Journal of Power and Energy Engineering, 2024, 12, 75-98
https://www.scirp.org/journal/jpee
ISSN Online: 2327-5901
ISSN Print: 2327-588X

Correlation Study of Operational Data and System Performance of District Cooling System with Ice Storage

Hui Cao[1], Nan Li[1,2*], Jiajing Lin[1]

[1]School of Civil Engineering, Chongqing University, Chongqing, China
[2]National Centre for International Research of Low-Carbon and Green Buildings, Ministry of Science & Technology, Chongqing University, Chongqing, China
Email: *nanlicqu@126.com

How to cite this paper: Cao, H., Li, N. and Lin, J.J. (2024) Correlation Study of Operational Data and System Performance of District Cooling System with Ice Storage. *Journal of Power and Energy Engineering*, **12**, 75-98.
https://doi.org/10.4236/jpee.2024.123006

Received: January 4, 2024
Accepted: March 26, 2024
Published: March 29, 2024

Abstract

The district cooling system (DCS) with ice storage can reduce the peak electricity demand of the business district buildings it serves, improve system efficiency, and lower operational costs. This study utilizes a monitoring and control platform for DCS with ice storage to analyze historical parameter values related to system operation and executed operations. We assess the distribution of cooling loads among various devices within the DCS, identify operational characteristics of the system through correlation analysis and principal component analysis (PCA), and subsequently determine key parameters affecting changes in cooling loads. Accurate forecasting of cooling loads is crucial for determining optimal control strategies. The research process can be summarized briefly as follows: data preprocessing, parameter analysis, parameter selection, and validation of load forecasting performance. The study reveals that while individual devices in the system perform well, there is considerable room for improving overall system efficiency. Six principal components have been identified as input parameters for the cold load forecasting model, with each of these components having eigenvalues greater than 1 and contributing to an accumulated variance of 87.26%, and during the dimensionality reduction process, we obtained a confidence ellipse with a 95% confidence interval. Regarding cooling load forecasting, the Relative Absolute Error (RAE) value of the light gradient boosting machine (lightGBM) algorithm is 3.62%, Relative Root Mean Square Error (RRMSE) is 42.75%, and R-squared value (R^2) is 92.96%, indicating superior forecasting performance compared to other commonly used cooling load forecasting algorithms. This research provides valuable insights and auxiliary guidance for data analysis and optimizing operations in practical engineering applications.

Keywords

DCS, Correlation Coefficient, PCA, Hourly Cooling Load, System Performance

1. Introduction

Since the Industrial Revolution, human civilization has mainly relied on and consumed fossil energy, allowing global industrialization and urbanization to develop rapidly, with civilization and prosperity growing by the day, while at the same time, the rapid development has also brought about environmental pollution, energy shortages, climate change and many other problems [1]. The Paris Agreement's central aim is to strengthen the global response to the threat of climate change by keeping global temperature rise this century well below 2˚C above pre-industrial levels and to pursue efforts to limit the temperature increase even further to 1.5˚C (United Nations, 2015). Based on climate model simulations, experts suggest that we will reach 1.5˚C of global warming between 2030 and 2052, as show in **Figure 1** [2]. Approximately one-third of the world's total energy consumption is used in buildings, heating ventilation and air conditioning energy consumption in developed countries accounts for half of the energy use in buildings [3] [4], as show in **Figure 2**. At a global level, space cooling makes up only 6% of energy demand in buildings. This is because of low ownership of air-conditioners in many of the world's warmest regions due to low incomes. As incomes increase over the Outlook in these regions, the number of households with space cooling rises rapidly, adding to the growth in electricity demand [4].

The electricity consumption for building operations accounts for nearly one quarters of the total electricity consumption in china, of which 60% can be attributed to heating, ventilation, and air conditioning equipment [5]. Cooling and heating loads are significant portions that can be shifted throughout time, as power grids are increasingly becoming stressed, commercial companies and institutions are hoping that incentives and electric cost plans to be encouraged on demand side load reduction [5] [6]. DCS techniques provide an alternative solution to enhance the energy efficiency of a central cooling system during part load conditions, it has been recognized as one of the effective methods to enhance the energy efficiency in buildings, especially for places where different tariff is adopted for top/peak and flat/valley section of energy used [7], At the same time, DCS in conjunction with the use of renewable energy sources will also help to reduce the rise in global temperature.

To ensure the efficiency of DCS, the control of the system needs to adapt to the aging of the system, require the least human intervention, and adapt to the changes of electricity prices in real time [8]. With the emergence of smart grid and the more penetration of unstable and intermittent renewable energy, DCS

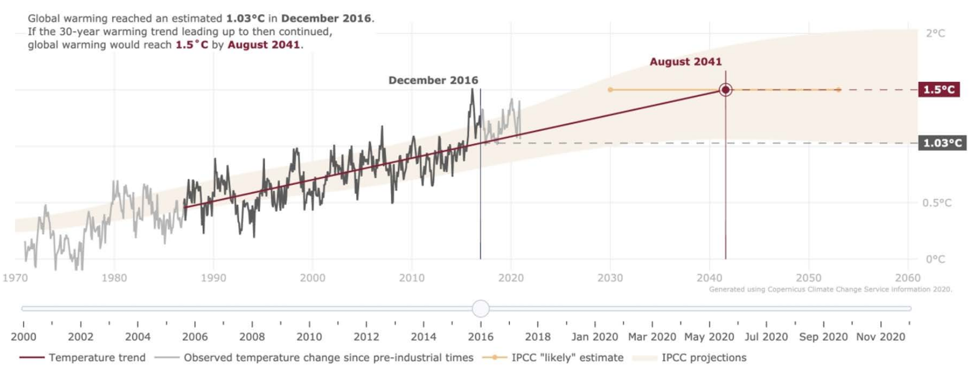

Figure 1. A simple model for estimating the time when global warming is likely to reach 1.5˚C.

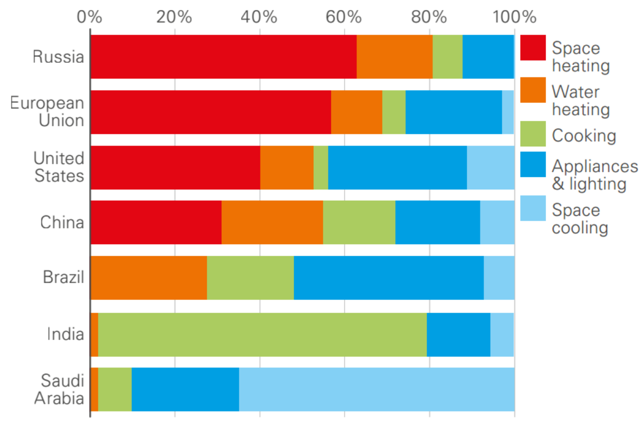

Figure 2. Energy demand by service in key countries and regions.

will be highly needed to balance the supply and demand sides of the power grid [9] [10]. Various factors influence the DCS loads such as building type, occupancy, meteorological conditions, equipment efficiency, thermostat setpoints, season and building controls [11] [12] [13]. Many artificial intelligence (AI) algorithms are used for data-driven load forecasting models. Such as artificial neural networks (ANN), random forests (RF), generalized regression neural network (GRNN), eXtreme gradient boosting (XGBoost), suppport vector machine regression (SVMR), long and short term memory(LSTM), deep neural networks(DNN) and light gradient boosting machine (LightGBM) algorithm, etc. [14]-[22]. For AI load forecasting models, the quality of the training samples largely determines the forecasting accuracy. There are always some algorithms applies better, that can produce better results for lower dimensional data inputs, with greater interpretability and smaller computational resources [23]. Based on a large number of literature studies, we chose Algorithm LSTM to predict the hourly cooling loads and to compare the study with some other load forecasting algorithms that are commonly used.

In the evaluation of the system, the decision variables of load forecasting should be determined first. Given the new trends of data collection and data analytics, data-driven solution can more accurately solve the optimization problem for the DCS. Our research evaluated the system's current operating status and contributed to the subsequent improvement of performance and to the full realization of the savings potential of the system. Also, the field measurements are collected and will be used to develop a mathematical model for the DCS and to evaluate the proposed optimization control strategy under a demand response program. Laying the groundwork for the optimize control research process in the future which runs in real time to properly update the ice storage charge and adapt to weather forecasts and tariff changes.

2. Methodology

2.1. Information of the system

The operational structural configurations for the primary and secondary sides of the system are depicted in **Figure 3**. The system is designed with a specified cooling total load of 187,517 kW and a heating total load of 85,465 kW. Its structural arrangement features a conventional setup where refrigeration units are positioned upstream, followed by a series arrangement of open-type discharge ice systems with connected ice tanks. The designed capacity of these ice tanks is 330,785 kWh.

The primary pumps mainly address the resistance encountered from equipment such as refrigeration units and filters. In contrast, the secondary pumps are designed to counteract the resistance from the plate heat exchangers at the terminal user and the associated distribution network. Given the configuration of the system, where refrigeration units are positioned upstream, followed by a series-connected open-type discharge ice system with ice tanks, when the units collaborate with the ice storage tanks for cooling supply, the network's return water undergoes primary cooling through plate heat exchangers corresponding to both the heat pump unit and the duplex unit. Subsequently, it undergoes secondary cooling and heat exchange via the discharge ice supply plate heat exchanger, ensuring the maintenance of a relatively stable supply water temperature.

The system mainly has 18 heat pump units, including 10 heat pump units and 8 duplex chiller units. The system is also equipped with 4 river water intake pumps, 5 primary chilled water pumps, 8 external network circulation pumps, 8 sets of cooling plate heat exchange, 6 sets of ice discharge plate heat exchange, and 8 ice discharge chilled water pumps. There is also a comprehensive energy consumption monitoring and control platform, which is the basis for efficient energy saving and safe operation and maintenance of the energy station. The platform can monitor and record the operating status parameters of DCS online 24 hours.

2.2. Data Preprocessing

The cold load calculation formula is

$$Load = Q\rho C(T_{in} - T_{out}) / 3600 \tag{1}$$

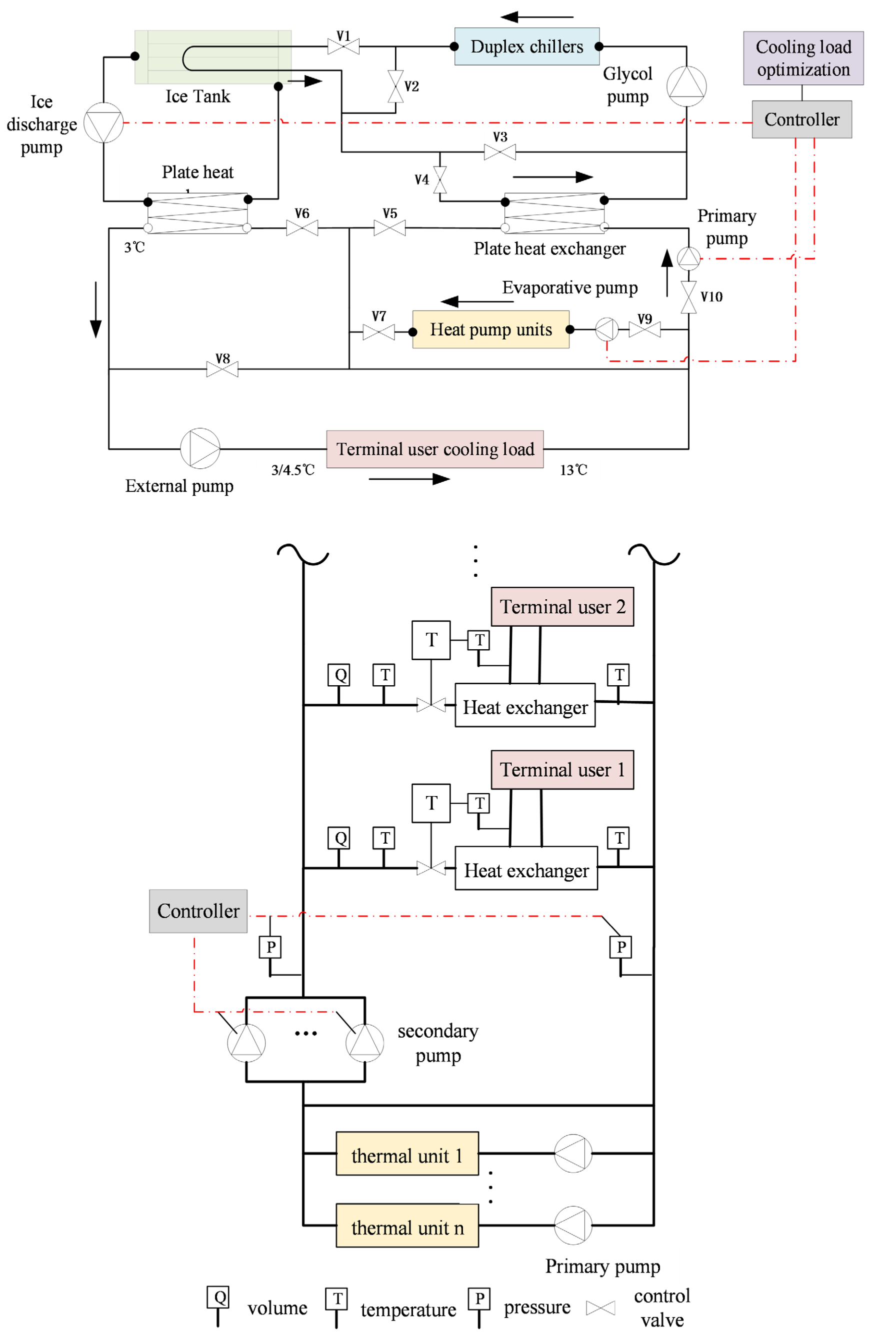

Figure 3. Sketch of the operating mode structure on the primary and secondary side of the system.

where Q is the chilled water flow (m³/h); ρ is the chilled water density (kg/m³); C is the specific heat capacity (kJ/kg ˚C); T_{out} is the supply temperature (˚C); T_{in} is the return temperature (˚C).

The preprocessing of load data primarily is data smoothing techniques: Determine whether the data in the load sequence for day d at time t exhibits any anomalies. Here, $t = 0,1,2,\ldots,23$, $d = 1,2,3,\ldots,n$. The specific steps for smoothing treatment are as follows:

(1) Calculate the mean $E(t)$ and standard deviation $\sigma(t)$ of the cooling load at time t over n days.

$$E(t) = \frac{1}{n}\sum_{d=1}^{n} x(d,t) \tag{2}$$

$$\sigma(t) = \sqrt{\frac{1}{n}\sum_{d=1}^{n}\left[x(d,t) - E(t)\right]^2} \tag{3}$$

(2) Calculate the deviation rate $\eta(d,t)$ of the cooling load at time t on day d and the maximum deviation rate $\eta_{\max}(d,t)$ at time t over n days.

$$\eta(d,t) = \frac{\left|x(d,t) - E(t)\right|}{\sigma(t)} \tag{4}$$

(3) Identify potential outliers. Set the maximum allowable deviation rate as C. If $\eta_{\max}(d,t) > C$, it is deemed that an outlier exists at time t. Subsequently, verify if $\eta(d,t) > C$ holds true, thereby completing the assessment of the deviation rates at time t across all historical days to pinpoint all outliers. In this study, the empirical value set for C is 3.

(4) Correct the identified outliers. Replace the outlier $x(d,t)$ with the mean $\overline{x}(d,t)$ of time t from one day prior and one day subsequent (or multiple days) to the outlier.

$$\overline{x}(d,t) = \frac{1}{2}\left[x(d-1(n),t + x(d+1(n),t)\right] \tag{5}$$

Repeat the process from steps 2 to 4 to identify all potential outliers, thereby completing the data smoothing procedure. It is also worth noting that in this study, data preprocessing is not solely limited to the cooling load data; it is also applicable to the data concerning factors affecting the cooling load. When necessary, preprocessing techniques can be employed for these datasets as well.

2.3. Correlation Study

Given our data collection involves a preprocessing step, this study opts for the Spearman rank correlation coefficient. The Spearman rank correlation coefficient is a non-parametric measure, denoted as r_s, which is independent of the distribution [24]. It quantifies the linear correlation between ordinal variables. Assuming that the original sample data (x_i, y_i) is arranged in ascending order, denoting $(x_i^{'}, y_i^{'})$ as the positions of the sorted data from (x_i, y_i), $(x_i^{'}, y_i^{'})$ is referred to as the ranks of (x_i, y_i). $d_i = x_i^{'} - y_i^{'}$ represents the differences in the ranks of (x_i, y_i).

If there are no tied ranks, r_s is given by:

$$r_s = \frac{n\sum x_i^{'} y_i^{'} - \sum x_i^{'} \sum y_i^{'}}{\sqrt{n\sum x_i^{'2} - \left(\sum x_i^{'}\right)^2}\sqrt{n\sum y_i^{'2} - \left(\sum y_i^{'}\right)^2}} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n\left(n^2 - 1\right)} \tag{6}$$

The Spearman rank correlation coefficient is often regarded as the Pearson correlation coefficient between the variables after ranking. If tied ranks are present, it becomes necessary to compute the Pearson correlation coefficient between the ranks. In such a case, r_s is given by:

$$r_s = \frac{\sum_i \left(x_i' - \overline{x'}\right)\left(y_i' - \overline{y'}\right)}{\sqrt{\sum_i \left(x_i' - \overline{x'}\right)^2 \sum_i \left(y_i' - \overline{y'}\right)^2}} \tag{7}$$

For the aforementioned correlation coefficients, n ideally should not be less than 10. When n is around 10, a correlation coefficient of at least approximately 0.7 indicates a close relationship. For $n \geq 10$, a correlation coefficient greater than 0.3 suggests that the two variables have met the threshold for a close relationship. In this study, n corresponds to the volume of data over a two-month period, with the order of magnitude being 10^3. Therefore, concerning the degree of correlation between parameters studied in this paper, our findings are summarized in **Table 1**.

2.4. Principal Component Study

Regarding the input parameters for the cooling load, PCA is employed to perform linearly independent transformations and dimensionality reduction on the input parameters, ultimately determining the input parameters for the cooling load forecasting model.

It should be clarified that principal components can be derived either from the covariance matrix Σ (where variables are not standardized) or from the correlation matrix R (where variables are standardized). In this study, the option of standardized variables is chosen.

Suppose there are n samples, with each sample observing m variables. This yields the original sample data matrix:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}_{n\times m} \tag{8}$$

x is an m-dimensional vector composed of m random variables $x_1, x_2, \ldots, x_m$, denoted as $x = (x_1, x_2, \cdots, x_m)^{\mathrm{T}}$. Its mean vector is represented as μ, the covariance is Σ, and the linear combination is:

$$\begin{cases} F_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m \\ F_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m \\ \cdots \\ F_m = a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mm}x_m \end{cases} \tag{9}$$

Table 1. The values of $|r_s|$ and their corresponding degrees of correlation.

$\lvert r_s \rvert$	0	0~0.3	0.3~0.6	0.6~0.9	0.9~1	1
Correlation degree	Unrelated	Weakly correlated	Significantly correlated	Highly correlated	Extremely correlated	Perfectly correlated

To standardize the columns of *X*, we have:

$$x_{ij}{}^{*} = \frac{x_{ij} - \overline{x}_j}{\sigma_j} \tag{10}$$

where $i = 1, 2, \cdots, n$, $j = 1, 2, \cdots, m$, $\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$, $\sigma_j{}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_{ij} - \overline{x}_j\right)^2$.

Clearly:

$$\overline{x}_j{}^{*} = \frac{1}{n}\sum_{i=1}^{n}\left(x_{ij}{}^{*}\right) = 0 \tag{11}$$

$$\sigma^{*}{}_j{}^{2} = \frac{1}{n}\sum_{i=1}^{n}\left(x_{ij} - \overline{x}_j\right)^2 = 1 \tag{12}$$

$$\left\| x_j{}^{*} \right\| = \sqrt{x_j{}^{*\mathrm{T}} x_j{}^{*}} = \sqrt{n} \tag{13}$$

where $x_j{}^{*} = (x_{1j}{}^{*}, x_{2j}{}^{*}, \cdots, x_{nj}{}^{*})^{\mathrm{T}}$.

At this point, for the sake of convenience in notation, the standardization of the variable $x_j{}^{*}$ is represented without the subscript $*$. Consequently, in the sample covariance matrix *S*, the variances of variables x_j and x_k is:

$$\sigma_{jk} = \frac{1}{n}\sum_{i=1}^{n}\left(x_{ij} - \overline{x}_j\right)\left(x_{ik} - \overline{x}_k\right) = \frac{1}{n}\sum_{i=1}^{n} x_{ij}x_{ik} = \frac{1}{n} x_j{}^{\mathrm{T}} x_k \tag{14}$$

simultaneously, the correlation coefficient between the variables x_j and x_k is:

$$r_{jk} = \frac{\sigma_{jk}}{\sqrt{\sigma_{jj}}\sqrt{\sigma_{kk}}} = \sigma_{jk} = \frac{1}{n} x_j{}^{\mathrm{T}} x_k \tag{15}$$

From this, it can be deduced that after standardization, the sample covariance matrix *S* and the correlation matrix *R* of the variables are identical. Moreover:

$$R = S = \frac{1}{n} X^{\mathrm{T}} X \tag{16}$$

where *X* represents the data after standardization, and this property can be succinctly denoted as

$$\mathrm{cov}(X) = (V^{1/2})^{-1} \Sigma (V^{1/2})^{-1} \triangleq R \tag{17}$$

here $V^{1/2} = diag(\sqrt{\sigma_{11}}, \sqrt{\sigma_{22}}, \cdots, \sqrt{\sigma_{mm}})$. Then the problem of PCA transforms into the task of determining the principal components starting from the correlation matrix *R*.

2.5. Cooling Load Forecasting Algorithm

The XGBoost algorithm has been recognized as one of the most accurate methods for cooling load forecasting. Prior to the introduction of LightGBM, XGBoost was the most renowned gradient boosting decision tree (GBDT) tool available [25]. However, when dealing with large datasets with multiple features, training with XGBoost can be time-consuming and memory-intensive. LightGBM offers optimizations over traditional GBDT algorithms, addressing the aforementioned limitations of XGBoost [14]. Without compromising accuracy, LightGBM accelerates the training speed of GBDT models. Hence, this study adopts the LightGBM algorithm. LightGBM is a Gradient Boosting Machine (GBM) algorithm utilized for both classification and regression tasks. It employs an ensemble learning method based on trees and leverages gradient boosting to combine multiple weak learners, typically decision trees, into a robust model.

3. Results and Discussion

3.1. The Correlation between System Data and System Performance

3.1.1. Cooling Load Data with System Performance

Define After data preprocessing, the hourly cooling load from June 1st to August 31st during the cooling season of 2022 was determined. Within this dataset, we selected the hourly cooling load for a typical day. Subsequently, a comparative analysis was conducted with the hourly cooling load of the design day, as illustrated in **Figure 4**.

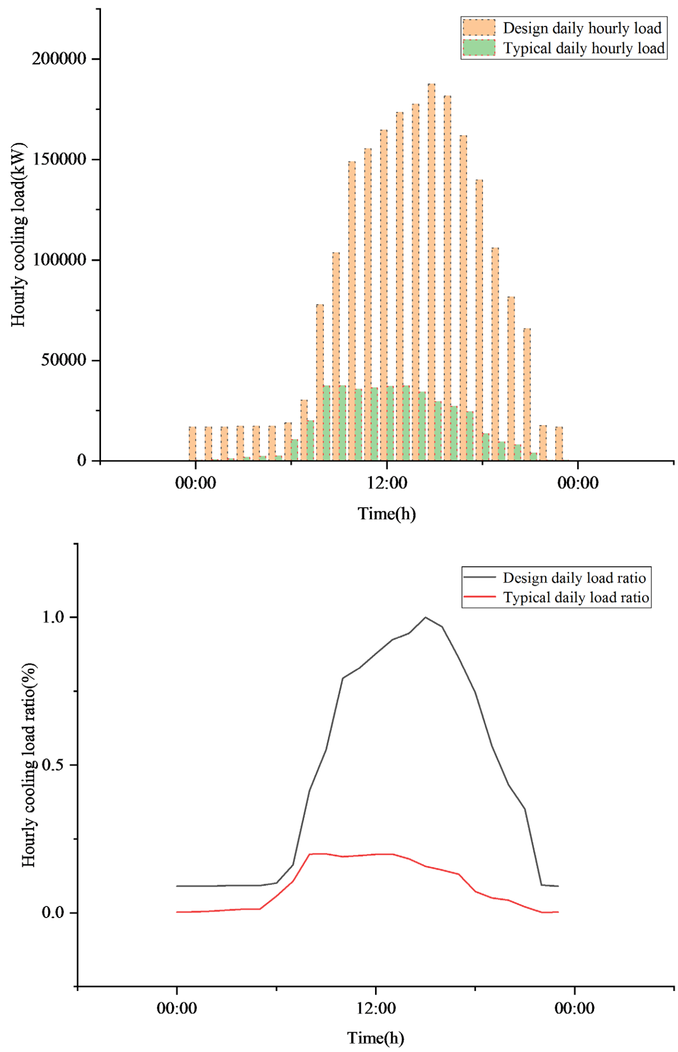

Figure 4. Hourly cooling loads and ratio for design day and typical day.

Based on the two figures above, it is evident that the typical daily load of the system closely aligns with the design day load, exhibiting a trend of increase followed by a decline. This pattern aligns with the work and life routines of individuals primarily associated with office buildings served by this system. On a daily basis, the peak load predominantly occurs between 10:00 AM and 5:00 PM, with a significant drop in load demand after 7:00 PM. Between 11:00 PM and 6:00 AM, there is minimal demand for load. The design day load is substantially higher than the typical daily load, with an average load rate difference of 0.4. This discrepancy arises because the district is an emerging economic zone where the DCS may potentially cater to a larger number of terminal-users in the future. This study, grounded in historical and current cooling load data of the DCS, takes into account the anticipated increased cooling demand due to regional growth. By integrating ice storage technology, it offers crucial parameter data and feasibility analysis for optimizing the operation of the DCS.

3.1.2. Refrigeration Unit Data with System Performance

Through the energy management platform, hourly data for the evaporator and condenser sides, including inlet and outlet flow rates, inlet and outlet water temperatures, unit power, current ratios, etc., of a typical cooling month (August 2022) for the refrigeration units were obtained. This study collects data from the typical month, focusing on the duplex chiller units during the nighttime (00:00-7:00) for ice charge and the heat pump units during the day (8:00-20:00) for cooling.

The hourly values and distribution of the COP for the unit throughout this month can be determined, as illustrated in **Figure 5**.

It can be observed that within this typical month, the hourly COP values for the duplex chiller units predominantly range between 3.5 and 4.4, with a monthly average COP value of 3.96. The rated COP value for the duplex chiller

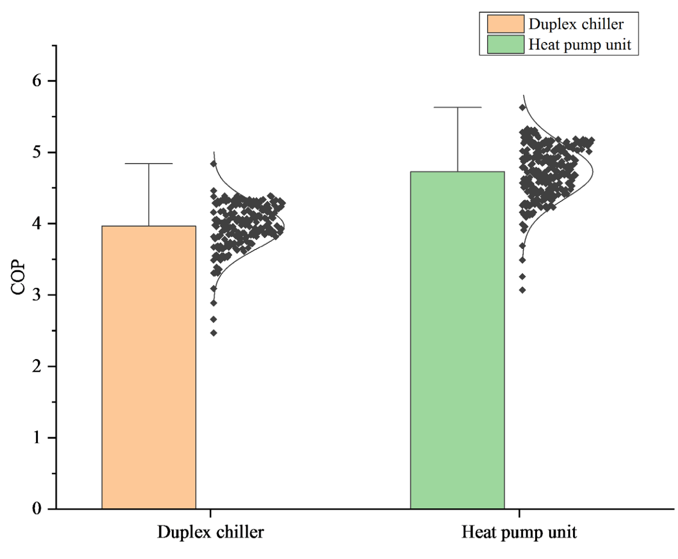

Figure 5. Visualization of hourly COP statistics for a typical month.

units is 4.37. The hourly and monthly average COP values for the duplex chiller units are quite close to their rated values, indicating their satisfactory operational performance. This is attributed to the fact that during nighttime ice charge, the duplex chiller units often operate at maximum capacity. Apart from start-up and shut-down phases, the load ratio of the duplex chiller units remains relatively high most of the time.

Similarly, for the heat pump units, the actual hourly COP values predominantly range between 4.3 and 5.4, with a monthly average COP value of 4.73. However, the rated COP value for the heat pump units under standard conditions is 5.40. There's a considerable deviation between the hourly and monthly average COP values and their rated values. One contributing factor could be that during the summer, the sand content in the river water exceeds the filtration capacity of the intake system, leading to reduced water supply on the condensing side of the units. Additionally, it's evident that the current DCS experiences relatively low-end user loads. During the daytime, cooling predominantly relies on discharging ice. The heat pump units are used to complement the loads that exceed the capacity of the ice storage tanks, resulting in a lower load ratio for these units. The primary reason behind this phenomenon is the poor self-regulation capability of the DCS. There's a frequent need to manually start or stop the heat pump units in response to fluctuations in the system's terminal demand, causing the units to operate at prolonged low load ratios. To enhance the operational efficiency of the heat pump units, improvements in the control system are essential. It's crucial to allocate the load responsibilities of the system equipment reasonably and plan the loads of the refrigeration units based on load forecasts. Optimizing the start-stop operations of the units will ensure their operation at higher load ratios.

3.1.3. Ice Tank Data with System Performance

Figure 6 depicts the cooling details of the DCS within this typical month for the district. The ice discharge for cooling is divided into two sections based on tariff: valley/flat section and top/peak section. Within this typical month, the cooling load for most days is predominantly managed by ice discharge cooling. However, there are only 9 days where the ice cooling ratio is at 1, and these days predominantly fall on weekends. On these specific days, the cooling units within the system remain inactive.

Furthermore, the cumulative ice cooling volume has never reached the designed storage capacity of the ice storage tanks. Regardless of whether the terminal-user cooling demand surpasses the designed storage capacity of the ice storage tanks, except for those 9 days, the refrigeration units are activated. This implies that the potential of the ice storage tanks is not fully utilized. This phenomenon often occurs during top/peak tariff periods when the system activates the refrigeration units to meet the terminal-user demands. Initiating the refrigeration units during top/peak tariff periods is typically uneconomical. The average daily ice cooling volume for this typical month stands at 212,908.87 kWh,

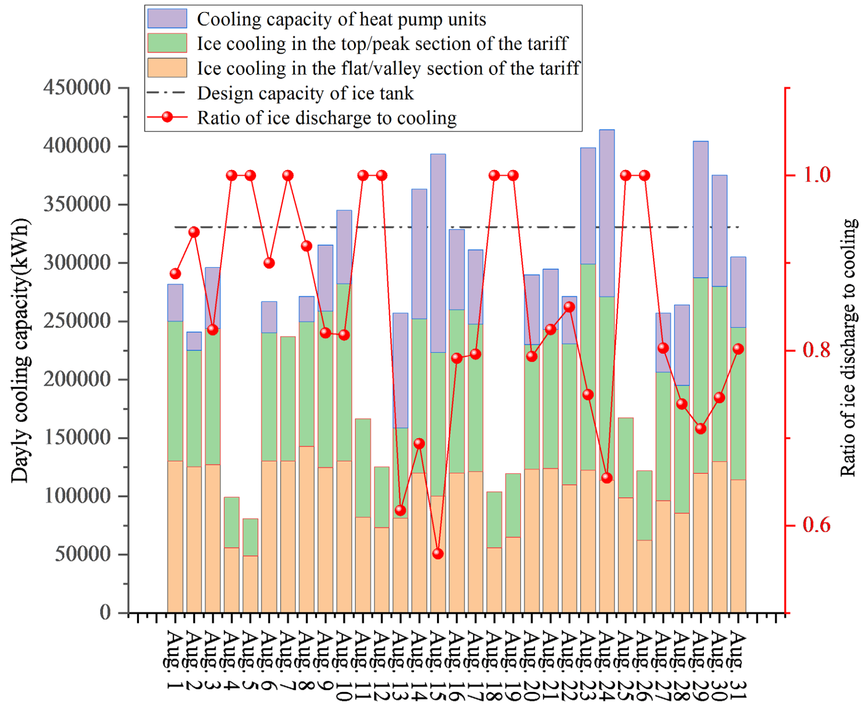

Figure 6. Details of cooling in a typical day and month for DCS.

representing only 64.36% of the ice storage tank's designed capacity. Given this scenario, the average ice cooling ratio for the ice storage tanks is 0.85, indicating excellent defrosting performance. However, the potential of the DCS to optimize peak shaving and load balancing remains underutilized, underscoring the need for an hourly load optimization for the system's components.

3.1.4. The current Overall Performance of the System

Figure 7 illustrates the daily fluctuations of the Energy Efficiency Ratio (EER) and the Cost Per Unit of Cooling Capacity (CPOC) during the typical summer month. EER represents the ratio of the system's cooling capacity to its total power consumption, while CPOC denotes the ratio of the system's operating cost to its cooling capacity. The EER predominantly falls within the range of 2.23 to 3.11, with an average of 2.79, slightly below the empirical average (3.02) derived from extensive studies and statistical analyses of relevant engineering data in the country. Meanwhile, the CPOC is mainly concentrated between 0.13 and 0.19 yuan/kWh, with an average of 0.16 yuan/kWh, also lower than the empirical average (0.2 yuan/kWh) based on certain engineering benchmarks. This indicates that the system's operational cost-effectiveness is relatively favorable.

Compared to the current control strategies, if this study can further enhance the energy-saving rate and cost-saving rate of the system while fully leveraging the advantages of peak-shifting and valley-filling through the ice storage system, it implies that by improving the system's EER, the CPOC can be further reduced. Therefore, there is considerable room for improvement in both the energy efficiency and economic aspects of the system.

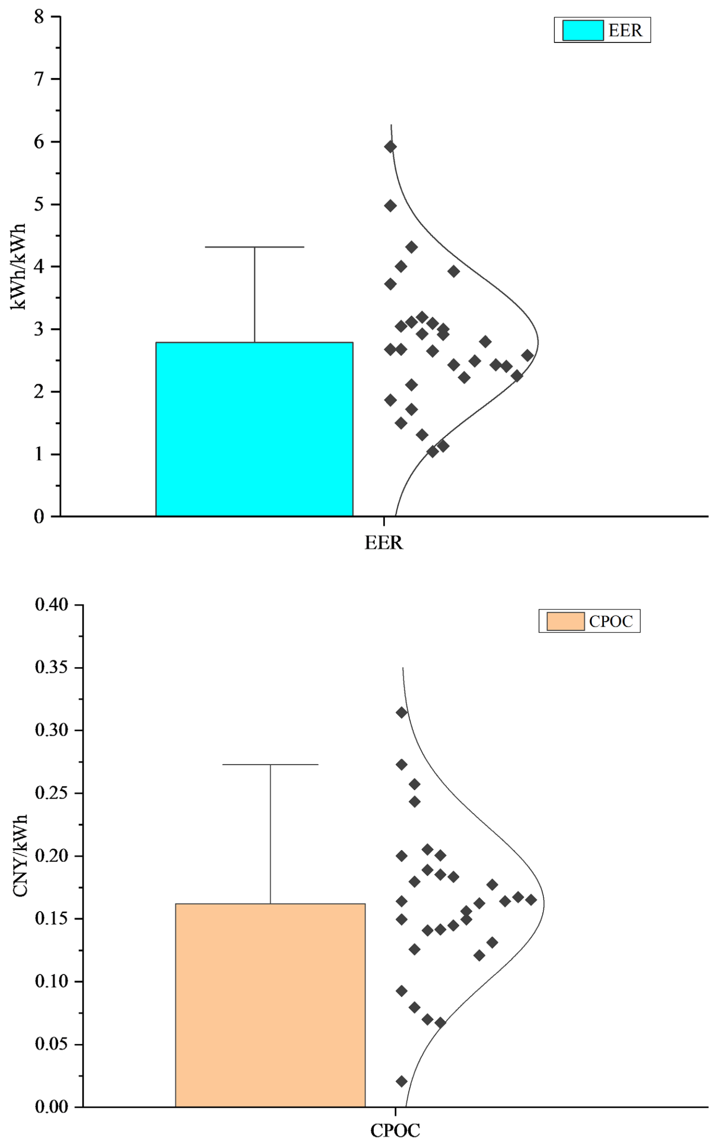

Figure 7. CPOC and EER for the DCS during the typical summer month.

3.2. Preliminary Determination of input Parameters

Regarding the DCS's cooling load and its forecasting models, the influencing factors on the cooling load primarily encompass meteorological factors and intrinsic factors of the DCS itself.

For meteorological influencing factors, this study primarily investigates the factors that relatively have a significant impact on the cooling load: dry bulb temperature T, relative humidity φ, solar radiation intensity R, precipitation amount W, atmospheric pressure P, and atmospheric wind speed V.

For the intrinsic factors of the DCS itself, factors such as indoor set parameters, internal disturbance parameters, occupancy rate of individuals indoors, and usage rate of indoor equipment cannot be overlooked. Given the broad supply range of the DCS and the multitude of terminal-users, obtaining comprehensive and precise data is impractical. However, the factors affecting the cooling load by the DCS itself are highly correlated with time. Therefore, we have comprehensively introduced the time point count factor t and the historical cooling load factor $L\text{-}n, n \in (1, 2, \cdots, 24)$.

In this study, data on cooling load and factors influencing the cooling load from July 1st to August 31st, 2022, were collected. After preprocessing this data, the Spearman rank correlation coefficient was utilized to perform a 24h correlation analysis on these load influencing factors, assessing the correlation level between each factor and the predicted cooling load L at the specific time point.

To avoid redundant analysis, this study focuses solely on the correlation analysis between T and L, as well as the selection of input parameters for the cooling load forecasting model related to T, as a case for investigation. **Figure 8** depicts the Spearman rank correlation coefficient between $T\text{-}n, n \in (0, 1, 2, \cdots, 24)$ and L, while **Figure 9** illustrates the correlation trend between $T\text{-}n, n \in (0, 1, 2, \cdots, 24)$ and L over a 24-hour period.

The correlation changes between non-adjacent factors exhibit corresponding characteristics. As indicated by **Figure 9**, with increasing time intervals, the variation trend of the correlation coefficient between T and L resembles a sine curve. Beyond a time interval of 16, the correlation reverts to positive, peaking at a correlation coefficient value of 0.57 when the time interval reaches 22.

Using an absolute value of the correlation coefficient between T and L set at 0.4 as the threshold, the values for n within set $T\text{-}n, n \in (0, 1, 2, \cdots, 24)$ should be 0, 1, 9, 10, 11, 19, 20, 21, 22, 23, and 24. From this, we deduce the input parameters for the cooling load forecasting model associated with T.

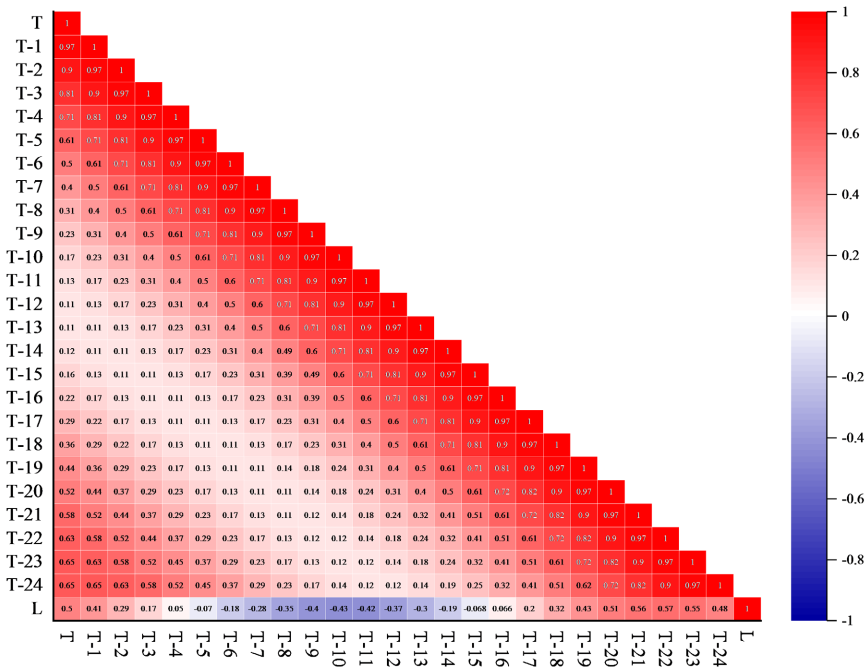

Figure 8. Plot of calculated Spearman rank correlation coefficients for T and L.

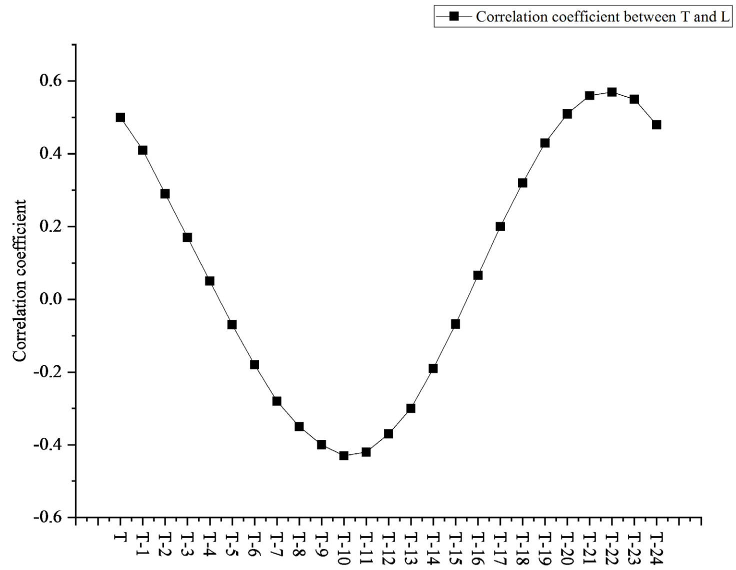

Figure 9. Trend curve of correlation between T-n and L over 24 hours.

Subsequently, we analyze the correlation between relative humidity φ-n, solar radiation intensity R-n, hourly precipitation W-n, atmospheric pressure P-n, atmospheric wind speed V-n, time point count t-n, and the historical cooling load L-n with the predicted cooling load L. This will allow us to preliminarily determine the input parameters for the cooling load forecasting model. The initially identified input parameters for the cooling load forecasting model total 44 items, as shown in **Table 2**.

3.3. Dimensionality Reduction of input Parameters

For machine learning-based cooling load forecasting model, having too many input parameters can increase computational complexity. In this study, after carefully considering various influencing factors, our focus was on selecting the parameters that have the most significant impact on the cooling load forecasting model to enhance both its accuracy and generalization capability. Among the 44 preliminary cooling load input parameters identified in **Table 2**, PCA was employed to perform a linearly independent transformation and dimensionality reduction of the inputs, ultimately determining the input parameters for the cooling load forecasting model.

Sequentially compute the mean and standard deviation for each input parameter. Then standardize the data to derive the correlation matrix R for the 44 input parameters. Based on R, determine its eigenvalues, resulting in a scree plot that illustrates the relationship between each parameter (principal component) and its corresponding eigenvalue. In the scree plot, the x-axis indicates the order of the eigenvalues (or principal components), while the y-axis represents the values of the eigenvalues, as illustrated in **Figure 10**. Alongside the eigenvalues, measures such as variance contribution rates and cumulative variance contribution rates are also obtained, detailed in **Table 3**.

Table 2. Preliminary determination of input parameters for the cooling load forecasting model.

Factor category	parameter variable	Selection of $\lvert r_s \rvert$	Correlation degree	parameters
Meteorological influencing factors	T	$0.4 \le \lvert r_s \rvert \le 0.57$	Significantly	*T*, *T*-1, *T*-9, T-10, *T*-11, *T*-19, *T*-20, *T*-21, *T*-22, *T*-23, *T*-24
	φ	$0.4 \le \lvert r_s \rvert \le 0.49$	Significantly	φ-19, φ-120 , φ-21 , φ-22 , φ-23
	R	$0.6 \le \lvert r_s \rvert \le 0.82$	Highly	*R*, *R*-1, *R*-2, *R*-3, *R*-11, *R*-12, *R*-13, *R*-14, *R*-15, *R*-23, *R*-24
	W	$\lvert r_s \rvert < 0.15$	Weakly	not take
	P	$\lvert r_s \rvert < 0.3$	Weakly	not take
	V	$\lvert r_s \rvert < 0.25$	Weakly	not take
Intrinsic factors of the DCS itself	t	$0.6 \le \lvert r_s \rvert \le 0.82$	Highly	*t*, *t*-7, *t*-8, *t*-9, *t*-19, *t*-20, *t*-21, *t*-22
	L	$0.6 \le \lvert r_s \rvert \le 0.93$	Highly & Extremely	*L*-1, *L*-2, *L*-3, *L*-11, *L*-12, *L*-13, *L*-22, *L*-23, *L*-24

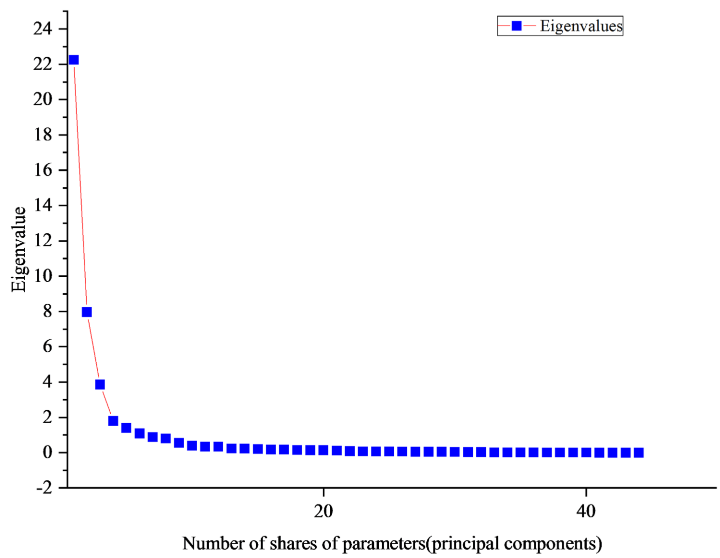

Figure 10. Scree plot for PCA of each parameter.

Table 3. The eigenvalues and variance contribution rates corresponding to the correlation matrix.

parameter (principal component)	eigenvalue	variance contribution rates	cumulative variance contribution rates
T	22.26533	50.60%	50.60%
T-1	7.96656	18.11%	68.71%
T-9	3.86014	8.77%	77.48%
T-10	1.79873	4.09%	81.57%
T-11	1.40627	3.20%	84.77%
T-19	1.09635	2.49%	87.26%

Continued

T-20	0.87887	2.00%	89.26%
T-21	0.80092	1.82%	91.08%
T-22	0.55145	1.25%	92.33%
T-23	0.40117	0.91%	93.24%
T-24	0.34111	0.78%	94.02%
φ-19	0.33078	0.75%	94.77%
φ-20	0.24468	0.56%	95.32%
φ-21	0.23031	0.52%	95.85%
φ-22	0.20151	0.46%	96.30%
φ-23	0.17887	0.41%	96.71%
R	0.16914	0.38%	97.10%
R-1	0.15062	0.34%	97.44%
R-2	0.14576	0.33%	97.77%
R-3	0.14377	0.33%	98.10%
R-11	0.11042	0.25%	98.35%
R-12	0.08601	0.20%	98.54%
R-13	0.07914	0.18%	98.72%
R-14	0.07382	0.17%	98.89%
R-15	0.06657	0.15%	99.04%
R-23	0.05882	0.13%	99.18%
R-24	0.055	0.13%	99.30%
t-6	0.04722	0.11%	99.41%
t-7	0.04507	0.10%	99.51%
t-8	0.04395	0.10%	99.61%
t-9	0.03442	0.08%	99.69%
t-19	0.03085	0.07%	99.76%
t-20	0.01936	0.04%	99.80%
t-21	0.01418	0.03%	99.83%
t-22	0.01287	0.03%	99.86%
L-1	0.01253	0.03%	99.89%
L-2	0.0113	0.03%	99.92%
L-3	0.00923	0.02%	99.94%
L-11	0.00784	0.02%	99.96%
L-12	0.00751	0.02%	99.97%
L-13	0.005	0.01%	99.99%
L-22	0.0034	0.01%	99.99%
L-23	0.00196	0.00%	100.00%
L-24	0.00117	0.00%	100.00%

For all parameters (principal components), one can obtain an eigenvector corresponding to each eigenvalue. The parameter loading matrix is a matrix composed of eigenvectors corresponding to the eigenvalues. In simpler terms, the loading vector is synonymous with the eigenvector and simultaneously represents the direction of the principal component. Mathematically speaking, the score of a principal component can be understood as the projection of a point onto the loading vector, representing the weights of each parameter point with respect to that principal component. The principal component is essentially a combination of these weights.

Firstly, corresponding to the 44 eigenvalues presented in **Table 3**, a biplot is generated as illustrated in **Figure 11**. The distances between points representing parameters (principal components) in the graph approximately signify the similarity between those parameters. The cosine value of the angle between the loading vectors approximately indicates the correlation between variables of the parameters. The projection of a point onto a vector approximately represents the interaction, or weight, between the parameter (principal component) and its variables. We obtain a 95% confidence ellipse, implying that with a probability as high as 95%, this is a relatively small three-dimensional ellipse for variables PC1(50.6%), PC2(18.1%), and PC3(8.8%). This indicates that the values of the parameters we have chosen closely approximate the true values, suggesting a high precision in estimating the sample parameters from the population parameters.

Secondly, as observed back from **Figure 10**, the scree plot starts to plateau after

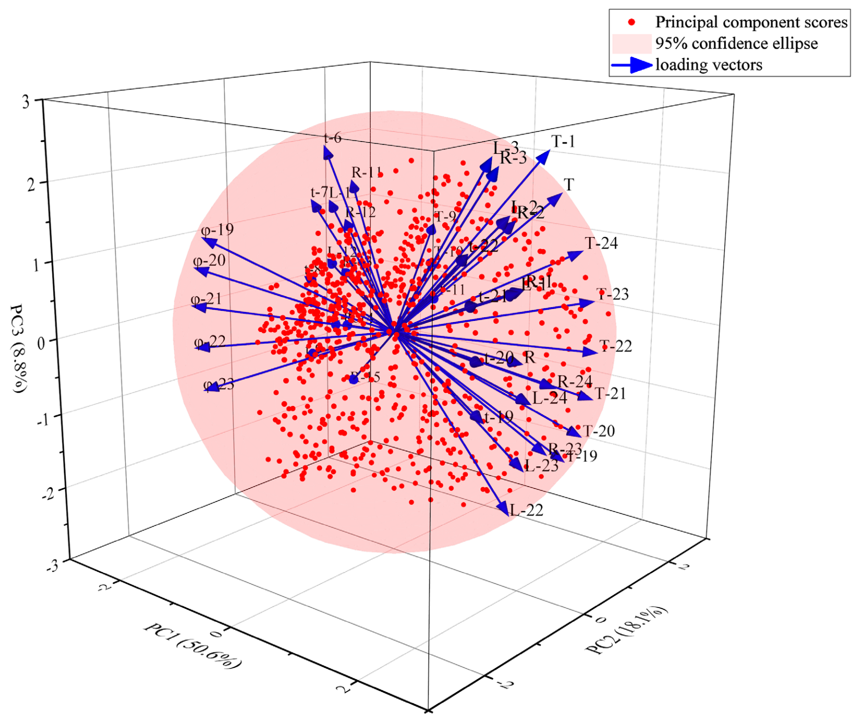

Figure 11. Biplot for PCA.

the sixth point, suggesting it can be theoretically disregarded. The criteria for selecting principal components are based on: identifying inflection points of eigenvalues, considering the magnitude of these eigenvalues, and evaluating the contribution rate of variance. It is evident that the sixth point in the plot signifies the inflection point of the eigenvalues. Referring to **Table 3**, this sixth point corresponds to T-19. Furthermore, if we focus solely on parameters with eigenvalues exceeding 1, it's evident that all eigenvalues are greater than 1 before the inflection point; post the inflection point, the eigenvalues corresponding to the eigenroots are all less than 1. Considering the cumulative variance contribution rate, when this rate surpasses 80%, it implies that the principal components essentially encapsulate all the information inherent in the parameters. The cumulative variance contribution rate corresponding to the sixth point is 87.26%. Consequently, in descending order of eigenvalues, this study selects the first six principal components from **Table 3** as input parameters for the cooling load forecasting model.

Lastly, let's denote the eigenvalues of the first six principal components as $\lambda_1 \ge \lambda_2 \ge \lambda_3 \ge \lambda_4 \ge \lambda_5 \ge \lambda_6$ corresponding to eigenvectors $e_1, e_2, e_3, e_4, e_5, e_6$, respectively. With this, dimensionality reduction can be carried out. Additionally, the first six principal components are denoted as $F_1, F_2, F_3, F_4, F_5, F_6$. Listing the 44 parameters from **Table 3** sequentially as $x_i (i = 1, 2, \cdots, 44)$, we conclusively determine the relationship between these 6 principal components and the selected 44 parameters. Essentially, they encapsulate the entirety of the information possessed by all parameters. These 6 principal components will serve as the input parameters for the cooling load prediction model, as illustrated by Equation (18).

$$F_n = \begin{pmatrix} x_1 & x_2 & \cdots & x_{44} \end{pmatrix} e_n, n \in (1, 2, \cdots, 6) \tag{18}$$

3.4. The Performance of the Cooling Load Forecasting Model

Using the LightGBM algorithm and based on operational data from June 1st to August 31st, 2022, as the sample set, the model was trained to forecast the hourly cooling load for the entire month of August 2022. **Figure 12** presents the hourly cooling load forecasting results over 31 days. For the majority of the time periods, the forecasted values of the hourly cooling load align closely with the actual measured values. Given that the end-users of the DCS are predominantly office buildings, there is a distinct difference in curve variations between weekdays and weekends.

We compared the forecasting performance of the LightGBM algorithm for cooling load with other algorithms mentioned in current relevant literature, specifically the GRNN, SVMR, and XGBoost. Using data from the entire month, we evaluated their forecasting capabilities using three commonly used metrics, the metrics are RAE(Relative Absolute Error), RRMSE(Relative Root Mean Square Error), and R-squared value (R^2), they are defined as [8] [26]:

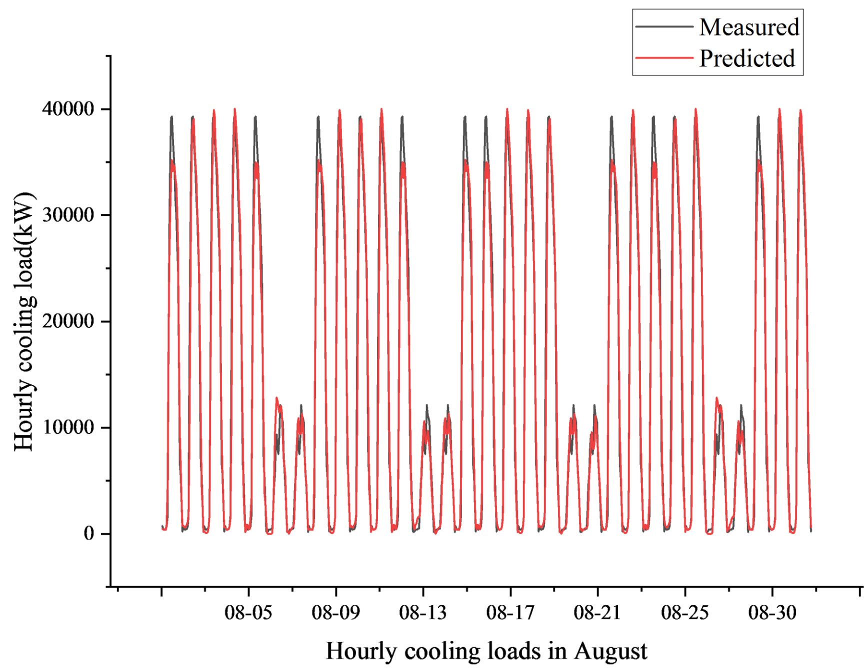

Figure 12. Visual comparison of predicted and measured hourly cooling loads.

$$\mathrm{RAE} = \frac{\frac{1}{n}\sum_{m=1}^{n}\left|y_m - y_{pre,m}\right|}{\frac{1}{n}\sum_{m=1}^{n} y_{pre,m}} \times 100\% \tag{19}$$

$$\mathrm{RRMSE} = 100 \times \frac{\sqrt{\frac{\sum_{m=1}^{n}\left(y_m - y_{pre,m}\right)^2}{n}}}{\frac{1}{n}\sum_{m=1}^{n} y_{pre,m}} \times 100\% \tag{20}$$

$$\mathrm{R}^2 = 100 \times \left(1 - \frac{\frac{\sum_{m=1}^{n}\left(y_m - y_{pre,m}\right)^2}{n}}{\frac{1}{n}\sum_{m=1}^{n}\left(y_{pre,m} - \frac{1}{n}\sum_{m=1}^{n} y_{pre,m}\right)^2}\right) \times 100\% \tag{21}$$

where y_m is the actual value, $y_{pre,m}$ is the predicted value, and *n* is the number of sample, $m \in [1, n]$.

RAE and RRMSE are used to measure the deviation between forecasted values and actual measurements, with lower values indicating better performance. R^2 quantifies the degree of correlation between forecasted values and measured values, with a higher value suggesting better correlation. **Figure 13** illustrates the performance of various load forecasting algorithms. It is evident that the LightGBM algorithm outperforms other forecasting methods. Specifically, its RAE is 3.62%, RRMSE is 42.75%, and R^2 is 92.96%. Additionally, it is observed that the SVM algorithm exhibits the least accuracy in the cooling load forecasting process.

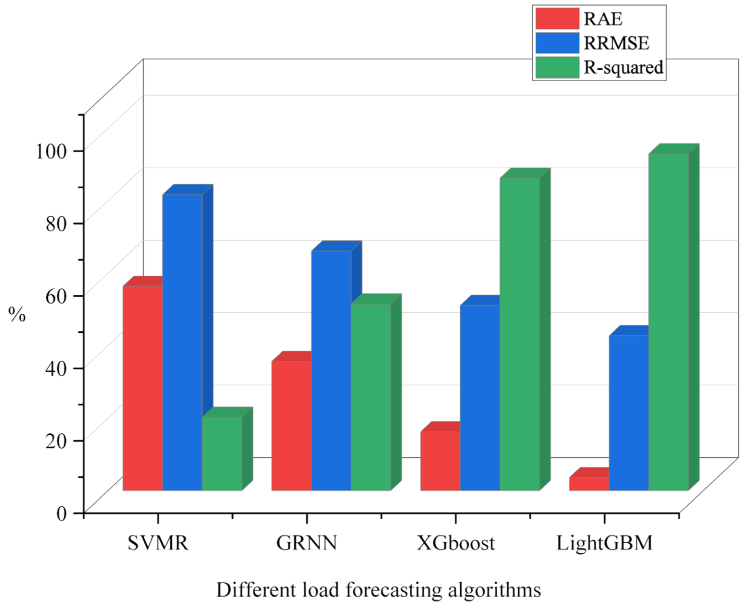

Figure 13. Forecasting performance of various algorithms.

4. Conclusions

Within a typical month, the hourly COP values for the duplex chiller units predominantly range between 3.5 and 4.4, with a monthly average COP value of 3.96. Both the hourly and monthly average COP values for the duplex chiller units closely align with their rated values. For the heat pump unit, the actual hourly COP values primarily lie between 4.3 and 5.4, resulting in a monthly average COP value of 4.73, indicating a significant deviation from its rated value. Throughout the typical month, the system's daily EER is primarily concentrated within the range of 2.23 to 3.11, with an average value of 2.79. Similarly, the CPOC is mainly situated within the range of 0.13 to 0.19 yuan/kWh, averaging at 0.16 yuan/kWh. Both average values are below the empirical benchmarks. If there is an opportunity to further enhance the system's EER, the CPOC for cooling capacity would consequently decrease.

Performing PCA and dimensionality reduction on the input parameters of the cooling load forecasting model, we derived principal components from the correlation matrix R (with variables standardized). Out of the 44 input parameters (principal components), 6 principal components were ultimately selected as the input parameters for the cooling load forecasting model. The eigenvalues of these six principal components are all greater than 1, and the cumulative variance contribution rate reaches 87.26%, essentially encapsulating the information contained in all parameters. During the dimensionality reduction process, we obtained a confidence ellipse with a 95% confidence interval. With a probability as high as 95%, this represents a smaller three-dimensional ellipse spanned by PC1 (50.6%), PC2 (18.1%), and PC3 (8.8%). This indicates that the parameter values we selected closely approximate the true values, suggesting a high precision in estimating population parameters based on sample parameters.

In this research, we initially delved into the correlation between the data and the system's performance to discern their interrelationships. Upon acquiring a comprehensive understanding of the system, we proceeded to investigate hourly cooling load forecasting using a data-driven approach. Among the commonly employed algorithms for cooling load forecasting, the lightGBM algorithm demonstrated superior performance. Specifically, the RAE value for the lightGBM algorithm was 3.62%, the RRMSE value stood at 42.75%, and the R^2 value reached 92.96%. By analyzing, organizing, and categorizing the data, as well as refining and optimizing the input parameters of the load forecasting model, we lay the groundwork for further research and insights. This paves the way for enhancing the operational efficiency of regional cooling systems, achieving cost savings, and elevating the comfort levels for terminal-users through various optimization measures.

Acknowledgements

The authors would like to thank the Chongqing Construction Science and Technology Plan Project of China (2019 No. 2-2-3, No. 2-2-4) for their financial support.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] IEA. World Energy Outlook 2022. http://www.iea.org

[2] B-Open/Aureliana Barghini. Copernicus Climate Change Service Application Documentation Global Temperature Trend Monitor 2021). http://www.copernicus.eu

[3] Angelidis, O., Ioannou, A., Friedrich D., Thomson, A. and Falcone, G. (2023) District Heating and Cooling Networks with Decentralised Energy Substations: Opportunities and Barriers for Holistic Energy System Decarbonisation. *Energy*, 269. https://doi.org/10.1016/j.energy.2023.126740

[4] bp. bp Energy Outlook 2023. http://www.bp.com

[5] Zhao, J. and Liu, X. (2018) A Hybrid Method of Dynamic Cooling and Heating Load Forecasting for Office Buildings Based on Artificial Intelligence and Regression Analysis. *Energ Buildings*, **174**, 293-308. https://doi.org/10.1016/j.enbuild.2018.06.050

[6] Gelazanskas, L. and Gamage, K.A.A. (2014) Demand Side Management in Smart Grid: A Review and Proposals for Future Direction. *Sustain Cities Soc*, **11**, 22-30. https://doi.org/10.1016/j.scs.2013.11.001

[7] Lei, Y., Wang, D., Jia, H., Chen, J., Li, J., Song, Y., *et al*. (2020) Multi-Objective Stochastic Expansion Planning Based on Multi-Dimensional Correlation Scenario Generation Method for Regional Integrated Energy system Integrated Renewable Energy. *Appl Energ*, 276. https://doi.org/10.1016/j.apenergy.2020.115395

[8] Luo, N., Hong, T.Z., Li, H., Jia, R.X. and Weng, W.G. (2017) Data Analytics and Optimization of an Ice-Based Energy Storage System for Commercial Buildings. *Appl*

Energ, **204**, 459-475. https://doi.org/10.1016/j.apenergy.2017.07.048

[9] Inayat, A. and Raza, M. (2019) District Cooling System via Renewable Energy Sources: A Review. *Renew Sust Energ Rev*, **107**, 360-373. https://doi.org/10.1016/j.rser.2019.03.023

[10] Vandermeulen, A., van der Heijde, B. and Helsen, L. (2018) Controlling District Heating and Cooling Networks to Unlock Flexibility: A Review. *Energy*, **151**, 103-115. https://doi.org/10.1016/j.energy.2018.03.034

[11] Sanaye, S. and Hekmatian, M. (2016) Ice Thermal Energy Storage (ITES) for Air-Conditioning Application in Full and Partial Load Operating Modes. *International Journal of Refrigeration-Revue Internationale Du Froid*, **66**, 181-197. https://doi.org/10.1016/j.ijrefrig.2015.10.014

[12] Sait, H.H. and Selim, A.M. (2014) Charging and Discharging Characteristics of Cool Thermal Energy Storage System with Horizontal Pipes Using Water as Phase Change Material. *Energy Conversion and Management*, **77**, 755-762. https://doi.org/10.1016/j.enconman.2013.10.034

[13] Ruan, Y.J., Liu, Q.R., Li, Z.W. and Wu, J.Z. (2016) Optimization and Analysis of Building Combined Cooling, Heating and Power (BCHP) Plants with Chilled Ice Thermal Storage System. *Appl Energ*, **179**, 738-754. https://doi.org/10.1016/j.apenergy.2016.07.009

[14] Kang, X., Wang, X., An, J. and Yan, D. (2022) A Novel Approach of Day-Ahead Cooling Load Prediction and Optimal Control for Ice-Based Thermal Energy Storage (TES) System in Commercial Buildings. *Energ Buildings*, 275. https://doi.org/10.1016/j.enbuild.2022.112478

[15] Hu, J., Zheng, W., Zhang, S., Li, H., Liu, Z., Zhang, G., *et al.* (2021) Thermal Load Prediction and Operation Optimization of Office Building with a Zone-Level Artificial Neural Network and Rule-Based Control. *Appl Energ*, 300. https://doi.org/10.1016/j.apenergy.2021.117429

[16] Wang, Z., Hong, T. and Piette, M.A. (2020) Building Thermal Load Prediction through Shallow Machine Learning and Deep Learning. *Appl Energ*, 263. https://doi.org/10.1016/j.apenergy.2020.114683

[17] Xu, Y., Li, F. and Asgari, A. (2022) Prediction and Optimization of Heating and Cooling Loads in a Residential Building Based on Multi-Layer Perceptron Neural Network and Different Optimization Algorithms. *Energy*, 240. https://doi.org/10.1016/j.energy.2021.122692

[18] Ben-Nakhi, A.E. and Mahmoud, M.A. (2004) Cooling Load Prediction for Buildings Using General Regression Neural Networks. *Energ Convers Manage*, **45**, 2127-2141. https://doi.org/10.1016/j.enconman.2003.10.009

[19] Cao, L., Li, Y., Zhang, J., Jiang, Y., Han, Y. and Wei, J. (2020) Electrical Load Prediction of Healthcare Buildings through Single and Ensemble Learning. *Energy Rep*, **6**, 2751-2767. https://doi.org/10.1016/j.egyr.2020.10.005

[20] Li, Q., Meng, Q., Cai, J., Yoshino, H. and Mochida, A. (2009) Predicting Hourly Cooling Load in the Building: A Comparison of Support Vector Machine and Different Artificial Neural Networks. *Energ Convers Manage*, **50**, 90-96. https://doi.org/10.1016/j.enconman.2008.08.033

[21] Cui, B., Gao, D.-C., Xiao, F. and Wang, S. (2017) Model-Based Optimal Design of Active Cool Thermal Energy Storage for Maximal Life-Cycle Cost Saving from Demand Management in Commercial Buildings. *Appl Energ*, **201**, 382-396. https://doi.org/10.1016/j.apenergy.2016.12.035

[22] Shan, K., Fan, C. and Wang, J.Y. (2019) Model Predictive Control for Thermal

Energy Storage Assisted Large Central Cooling Systems. *Energy*, **179**, 916-927. https://doi.org/10.1016/j.energy.2019.04.178

[23] Tang, H., Yu, J., Geng Y., Liu, X. and Lin, B. (2023) Optimization of Operational Strategy for Ice Thermal Energy Storage in a District Cooling System Based on Model Predictive Control. *J Energy Storage*, 62. https://doi.org/10.1016/j.est.2023.106872

[24] Cao, H., Lin, J. and Li, N. (2023) Optimal Control and Energy Efficiency Evaluation of District Ice Storage System. *Energy*, 276. https://doi.org/10.1016/j.energy.2023.127598

[25] Ma, X., Sha, J., Wang, D., Yu, Y., Yang, Q. and Niu, X. (2018) Study on a Prediction of P2P Network Loan Default Based on the Machine Learning LightGBM and XGboost Algorithms According to Different High Dimensional Data Cleaning. *Electronic Commerce Research and Applications*, **31**, 24-39. https://doi.org/10.1016/j.elerap.2018.08.002

[26] Heine, K., Tabares-Velasco, P.C. and Deru, M. (2021) Design and Dispatch Optimization of Packaged Ice Storage Systems within a Connected Community. *Appl Energ*, 298. https://doi.org/10.1016/j.apenergy.2021.117147