A novel clustering-enhanced adaptive artificial neural network model for predicting day-ahead building cooling demand

X.J. Luo*

Big Data Enterprise and Artificial Intelligence Laboratory (Big-DEAL)
University of the West of England (UWE), Frenchay Campus, Bristol, United Kingdom
*Corresponding author: xiaojun.luo@uwe.ac.uk

Abstract

To accurately predict hourly day-ahead building cooling demand, year-round historical weather profile needs to be evaluated. The daily weather profiles among different time periods result in various features of historical datasets. The different appropriate structure and parameters of artificial neural network models may be identified for training datasets with different features. In this study, a novel clustering-enhanced adaptive artificial neural network (C-ANN) model is proposed to forecast 24h-ahead building cooling demand in subtropical areas. The uniqueness of the proposed adaptive model is that *k*-means clustering is implemented to recognise representative patterns of daily weather profile and thus categorize the annual datasets into featuring clusters. Each cluster of the weather profile, along with the corresponding time variables and cooling demand, is adopted to train one ANN sub-model. The optimal structure and parameters of each ANN sub-model are selected according to its featuring training datasets; thus the ANN sub-models are adaptive. The proposed C-ANN model is tested on a representative office building in Hong Kong. It is found that the mean absolute percentage error of the training and testing cases of the proposed predictive model is 3.59% and 4.71%, which has 4.2% and 3.1% improvement compared to conventional ANN model with a fixed structure. The proposed adaptive predictive model can be applied in building energy management system to accurately predict day-ahead building cooling demand using the latest forecast weather profile.

Keywords

Cooling demand; Prediction; Clustering; Artificial neural network; Data-driven model.

1. Introduction

The emergence of energy crisis and the increase in energy demand has forced people to use energy in a more efficient and effective way. Buildings have become the largest energy consumer worldwide since most people spend over 90% of their daily lives indoors. Generally, building energy demands include cooling, heating and electricity demands. Long-term forecasting is crucial in building design evaluation, while short-term prediction is important in effective building energy management [1-4]. There are a variety of factors affecting building energy demands, such as physical properties, energy devices, weather conditions and occupant behaviours. Due to the nonlinear and complex relationship between these affecting factors and building energy demands, it is challenging to predict day-ahead hourly building energy demands accurately.

1.1 Related works

Analytical thermal models and machine learning techniques are two categories of approaches in building energy demands prediction [5-7]. Fundamental thermodynamic equations are usually used to construct the analytical thermal models of different types of buildings. The cooling and heating demands are estimated based on the detailed building and environmental parameters, including building envelop information, climatic weather profile and internal operating schedules. However, such information is generally difficult to obtain. Commercial energy simulation software, such as TRNSYS [8], EnergyPlus [9] and ESP-r [10] are generally adopted for developing such analytical thermal models. The complicated building envelopes, especially in high-rise buildings, will make the analytical thermal models computationally expensive [5].

On the other hand, machine learning techniques, especially artificial neural network (ANN) models, are widely used in investigating the complex relationship between affecting factors and energy demands. Historical measurement is usually adopted to train the data-driven models. The training process of ANN models is to determine its structure and parameters, thus to reveal the nonlinear relationship between various output and input datasets. In [11], four single-hidden-layer ANN predictive models were developed to forecast the daily average heating and cooling demands in buildings for four different seasons, respectively. To represent the daily weather patterns, the mean, standard deviation, maximum, and minimum values of daily outdoor air dry-bulb temperature and humidity ratio were adopted as inputs to the proposed ANN predictive models. It was pointed out by the authors that seasonal effects might highly influence the parameters of the predictive model. Therefore, the manual approach was adopted to split and select data, which is based on pre-determined time periods and daily average dry-bulb temperature. In [12], various single-hidden-layer ANN models were ensembled to predict the daily heating energy consumption of a university campus. The ANN models included feed-forward

backpropagation neural network, radial basis function network and adaptive neuro-fuzzy interference system. The whole year was manually divided into three periods: cold, mild and warm periods, while only the cold period was studied. In [13], the single-hidden-layer ANN predictive models were constructed for daily cooling demand in institutional buildings. To overcome the problem of high variation in one-year measurement data, the energy consumption values were assigned to five different groups through manual classification: very low, low, medium, high and very high. Each group of the database was adopted to train one ANN predictive model.

In [14], a single-hidden-layer ANN predictive model with fixed structure was developed to predict the hourly building energy consumption. The inputs to the ANN model were outdoor air dry-bulb and wet-bulb temperatures, while the output was the electrical energy consumption. Several months' simulation data was adopted to train the ANN predictive model. In [15, 16], a single-hidden-layer ANN predictive model with fixed structure was developed to predict the hourly building cooling demand using the simulation data from a week. In [17], a single-hidden-layer ANN model was ensembled with other data-mining methods to predict the building cooling demand using one-month's simulation data. In [18], a support vector machine model and a single-hidden-layer ANN model was used to predict office cooling demand using one month's measurement data. In [19], a hybrid teaching-learning optimization based-ANN predictive model with the fixed structure was developed for building electrical energy consumption using several months' measurement. In [20], the single-hidden-layer ANN models with several different fixed structures for building energy consumption were trained using 3 years' measurement data collected from the building management system of a specific building.

1.2 Limitation of existing approaches

In terms of daily energy demand prediction, Jovanovic *et al.* [12] implicated that higher accuracy could be obtained through using separate data-driven models for each period compared to adopting one model for the whole year. However, in previous research works, the year-round weather profile database was generally divided into different groups according to natural seasons or outdoor air dry-bulb temperature range. Since there exist different features of weather profile during the same season, those manual classification approaches were actually not accurate enough to adequately distinguish the featuring patterns of weather profile. Moreover, for hourly energy demand prediction, most of the previous research studies used one week's [15, 16] one month's [17, 18] or several months' [14, 19] meteorological weather profile to train the ANN predictive model with a fixed structure. However, data within a short period was not sufficient to represent the variation of building energy demands. Although the building operating pattern might be similar year-round, the meteorological weather profile varies [1]. It is not appropriate to use several months' historical data to train ANN predictive models and then adopt them for building energy demand prediction in other months.

Clustering approach is an effective tool in pattern recognition [21]. Through clustering, the datasets within the same cluster are similar to each other, while the datasets among different clusters have little similarity. Panagiota *et al.* [22] used the *k*-means clustering algorithm to identify the typical daily load profiles from 8293 Danish single-family households in Aarhus. It was found that Danish district heating customers could be segmented into five clusters with regards to the energy consumption intensity. Samira *et al.* [23] used the *k*-means clustering algorithm to group the year-round energy demand profile into several typical periods, while the featuring characteristics of the annual energy demand were adequately observed. Schiefelbein *et al.* [24] adopted the *k*-means clustering algorithm to identify the typical daily energy demands profile, thus to reduce computational time in the design and optimization of thermal energy systems. Fernando *et al.* [25] used the *k*-means clustering algorithm to reduce the whole year energy demands profile into a few representative days, while the unique characteristics, including the peak demands, the demand duration curves, and the temporal inter-relationship among heating, cooling and electricity demands, were retained. Luo *et al.* [26] proposed the Gaussian mixture model-based clustering method to identify representative building cooling load. Adesoji *et al*. [27] developed a hybrid clustering and multi-linear regression model to identify the representative archetypes.

However, as far as it is concerned, the clustering approach was seldom hybrid with machine learning models for building energy demands prediction. Only Chen *et al*. [28] proposed a hybrid support vector regression model to predict the day-ahead hourly electric demand based on feature identification and better organization of training samples. The training samples were clustered into different groups according to the daily trend of electric consumption. However, it did not mention how to implement such clustering in future energy consumption prediction since its trend would be unknown.

1.3 Contribution

In the above-mentioned literature review, most of the ANN-based predictive models used only several months' historical data as training samples. Although the entire-year weather profile was adopted in some of the prediction models, the year-round weather profile was generally divided into different groups according to natural seasons or outdoor air dry-bulb temperature range. Since there exist different features of weather profile during the same season, these manual classification approaches were actually not accurate enough to adequately distinguish the featuring patterns of annual weather profile. On the other hand, *k*-means clustering analysis is effective in pattern recognition of different energy profiles. The gap in knowledge inspires the idea of using clustering-enhanced adaptive artificial neural network (C-ANN) model for predicting cooling energy demand in building applications.

The most distinctive feature of the proposed predictive model is that clustering analysis is firstly used to detect the featuring patterns of one-day weather profile and group the yearly weather profile into several different clusters. After that, each sub-ANN predictive model with its distinctive structure and parameters is trained by the datasets which share the same featuring pattern. It is expected that the proposed C-ANN predictive model will serve as an effective and accurate prediction approach for various building energy demands in practical application.

The rest of the paper is structured like this: The next section illustrates the proposed clustering-enhanced adaptive artificial neural network model; the following section illustrates the generation of the historical weather data profile and corresponding cooling demand database; the fourth section discusses the prediction results; the fifth section details the implication for practice and future study while the last section points out the conclusion.

2. Proposed clustering-enhanced adaptive artificial neural network model

The proposed clustering-enhanced adaptive artificial neural network model consists of four key procedures: data pre-processing, data clustering, data training and data post-processing. The outline of the proposed predictive model is illustrated in Fig. 1. Outdoor air dry-bulb temperature and relative humidity play an important role in determining convective heat gain from walls and windows, infiltration heat gain and ventilation heat gain; Total horizontal radiation and horizontal beam radiation decide the solar heat gain; Building operating schedules are decisive in various internal heat gains. Therefore, weather data profile and time variables are consolidated as input datasets to the clustering-enhanced adaptive artificial neural network model.

- The outdoor weather condition mainly includes outdoor air dry-bulb temperature T_{db}, outdoor air relative humidity RH, total horizontal radiation THR and horizontal beam radiation HBR. Therefore, at the k^{th} hour on the i^{th} day: $x'_{i,1,k} = T_{db,i,k}, x'_{i,2,k} = RH_{i,k}, x'_{i,3,k} = THR_{i,k}, x'_{i,4,k} = HBR_{i,k}$.
- In office buildings, the operating schedules (i.e. occupants, office equipment, lighting and lifts) are relatively constant during weekdays and weekends, respectively. Thus, time variables mainly consist of hour of the day *hd* and day of the week *dw*, and $t_{i,1,k} = hd, t_{i,2,k} = dw$.

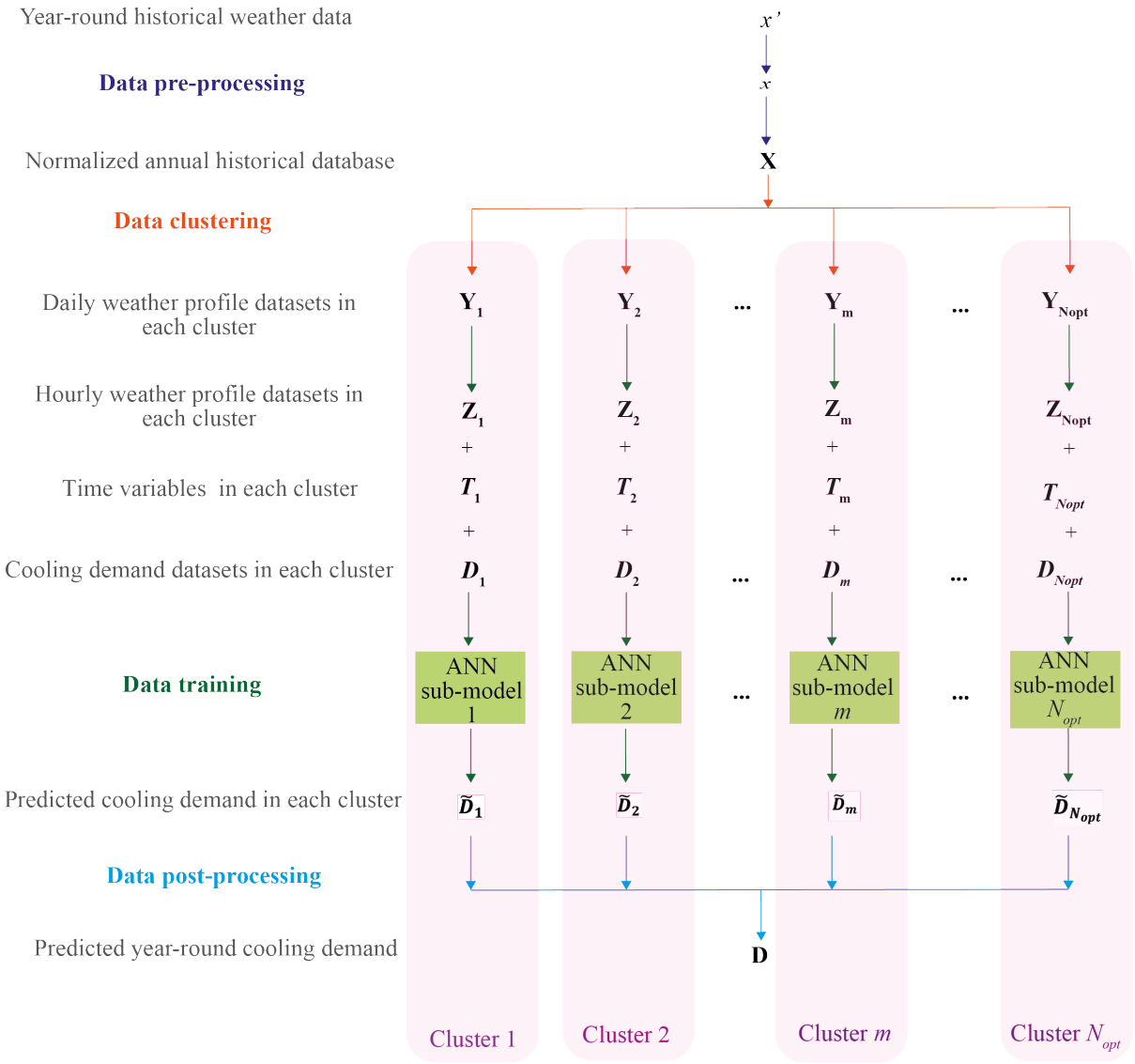


Fig. 1. Outline of the proposed clustering-enhanced adaptive artificial neural network model.

2.1 Data pre-processing

Since the value range of each parameter $x'_{i,j,k}$ in the weather profile is quite different, min-max normalization is conducted to ensure that the training process of the ANN model does not suffer from numerical round-off effects [29]. Therefore, all the values in the historical database are normalized to values between 0 and 1.

$$x_{i,j,k} = \frac{x'_{i,j,k} - \min\limits_{1\le i\le N_d, 1\le k\le N_h} x'_{i,j,k}}{\max\limits_{1\le i\le N_d, 1\le k\le N_h} x'_{i,j,k} - \min\limits_{1\le i\le N_d, 1\le k\le N_h} x'_{i,j,k}} \tag{1}$$

where,

$x'_{i,j,k}$: the original value of the j^{th} type of weather data at the k^{th} hour on the i^{th} day;

$x_{i,j,k}$: the normalized value;

N_h: total hours of the day, and $N_h = 24$;

N_d: total days of the year, and $N_d = 365$.

2.2 Data clustering

The year-round weather profile $\mathbf{X}$ consists of the daily weather profile of $N_d = 365$ days and $\mathbf{X} = [\boldsymbol{X}_1; \boldsymbol{X}_2; \ldots \boldsymbol{X}_i; \ldots; \boldsymbol{X}_{N_d}]$; $\boldsymbol{X}_i$ is the daily profile of various weather data, and $\boldsymbol{X}_i = [X_{i,1}\, X_{i,2} \ldots X_{i,j} \ldots X_{i,N_p}]$; $X_{i,j}$ is the daily profile of the j^{th} type of weather data, and $X_{i,j} = [x_{i,j,1}\, x_{i,j,2} \ldots x_{i,j,N_h}]$. N_p is the total quantity of weather data types, and $N_p = 4$.

By using the *k*-means clustering algorithm, the daily profile $\boldsymbol{X}_i$ (i = 1, 2, …, N_d) is classified into certain number (m) of clusters (m = 1, 2, …, N_c). Therefore, the dataset of cluster centroids C_m in each cluster is determined by minimizing the total Euclidian distance between each dataset to its corresponding cluster centroid as described in Eq. (2)

$$J = \sum_{m=1}^{Nc} \sum_{i=1}^{N_d} \left(\sum u_{i,m} \| \boldsymbol{X}_i - C_m \| \right) \tag{2}$$

where,

$\| \|$: Euclidean distance between two data points;

m: cluster number;

N_C: total quantity of clusters;

$u_{i,m}$: clustering factor, $u_{i,m}$ = 1 indicates the i^{th} day belonging to the m^{th} cluster, $u_{i,m}$ = 0 otherwise.

The cluster centroid dataset $C_m = \{c_{m,j,k} | j = 1,2, \ldots, N_p;\ k = 1,2, \ldots, N_h\}$ is determined by Eq. (3):

$$C_m = \frac{\sum_{i=1}^{i=Nd} (\boldsymbol{X}_i\, u_{i,m})}{\sum_{i=1}^{i=Nd} u_{i,m}} \tag{3}$$

The optimum quantity of clusters N_{opt} is decided through the lowest Davies-Bouldin index evaluation value [30].

$$DB_{N_c} = max \frac{\left[\frac{\sum u_{i,m1}=1 (|X_i - C_{m1}| + |X_i - C_{m2}|)}{N_{m1}} \right]}{|C_{m1} - C_{m2}|} \tag{4}$$

Hence the cluster centroid database **C** can be formulated as $\mathbf{C} = [C_1; C_2; \ldots; C_m; \ldots; C_{N_{opt}}]$. According to clustering factor $u_{i,m}$, the historical database **X** are grouped into N_{opt} clusters: $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_m, \ldots, \mathbf{Y}_{N_{opt}}$, and $\mathbf{X} = \mathbf{Y}_1 \cup \mathbf{Y}_2 \cup \ldots \mathbf{Y}_m \ldots \cup \mathbf{Y}_{N_{opt}}$. The procedure of the data clustering is shown in Fig. 2.

```
for m = 1, 2, ..., N_opt
i' = 1
    for i = 1, 2, ..., N_d
        if u_i,m = 1
        Y_m,i' = X_i
            for j = 1, 2, ..., N_p
                for k = 1, 2, ..., N_h
                    y_m,i',j,k = x_i,j,k
                end for
            end for
        i' = i' + 1
        end if
    end for
N_m = i'
end for
```

Fig. 2. Procedure of weather profile clustering.

Thus, in the m^{th} cluster, $\mathbf{Y_m} = [\boldsymbol{Y}_{m,1}; \boldsymbol{Y}_{m,2}; \ldots; \boldsymbol{Y}_{m,i'}; \ldots; \boldsymbol{Y}_{m,N_m}]$, $\boldsymbol{Y}_{m,i'} = \{y_{m,i',j,k} | j = 1, 2, \ldots, N_p;\ k = 1,2, \ldots N_h\}$, and N_m is the total quantity of days in the m^{th} cluster. Based on the clustering result $u_{i,m}$ from weather profile clustering, the year-round time variables $\mathbf{T} = \{t_{i,j,k} | i = 1,2, \ldots N_d; j = 1,2;\ k = 1,2, \ldots N_h\}$ and cooling demand profile $\mathbf{D} = \{d_{i,k} | i = 1,2, \ldots N_d;\ k = 1,2, \ldots N_h\}$ are assigned into corresponding groups according to its number of days, as shown in Fig. 3.

```
for m = 1, 2, ..., N_opt
i' = 1
    for i = 1, 2, ..., N_d
        if u_i,m = 1
            for k = 1, 2, ..., N_h
                D_m,i',k = d_i,k
                for j = 1, 2
                    T_m,i',j,k = t_i,j,k
                end for
            end for
        i' = i' + 1
        end if
    end for
end for
```

Fig. 3. Procedure for grouping time variables and cooling demands.

Therefore, in the m^{th} cluster, $\boldsymbol{D_m} = \{D_{m,i',k} | i' = 1,2,\dots,N_m; k = 1,2,\dots,N_h\}$, $\boldsymbol{T_m} = \{T_{m,i',j,k} | i' = 1,2,\dots,N_m; j = 1,2; k = 1,2,\dots,N_h\}$.

2.3 Data training

Each cluster of the weather profile $\mathbf{Y_m}$, along with the corresponding time variables $\boldsymbol{T_m}$ and cooling demand $\boldsymbol{D_m}$, is adopted to train one ANN sub-model. To facilitate the training of sub-ANN models, the $N_m \times (N_h \times N_p)$ matrix $\mathbf{Y}_m$ is transformed into the $(N_m \times N_h) \times N_p$ matrix $\mathbf{Z}_m$. The transforming process is shown in Fig. 4.

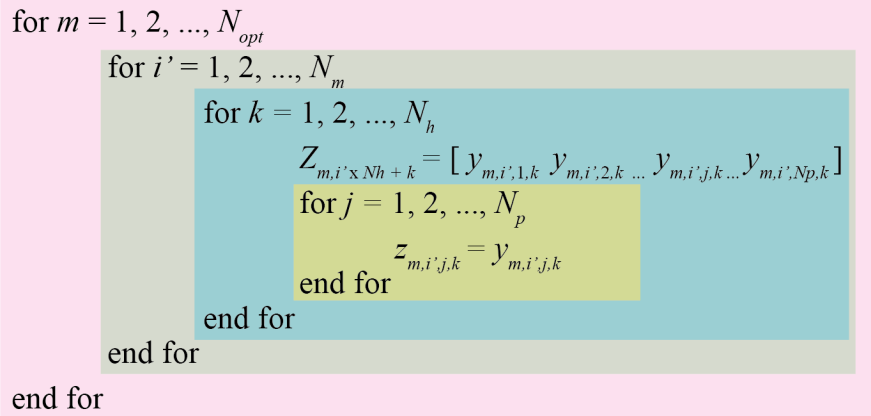


Fig. 4. Transforming process of $\mathbf{Y}_m$ into $\mathbf{Z}_m$.

Thus, the training database $\mathbf{Z}_m$ is formulated as $\mathbf{Z_m} = \{z_{m,i',j,k} | i' = 1,2,\dots,N_m; j = 1,2,\dots,N_p; k = 1,2,\dots,N_h\}$. In summary, in each cluster m, there are $N_m \times N_h$ time steps, and thus $N_m \times N_h$ samples of data for ANN training. At each time step, for example, the k^{th} h hour of the i'^{th} day, the training sample is represented as: $v_1 = z_{m,i',1,k}, v_2 = z_{m,i',2,k}, v_3 = z_{m,i',3,k}, v_4 = z_{m,i',4,k}, v_5 = T_{m,i',1,k}, v_6 = T_{m,i',2,k}$.

ANN algorithm was first introduced by McCulloch and Pitts [31], which is based on mimicking the function of the human brain. Similar to the central nervous system of mammals, the ANN attempts to simulate the nonlinear and nonstationary multivariate dataset through the networks. The schematic diagram of the ANN algorithm is shown in Fig. 5.

The representative ANN model constitutes of an input layer, a hidden layer and an output layer. The essential component in the ANN model is the neuron, which is affiliated in layers while tied to neurons in other layers through synaptic weights. The values of these weights are determined during the training process. The quantity of neurons in the input and output layers are determined by the total types of input datasets and output datasets, respectively. The hidden layer plays an important role in determining the accuracy and effectiveness of the predictive model. The neurons in the hidden layer allow the neural

networks to detect the feature, to capture the pattern in the dataset, and to perform the complicated nonlinear mapping between input and output variables. According to [11-20], the single hidden layer is sufficient for the ANN models to approximate any complex nonlinear function with satisfactory accuracy. Therefore, one hidden layer is adopted in each of the ANN sub-models in this study. The optimal quantity of neurons in the hidden layer is chosen according to its prediction performance (i.e. mean absolute percentage error between the predicted and actual result); thus each ANN sub-model is configured with the unique feature of the weather database $\mathbf{Z}_m$.

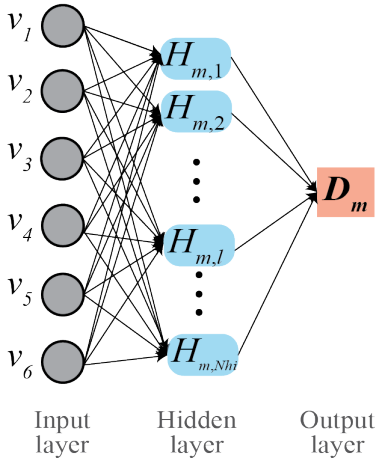


Fig. 5. Schematic diagram of the ANN algorithm.

The l^{th} neuron $H_{m,l}$ in the hidden layer of the m^{th} ANN sub-model is defined as:

$$H_{m,l} = f\left(\sum_{j=1}^{j=6}(w_{m,j,l}\ v_j)\right) \tag{5}$$

where $w_{m,j,l}$ is the connection weighting factor between the j^{th} input and the l^{th} neuron, while Rectified Linear Unit is adopted as the activation function f:

$$f(t) = \begin{cases} 0 \ \ for\ t < 0 \\ t\ for\ t \geq 0 \end{cases} \tag{6}$$

The cooling demand dataset $\boldsymbol{D_m}$ is considered as the only neuron in the output layer:

$$\widetilde{\boldsymbol{D}}_{\boldsymbol{m}} = f\left(\sum_{l=1}^{l=N_{hi}}(w_{m,l}H_{m,l})\right) \tag{7}$$

where $\widetilde{\boldsymbol{D}}_{\boldsymbol{m}} = \left\{\widetilde{D}_{m,i',k}\middle|\ i' = 1,2,\dots,N_m; k = 1,2,\dots,N_h\right\}$ is the predicted cooling demand through the m^{th} ANN sub-model. The aim of the training process of each ANN sub-model is to minimize the squared error E_m between $\widetilde{\boldsymbol{D}}_{\boldsymbol{m}}$ and $\boldsymbol{D_m}$. Both Levenberg-Marquardt (LM) and Bayesian Regularization (BR)

approaches are adopted to minimize E_m. Therefore, the weighting factor database $\mathbf{W} = \{W_m | m = 1, 2, \dots, N_{opt}\}$, and $W_m = \{w_{m,j,l}; w_{m,l} | j = 1,2, \dots,6; l = 1,2, \dots, N_{hi}\}$ can be determined for each ANN sub-model.

$$E_m = \sum(\boldsymbol{D_m} - \widetilde{\boldsymbol{D}}_{\boldsymbol{m}})^2 \tag{8}$$

2.4 Data post-processing

After obtaining the predicted cooling demand from each cluster, $\widetilde{\boldsymbol{D}}_{\boldsymbol{m}}$ is regrouped to formulate yearly predicted cooling demand $\widetilde{\mathbf{D}}$, as illustrated in Fig. 6.

```
for m = 1, 2, ..., N_opt
i' = 1
    for i = 1, 2, ..., N_d
        if u_i,m = 1
            for k = 1, 2, ..., N_h
                d̃_i,k = D̃_m,i',k
            end for
        i' = i' + 1
        end if
    end for
end for
```

Fig. 6. Process of data post-process.

After obtaining the cluster centroid database **C** and weighting factor database **W**, the proposed C-ANN predictive model can be adopted to predict future building cooling demand. The day-ahead weather profile forecasted by the local weather station would be assigned to the appropriate cluster according to its Euclidean distance to the centroid dataset C_m. Thus, the weighting factor dataset W_m of the m^{th} sub-ANN model can be adopted to determine the day-ahead cooling demand.

2.5 Performance evaluation index

Overestimating the cooling demand will result in extra energy consumption, while underestimation may cause thermal comfort problems. Mean absolute percentage error (MAPE) is a scale-independent metric which proficiently reflects the relative prediction error. To evaluate the performance of the m^{th} cluster and the overall year, Eq. (8) and Eq. (9) is adopted, respectively:

$$MAPE_m = \frac{1}{N_m \times N_h} \sum_{i'=1}^{i'=N_m} \sum_{k=1}^{k=N_h} \left| \frac{\tilde{D}_{m,i',k} - D_{m,i',k}}{D_{m,i',k}} \right| \times 100\% \tag{9}$$

$$MAPE_o = \frac{1}{N_d \times N_h} \sum_{i=1}^{i=N_d} \sum_{k=1}^{k=N_h} \left| \frac{\tilde{d}_{i,k} - d_{i,k}}{d_{i,k}} \right| \times 100\% \quad (10)$$

3. Generation of the historical database using synthetic data

To assess the performance of the proposed C-ANN model, it was implemented on a representative office building in Hong Kong. To obtain the year-round cooling demand, the validated simulation platform TRNSYS 18 was adopted to develop the analytical building model. The building description, the TRNSYS simulation model and the structure of the database are discussed in each sub-section.

3.1 Building description

The representative office building in Hong Kong as detailed in the guidelines from Electrical and Mechanical Services Department [32, 33] is selected as the case study. It is a typical high-rise 30-floor office building while each floor has the same layout, as shown in Fig. 7. The floor area is 40.8 × 40.8 m^2 with a floor-to-floor distance of 3.6 m. Each floor is divided into three zones: the exterior zone, the interior zone and the lift lobby. The exterior zone is defined as the region within 4 m from the building external wall. Since the thickness of the external wall is 0.2 m, the area of an exterior zone is 556 m^2. The area of the lift lobby is based on the approximation of dividing the non-office region into a 7 × 7 grid sections and the total area of lift lobby measured 17 sections. Thus, the area of the lift lobby is 150 m^2. The area of the interior zone is 926 m^2, which is calculated as the total area minus those for the exterior zone and lift lobby. The building operating schedules are based on the local design practice [32, 33]. The operating schedules of the occupant, lighting air conditioning and lift of a weekday are presented in Fig. 8, while the schedules during the weekend are kept consistent as the off-peak hours (i.e. 1-5 h, 23-24 h). Meanwhile, the indoor design guideline of the office building is summarized in Table 1.

Table 1. Indoor design guideline.

Items	Value	Unit
Area per occupancy	8	m^2/person
Lighting power intensity	15	W/m^2
Office equipment power intensity	10	W/m^2
Indoor dry-bulb temperature	24	°C
Indoor relative humidity	50	%
Fresh air requisite	8	L/s/person
Occupant sensible heat gain	75	W/person
Occupant latent heat gain	75	W/person
Infiltration	0.0408	ACH

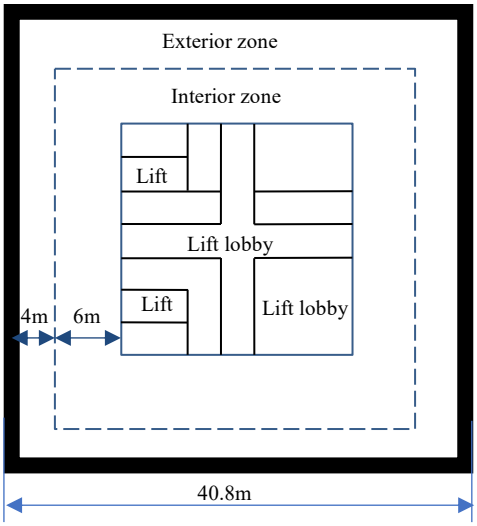


Fig. 7. Floor layout of the reference office building.

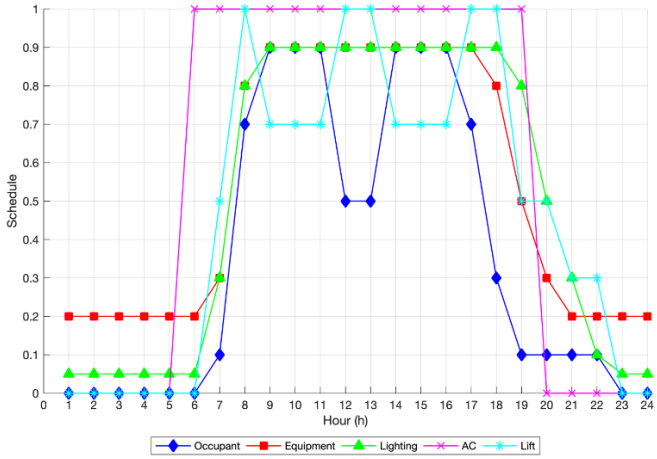


Fig. 8. Building operating schedules.

3.2 TRNSYS simulation model

In this study, the historical database is obtained through synthetic data. Owing to the yearly changing weather profile and building operating schedules, transient simulation is required to get the year-round hourly building cooling demand. Hence, TRNSYS 18 is implemented as the dynamic simulation platform, while Type 56 is adopted to construct the analytical thermal model of the representative office building.

The TRNSYS program has been continuously developed and improved by the Solar Energy Laboratory at the University of Wisconsin since 1975, and is widely adopted to simulate the transient performance of thermal energy systems. The Multi-zone Building model Type 56 can calculate the interaction among different zones through solving various coupled differential equations. In Type 56, ASHRAE transfer function approach is adopted to simulate the thermal performance of building envelopes [34]. TRNSYS program and Type 56 have been verified as a reliable approach for calculating various building energy demands in previous studies [35-42]. Moreover, in [43], the performance comparison of modelling building thermal behaviour was conducted through various popular simulation software, including EnergyPlus, DOE-2.1E and TRNSYS, the mean absolute percentage error from the experiment is found to be 9.0%, 7.7% and 6.6%, respectively. Therefore, the TRNSYS simulation gives the results with the most satisfactory accuracy.

To further validate the developed analytical building model, relative errors between the simulated peak cooling demand in each zone and those from [44] are calculated as shown in Table 2. [44] is also a simulation study with the same building. It is found that the relative error of the peak cooling demand

in each zone is smaller than 6%. Hence the TRNSYS simulation model for the office building is reckoned to be validated.

Table 2. Validation of the simulation model

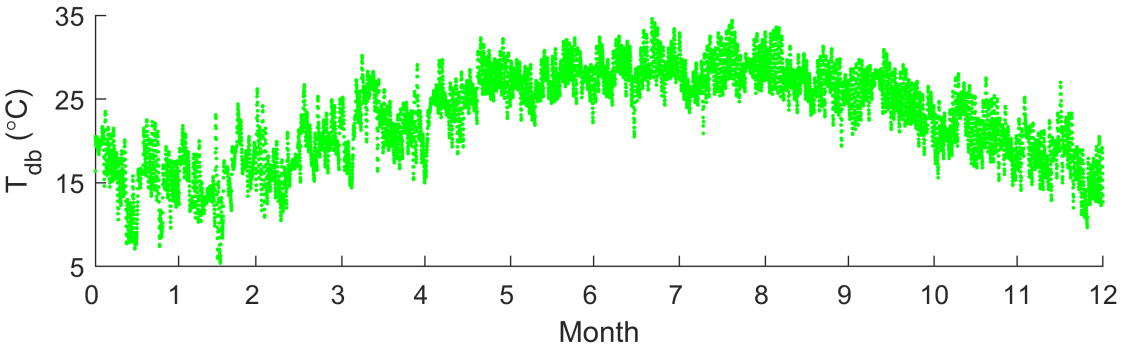
Zone	Simulated peak cooling demand (kW)		Relative error
	[44]	TRNSYS simulation model	
Exterior	82.79	77.94	5.86%
Interior	84.39	80.00	5.20%
Lift lobby	18.88	19.97	5.77%

3.3 Structure of the generated historical database

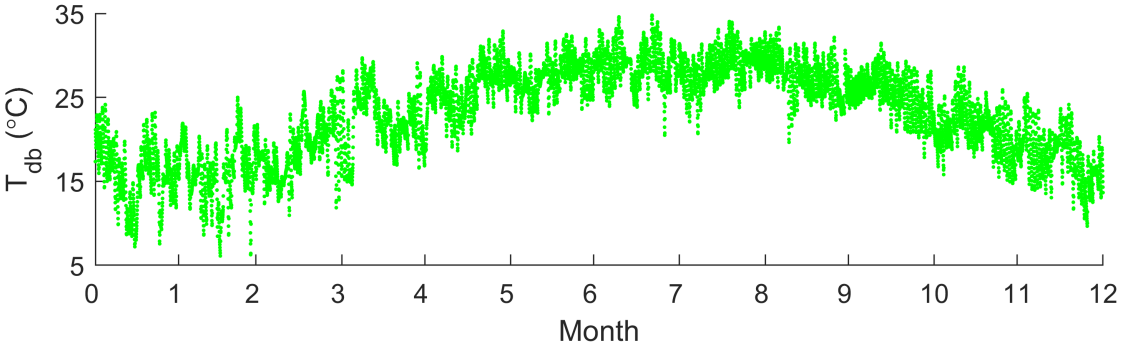
For training and testing purposes of the proposed clustering-enhanced C-ANN model, two sets of historical databases are generated. The typical meteorological year weather profiles from Hong Kong Observatory Station are used for training purpose while the ones from Hong Kong King's Park are used for testing purpose, respectively. The two year-round weather profile are shown in Fig. 9, while the maximum, minimum and average value of T_{db} and *RH*, along with the maximum and total value of *THR* and *HBR* in each month are summarized in Table 3.

The outdoor air dry-bulb temperature has the largest variation among these four types of weather profiles, which reaches about 34 ℃ in summer, while drops to 6 ℃ in winter. During most of the time, the range of outdoor air relative humidity is between 65-90% among the four different seasons, although it drops to 40% at a certain period. Total horizontal radiation and horizontal beam radiation have the largest values in summer while the smallest values in winter. Overall speaking, the weather profiles, including outdoor air dry-bulb temperature, relative humidity, total horizontal radiation and horizontal beam radiation, vary all over the year.

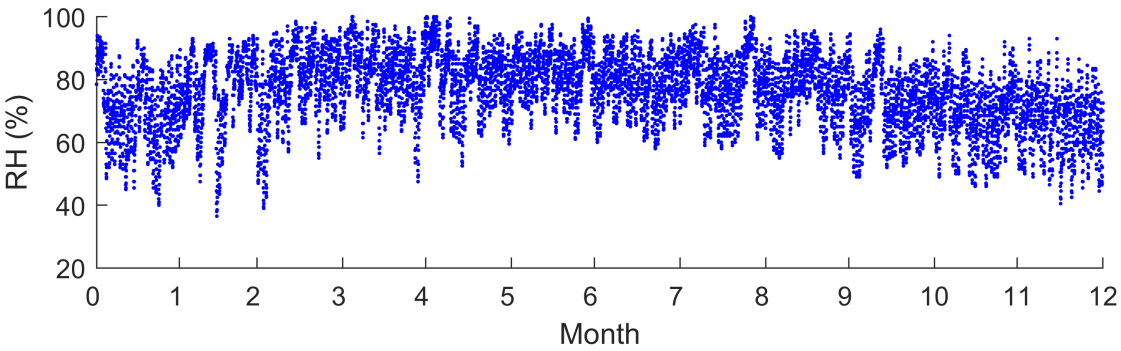
On the other hand, the two weather data profiles share similar average, minimum and maximum outdoor air dry-bulb temperature and relative humidity in the same months. Meanwhile, the maximum and total horizontal radiation and horizontal beam radiation are also similar among these two weather profiles. Therefore, these two sets of weather profiles can be served as training and testing databases, respectively.



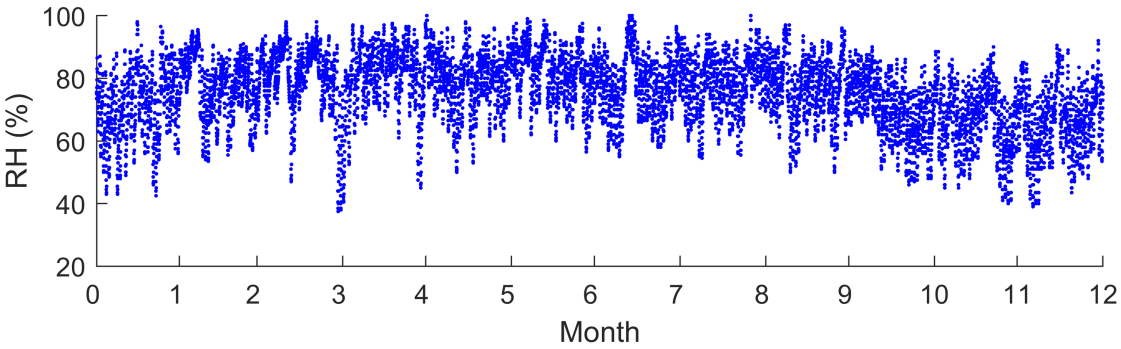
(a) Dry-bulb temperature, Hong Kong Observatory



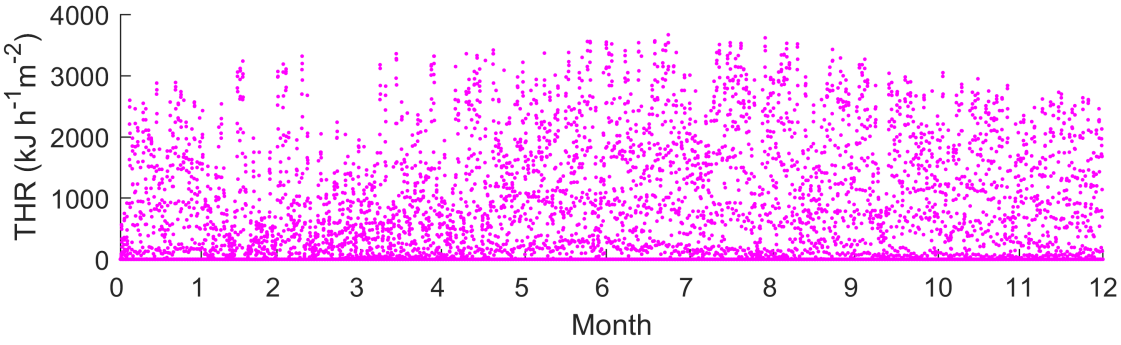
(b) Dry-bulb temperature, Hong Kong King's Park



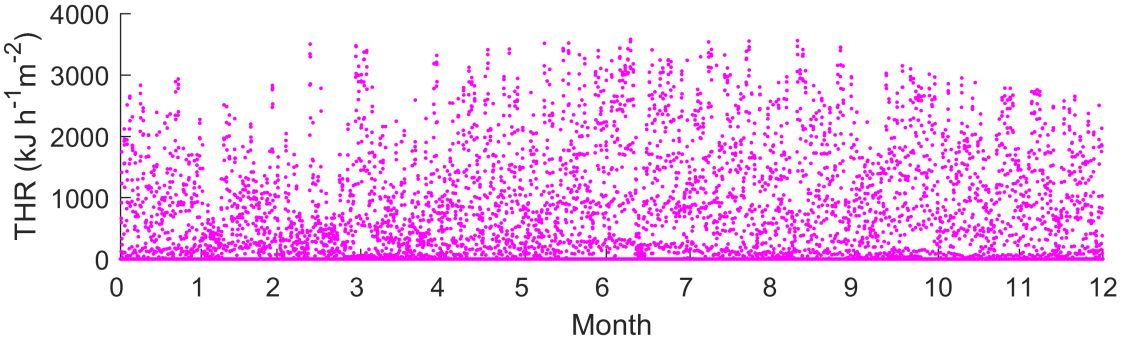
(c) Relative humidity, Hong Kong Observatory



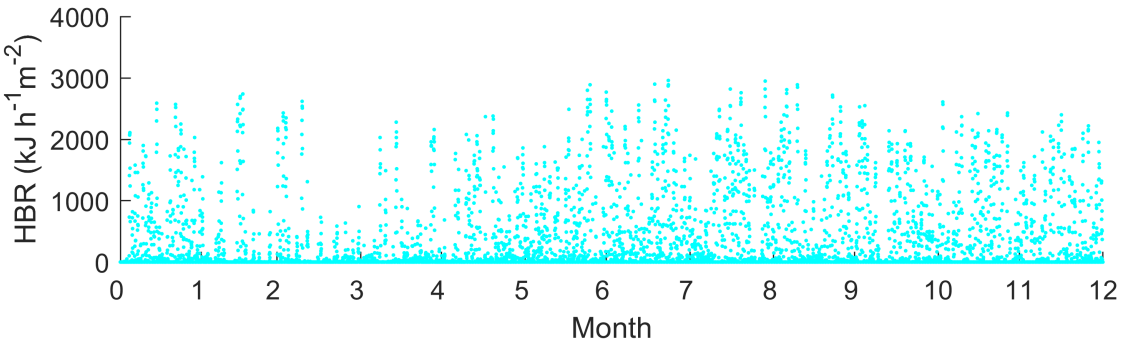
(d) Relative humidity, Hong Kong King's Park



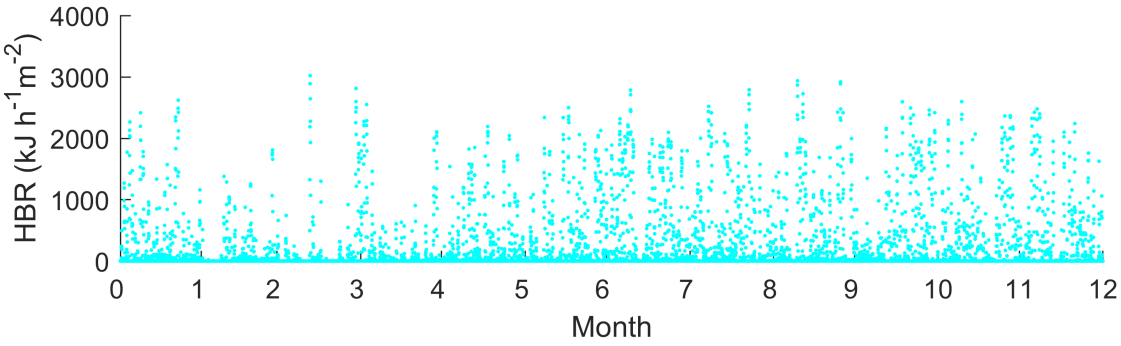
(e) Total horizontal radiation, Hong Kong Observatory



(f) Total horizontal radiation, Hong Kong King's Park



(g) Horizontal beam radiation, Hong Kong Observatory



(h) Horizontal beam radiation, Hong Kong King's Park

Fig. 9. Year-round weather profiles.

Table 3. Brief comparison weather condition in each month.

Weather variables			Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
T_{db}	Maximum (°C)	O	23.5	26.2	26.7	30.2	32.3	32.3	34.6	34.4	33.6	30.9	28.0	27.0
		K	24.2	25.0	28.2	29.7	32.9	32.5	34.8	34.1	33.4	31.5	31.5	28.6
	Minimum (°C)	O	7.1	5.4	10.5	14.4	18.5	21.9	20.5	20.9	19.4	18.4	14.7	9.7
		K	7.2	6.1	11.0	14.6	18.7	22.3	20.5	20.7	19.7	17.9	14.8	9.7
	Average (°C)	O	15.8	15.8	18.6	21.9	26.0	27.4	28.8	28.4	27.3	25.3	21.1	17.7
		K	16.0	16.1	18.9	22.2	26.3	27.6	29.0	28.7	27.5	25.4	21.3	17.9
RH	Maximum (%)	O	94	94	98.5	100	100	99.5	95.5	100	95.5	96	94	93
		K	98	95.5	98.0	100	98	99	100	100	97	90	90	92
	Minimum (%)	O	40	36.5	39	47.5	52.5	61	58	58	55	49	46	40.5
		K	42.5	53.5	37.5	45.0	50.0	56.5	55.0	54.5	50.0	46.0	40.0	39.0
	Average (%)	O	70.0	77.3	79.3	82.2	81.0	82.4	78.9	80.2	78.3	72.3	69.4	66.3
		K	71.4	77.2	79.2	81.2	79.3	81.8	78.2	78.5	76.8	70.2	66.7	66.4
R_g	Maximum ($kJ \cdot h^{-1}\ m^{-2}$)	O	2.89	3.24	3.32	3.36	3.42	3.56	3.67	3.62	3.53	3.27	3.05	2.73
		K	2.93	2.83	3.50	3.40	3.42	3.52	3.58	3.55	3.56	3.15	2.95	2.76
	Total ($kJ \cdot m^{-2}$)	O	37.7	23.0	28.1	33.9	46.9	52.8	60.5	56.8	49.5	47.1	39.1	37.4
		K	33.5	23.5	25.7	32.4	44.5	46.7	56.5	50.9	44.8	42.5	36.2	34.5
R_h	Maximum ($kJ \cdot h^{-1}\ m^{-2}$)	O	2.59	2.74	2.62	2.28	2.38	2.89	2.96	2.95	2.89	2.55	2.61	2.40
		K	2.62	1.81	3.02	2.55	2.19	2.50	2.79	2.79	2.94	2.60	2.60	2.48
	Total ($kJ \cdot m^{-2}$)	O	17.6	8.4	7.2	8.5	15.6	20.7	26.3	26.2	22.9	22.9	20.0	19.6
		K	13.0	4.8	7.4	8.0	13.0	16.2	25.4	18.4	18.4	17.8	16.4	16.1

O: Hong Kong Observatory Station K: Hong Kong King's Park

The two sets of year-round weather profiles, along with the time variables, are used in the TRNSYS simulation model to obtain two sets of cooling demand profiles. The detailed information regarding the inputs and output datasets is summarized in Table 4.

Table 4. Input and output datasets to the proposed C-ANN.

Historical database			Train	Test
Input datasets	Weather profile	Outdoor air dry-bulb temperature	Typical meteorological year weather profile from Hong Kong Observatory Station	Typical meteorological year weather profile from Hong Kong King's Park
		Outdoor air relative humidity		
		Global horizontal radiation		
		Horizontal beam radiation		
	Time variables	Hour of the day	Simulation calendar	
		Day of the week		
Output datasets	Simulated cooling demand		Simulation results from TRNSYS model based on weather profile from Hong Kong Observatory Station	Simulation results from TRNSYS model based on weather profile from Hong Kong King's Park

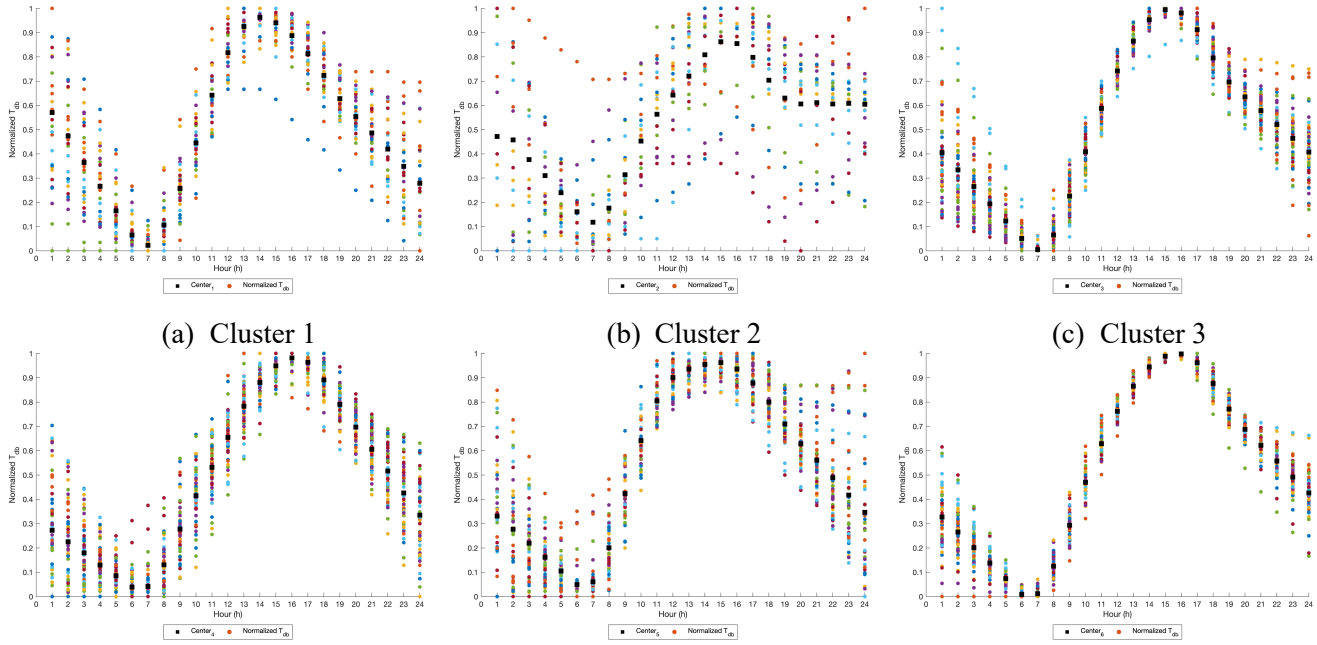
4. Results and discussion

So as to assess the accuracy and effectiveness of the proposed C-ANN predictive model, a deep insight into the clustering results of the daily weather data profiles is taken. After that, the *MAPE* values of each ANN sub-model and the year-round *MAPE* are assessed, while its improvement in prediction accuracy is investigated. Finally, the predicted cooling demands of the four representative days in each season are evaluated.

4.1. Clustering result of weather data profiles

As the lowest value of DB is obtained when $N_c = 6$, thus $N_{opt} = 6$. The year-round database **X** is divided into 6 clusters, while the datasets in each cluster is adopted to train one ANN sub-model. In order to evaluate the representative patterns of the daily weather data profiles, the normalized values of outdoor air dry-bulb temperature, relative humidity, total horizontal radiation and horizontal beam radiation at each hour during the 24-hour duration are summarized in Figs. 10-13, along with the corresponding centroid values. The black square markers in Figs. 10-13 represent the centroid value, while the colorful round ones stand for the values from each dataset $\boldsymbol{Y}_{m,i'}$. Through *k*-means clustering, representative patterns are identified in each cluster, covering the lowest and highest value occurring hour, as well as the most concentrated or scatted distribution.

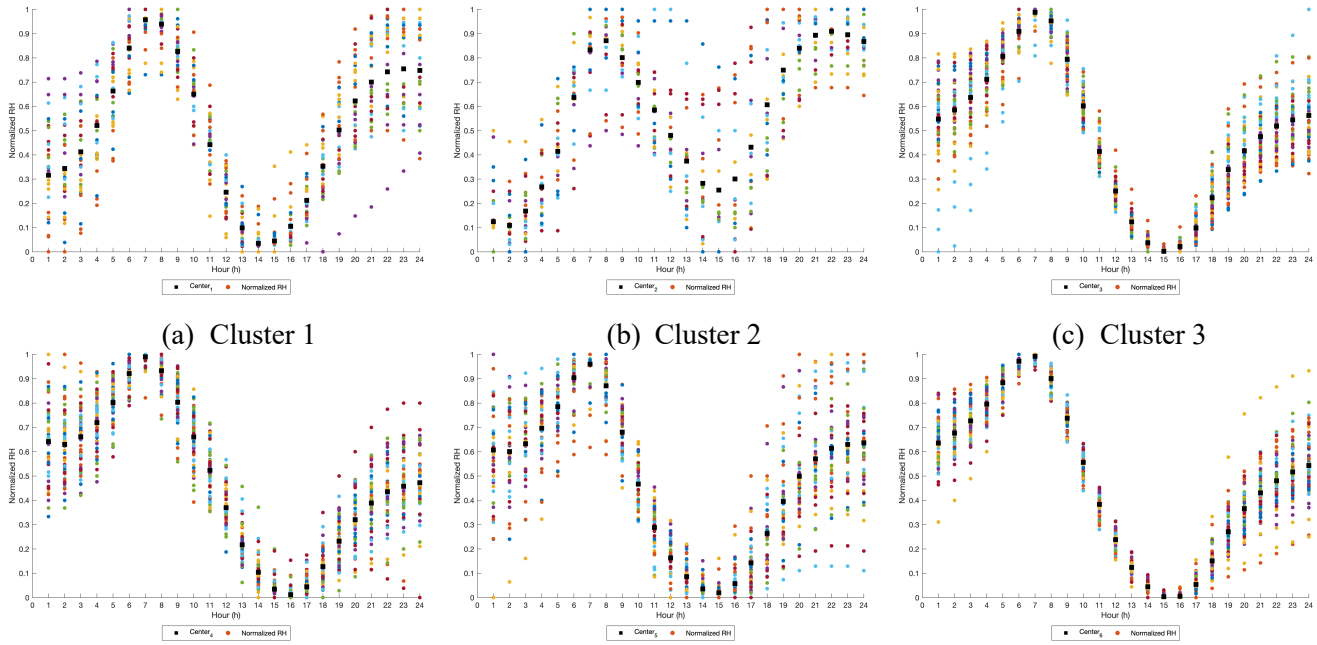
Generally, the outdoor air dry-bulb temperature decreases during the first 6th or 7th h, then increases till the 15th or 16th h, and finally decreased till the end of the day. It is found that the lowest value of dry-bulb temperature happened at the 7th h in Clusters 1-3, while at the 6th h in Clusters 4-6. The highest value occurs at the 14th, 15th, 15th, 16th, 15th and 16th h in Clusters 1, 2, 3, 4, 5 and 6, respectively. The largest daily temperature difference and the most concentrated points distribution is found in Cluster 6, while the smallest daily temperature difference and the most scatted points distribution is identified in Cluster 2.



(a) Cluster 1 (b) Cluster 2 (c) Cluster 3

(d) Cluster 4 (e) Cluster 5 (f) Cluster 6

Fig. 10. Clustering result of outdoor air dry-bulb temperature.



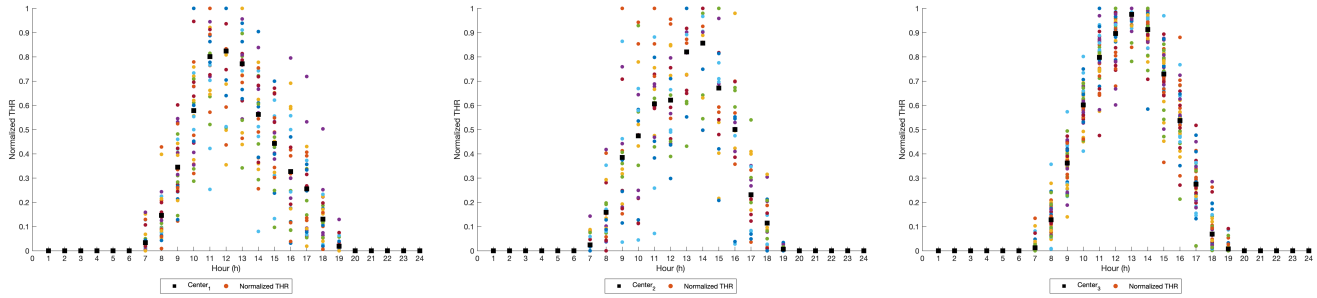
(a) Cluster 1 (b) Cluster 2 (c) Cluster 3

(d) Cluster 4 (e) Cluster 5 (f) Cluster 6

Fig. 11. Clustering result of outdoor air relative humidity.

The daily variation of outdoor air relative humidity is roughly inverse to that of outdoor air dry-bulb temperature: it increases during the first 6th or 7th h, then decreases till the 15th or 16th h, and increases at the end of the day. Although the highest value happens at the 7th h in each cluster, different representative patterns are identified in each cluster. The lowest value happens at the 14th, 15th, 15th, 16th, 15th and 15th h in Clusters 1, 2, 3, 4, 5 and 6, respectively. Similar to the phenomenon found in outdoor air dry-bulb temperature, the largest daily relative humidity difference and the most

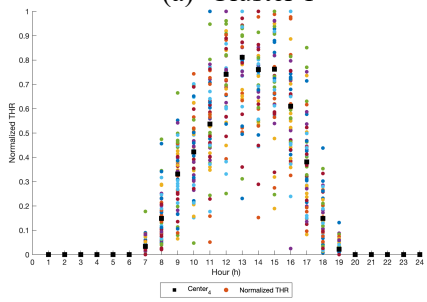
concentrated points distribution is detected in Cluster 6, while the smallest daily relative humidity difference and the most scatted points distribution is spotted in Cluster 2.



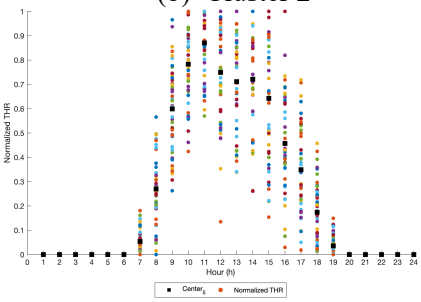
(a) Cluster 1

(b) Cluster 2

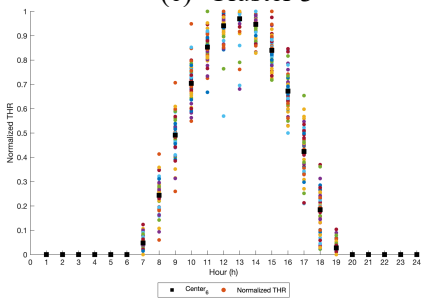
(c) Cluster 3



(d) Cluster 4

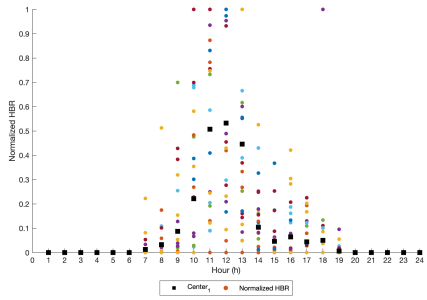


(e) Cluster 5

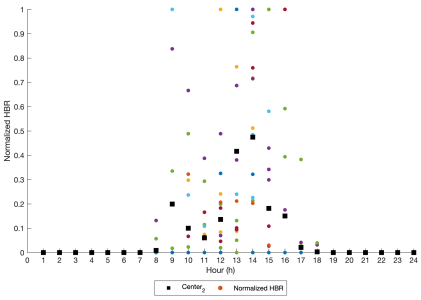


(f) Cluster 6

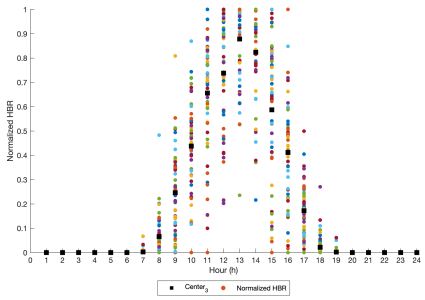
Fig. 12. Clustering result of total horizontal radiation.



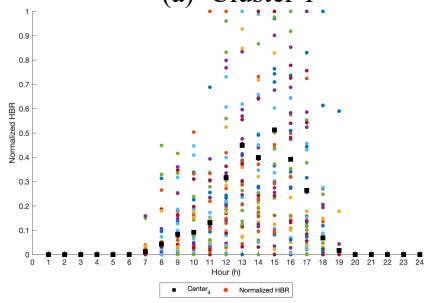
(a) Cluster 1



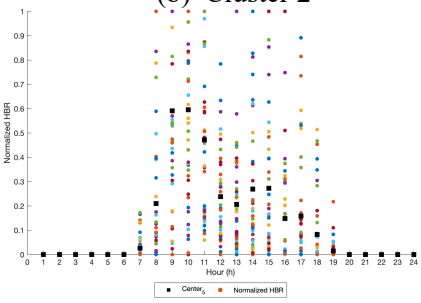
(b) Cluster 2



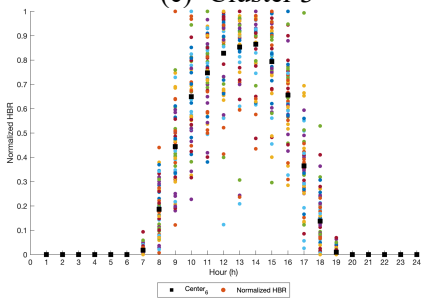
(c) Cluster 3



(d) Cluster 4



(e) Cluster 5



(f) Cluster 6

Fig. 13. Clustering result of horizontal beam radiation.

In Hong Kong, sunrise usually happens at around the 6th h, while sunset is at about the 19th h each day. There is no solar radiation during the 1-6th h and 20-24th h each day. The highest value of total horizontal radiation is found at the 12th, 14th, 13th, 13th, 11th and 13th h in Clusters 1, 2, 3, 4, 5 and 6, respectively.

The most concentrated distribution is found in Cluster 6, while the most scatted distribution is identified in Cluster 4.

Similarly, the horizontal beam radiation is 0 during the 1-6th h and the 20-24th h in each day. The highest value of horizontal beam radiation locates at the 12th, 14th, 13th, 15th, 10th and 14th h in Clusters 1, 2, 3, 4, 5 and 6, respectively. The most concentrated distribution is found in Cluster 6 while the most scatted distribution is identified in Cluster 5.

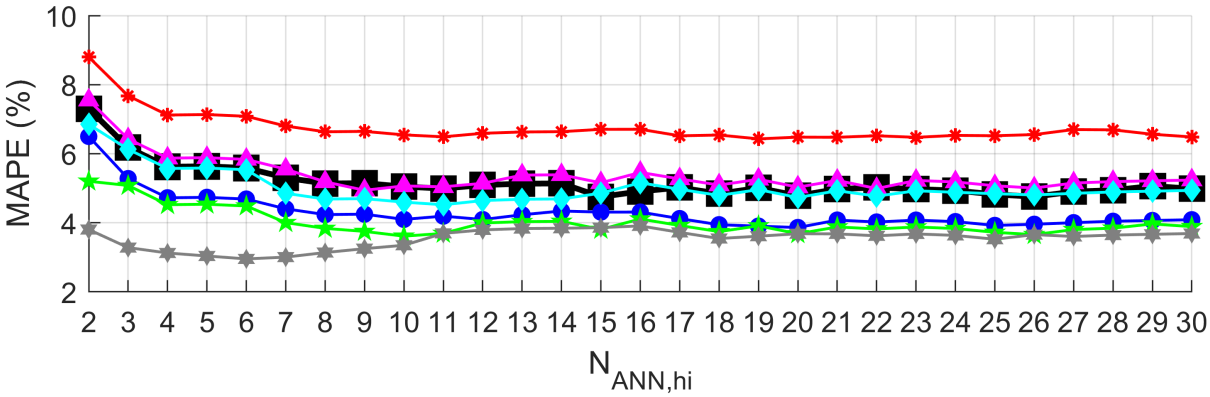
To summarize the above discussion, the featuring pattern identified in each cluster is summarized in Table 5.

Table **5**. Summary of featuring patterns in each cluster.

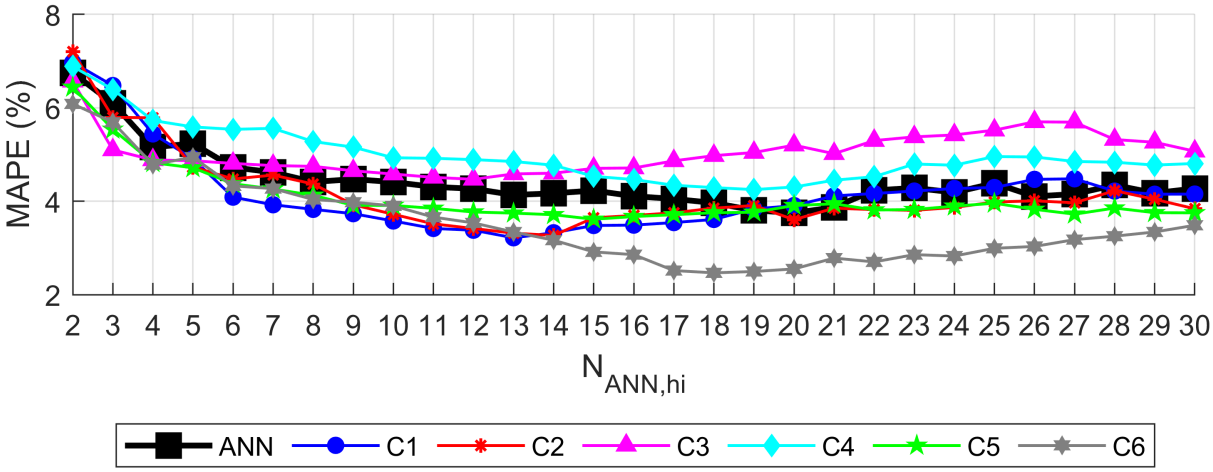
		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Dry-bulb temperature	Lowest value occurred time	7	7	7	6	6	6
	Highest value occurred time	14	15	15	16	15	16
	Most concentrated						
	Most scatted						
Relative humidity	Lowest value occurred time	14	15	15	16	15	16
	Highest value occurred time	7	7	7	7	7	7
	Most concentrated						
	Most scatted						
Global horizontal radiation	Highest value occurred time	12	14	13	13	11	13
	Most concentrated						
	Most scatted						
Horizontal beam radiation	Highest value occurred time	12	14	13	15	10	14
	Most concentrated						
	Most scatted						

4.2 Evaluation of MAPE

As sub-ANN models 1-6 are trained using different featuring datasets in clusters 1-6 respectively, the different optimal quantity of neurons in the single hidden layer N_{hi} is identified. The variation between *MAPE* and N_{hi} in each sub-ANN model is illustrated in Fig. 15, along with the conventional ANN model. The reference conventional ANN model is trained using the datasets from one entire year. For LM optimization, the optimal quantity of neurons for ANN sub-models 1-6 are 20, 19, 9, 11, 10 and 6, respectively, while the optimal quantity for the reference conventional ANN model is 15; For BR optimization, the optimal quantity of neurons for ANN sub-models 1-6 are 13, 14 12, 19, 15 and 18, respectively, while the optimal quantity for the reference conventional ANN model is 20.



(a) Levenberg-Marquardt optimization



(b) Bayesian Regularization optimization

Fig. 15. Variation between *MAPE* and N_{hi} in conventional ANN model and proposed sub-ANN models.

According to the clustering result, there are 42, 29, 74, 77, 61 and 82 days in clusters 1-6 for the training case, respectively; while there are 43, 31, 73, 79, 59 and 80 days in clusters 1-6 for the testing case, respectively. The *MAPE* values of the conventional ANN model and the proposed C-ANN predictive models are summarized in Table 6. For training case of the LM optimization-based C-ANN, the *MAPE* value from sub-models 1, 4, 5 and 6 are smaller those from the conventional ANN model; For testing case of the LM optimization-based C-ANN, the *MAPE* value from sub-models 1, 2, 4 and 6 are smaller those from the conventional ANN model; For training case of the BR optimization-based C-ANN, the *MAPE* value from sub-models 1, 2, 5 and 6 are smaller those from the conventional ANN model; For testing case of the BR optimization-based C-ANN, the *MAPE* value from sub-models 1, 2 and 6 are smaller those from the conventional ANN model. Therefore, in training cases, the reduction of overall *MAPE* value is 12.1% and 4.2% when using LM and BR optimization, respectively. In testing cases, the reduction of annual average *MAPE* value is 4.2% and 3.1% when using LM and BR optimization, respectively. Overall, the BR optimization resulted in the best performance, with the *MAPE* value of 3.59% and 4.71% at training and testing cases, respectively.

Table 6. *MAPE* of conventional ANN and proposed C-ANN predictive models.

Predictive models		Quantity of days in each cluster	Quantity of days in each cluster	LM optimization		BR optimization	
				MAPE (%)			
				Training	Testing	Training	Testing
Conventional ANN model		N.A.	N.A.	4.75	4.98	3.75	4.86
C-ANN sub-models	1	42	43	3.86	4.71	3.22	4.51
	2	29	31	6.43	4.02	3.27	2.34
	3	74	73	4.94	5.06	4.47	5.16
	4	77	79	4.52	4.89	4.25	5.20
	5	61	59	3.61	5.42	3.61	5.10
	6	82	80	2.95	4.07	2.47	4.56
	Overall			4.16	4.73	3.59	4.71
	Reduction			12.1%	4.9%	4.2%	3.1%

In addition, a two-hidden-layer C-ANN model trained by BR optimization is adopted as reference. The same quantity of neurons is adopted in the first and second layer in each of the C-ANN sub-models. According to the result in Fig. 15(b), the quantity is designed as 7, 7, 6, 10, 8 and 9 for sub-models 1-6, respectively. The overall performance and the *MAPE* value of each sub-model is summarized in Table 7. Although the *MAPE* value of each ANN sub-model and the overall *MAPE* in training case is smaller than the proposed C-ANN model with a single hidden layer, the overall *MAPE* value from the testing case is similar. It might be due to the over-fitting of the ANN training. Therefore, it is concluded that the proposed single-hidden-layer C-ANN model is sufficient and accurate enough for building cooling demand prediction.

Table 7. *MAPE* of the reference 2-hidden-layer C-ANN predictive model.

MAPE (%)	C-ANN sub-models						Overall
	1	2	3	4	5	6	
Training	3.14	3.20	4.41	4.18	3.54	2.41	3.45
Testing	4.54	2.32	5.25	5.14	5.03	4.61	4.73

4.3 Cooling demand prediction

To take a deep look into the day-ahead prediction performance, the cooling demand prediction results from 4 representative days in each season are presented in Fig. 14 and Fig. 15, for training and testing cases, respectively. Compared to the conventional single ANN predictive model, it is found that the predicted cooling demand of proposed predictive model, no matter through LM and BR optimization, is closer to the simulated cooling demand from TRNSYS model. This is consistent with the discussion in Section 4.2.

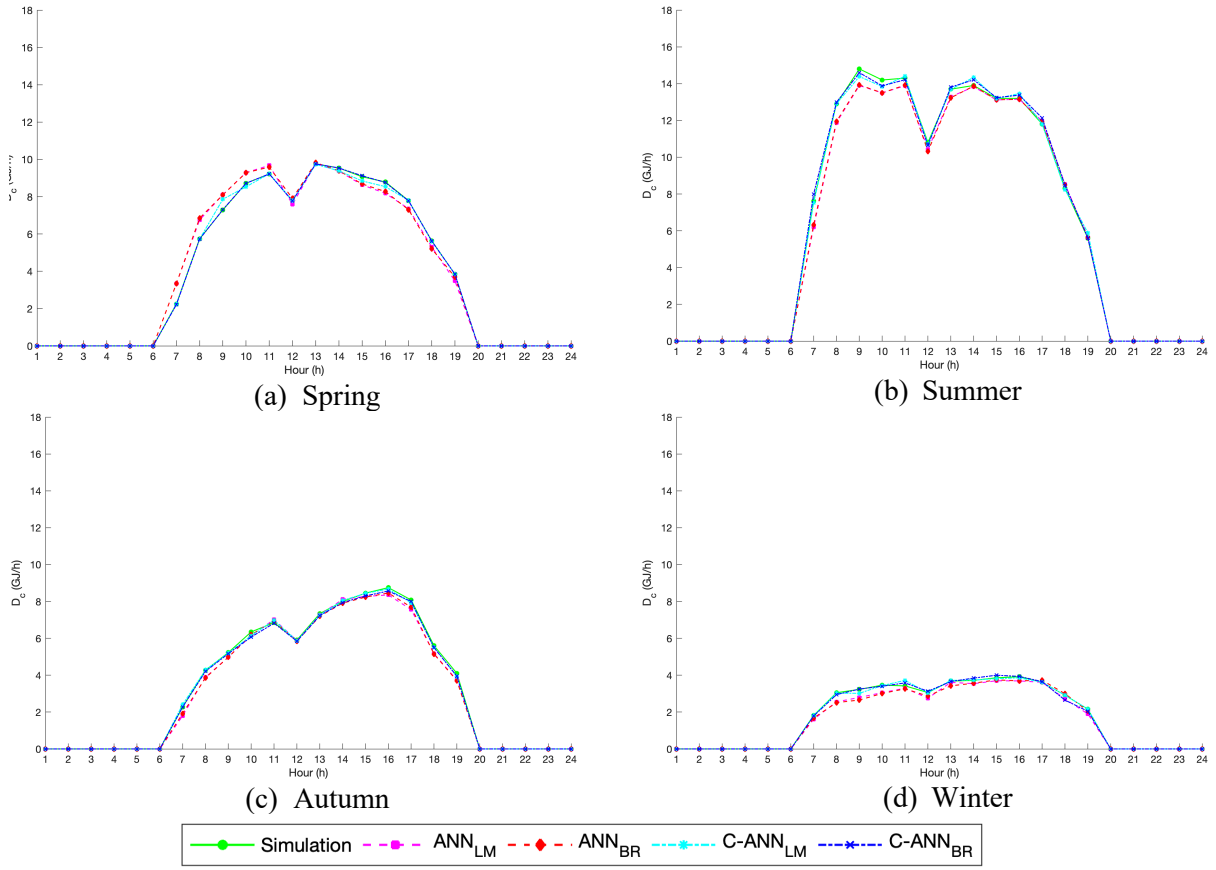


Fig. 13. Training result of cooling demand prediction.

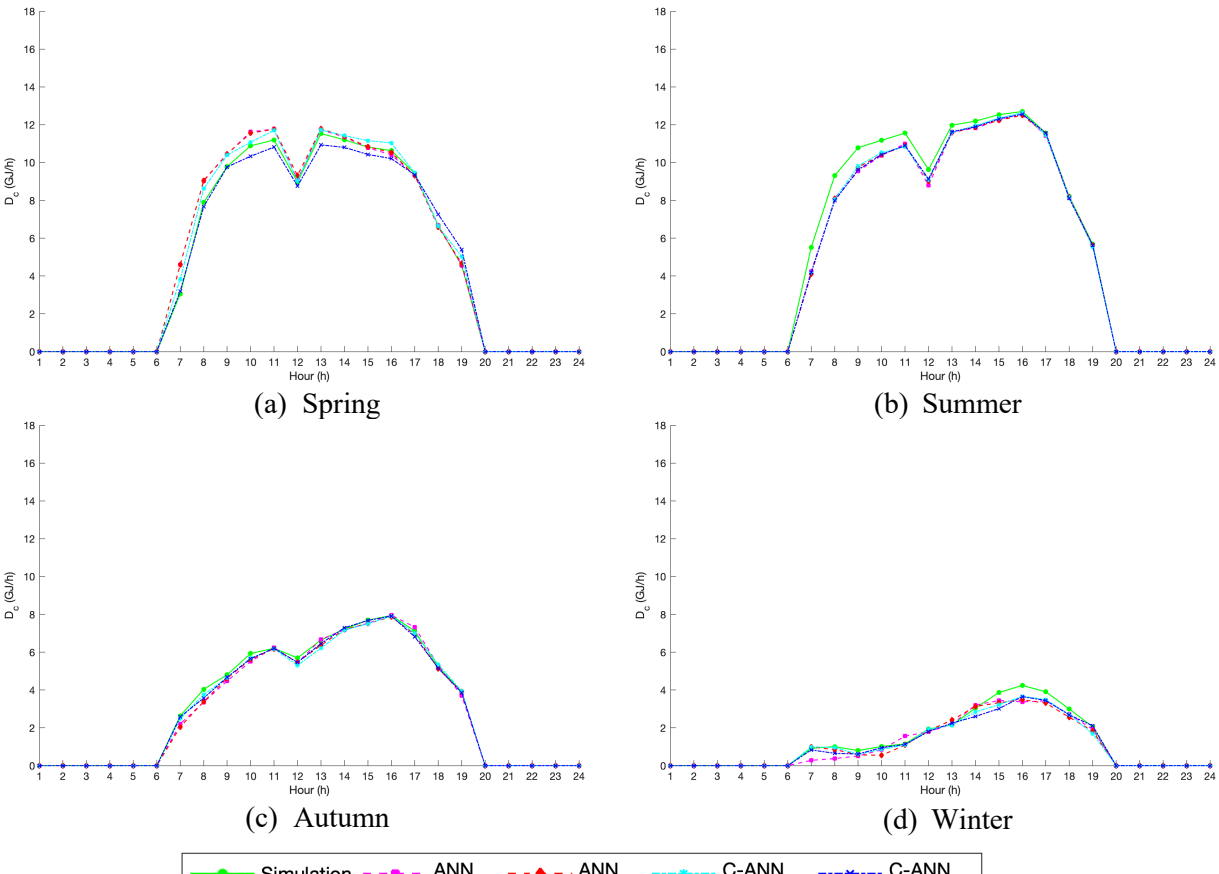


Fig. 14. Testing result of cooling demand prediction.

5. Implication for practice and future direction

In this study, to generate the historical database for training and testing cases for the proposed C-ANN predictive model, the year-round weather profile is assumed to be the same as those of the typical meteorological year from Hong Kong Observatory and Hong Kong King's Park; the building operating schedules are assumed to follow the local guidelines; while the cooling demand is calculated using the TRNSYS simulation model. In practical application, the historical outdoor air dry-bulb temperature, relative humidity, total horizontal radiation and horizontal beam radiation should be obtained from the local weather station. The cooling demand should be estimated from the operating parameters of the chilled water system in the building management system. After training the C-ANN model using past year's historical data, the day-ahead building cooling demand can be predicted using the latest forecast of weather profile acquired from the weather reporting website [45]. It can be adopted in the building management system in predicting the day-ahead cooling demand, thus determining the operating schedules of various equipment units such as chilled water system and air handling units.

Compared to practical measurement data which may have sensor or equipment faults, the simulated data from TRNSYS 18 is noise-free. Since the simulation data is generated from the well-validated model, it provides an ideal scenario under which the variation of the cooling demand is more predictable. It can serve as the first step towards the understanding, developing and testing the proposed C-ANN model. In the future study, the performance evaluation of the proposed C-ANN model should be tested on real measurement data. The predictive model should be further refined to tackle the problem caused by the probable faulty data owing to sensor and equipment faults.

Moreover, *k*-means clustering analysis is hybrid with ANN predictive model. It remains to be seen whether it will result in better prediction performance when *k*-means clustering is integrated with support vector regression, long-short term method and other data-driven predictive models. Moreover, the performance of other types of clustering algorithms in pattern identification, such as density-based, distribution-based, connectivity-based and hierarchical-based clustering algorithm should be investigated.

6. Conclusion

To improve the accuracy, effectiveness and robustness of cooling demand prediction, a novel clustering-enhanced adaptive artificial neural network predictive model is proposed in this study. The major contribution of this study is the unique adaptability feature of the proposed model. To be more specific, *k*-means clustering analysis is firstly used to recognise the representative patterns of daily weather data profile and group the year-round working days' profiles into different clusters. After that,

datasets in each cluster are used to train one specific adaptive ANN sub-model. The optimal structure and parameters are determined for each ANN sub-model based on the distinctive pattern of daily weather profile in each cluster. Since each ANN sub-model is adaptive to the features of database of input variables, the prediction performance of the proposed clustering-enhanced adaptive artificial neural network predictive model is improved compared to the conventional ANN predictive model with a fixed structure.

To evaluate the performance of the clustering-enhanced adaptive artificial neural network predictive model, the whole-year typical weather data profiles from two different locations, along with building the corresponding cooling demand profile, are established as two databases for training and testing purposes, respectively. In training cases, the reduction of annual *MAPE* value from the proposed adaptive model is 12.1% and 4.2% when using LM and BR optimization, respectively, compared to the conventional fixed predictive model; In testing cases, the reduction of annual average *MAPE* value is 4.9% and 3.1% when using LM and BR optimization, respectively, compared to the conventional fixed predictive model. It is also found that the proposed adaptive C-ANN model with BR optimization has the best prediction performance, with *MAPE* value of 3.59% and 4.71% at training and testing cases, respectively. Through comparison with the C-ANN model with two hidden layers, it is also demonstrated that the single-layer is sufficient enough.

Different consumption patterns result in different energy sequences. In general office buildings, the similar operating schedules of lighting, office equipment, air conditioner and lift can be identified on each weekday. Therefore, time variables can be adopted to represent those operating schedules. The representative high-rise office building is tested in this study, and it is expected that the proposed C-ANN model is general and can be adopted in other office buildings. However, for special office buildings such as event venues, the operating schedule is not consistent. Thus, time variables cannot simply represent complicated operating schedules. The cooling demand prediction lies in the comprehensive effects of both exogenous parameters (weather conditions) and endogenous parameters (stochastic events). The schedules of the stochastic events should be adopted to train the ANN model instead of general time variables.

Nomenclature

c Data point of cluster centroid
C Dataset of cluster centroid
$\mathbf{C}$ Database of cluster centroid
d Data element in output database $\mathbf{D}$
dw Day of the week
$\tilde{d}$ Predicted data element in output database $\widetilde{\mathbf{D}}$
D Dataset in database $\mathbf{D}$
$\widetilde{D}$ Predicted dataset in output database $\widetilde{\mathbf{D}}$

D Cooling demand database
$\widetilde{\mathbf{D}}$ Predicted cooling demand database
E Squared error
hd Hour of the day
l Number of the neuron in the hidden layer of sub-ANN model
m Number of the cluster
N Quantity
u Clustering result
w Weight of the connection in ANN
x' Original data in database
x Normalized data in database
X Daily profile of a certain type of weather data
$\boldsymbol{X}$ Daily profile of weather data
X Historical database of weather profile before clustering
y Data element in database **Y**
$\boldsymbol{Y}$ Dataset in each group
Y Database in each group
Z Dataset in database **Z**
Z Database for ANN
$\| \|$ Euclidean distance

Subscripts
c Cluster
d Day
h Hour
hi Hidden layer of ANN model
i Number of working days
in Input layer of ANN model
j Type of weather data
k Hour of the day
m Cluster number
opt Optimal quantity of clusters

Abbreviations

ANN Artificial neural network
BR Bayesian Regularization
C-ANN Clustering enhanced artificial neural network
LM Levenberg-Marquardt
MAPE Mean absolute percentage error

References

[1] 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.
[2] Luo XJ and Fong KF. Control optimization of combined cooling and power system with prime mover of solid oxide fuel cell-gas turbine for building application. Energy Procedia. 105(2017)1883-1888.
[3] 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.
[4] Luo XJ, Lukumon O, Anuoluwapo A, Chukwuka M, Olugbenga A and Lukman A. Development of an IoT-based big data platform for day-ahead prediction of building heating and cooling demands. Advanced Engineering Informatics. 41(2019)100926.

[5] Zhao HX and Frédéric M. A review on the prediction of building energy consumption. Renewable and Sustainable Energy Reviews. 16(2012)3586-3592.
[6] Wei Y, Zhang X, Shi Y, Xia L, Pan S, Wu J and Zhao X. A review of data-driven approaches for prediction and classification of building energy consumption. Renewable and Sustainable Energy Reviews. 82(2018)1027-1047.
[7] Amasyali Kadir and Nora M. A review of data-driven building energy consumption prediction studies. Renewable and Sustainable Energy Reviews. 81(2018)1192-1205.
[8] TRNSYS 18: A Transient System Simulation Program. Solar Energy Laboratory University of Wisconsin, Madison, USA.
[9] EnergyPlus Engineering Reference: The Reference to EnergyPlus Calculations. Ernest Orlando Lawrence Berkeley National Laboratory. (2013)
Available at: http://apps1.eere.energy.gov/buildings/energyplus/pdfs/engineeringreference.pdf
[10] ESP-r. 2000. ESP-r Overview. http://www.esru.strath.ac.uk/Programs/ESP-r.htm
[11] Kusiak A, Li MY and Zhang ZJ. A data-driven approach for steam load prediction in buildings. Applied Energy. 87(2010)925-933.
[12] Jovanović RŽ, Sretenović AA and Živković BD. Ensemble of various neural networks for prediction of heating energy consumption. Energy and Buildings. 94(2015)189-199.
[13] Deb C, Eang LS, Yang J and Santamouris M. Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks. Energy and Buildings. 121(2016)284-297.
[14] Yang J, Hugues R and Radu Z. On-line building energy prediction using adaptive artificial neural networks. Energy and buildings. 37(2005)1250-1259.
[15] Ahmad T, Chen H, Shair J and Xu C. Deployment of data-mining short and medium-term horizon cooling load forecasting models for building energy optimization and management. International Journal of Refrigeration. 98(2019)399-409.
[16] Ahmad T and Chen HX. 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.
[17] Wang L, Eric WML and Richard KKY. Novel dynamic forecasting model for building cooling loads combining an artificial neural network and an ensemble approach. Applied Energy. 228(2018)1740-1753.
[18] Ding Y, Zhang Q, Yuan T and Yang F. Effect of input variables on cooling load prediction accuracy of an office building. Applied Thermal Engineering. 128(2018)225-234.
[19] Li K, Xie X, Xue W, Dai X, Chen X and Yang X. A hybrid teaching-learning artificial neural network for building electrical energy consumption prediction. Energy and Buildings. 174(2018)323-334.
[20] Katsatos AL, Moustris KP. Application of artificial neuron networks as energy consumption forecasting tool in the building of regulatory authority of energy, Athens, Greece. Energy Procedia. 157(159)851-861.
[21] Luo XJ, Fong KF, Sun YJ and Leung MKH. Development of clustering-based sensor fault detection and diagnosis strategy for chilled water system. Energy and Buildings. Energy and Buildings. 186(2019)17-36.
[22] 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.
[23] Fazlollahi S, Bungener SL, Mandel P, Becker G and Maréchal F. Multi-objectives, multi-period optimization of district energy systems: I. Selection of typical operating periods. Computers & Chemical Engineering. 65(2014)54-66.
[24] Schiefelbein J, Tesfaegzi J, Streblow R and Müller D. Design of an optimization algorithm for the distribution of thermal energy systems and local heating networks within a city district. Proceedings of ECOS (2015).
[25] Domínguez-Muñoz F, Cejudo-López JM, Carrillo-Andrés A and Gallardo-Salazar M. Selection of typical demand days for CHP optimization. Energy and buildings. 43(2011)3036-3043.
[26] Luo XJ, Oyedele L, Akinade G and Ajayi A. 3D pattern identification approach for cooling load profiles in different buildings. Journal of Building Engineering, 31(2020)101339.
[27] Famuyibo AA, Duffy A and Strachan P. Developing archetypes for domestic dwellings—An Irish case study. Energy and Buildings, 50(2012)150-157.

[28] Chen YB, Tan H and Berardi U. Day-ahead prediction of hourly electric demand in non-stationary operated commercial buildings: A clustering-based hybrid approach. Energy and Buildings, 148(2017)228-237.
[29] Luo XJ, Oyedele L, Ajayi A, Monyei C, Akinade G and Akanbi L. Development of an IoT-based big data platform for day-ahead prediction of building heating and cooling demands. Advanced Engineering Informatics, 41(2019)100926.
[30] Davies DL and Bouldin DW. A cluster separation measure. IEEE transactions on pattern analysis and machine intelligence. 1979(2)224-7.
[31] McCulloch WS and W Pitts. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics. 5(1943)115-133.
[32] EMSD. Performance-based building energy code. Electrical and Mechanical Services Department, the Hong Kong Special Authority Region. 2007.
[33] EMSD. Code of practice for energy efficiency of building services installation. Electrical and Mechanical Services Department. The Hong Kong Special Authority Region. 2012.
[34] American Society of Heating, Refrigerating and Air-Conditioning Engineers. ASHRAE handbook: fundamentals (Inch-pound ed.). ASHRAE. Atlanta (GA) (2013) 4.15-4.16.
[35] Al-ajmi F, Hanby V. Simulation of energy consumption for Kuwaiti domestic buildings. Energy and Buildings. 40 (2008)1101-1109.
[36] Chargui R, Sammouda H. Modeling of a residential house coupled with a dual source heat pump using TRNSYS software. Energy Convers Manage. 81(2014)384-399.
[37] Cacabelos A, Eguia P, Migues J, Granada E, Arce M. Calibrated simulation of public library HVAC system with a ground-source heat pump and a radiant floor using TRNSYS and GeoOpt. Energy and Buildings. 108(2015)114-126.
[38] Djedjig R, Bozonnet E, Belarbi R. Analysis of thermal effects of vegetated envelopes: Integration of a validated model in a building energy simulation program. Energy and Buildings. 86 (2015)93-103.
[39] Ruiz-Calvo F, Montagud C, Cazorla-Marin A, Corberan J. Development and experimental validation of a TRNSYS dynamic tool for design and energy optimization of ground source heat pump systems. Energies. 10(2017)1510.
[40] Lomas KJ, Eppel H, Martin CJ and Bloomfield DP. Empirical validation of building energy simulation programs. Energy and buildings. 26(1997)253-275.
[41] Blair N and Holst S. BESTEST Results and Experiences using the latest TRNSYS Building Model (TRNSYS Version 14.2). Sophia Antipolis: Centre Scientifique et Technique du Bâtiment. 1998.
[42] Holst S. Heating load of a building model in TRNSYS with different heating systems. ZAE Bayern, Abt, 4. 1993.
[43] Loutzenhiser P, Manz H, Felsmann C, Strachan P, Frank T and Maxwell G. Empirical validation of models to compute solar irradiance on inclined surfaces for building energy simulation. Solar Energy. 81(2007)254-267.
[44] Fong KF and Lee CK. Performance analysis of internal-combustion-engine primed trigeneration systems for use in high-rise office buildings in Hong Kong. Applied Energy. 160 (2015)793-801.
[45] AccuWeather. Website: http://www.accuweather.com/en/hk/hong-kong/1123655/hourly-weather-forecast/1123655. (last accessed on 26 Feb 2020).