Journal of Building Engineering 44 (2021) 103379

Contents lists available at ScienceDirect

Journal of Building Engineering

journal homepage: www.elsevier.com/locate/jobe

Operating optimization of air-conditioning water system in a subway station using data mining and dynamic system models

Xing Su [a,*], Yixiang Huang [a], Lei Wang [a], Shaochen Tian [a], Yanping Luo [b]

[a] School of Mechanical Engineering, Tongji University, Shanghai, 200092, China
[b] Guangzhou Metro Design and Research Institute Company Limited, Guangzhou, 440104, China

ARTICLE INFO

Keywords:
Air conditioning water system
Subway station
Parameters optimizing
Feedforward control

ABSTRACT

Energy-conservation potential in the air-condition water system for subway stations is huge due to its conservative design method. Also, for operation strategy of such systems, the operation modes are formulated with the fixed schedule. This paper presents a data-based optimization method to obtain optimal parameters of the system for feedforward control. The data mining models are established by using the data from energy consumption platform of the refrigerating system. The study utilized the box-plot method, kNN algorithm and k-means algorithm to process and repair original data. Then Artificial Neural Network (ANN) model is adopted to developed the forecasting model to assess load, performance and energy consumption of the system. The input features of the models are determined by the existed models and clustering analysis. The optimal parameters under the conditions of different load-ratio range and ambient thermal environments are calculated via Genetic Algorithm and trained equipment models. And the optimal parameters are applied to establish operation schedule based on feedforward control and response time. The optimal feedforward control method is verified by a validated TRNSYS model. When the parameters are optimized, the water system energy consumption can be save by 9.5% in a cooling season.

1. Introduction

With the urbanization and urban expansion of many large-scale cities in China, there exists great demand for urban transport. Subway, an urban mass transit system owing huge passenger capacity and easing traffic pressure on the ground, is taken into consideration in an essential part of infrastructures construction in series of major cities [1]. By the end of 2019, the total length of unbar subway transit lines constructed in China reached 6730.24 km. The energy consumption of subway station is increasing with the development and construction of unbar subway transit lines. According to the relative report, total power consumption of unbar subway transit lines reached 12.2 billion kWh, and taken up approximately 1.94‰ of national power consumption. Power consumption index of subway stations, of more than approximately 200 kWh/m^2, is 2–4 times higher than it of regular buildings [2]. And the power consumption of ventilation and air conditioning (VAC) systems of subway stations account for 30–50% of the total energy consumption [3,4]. Therefore, improving energy efficiency of VAC system is crucial to reduce the energy consumption of the whole subway system.

As for design of relative equipment of VAC system, the capacity of the equipment is determined by the load under the full-load condition. And the capacity of subway systems is usually designed according to peak of passenger flow and extreme weather conditions. Also, the capacity of VAC equipment is designed and determined by the maximum long-term load [5]. Therefore, the designed capacity of systems is higher than the demand of actual operation process. In the operation process, the time of VAC equipment working on full-load condition is less than 5% of whole operation time [6]. So, it necessary to carry out the optimization research on operating parameters and operation strategies of VAC systems in the subway stations.

With the development of data mining method, the relative data mining algorithms have been used to find potential operating information of VAC in the field of building physics. The VAC systems operating data collected from various approaches was processed and applied in these fields of building energy consumption analysis, system fault diagnosis, etc. The data mining model with well forecasting performance has to be trained by high-quality data. Moreover, the data sources of VAC system are the building auto-control system and building power management system. The data obtained from the approaches are complicated. Therefore, appropriate data cleansing methods and data preprocessing methods should be carried out to provide high-quality

* Corresponding author. School of Mechanical Engineering, Tongji University, Shanghai, China.
E-mail address: suxing@tongji.edu.cn (X. Su).

https://doi.org/10.1016/j.jobe.2021.103379
Received 20 August 2021; Received in revised form 17 September 2021; Accepted 23 September 2021
Available online 25 September 2021

Nomenclature

f	frequency of Variable-frequency Drive, Hz
H	pumping head of pumps, m
H_{rated}	pumping hear under the rated condition, m
m_{chw}	chilled water flow, kg/s
m_{cw}	cooling water flow, kg/s
m_a	air flow driven by fans of cooling towers, kg/s
n	pump speed, rpm
n_{rated}	pump speed under the rated condition, rpm
P_{in}	number of persons entering the station, persons, persons
P_{out}	number of persons leaving the station, persons, persons
P_{chwp}	energy consumption of chilled water pumps, kW
P_{cwp}	cooling water pumps, kW
P_{tower}	energy consumption of cooling towers, kW
P_{pump}	power of pumps, kW
P_{tower}	energy consumption of cooling towers, kW
$Q_{chiller}$	refrigerating output of chillers, kW
Q_{tower}	condensation heat removed by cooling towers, kW
$Q_{rated,i}$	rated cooling capacity of the *i*th chiller, kW
Q_{cl}	cooling demand of air conditioning terminals, kW
T_1	residence time in the hall of persons to enter the station, min
T_2	residence time in the hall of persons to leave the station
T_3	residence time in the hall of persons to transfer among the halls, min
T_4	residence time in the hall of persons to enter the station, min
T_5	alighting time of persons staying in the plat, min
$T_{chw,out}$	outlet temperature of chilled water, °C
$T_{cw,in}$	inlet temperature of cooling water, °C
T_{wb}	ambient wet bulb temperature
V	volume flow of pumps, m^3/h
V_{rated}	volume flow of pumps under the rated condition, m^3/h
X_{tr}	transfer passenger flow of line X, person
Y_{tr}	transfer passenger flow of line Y, person

Greek symbols

η	pump efficiency, %
η_p	transmission efficiency, %
η_m	motor efficiency, %
η_{VFD}	variable-frequency drive efficiency, %
μ	penalty factor of the objective function

Subscripts

chw	Chilled water
chw	Chilled water pump
cw	Cooling water
cwp	Cooling water pump
in	inlet
out	outlet

Abbreviations

ACWS	Air conditioning water system
ANN	Artificial Neural Network
GRA	Grey Relation Analysis
PLR	Part Load Ratio
SSE	Sum of the Squared Errors
VAC	Ventilation and air conditioning
VDF	Variable-Frequency Drive

data for the data mining model. Fan et al. [7–9] proposed series of data cleansing methods for the massive data of the building systems based on the actual cases. Wavelet analysis method [10], identify various variations in measuring data, was used to distinguish the abnormal data of the chilling systems. As for the variable air volume system, the wavelet analysis method was used to process the original data obtained from the building automation system [11]. Based on the decomposition results of wavelet analysis, the kinds of information representing different operating conditions. The outliers of the energy consumption data from the building automation system were detected by kind of methods (i.e., generalized extreme studentized deviate and canonical variables analysis [12]) to improve this quality of data [13,14]. Besides, *k-means* method, an algorithm clustering data into series of sets by the initial clustering centroids, was used to identify outliers by the distance of data points to clustering centroids [15]. *KNN* method was also applied in the field of filling missing data by finding the *k* nearest distance of selected data points [16]. As for the data cleansing method, the method should be selected according to the actual application, the volume and mathematical characteristics of the given data set.

The air conditioning load is the basic information of the design and optimization for the VAC system. Due to the nonlinear and time-varying property of the load, its influencing factors are coupling and of complex structure. In the existed studies, the white-box model, grey-box model and black-box model are widely applied in the load forecasting and equipment performance forecasting. Linear regression method, a simplified data-fitting method, was used to calculate the design load of under-floor air-conditioning system [17]. Because of the simple form of linear regression, this method cannot fit the data points well in actual application. The linear method is not able to describe the detailed information of nonlinear model. The black-box model (i.e., artificial neuron net [18] and support vector machine [19]), owning physical model-free characteristic and nonlinear function fitting performance, was widely adopted to forecasting dynamic air condition load in different kinds of buildings. Yang et al. [20] proposed an adaptive neuron net model to forecasting the load online by updating input real-time data. The model owes better dynamic responses performance compared with static forecasting model. R. Mena et al. [21] established a short-term load forecasting model of a bioclimatic build by the artificial neural network (ANN). The field measured data was regarded as the input data set to train and validate the model. By the advantage of physical rule-based method, a rule-based ANN model was developed to forecast the cooling load of a building in the improved accuracy.

Optimization of operating parameter at the VAC system, in essence, is finding the parameters with the objectives of minimizing energy consumption and maintaining the comfortable environment. Therefore, the discussion about the optimal objectives of VAC system focuses on the optimization of both local parameters for a kind of equipment and it of global parameters for the whole system [22]. As for the optimization of local parameters, the load ratio distribution and ON/OFF control schedule of chillers group are the research emphasis. Bo et al. [23] estimated the probability density function of cooling load ratio to optimize the load distribution of chillers. J. Li and Z. Li [24] discussed the free cooing switchover temperature for data center cooling system with water-side economizer to minimize the energy consumption of the cooling system. Congradac et al. [25] optimized the outlet temperature of chilled water with the optimal objective of energy consumption minimization using the degree-days model. As for the optimization of global parameters, Lu et al. [26,27] developed a system model of VAC system, and arranged the optimization process with the minimization of system energy consumption based on the established model. The optimal operation parameters were validated by the field experiment. The energy conservation rate of 6–20% can be achieve. Andrew et al. [28,29] developed the system model by dynamic ANN model, and the performance of air handle unit was established by support vector

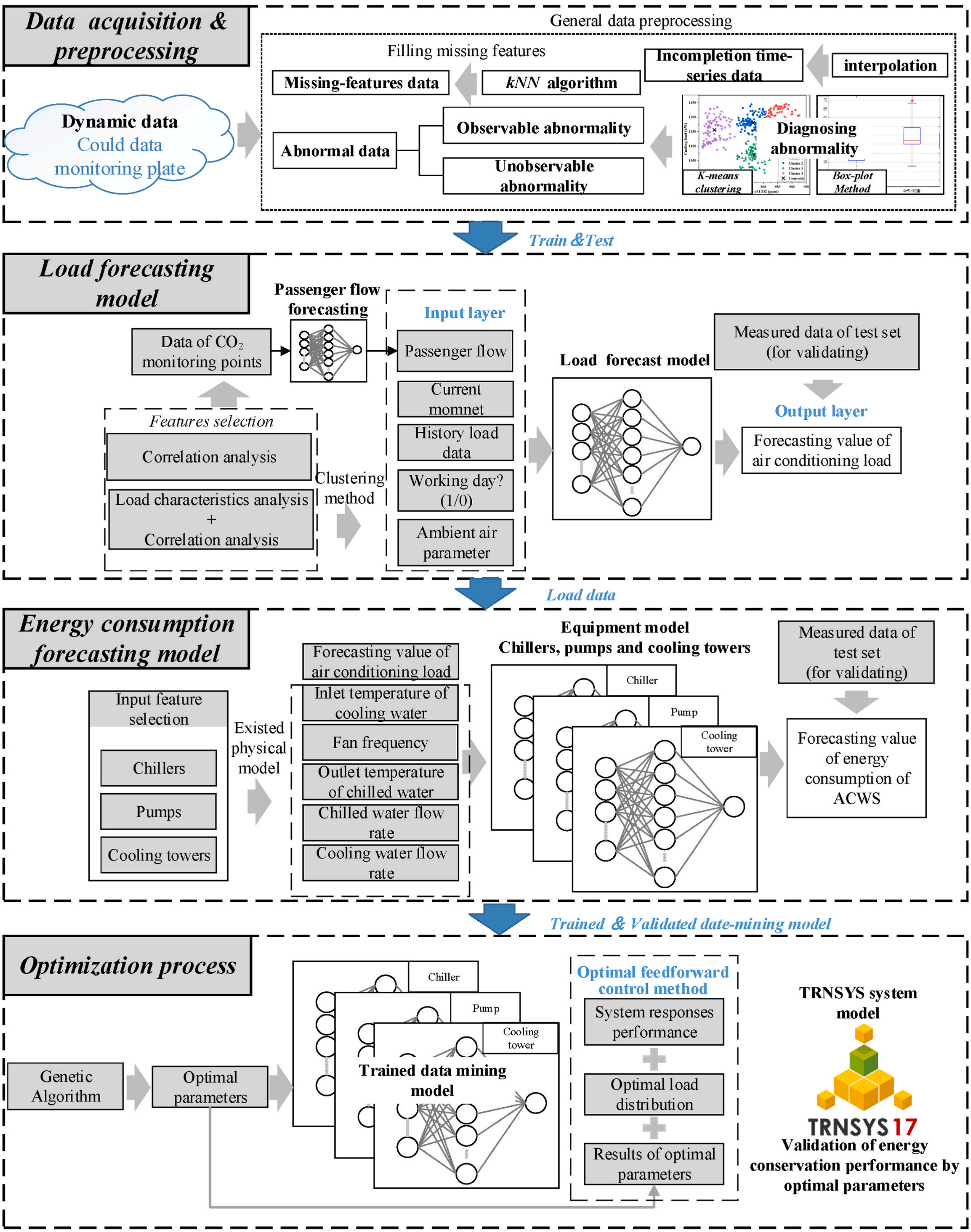

Fig. 1. Research framework.

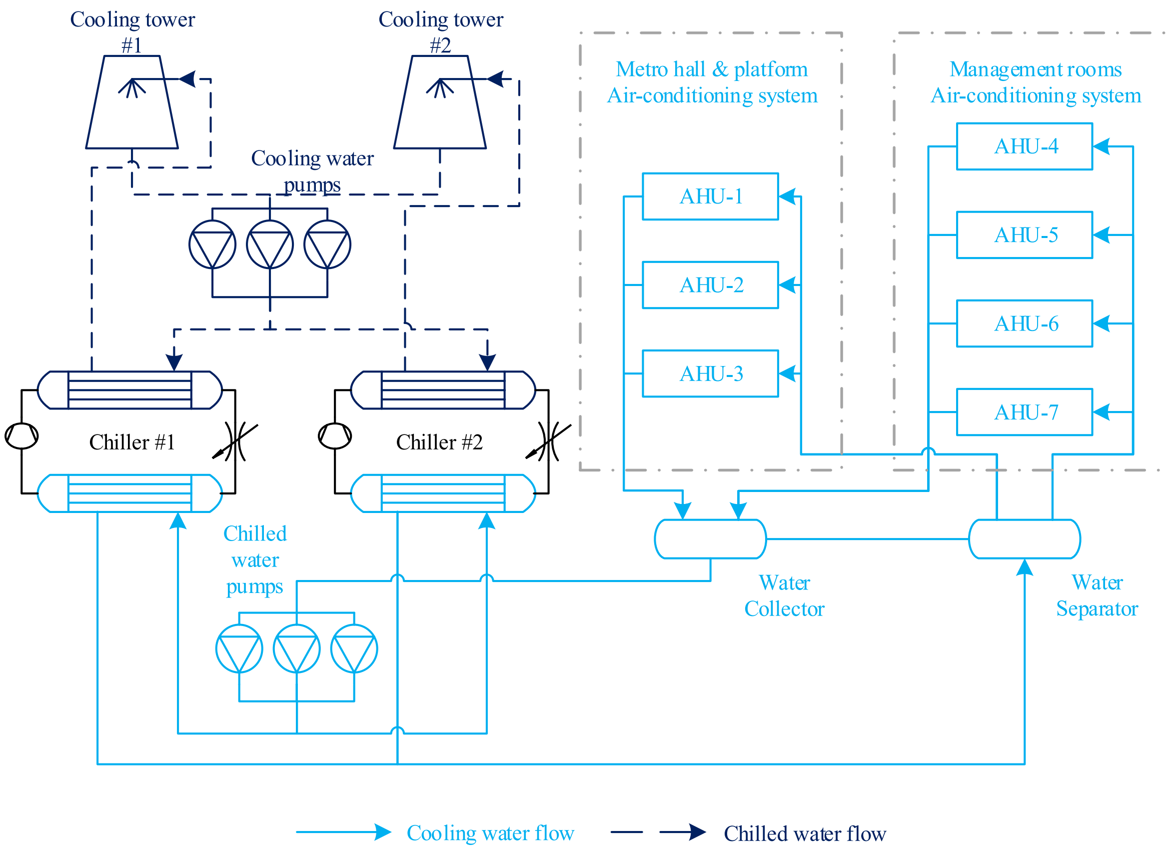

Fig. 2. Schematic of the basic air conditioning water system in the subway station.

machine method. Then, the operating parameters were obtained using particle swarm optimization algorithm. Compared with the existed control strategy, the energy conservation rate with optimal parameters reaches 23%. The aforementioned studies focused on updating the parameters of existed strategy and schedule. However, for most of the VAC systems, feedback control method was adopted to adjust the system state. This control performance depends on responses characteristic of the equipment in the system. Therefore, model predictive control was proposed to optimize the whole system combined with the relative forecasting models and responses performance of the system. The process of model predictive control is establishing the system model by data mining method, then proposing the control schedule based on the system model and optimal parameters to minimize the energy consumption of relative systems [30–32].

In this study, the authors conducted optimization works on operating parameters of air-conditioning water system (ACWS) in a subway station using data mining method. First, real-time monitoring data, collected from a energy consumption monitoring platform, was cleaned by boxplot method, *KNN* method and *k-means* method to provide high-quality input data set of data mining model. Second, based on the preprocessed data set, typical conditions were obtained by load characteristics analysis using clustering algorithm. Also, passenger flow, a significant influencing factor of load, was estimated by a real-time data-trained Artificial Neural Network (ANN) model. Third, data mining model was selected to establish air conditioning load forecast model and energy consumption model of ACWS. Finally, optimal operation parameters were obtained based on the trained models. And the energy conservation performance of these optimal parameters was tested by a validated TRNSYS model.

2. Methodology

Fig. 1 shows the process and framework of our work including the main information of methodology, various models developing, established model validating, model-based optimization and verifying optimizing performance, etc.

In the section of methodology, Section 2.1 describes the kind of data need to be obtained from the energy consumption monitoring platform. Also the basic information about the subway station and its ACWS is introduced in the part. Section 2.2 describes the data preprocessing methods used in elevating date quality for data mining. Sections 2.3 and 2.4 describe the process of establishing data mining model to forecast air condition load and the equipment performance. The input features selection methods for various model are mainly introduced. In section 2.5, the optimized parameter and the mathematical expression of the system optimization problem are focused on. Section 2.6 indicates a simulating model using TRNSYS. Additionally, Fig. 1 provides the general information in this work.

2.1. Data acquisition

In this study, a subway station, located in Guangzhou, China, was regarded as an optimizing object. And the subway station is a T-type interchange station with 4 exits/entrances, whose design peak passage flow in the morning rush hour and evening rush hour is approximately 40,200 persons/hour and 47,600 persons/hour, respectively. The plan sketch is provided in Fig. S1. In this station, an energy consumption data acquisition system with real-time monitoring has been installed to monitor the real-time dynamic performance indicators of ACWS (water flow, power of different equipment, frequency of fans or pumps, inlet/outlet temperature of chillers and cooling towers, etc.). Also, the platform has been taken into service for at least 3 years. Large of history data was collected and kept in the Cloud, providing a good foundation for subsequent modelling by data mining method. The user interface of the data acquisition system is shown in Fig. S2 in the Supplementary Materials.

Table 1
Extracting historical data.

Purposes		Extracting data	Data sources
Load forecasting		Time label, ambient dry bulb temperature, ambient wet bulb temperature, concentration of CO_2 measured from series of points and refrigerating output of ACWS	platform
		Number of train pairs	Station operation schedule
		Passenger flow	Field measurement
Equipment modelling of ACWS	Energy consumption forecasting of chillers	Dynamic power of chillers, dynamic input and output temperature of chilled water, dynamic refrigerating output of chillers	platform
	Energy consumption forecasting of water pumps	Dynamic power of pumps and dynamic water flow	platform
	Energy consumption forecasting of cooling towers	Fan energy consumption forecasting of cooling towers and frequency of fans of cooling towers	platform
	Cooling capacity forecasting of cooling tower	Dynamic cooling capacity of cooling towers, outlet water temperature of cooling tower, frequency of fans of cooling towers, outdoor wet bulb temperature	platform
Response performance of control system		Time label, inlet and outlet water temperature of chillers and cooling towers, set outlet temperature of chillers, frequency of fans and pumps	platform
Others		On/off signals of chillers, pumps and fans and Rotation speed of fans	platform

The cooling type of air conditioning systems used in the subway stations can be classified into two kinds: district cooling and substation cooling. In this subway station, substation cooling was adopted to cool the air condition system. The refrigeration system of ACWS used in this subway station is made of two screw chillers, two cross-flow cooling towers, three cooling water pumps (dual purpose) and three chilled water pumps (dual purpose). The detailed information of the equipment installed in ACWS is shown in Table S1 in the Supplementary Materials part. Fig. 2 shows the system configuration of the ACWS.

Due to the system operation data lacking of screening and classifying during the collecting process, there exists lots of redundant information stored in the cloud database. It is necessary to select appreciate kinds of data from the database according to different modelling and optimization targets (i.e., cooling load forecasting, equipment of ACWS modelling, studying on the response performance of control system and optimizing operation parameters of ACWS). It is worth noting that cooling load of the air conditioning system cannot be measured and obtain directly. But the refrigerating output can be regarded as cooling load when the system reached a balance condition (the air humidity and air temperature of buildings indoor environment maintain stable). The air temperature and humidity variation of the station platform and the management room are shown in Fig. S3 of the Supplementary Materials part.

The kind of historical data extracted from the cloud platform or field measurement is shown in Table 1. It should be cleared that the types of data can be classified into two types: discrete type and continuous type. Therefore, the data should be processed in accordance with the types of different input variables.

2.2. Data preprocessing

2.2.1. General data preprocessing methods

As indicated in Section 2.1, a complete sample of data with the time label usually consist of both discrete and continuous input features. For discrete input features, step change is usually the dynamic change form of it, which usually are set temperature and ON/OFF star-stop signals. Therefore, the problems of discrete data are usually incompletion among the whole time series (i.e., missing values and zero values). For continuous input feature, the value of it changes continuously in the time series. Because there exist blackout conditions and abnormal data transportation of the measuring process of terminal sensors, frequency of occurring data missing and abnormality is high. These problems are obvious, and these can be defined as "observable abnormality". Also, there exist "unobservable abnormality" of the continuous data. In the case of relationship between the frequency and energy consumption of pumps, the frequency of pumps changes in a large range with a constant energy consumption level of them. However, the relationship of the two indicators is exactly a form of single-value function. These data contribute adverse effect on the processes of training and establishing model. The problem of such data can be found only in the conditions of relating the features of the data. Compared with the data with observable abnormality, the data can be defined as problem data with "unobservable abnormality".

In accordance with the classification of problem data, there exists corresponding solutions to repair these data series. For missing-features data, k-Nearest Neighbor (kNN) algorithm was adopted to fill the relative features of missing-features data. kNN algorithm, a kind of supervised learning algorithm, can find the k closest sample points to the set sample point of the test set. For a piece of complete data sample, its features are complete so that they can be applied in training process. However, the features of missing-features data sample are incomplete: Parts of these features are unknown and some of these are known. kNN algorithm can be used to obtain the k closed sample points with complete features based on the known features of missing-features data. Then the missing features of these data can be filled according to the average value of corresponding features of the k select sample points.

For the set of data with observable abnormality, these data can be detected by setting normal range of relative features. Also, the detected observable abnormal data can be repaired by data interpolation among the continuous time series. For the data with unobservable abnormality, k-means clustering algorithm and box-plot method were used to processing these data. The box-plot method is used to determine abnormal data by the distribution of data. The unobservable abnormal data was distinguished by bounds of box plots calculated by interquartile range. And the box-plot method can be applied to process huge amounts of data and owe strong robustness. As for the k-means clustering algorithm, it was adopted to divide the data set into k clusters of sample points. Therefore, the cluster of sample points, whose number of points is small, is regarded as unobservable abnormal data. When the number of data point of a sample cluster is huge, these points, far away from the cluster center, are treated as unobservable abnormal data as well.

2.2.2. Specific preprocessing methods for different data-mining targets

Based on the aforementioned data processing methods and data mining targets, we applied these methods in preprocessing original data. The detail information is provided as following:

(1) Load forecasting model

These kinds of data, indicated in Table 1, were used in the establishing process of loading forecasting model. And the historical data was also provided to train the load forecasting model and describe time-series characteristic of dynamic loading. Therefore, a complete time-series data set with high quality should be obtained by the preprocessing methods. As for missing-features data, it is necessary to ensure the train set owing continuous time series. The missing features must be filled based on the aforementioned *k*NN method and interpolation method. For the observable abnormal data, appropriate range of relative features was set in accordance with the actual conditions and physical principles. The ambient dry bulb temperature, for instance, was initially screened based on the standard outdoor weather parameters: the normal range of the ambient dry bulb temperature is 15–40 °C. If the value of this temperature is out of this range, the feature will be deleted and filled by the filling principle of missing features. For unobservable abnormal data, as all the value of features of these data is within the normal range, the data cannot be detected directly. However, the data did not conform to the actual operation conditions. The relationship among existed features were used to diagnose the unobservable abnormality. Considering the huge numbers of conditions for load data and, the existing of similarity of the operation conditions with the same load interval, the box-plot method was used to detect unobservable abnormality by load interval. The eliminated abnormal data was regarded as missing-feature kind and filled by aforementioned method. Based on primary screening by box-plot method, *k*-means algorithm was applied in the following data processing. The load feature was coupled with the other selected features (i.e., concentration of CO2), and then outliers of these k clusters were deleted to improve the data quality. Therefore, the general process of preprocessing dynamic data used in load forecasting is that different features were selected to be combined with the load data, and then abnormal data was diagnosed and smoothed based on aforementioned methods.

(2) Energy consumption forecasting model

In the process of establishing energy consumption forecasting model based on actual data, models for chillers, pumps and cooling towers should be found firstly. In addition, the total system model of ACWS was made up of the found models. The detailed processing methods of different kinds of data preprocessing are provided respectively in this subsection.

As for missing-feature data, the time label of these data is not necessary for the energy consumption forecasting process. Also, the complete time data series is not required. This represents that the filling operation for the data without complete time series is unnecessary for the train set of energy forecasting model. Because the number of data which can be used for establishing energy consumption models, is adequate, the data with missing features can be deleted directly. Similarly, because complete time series are unnecessary, the missing-features data with observable abnormality can be deleted directly based on large number of data obtained from the platform. As for data with unobservable abnormality, the energy consumption data and the equipment performance data coupled with different input features was preprocessed by the box-plot method and the *k*-means clustering algorithm. The outliers diagnosed by the box-plot method and *k*-means clustering algorithm was deleted.

2.3. Load forecasting model

2.3.1. Input features selection of load forecasting

In the process of establishing a data mining model, the input features should be identified at first. The air conditioning system load is made up of dynamic heat gain and static heat gain. The dynamic heat gain fluctuates according to different factors during the operation period. It depends on time variable, number of train pairs, passenger flow and outdoor environment parameters, etc. And the static heat gain can be regarded as a stable value under the normal conditions. Therefore, dynamic heat gain should be focused on in the process of establishing load forecasting model. The air conditioning heat gain varies with the changing of dynamic heat gain, including heat gain from passengers, heat gain from mechanical fresh air, heat gain from the infiltration air through platform doors and heat gain from the infiltration air through entrances. The heat gain from the mechanical fresh air can be determined by the passenger flow of halls and platforms zone. The heat gain from mechanical fresh air can be determined by the passenger flow and outdoor weather parameter (i.e., ambient dry- and wet-bulb temperature). It should be noted that ambient dry bulb temperature and ambient wet bulb temperature are used for the input feature of load characteristics and load forecasting model, respectively. The heat gains from infiltration air through the entrances and the platform doors are determined by outdoor thermal environment parameters. These heat gain parameters depend on the dynamic input features (i.e., passenger flow, number of train pairs, outdoor thermal environment parameters). Therefore, the air conditioning load can be determined based on these input features.

2.3.2. Passenger flow forecasting model

Based on the founded input features of load forecasting model, passenger flow, a curial feature, should be calculated in this forecasting processing. However, there exist difference in retention time of the passengers during different periods. The dynamic changes of passenger flow were not taken into consideration. A dynamic passenger flow forecasting model was established in this study. CO_2 concentration, which can be easily obtained from the cloud platform, can reflect the dynamic characteristics of passenger flow in the hall and platform of the subway station. Therefore, the dynamic CO_2 concentration data of different sensors was adopted to be input to regress and forecast the passenger flow.

It should be pinpoint that the passenger flow cannot be obtained from the monitoring plat. A passenger flow calculation method was adopted to calculate passenger flow. The passenger flow of platform, hall and transfer plat can be calculated based on the followed formulas [33]:

$$G_1 = \frac{P_{in}}{60} \cdot T_1 + \frac{P_{out}}{60} \cdot T_2 \quad (1)$$

$$G_2 = \frac{P_{in} + X_{tr} + Y_{tr}}{60} \cdot T_3 + \frac{P_{out} + X_{tr} + Y_{tr}}{60} \cdot T_4 \quad (2)$$

$$G_3 = \frac{X_{tr} + Y_{tr}}{60} \cdot T_5 \quad (3)$$

$$G_{total} = G_1 + G_2 + G_3 \quad (4)$$

where G_1, G_2, G_3 and G_{total} are the number of passengers in the hall, plat, transfer plat and whole subway station, respectively, in persons;P_{in} and P_{out} are number of persons entering and leaving the station, respectively, in persons; X_{tr} is residence time of persons entering the station and staying in the hall in min; T_1 is residence time in the hall of persons to enter the station in min; T_2 is residence time in the hall of persons to leave the station in min;T_3 is residence time in the hall of persons to transfer among the halls in min;T_4 is boarding time of persons staying in the plat in min; T_5 is alighting time of persons staying in the plat in min; X_{tr} and Y_{tr} is the transfer passenger flow of line X and line Y in persons.

In accordance with data from the field test and automatic fare collection system during August 2019, the transfer passenger flow and the number of persons entering and leaving the station can be obtained. Also, combined with the subway operation schedule, value of T_1, T_2, T_3, T_4 and T_5 can be assumed as 2 min, 1.5 min, 5 min, 1.5 min and 1.5 min, respectively. The hourly passenger flow was determined by the aforementioned information.

Due to the larger number of CO_2 measuring points installed in

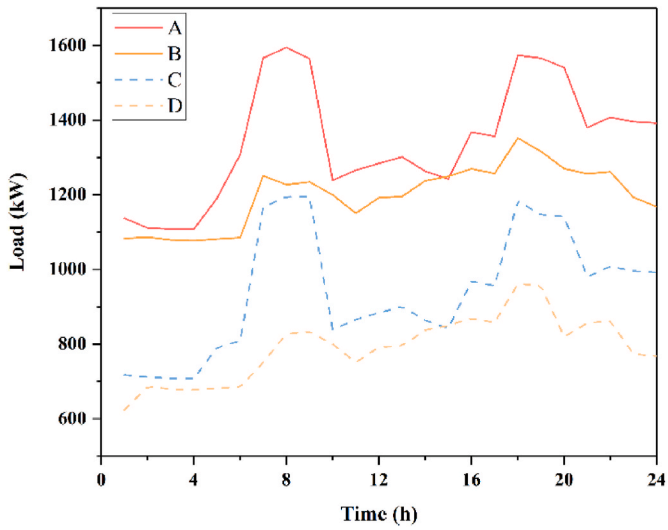

Fig. 3. Daily load curves of 4 different cluster centroids.

different locations of the subway station (platform, hall and transfer plat, etc.), in order to select input data features of reflecting the dynamic passenger flow, Grey Relation Analysis (GRA) method [34], assessing the relativity between the output variable and series input variables, was applied to select significant measuring points among total 16 point (C1 to C16). Based on the grey correlation results of 16 points in the Supplementary Materials (Fig. S4), C7, C8 and C11, whose values of grey correlation are over 0.75, were selected as input features to regress the relationship between CO_2 concertation and passenger flow.

2.3.3. Load characteristics analysis based on clustering method

The air condition load is influenced by various factors (i.e., outdoor meteorological parameters and dynamic changes of passenger flow etc.). Moreover, the factors often follow a cyclical pattern in the season- and day-scale period. The passenger flow, for instance, usually have the similar tendency on the working days and holidays. And in this subway station, the operation strategies of cooling season and transitional season of air conditioning system differ by the outdoor meteorological parameters. The time-series similar changes of meteorological parameters and passenger flow lead to similarity existing among the large number of load data. It is necessary to find such unobservable relationship and similarity, as well as clarify details of input features.

In this part, *k*-means clustering algorithm was used to divide the daily changes of air conditioning load into various kinds. The daily changes tendency of the load within the same kind are similar. The number of clusters *k* should be determined at first in the analysis process. Elbow method, a method to determine optimal clusters number *k*, was used. General speaking, sum of the squared errors *SSE* will decrease with the number of clusters increasing. However, there exists an optimal number of cluster. When the number of clusters is larger than the optimal value *k*, the decrease of sum of the squared errors *E* tends to be slow. The variation trend of *SSE* with different clustering number is the shape of 'elbow'. The results of *SSE* with different clusters numbers are provided in Fig. S5. It can be seen that the value of SSE is improved significantly when the number of clusters is 4. The number of clusters *k* was set as 4 by the Elbow method.

Fig. 3 shows that daily load curves of cluster centroids of 4 clustering kind A, B, C and D. The load value of cluster A and B is obviously higher than of cluster C and D. As for the change trend of the four kinds, there exist peaks and troughs among the load changes of cluster A and cluster C. As indicated in the aforementioned analysis, the passenger flow and

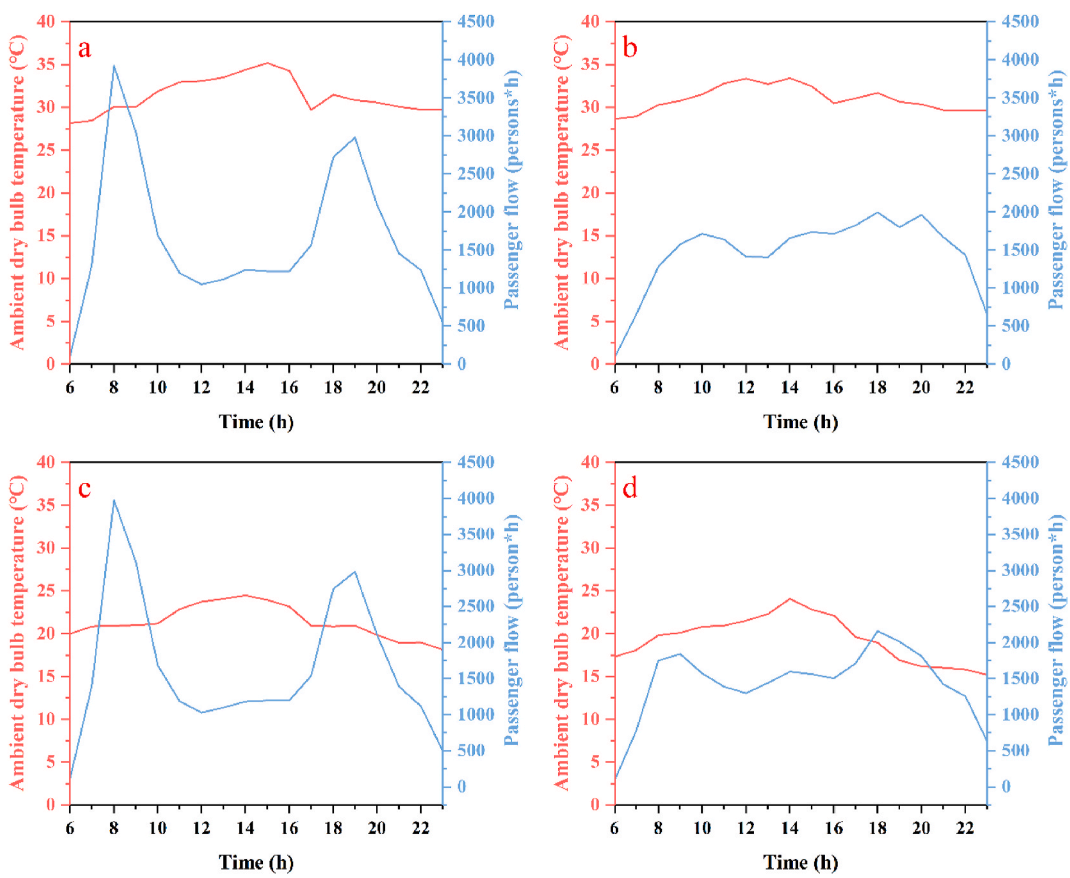

Fig. 4. Results of dynamic passenger flow and ambient dry bulb temperature of (a) working days of cooling season, (b) holidays of cooling season, (c) working days of transitional season and (d) holidays of transitional season.

outdoor meteorological parameters are crucial to the dynamic load. The ambient dry bulb temperature and the passenger flow of four kinds of conditions (i.e., working days of cooling season, holidays of cooling season, working days of transitional season and holidays of transitional season) were downloaded and analyzed in Fig. 4. As shown in Figs. 3 and 4(a) and 4(c), the load peaks are corresponding to the passenger flow peaks on the working days. Also, it can be concluded that the overall level of the load is determined by the outdoor meteorological. The higher load of A and B than it of C and D is due to higher outdoor temperature of the cooling season compared with it of transitional season. Therefore, the dynamic changes of load origin from the fluctuation of passenger flow and outdoor air thermal parameters.

In a word, the air conditioning load characteristic of subway stations is that the load of hall & platform system is influenced by passenger flow and outdoor meteorological parameters. There exists obvious load difference among the holidays and working days. It is necessary to forecast the air-conditioning load by two conditions of working days and holidays. The input features adopted in forecasting air-conditioning load are passenger flow (obtained by CO_2-based model forecasting), outdoor air enthalpy and holiday/working day. However, these features of only a single moment cannot describe the dynamic characteristic of the air-conditioning load. The historical data contained various complicated information of dynamic indicators (i.e., solar radiation flux and rate of utilization for equipment). To forecast the dynamic load accurately, the historical data of load and these input features was taken into consideration to reflect time-series information. So an appropriate time span of historical data should be confirmed firstly. The GRA method was also used to determine the time span of historical load data. The autocorrelation of load within different time intervals is shown in Fig. S6. The threshold value of the correlation was set as 0.8.1 h, 2 h, 3 h, 23 h and 24 h were selected to describe the historical information. As shown in the part of load forecasting model in Fig. 1, passenger flow, ambient wet bulb temperature (enthalpy of outdoor air), current moment and historical load date of selected time span were select to forecast the dynamic load. The selection of ambient thermal parameters (wet bulb temperature and dry bulb temperature) should be clarified. As for the input feature of load characteristic analysis, ambient dry bulb temperature, which is regarded as a classical outdoor meteorological parameter, is used in clustering analysis to describe the relationship between passage of the seasons and dynamic air conditioning load. However, as for forecasting air conditioning load, the load, composed of sensible load and latent load, depends on both temperature and humidity of ambient air. And enthalpy of outdoor environment, deduced by ambient wet bulb air in accordance with Lewis relation, is the indictor used to describe total air conditioning load. Therefore, ambient wet bulb temperature is selected as one of the input features for load forecasting model.

2.4. Energy consumption forecasting model

As aforementioned analysis, the historical data includes various operation information. In the actual operation conditions, the performance of equipment is usually different from it of rated condition. And there exist depreciation of equipment resulting in the performance loss with the process of equipment using. In this section, a data driven model based actual historical data was established to describe the real-time performance of the equipment of ACWS. In the process of establishing the data driven model of the equipment of ACWS, the first task is to select appropriate input features as input of the model. As shown in Fig. 2, the main equipment of ACWS are chillers, cooling towers and variable frequency pumps etc. So the feature selection of the models for the aforementioned equipment were arranged.

2.4.1. Chillers

The existed physical models of chillers are based on empirical fitting of various input features (i.e., GN model [34], linear regression model [35], MP model [36], BQ model [36] and ASHRAE Handbook model [37] etc.). The performance and energy consumption can be determined by the followed features: refrigerating output of chillers, outlet water temperature of cooling towers, inlet water temperature of cool towers, inlet temperature of chilled water and outlet temperature of chilled water. Chilled water flow and the cooling water flow, which also influence energy consumption of the chillers, were taken into consideration. GRA method was adopted in the part to select significant features. 10,000 pieces of data samples were used for the original data of the correlation analysis. The results table can be obtained from Table S2 of the Supplementary Materials. The GRA results show that the cooling water flow and the chilled water flow have great influence on the dynamic power of chillers. Moreover, the outlet temperature of chilled water and cooling water were highly correlated with the inlet temperature of chilled water and cooling water. There existed information redundancy of the inlet and outlet temperature of chilled or cooling water. Also, the model was established to optimize the operation parameters. The dynamic refrigerating output of chillers $Q_{chiller}$, outlet temperature of chilled water $T_{chw,out}$, outlet temperature of cooling water $T_{cw,out}$, chilled water flowm_{chw} and cooling water flow m_{cw} were selected as input features of the data driven model of the chillers.

2.4.2. Pumps

In the water system, transportation system, whose energy consumption cannot be neglected, is used to transport the chilled and cooling water to relative equipment. Variable frequency pumps are adopted in many conditions and in this subway station ACWS.

In this section, the feature selection of variable frequency pumps was arranged. The power of variable frequency pumps can be calculated by Eq. (5):

$$P_{pump} = \frac{\rho g V H}{3.6 \times 10^6 \eta} \tag{5}$$

where P_{pump} is the power of pumps in kW, V is volume flow of pumps in m^3/h, η is the pump efficiency, H is pumping head in m, ρ is density of water in kg/m3; and g is gravitational acceleration in m^2/s.

The pumping head depends on pipeline resistance and the water flow of relative terminals. In accordance with the scaling law of the pumps, the relationship between water flow and pipeline resistance can be written as Eq. (6):

$$\left(\frac{H}{H_{rated}}\right)^2 = \left(\frac{V}{V_{rated}}\right)^3 = \left(\frac{n}{n_{rated}}\right)^3 \tag{6}$$

where H_{rated} is pumping hear under the rated condition in m, V_{rated} is volume flow of pumps under the rated condition in m^3/h, n is pump speed in rpm; and n_{rated} is pump speed under the rated condition in rpm.

And for different pump typess, there exist differences of the pump efficiency. The pump efficiency can be described as the form of Eq. (7):

$$\eta = \eta_p \cdot \eta_m \cdot \eta_{VFD} \tag{7}$$

where η_p is transmission efficiency in %, η_{VFD} is variable-frequency drive efficiency in %; and η_m is motor efficiency in %.

As for the three different efficiency, the transmission efficiency can be defined by given water flow and pump speed. The variable-frequency drive efficiency (VDF) and motor efficiency depend on the pump speed. In general, the relationship of the pump water flow, the pumping head and the pump frequency can be deduced as a single-factor relationship of pump water flow. Therefore, the water flow, a controllable and obtainable parameter, is set as the input feature of data mining model for variable frequency pumps. The pumps frequency can be obtained in accordance with given water flow by Eq. (6).

2.4.3. Cooling towers

Cooling towers, which are used to remove heat of condensation, are important parts of ACWS. The energy consumption of the cooling towers

Table 2
Input features selected for energy consumption and performance forecasting model.

Equipment		Input features
Chillers	Energy consumption	Dynamic refrigerating output of chillers $Q_{chiller}$, outlet temperature of chilled water $T_{chw,out}$, intel temperature of cooling water $T_{cw,in}$, chilled water flowm_{chw} and cooling water flow m_{cw}
Pumps	Energy consumption	Cooling water flow m_{cw} or chilled water flow m_{chw}
Cooling towers	Energy consumption	Fan frequency f_{fan}
	Condensation heat	Fan frequency f_{fan}, cooling water flow m_{cw}, inlet temperature of cooling water $T_{cw,in}$, ambient wet bulb temperature T_{wb}

was established firstly. The existed energy model of cooling towers is based on polynomial fitting. And the energy consumption of cooling towers mainly origins from the fans of cooling towers. The energy consumption of fans depends on air flow driven by the fans. In this study, according to the scaling law, the air flow of fans can be deduced by given fan frequency. The fan frequency of cooling towers can be obtained from the monitoring platform. The fan frequency of cooling tower is used to be a single input feature.

As indicated in section 2.4.1, it is necessary to establish a model for forecasting condensation heat of cooling towers Q_{tower} to close the whole model of ACWS. An experimental model for calculating the condensation heat is written as Eq. (8):

$$Q_{tower} = \frac{d_1 m_a^{d_3}}{1 + d_2\left(\frac{m_a}{m_{cw}}\right)^{d_3}} \times \left(T_{cw,out} - T_{db}\right) \tag{8}$$

where Q_{tower} is the condensation heat removed by cooling towers in kW, m_a is air flow driven by fans of cooling towers in kg/s, m_{cw} is cooling water flow in kg/s, T_{db} is the ambient wet bulb temperature in °C; and d_1, d_2, and d_3 are the coefficients of the experimental model.

As shown in Eq. (8), the condensation heat removed by cooling towers was concluded as the followed features: air flow driven by fans of cooling towers m_a, cooling water flow m_{cw}, inlet temperature of cooling water $T_{cw,in}$ and ambient wet bulb temperature T_{db}. However, the air flow driven by fans of cooling towers cannot be replaced by fan frequency, which can be obtained from the monitoring platform. Fan frequency f_{fan}, cooling water flow m_{cw}, outlet temperature of cooling water from cooling towers $T_{cw,out}$ and ambient wet bulb temperature T_{wb} were selected to forecast the input features for forecasting condensation heat.

In a word, the input features can be obtained based on aforementioned analysis. The detail input features of various equipment used in ACWS of the subway station are indicated in Table 2.

2.5. Optimization of operation parameters of ACWS

2.5.1. Optimization of load distribution of chillers

Due to the design method of air conditioning system, the chillers usually works on the part load condition. Moreover, the air conditioning system was supported by series chillers. It is necessary to optimize load distribution of different chillers of ACWS to improve the operation efficiency of ACWS in the subway station.

In this section, the optimization goal of load distribution was to find a load distribution scheme to minimize of energy consumption of chillers. The Genetic Algorithm (GA) was used to find optimal solution in non-liner and discontinuous design space. The maximum number of generations was selected as optimization stopping criterion. The value of final generations was set as 200. And 200 population members with 100 crossovers (crossover probability: 0.5) and 2 mutants (mutation probability: 0.01) were used for the optimization parameters. The optimal formulation of multi-objective optimization problem is given as follows:

Objective function:

$$F = \sum_{1}^{2} P_{chiller,i} \tag{9}$$

The dynamic refrigerating output of chillers should cover the cooling demand of air conditioning terminals. The relationship between load

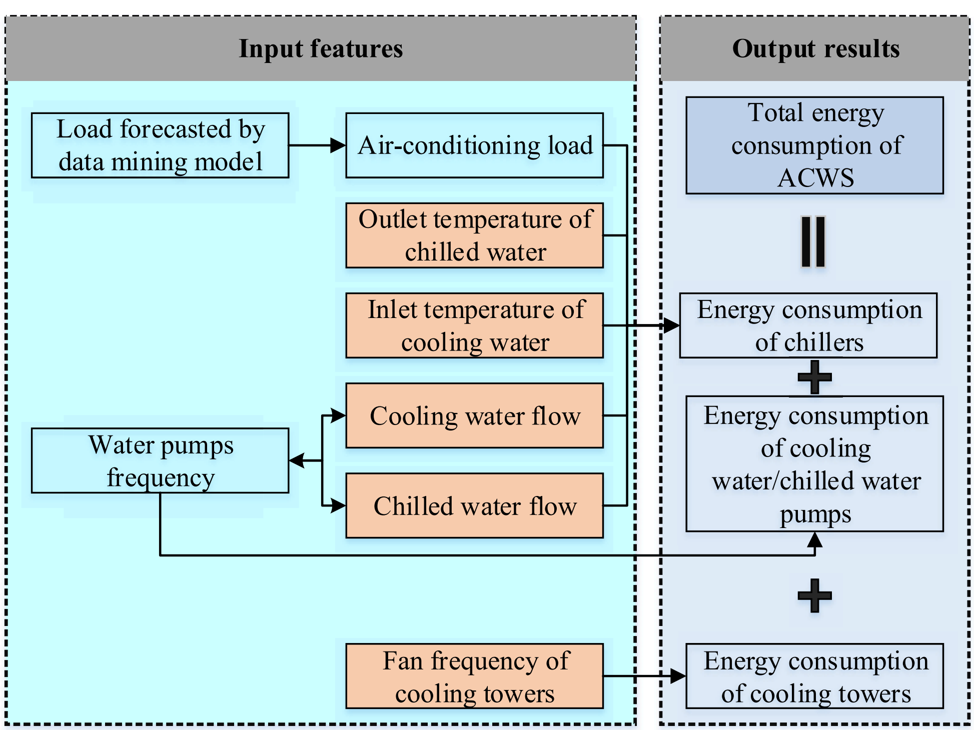

Fig. 5. Details of input features and energy consumption of ACWS.

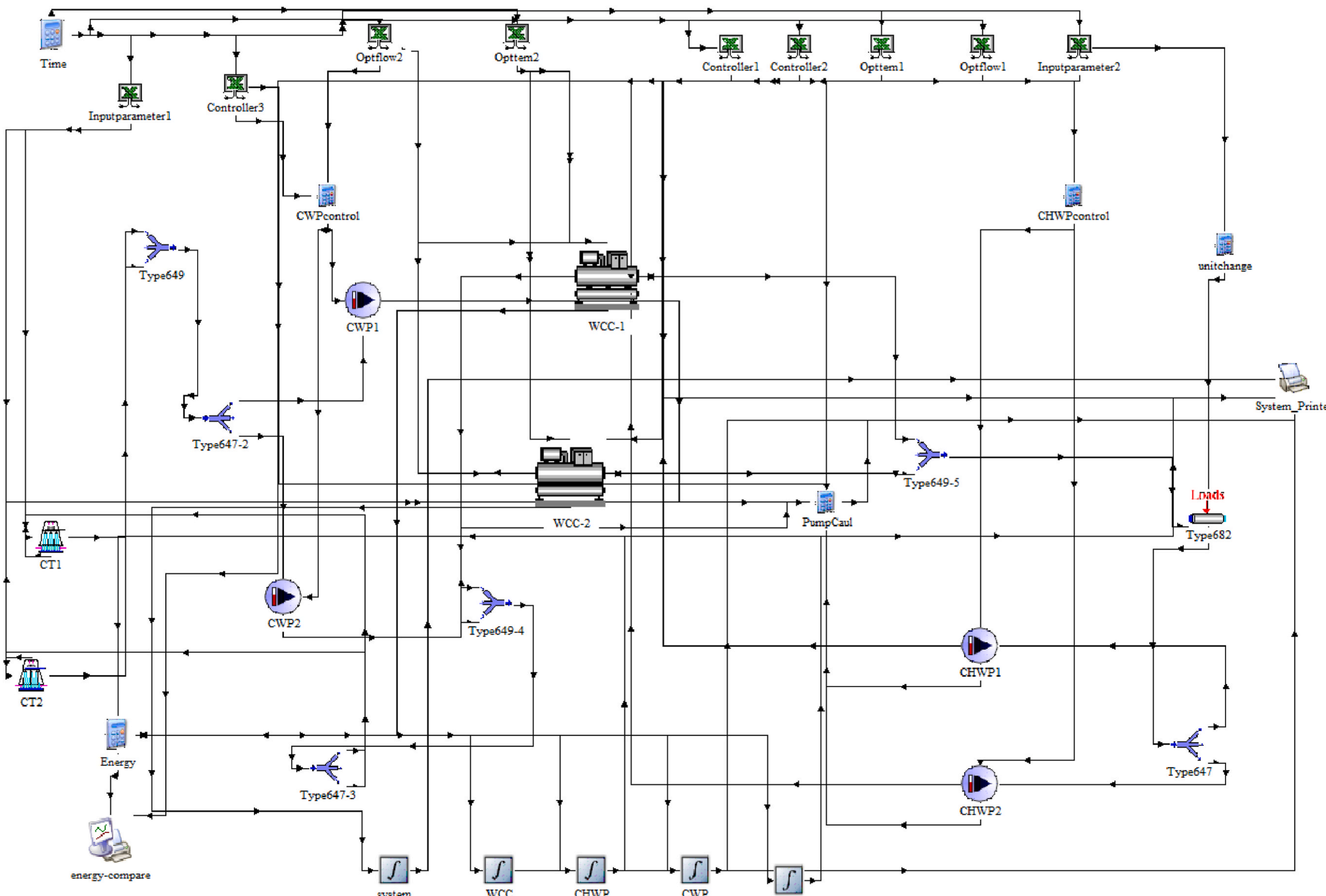

Fig. 6. TRNSYS simulating model of ACWS of the subway station.

and refrigeration output can be expressed as a form of equality constraint of Eq. (10):

$$Q_{cl} = \sum_{1}^{2} (PLR_i \cdot Q_{rated,i}) \tag{10}$$

In the optimization problems, penalty function method was used for transferring the optimal problem with equality constraint into unconstrained optimization problem. The objective function of this problem with penalty functions is shown in Eq. (11):

$$F = \sum_{1}^{2} P_{chiller,i} + \mu \left| Q_{cl} - \sum_{1}^{2} (PLR_i \cdot Q_{rated,i}) \right| \tag{11}$$

where $P_{chiller,i}$ is the energy consumption of the ith chiller in kW, Q_{cl} is the cooling demand of air conditioning terminals in kW, PLR_i is the part load ratio of the ith chiller, $Q_{rated,i}$ is the rated cooling capacity of the ith chiller; and μ is the penalty factor of the objective function.

The energy consumption of chillers was estimated by the model of Section 2.4.1. And the optimization was arranged under the normal condition of chillers. The return temperature of chilled water was set at 12 °C. The input temperature of cooling water was set at 30 °C. Also, the chilled water flow and cooling water flow were set same with the value of the rated condition.

2.5.2. Optimization of operating parameters in ACWS

In section 2.3 and section 2.4, the air conditioning load and energy consumption of ACWS were modeled by their physical characteristic and actual data driving. The optimal parameters should be confirmed before the optimization process. Fig. 2 shows that the ACWS of the subway station is made up by four kinds of equipment: chillers, cooling water pumps, chilled water pumps and cooling towers. There exist the same input features of different equipment models. For example, the fan frequency not only influenced the energy consumption of cooling towers, but also affect the energy consumption of chillers by changing inlet temperature of cooling water.

As shown in Fig. 5, the air conditioning load changes with different outdoor boundary conditions. The air conditioning load is an unknown feature in actual operation condition. However, in this study, the load can be forecasted by the relative forecasting models (i.e., the load forecasting model and the passenger flow forecasting model) based on the outdoor weather parameters and actual measured data, e.g. CO_2 concentration. The load data was able to be known in the optimization process. And the outlet temperature of chilled water can be regulated by changing the set value of chillers. The cooling water flow and the chilled water flow also can be changed by regulating the frequency of pumps with the variable-frequency drive. The inlet temperature of cooling water can be regulated by changing air flow of cooling towers with different fan frequencies. Also the fan frequency can be regulated directly. Therefore, considering the controllability, adjustability and obtainability of the features, outlet temperature of chilled water $T_{chw,out}$, intel temperature of cooling water $T_{cw,in}$, chilled water flow m_{chw}, cooling water flow m_{cw} and fan frequency of cooling towers f_{fan} were selected as optimal parameters. The water flow of pumps can be used to calculate pumps frequency by the scaling law.

The total energy consumption of ACWS is made of four part: energy of chillers, cooling water pumps, chilled water pumps and cooling towers. Based on the single equipment model, the system energy consumption can be summed as Eq. (9):

$$P_{system} = P_{chiller} + P_{chwp} + P_{cwp} + P_{tower} \tag{12}$$

where P_{system} is the energy consumption of the whole ACWS in kW, $P_{chiller}$ is the energy consumption of chillers, P_{chwp} is the energy consumption of chilled water pumps in kW, P_{cwp} is the energy consumption of cooling water pumps in kW; and P_{tower} is the energy consumption of cooling towers in kW.

The optimal objective was minimizing total energy consumption of ACWS for defined design space. The system energy consumption in Eq. (9) was estimated by the model of aforementioned section 2.4. Also the range of 5 optimal input features should be defined.

As for the parameters of chillers (i.e., input temperature or output temperature of chilled water), the low temperature of chilled water will result in chiller fault. Usually, 5 °C was set as the lower limit of outlet temperature of chilled water. In accordance with actual operation conditions, the upper limit of chilled water outlet temperature was set as 12 °C. The range of outlet temperature of chilled water $T_{chw,out}$ can be described as follows:

$$5^{\circ}C \leq T_{chw,out} \leq 12^{\circ}C \tag{13}$$

And due to the outlet water temperature of cooling towers equals to the inlet temperature of cooling water for the condenser of chillers $T_{cw,in}$. the value of $T_{cw,in}$ can be set in a specific range based on the condition of ambient wet bulb temperature. Based on the actual operation data of cooling towers, the range of cooling tower approach temperature was 1–6 °C. In the optimization process, the constraint of $T_{cw,in}$ is as follows:

$$T_{wb} + 1^{\circ}C \leq T_{cw,in} \leq T_{wb} + 6^{\circ}C \tag{14}$$

As for pumps and fans with VFD, the limit value of water flow and air flow can be deduced by the limit of the VFD frequency according to the scaling law. As for common VFD, the upper limit and lower limit of frequency f are 30 Hz and 50 Hz, respectively.

$$30\ Hz \leq f \leq 50\ Hz \tag{15}$$

In addition, the relationship among condensation heat removed by cooling towers Q_{tower}, refrigeration output of chillers $Q_{chiller}$ and energy consumption of chillers is expressed as follow:

$$Q_{tower} = Q_{chiller} + P_{chiller} \tag{16}$$

The penalty function method was used to process the equality constraint and the objective function. The processed objective function is shown in Eq. (14).

$$\min f(T_{chw,out}, T_{cw,in}, m_{chw}, m_{cw}, f_{tower}) = \min(P_{system} + \mu |Q_{tower} - (P_{chiller} + Q_{chiller})|) \tag{17}$$

2.6. Dynamic model of ACWS based on TRNSYS

The optimal parameters were obtained based on the optimization algorithm and data mining model. In this section, a TRNSYS system simulating model was established to validate the performance of the calculated optimal parameters. Also a feedforward control strategy with optimal parameters was arranged and assessed based on the simulating model. As for the equipment model used in the TRNSYS model, MP model was adopted to describe the performance of chillers and pumps. Technical parameters of cooling towers obtained from handbooks were set to be input features of this model. The connect relationship among different equipment is shown in Fig. 6. The operating parameters, ambient thermal parameters (i.e., water temperature, water flow, air temperature and humidity etc.) and control signal were transferred by the lines of the model.

In this simulating process of simulation, the timer was set into initial condition. The EXCEL files were used to storage input parameters of different equipment (i.e., optimal parameter, air conditioning load, ambient wet bulb temperature and control schedules etc.). The load data was inputted into the calculator part to disturb the load ratio, and ambient wet bulb temperature was applied to be input feature of cooling towers model. The ON/OFF schedules were used to control the ON/OFF state of the equipment. Finally, the simulation results and relative dynamic parameters were obtained from the printer part and calculators.

It is necessary to obtain response characteristic of the equipment of ACWS to applied the optimal parameter in the feedforward controlling method. According to the actual operation data of $T_{chw,out}$, $T_{cw,in}$, m_{cw} and

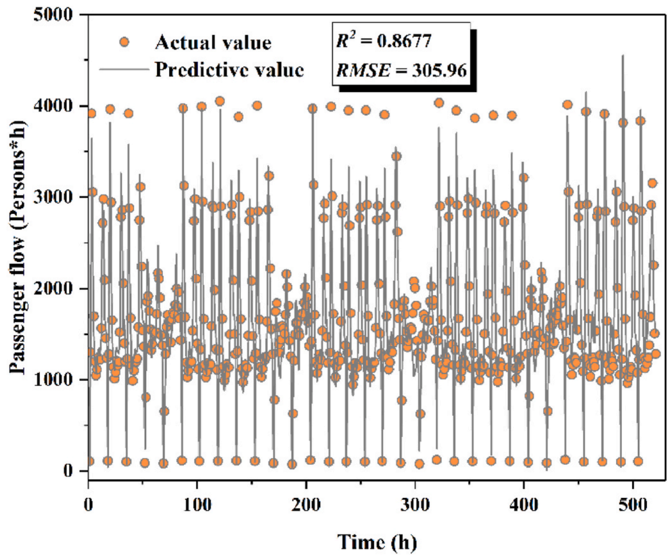

Fig. 7. Passenger flow forecasting results of ANN model.

m_{chw}, the response indicators of chillers, cooling towers and pumps should be calculated.

3. Validation for the established models

3.1. Data mining model

In this study, ANN, owing well learning ability and nonlinear fitting ability, was used to mine the information of relative measured data. As for practice of ANN method, the activation function should be determined by actual mathematical properties of forecasted model. Sigmoid function, a nonlinear relationship fitting function, was used as the activation function of ANN model. Then, the structure parameters of ANN (the number of hidden layers and neurons) were confirmed. Finally, the measured data set was divided into train set and validation set. The comprise of the validation data was arranged to validate the trained model. In the section, correlation coefficient R^2 and Root Mean Squared Error (RMSE) were adopted to assess the validation of trained model.

3.1.1. Load forecasting model

As indicated in section 2.3, the passenger flow, ambient wet bulb temperature, working day/holiday, historical data (1 h, 2 h, 3 h, 23 h, 24 h) and current forecasted time were select as input features. The passenger flow, a curial parameter, was forecasted by an ANN model. The input features of passenger flow forecasting were dynamic concentration of CO_2 from measured points C7, C8 and C11. A single-layer ANN model with seven neurons was trained. The measured data of August 2019 was used to trained and validated the passenger flow forecasting model. The passenger flow forecasting results are shown in Fig. 7. The forecasting results show the ANN model can forecast the dynamic passenger flow by CO_2 concentration accurately. The trained passenger flow forecasting model was combined into the load forecasting model.

The data of a week of August 2019 was selected as test set of load forecasting model. Moreover, the amount of data used for training model was studied in the modelling process. The data of 1 month (July 2019), 2 months (June to July 2019), 3 months (May to July 2019) and 4 months (April to July 2019) was used to train the model. The forecasting performance of different data amount is shown in Fig. 8. There existed no difference of forecasting results calculated by the models, which were trained by data of different time scale. For the predictive performance of all four models, the value of R^2 is more than 0.96, and the range of RMSE

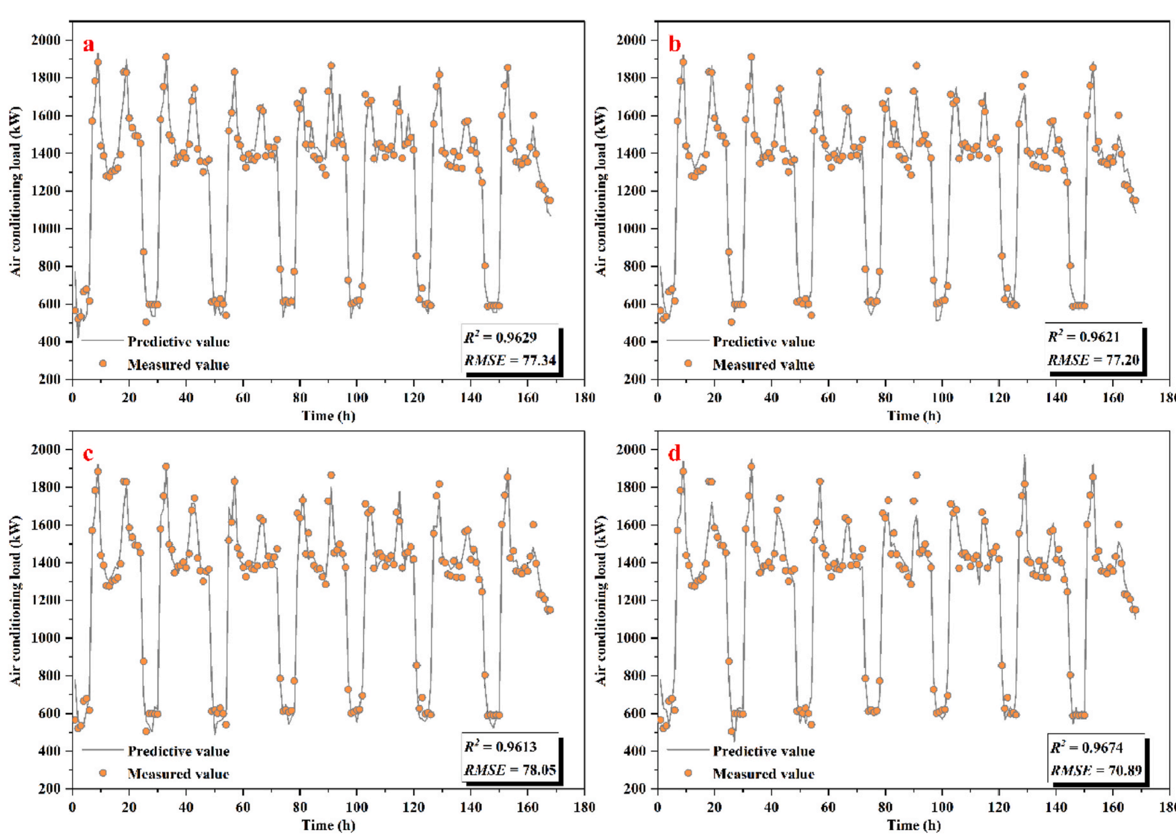

Fig. 8. Comparisons of the load between predictive value and actual value with four data amounts: (a) 1 month, (b) 2 months, (c) 3 months; and (d) 4 months.

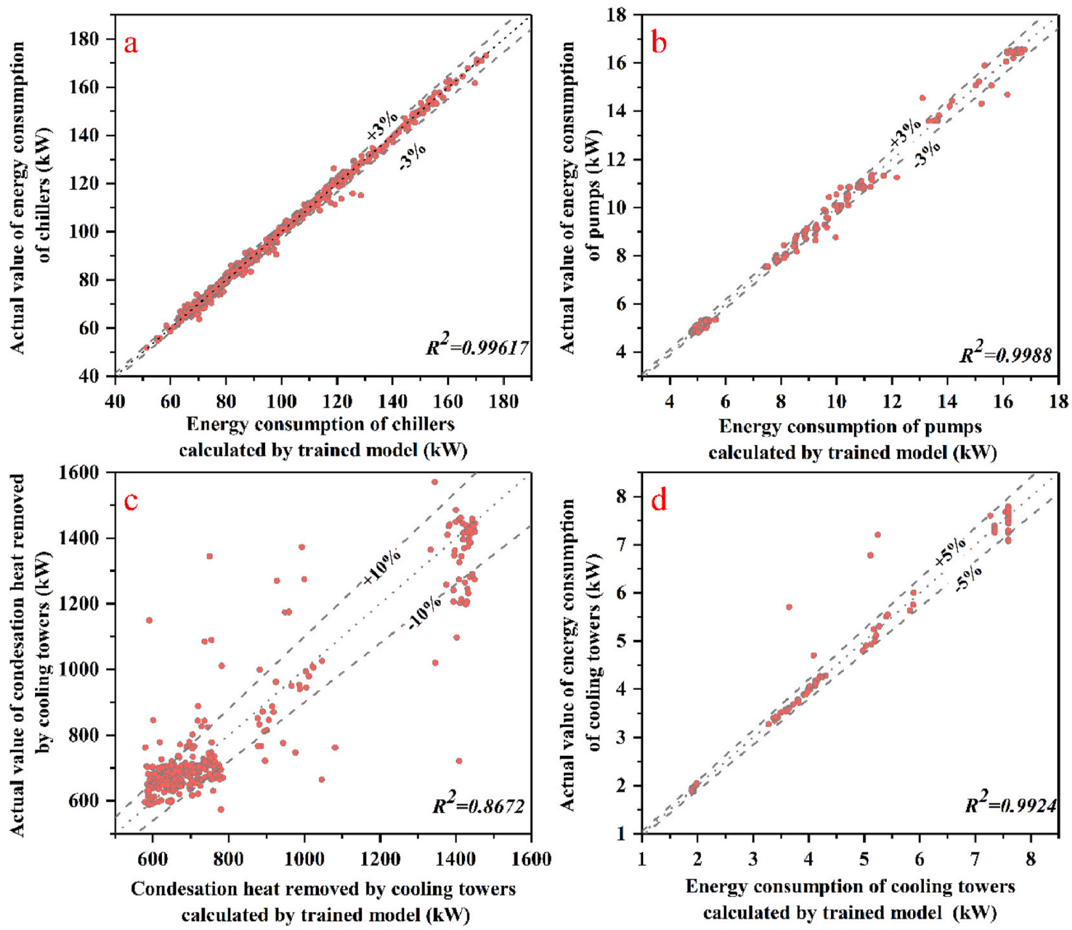

Fig. 9. Comparison of ANN simulation results of cooling towers model and actual measured results: (a) chillers, (b) pumps, (c) condensation heat removed by cooling towers.

is 70–80 kW. The forecasting performance of load is not sensible to the time scale of train set.

3.1.2. Equipment models

The ANN model was adopted to describe chillers, pumps and cooling towers. As for the model of chillers, the refrigeration output of chillers Q_{cl}, chilled water flow m_{chw}, cooling water flow m_{cw}, inlet temperature of cooling water $T_{cw,in}$ and outlet temperature of chilled water were $T_{chw,out}$ taken as input features of ANN model for chillers. The energy consumption of chillers $p_{chiller}$ was set as output feature of the model. The data of March 1st – October 31st' 2019 was processed and used for data set. Stratified sampling was used to construct train set and test set by the time range. The number of data points in the train set and test set was 2872 and 718, respectively. The stopping criterion of training process was as follows: set number of iterations was 1000 Epochs, and target error was 0.001 and learning rate was 0.01. For the number of hidden neurons, it was determined by comparison of fitting performance of various neurons number. Fig. 9(a) shows the comparison of chillers-ANN forecasting results and actual results. The value of R^2 was approximately 0.99. The forecasting results cover the operating conditions of chillers. The relative error of approximate 95% of simulating results is less than 3%. The trained ANN model for chillers can forecast the actual dynamic energy consumption of chillers.

For the model of pumps, the model is established to reflect the relationship between the dynamic power of pumps P_{pump} and the water flow. The two factors were set as the input feature and the output feature of pumps model, respectively. The operating data of March 1st – October 31st' 2019 was downloaded from the monitoring platform to construct the train set and the test set. Also, 3000 samples and 750 samples were used to be the train set and the test set by time-series stratified random sampling. The number of hidden layer neurons was set to be 6 by comparison of forecasting performance. Fig. 9(b) shows the forecasting performance of the trained model for pumps. Compared with the actual data, the simulation results can forecast the working characteristics of pumps. The value of R^2 was approximate 0.99 in the test set. As shown in Fig. 9(b), the number of data points under the low-power condition is small. According to the actual data, the operating frequency of pumps is more than 30 Hz. And the relative error of approximate 97% of simulating results is less than 3%. Therefore, the trained ANN model of pumps can forecast the actual energy consumption of pump in the case of normal operating conditions, and can be used for the following optimization process.

For the models of cooling towers, the models of cooling towers are established to forecast the energy consumption $P_{chiller}$ and the condensation heat removed by cooling towers Q_{tower}. The data source of the parameters for cooling towers is similar with the data source of other equipment models.

For energy consumption model, by the comparisons of different setting parameters, the number of hidden-layer neurons was set as 6.900 data points and 350 data points were sampled by stratified sampling method as the train set and the test set, respectively. As for condensation heat removed by cooling towers, the number of hidden-layer neurons was set as 11. Fig. 9(c) shows the comparisons between forecasting value and actual value of the consumption power of cooling towers. There exists concertation of data points in the energy consumption of 4 kW and 7.5 kW. In accordance with the historical data of cooling towers, the fan frequency of cooling towers is usually set as two standard value: 30 Hz

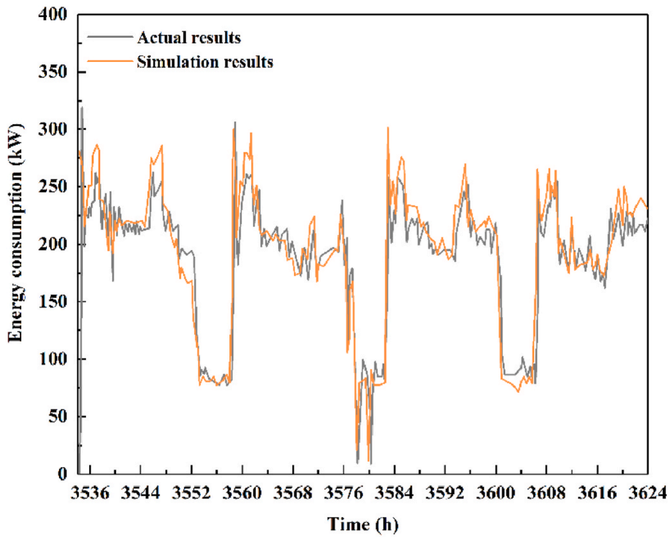

Fig. 10. Validation for TRNSYS model using the actual historical data.

and 50 Hz (the corresponding dynamic energy consumption of fans are 4 kW and 7.5 kW, respectively). And except for few of data points, the relative error of most of data point is within 5%.

For the model of condensation heat removed by cooling towers, the number of hidden-layer neurons was set as 11. As shown in Fig. 9(d), the relative error of 65% of data points was within 10%. And the mean relative error is approximate 5%. Therefore, it can be concluded that ANN model can be used to forecast energy consumption behaviors and condensation removing performance of cooling towers.

3.2. TRNSYS model

In this section, to validate the TRNSYS system model, the operation data of single chiller operating condition and two chillers operating condition during May 2019 downloaded from the platform was adopted. In the single chiller operating condition, the operating equipment of ACWS contains a chiller, a chilled water pump, a cooling water pump and a cooling tower. In the two chillers operating condition, the number of chillers, chilled water pumps, cooling water pumps and cooling towers are two. And the actual controlling schedule and set parameters were applied in the TRNSYS model. The single- and two-chiller condition was used for the day time and night time, respectively. The comparisons between actual results and simulation results by TRNSYS model are shown in Fig. 10. The relative error of most data points, accounting for approximately 85% of all points, is within 10%. The number of simulating data points is few, of which the relative error is more than 20%. The main difference between actual value and simulation value occurs at the moment of switching conditions among single-chiller mode and two-chiller mode. However, the duration of switching conditions is short. The error caused by changing operating modes can be negligible for the whole simulation process. The validated TRNSYS model can be used to assess the energy conservation performance of feedforward controlling with optimal parameters.

4. Results

4.1. Optimization results of regulating chillers load ratio

In section 2.5.1, the objective function and optimal parameters have been defined. The optimization process was based on the rated condition of chillers: the chilled water flow and cooling water flow were the flow value of rated condition. And the inlet temperature of cooling water and returned temperature of chilled water were set as 30 °C and 12 °C, respectively. Because the air conditioning load remain stable during one specific time, a series of load intervals were divided based on various total load ratio of ACWS. When the load ratio of ACWS was 50%, 60%, 70%, 80%, 90% and 100%, the part load ratio (PLR) of the two chillers was obtained by GA. The optimal PLR of two chillers is shown in Table 3.

Table 3
Optimal PLR results of two chillers.

Total load ratio of ACWS/%	PLR_1/ %	PLR_2/ %	Total energy consumption of chillers/kW
50%	48.85%	51.15%	208.53
60%	63.70%	56.30%	242.38
70%	70.37%	69.63%	277.29
80%	78.95%	81.05%	318.83
90%	86.38%	93.62%	365.11
100%	99.99%	99.99%	409.46

The existed study [38] proposed an optimal PLR controlling strategy: the PLR of two chiller remain equal to minimize the energy consumption of two chillers. By the optimal strategy, the total energy consumption of chillers was 208.69 kW, 242.28 kW, 277.27 kW, 318.81 kW, 365.26 kW and 409.30 kW by the aforementioned load dividing method. The optimal PLR results were approximately same with the results of the equal-PLR strategy. Therefore, it can be concluded that the load distribution with equal PLR of chillers is optimal PLR regulating method. The followed optimization process was based on this PLR regulating method.

4.2. Optimization results of operating parameters of ACWS

The air conditioning load maintains stable with small variation during a period. The load of air conditioning system of management room, for instant, is usually constant in the evening. Switching the operation parameter frequently will lead to fluctuation of relative parameters influencing the normal operation of the system. Therefore, the optimization of parameters by specific load intervals is used in the study.

The inlet temperature of cooling water $T_{cw,in}$ is influenced by the ambient web ball temperature T_{wb}. The wet bulb temperature changes with different locations and different periods. The ambient wet bulb temperature of Guangzhou, for example, fluctuate in the range of 24–28 °C. The diurnal range of ambient wet bulb temperature is approximate 1 °C. The range of fluctuation in ambient wet bulb temperature is far lower than it of load fluctuation. Therefore, the optimization of operating parameters was arranged in various load ratio intervals with a constant ambient wet bulb temperature. And then, the optimal parameters of various load ratio intervals are updated by changing ambient wet bulb temperature. In accordance with the load distribution schedule of section 4.1, the load ratio of 50% is regard as a critical point of switching conditions of single- or two-chiller operation. The optimization results of various load ratio intervals can be found in Fig. S7 in the Supplementary Material.

4.3. Responses performance analysis of ACWS

The feedforward control was adopted to overcome the delaying effect of feedback control method. The indicator, relaxation time, was introduced to assess the response time of different equipment. As for feedforward control method, the regulated parameters should be stable and achieve the set value. However, there exist responses problems of the actual controlling actions. It should take some time to achieve stable state after regulating operating parameters. Therefore, the relaxation time has to be obtained to guide the advanced time of feedforward controlling. The pump water flow, inlet temperature of cooling water and outlet temperature of chilled water were regulated indirectly by adjusting the frequency of VFD and the opening value of guide vanes. The responses performance of three indicators was analyzed as follows.

According to the actual operating data of ACWS, due to lack of reasonable control strategy, outlet temperature of chilled water, inlet

Table 4
Responses analysis of different cases.

Case No.	Set outlet temperature of chilled water/°C	Pumps frequency/Hz	Frequency of fans of cooling towers/Hz
1	10.5 to 8.5	50	50
2	8.5 to 10.5	50	50
3	10.5	30 to 50	50
4	10.5	50 to 30	50
5	10.5	50	30->50
6	10.5	50	50->30

temperature of cooling water and pumps water flow were regulated by a predetermined schedule. The set outlet temperature of chilled water of working-days condition, for instant, is usually set to be 8.5 °C during the morning rush hours, 9 °C during the evening rush hours and 10 °C during other hours. The frequency of fans and pumps was adjusted in the range of 30–50 Hz.

Therefore, the actual data of 6 conditions were extracted from the platform for analyzation. The detail information of these conditions is shown in Table 4.

Based on the operating data of ACWS, the dynamic responses of various equipment are shown in Fig. 11. Fig. 11(a) and (b) show the dynamic outlet temperature of chilled water changes after the regulating action. When the outlet temperature of chilled temperature is regulated from 10.5 °C to 8.5 °C and from 8.5 °C to 10.5 °C, respectively, the relaxation time of adjusting temperature of chilled water is approximate 30 min. Similarly, the relaxation time of adjusting frequency of pumps and fans are 15–20 min and 35min~40min, respectively. To obtain

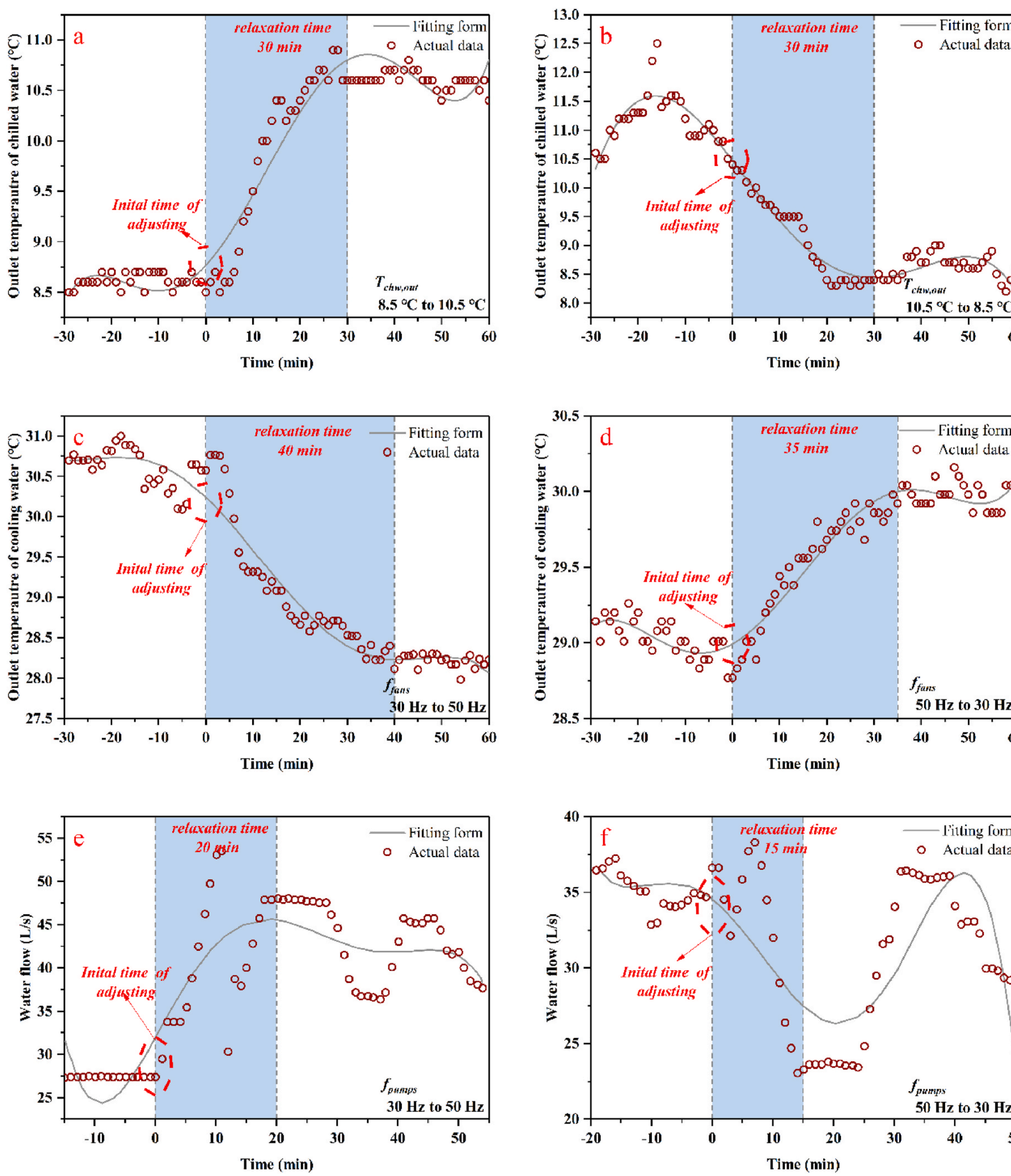

Fig. 11. Response results of (a) Case No.1, (b) Case No.2, (c) Case No.3, (d) Case No.4, (e) Case No.5; and (f) Case No.6.

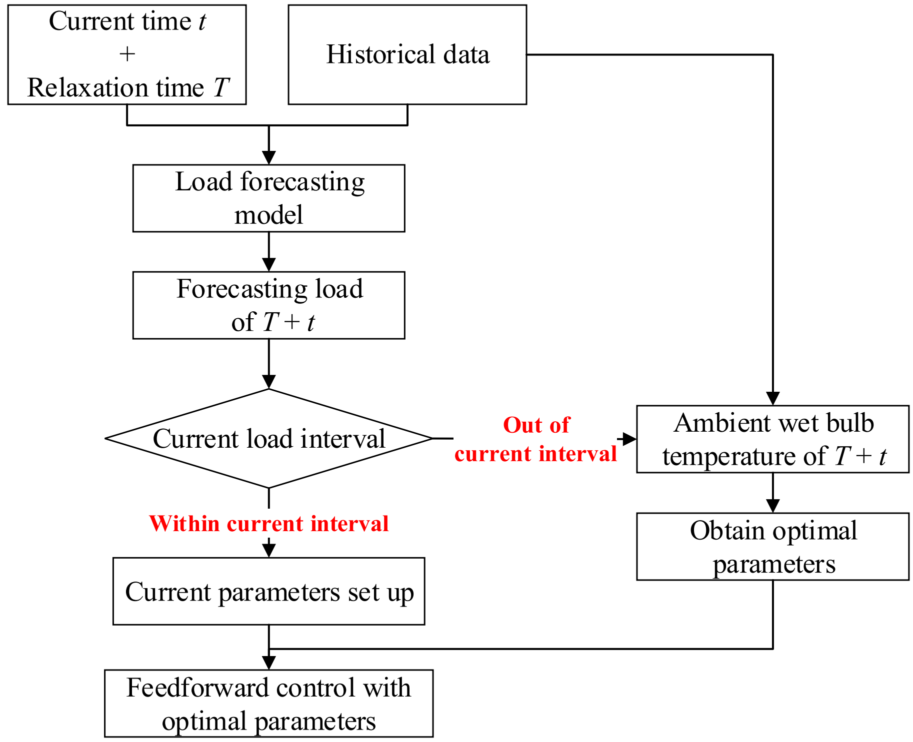

Fig. 12. Flow chart of optimal feedforward control.

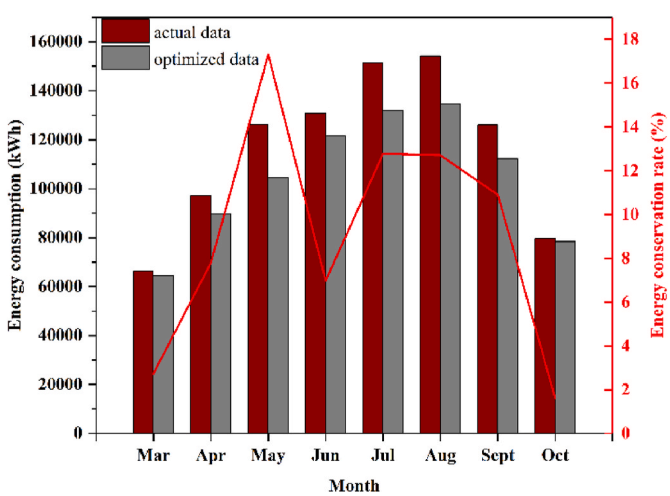

Fig. 13. Energy conservation performance of optimal feedforward control.

responses characteristic of different equipment, based on the actual operating data of July to September 2019, the continuous function was obtained by fitting method. And the relaxation time was determined by the derivative of the fitting function being zero. It can be concluded that the relaxation time of changing outlet temperature of chilled water remains approximate 30min. It of changing frequency of pumps varies in the range of 10–30 min. And it takes approximate 20–40 min to reach the stable value of outlet temperature of cooling water after regulating frequency of fans of cooling towers. The range of the relaxation time of operating parameters is 10–40 min. The forecasting time should be 40 min earlier than current time to reach stable state and eliminate delaying effect of regulating actions. Also, to ensure ACWS working on the high-efficiency condition, the time interval between forecasting time and current time cannot be set too long. 1 h was taken into consideration of using it as the time interval of feedforward controlling.

4.4. Energy conservation performance of feedforward controlling

In this section, the feedforward control strategy is introduced. The optimal operating parameters can be found by the divided load interval. The feedforward controlling process can be adopted as followed: First, based on the time interval of 1 h, the existed parameters remained constant while the forecasting load will be within current load ratio range; and the operating parameters will be updated into the new value by ambient web bulb temperature of current time while the forecasting load is out of current load interval. The flow chart of feedforward control method by obtained optimal parameters is shown in Fig. 12.

Fig. 13 shows the energy consumption comparisons between actual conditions and the condition with proposed feedforward control method. The actual data is the energy consumption data of the cooling season in 2019. And the optimal data is calculated by the validated TRNSYS model. The energy conservation rate of the cooling season is 9.48%. And the energy conservation rate of Mar. and Oct. is on the low side, within the range of 1–3%. The main reason is that the load of these months is on the low level, and the ACWS was operating on the single-chiller condition. The number of chillers, cooling towers, chilled- and cooling-water pumps is one single. Also, the equipment of ACWS in the studied subway station were designed in accordance with design guide of energy-efficiency demonstration project. These equipment was of high efficiency. The regulation performance of the relative equipment under the condition of the low and stable PLR is well. In case of the single-chiller condition, during the transition season, the dynamic change of air conditioning load is sight. The schedule-based operating parameters of the ACWS is suitable for the actual operation. The energy conservation potential with optimized parameters on this condition is limited. As for the other months, all the chillers, pumps and cooling towers should be opened and operates with high PLR. And there exists the variance of load distribution among the chillers. Also the control strategy with a schedule of fixed parameters is unreasonable. This lead to the huge energy conservation potential of feedforward control with optimal parameters. As for the whole cooling season, the COP of chillers used in the subway station was approximate 6.4, and the COP of the whole ACWS was approximate 5.6. The energy conservation performance of this system was well. In case of using feedforward control method, the COP of chillers increases to 7.5, and the COP of ACWS has reached to 6.2. The energy conversation rate has reach to 9.48% based on the existed high efficiency equipment.

5. Conclusions

In this work, the authors aim to develop a data-mining-based optimal feedforward control for ACWS in the subway station. The data mining models were adopted and validated to forecasting air conditioning load and equipment characteristics of the ACWS. Based on the well-trained models, the optimal PLR distribution and optimal operating parameters were obtained by GA. With the combination of ACWS responses performance and obtained optimal operating parameters, the optimal feedforward operation control strategy was assessed in a validated TRNSYS system model. The main conclusions were summarized below:

(1) The ANN predictions of passenger flow, air conditioning load, equipment operating parameters were well accepted when compared with measured data. T
(2) As for load forecasting model, the predictive performance is independent of the duration of input historical data.
(3) The equal-PLR strategy is the optimal load distribution schedule for the chillers group.
(4) The relaxation time of regulating actions is in the range of 10–40 min, and the required value of lag time for feedforward control is 40 min.
(5) Based on the optimal schedule and system responses time, energy consumption of can be save by approximately 9.5% during the whole cooling season with the adoption of optimal feedforward control strategy.

Author statement

Xing Su: Conceptualization, Methodology, Data curation, Funding acquisition, Writing - review & editing, Supervision. Yixiang Huang: Writing - original draft, Methodology, Data curation. Lei Wang: Methodology, Investigation, Data curation. Shaochen Tian: Writing - review & editing. Yanping Luo: Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research has been supported by the National Natural Science Foundation of China under Grant No. 52178085.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jobe.2021.103379.

References

[1] L. Zhao, L. Shen, The impacts of rail transit on future urban land use development: a case study in Wuhan, China, Transp, Policy 81 (2019) 396–405, https://doi.org/10.1016/j.tranpol.2018.05.004.
[2] B. Guan, T. Zhang, X. Liu, Performance investigation of outdoor air supply and indoor environment related to energy consumption in two subway stations, Sustain. Cities Soc. 41 (2018) 513–524, https://doi.org/10.1016/j.scs.2018.06.009.
[3] B. Guan, T. Zhang, X. Liu, Performance investigation of outdoor air supply and indoor environment related to energy consumption in two subway stations, Sustain. Cities Soc. 41 (2018) 513–524, https://doi.org/10.1016/j.scs.2018.06.009.
[4] Y. Li, Q. Zhang, J. Wang, C. Yue, L. Che, J. Zhao, Y. Wang, F. Duan, Field investigation on operation parameters and performance of air conditioning system in a subway station, Energy Explor. Exploit. 38 (2020) 235–252, https://doi.org/10.1177/0144598719881224.
[5] S. Pan, F. Pei, Y. Wei, H. Wang, J. Liu, X. Zhang, G. Li, Y. Gu, Design and experimental study of a novel air conditioning system using evaporative condenser at a subway station in Beijing, China, Sustain. Cities Soc. 43 (2018) 550–562, https://doi.org/10.1016/j.scs.2018.09.013.
[6] Z. Yang, Z. Yu, L. Yu, F. Ma, Research on frequency conversion technology of metro station's ventilation and air-conditioning system, Appl. Therm. Eng. 69 (2014) 123–129, https://doi.org/10.1016/j.applthermaleng.2014.04.016.
[7] C. Fan, F. Xiao, H. Madsen, D. Wang, Temporal knowledge discovery in big BAS data for building energy management, Energy Build. 109 (2015) 75–89, https://doi.org/10.1016/j.enbuild.2015.09.060.
[8] F. Xiao, C. Fan, Data mining in building automation system for improving building operational performance, Energy Build. 75 (2014) 109–118, https://doi.org/10.1016/j.enbuild.2014.02.005.
[9] C. Fan, F. Xiao, C. Yan, A framework for knowledge discovery in massive building automation data and its application in building diagnostics, Autom. ConStruct. 50 (2015) 81–90, https://doi.org/10.1016/j.autcon.2014.12.006.
[10] Y. Chen, X. Hao, G. Zhang, S. Wang, Flow meter fault isolation in building central chilling systems using wavelet analysis, Energy Convers. Manag. 47 (2006) 1700–1710.
[11] Z. Du, X. Jin, Y. Yang, Fault diagnosis for temperature, flow rate and pressure sensors in VAV systems using wavelet neural network, Appl. Energy 86 (2009) 1624–1631, https://doi.org/10.1016/j.apenergy.2009.01.015.
[12] J.E. Seem, Using intelligent data analysis to detect abnormal energy consumption in buildings, Energy Build. 39 (2007) 52–58, https://doi.org/10.1016/j.enbuild.2006.03.033.
[13] O. Ogunsola, L. Song, Application of a simplified thermal network model for real-time thermal load estimation, Energy Build. 96 (2015) 309–318, https://doi.org/10.1016/j.enbuild.2015.03.044.
[14] I. Pardoe, R.D. Cook, X. Yin, Graphical tools for quadratic discriminant analysis, Technometrics 49 (2007) 172–183, https://doi.org/10.1198/004017007000000074.
[15] A. Likas, N. Vlassis, J.J. Verbeek, The global k-means clustering algorithm, Pattern Recogn. 36 (2003) 451–461, https://doi.org/10.1016/S0031-3203(02)00060-2.
[16] S.X. Hao, H. Song, X.F. Zhou, Predicting missing values with KNN based on the elimination of neighbor noise, Comput. Simul. 31 (2014) 264–268.
[17] S. Schiavon, K.H. Lee, F. Bauman, T. Webster, Simplified calculation method for design cooling loads in underfloor air distribution (UFAD) systems, Energy Build. 43 (2011) 517–528, https://doi.org/10.1016/j.enbuild.2010.10.017.
[18] R. Kumar, R.K. Aggarwal, J.D. Sharma, Energy analysis of a building using artificial neural network: a review, Energy Build. 65 (2013) 352–358, https://doi.org/10.1016/j.enbuild.2013.06.007.
[19] F. Lai, F. Magoulès, F. Lherminier, Vapnik's learning theory applied to energy consumption forecasts in residential buildings, Int. J. Comput. Math. 85 (2008) 1563–1588, https://doi.org/10.1080/00207160802033582.
[20] J. Yang, H. Rivard, R. Zmeureanu, On-line building energy prediction using adaptive artificial neural networks, Energy Build. 37 (2005) 1250–1259, https://doi.org/10.1016/j.enbuild.2005.02.005.
[21] R. Mena, F. Rodríguez, M. Castilla, M.R. Arahal, A prediction model based on neural networks for the energy consumption of a bioclimatic building, Energy Build. 82 (2014) 142–155, https://doi.org/10.1016/j.enbuild.2014.06.052.
[22] Y. Ye, C. Jing, Global optimization of a central air-conditioning system using decomposition–coordination method, Energy Build. 42 (2010) 570–583.
[23] F. Bo, X. Jin, Z. Du, Optimal control strategies for multi-chiller system based on probability density distribution of cooling load ratio, Energy Build. 43 (2011) 2813–2821.
[24] J. Li, Z. Li, Model-based optimization of free cooling switchover temperature and cooling tower approach temperature for data center cooling system with water-side economizer, Energy Build. 227 (2020) 110407, https://doi.org/10.1016/j.enbuild.2020.110407.
[25] V. Ongradac, F. Kuli, Recognition of the importance of using artificial neural networks and genetic algorithms to optimize chiller operation, Energy Build. 47 (2012) 651–658.
[26] L. Lu, W. Cai, Y.S. Chai, L. Xie, Global optimization for overall HVAC systems—Part I problem formulation and analysis, Energy Convers. Manag. 46 (2005) 999–1014.
[27] L. Lu, W. Cai, Y.C. Soh, L. Xie, Global optimization for overall HVAC systems—Part II problem solution and simulations, Energy Convers. Manag. 46 (2005) 1015–1028.
[28] A. Kusiak, Y. Zeng, G. Xu, Minimizing energy consumption of an air handling unit with a computational intelligence approach, Energy Build. 60 (2013) 355–363.
[29] A. Kusiak, G. Xu, Modeling and optimization of HVAC systems using a dynamic neural network, Energy 42 (2012) 241–250.
[30] W. Gang, J. Wang, S. Wang, Performance analysis of hybrid ground source heat pump systems based on ANN predictive control, Appl. Energy 136 (2014) 1138–1144.
[31] W. Gang, J. Wang, Predictive ANN models of ground heat exchanger for the control of hybrid ground source heat pump systems, Appl. Energy 112 (2013) 1146–1153.
[32] E. Biyik, A. Kahraman, A predictive control strategy for optimal management of peak load, thermal comfort, energy storage and renewables in multi-zone buildings, J. Build. Eng. 25 (2019) 100826, https://doi.org/10.1016/j.jobe.2019.100826.

[33] Liang Yuan, Study of the computational person in public condition system of transfer subway station, refrig, Air Cond 27 (2013) 158–161 (In Chinese).
[34] Y. Kuo, T. Yang, G.W. Huang, The use of grey relational analysis in solving multiple attribute decision-making problems, Comput. Ind. Eng. 55 (2008) 80–93.
[35] D.J. Swider, M.W. Browne, P.K. Bansal, V. Kecman, Modelling of vapour-compression liquid chillers with neural networks - ScienceDirect, Appl. Therm. Eng. 21 (2001) 311–329.
[36] T.A. Reddy, K. Andersen, An evaluation of classical steady state off line linear parameter estimation methods applied to chiller performance data, HVAC R Res. 8 (2002) 101–124.
[37] R. Engineers, ASHRAE Handbook. Heating, Ventilating, and Air-Conditioning Systems and Applications, ASHRAE Handbook. Heating, Ventilating, and Air-Conditioning Systems and Applications, 2012.
[38] J.E. Braun, S.A. Klein, J.W. Mitchell, W. Beckman, Applications of optimal control to chilled water systems without storage, Ashrae Trans 95 (1989) 663–675.