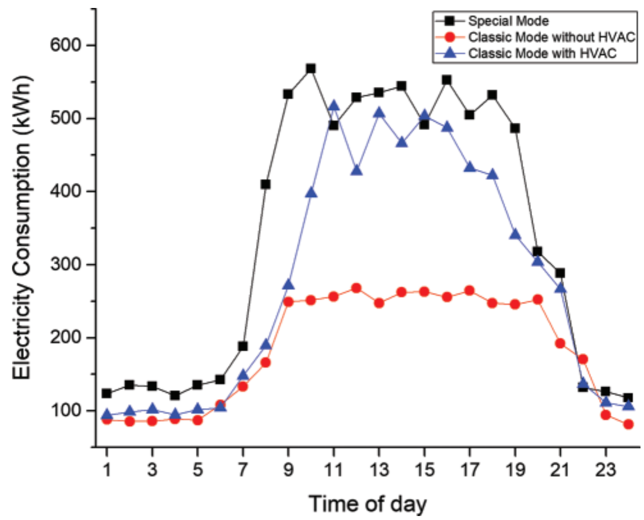

Figure 17: The diurnal pattern of hourly electricity consumption in this library

3.3 Clustering Coupling Analysis of Library Electricity Data Based on Association Rule Algorithms

This section will use generalized data to divide the building's electricity consumption and temperature from August 31 to October 28 into categories and reduce the number of values for a given continuous attribute. The continuous data is discretely clustered using a systematic clustering method, replacing the category labels' actual data values [54].

The daily total consumption, HVAC consumption, particular electricity consumption, daily maximum temperature, and minimum temperature of the building from August 31 to October 28 are systematically clustered. Each of them will be classified as high, medium, and low, so the number of clusters is selected as three. Figs. 18a–18e show the clustering results of these five types of data. Based on the cluster analysis results, the total consumption, HVAC consumption, particular electricity consumption and the temperature are generalized into three levels: low, medium, and high, as shown in Table 3. The consumption and temperature in the original dataset are classified and obtained from the generalized dataset, as shown in Table 4. TL, TM, and TH represent the low, medium, and high electricity consumption levels. AL, AM, and AH represent the low, medium, and high levels of the HVAC consumption. SL, SM, and SH represent the low, medium, and high levels of particular electricity consumption. TMAXL, TMAXM, and TMAXH represent the low, medium, and high levels of the highest temperatures outside. TMINL, TMINM, and TMINH represent the low, medium, and high levels of the lowest temperatures outside.

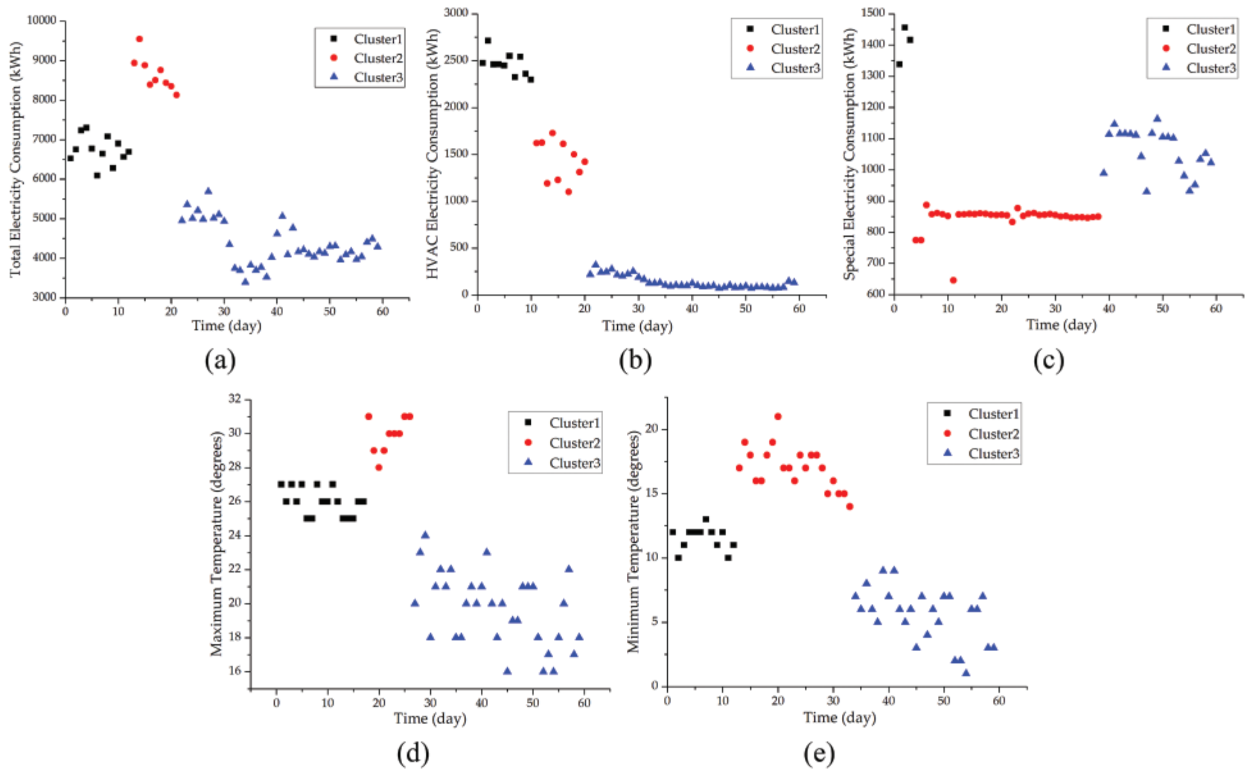

Figure 18: Clustering results of five variables: (a) Daily total electricity consumption; (b) Daily HVAC consumption; (c) Daily particular electricity consumption; (d) Daily maximum temperature; (e) Daily minimum temperature

Table 3: Ranges of five variables generalization grade

Generic terms	Generalization levels		
	Low	Medium	High
Total building electricity consumption	3300~6000	6000~7500	7500~10000
Air conditioning electricity consumption	0~1000	1000~2000	2000–3000
Special electricity consumption	600~900	900~1200	1200~1500
Maximum temperature	15~24.5	24.5~27.5	27.5~35
Minimum temperature	0~9	10~13	14~21

Table 4: Generalization datasets

Dates	Total	HVAC electricity	Special electricity	Maximum temperature	Minimum temperature
08/31	TH	AH	SH	TMAXH	TMINH
09/01	TH	AH	SH	TMAXM	TMINH

(Continued)

Table 4 (continued)

Dates	Total	HVAC electricity	Special electricity	Maximum temperature	Minimum temperature
……	……	……	……	……	……
09/25	TL	AM	SL	TMAXM	TMINM
……	……	……	……	……	……
10/26	TL	AL	SM	TMAXL	TMINL
10/27	TL	AL	SM	TMAXL	TMINL

Apriori algorithms are used to mine the dataset for association rules. Parameters are set as follows: support is 0.1, confidence is 0.6, and the maximum number of items in each item set is 2. The correlation results are sorted according to the degree of lift. Table 5 shows correlation rules sorted by lift. From Table 5, we can see that:

(1) Effect of temperature on total electricity consumption. Rules 3, 8, and 10 are analyzed. Combined with the generalization level, it can be seen from rule 8 that the daily minimum temperature is low and has a strong correlation with the low total consumption. When the daily minimum temperature is between 0 and 9 degrees Celsius, almost 100% of the total consumption is the lowest. Similarly, rule 10 shows that the daily maximum temperature is low in most of the time interval studied in this paper, and the total consumption is also at a low level. When the daily maximum temperature is in the range of 15–24.5 degrees Celsius, the probability of the total consumption is 93.93%, which is at a low level of 3300–6000 kWh. According to rule 3, when the maximum temperature is 27.5–35 degrees Celsius, the probability of total consumption is 66.67%, 7500–10000 kWh. Therefore, the corresponding demand response mechanism can be considered when the temperature is high. Prevent excessive building power consumption due to elevated temperatures and reduce damage to power grids and equipment lines.

(2) Impact of HVAC consumption on total consumption. Rule 2, rule 5, and rule 13 are analyzed. When HVAC consumption is at high, middle, and low levels, building total consumption is at corresponding high, middle, and low levels. The membership degrees at the corresponding level are 0.153, 0.102, and 0.576, respectively, that is, the rule has a certain number; The corresponding levels of confidence were 0.900, 0.872, 0.600, which reached a high degree of confidence, indicating that the 'quality' requirements are met. The conclusion is consistent with the existing law of building electricity. In our data, HVAC consumption has a crucial impact on the total consumption of the building, verifying the correctness of the method. Further, if the energy consumption data of each air conditioning unit can be obtained, and based on the correlation analysis between them, the joint regulation of multiple units with high correlation can achieve accurate load scheduling.

(3) Relationship between HVAC consumption and particular electricity consumption. Special electricity consumption includes power consumption of the pump room, exhaust fan, and equipment in the engine room. By analyzing rules 6, 16, and 18, it can be found that there is a specific positive correlation between HVAC consumption and particular electricity consumption. When the HVAC electricity is at a high level (2000–3000 kWh per day), the special electricity is at a medium level (900–1200 kWh per day). Furthermore, when HVAC

consumption is at a low level (0–2000 kWh per day), the special electricity consumption is low (600–900 kWh per day).

Table 5: Results of correlation rule mining

No.	Prerequisite	Association results	Support	Confidence	Lift
1	TATOL = TH	AIRCONDITION = AH	0.153	1.000	5.900
2	AIRCONDITION = AH	TATOL = TH	0.153	0.900	5.900
3	MAXTEMPERATURE = TMAXH	TATOL = TH	0.102	0.667	4.370
4	MAXTEMPERATURE = TMAXH	AIRCONDITION = AH	0.102	0.667	3.933
5	AIRCONDITION = AM	TATOL = TM	0.102	0.600	2.950
6	AIRCONDITION = AH	SPECIAL = SM	0.119	0.700	1.967
7	TATOL = TH	SPECIAL = SM	0.102	0.667	1.873
8	MINTEMPERATURE = TMINL	TATOL = TL	0.441	1.000	1.553
9	MINTEMPERATURE = TMINL	AIRCONDITION = AL	0.441	1.000	1.513
10	MAXTEMPERATURE = TMAXL	TATOL = TL	0.525	0.939	1.459
11	MAXTEMPERATURE = TMAXL	AIRCONDITION = AL	0.525	0.939	1.421
12	TATOL = TL	AIRCONDITION = AL	0.576	0.895	1.354
13	AIRCONDITION = AL	TATOL = TL	0.576	0.872	1.354
14	MINTEMPERATURE = TMINL	SPECIAL = SL	0.339	0.769	1.297
15	MAXTEMPERATURE = TMAXL	SPECIAL = SL	0.424	0.758	1.277
16	AIRCONDITION = AL	SPECIAL = SL	0.492	0.744	1.253
17	SPECIAL = SL	AIRCONDITION = AL	0.492	0.829	1.253
18	AIRCONDITION = AM	SPECIAL = SL	0.102	0.600	1.11

Through the above analysis, the proposed association rule mining can provide practical guidance and help for the application of building consumption management strategy for demand response. Several significant limitations need to be considered. The lack of equipment data makes it impossible to explore the energy consumption relationships between various electrical equipment or subsystems. For example, the air conditioning unit, lighting system, fresh air system, monitoring system, personnel behavior (such as students, staff), and other related data can be introduced, and the correlation analysis can be carried out by using this method. In the future, when the building participates in the demand response, according to the correlation between each equipment and system, the joint regulation of multiple types of equipment and multiple systems can achieve accurate load scheduling and achieve the purpose of optimizing electricity under the premise of satisfying the user's comfort.

4 Conclusions

This paper presented a systematic approach to characterizing the electricity consumption of a library building in Beijing using standard seasonal decomposition, hierarchical clustering and apriori correlation mining. The data on monthly/daily/hourly energy consumption was profiled and analyzed. The following conclusions were drawn:

(1) Verified that the winter and summer holidays are the two lowest energy consumption periods for this type of building. The month of Spring Festival has the lowest energy use of the year. June is the highest energy consumption month of the year for buildings. It was found that educational schedules, social examinations, and weather conditions all had a significant impact on energy consumption.

(2) Three typical daily load patterns were summarized. They are HVAC closed mode, HVAC opening mode and special mode. The hourly analysis found that whatever the mode was divided into three phases: off, transition and smooth operation.

(3) Correlations between different energy consumption, different temperatures were discovered. The daily data analysis revealed a strong relationship between HVAC use and total building consumption, with a lift value of 5.9. It demonstrates that HVAC consumption plays a significant influence in total consumption in this building during this time period.

The data analysis methodology described in this study can be utilized as a guideline for future research on similar structures. It also serves as a foundation for library energy-efficiency upgrading. Time-share tariffs and building equipment optimization strategies can be developed for library buildings. Future study should include: increasing the amount of data in the model; experimenting with various models (for example, using k-shape approaches for cluster analysis); investigating variations in energy use for different energy sources (natural gas, water, and others). The conclusion can provide the research basis for analyzing similar buildings.

Funding Statement: This research was funded in part by the Doctoral Scientific Research Foundation of Beijing University of Civil Engineering and Architecture under Grant ZF15054, in part by the Fundamental Research Funds for Beijing University of Civil Engineering and Architecture under Grant X18066, and in part by the 2021 BUCEA Post Graduate Innovation Project under Grant PG2021011.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Li, K., Wang, F., Zhen, Z., Mi, Z., Sun, H. et al. (2016). Analysis on residential electricity consumption behavior using improved k-means based on simulated annealing algorithm. *IEEE Power and Energy Conference at Illinois (PECI)*, vol. 2016, pp. 1–6. Urbana, USA.
2. Pan, S., Wang, X., Wei, Y., Zhang, X., Gal, C. et al. (2017). Cluster analysis for occupant-behavior based electricity load patterns in buildings: A case study in shanghai residences. *Building Simulation, 10,* 889–898. DOI 10.1007/s12273-017-0377-9.
3. IEA (2021). Buildings–topics. https://www.iea.org/topics/buildings.
4. Fang, J., Ma, R., Deng, Y. (2020). Identification of the optimal control strategies for the energy-efficient ventilation under the model predictive control. *Sustainable Cities and Society, 53,* 101908. DOI 10.1016/j.scs.2019.101908.

5. Capozzoli, A., Piscitelli, M. S., Gorrino, A., Ballarini, I., Corrado, V. (2017). Data analytics for occupancy pattern learning to reduce the energy consumption of HVAC systems in office buildings. *Sustainable Cities and Society, 35,* 191–208. DOI 10.1016/j.scs.2017.07.016.
6. Homod, R. Z. (2018). Analysis and optimization of HVAC control systems based on energy and performance considerations for smart buildings. *Renewable Energy, 126,* 49–64. DOI 10.1016/j.renene.2018.03.022.
7. Homod, R. Z., Togun, H., Abd, H. J., Sahari, K. S. M. (2020). A novel hybrid modelling structure fabricated by using takagi-sugeno fuzzy to forecast HVAC systems energy demand in real-time for Basra city. *Sustainable Cities and Society, 56,* 102091. DOI 10.1016/j.scs.2020.102091.
8. Jiang, J. (2016). China's urban residential carbon emission and energy efficiency policy. *Energy, 109,* 866–875. DOI 10.1016/j.energy.2016.05.060.
9. Xinhuanet (2021). U.S.-China joint presidential statement on climate change. http://www.xinhuanet.com//world/2015-09/26/c_1116685873.htm.
10. Hu, S., Yan, D., Guo, S., Cui, Y., Dong, B. (2017). A survey on energy consumption and energy usage behavior of households and residential building in urban China. *Energy and Buildings, 148,* 366–378. DOI 10.1016/j.enbuild.2017.03.064.
11. Li, L. (2021). Impacts of multiple factors on energy consumption of aging residential buildings based on a system dynamics model–taking Northwest China as an example. *Journal of Building Engineering, 14,* 10295. DOI 10.1016/j.jobe.2021.102595.
12. Homod, R. Z., Almusaed, A., Almssad, A., Jaafar, M. K., Goodarzi, M. et al. (2021). Effect of different building envelope materials on thermal comfort and air-conditioning energy savings: A case study in Basra city, *Iraq. Journal of Energy Storage, 34,* 101975. DOI 10.1016/j.est.2020.101975.
13. Yilmaz, S., Chambers, J., Patel, M. K. (2019). Comparison of clustering approaches for domestic electricity load profile characterisation-Implications for demand side management. *Energy, 180,* 665–677. DOI 10.1016/j.energy.2019.05.124.
14. Godina, R., Rodrigues, E., Pouresmaeil, E., Matias, J., Catalão, J. (2018). Model predictive control home energy management and optimization strategy with demand response. *Applied Sciences, 8,* 408. DOI 10.3390/app8030408.
15. IEA (2021). Demand response. https://www.iea.org/reports/demand-response.
16. Parra, D., Patel, M. K. (2016). Effect of tariffs on the performance and economic benefits of pv-coupled battery systems. *Applied Energy, 164,* 175–187. DOI 10.1016/j.apenergy.2015.11.037.
17. Barbour, E., Parra, D., Awwad, Z., González, M. C. (2018). Community energy storage: A smart choice for the smart grid? *Applied Energy, 212,* 489–497. DOI 10.1016/j.apenergy.2017.12.056.
18. Zhou, K., Yang, S., Shao, Z. (2017). Household monthly electricity consumption pattern mining: A fuzzy clustering-based model and a case study. *Journal of Cleaner Production, 141,* 900–908. DOI 10.1016/j.jclepro.2016.09.165.
19. Fischer, C. (2008). Feedback on household electricity consumption: A tool for saving energy? *Energy Efficiency, 1,* 79–104. DOI 10.1007/s12053-008-9009-7.
20. Gottwalt, S., Ketter, W., Block, C., Collins, J., Weinhardt, C. (2011). Demand side management—A simulation of household behavior under variable prices. *Energy Policy, 39,* 8163–8174. DOI 10.1016/j.enpol.2011.10.016.
21. Lee, S., Kwon, B., Lee, S. (2014). Joint energy management system of electric supply and demand in houses and buildings. *IEEE Transacion on Power System, 29,* 2804–2812. DOI 10.1109/TPWRS.2014.2311827.
22. Rajabi, A., Eskandari, M., Jabbari Ghadi, M., Ghavidel, S., Li, L. et al. (2019). A pattern recognition methodology for analyzing residential customers load data and targeting demand response applications. *Energy and Buildings, 203,* 109455. DOI 10.1016/j.enbuild.2019.109455.

23. Zhou, K., Yang, C., Shen, J. (2017). Discovering residential electricity consumption patterns through smart-meter data mining: A case study from China. *Utilities Policy, 44,* 73–84. DOI 10.1016/j.jup.2017.01.004.
24. Zhou, K., Fu, C., Yang, S. (2016). Big data driven smart energy management: Rom big data to big insights. *Renewable and Sustainable Energy Reviews, 56,* 215–225. DOI 10.1016/j.rser.2015.11.050.
25. Kong, W., Dong, Z. Y., Ma, J., Hill, D. J., Zhao, J. et al. (2018). An extensible approach for non-intrusive load disaggregation with smart meter data. *IEEE Transaction on Smart Grid, 9,* 3362–3372. DOI 10.1109/TSG.2016.2631238.
26. Ponocko, J., Milanovic, J. V. (2017). Application of data analytics for advanced demand profiling of residential load using smart meter data. *IEEE Manchester PowerTech, 2017,* 1–6. DOI 10.1109/PTC.2017.7980821.
27. Hannan, M. A., Faisal, M., Ker, P. J., Mun, L. H., Parvin, K. et al. (2018). A review of internet of energy based building energy management systems: Issues and recommendations. *IEEE Access, 6,* 38997–39014. DOI 10.1109/Access.6287639.
28. Homod, R. Z., Sahari, K. S. M., Nagi, F., Mohamed, H. A. F. (2010). Modeling of heat and moisture transfer in building using RLF method. *IEEE Student Conference on Research and Development (SCOReD)*, vol. 2010*,* pp. 287–292. Kuala Lumpur, Malaysia.
29. Annibaldi, V., Cucchiella, F., Rotilio, M. (2020). A sustainable solution for energy efficiency in Italian climatic contexts*. Energies, 13,* 2817. DOI 10.3390/en13112817.
30. Zhao, T. (2021). A monitoring data based bottom-up modeling method and its application for energy consumption prediction of campus building. *Journal of Building Engineering, 9,* 101962. DOI 10.1016/j.jobe.2020.101962.
31. Dong, Z., Liu, J., Liu, B., Li, K., Li, X. (2021). Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification. *Energy and Buildings, 241,* 110929. DOI 10.1016/j.enbuild.2021.110929.
32. Chicco, G. (2012). Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy, 42,* 68–80. DOI 10.1016/j.energy.2011.12.031.
33. Tsekouras, G. J., Kotoulas, P. B., Tsirekis, C. D., Dialynas, E. N., Hatziargyriou, N. D. (2008). A pattern recognition methodology for evaluation of load profiles and typical days of large electricity customers. *Electric Power Systems Research, 78,* 1494–1510. DOI 10.1016/j.epsr.2008.01.010.
34. Ozawa, A., Furusato, R., Yoshida, Y. (2016). Determining the relationship between a household's lifestyle and its electricity consumption in Japan by analyzing measured electric load profiles. *Energy and Buildings, 119,* 200–210. DOI 10.1016/j.enbuild.2016.03.047.
35. Pereira, R., Fagundes, A., Melício, R., Mendes, V. M. F., Figueiredo, J. et al. (2016). A fuzzy clustering approach to a demand response model. *International Journal of Electrical Power & Energy Systems, 81,* 184–192. DOI 10.1016/j.ijepes.2016.02.032.
36. Biscarri, F., Monedero, I., García, A., Guerrero, J. I., León, C. (2017). Electricity clustering framework for automatic classification of customer loads. *Expert Systems with Applications, 86,* 54–63. DOI 10.1016/j.eswa.2017.05.049.
37. Zhang, T., Zhang, G., Lu, J., Feng, X., Yang, W. (2012). A new index and classification approach for load pattern analysis of large electricity customers*. IEEE Transaction on Power System, 27,* 153–160. DOI 10.1109/TPWRS.2011.2167524.
38. Räsänen, T., Ruuskanen, J., Kolehmainen, M. (2008). Reducing energy consumption by using self-organizing maps to create more personalized electricity use information. *Applied Energy, 85,* 830–840. DOI 10.1016/j.apenergy.2007.10.012.
39. Crow, M. L. (2014). Clustering-based methodology for optimal residential time of use design structure. *North American Power Symposium (NAPS)*, vol. 2014, pp. 1–6. Pullman, USA.
40. Haben, S., Singleton, C., Grindrod, P. (2016). Analysis and clustering of residential customers energy behavioral demand using smart meter data. *IEEE Transaction on Smart Grid, 7,* 136–144. DOI 10.1109/TSG.2015.2409786.

41. Wang, F., Li, K., Duić, N., Mi, Z., Hodge, B. M. et al. (2018). Association rule mining based quantitative analysis approach of household characteristics impacts on residential electricity consumption patterns. *Energy Conversion and Management, 171,* 839–854. DOI 10.1016/j.enconman.2018.06.017.
42. Yang, J., Zhao, J., Wen, F., Dong, Z. (2019). A model of customizing electricity retail prices based on load profile clustering analysis. *IEEE Transaction on Smart Grid, 10,* 3374–3386. DOI 10.1109/TSG.5165411.
43. Fahiman, F., Erfani, S. M., Rajasegarar, S., Palaniswami, M., Leckie, C. (2017). Improving load forecasting based on deep learning and k-shape clustering. *International Joint Conference on Neural Networks (IJCNN)*, vol. 2017, pp. 4134–4141. Anchorage, USA.
44. Quintana, M., Arjunan, P., Miller, C. (2021). Islands of misfit buildings: Detecting uncharacteristic electricity use behavior using load shape clustering. *Building Simulation, 14,* 119–130. DOI 10.1007/s12273-020-0626-1.
45. Wen, L., Zhou, K., Yang, S. (2019). A shape-based clustering method for pattern recognition of residential electricity consumption. *Journal of Cleaner Production, 212,* 475–488. DOI 10.1016/j.jclepro.2018.12.067.
46. Yan, L., Qian, F., Li, W. (2018). Research on key parameters operation range of central air conditioning based on binary k-means and apriori algorithm. *Energies, 12,* 102. DOI 10.3390/en12010102.
47. Jia, X., Wu, F. (2017). *Electromechanical control technology and transportation*. China: CRC Press.
48. Belessiotis, V., Mathioulakis, E. (2002). Analytical approach of thermosyphon solar domestic hot water system performance. *Solar Energy, 72,* 307–315. DOI 10.1016/S0038-092X(02)00011-7.
49. Fabi, V., Buso, T., Andersen, R. K., Corgnati, S. P., Olesen, B. W. (2013). Robustness of building design with respect to energy related occupant behaviour. *13th Conference of International Building Performance Simulation Association (IBPSA)*, pp. 1999–2006. Chambery, France.
50. Zang, Z., Dai, K., Wu, H., Gao, Z., Yang, G. (2019). Spring festival corrected electricity forecasting model based on x12 seasonal decomposition. *Annual Meeting of the Electricity Market Special Committee of CSEE*, pp. 1–6. Chengdu, China.
51. Luo, X. J., Oyedele, L. O., Olugbenga, O. O., Ajayi, A. O. (2020). 3D pattern identification approach for cooling load profiles in different buildings. *Journal of Building Engineering, 31,* 101339. DOI 10.1016/j.jobe.2020.101339.
52. Qi, Y., Luo, B., Wang, X., Wu, L. (2017). Load pattern recognition method based on fuzzy clustering and decision tree. *IEEE Conference on Energy Internet and Energy System Integration (EI2)*, vol. 2017, pp. 1–5. Beijing, China.
53. Rdusseeun, L., Kaufman, P. (1987). Clustering by means of medoids. *Statistical Data Analysis Based on the L1 Norm Conference*, vol. 31, pp. 405–416. Neuchatel, Switzerland.
54. Lin, S., Hu, F., Hao, C., Li, D., Fu, Y. (2018). Study on the power consumption analysis model of building based on data mining. *Electrical Measurement & Instrumentation, 55,* 52–59. DOI 10.3969/j.issn.1001-1390.2018.20.009.