**3D pattern identification approach for cooling load profiles in different buildings**

X.J. Luo, Lukumon O. Oyedele*, Olugbenga O. Olugbenga and Anuoluwapo O. Ajayi

Big Data Enterprise and Artificial Intelligence Laboratory
University of the West of England, Frenchay Campus, Bristol, United Kingdom
*Corresponding author: l.oyedele@uwe.ac.uk

**Abstract**

Building energy conservation has gained increasing concern owing to its large portion of energy consumption and great potential of energy saving. In-depth understanding of representative patterns of daily cooling load profile will facilitate effective building energy system scheduling, fault detection and diagnosis, as well as demand and supply side management. In this study, a novel three-stage approach is proposed for pattern identification of cooling load profiles in different types of buildings. The three stages include data preparation, data clustering and data visualization. The initial measurement in the building energy management system is conducted at the time step of 15 minutes. To further explore the characteristics of the building cooling load trend, 1-h mean pattern, 4-h mean pattern and daily statistical information (i.e. average, minimum and maximum values) of cooling load are also adopted for data clustering, respectively. To test the generality and robustness of the proposed approach, one-year historical measurement data collected from the practical chilled water system in two different buildings are adopted, respectively. The analysis demonstrates that the 3D pattern identification approach can effectively discover the representative characteristics of the daily cooling load profiles in both buildings. It is also expected that the proposed 3-stage pattern identification approach is general in adoption and can be potentially adopted in various types of buildings in different climate zones.

**Keywords:**
Pattern identification; Gaussian mixture model clustering; cooling load; data visualization; energy management.

## 1. Introduction

As the largest energy-consuming sector, building accounts for over one-third of primary energy consumption globally [1]. To improve building energy efficiency, various technologies such as energy demand prediction [2], system design optimization [3], model predictive control [4], demand side management [5] as well as energy-efficient prime movers [6, 7] have been investigated.

### *1.1 Literature review*

Energy modelling in buildings seeks to quantify energy consumption. In recent years, a large number of studies for energy modelling approaches have been proposed for different types of buildings, which can be generally categorised into engineering-based methods and data-driven methods [8]. Engineering methods adopt the fundamental thermodynamic and physics models to investigate the energy behaviour of the building. On the other hand, data-driven methods use the statistical or machine learning methods to investigate building energy performance using historical measurement data [9,10]. The newly developed statistical models include multiple linear regression and various gaussian process regression models [11, 12]; while the newly developed machine learning models include artificial neuron networks, support vector machines, decision trees and genetic algorithms. On the other hand, pattern identification of representative energy load profiles will further enhance the performance of energy demand prediction, system design optimization as well as demand and supply side management. Clustering-based data-driven models are an unsupervised energy consumption data analysis approach including the purpose of discovering unlabelled datasets and hidden information, which is widely used within the studies of building electricity, heating and cooling load profiles.

Pattern identification of electricity consumption profile was mainly investigated on residential and office buildings. For residential buildings, Wen *et al.* [13] proposed a principle component analysis-based *k*-means clustering algorithm and a shape-based clustering method for pattern recognition of residential electricity consumption. Li *et al.* [14] proposed a simulated annealing-based *k*-means clustering approach to extract typical electricity consumption patterns of different residential buildings. Guo *et al.* [15] used the *k*-means clustering method to identify the electricity consumption patterns in different households within an area in China. In these three studies, year-round electricity consumption profiles collected from a large quantity of residential households were adopted as database. Through various *k*-means clustering methods, various residential buildings were grouped into different clusters, while the representative electricity consumption patterns were recognized from each cluster. For office buildings, Fan *et al.* [16] proposed a temporal knowledge discovery methodology for identifying electrical energy consumption of an office building. Several time series data mining techniques were explored and assembled, including the symbolic aggregate approximation, motif discovery, and temporal association rule mining. The year-round electricity consumption profile of the building was grouped into 4 clusters, while each cluster represented a climate (i.e. cold and hot seasons) and day type (i.e. weekdays and weekends). Xiao *et al.* [17] adopted the entropy-weighted *k*-means clustering algorithm to identify the typical electrical power consumption patterns in an office building. It was found that electricity consumption patterns in weekdays were similar to each other, while there existed distinctive differences among weekdays, Saturdays and Sundays. Li *et al.* [18] adopted the Gaussian mixture model (GMM) clustering to identify the typical daily electricity usage profiles of each individual building. Through GMM clustering, the year-round daily electricity consumption profile was grouped into two or three clusters, which indicate the cold, medium and hot seasons, respectively. In addition, an agglomerative hierarchical clustering to identify electricity usage profiles of multiple buildings. As a result, typical patterns and periodical variation of daily electricity consumption could be discovered. Furthermore, Wang *et al*. [19] adopted the *k*-means clustering for pattern recognition of water-source heat pump operation using its electrical power consumption. The one-month profile of pump power consumption was grouped into 6 different clusters, each of which represented a typical daily operating schedule of heat pumps. In these four studies, the electricity consumption profile was grouped mainly according to its intensity.

Building heating loads are mainly provided by heat pumps and district heating system. Carolina *et al.* [20] adopted the *k*-means clustering to explore patterns in daily heating load profiles of heat pumps in Danish dwellings. Two main clusters were identified: a main cluster with a relatively constant load profile and a minor cluster with a more distinct variation during the daytime. Panagiota *et al.* [21] proposed a clustering-based knowledge discovery approach for evaluating heating consumption data of the residential buildings from district heating system. It was found that the daily heating load profiles were grouped into 5 clusters with regards to their consumption intensity. The clusters were characterized by relatively constant load profiles with a peak in the early morning and a peak in the evening, respectively. Lu *et al.* [22] adopted the GMM clustering approach to identify the occur time and energy signature of daily heating load patterns in 6 office buildings. Four operating patterns of the district heating system were identified, including working, on-duty, daytime-nighttime and nighttime-daytime patterns.

Building cooling load can be estimated from various measurement of the chilled water system. However, there were quite limited studies regarding the pattern identification of building cooling loads. Yu *et al.* [23, 24] used the *k*-means clustering algorithm to evaluate the performance of the chilled water system. The operating variables including load of each operating chiller, the total system load, the quantity of operating primary chilled water pumps, the quantity of operating condenser water pumps, and the quantity of operating cooling towers were adopted as database for clustering analysis. The operating data at each time step was grouped into 3 clusters, indicating the small, medium and heavy loading condition of the chilled water system, respectively. An *et al.* [25] adopted the *k*-means clustering algorithm to analysis the air-conditioning intensity and usage patterns in residential buildings. The daily usage of air conditioning system was grouped into different clusters based on its operating probability. Although these three studies investigated the chilled water and air conditioning systems, it mainly focused on the operating parameters of the chilled water system and the air conditioning systems. The variation trend of daily cooling load profiles was not investigated.

## *1.2 Research gaps and Contribution*

From the above-discussed comprehensive literature review, three major research gaps were identified:

- Engineering building energy simulation methods require both high level of technical knowledge and detailed information of building materials to improve its modelling accuracy;
- Although data clustering approaches have been widely adopted for pattern identification of energy load profiles in residential and office buildings, they mainly focused on electricity and heating loads. In previous cooling load studies, only the operating parameters of the chilled water system were investigated, while there is a lack of study regarding the daily profile of the building cooling loads.
- Owing to its low computational complexity, *k*-means clustering was widely adopted in previous research works for pattern identification. However, the initial clustering centroids were randomly selected, which might lead to convergence to local minimum points and difficulty in achieving global optimal solutions. However, GMM clustering can approximate any probability distribution by adjusting the quantity of mixture models, while there is a lack of study in exploring the effectiveness of GMM clustering in pattern identification of daily cooling load profile.
- Furthermore, previous studies mainly focus on the residential and office buildings. Due to the unique functions, the cooling load profile in campus and hotel buildings also worth investigation. However, none of the previous studies considered the energy loads in campus and hotel buildings.
- Last but not least, most of the previous research works mainly focused on the methodology of data clustering. However, data preparation is critical in providing the high-quality database while data visualization is significant in analyzing the clustering results.

Therefore, the contribution of this study is to propose a robust and generalized 3D pattern identification approach for daily cooling load profiles which can be adopted in different types of buildings. The 3D indicates the three-stage of the pattern identification approach: data preparation, data clustering and data visualization, while GMM is adopted for data clustering. To demonstrate the robustness and generality of the proposed pattern identification approach, the historical measurement data from a campus and a hotel building from the same hot-tropic area are adopted to test the performance of the proposed pattern identification approach. It is also expected that the proposed approach can be generalized and adopted in diverse climate zones.

The remainder of the paper is organized as follows: the next section illustrates the structure of the proposed pattern identification approach; the third section presents the information regarding the chilled water system; the fourth section discusses the identified representative patterns of building cooling loads; the fifth section shows implication for practice and future direction while the last section draws the conclusion.

## 2. Structure of the pattern identification approach

In this study, a three-stage pattern identification approach is proposed to process and analyse the building cooling load profile. The structure of the proposed pattern identification approach is shown in Fig. 1. The three stages consist of data preparation, data clustering and data visualization.

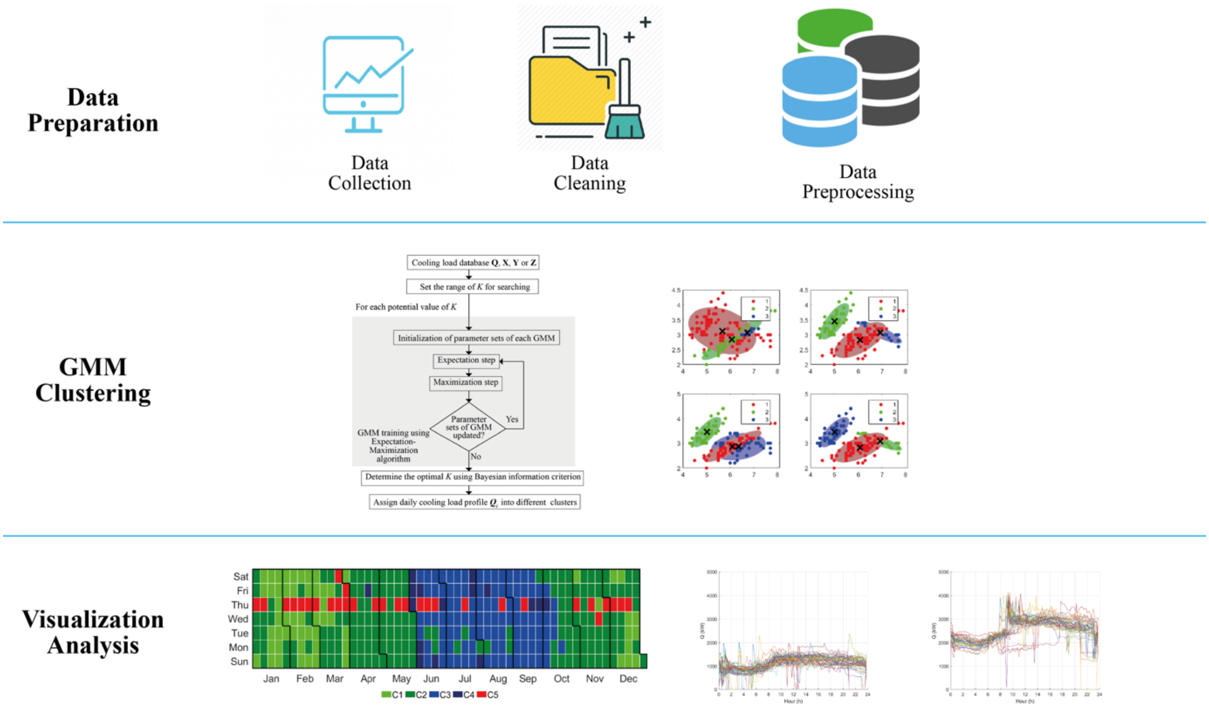


Fig. 1. Structure of 3D pattern identification approach.

### 2.1 *Data preparation*

To provide high-quality data for data clustering and data visualization, data preparation is conducted through three steps: data collection, data cleaning and data pre-processing.

#### 2.1.1 Data collection

Due to heat loss on various pipes and circulation pumps, the actual cooling load provided by the air handling unit would be smaller than the initial cooling energy generated by the chiller plant. However, operating schedule of the chilled water system should be determined by the initial cooling energy demand at the chiller side. Therefore, in this study, the cooling load profile is considered to be the initial cooling energy generated by the chiller plant. To gain the information about the initial cooling load profile, historical measurement data of the chilled water system should be collected from the building energy management system, covering total quantity of operating chillers $G$; leaving temperature $T_{l,g,i,j}$, return temperature $T_{r,g,i,j}$ and mass flow rate $M_{g,i,j}$ of each chiller; as well as the leaving temperature $T_{l,b,i,j}$, return temperature $T_{r,b,i,j}$ and mass flow rate $M_{b,i,j}$ at the main branch. $i$ stands for the number of the day, while $j$ represents the time step in each day.

Since the temperature difference between the inlet and outlet of the circulation water pump is relatively low, the sum of cooling load from each chiller should be approximate to the cooling load from the main branch. Therefore, the chilled water system can be estimated using two methods: from each operating chiller $Q_{c,i,j}$ and from main branch $Q_{b,i,j}$:

$$Q_{c,i,j} = \sum_{g=1}^{g=G} M_{g,i,j}\ C_p\ (T_{r,g,i,j} - T_{l,g,i,j}) \quad (1)$$

$$Q_{b,i,j} = M_{b,i,j}\ C_p\ (T_{r,b,i,j} - T_{l,b,i,j}) \quad (2)$$

where $C_p$ is the specific heat capacity of chilled water.

### 2.1.2 Data cleaning

Missing and outlier values are mainly caused by sensor malfunctions and signal transmission problems in the building management system. The missing and outlier values of the measurement data have a negative impact on data analysis. Hence, the aim of data cleaning is to improve the quality of raw measurement data by excluding missing and outlier values. In this study, after detecting the missing values, outliers of the measurement data are identified through the following procedures:

- When the chiller is operating, the temperature of the leaving chilled water temperature should be lower than the return chilled water owing to the heat exchange with the cooling tower. Therefore, chilled water can be delivered to the air handling units for space cooling purposes. When space cooling is not needed, chillers and cooling towers are turned off, thus the temperature of leaving chilled water is equal to that of the returning chilled water. Therefore, at each time step, $Q_{c,i,j}$ and $Q_{b,i,j}$ should not be lower than 0. The negative value of $Q_{c,i,j}$ and $Q_{b,i,j}$ might be caused by the faulty sensor measurement thus should be excluded.
- Due to the heat loss in pipes and circulation pumps, there would be little difference between$Q_{c,i,j}$ and $Q_{b,i,j}$. Therefore, the relative error $r$ between $Q_{c,i,j}$ and $Q_{b,i,j}$ should not be higher than 10%:

$$r = \begin{cases} \frac{2|Q_{c,i,j} - Q_{b,i,j}|}{Q_{c,i,j} + Q_{b,i,j}} & \max(Q_{c,i,j}, Q_{b,i,j}) > 0 \\ 0 & else \end{cases} \tag{3}$$

After clearing out the missing and outlier values, the cooling load of the chilled water system $Q_{i,j}$ is calculated as:

$$Q_{i,j} = \frac{Q_{c,i,j} + Q_{b,i,j}}{2} \tag{4}$$

2.1.3 Data pre-processing

In building energy management system, sensor measurement is generally recorded at 15 minutes. From the 15-min measurement data, the database $\mathbf{Q}$ for the year-round cooling load is consolidated as a 365 × 96 matrix. $\mathbf{Q} = [\boldsymbol{Q_1};\ \boldsymbol{Q_2}; \dots \boldsymbol{Q_i}; \dots; \boldsymbol{Q_{365}}]$ , each $\boldsymbol{Q_i}$ is a vector representing daily cooling load profile of the $i^{th}$ day, and $\boldsymbol{Q_i} = [Q_{i,1}\ Q_{i,2} \dots Q_{i,j} \dots\ Q_{i,96}]$. To provide an in-depth investigation on the daily cooling load profile, the initial 15-min measurement data is averaged at every 1 h and 4 h, respectively. In addition, daily brief information (i.e. daily minimum, maximum and average value) of the cooling load profile is also obtained.

To make the cooling load relatively flatten, the average value of the 15-min cooling load $Q_{i,j}$ is obtained at every 4-time-steps to obtain the 1-h mean value.

$$X_{i,j} = \frac{Q_{i,(j-1)\times4+1} + Q_{i,(j-1)\times4+2} + Q_{i,(j-1)\times4+3} + Q_{i,(j-1)\times4+4}}{4} \tag{5}$$

Thus, the daily 1-h mean cooling load profile $\boldsymbol{X_i}$ is a vector consisting of 24 elements, and $\boldsymbol{X_i} = [X_{i,1}\ X_{i,2} \dots X_{i,j} \dots\ X_{i,24}]$. Meanwhile, the database $\mathbf{X}$ for the 1-h mean value is consolidated as a 365 × 24 matrix, and: $\mathbf{X} = [\boldsymbol{X_1};\ \boldsymbol{X_2}; \dots; \boldsymbol{X_{365}}]$.

Furthermore, to gain briefer understanding of the overall trend of the daily cooling load profile, the average value of the 1-h mean cooling load $X_{i,j}$ is calculated at every 4 time steps to obtain the 4-h mean value.

$$Y_{i,j} = \frac{X_{i,(j-1)\times4+1} + X_{i,(j-1)\times4+2} + X_{i,(j-1)\times4+3} + X_{i,(j-1)\times4+4}}{4} \tag{6}$$

Therefore, the daily 4-h mean cooling load profile $\boldsymbol{Y_i}$ is a vector consisting of 6 elements, and $\boldsymbol{Y_i} = [Y_{i,1}\ Y_{i,2}\ \dots\ Y_{i,6}]$. After that, the database $\mathbf{Y}$ for the 4-h mean value is consolidated as a 365 × 6 matrix, and $\mathbf{Y} = [\boldsymbol{Y_1};\ \boldsymbol{Y_2};\ \dots;\boldsymbol{Y_{365}}]$.

To summarize the daily trend of the cooling load profile, its average, minimum and maximum values in each day is consolidated as the daily brief information.

$$Z_{i,1} = \frac{Q_{i,1}+Q_{i,2}+\cdots+Q_{i,96}}{96} \tag{7}$$

$$Z_{i,2} = \max\left(Q_{i,1}, Q_{i,2}, \dots, Q_{i,96}\right) \tag{8}$$

$$Z_{i,3} = \min\left(Q_{i,1}, Q_{i,2}, \dots, Q_{i,96}\right) \tag{9}$$

As a result, the daily brief information $\boldsymbol{Z_i}$ is a vector consisting of 3 elements, and $\boldsymbol{Z_i} = [Z_{i,1}\ Z_{i,2}\ Z_{i,3}]$. The database $\mathbf{Z}$ for the daily brief information is consolidated as a 365 × 3 matrix, and $\mathbf{Z} = [\boldsymbol{Z_1};\ \boldsymbol{Z_2};\ \dots;\boldsymbol{Z_{365}}]$.

### *2.2 Data clustering*

GMM-based clustering uses the feature of GMM to group multiple observations from the database into different clusters. In order to investigate the representative patterns of the cooling load profile, GMM clustering is conducted to group the databases of **Q**, **X**, **Y** and **Z** into different clusters.

#### 2.2.1 GMM clustering

The GMM is a probabilistic model, which uses a weighted combination of several Gaussian distributions to represent the database [26-28]. The procedure of GMM clustering is illustrated in Fig. 2. In this subsection, database **Q** is illustrated as an example. The procedure of GMM clustering for **X**, **Y** and **Z** is similar, while the dimension *d* is 96, 24, 6 and 3 for **Q**, **X**, **Y** and **Z**, respectively.

The *d*-dimensional GMM $\Psi(\mathbf{Q})$ with *K* mixture components can be described as:

$$\Psi(\mathbf{Q}) = \sum_{k=1}^{K} \omega_k \phi_k(\mathbf{Q}|\mu_k, \Sigma_k) \tag{10}$$

$$\phi_k(\mathbf{Q}|\mu_k, \Sigma_k) = \left(2\pi|\Sigma_k|\right)^{-\frac{d}{2}} \exp\left\{-\frac{1}{2}(\mathbf{Q}-\mu_k)^T \Sigma_k^{-1}(\mathbf{Q}-\mu_k)\right\} \tag{11}$$

where

*K* quantity of mixture models

$\omega_k$ weight of the $k^{\text{th}}$ mixture component, $0 < \omega_k < 1$, and $\sum_{k=1}^{K} \omega_k = 1$

$\phi_k(\mathbf{Q}|\mu_k, \Sigma_k)$ Gaussian probability density function of the $k^{\text{th}}$ mixture component, where $\mu_k$ denotes the mean matrix, while $\Sigma_k$ is the covariance matrix. $\phi_k$ can specify the spread and orientation of the distribution:

$$\phi_k\left[\begin{pmatrix}\mu_1\\ \mu_2\\ \ldots\\ \mu_k\end{pmatrix}, \begin{pmatrix}\sigma_1^2\ \sigma_{12}\ \ldots\ \sigma_{1k}\\ \sigma_{21}\ \sigma_2^2\ \ldots\ \sigma_{2k}\\ \ldots\\ \sigma_{k1}\ \sigma_{k2}\ \ldots\ \sigma_k^2\end{pmatrix}\right]$$

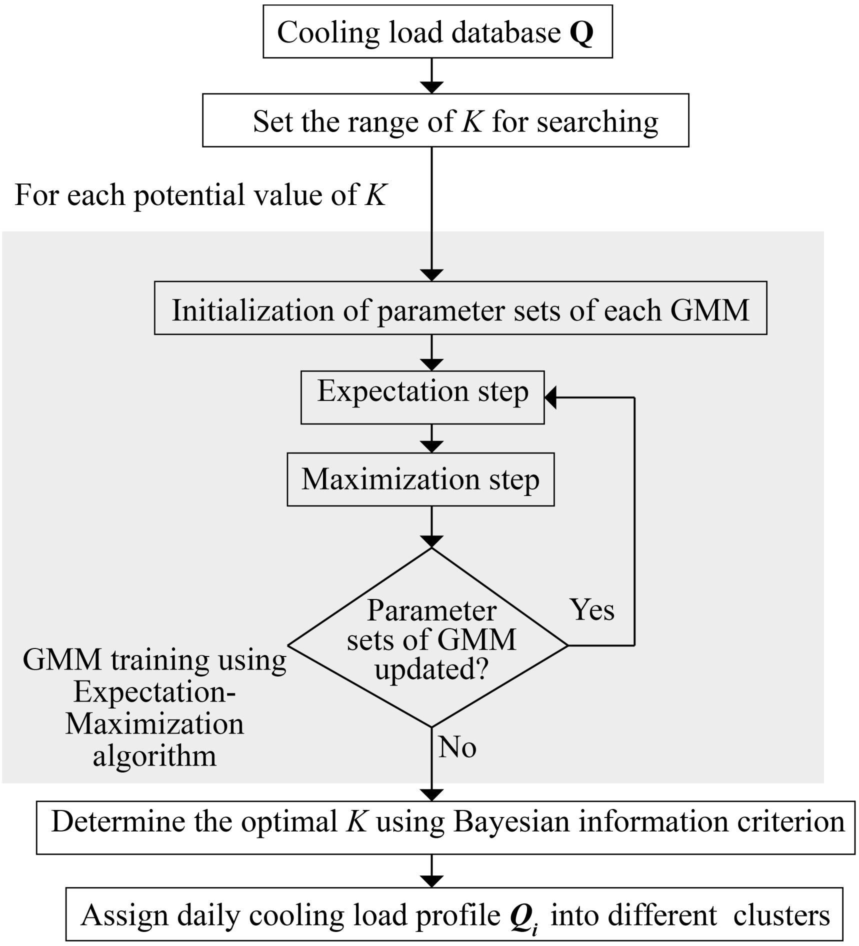


Fig. 2. Procedure of GMM clustering.

In this study, each cluster is assumed to be represented by either identical or different mixture models. Namely, for "true" case, all the mixture models share the same covariance matrix; while for "false" case, different covariance matrix are assigned to various mixture models. For both cases, both diagonal and full Gaussian distribution are studied.

Given the database **Q** and pre-set $K$, a GMM fitting is to estimate the values of parameter sets $\{\alpha_k, \mu_k, \Sigma_k\}(k = 1, 2, \ldots K)$ to ensure that the GMM has the maximum likelihood. The fitting procedure is carried out using Expectation Maximization algorithms, which consists of three steps: initialisation, expectation step and maximisation step. The initial values of parameter sets are randomly selected at first. After that, an iteration of expectation step and maximization step is conducted to improve the estimation of the parameter sets.

In the expectation step, each daily cooling load profile $\boldsymbol{Q_i}$ at the $i^{\text{th}}$ day is assigned to one of the mixture model $k$ which has the highest probability $\gamma_{ik}$. The posteriori probability $\gamma_{ik}$ is calculated based on the given values of parameter set $\{\alpha_k, \mu_k, \sum_k\}$:

$$\gamma_{ik} = \frac{\omega_k \phi_k(\boldsymbol{Q_i}|\mu_k, \Sigma_k)}{\sum_{k=1}^{K} \alpha_k \phi_k(\boldsymbol{Q_i}|\mu_k, \Sigma_k)} \tag{12}$$

Based on the $\gamma_{ik}$ obtained from the expectation step, the parameter sets $\{\alpha_k, \mu_k, \sum_k\}$ are updated in the maximisation step:

$$\omega_k = \frac{\phi_k}{N} \tag{13}$$

$$\mu_k = \frac{1}{\phi_k} \sum_{j=1}^{N} \gamma_{ik}\, \boldsymbol{Q_i} \tag{14}$$

$$\textstyle\sum_k = \frac{1}{\phi_k} \sum_{j=1}^{N} \gamma_{ik}\, (\boldsymbol{Q_i} - \mu_k)(\boldsymbol{Q_i} - \mu_k)^T \tag{15}$$

where $\phi_k = \sum_{i=1}^{N} \gamma_{ik}$, $N$ denotes the quantity of vectors in the database, and $N = 365$.

The expectation step and maximization step are iteratively computed until the updated parameters of all the mixture models do not change further. Therefore, the likelihood function is determined as:

$$L(\theta|\boldsymbol{Q}_i) = \sum_{i=1}^{N} \ln\left\{\sum_{k=1}^{K} \omega_k \phi_k(\boldsymbol{Q}_i|\mu_k, \textstyle\sum_k)\right\} \tag{18}$$

Once the GMM has been fitted, the observations belonging to the same mixture component are considered to be grouped in the same cluster.

### 2.2.2 Performance evaluation criterion

Bayesian Information Criterion (BIC) is the most widely adopted indicator for statistical model selection. Assuming that the input data is originally generated according to an unknown GMM (i.e. $\psi_{\text{true}}$) and the GMM is fitted using the generated input data (i.e. $\psi_{\text{candidate}}$), BIC is adopted to measure the difference between $\psi_{\text{true}}$ and $\psi_{\text{candidate}}$. Therefore, the GMM with an optimal $K$ value will result in the lowest *BIC*. The optimal $K$ value in the intra-building clustering used in this study was determined based on the range of 2-20:

$$BIC = -L(\theta|\boldsymbol{Q}_i) + \frac{A}{2}\log(N) \tag{19}$$

where $A$ is the total quantity of free parameters, and can be calculated as:

$$A = Kd + \frac{1}{2}Kd(d+1) + (K-1) \tag{20}$$

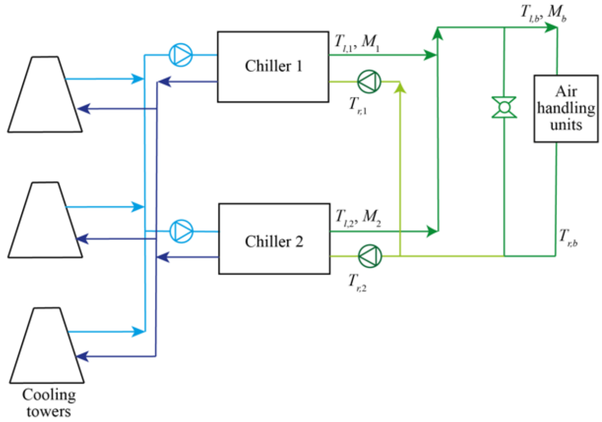
### *2.3 Data visualization*

To investigate the clustering results of year-round cooling profile, the calendar view is adopted to illustrate the load distribution of the whole year. Meanwhile, through plotting the cooling load profile based on the cluster number, the representative and characteristics of each cluster can be identified.

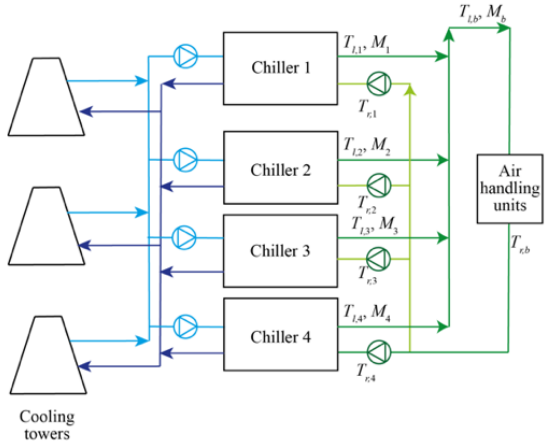
## 3. Chilled water system under study

To keep track of the changing cooling load, chilled water systems must respond by varying chilled water flow rate or chilled water temperature differential. Variable chilled water pumps and bypass valve are two representative approaches to adjust the cooling load of the water-cooled chilled water system. Variable chilled water pumps can be used to adjust the flow rate of chilled water while bypass valve is generally adopted to change the chilled water temperature differential.

To test the performance and demonstrate the generality of the proposed pattern identification approach, it is implemented on two typical water-cooled chilled water systems installed in a campus building and a hotel building, respectively. The basic building information is summarized in Table 1. The chilled water system 1 used in the high-rise campus building consisted of two identical centrifugal water-cooled chillers. The differential pressure bypass was adopted in chilled water system 1 to allow surplus chilled water passing through when the air handling units were under part-load operation. The chilled water system 2 adopted in the high-rise hotel building was comprised of four identical chillers, while the maximum operating quantity of chillers was two. Each chiller had an associated chilled water pump and a cooling water pump, respectively. The variable chilled water pumps were equipped in the chilled water system 2. The same type of water-cooled chillers and cooling towers were installed in both chilled water systems. The schematic diagrams of the two chilled water systems are shown in Fig. 3, while its key design parameters are summarized in Table 2.




(a) Chilled water system 1 in the campus building.




(b) Chilled water system 2 in the hotel building.

Fig. 3. Schematic diagram of the chilled water systems.

Table 1. Building information

| | | |
|---|---|---|
| Building type | Campus | Hotel |
| Number of floors | 37 | 15 |
| Floor area | 63500 | 160000 |
| Location | Hong Kong Island | Hong Kong Island |

Table 2. Key design parameters of the chilled water system.

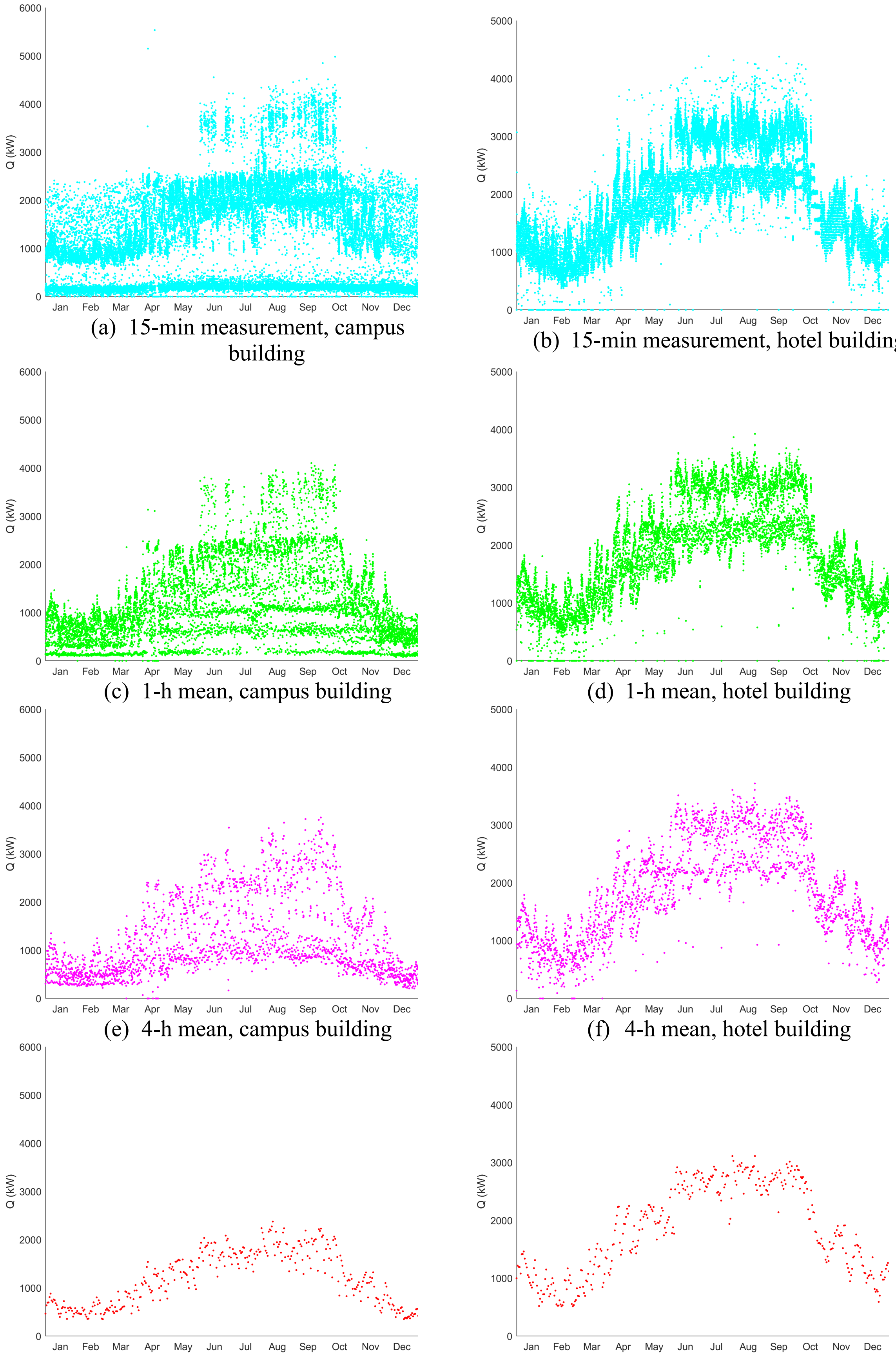
| | |
|---|---|
| Rated cooling capacity of each chiller (kW) | 2103 |
| Rated chilled water flow rate of each chiller (L/s) | 100.4 |
| Set-point of chilled water leaving temperature (°C) | 7 |
| Design chilled water return temperature (°C) | 12 |
| Design *COP* | 6.2 |
| Rated condenser water flow rate of each chiller (L/s) | 110.0 |
| Set-point of cooling water leaving temperature (°C) | 37.37 |
| Design cooling water return temperature (°C) | 32.0 |

## 4. Results and analysis

In this section, the performance of each stage of the proposed 3D pattern identification approach was evaluated, respectively.

### *4.1 Performance assessment of data preparation*

To assess the performance of data collection, data cleaning and data pre-processing, the year-round cooling load of the two buildings were investigated. The 15-min measurement, 1-h mean, 4-h mean, and daily average cooling load of the two buildings are shown in Fig. 4.




(a) 15-min measurement, campus building

(b) 15-min measurement, hotel building

(c) 1-h mean, campus building

(d) 1-h mean, hotel building

(e) 4-h mean, campus building

(f) 4-h mean, hotel building

(g) Daily brief, campus building

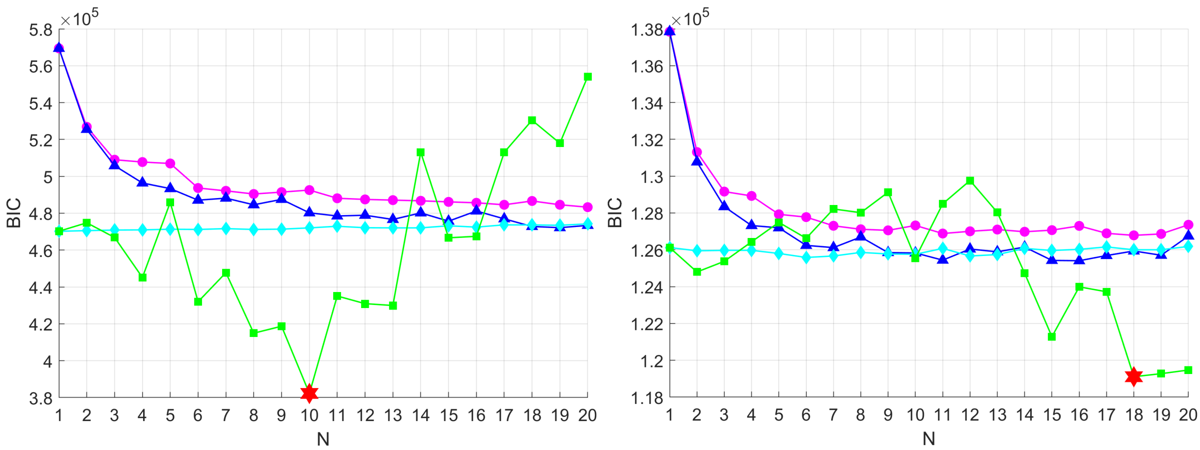
(h) Daily brief, hotel building

Fig. 4 Year-round cooling load of the two buildings.

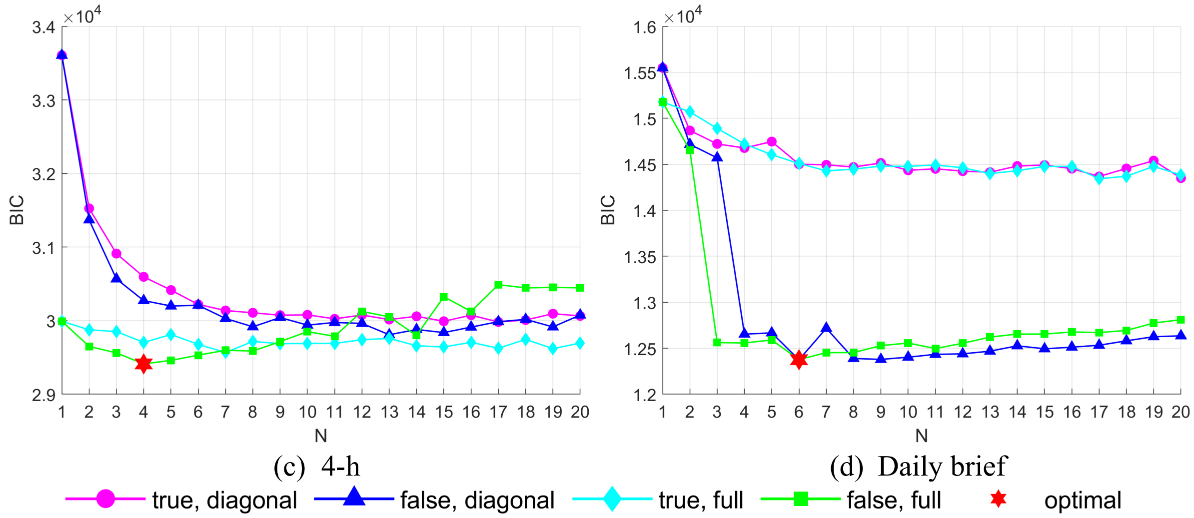
For both the campus and hotel buildings, cooling loads were higher from mid-May till mid-October. From the 15-min measurement, it can be seen that the variation of the cooling load in the campus building was quite large in each day. In other words, the cooling load tended to be 0 during night time. For the campus building, there were approximately three types of cooling load during mid-May to mid-October, while two types in other period of the year. For the hotel building, there were about two categories of cooling load from mid-May to mid-October, while only one type in other time periods.

### *4.2 Performance evaluation of GMM clustering*

The Bayesian Information Criterion values introduced in Section 2.2.2 were adopted to evaluate the GMM clustering performance. The variation between BIC values and quantity of components in the Gaussian mixture model *K* under different cases are summarized in Figs. 5 and 6. For each building under four different cases (i.e. shared covariance is true while sigma is diagonal, shared covariance is false while sigma is diagonal, shared covariance is false while sigma is full, shared covariance is false while sigma is full), the *BIC* values were evaluated. For both the campus and the hotel buildings, the optimal *K* was found when shared covariance was false while sigma was full. For the campus building, the smallest *BIC* value was $4.20 \times 10^5$, $1.19 \times 10^5$, $2.94 \times 10^4$ and $1.24 \times 10^4$ when *K* was 9, 18, 4 and 6 for 15-min, 1-h, 4-h, and daily cooling load, respectively. For the hotel building, the smallest *BIC* value was $3.80 \times 10^5$, $1.06 \times 10^5$, $2.84 \times 10^4$ and $1.45 \times 10^4$ when *K* was 10, 19, 5 and 5 for 15-min, 1-h, 4-h, and daily cooling load, respectively. Therefore, the similar smallest *BIC* value and optimal *K* was identified for the two cooling load profiles under the same time-step value.

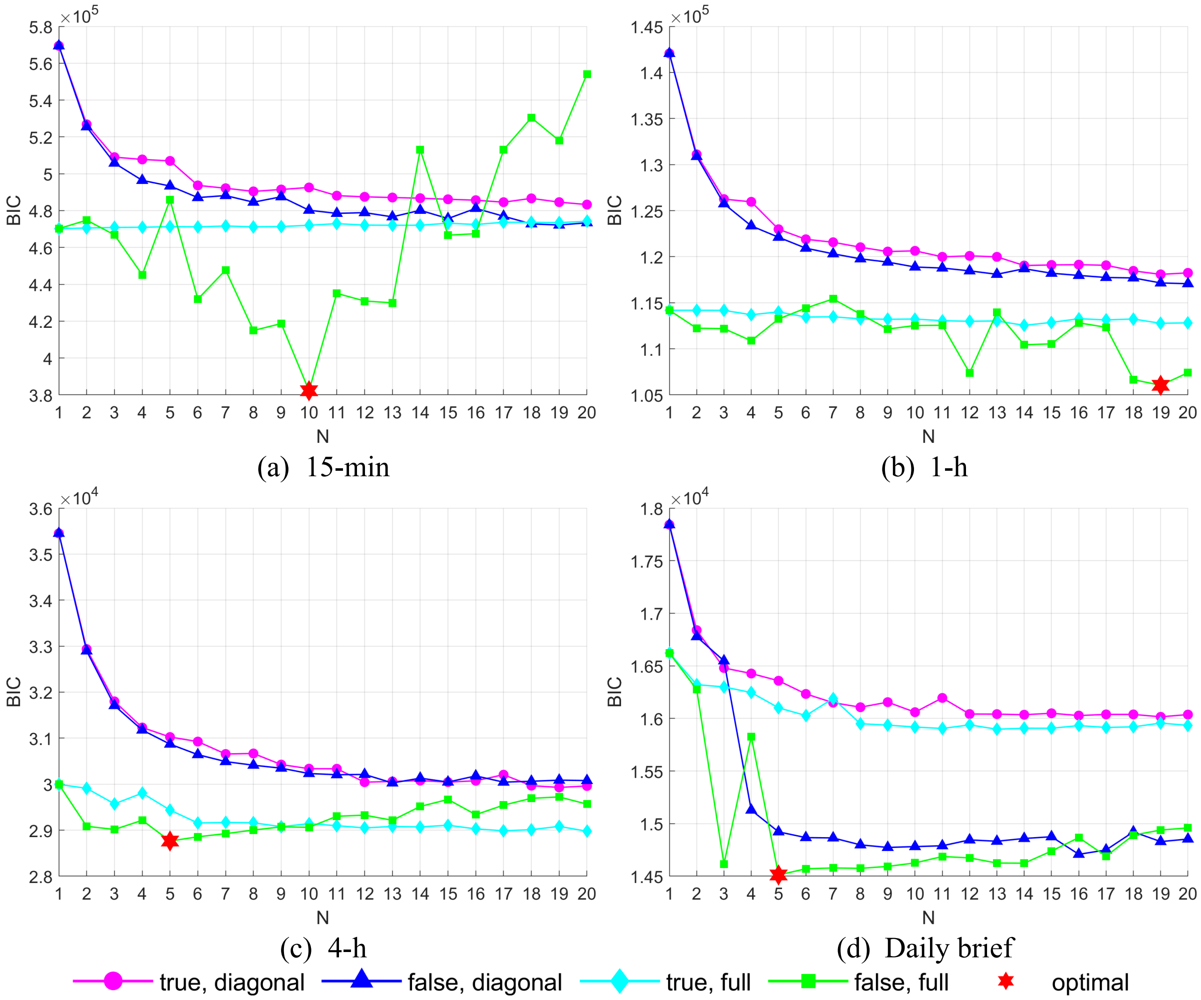


(a) 15-min (b) 1-h




(c) 4-h (d) Daily brief

Fig. 5. *BIC* values of campus building for different time steps.




(a) 15-min (b) 1-h

(c) 4-h (d) Daily brief

Fig. 6. *BIC* values of hotel building for different time steps.

### *4.3 Data visualization and analysis*

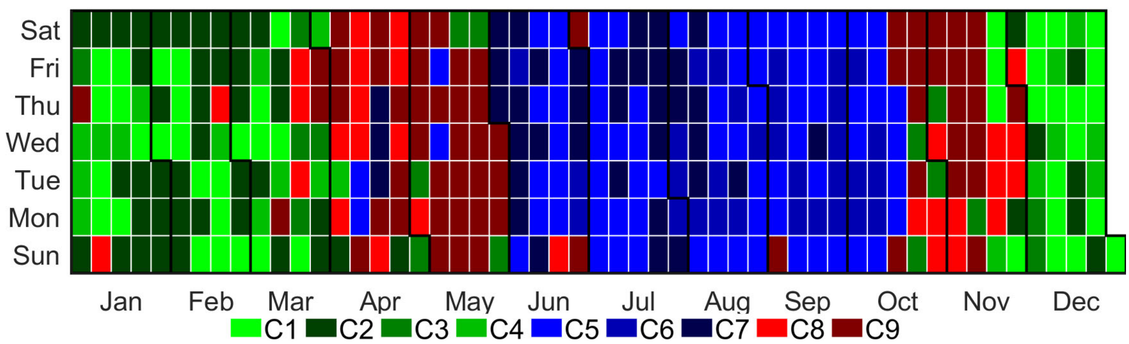
To investigate the characteristics of the year-round cooling load of the campus and the hotel building, the distribution of the cooling load profile in a calendar view was visualised to explore the different load distribution among the whole year while the representative cooling load in each cluster were identified through plotting the clustered cooling load profiles.

### 4.3.1 Chilled water system in the campus building

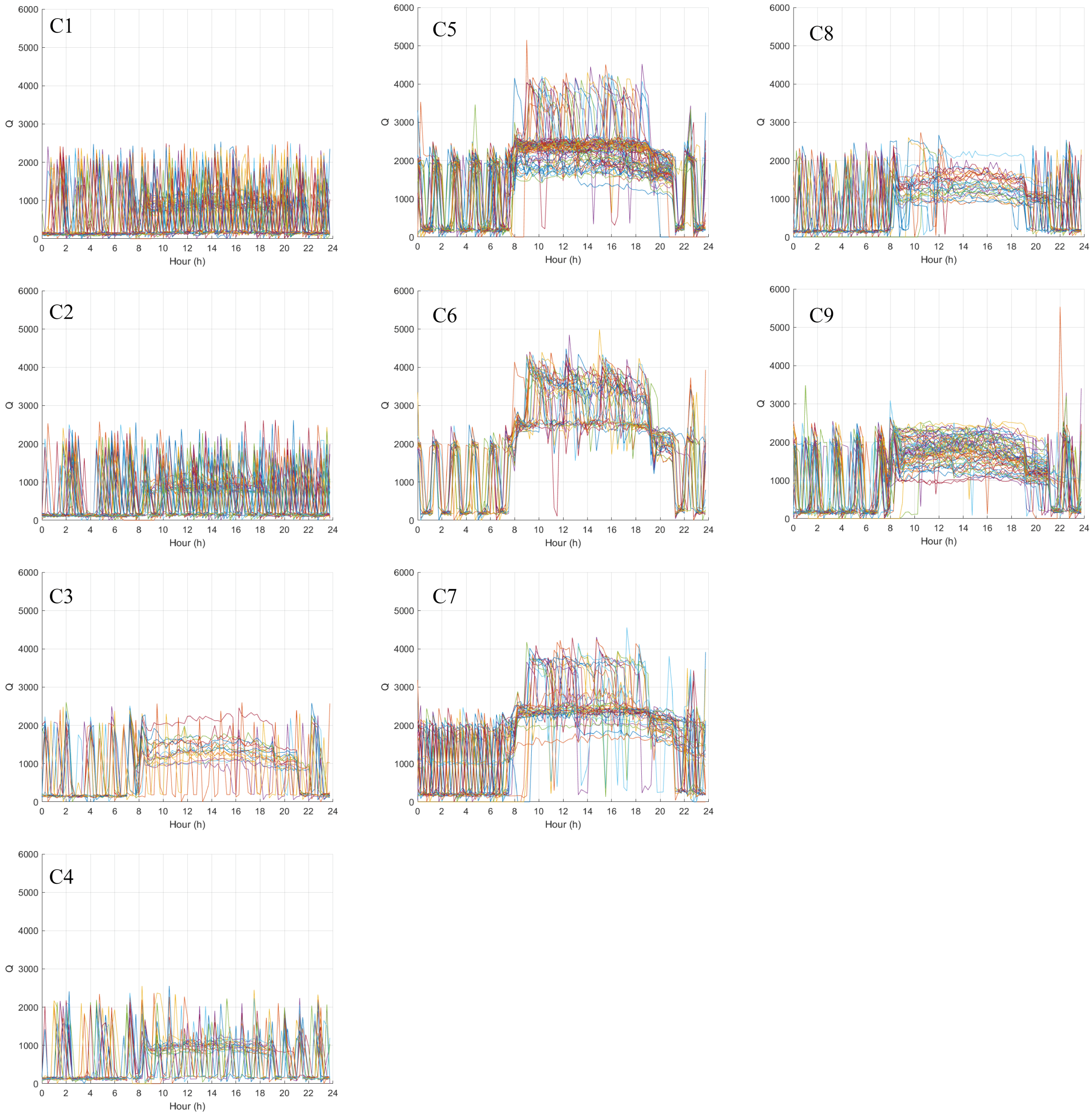
The clustering results of the 15-min measurement, 1-h mean value, 4-h mean value and daily brief are summarized in this section.

As shown in Fig. 5(a), the optimal *K* value was 9 the 15-min measurement of daily cooling load profile was grouped into 9 different clusters. The cluster distribution at each day is shown in Fig. 7(a), while the grouped daily cooling load profile in different clusters are summarized in Fig. 7(b). It is found that:

- Clusters C1-C4 were mainly distributed among the cold season (i.e. January-March and December), during which the cooling load varied from 0 to 2000 kW. Due to the relatively low cooling load during the cold season, the chillers were continuously switched on-and-off, which thus resulted in the persistent variation of cooling load in each day.
- Clusters C5-C7 were generally found in the hot season (i.e. June to early October). During the night time, there existed continuous variation of cooling load between 0 to 2000 kW. During the daytime (8-19 h), it jumped to about 4000 kW, and varied between 2000 to 4000 kW.
- Clusters C8 and C9 represented the mid-season (i.e. April, May, late October and November). During the mid-season, there also existed continuous variation of cooling load between 0 to 2000 kW during night-time, while it was relatively constant between 1000 to 2000 kW during the daytime (8-19 h).



(a) Calendar view of cooling load distribution.




(b) Clustered cooling load profiles.

Fig. 7 Clustering results of the campus building at 15-min time step.

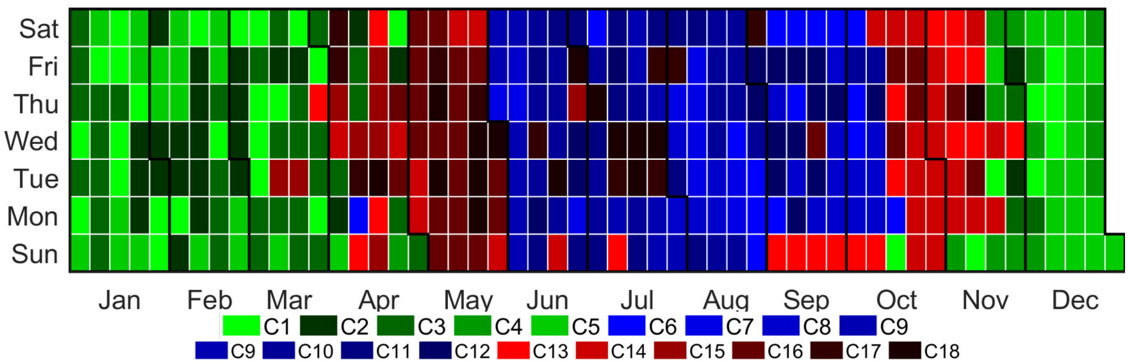
As shown in Fig. 5(b), the optimal *K* value was 18 thus the 1-h mean value of daily cooling load profile was grouped into 18 different clusters. The cluster distribution at each day is shown in Fig. 8(a), while the daily cooling load profile in corresponding clusters are shown in Fig. 8(b). It is found that:

- Clusters C1-C5 were mainly distributed among the cold season (i.e. January-March and December), while there were more days assigned into the cold season compared to those from 15-min measurement. In the cold season, the 1-h mean value of cooling load varied from 0 to 1200 kW. Due to the relatively low cooling load during the cold season, the chillers were continuously switched on-and-off, which resulted in the continuous waves in each day.
- Clusters C6-C12 were generally found in the hot season (i.e. June to early October). Compared to the 15-min measurement, there were fewer days grouped into the hot season. During this period,

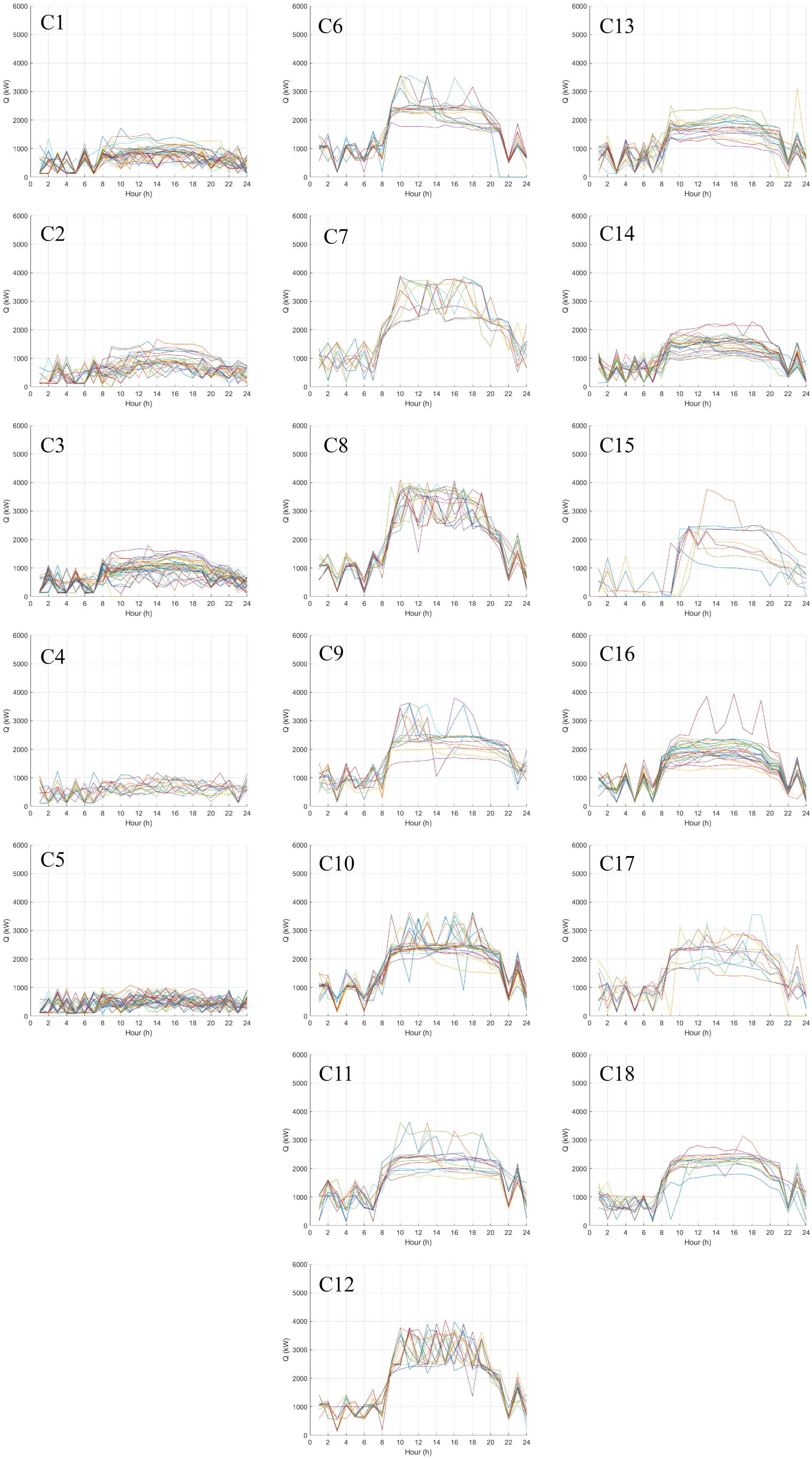
there also existed continuous variation between 0 to 1200 kW during night-time, while the 1-h mean value of cooling load was approximately between 2200 to 3800 kW during daytime (8-20 h).

- Clusters C13-C18 represented the mid-season (i.e. April, late October and November). There were more days clustered into the mid-season compared to that from the 15-min measurement, especially the Sundays in September. During this period, the 1-h mean value of cooling load was low during night-time while almost constant at about 1800 kW during daytime (8-20 h).

As shown in Fig. 5(c), the optimal *K* value was 4 while the 4-h mean value of daily cooling load profile was grouped into 4 different clusters. The cluster distribution at each day is shown in Fig. 9(a), while the daily cooling load profile in corresponding clusters are shown in Fig. 9(b). Different from the 15-min measurement and 1-h mean value, the whole year cooling load profile was briefly divided into two seasons: cold and hot seasons. It is found that most of the days in the cold season (i.e. January to March, late November and December) were grouped into the single cluster C1, during which the 4-h mean value of cooling load was around 200 to 1200 kW. Meanwhile, most of the days during the hot season (i.e. April-early November) are grouped into Clusters C2 and C3, during which the 4-h mean value of cooling load was around 1000 kW at night and 2000 kW at daytime, respectively. Cluster C4 stood for few days during which there existed a decreasing trend from 2-6 h. Moreover, the 4-h mean value of cooling load profiles in each group were more similar to each other in the respective clusters. It is because that the effects of abrupt variation caused by the continuously switching on-and-off were minimized through the long time-step average.

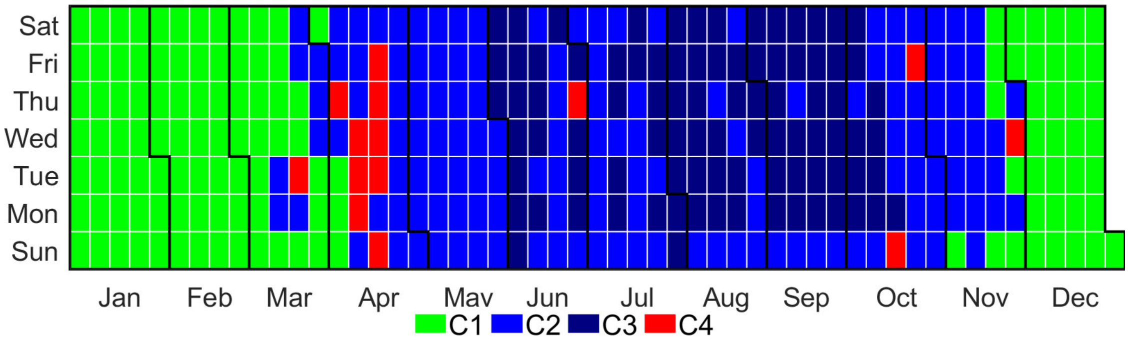


(a) Calendar view of cooling load distribution.

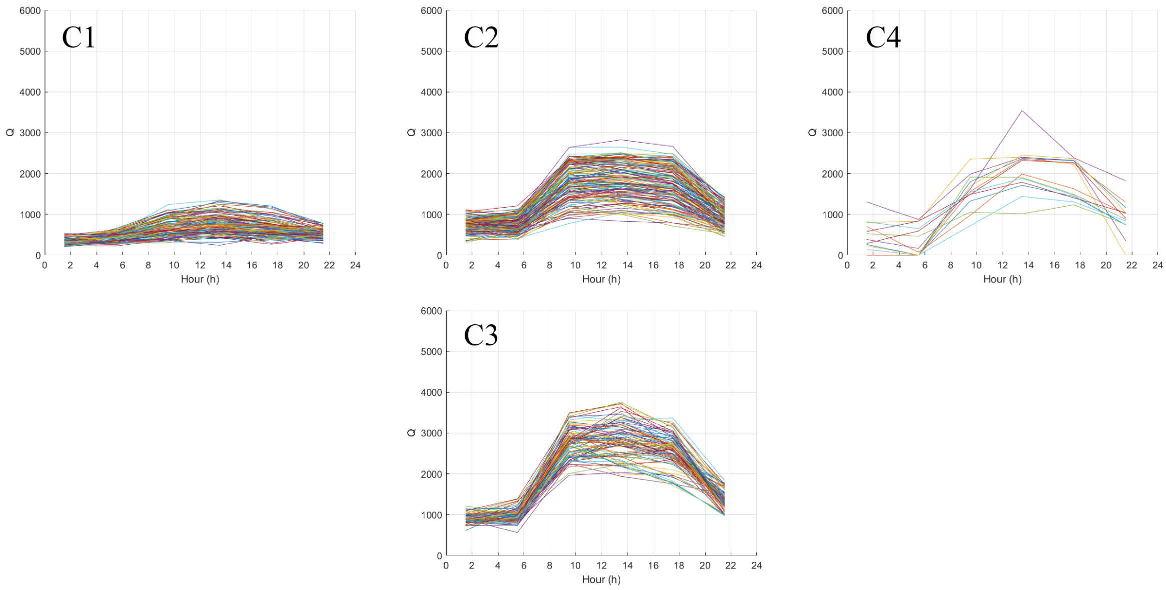



(a) Clustered cooling load profiles.

Fig. 8. Clustering results of the campus building of 1-h mean value.



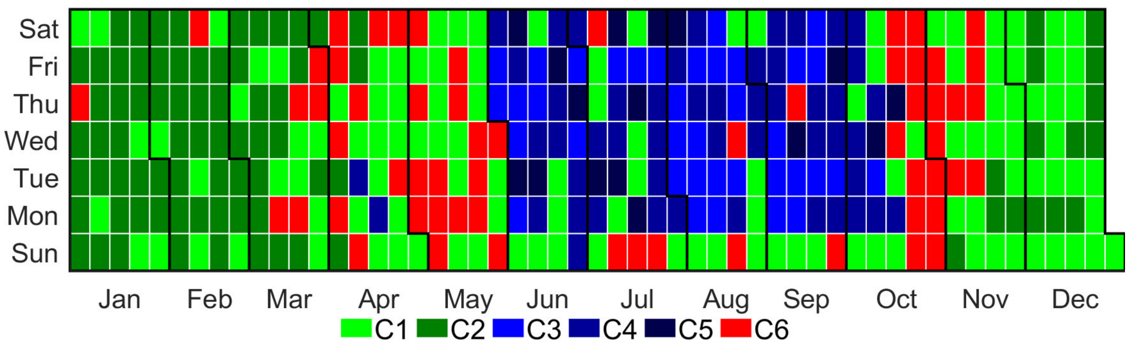

(a) Calendar view of cooling load distribution.



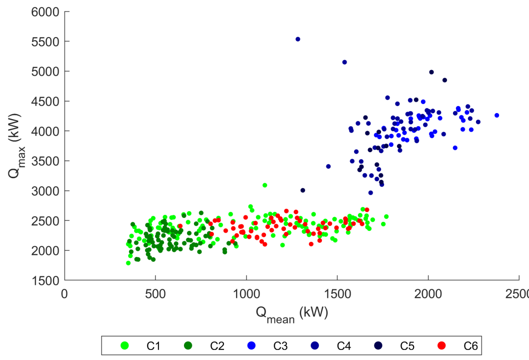

(a) Clustered cooling load profiles.

Fig. 9. Clustering results of the campus building of 4-h mean value.

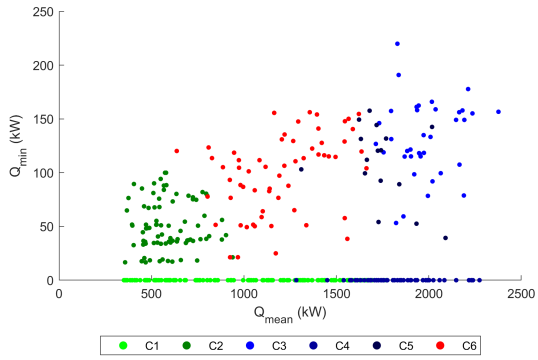
As shown in Fig. 5(d), the optimal *K* value was 6 while the sets of daily average, minimum and maximum cooling load was grouped into 6 different clusters. The cluster distribution at each day is shown in Fig. 10(a), while the daily minimum-average and maximum-average values of cooling load are shown in Fig. 10(b) and (c), respectively. Similar to the 4-h mean value of cooling load, the whole year was also roughly divided into two seasons: cold and hot seasons. It is found that Clusters C1 and C2 represented the relatively cold months (i.e. January-May, November and December), in which the daily average, minimum and maximum cooling load were in the range of [400, 900], [0, 100], [1700, 2700], respectively. Clusters C3-C5 stood for the relatively hot months (i.e. June-October), while the daily average, minimum and maximum cooling load were in the range of [1700, 2300], [50, 250], [3500, 4500], respectively. Cluster C6 was spread among both hot and cold seasons. From Fig. 10(b), the daily minimum value of cooling load was higher than that in the cold season while lower than that in the hot season. However, the daily maximum value of cooling load was similar to that in the cold season.




(a) Calendar view of cooling load distribution.




(a) Max-mean cooling load.




(b) Min-mean cooling load.

Fig. 10. Daily average, minimum and maximum cooling load of the campus building.

#### 4.3.2 Chilled water system in the hotel building

The clustering results of the 15-min measurement, 1-h mean value, 4-h mean value and daily brief are summarized in this section.

As shown in Fig. 6(a), the optimal *K* value was 10, therefore, 15-min daily cooling load profile was grouped into 10 different clusters. The cluster distribution at each day is shown in Fig. 11(a), while the daily cooling load profile in corresponding clusters are shown in Fig. 11(b). It is found that:

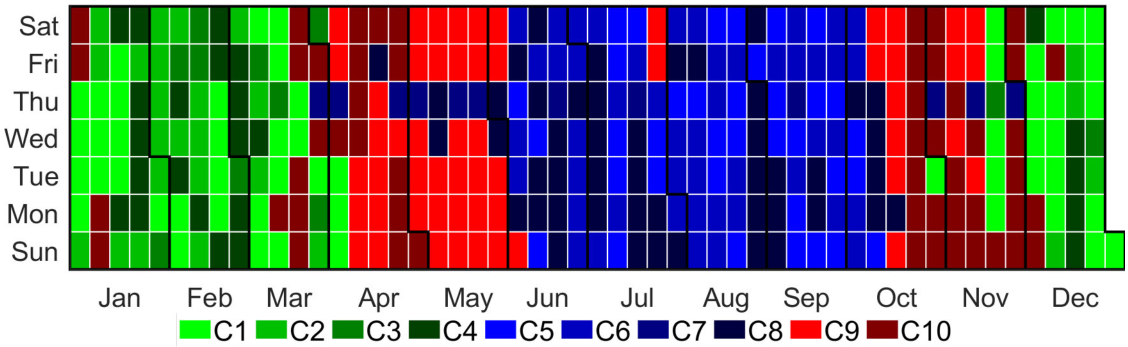
- Clusters C1-C4 were mainly distributed among the cold season (i.e. January-March, November and December), during which the cooling load was around 1000 kW from 7-24 h. Compared to that in the campus building, the constant cooling load period was longer while the frequency of switch on-and-off was lower.
- Clusters C5-C8 represented the hot season (i.e. June to early October). During this period, the cooling load was about 2200 kW and 3000 kW during 0-6 h and 6-23 h, respectively. During this season, there seemed to be less frequency of switching on-and-off chillers.
- Clusters C9 and C10 were generally found in the middle season (i.e. April, May and late). Similar to that in the hot season, there was less frequency of switching on-and-off in this season. Moreover, in Cluster C9, the cooling load was about 1600 kW and 2200 kW during 0-7 h and 7-24 h, respectively; in Cluster 10, the cooling load was about 1200 kW and 1800 kW during 0-7 h and 7-

24 h, respectively. In other words, the cooling load difference between night-time and day-time was less than that in the hot season.

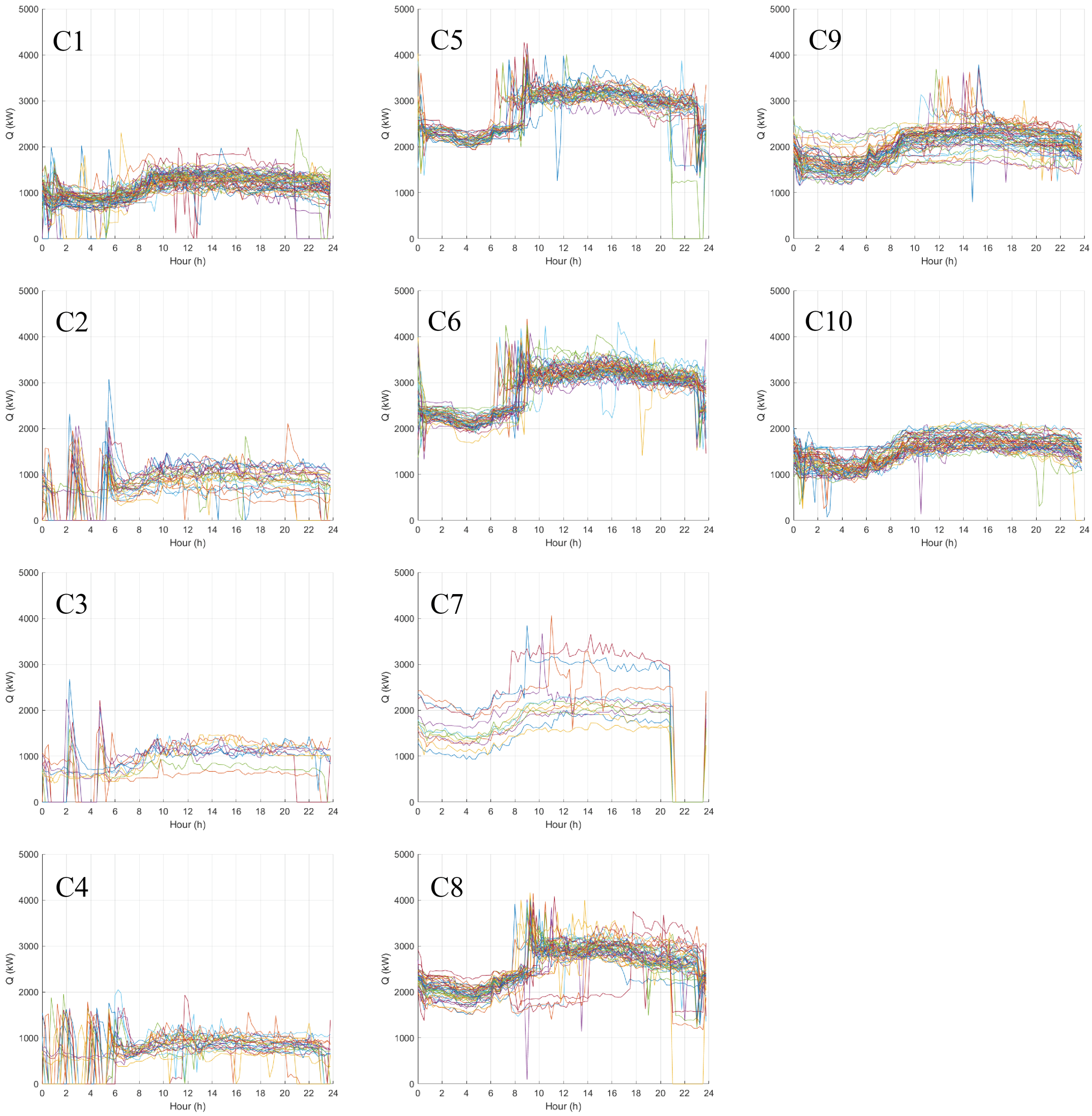
As shown in Fig. 6(b), the optimal *K* value was 19 thus the 1-h mean value of daily cooling load profile of the hotel building was grouped into 19 different clusters. The cluster distribution at each day is shown in Fig. 12(a), while the daily cooling load profile in corresponding clusters are shown in Fig. 12(b). It is found that:

- Clusters C1-C7 were mainly distributed among the cold season (i.e. January-March, late November and December), during which the cooling load varied from 0 to 1200 kW. Due to the relatively low cooling load during the cold season, the chillers were continuously switched on-and-off during some of the days during 0-7 h, which resulted in the waves of 1-h mean value of cooling load. During 7-24 h, the cooling load was relatively constant at 1100 kW.
- Clusters C8-C15 were generally found in the hot season (i.e. June to early October). During this period, the cooling load was about 2000 kW during night, while it increased to 3000 kW from the 9-22 h, and dropped to 2000 kW after the 23$^{rd}$ h.
- Clusters C16-C19 represented the mid-season (i.e. April, late October and early November). During this period, the cooling load was about 1500 kW during night, while it increased to 2000 kW during daytime (7-24 h).

As shown in Fig. 6(c), the optimal *K* value was 5 while the 4-h mean value of daily cooling load profile was grouped into 5 different clusters. Different from the 15-min measurement and 1-h mean value of cooling load, the whole year was briefly divided into two seasons: cold and hot seasons. Most of the days in the cold season (i.e. January-May, late October-December) were grouped into Clusters C1 and C2, during which the 4-h mean value of cooling load was around 500 to 2500 kW. Most of the days during the hot season (i.e. June-early October) were grouped into Clusters 3 and 4, during which the cooling load was around 2200 kW and 3000 kW at night-time and daytime, respectively. Cluster C5 stood for few days during which there existed a decreasing trend at the end of each day. Most of the days which were grouped into Cluster C5 were Thursdays. During these days, the cooling load dropped to 0 during 21-24 h. This might be caused by the pre-scheduled maintenance of the related system. Moreover, the 4-h mean value of cooling load in each group were more similar to each other in the respective clusters. It was because that the effects of abrupt variation caused by the continuously switching on-and-off were minimized through the long time-step average.

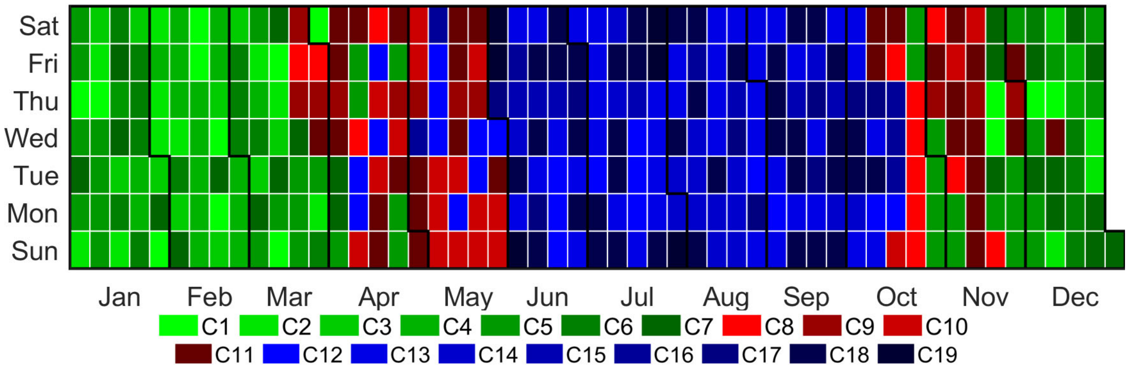


(a) Calendar view of cooling load distribution.



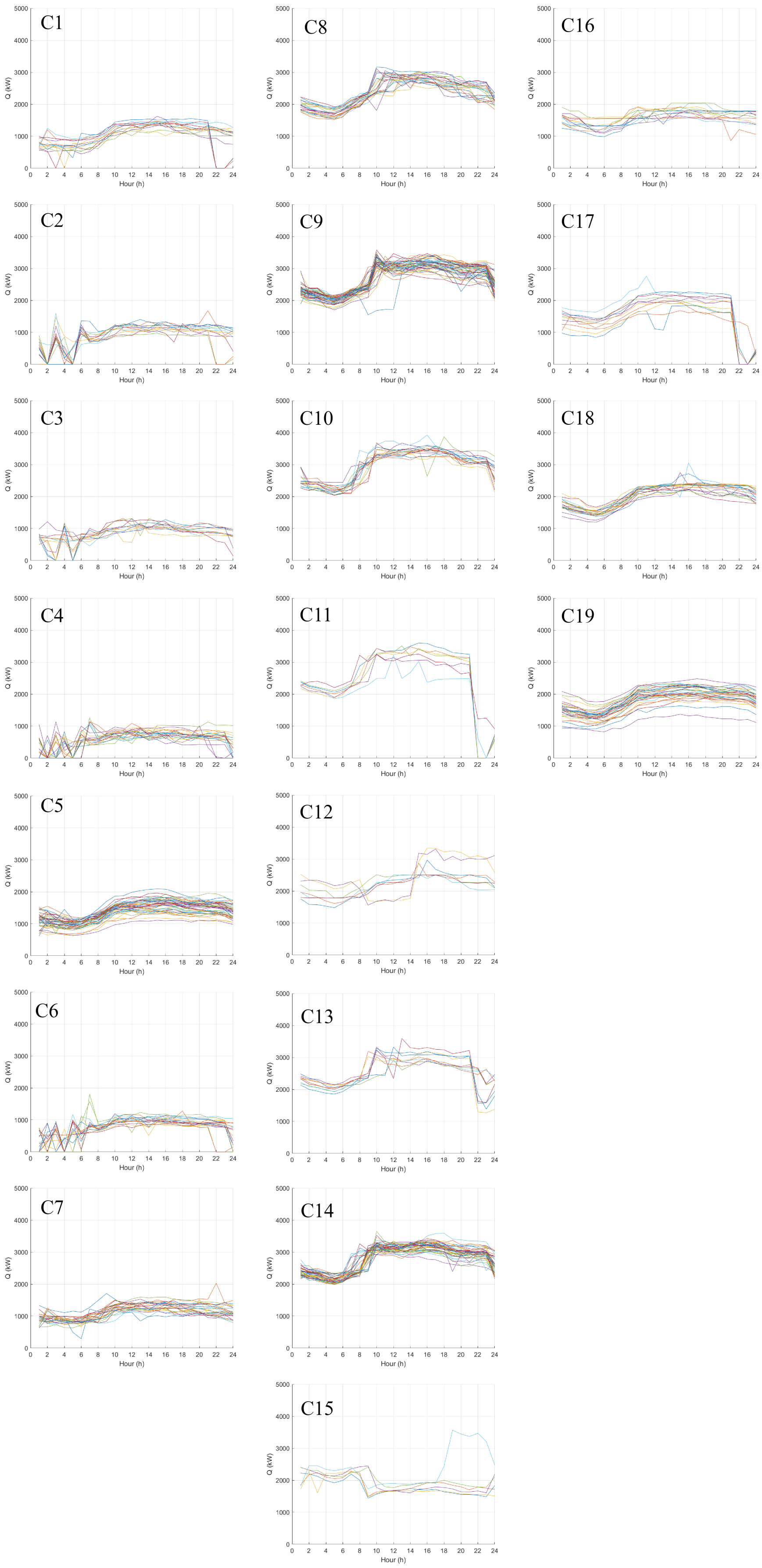
(a) Clustered cooling load profiles

Fig. 11. Clustering results of the hotel building at 15-min time step.



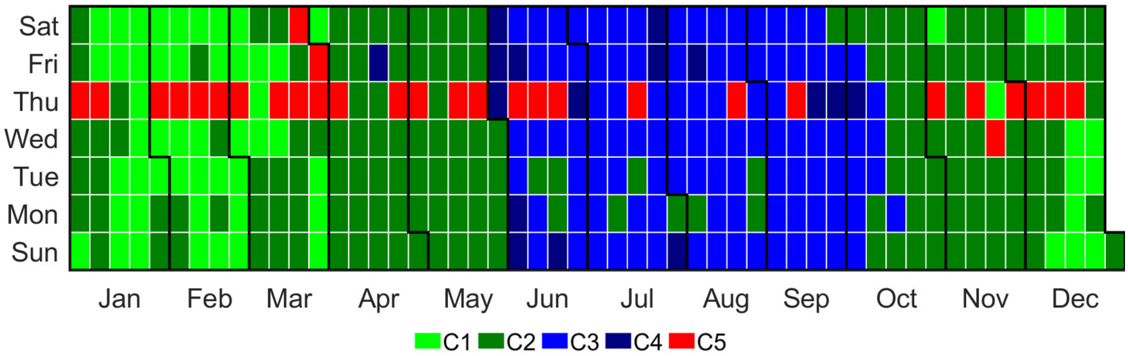

(a) Calendar view of cooling load distribution.

As shown in Fig. 6(d), since the optimal $K$ value was 5, the sets of daily average, minimum and maximum cooling load was grouped into 5 different clusters. The cluster distribution at each day is shown in Fig, 14(a), while the daily minimum-average and maximum-average values of cooling load are shown in Fig. 14(b) and (c), respectively. Similar to the 4-h averaged cooling load, the whole year was also roughly divided into two seasons: cold and hot seasons. It is found that Clusters C1-C4 represented the relatively cold months (i.e. January-May, November and December), in which the daily average, minimum and maximum cooling load were in the range of [600, 2400], [0, 1800] and [1000, 2500], respectively. Cluster C5 stood for the relatively hot months (i.e. June-October), in which the daily average, minimum and maximum cooling load were in the range [2400, 3000], [1200, 2300] and [3200, 4500], respectively. Moreover, due to the switching on-and-off of chillers at the 21$^{st}$ h, those mentioned Thursdays in C5 from 4-h mean cooling load data were clustered into C1 due to the low minimum and average cooling load in each day.

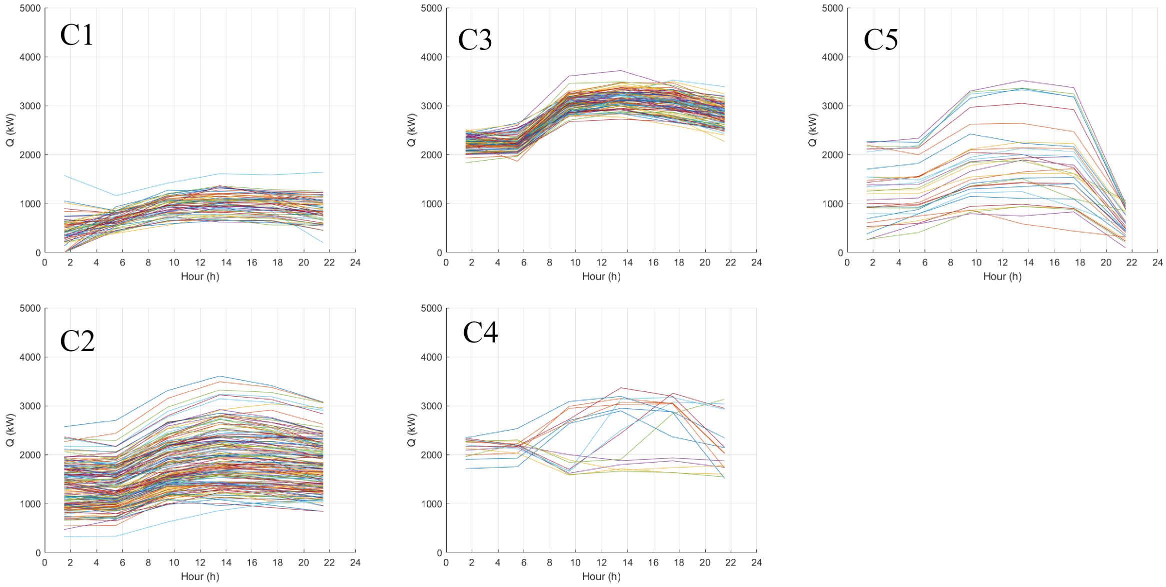



(b) Clustered cooling load profiles.

Fig. 12. Clustering results of the hotel building of 1-h mean cooling load.

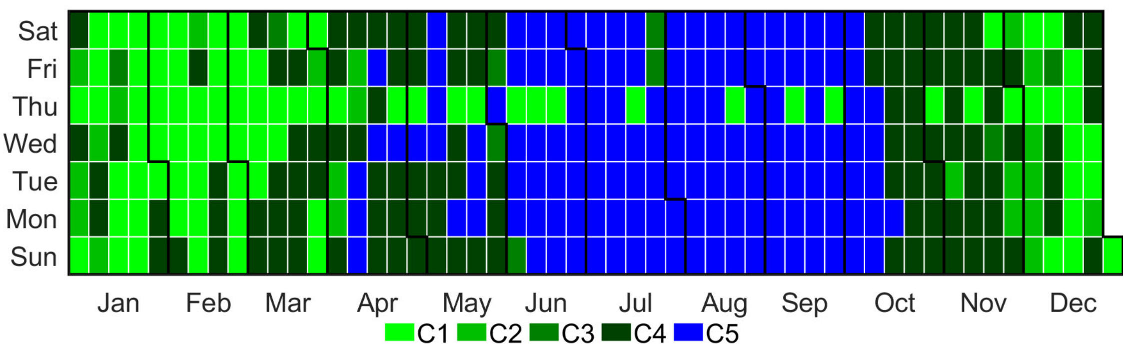



(a) Calendar view of cooling load distribution.

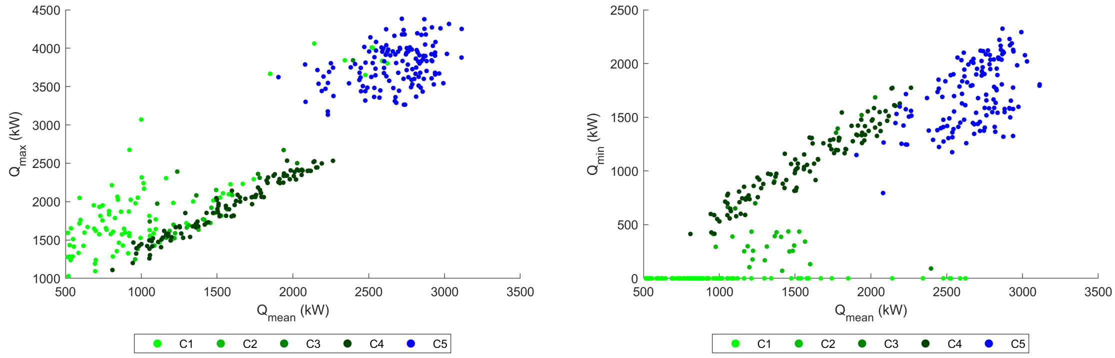



(a) Clustered cooling load profiles

Fig. 13. Clustering results of the hotel building at 4-h time step.




(a) Calendar view of cooling load distribution.




(b) Max-mean cooling load (a) Min-mean cooling load

Fig. 14. Clustering results of hotel building for daily average cooling load.

### *4.4 Summary of the proposed 3D pattern identification approach*

In this study, the 3D pattern identification approach was proposed for the building cooling loads.

- At the first stage, the historical measurement data of the chilled water system was collected from the building management system, while data cleaning was conducted to improve the quality of raw measurement by excluding the missing and outlier values. Meanwhile, measurement data was processed at different time steps (i.e. the 1-h mean value, 4-h mean value, daily average, minimum and maximum value) to evaluate the different trends of the daily cooling load profile.
- At the second stage, GMM clustering was conducted to group the 15-min measurement, 1-h mean value, 4-h mean value and the daily brief value of the daily cooling load profile into different clusters.
- At the third stage, data visualization of clustering group in a calendar view and the grouped daily cooling load profile was implemented to investigate the characteristics of the cooling load profile around the whole year.

Through the proposed pattern identification approach, the similarity and differences of the cooling load among the campus and the hotel building were found and summarized in Table 3.

From the above analysis, it is seen that the 15-min measurement could provide the detailed trends of the daily cooling load, the 1-h mean value was able to illustrate the overall trend of the daily cooling profile, the 4-h mean could show the general trend of the building cooling load, and the daily brief load profile demonstrated the maximum, minimum and average cooling load in different seasons. For both the campus and the hotel building, the 1-h mean value resulted in the most clusters while both the 4-h mean value and daily brief values resulted in the fewest clusters through GMM clustering. By partially avoiding the frequent changes of cooling load, the 1-h mean value resulted in the most accurate

Table 2. Summary of the similarity and differences among the two chilled water system.

| Building type | Load profile | Clusters | Load characteristics | | |
|---|---|---|---|---|---|
| | | | Cold season | Middle season | Hot season |
| Campus building | 15-min measurement | 9 | Jan-Mar, Dec: 0-2000 kW, continuously on-and-off | Apr, May, late Oct, Nov: 0-2000 kW during night, continuously on-off; 1000-2000 kW during daytime (8-20 h). | June-early Oct: 0-2000 kW during night, continuously on-off; 2000-4000 kW during daytime (8-20 h). |
| | 1-h mean | 18 | Jan-Mar, Dec: 0-1200 kW, continuously on-and-off | Apr, May, late Oct, Nov: 0-1000 kW during night, continuously on-off; 1000-2200 kW during daytime (8-20 h). | June-early Oct: 0-1200 kW during night, continuously on-off; 2200-3800 kW during daytime (8-20 h). |
| | 4-h mean | 4 | Jan-Mar, late Nov, Dec: around 200-1200 kW | N.A. | Apr-early Nov: around 1000 kW during night while 2000-2800 kW during daytime (8-20 h) |
| | Daily brief | 6 | Jan-May, Nov, Dec:<br>Average [400-900]<br>Minimum [0-100]<br>Maximum [1700-2700] | N.A. | June-Oct:<br>Average [1700-2300]<br>Minimum [50-250]<br>Maximum [3500-4500] |
| Hotel building | 15-min measurement | 10 | Jan-Mar, late Nov, Dec: continuously on-off during 0-7 h, while relatively constant at 1000 kW during 7-24 h. | April, May and late Oct: relatively constant around 1600 kW during night (0-7 h), while around 1800-2000 kW during daytime (7-24 h). | June-early Oct: relatively constant around 2000 kW during night (0-7 h), while around 3000 kW during daytime (7-24 h) |
| | 1-h mean | 19 | Jan-Mar, late Nov, Dec: almost constantly at 1100 kW during 7-24 h, while several ups-and-downs during 0-7 h. | April, May, late Oct and early Nov: about 1500 kW during night (0-7 h), while 2000 kW during daytime (7-24 h). | June-early Oct: about 2000 kW during night (0-7 h), while 3000 kW during daytime (7-24 h). |
| | 4-h mean | 5 | Jan-May, late Oct-Dec: around 500-2500 kW. | | June-early Oct: about 2200 kW during night (0-7 h) while 3000 kW during daytime (7-24 h). |
| | Daily brief | 5 | Jan-May, Nov and Dec: Average [600-2400]<br>Minimum [0-1800]<br>Maximum [1000-2500] | | June-Oct:<br>Average [2400-3000]<br>Minimum [1200-2300]<br>Maximum [3200-4500] |

clustering thus could explore the most of characteristics of the daily cooling load profile. Moreover, 4-h mean value and daily brief value is also helpful when simpler and quicker analysis is needed.

It is also discovered that different characteristics of the cooling load can be identified in hot, cold and middle seasons, respectively:

- During the cold season, there usually existed ups-and-downs of the cooling load during night in the two chilled water system, while it was relatively constant during daytime;

- During the middle and hot seasons, there were still some ups-and-downs during the night in the campus building, while the cooling load was relatively constant during daytime and night in the hotel building;
- For the campus building, the daytime was around 8-20 h, while the daytime was much longer in the hotel building: 7-24 h.

From the above analysis, it is seen that when the cooling load was relatively low, there existed a lot of ups-and-downs of the cooling load. In other words, the chillers were working at low part-load-ratio during most of the time, which resulted in low energy conversion efficiency. It is suggested that chilled water storage could be installed to shift the peak cooling demand and help the chillers operate at relatively constant load.

## 5. Implication for practice and future direction

In this study, a 3D pattern identification approach was proposed for building cooling loads. Due to the limitation of available historical data, only one-year measurement data was adopted as the database in this study. In practical application, several years' measurement data can be adopted as the database of the proposed pattern identification approach to further improve its accuracy and effectiveness.

Based on the recorded profile of historical operating data of the chilled water system, the proposed pattern identification approach can be adopted to discover the characteristics and identify the representative cooling loads in campus and hotel buildings. Through effective identification of representative daily cooling load profile,

- Energy-efficient system management can be implemented: During low cooling load period, chilled water storage or other methods should be adopted to avoid frequent switching on-and-off of chillers hence to improve chiller operating efficiency;
- Effective demand and supply side management can be adopted: Due to the different characteristics of cooling loads in different types of buildings, the power grid could make personalized intervention

strategies for users during different periods from the point of view of demand and supply side management;

- Fault detection and diagnosis can be implemented for the energy management system: based on the representative daily cooling load, abnormal energy consumption of the cooling system can be diagnosed through comparing the real-time operating data to the representative cooling load profiles.

Both the outdoor weather conditions (i.e. outdoor air dry-bulb temperature, relative humidity, solar radiation, wind speed, etc) and indoor operating schedules (i.e. occupancy, lighting, etc) have effect on the actual cooling load [2]. It is worthwhile to see the relationship between clustering result of outdoor weather conditions, indoor operating schedules and the cooling load profile. It also worth investigating the clustering result of an integrated information among outdoor weather conditions, indoor operating schedules and cooling load profile.

In future study, the applicability of the proposed pattern identification approach should be tested on other types of building such as residential, office and hospital buildings. Moreover, the performance of other types of clustering algorithms in feature identification, such as connectivity-based clustering, centroid-based clustering and density-based clustering should be evaluated. Meanwhile, it is worthwhile to see whether this pattern identification approach is also effective on other types of chilled water system (i.e. air-cooled chilled water system) and in different climatic regions. It is also expected that the proposed 3D pattern identification can formulate a standard tool to facilitate feature identification of cooling loads in various types of buildings and even district area thus to provide valuable retrofiting suggestion for chilled water system and buildings.

## 6. Conclusion

This paper proposes a 3D pattern identification approach for building cooling load using Gaussian mixture model-based clustering.

- At the first stage, various historical measurement data of the chilled water system is collected from the building energy management system; Data cleaning is then conducted to improve the quality of measurement data by excluding the missing values and detecting outliers in raw measurement data; Furthermore, data pre-processing is conducted to obtain the 1-h mean value, 4-h mean value as well as daily average, minimum and maximum value of cooling load to have better understanding of the different trends of daily cooling load.
- At the second stage, data clustering is implemented on the 15-min measurement, 1-h mean value, 4-h mean value as well as daily average, minimum and maximum value of cooling load,

respectively. Bayesian Information Criterion (BIC) value is then adopted to select the optimal quantity of components of Gaussian mixture model in each case. Therefore, each day of the year is grouped into different clusters based on its daily cooling load profile.

- At the third stage, based on the year-round distribution and clustered cooling load profiles, data visualization of calendar view and grouped cooling load profile is conducted to extract the unique feature of each cluster.

The successful implementation of the proposed pattern identification approach on two chilled water systems in different types of buildings has demonstrated its robustness and generality. The main findings from the pattern identification are highlighted as follows:

- Through the proposed pattern identification approach, the year-round cooling load profile can be grouped into different clusters, with each cluster illustrating a distinctive characteristic of daily cooling load profile;
- The 15-min measurement, 1-h mean value, 4-h mean value and the daily brief load profile could provide the detailed trend, overall trend, general trend and brief information of the daily cooling load, respectively;
- The 1-h mean value resulted in the most clusters while the 4-h mean value and daily brief values resulted in the fewest clusters through GMM clustering. Therefore, the 1-h mean value resulted in the most accurate clustering thus could explore the most of features of the daily cooling load profile. On the other hand, 4-h mean value and daily brief value is helpful in quicker analysis.

Through comparing the clustering results of the campus and hotel building, it is also found out that:

- Compared to that in the hotel building, there are more frequent switching on-and-off of chillers in the campus building;
- There are roughly 3 seasons resulted from 15-min measurement and 1-h mean value of cooling load profile, while 2 seasons from 4-h mean value and daily brief information (i.e. average, minimum and maximum value of daily cooling load profile);
- The optimal quantity of components in the Gaussian mixture model for each time-step’s load profile is similar between the campus and the hotel building, which also demonstrates the robustness and generality of the proposed approach.

## Acknowledgement

The authors would like to acknowledge and express their sincere gratitude to The Department for Business, Energy & Industrial Strategy through grant project number TEIF-101-7025. Opinions expressed and conclusions arrived at are those of the authors and are not to be attributed to BEIS.

## Nomenclature

| | |
|---|---|
| $A$ | Total quantity of free parameters in GMM |
| $BIC$ | Bayesian Information Criterion |
| $d$ | Dimension, or quantity of elements in each vector |
| $C_p$ | Specific heat of chilled water (J kg$^{-1}$ K$^{-1}$) |
| $G$ | Quantity of operating chillers |
| $K$ | Quantity of components in Gaussian mixture models |
| $L$ | Likehood function |
| $M$ | Mass flow rate (L/s) |
| $N$ | Quantity of vectors in the database |
| $Q$ | 15-min cooling load |
| $\boldsymbol{Q}$ | Vector of database for 15-min cooling load |
| $\mathbf{Q}$ | Database of 15-min cooling load |
| $r$ | Relative error |
| $T$ | Temperature (K) |
| $X$ | 1-h mean cooling load |
| $\boldsymbol{X}$ | Vector of database for 1-h mean cooling load |
| $\mathbf{X}$ | Database of 1-h mean cooling load |
| $Y$ | 4-h mean cooling load |
| $\boldsymbol{Y}$ | Vector of database for 4-h mean cooling load |
| $\mathbf{Y}$ | Database of 4-h mean cooling load |
| $Z$ | Average, minimum or maximum value of daily cooling load |
| $\boldsymbol{Z}$ | Vector of database for daily brief information of cooling load |
| $\mathbf{Z}$ | Database of daily brief information of cooling load |
| | |
| $\omega$ | Weight of mixture component |
| $\phi$ | Gaussian probability density function |
| $\mu$ | Mean matrix of GMM |
| $\Sigma$ | Covariance matrix of GMM |
| $\gamma$ | Posteriori probability |

## Subscripts

| | |
|---|---|
| $b$ | Branch |
| $c$ | Chiller |
| $g$ | Number of operating chiller |
| $i$ | Day of the year |
| $j$ | Time step of the day |
| $k$ | Number of GMM |
| $l$ | Leaving |
| $r$ | Return |

## Abbreviations

| | |
|---|---|
| GMM | Gaussian mixture model |

## References


[1] Luo XJ, Fong KF, Sun YJ and Leung M.K.H. Development of clustering-based sensor fault detection and diagnosis strategy for chilled water system. Energy and Buildings 186(2019)17-36.

[2] Luo XJ, Oyedele LO, Ajayi AO, Monyei CG, Akinade OO and Akanbi LA. Development of an IoT-based big data platform for day-ahead prediction of building heating and cooling demands. Advanced Engineering Informatics. 41(2019)100926.

[3] Zhu GY, Chow TT, Fong, KF, Lee, C and Luo XJ. Design optimisation and performance appraisal of a combined cooling, heating and power system primed with Maisotsenko combustion turbine cycle. Energy Conversion and Management. 177(2018)91-106.

[4] Luo XJ and Fong KF. Development of multi-supply-multi-demand control strategy for combined cooling, heating and power system primed with solid oxide fuel cell-gas turbine. Energy Conversion and Management. 154(2017)538-561.

[5] Luo XJ and Fong KF. Development of integrated demand and supply side management strategy of multi-energy system for residential building application. Applied Energy. 242(2019)570-587.

[6] Luo XJ and Fong KF. Development of 2D dynamic model for hydrogen-fed and methane-fed solid oxide fuel cells. Journal of Power Sources. 328(2016)91-104.

[7] Luo XJ and Fong KF. Investigation on part-load performances of combined cooling and power system primed by solid oxide fuel cell with different bottoming cycles. Journal of Power Sources. 429(2019)127-148.

[8] Ortega A and Manana M. 2017. Characterization and analysis of energy demand patterns in airports. Energies. 10(1)119.

[9] Ahmad T and Chen H. 2018. Short and medium-term forecasting of cooling and heating load demand in building environment with data-mining based approaches. Energy and Buildings. 166(2018)460-476.

[10] Ahmad T and Chen H. 2019. A review on machine learning forecasting growth trends and their real-time applications in different energy systems. Sustainable Cities and Society. (2019)102010.

[11] Ahmad T, Chen H, Huang R, Yabin G, Wang J, Shair J, Akram HM, Mohsan SA and Kazim M. Supervised based machine learning models for short, medium and long-term energy prediction in distinct building environment. Energy. 158(2018)17-32.

[12] Ahmad T, Chen H, Guo Y, Wang J. A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review. Energy and Buildings. 165(2018)301-20.

[13] Wen L, Zhou K and Yang S. A shape-based clustering method for pattern recognition of residential electricity consumption. Journal of Cleaner Production. 212(2019)475-488.

[14] Kang L, Fei W, Zhao Z, Zeng M, Hong S, Chun L and Jing L. Analysis on residential electricity consumption behavior using improved *k*-means based on simulated annealing algorithm. In Power and Energy Conference at Illinois, IEEE1-6.

[15] Guo Z, Zhou K, Zhang X, Yang S, Shao Z. Data mining based framework for exploring household electricity consumption patterns: A case study in China context. Journal of cleaner production. 195(2018)773-85.

[16] Fan C, Xiao F, Madsen H, Wang D. Temporal knowledge discovery in big BAS data for building energy management. Energy and Buildings. 109(2015)75–89.

[17] Xiao F and Fan C. Data mining in building automation system for improving building operational performance. Energy and Buildings. 75(2014)109–118.

[18] Li K, Ma Z, Robinson D and Ma J. Identification of typical building daily electricity usage profiles using Gaussian mixture model-based clustering and hierarchical clustering. Applied energy. 231(2018)331-342.

[19] Wang J, Li G, Chen H, Liu J, Guo Y, Sun S and Hu Y. Energy consumption prediction for water-source heat pump system using pattern recognition-based algorithms. Applied Thermal Engineering. 136(2018)755-766.

[20] Carmo C.M.R. and Christensen T.H. Cluster analysis of residential heat load profiles and the role of technical and household characteristics. Energy and Buildings. 125(2016)171-180.

[21] Gianniou P, Liu X, Heller A, Nielsen PS and Rode C. Clustering-based analysis for residential district heating data. Energy Conversion and Management. 165(2018)840-850.

[22] Lu Y, Tian Z, Peng P, Niu J, Li W and Zhang H. GMM clustering for heating load patterns in-depth identification and prediction model accuracy improvement of district heating system. Energy and Buildings. 190(2019)49-60.

[23] Yu FW and Chan KT. Using cluster and multivariate analyses to appraise the operating performance of a chiller system serving an institutional building. Energy and Buildings. 44 (2012)104–113.

[24] Yu FW and Chan KT. Assessment of operating performance of chiller systems using cluster analysis. International Journal of Thermal Sciences. 53 (2012)148–155.

[25] An J, Yan D, Hong T. Clustering and statistical analyses of air-conditioning intensity and use patterns in residential buildings. Energy and Buildings. 174(2018)214-27.

[26] Han J, Pei J and Kamber M. Data mining: concepts and techniques. Elsevier. 2011.

[27] Bishop CM. 2006. Pattern recognition and machine learning. springer.

[28] Banfield JD and Raftery AE. Model-based Gaussian and non-Gaussian clustering. Biometrics. (1993)803-821.