Available online at www.sciencedirect.com

SCIENCE DIRECT

International Journal of Thermal Sciences 43 (2004) 1107–1118

International Journal of Thermal Sciences

www.elsevier.com/locate/ijts

Hourly cooling load prediction by a combined forecasting model based on Analytic Hierarchy Process

Ye Yao [a], Zhiwei Lian [a,*], Shiqing Liu [b], Zhijian Hou [a]

[a] *Institute of Refrigeration & Cryogenics, Shanghai Jiao Tong University, Shanghai 200030, China*
[b] *Institute of Mathematics and Physics, Nanhua University, Hunan 421001, China*

Received 16 July 2003; received in revised form 13 February 2004; accepted 23 February 2004

Available online 7 May 2004

Abstract

Predicting the next-24-hour load in a building is essential for the optimal control of heating, ventilating and air-conditioning (HVAC) systems that use thermal/cool storage technology and also for cost and energy reduction of the non-storage systems. To fully integrate the advantages of several models and improve the accuracy of forecasting load, the application of the combined forecasting method to hourly load forecasting is presented in this paper. The method of Analytic Hierarchy Process (AHP) is employed to deduce the weights of each model. A case study shows that the combined forecasting model based on AHP may be better than the individual ones in predicting the building's hourly load for the future hours.

Keywords: Air-conditioning; Load; Combined forecasting; Analytic Hierarchy Process (AHP)

1. Introduction

Accurate prediction of the dynamic air-conditioning load in a building is a key for HVAC system design. It is also useful in HVAC operations including adjusting the starting time of cooling to meet start-up loads, minimizing or limiting the electric on-peak demand, load prediction to optimize costs and energy use for cool storage systems, and related energy and cost optimization needs in other HVAC systems [1]. MacArthur et al. [2] described a method for optimal control of cool storage systems that requires forecasts of both cooling loads and non-cooling electrical demand. Stoecker et al. [3] and Braun [4] have showed that the load requirements of a building's might be shifted significantly through management of the building's thermal storage. Forecasting cooling load for the future hours is very necessary in order to determine the optimal control that minimizes the total operating cost of the thermal energy storage systems.

Several prediction techniques have been previously investigated. Forrester and Wepfer [5] presented a method based on an extensive multiple linear regression (LR) technique (*see Appendix A for a brief review of the LR model*) that predicts electrical demand up to twenty-four hours in advance. MacArthur et al. [2] and Spethmann [6] developed a prediction method based on the autoregressive integrated moving average (ARIMA) model (*see Appendix B for a brief review of the ARIMA model*) and applied it to an optimal cold storage controller. Minoru Kawashima et al. [7] described an artificial neural network (ANN) model (*see Appendix C for a brief review of the ANN model*) to predict the next day's total cooling load. Kreider and Wang [8] demonstrated an automated load predictor using the ANN model. Anstett and Kreider [9] examined the accuracy of the ANN model for energy predictions. The grey system theory was initially presented by Deng [10,11] and it has been successfully used in the forecasting. The advantages of the grey model (GM) (*see Appendix D for a brief review of the GM model*) include: (a) it can be used in circumstances with relatively little data; as low as four observations were reported [12] to estimate the outcome of an unknown system; and (b) it can used a first-order differential equation to characterize a system. Therefore, only a few discrete data are sufficient to characterize an unknown system. Hwang et al. [13] used the grey relation to select the influential factors for power-load forecasting and build the forecasting model.

* Corresponding author.
E-mail addresses: zwlian@sjtu.edu.cn, yeyao10000@sohu.com (Z. Lian).

1290-0729/$ – see front matter © 2004 Elsevier SAS. All rights reserved.
doi:10.1016/j.ijthermalsci.2004.02.009

Though there are various forecasting models mentioned above, no single one has performed well enough because each model can take just several or usually only one relevant factor into consideration. In practical applications, engineers often try several kinds of models to satisfy the actual need better. The result of each forecasting model is compared and analysis has to be done by experienced forecasters to get the best forecasting result.

To fully utilize the useful information from the models, the combined forecasting method is introduced in this paper. It is one of the most popular subjects in the field of forecasting methods [14–16]. The theory of the combined forecasting method is based on a certain linear combination of various results from different forecast models. The fitting capacity of the combined forecasting model is greatly improved, and the forecasted result will show higher precision [17]. Formulations have been developed in the past literatures [18] for the optimal combined forecasting method, whose deviation reaches the minimum and is less than that of each single forecasting method. The application of the combined forecasting method can combine separate methods and integrate merits of each model to provide a more accurate result.

2. Principles of the combined forecasting method

For a certain forecasting problem, assume the actual value in period t is y_t $(t = 1, 2, \dots, n)$ and there are m kinds of forecasting models. Let the forecasting value in period t by model i is f_{it} $(i = 1, 2, \dots, m)$, then the corresponding deviation is $e_{it} = y_t - f_{it}$. Suppose the weights vector is $W = [w_1, w_2, \dots, w_m]^{\mathrm{T}}$, the combined forecasting model can be expressed as follows:

$$\hat{y}_t = \sum_{i=1}^{m} w_i f_{it} \quad (t = 1, 2, 3, \dots, n) \tag{1}$$

$$\sum_{i=1}^{m} w_i = 1 \tag{2}$$

Eq. (1) can also be substituted by Eq. (2):

$$\widehat{Y} = FW \tag{3}$$

where, $\widehat{Y} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n]^{\mathrm{T}}$, $F = [f_{it}]_{n\times m}$.

The forecasting error of combined model can be written as:

$$\begin{aligned} e_t = y_t - f_t &= \sum_{i=1}^{m}(w_i y_t) - \sum_{i=1}^{m}(w_i f_{it}) \\ &= \sum_{i=1}^{m} w_i (y_t - f_{it}) = \sum_{i=1}^{m}(w_i e_{it}) \end{aligned} \tag{4}$$

Although the combined model cannot improve the forecasting accuracy essentially, it may take advantage of the "randomness" of the errors to reduce the forecasting error. For example, when the deviations of all the models are not in the same direction, the errors can counteract partially each other in the combined forecasting.

The key of the combined forecasting method is to determine the weights of each model. There are a variety of methods available to determine the weights used in the combination of forecasts. The equal weights (EW) method that uses an arithmetic average of the individual forecasts is a very simple approach. It does not require information about the precision of the forecasts or the correlations between their errors. However, the method treats the forecasts as though they are exchangeable and indistinguishable from one another. While this may be a reasonable assumption when the models have similar error variances, it is in general not appealing. The minimum-variance (MV) method is a Bayesian approach for combining individual forecasts. The combination weights proposed by the MV method are less reliable when the data are sparse or unstable [14]. In this paper, Analytic Hierarchy Process (AHP) is employed to deduce the weights of each model.

3. Determining weights by Analytic Hierarchy Process (AHP)

AHP is an intuitive method for formulating and analyzing decisions. AHP has been applied to numerous practical problems in the last few decades [19]. Because of its intuitive appeal and flexibility, many corporations and governments routinely use AHP for making major policy decisions [20]. A brief discussion of AHP is provided in this section. More detailed description of AHP and application issues can be found elsewhere [21–24]. Application of AHP to a decision problem involves four steps (see below).

Step 1: structuring of the decision problem into a hierarchical model

It includes decomposition of the decision problem into elements according to their common characteristics and the formation of a hierarchical model having different levels. Each level in the hierarchy corresponds to the common characteristic of the elements in that level. The topmost level is the 'focus' of the problem. The intermediate levels correspond to criteria and sub-criteria, while the lowest level contains the "decision alternatives".

Step 2: making pair-wise comparisons and obtaining the judgment matrix

In this step, the elements of a particular level are compared pair-wise, with respect to a specific element in the immediate upper level. A judgment matrix is formed and used for computing the priorities of the corresponding elements. First, criteria are compared pair-wise with respect to the goal. A judgment matrix, denoted as B, will be formed

Table 1
The semantic scale used in AHP

Intensity of importance	Definition	Description
1	Equal importance	Elements B_i and B_j are equally important
3	Weak importance of B_i over B_j	Experience and Judgment slightly favor B_i over B_j
5	Essential or strong importance	Experience and Judgment strongly favor B_i over B_j
7	Demonstrated importance	B_i is very strongly favored over B_j
9	Absolute importance	The evidence favoring B_i over B_j is of the highest possible order of affirmation
2, 4, 6, 8	Intermediate	When compromise is needed, values between two adjacent judgments are used
Reciprocals of the above judgments	If B_i has one of the above judgments assigned to it when compared with B_j, then B_j has the reciprocal value when compared with B_i	A reasonable assumption

using the comparisons. Each entry b_{ij} of the judgment matrix is formed comparing the row element B_i with the column element B_j:

$$B = (b_{ij}) \quad (i, j = 1, 2, \ldots, \text{the number of criteria}) \tag{5}$$

The comparison of any two criteria C_i and C_j with respect to the goal is made using questions of the type: "of the two criteria C_i and C_j, which is more important and how much more?" Saaty [24] suggests the use of a 9-point scale to transform the verbal judgments into numerical quantities representing the values of b_{ij}. The scale is explained in Table 1. Larger number assigned to the pair-wise comparisons means larger differences between criterion levels. Thus, in comparison to the numerical mode, the verbal mode is expected to predict larger differences between criterion levels. This implies a larger range between the weights of the most preferred criterion level and the least preferred criterion level. Provided the example of a decision maker who prefers alternative "A" slightly to alternative "B", the AHP interprets this verbal statement as the numerical score 3, implying that the decision maker prefers alternative A three times as much as alternative B. Given the meaning of the word 'slightly' in the regular use of language, the score 3 is probably an overestimation of the difference as perceived by the decision maker. The same applies to the other verbal judgments in the AHP.

The entries b_{ij} are governed by the following rules:

$$b_{ij} > 0, \qquad b_{ij} = 1/b_{ji}, \qquad b_{ii} = 1 \quad \text{for all } i \tag{6}$$

Because of the above rules, the judgment matrix B is a positive reciprocal pair-wise comparison matrix.

Table 2
The average consistencies of random matrices (RI) [21–24]

Size (n)	1	2	3	4	5	6	7	8	9
RI	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45

Step 3: local priorities and consistency of comparisons

Once the judgment matrix of comparisons of criteria with respect to the goal is available, the local priorities of criteria are obtained and the consistency of the judgments is determined. It has been generally agreed that priorities of criteria can be estimated by finding the principal eigenvector w of the matrix B. That is:

$$Bw = \lambda_{\max} w \tag{7}$$

When the vector w is normalized, it becomes the vector of priorities of the criteria with respect to the goal. $\lambda_{\max}$ is the largest eigenvalue of the matrix B and the corresponding eigenvector w contains only positive entries. The consistency of the judgment matrix can be determined by a measure called the consistency ratio (CR), defined as:

$$CR = \frac{CI}{RI} \tag{8}$$

where, *CI* is called the consistency index and *RI*, the Random Index.

CI is defined as:

$$CI = \frac{(\lambda_{\max} - n)}{(n - 1)} \tag{9}$$

where, n is the matrix size.

RI is the consistency index of a randomly generated reciprocal matrix from the 9-point scale, with reciprocals forced. Saaty [21–24] has provided average consistencies (*RI* values) of randomly generated matrices (up to size 11×11) for a sample size of 500. The *RI* values for matrices of different sizes are shown in Table 2.

If *CR* of the matrix is higher, it means that the input judgments are not consistent, and hence are not reliable. Generally, it is acceptable only if $CR < 0.10$. Using a very similar procedure, the local priorities of alternatives with respect to each criterion can be estimated.

Step 4: aggregation of local priorities

Once the local priorities of elements of different levels are available as outlined in the previous step, they are aggregated to obtain final priorities of the alternatives. For aggregation, the following principle of hierarchic composition [24] is used:

$$\begin{aligned}&\text{Final local priority of decision alternative}\\ &\quad = \sum_{i=1}^{n} (\text{Local priority of decision alternative}\\ &\qquad \text{with respect to } C_i \times \text{Local priority of } C_i\\ &\qquad \text{with respect to the goal})\end{aligned} \tag{10}$$

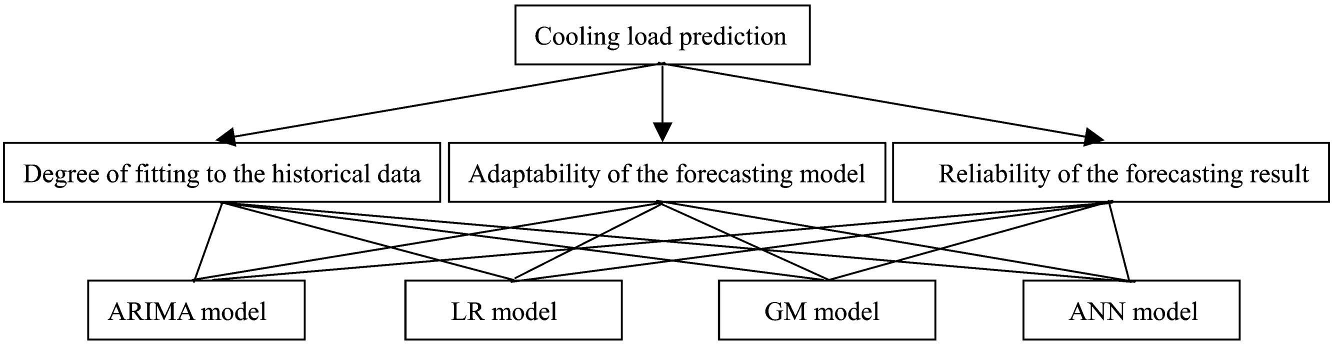

Fig. 1. The AHP model for cooling load combined forecasting.

Note that the above is a simple weighted summation. The final priorities thus obtained represent the rating of the alternatives in achieving the focus of the problem. In this study, local priorities of decision alternative stand for the weight of each forecasting model in the combined forecasting, respectively.

4. Combined forecasting model for Hourly cooling load prediction using AHP

To establish the combined forecasting model using AHP, the common characteristics of cooling load prediction ought to be known. Generally, forecasting is made on the basis of the historical data. So the degree of fitting to the historical data is one of the elements that are under consideration during the forecasting. In addition, the adaptability and the reliability are another two important elements that are taken into account in evaluating an individual forecasting model. Adaptability refers to the ability the forecasting model has to adapt to the fickle environments, and the reliability refers to the accuracy of forecasting.

In this study, authors only consider the three elements (degree of fitting to the historical data, adaptability and reliability) that impact on the effect of cooling load prediction. Thus, a hierarchical model having three levels for cooling load combined forecasting can be formed, as is shown in Fig. 1. In this model, cooling load prediction is reckoned as the 'focus' of the problem, which is in the topmost level. The intermediate level corresponds to criteria that include the three elements mentioned above, while the lowest level contains four forecasting models that are reckoned as the "decision alternatives". In order to predict cooling load using the AHP combined model, the weights of each individual forecasting model must be obtained in advance by the method of AHP according to the actual situation. The following can be done manually or automatically by the AHP software:

(1) Constructing the pair-wise comparison matrices based on the experiments or expert;
(2) Calculating the priority vector for a criterion;
(3) Calculating λ_{max};
(4) Calculating the consistency index, *CI*;
(5) Selecting appropriate value of the random consistency ratio from Table 2;
(6) Calculating the consistency ratio, *CR*;
(7) Checking the consistency of the pair-wise comparison matrix using the value of *CR* to check whether the decision-maker's comparisons were consistent or not;
(8) Obtaining the weights of each model and using Eq. (1) to make the combined forecasting.

To further illustrate the AHP combined forecasting model, an example of hourly cooling load prediction for an office room is presented as follows.

To begin with, it is necessary to have a brief description of the room. The room is about 321.5 square meters; the exposed walls are all made of gravel concrete; the exposed windows are all double-glazing windows; the glazing rate is 45% on the southern wall and 35% on the northern wall; indoor heat source mainly come from the computers and the occupants, which is about 60–70 W per square meters. The room is air-conditioned by fan-coil units that are equipped with thermal meters from which the cooling load can be recorded. The thermal meter uses an ultrasonic flowmeter (type TFX; measure precision is $\pm 1.0\%$) and two Pt100 temperature sensors (measure precision is) to detect the chilled water flow rates and the inlet/outlet water temperatures, respectively, when the fan-coil unit is running. The actual cooling load of the room may be calculated by:

$$Q_0(\tau) = c_p \cdot G(\tau) \cdot \left[t_0(\tau) - t_i(\tau)\right] \tag{11}$$

where,

$Q_0(\tau)$ = The actual cooling load at the time τ, W
$G(\tau)$ = The mass flow rate of the chilled water passing through fan-coil at the time τ, $\text{kg}\cdot\text{s}^{-1}$
$t_0(\tau)$ = The temperature of outlet water of fan-coil at the time τ, °C
$t_i(\tau)$ = The temperature of inlet water of fan-coil at the time τ, °C
c_p = The mass specific heat of the chilled water, $\text{J}\cdot\text{kg}^{-1}\cdot{}^\circ\text{C}^{-1}$

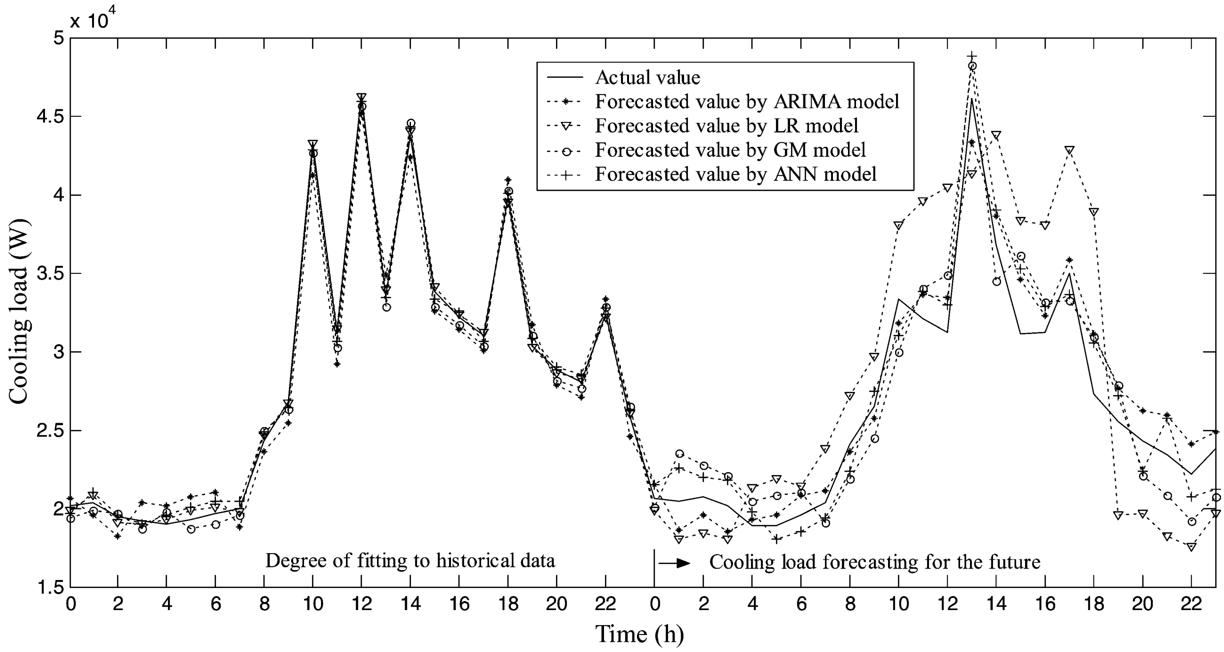

Fig. 2. Cooling load prediction by different forecasting models.

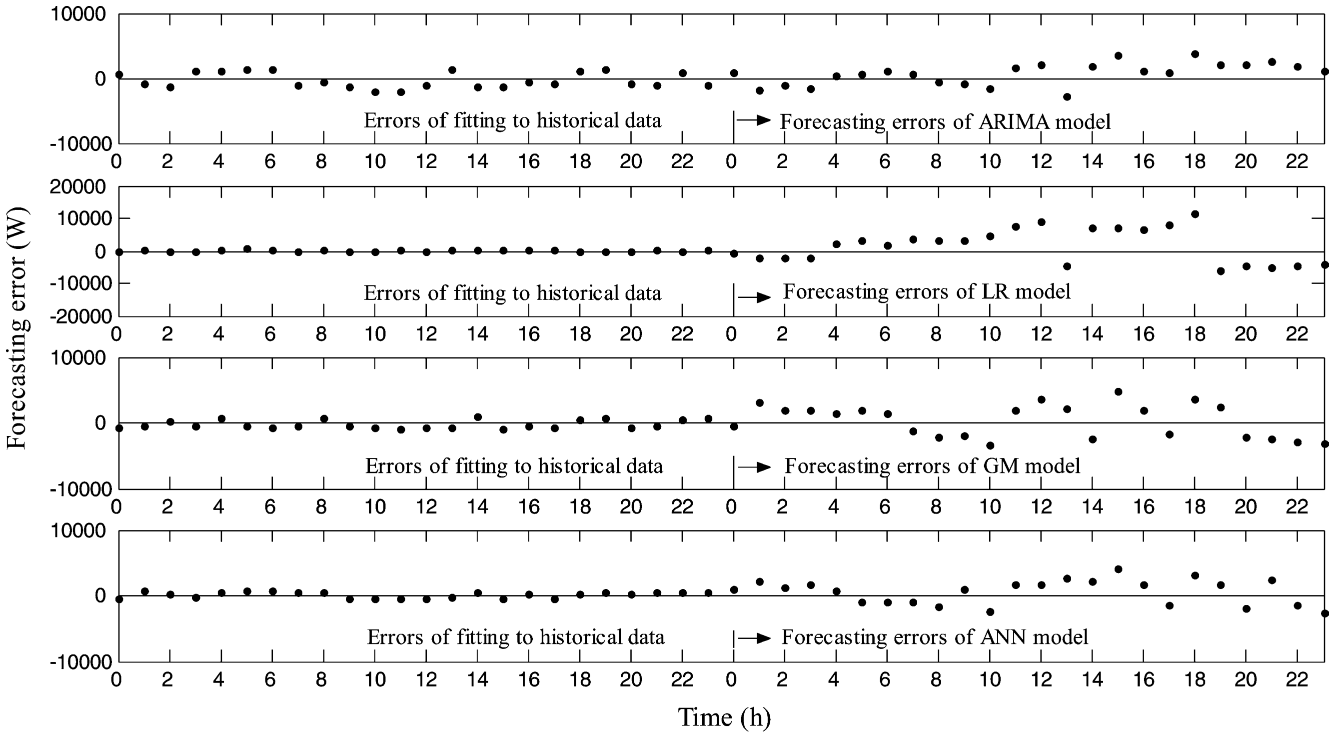

Fig. 3. Comparisons of errors between different forecasting models.

To provide enough information for each forecasting model, some other thermal parameters were real-timely monitored including the indoor temperatures, the outdoor temperatures and the values of solar intensity. The solar intensity was measured by solar radiometer (type CE183; manufactured by CIMEL Company of France; measurement precision ±1.0%).

The four individual forecasting models are used to practise the prediction in advance, respectively, to make their pair-wise comparisons of the relative priority of the criteria in the intermediate level. Fig. 2 is the results of one summer day's hourly cooling load prediction by different individual forecasting model. Fig. 3 shows the comparisons of forecasted errors among these models. In this paper,

Table 3
Pair-wise comparison of four forecasting models with respect to the criterion I (degree of fitting to the historical data)

Degree of fitting to the historical data	ARIMA model	LR model	GM model	ANN model
ARIMA model	1	1/7	1/3	1/3
LR model	7	1	3	3
GM model	3	1/3	1	1
ANN model	3	1/3	1	1

Table 4
Local priority of four forecasting models with respect to the criterion I (degree of fitting to the historical data)

Degree of fitting to the historical data	ARIMA model	LR model	GM model	ANN model	Local priority
ARIMA model	1	1/7	1/3	1/3	0.069
LR model	7	1	3	3	0.545
GM model	3	1/3	1	1	0.193
ANN model	3	1/3	1	1	0.193
$\lambda_{max} = 4.008$; $CI = 0.00267$; $RI = 0.90$; $CR = 0.00297 < 0.1$ OK					

the forecasted errors are defined as "the actual cooling loads minus the forecasted ones". Known from Fig. 2 and Fig. 3, LR model has the best fitting to the historical data, GM model and ANN model have a parallel better one, while ARIMA has the worst one. Thus, the pair-wise comparison matrices for the criterion I (Degree of Fitting to the Historical Data) can be obtained, as is shown in Table 3. The calculations for these items will be explained next for illustration purposes.

At first, the largest eigenvalue, λ_{max}, and the corresponding principal eigenvector of the judgment matrix (B_1) can be calculated, respectively, with the help of MATLAB 5.0 software [25] as follows:

$$\lambda_{max} = 4.008 \tag{12a}$$

$$\overline{W} = [0.1120, 0.8884, 0.3148, 0.3148]^{T} \tag{12b}$$

Finally, the priority vector w is obtained by normalizing $\overline{W}$:

$$w = \left[\frac{0.1120}{\sum_{i=1}^{4}\overline{W}_i}, \frac{0.8884}{\sum_{i=1}^{4}\overline{W}_i}, \frac{0.3148}{\sum_{i=1}^{4}\overline{W}_i}, \frac{0.3148}{\sum_{i=1}^{4}\overline{W}_i}\right]^{T} = [0.069, 0.545, 0.193, 0.193]^{T} \tag{13}$$

Now, estimating the consistency ratio is as follows:
The consistency index, CI, can be calculated:

$$CI = \frac{\lambda_{max} - n}{n-1} = \frac{4.008 - 4}{4-1} = 0.00267 \tag{14}$$

Selecting appropriate value of random consistency ratio, RI, for a matrix size of four using Table 2, we find $RI = 0.90$. The consistency ratio, CR, can be calculated as follows:

$$CR = \frac{CI}{RI} = \frac{0.00267}{0.90} = 0.00297 \tag{15}$$

As the value of CR is less than 0.1, the judgments are acceptable. The results are shown in Table 4.

From the characteristics of these models, the authors think that GM model and ANN model equally have the best adaptability in the cooling load prediction, and ARIMA model is slightly worse than them, while LR model has the least adaptability. Thus, the pair-wise comparison matrices and priority vectors for adaptability can be found in Table 5.

Known from the forecasting results in Figs. 2 and 3, ARIMA model is the best in terms of reliability, GM model and ANN model are equally reliable. However, they are both slightly worse than ARIMA model. LR model is the worst in reliability among the four forecasting models. Thus, the pair-wise comparison matrices and priority vectors for reliability can be found in Table 6.

Table 5
Pair-wise comparison of four forecasting models with respect to the criterion II (adaptability of the forecasting model)

Adaptability of the forecasting model	ARIMA model	LR model	GM model	ANN model	Local priority
ARIMA model	1	3	1/3	1/3	0.143
LR model	1/3	1	1/5	1/5	0.064
GM model	3	5	1	1/3	0.288
ANN model	3	5	3	1	0.505
$\lambda_{max} = 4.198$; $CI = 0.066$; $RI = 0.90$; $CR = 0.0733 < 0.1$ OK					

Table 6
Pair-wise comparison of four forecasting models with respect to the criterion II (reliability of the forecasting results)

Reliability of the forecasting result	ARIMA model	LR model	GM model	ANN model	Local priority
ARIMA model	1	5	3	3	0.522
LR model	1/5	1	1/3	1/3	0.078
GM model	1/3	3	1	1	0.200
ANN model	1/3	3	1	1	0.200
$\lambda_{max} = 4.044$; $CI = 0.0147$; $RI = 0.90$; $CR = 0.0163 < 0.1$ OK					

In addition to the pair-wise comparison for the decision alternatives (the four forecasting models), it can be also used to set priorities for all three criteria in terms of importance of each in contributing to the overall goal (cooling load prediction). Among the three criteria, the criterion of reliability should have the topmost priority to be considered, the adaptability has the lower one, while the degree of fitting to historical data is the most subordinate factor to be taken into account in the cooling load forecasting. Thus, the pair-wise comparison matrices and priority vectors for all the three criteria can be obtained in Table 7.

Now, the weight of each model in the combined forecasting can be found by combining the criterion priorities and the priorities of each model to each criterion, as is shown in Table 8. The calculations are given below for illustration purposes.

$$\text{Weight of ARIMA model} = 0.078 \times 0.069 + 0.435 \times 0.143 + 0.487 \times 0.522 = 0.322 \tag{16a}$$

Weight of LR model

$$= 0.078 \times 0.545 + 0.435 \times 0.064 + 0.487 \times 0.078 = 0.108 \quad (16b)$$

Weight of GM model

$$= 0.078 \times 0.193 + 0.435 \times 0.288 + 0.487 \times 0.200 = 0.238 \quad (16c)$$

Weight of ANN model

$$= 0.078 \times 0.193 + 0.435 \times 0.505 + 0.487 \times 0.200 = 0.332 \quad (16d)$$

From the weights above, it is indicated that ANN model may be considered as the best model in cooling load forecasting, and ARIMA model takes second place, then GM model and LR model next.

Using the weights of each forecasting model in Table 8 and the combined forecasting formula (Eq. (1)) as well as the forecasted results by each individual forecasting model, the cooling load prediction at the same time for the office room is made once again by the combined forecasting method, as is shown in Fig. 4. It is easy to see from Fig. 4 that the forecasted results obtained by the AHP combined forecasting model have a favorable agreement with the actual ones. To further demonstrate the validity of the model established in this paper, comparisons of forecasting errors are made between the AHP combined forecasting model and the other forecasting ones, respectively (Please see Fig. 5). As is shown in Fig. 5, the AHP combined model has much better forecasting results than the LR model in the future 24-hour forecasting. In addition, it is easy to see from Fig. 5 that in the beginning of the forecasting, the AHP combined model is more accurate than the other three models (ARIMA, GM and ANN).

Table 7
Pair-wise comparison of criteria with respect to the overall objective (cooling load prediction)

Cooling load prediction	Degree of fitting to the historical data	Adaptability of the forecasting model	Reliability of the forecasting result	Local priority
Degree of fitting to the historical data	1	1/5	1/7	0.078
Adaptability of the forecasting model	5	1	1	0.435
Reliability of the forecasting result	7	1	1	0.487
$\lambda_{max} = 3.013$; $CI = 0.0065$; $RI = 0.58$; $CR = 0.012 < 0.1$ OK				

Table 8
Weight of each model in the combined forecasting

Forecasting model	Criterion			Final weight of each forecasting model
	Degree of fitting to the historical data	Adaptability of the forecasting model	Reliability of the forecasting result	
	0.078	0.435	0.487	
ARIMA model	0.069	0.143	0.522	0.322
LR model	0.545	0.064	0.078	0.108
GM model	0.193	0.288	0.200	0.238
ANN model	0.193	0.505	0.200	0.332

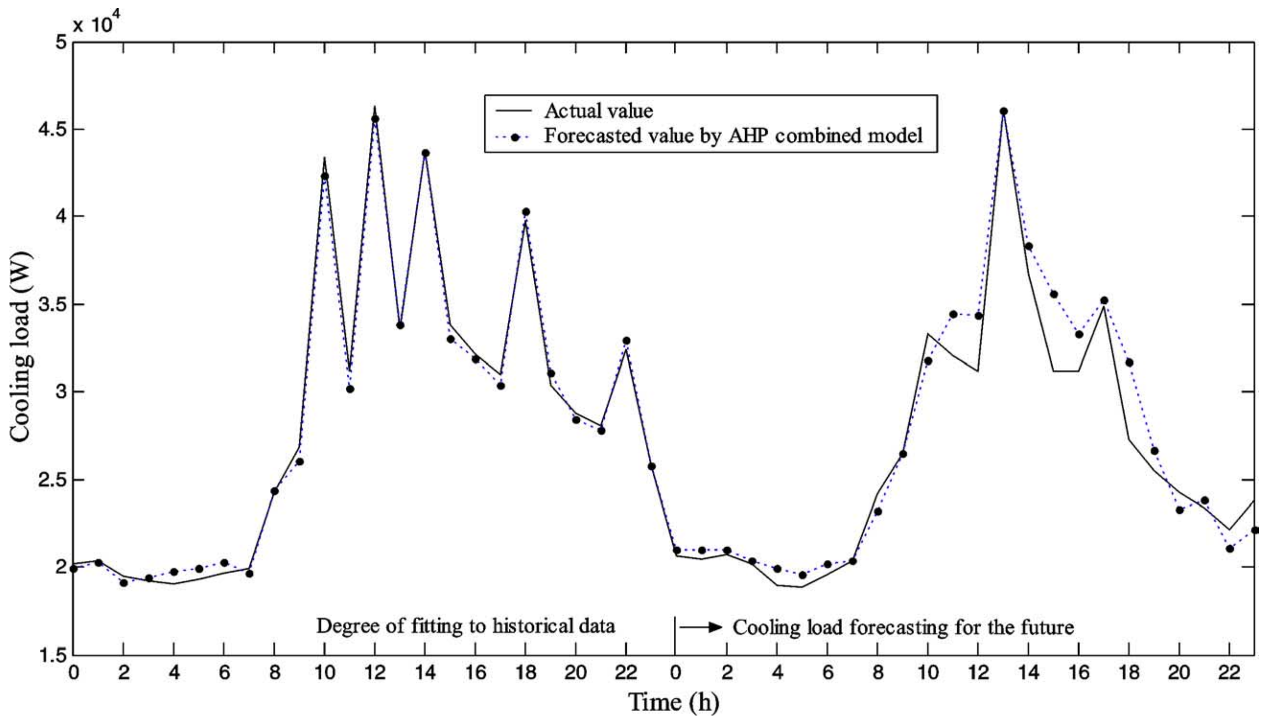

Fig. 4. Cooling load prediction by AHP combined forecasting model.

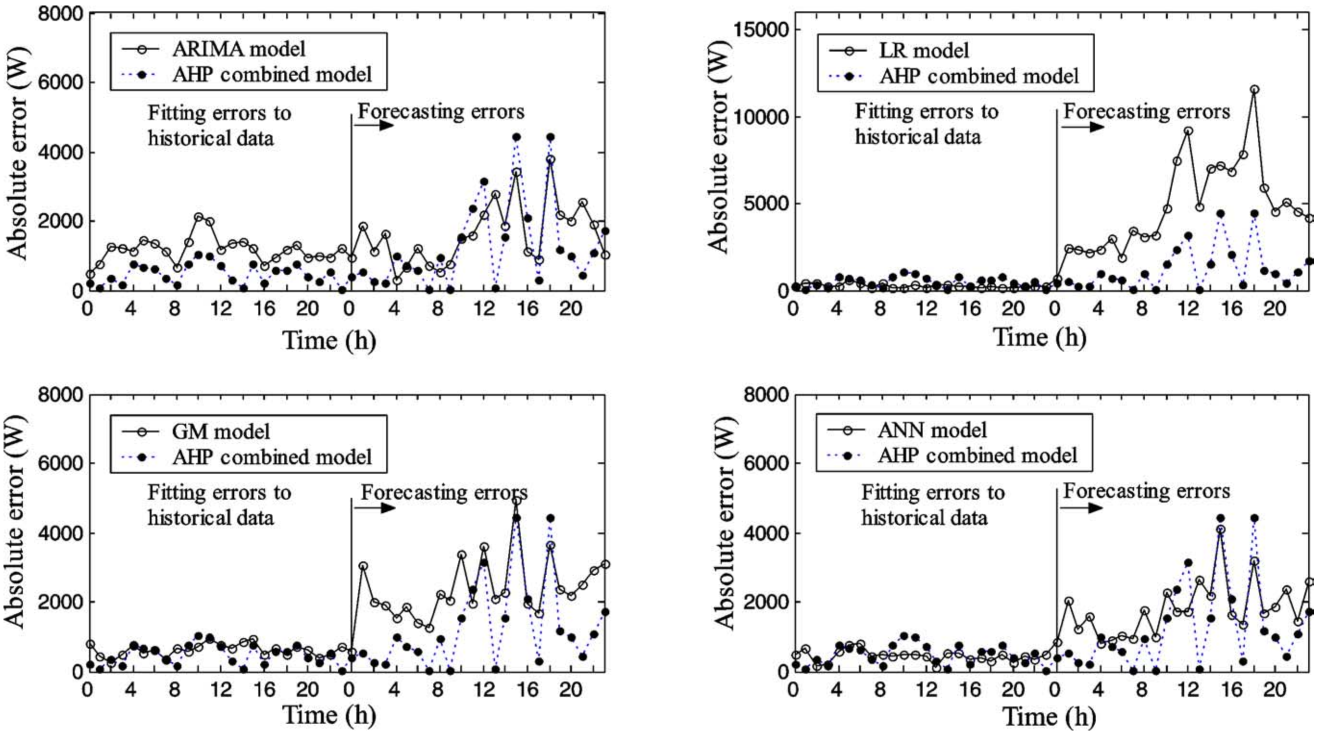

Fig. 5. Comparisons of forecasting errors between AHP combined model and the other ones.

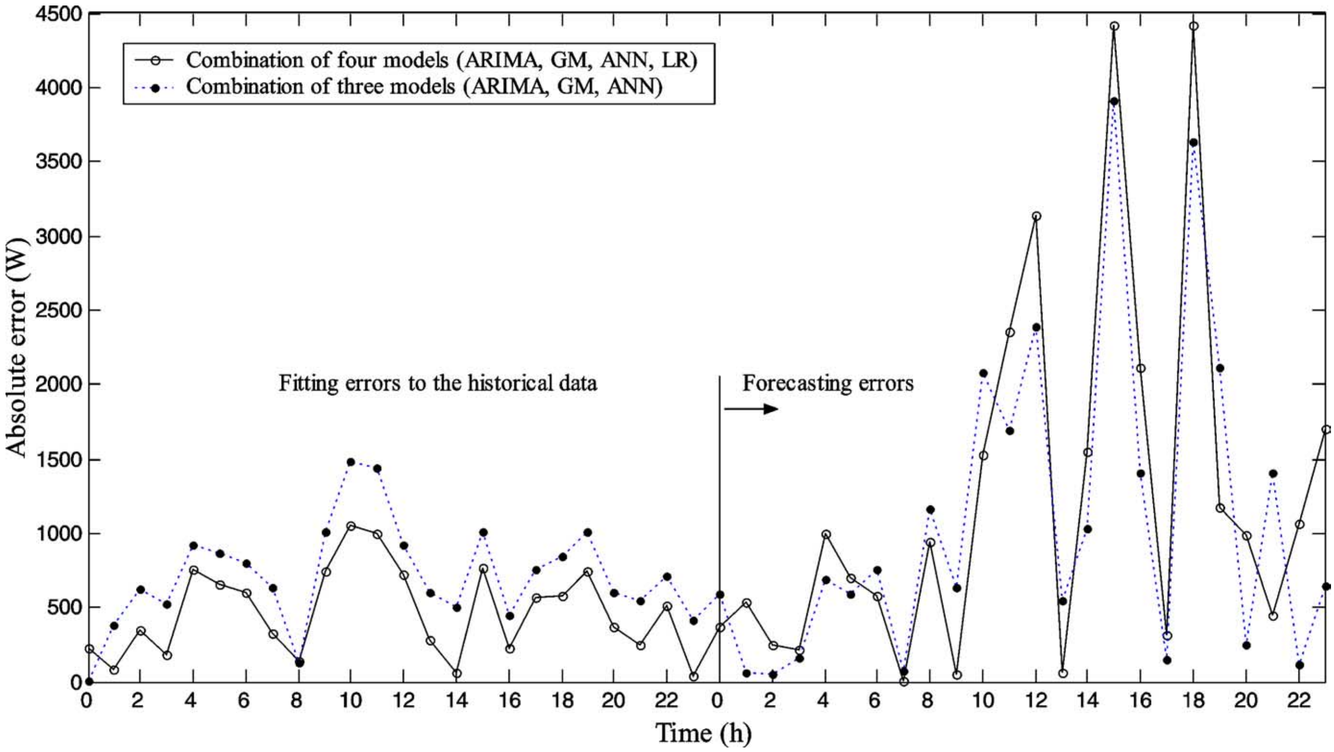

Fig. 6. Comparisons of forecasting errors between three-combined model and four-combined model.

However, in the later forecasting hours (about after 10 hours), big forecasting errors by AHP combined model occur uncertainly. Sometimes, they exceed those of ARIMA, GM and ANN. This is because the LR model will produce increasing large error with time going on. It indicates that the weights in Table 7 may be befitting for the combined model to forecast the future several-hour cooling load of the building.

To know the impact of LR model on the forecasting accuracy of the combined model, only three models (ARIMA, GM, ANN) are taken into account in the combination. Using above-mentioned AHP method, the weight of ARIMA, GM and ANN can be obtained. They are 0.564, 0.218 and 0.219, respectively. Fig. 6 shows the comparisons of forecasting errors between three-combined model and four-combined model. Seen from Fig. 6, although the forecasting errors of the three-combined model are smaller than those of the four-combined model in most cases, their gaps are very little. Therefore, it is suggested that LR model be kept in the combination for it may have useful information in the course of forecasting.

5. Conclusions

In this paper, the application of the combined forecasting method in cooling load forecasting is proposed and the preliminary results show that it has promised. It is necessary for load forecasting to take various relevant factors into consideration when evaluating each model. It is the precondition of comprehensively evaluating each model. AHP has the flexibility to combine quantitative and qualitative factors and to handle different groups of actors. By breaking a problem down in a logical fashion from the large, descending in gradual steps, to the smaller and smaller, several good forecasting models can be connected and combined into a better one through the simple pair-wise comparison judgments. The weights obtained in the paper may not be the only ones or the best ones in the combined forecasting. They should be amended from time to time based on different actual situations. In spite of this, The AHP combined forecasting model is still a valuable method to be advantageously employed in the hourly load prediction.

Appendix A. Brief review of LR model [1]

A multiple linear system model that has n inputs $(x_1, x_2, x_3, \dots, x_n)$ and one output (y) at time t can be described by the following equation:

$$y = k_1x_1 + k_2x_2 + \cdots + k_nx_n \tag{A.1}$$

where, the $k_1, k_2, \dots, k_n$ are constant unknown parameters.

At time $1 - m$, the system is shown by Eq. (A.2) using the vectors y and k and a matrix, X: where, $y = [y_1, y_2, \dots, y_m]^{\mathrm{T}}$,

$$k = [k_1, k_2, \dots, k_m]^{\mathrm{T}}$$

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{m1} & x_{n2} & \dots & x_{mn} \end{bmatrix}$$

Then, the error vector, e, between observed and predicted data is as follows:

$$e = y - Xk \tag{A.2}$$

The square of the e vector is denoted J:

$$\begin{aligned} J &= \sum_{i=1}^{m} e_i^2 = e^{\mathrm{T}}e = (y - Xk)^{\mathrm{T}}(y - Xk) \\ &= y^{\mathrm{T}}y - k^{\mathrm{T}}X^{\mathrm{T}}y - y^{\mathrm{T}}Xk + k^{\mathrm{T}}X^{\mathrm{T}}Xk \end{aligned} \tag{A.3}$$

To determine the estimate R that minimizes J, the derivative of J with respect to k is set to 0:

$$\frac{\partial J}{\partial k}\bigg|_{k=R} = -2X^{\mathrm{T}}y + 2X^{\mathrm{T}}Xk = 0 \tag{A.4}$$

Thus, k can be solved for

$$k = \left(X^{\mathrm{T}}X\right)^{-1}X^{\mathrm{T}}y \tag{A.5}$$

Then, prediction of y at time $m + 1$ is:

$$y_{m+1} = x_{1,m+1}k_1 + x_{2,m+1}k_2 + \cdots + x_{n,m+1}k_n \tag{A.6}$$

After obtaining new observed data at time $t + 1$, further parameter estimation can be done by the recursive least squares method. In this study, the model is defined to have 49 inputs and one output. The inputs are the load at $t - 24$ hours, 24 ambient temperatures and 24 solar insolation data from 23 hours before to the current time. One output is the hourly load. When the next 24 current observed data are obtained, the parameters are re-estimated by the recursive least-squares method. The hourly ambient temperatures and solar insolation for the next 24 hours are required to predict hourly loads for the next 24 hours.

Appendix B. Brief review of ARIMA model [2,6]

In the statistical approach, a statistical model is fitted to the observed data. By using an appropriate statistical model, the procedure discussed in this paper provides a model that takes into account the characteristics of both the load profiles and the noise.

When y_t denotes the observation at time t and e_t is a sequence of uncorrelated variables or residual error assumed white noise, the model may be written as:

$$\begin{aligned} &y_t + a_1y_{t-1} + a_2y_{t-2} + \cdots + a_py_{t-p} \\ &\quad = e_t + b_1e_{t-1} + b_2e_{t-2} + \cdots + b_qe_{t-q} \end{aligned} \tag{B.1}$$

This liner stochastic difference equation is called an autoregressive moving average model, denoted by ARMA(p, q). Using a time-delay operator, z^{-1} (defined by $z^{-1}y_t = y_{t-1}$), the following equation is obtained:

$$A\left(z^{-1}\right)y_t = B\left(z^{-1}\right)e_t \tag{B.2}$$

where,

$$A\left(z^{-1}\right) = 1 + a_1z^{-1} + a_2z^{-2} + \cdots + a_pz^{-p}$$
$$B\left(z^{-1}\right) = 1 + b_1z^{-1} + b_2z^{-2} + \cdots + b_qz^{-q}$$

The types of industrial time series that people wish to analyze frequently exhibit a particular kind of non-stationary behavior that can be represented by a stochastic model, which is a modified form of the ARMA model. The process is defined by the following two equations:

$$\begin{cases} A(z^{-1})y_t = B(z^{-1})e_t \\ y_t = \nabla^d x_t = (1-z^{-1})^d x_t \end{cases} \tag{B.3}$$

The model corresponds to assuming that the dth difference of the time series $\{x_t\}$ can be represented by a stationary ARMA model. An alternative way of looking at the model for $d = 1$ results from Eq. (B.3) to give:

$$y_t = \nabla x_t = x_t - x_{t-1} \tag{B.4}$$

where, ∇ is called the difference operator. In turn, ∇ has for its inverse the summation operator S, given by:

$$\begin{aligned} \nabla^{-1} y_t &= S y_t = y_t + y_{t-1} + y_{t-2} + \cdots \\ &= (1 + z^{-1} + z^{-2} + \cdots) y_t = (1 - z^{-1})^{-1} y_t \end{aligned} \tag{B.5}$$

The operator $S^2 y_t$ is similarly defined as:

$$\begin{aligned} x_t &= S^2 y_t = S y_t + S y_{t-1} + S y_{t-2} + \cdots \\ &= \sum_{i=-\infty}^{t} \sum_{h=-\infty}^{i} y_h \end{aligned} \tag{B.6}$$

Also,

$$x_t = S^3 y_t = \sum_{j=-\infty}^{t} \sum_{i=-\infty}^{j} \sum_{h=-\infty}^{i} y_h \tag{B.7}$$

and so on. Eq. (B.3) implies that the time series $\{x_t\}$ can be obtained by integrating the stationary time series $\{y_t\}d$ times. Therefore, the model mentioned above is called autoregressive integrated moving model, denoted by ARIMA(p, d, q).

When $\{x_t\}$ contains a period component with an elementary period of s, $\nabla_s = (1 - z^{-s})$ is applied to $\{x_t\}d_1$ times, and the ARMA(p, q) model is applied to time series $\{y_t\}$, the following model is obtained:

$$A(z^{-1}) \nabla_s^{d_1} x_t = B(z^{-1}) c_t \tag{B.8}$$

Next, the periodic variation pattern is obtained from the time series $\{c_t\}$. By taking $c_{t1}, c_{t1+s}, c_{t1+2s}, \ldots$ for any time t_1 within the elementary period, the ARMA(p_1, q_1) model is applied to this time series and the following model is obtained:

$$\begin{cases} P(z^{-s})c_t = Q(z^{-s})e_t \\ P(z^{-s}) = 1 + \alpha_1 z^{-s} + \cdots + \alpha_{p_1} z^{-p_1 s} \\ Q(z^{-s}) = 1 + \beta_1 z^{-s} + \cdots + \beta_{q_1} z^{-q_1 s} \end{cases} \tag{B.9}$$

The prediction model for a time series containing an elementary period of s is obtained from Eqs. (B.8) and (B.9) as follows:

$$P(z^{-s}) A(z^{-1}) \nabla_s^{d_1} x_t = Q(z^{-s}) B(z^{-1}) e_t \tag{B.10}$$

Furthermore, if $\{x_t\}$ has trend components and periodicity, Eq. (B.10) is rewritten as Eq. (B.11):

$$P(z^{-s}) A(z^{-1}) \nabla^d \nabla_s^{d_1} x_t = Q(z^{-s}) B(z^{-1}) e_t \tag{B.11}$$

where, $\{e_t\}$ is a white-noise sequence. This model is called ARIMA$(p, d, q) \times (p_1, d_1, q_1)$. The p, d, and q are order numbers of the processes for autoregressive, integrated, and moving average components, respectively. This means that the dth deviation of the time series data is expressed by the pth-order autoregressive term and the qth-order moving average term. The p_1, d_1, and q_1 refer to the same orders as p, d, and q at nth previous time. The value 24 for n is chosen since the time series data have a 24-hour cycle, which means that the loads at an hour are correlated with previous data a few hours before and one day before. In this study, the coefficients $a_1, a_2, \ldots, a_p, b_1, b_2, \ldots, b_q, \alpha_1, \alpha_2, \ldots, \alpha_{p_1}, \beta_1, \beta_2, \ldots, \beta_{q_1}$ are estimated using hourly loads of the previous day after the suitable order numbers (p, d, q, p_1, d_1, q_1) are chosen empirically. Then, hourly loads for the next day will be predicted. The order numbers are usually zero, one, and two.

Appendix C. Brief review of ANN model [7,8]

A neural network basically consists of interconnected neurons. Each neuron or node is an independent computational unit (Fig. 7), which works as per the following equation:

$$y = f\Big[\sum (x_1 w_1 + x_2 w_2 + x_3 w_3 + \cdots) + \beta\Big] \tag{C.1}$$

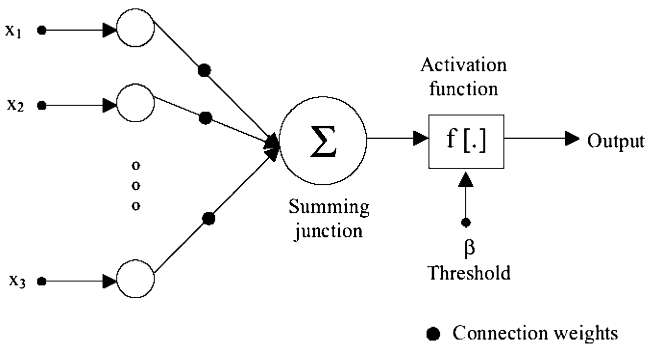

Fig. 7. Working of a neuron.

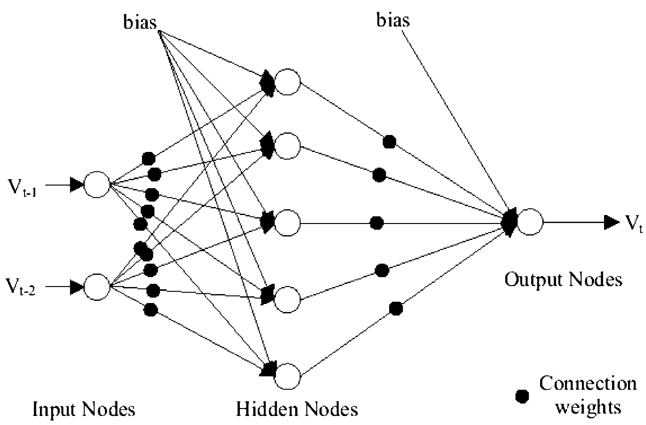

Fig. 8. Typical feed forward network.

where, y is the output from neuron; x_1, x_2, x_3, y are the input values; w_1, w_2, w_3 are the connection weights; β is the bias value; f is the transfer function, typically sigmoidal function given by

$$f[\bullet] = \frac{1}{1 + e^{-[\bullet]}} \tag{C.2}$$

A typical neural network used in the present study is shown in Fig. 8. This is called feed forward type of network where computations proceed along the forward direction only. There are three layers of neurons, namely input, hidden and output layer. The output obtained from the output neurons constitutes the network output.

The connection weights and bias values are initially chosen as random numbers and then fixed by the results of a training process. Many alternative training processes are available, out of which the present study adopted two popular schemes, namely back-propagation (BP) and cascade correlation (CC). The goal of any training algorithm is to minimize the global (mean sum squared) error E; defined below:

$$E = \frac{1}{2}\sum (O_n - t_n)^2 \tag{C.3}$$

where, O_n and t_n are network and target output for any nth output node. The summation has to be carried out over all output nodes for every training pattern. A pair of input and output values constitutes a training pattern.

In this study, the ANN model has 49 inputs and 1 output, which are the same as in the LR model, and 99 hidden layer neurons. The network is trained using hourly data of the past day. After the training, the hourly loads for the future 24 hours are predicted using the observed next-day weather data.

Appendix D. Brief review of GM model [10,11]

Grey forecasting model (GM) has three basic operations: (1) accumulated generation, (2) inverse accumulated generation, and (3) grey modeling. The grey forecasting model uses the operations of accumulated generation to build differential equations. Intrinsically speaking, it has the characteristics of requiring less data. The GM(1, 1) grey model, i.e., a single variable first-order grey model, is summarized as follows:

(a) *Step* 1: the initial sequence is

$$x^{(0)} = \left(x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(i), x^{(0)}(n)\right) \tag{D.1}$$

where, $x^{(0)}(i)$ is the time series data at time i.

(b) *Step* 2: based on the initial sequence $\boldsymbol{x}^{(0)}$, a new sequence $\boldsymbol{x}^{(1)}$ is generated by the accumulated generating operation (AGO), where,

$$x^{(1)} = \left(x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(i), x^{(1)}(n)\right) \tag{D.2}$$

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i) \tag{D.3}$$

(c) *Step* 3: the following first-order differential equation holds true:

$$\frac{\mathrm{d}x^{(0)}}{\mathrm{d}t} + ax^{(1)} = u \tag{D.4}$$

(d) *Step* 4: from step 3, we have

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{u}{a}\right)e^{-ak} + \frac{u}{a} \tag{D.5}$$

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \tag{D.6}$$

where

$$\hat{a} = \begin{bmatrix} a \\ u \end{bmatrix} = \left(B^{\mathrm{T}}B\right)^{-1}B^{\mathrm{T}}y_N \tag{D.7}$$

$$B = \begin{bmatrix} -0.5(x^{(1)}(1) + x^{(1)}(2)) & 1 \\ -0.5(x^{(1)}(2) + x^{(1)}(3)) & 1 \\ \vdots & \vdots \\ -0.5(x^{(1)}(n-1), x^{(1)}(n)) & 1 \end{bmatrix} \tag{D.8}$$

$$y_N = \left(x^{(0)}(2), x^{(0)}(3), \ldots, x^{(0)}(n)\right)^{\mathrm{T}} \tag{D.9}$$

$\hat{x}^{(1)}(k+1)$ is the predicted value of $x^{(1)}(k+1)$ at time $k+1$.

The GM(1, 1) grey model can be easily extended to a GM(1, N) grey model. Note that the second index in the GM(1, N) grey model stands for N variables ($x_1^{(0)}, x_2^{(0)}, \ldots, x_N^{(0)}$), and the differential equation can be written as follows:

$$\frac{\mathrm{d}x_1^{(0)}}{\mathrm{d}t} + ax_1^{(1)} = \sum_{i=2}^{N} b_{i-1}x_1^{(1)} \tag{D.10}$$

where, $a, b_1, b_2, \ldots, b_{N-1}$ are unknown parameters. According to step 4 of the GM(1, 1) grey model, these parameters can be estimated as follows:

$$\hat{a} = \left(\hat{a}, \hat{b}_1, \hat{b}_2, \ldots, \hat{b}_{N-1}\right) = \left(B^{\mathrm{T}}B\right)^{-1}B^{\mathrm{T}}y_N \tag{D.11}$$

where

$$B = \begin{bmatrix} -0.5(x_1^{(1)}(1) + x_1^{(1)}(2)) & x_2^{(1)}(2) & \ldots & x_N^{(1)}(2) \\ -0.5(x_1^{(1)}(2) + x_1^{(1)}(3)) & x_2^{(1)}(n) & \ldots & x_N^{(1)}(3) \\ \vdots & \vdots & \ddots & \vdots \\ -0.5(x_1^{(1)}(n-1) + x_1^{(1)}(n)) & x_2^{(1)}(n) & \ldots & x_N^{(1)}(n) \end{bmatrix} \tag{D.12}$$

$$y_N = \left(x^{(0)}(2), x^{(0)}(3), \ldots, x^{(0)}(n)\right)^{\mathrm{T}} \tag{D.13}$$

The forecasts of $x_1^{(1)}$ are as follow:

$$\hat{x}_1^{(1)}(k+1) = \left(x_1^{(0)}(1) - \sum_{i=2}^{N}\frac{b_{i-1}}{a}x_i^{(1)}(k+1)\right)e^{-ak} + \sum_{i=2}^{N}\frac{b_{i-1}}{a}x_i^{(1)}(k+1) \tag{D.14}$$

References

[1] M. Kawashima, C.E. Dorgan, J.W. Mitchell, Hourly thermal load prediction for the next 24 hours by ARIMA, EWMA, LR, and an artificial neural network, ASHRAE Trans. Part 1 101 (1995) 186–200.

[2] J.W. MacArthur, A. Mathur, J. Zhao, On-line recursive estimation for load profile prediction, ASHRAE Trans. Part 1 95 (1989) 621–628.

[3] W.F. Stoecker, R.R. Crawford, S. Ikeda, W.H. Dolan, D.J. Leverenz, Reducing the peaks of internal air-conditioning loads by use of temperature swings, ASHRAE Trans. Part 2 87 (1981) 599–600.

[4] J.E. Braun, Reducing energy costs and peak electrical demands through optimal control of building thermal storage, ASHRAE Trans. Part 1 96 (1990) 587–601.

[5] J.R. Forrester, W.J. Wepfer, Formulation of a load prediction algorithm for a Large commercial building, ASHRAE Trans. Part 2B 90 (1984) 536–551.

[6] D.H. Spethmann, Optimal control for cool storage, ASHRAE Trans. Part 1 95 (1989) 710–721.

[7] M. Kawashima, C.E. Dorgan, et al., Optimizing system control with load prediction by neural networks for an ice-storage system, ASHRAE Trans. Part 1 102 (1996) 1169–1178.

[8] J.F. Kreider, X.A. Wang, Improved artificial networks for commercial building energy use prediction, in: Proceedings of the ASME Annual Solar Engineering Meeting, Maui, HI, 1992, pp. 361–366.

[9] M. Anstett, J.F. Kreider, Application of neural networking model to predict energy use, ASHRAE Trans. Part 1 99 (1993) 505–517.

[10] J.L. Deng, Introduction to grey system theory, J. Grey System 1 (1) (1989) 1–24.

[11] J.L. Deng, Properties of multivariable grey model $GM(1, N)$, J. Grey System 1 (1) (1989) 5–141.

[12] C.M. Homg, C.T. Chiang, S.C. Lin, Design of grey prediction controller based on fuzzy reasoning, in: The Second National Conference on Fuzzy Theory and Application, Taiwan, 1994, pp. 66–71.

[13] R.C. Hwang, H.C. Huang, Y.J. Chen, J.G. Hsieh, H.C. Chen, C.W. Chuang, Selection of influencing factors for power load forecasting by grey relation analysis, in: Second National Conference on Grey Theory and Applications, Taiwan, 1997, pp. 109–113.

[14] H. Kang, Unstable weights in the combination of forecasts, Management Sci. 32 (1986) 683–695.

[15] S. Gupia, P. Wilton, Combination of forecasts: An extension, Management Sci. 33 (1987) 356–372.

[16] D.C. Schmittlein, et al., Combining forecasts: Operational adjustments to theoretically optimal rules, Management Sci. 36 (1990) 1044–1056.

[17] J.M. Bates, C. Gr, The combination of forecast, Operation Res. Quarterly 20 (1969) 451–463.

[18] C.W.J. Granger, R. Ramanathan, Improved methods of combining forecasts, J. Forecasting 3 (1984) 197–204.

[19] J.P. Shim, Bibliographical research on the analytic hierarchy process (AHP), Socio-Economic Planning Sci. 23 (1989) 161–167.

[20] F. Elkarni, I. Mustafa, Increasing the utilization of solar energy technologies (SET) in Jordan: Analytic hierarchy process, Energy Policy 21 (1993) 978–984.

[21] T.L. Saaty, The Analytic Hierarchy Process: Planning, Priority Setting and Resource Allocation, McGraw-Hill, New York, 1980.

[22] T.L. Saaty, Axiomatic foundations of the analytic hierarchy process, Management Sci. 32 (1986) 841–855.

[23] T.L. Saaty, Rank generation, preservation and reversal in the analytic hierarchy process, Decision Sci. 18 (1987) 157–177.

[24] T.L. Saaty, Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process, RWS Publications, Pittsburg, 2000.

[25] D. Redfern, The MATLAB 5 Handbook, Springer, New York, 1997.