Energy & Buildings 218 (2020) 110055

Contents lists available at ScienceDirect

Energy & Buildings

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

Model-free control method based on reinforcement learning for building cooling water systems: Validation by measured data-based simulation

Shunian Qiu[a], Zhenhai Li[a], Zhengwei Li[a,b,*], Jiajie Li[a], Shengping Long[c], Xiaoping Li[d]

[a] School of Mechanical Engineering, Tongji University, Shanghai, China
[b] Key Laboratory of Performance Evolution and Control for Engineering Structures of Ministry of Education, Tongji University, Shanghai, China
[c] Shanghai East Low Carbon Technology Industry CO., LTD., Shanghai, China
[d] China Academy of Building Research, Beijing, China

ARTICLE INFO

Article history:
Received 30 December 2019
Revised 18 March 2020
Accepted 6 April 2020
Available online 11 April 2020

Keywords:
Cooling water system
Cooling tower
Cooling water pump
Optimal control
Reinforcement learning
Model-free control

ABSTRACT

In the domain of optimal control for building HVAC systems, the performance of model-based control has been widely investigated and validated. However, the performance of model-based control highly depends on an accurate system performance model and sufficient sensors, which are difficult to obtain for certain buildings. To tackle this problem, a model-free optimal control method based on reinforcement learning is proposed to control the building cooling water system. In the proposed method, the wet bulb temperature and system cooling load are taken as the states, the frequencies of fans and pumps are the actions, and the reward is the system COP (i.e., the comprehensive COP of chillers, cooling water pumps, and cooling towers). The proposed method is based on Q-learning. Validated with the measured data from a real central chilled water system, a three-month measured data-based simulation is conducted under the supervision of four types of controllers: basic controller, local feedback controller, model-based controller, and the proposed model-free controller. Compared with the basic controller, the model-free controller can conserve 11% of the system energy in the first applied cooling season, which is greater than that of the local feedback controller (7%) but less than that of the model-based controller (14%). Moreover, the energy saving rate of the model-free controller could reach 12% in the second applied cooling season, after which the energy saving rate gets stabilized. Although the energy conservation performance of the model-free controller is inferior to that of the model-based controller, the model-free controller requires less a priori knowledge and sensors, which makes it promising for application in buildings for which the lack of accurate system performance models or sensors is an obstacle. Moreover, the results suggest that for a central chilled water system with a designed peak cooling load close to 2000 kW, three months of learning during the cooling season is sufficient to develop a good model-free controller with an acceptable performance.

1. Introduction

The heating, ventilation and air-conditioning (HVAC) system consumes more than half of the total energy required for a building [1–4]. The cooling water system, also known as the condenser water system, is an essential subsystem of HVAC system. The cooling water system includes cooling water pumps, cooling towers, chiller condensers, and possibly water side economizers [5]. Cooling water systems are intended to discharge the heat rejected by the chillers. Operation of a cooling water system is essential to the chiller COP (coefficient of performance), which significantly influences the energy consumption of the entire HVAC system [5,6]. However, a cooling water system also uses energy during its operation. To reduce the energy consumption of HVAC systems, the trade-off between cooling water system energy and chiller energy should be considered, thus revealing the importance of cooling water system optimal control [7]. The objective of this study is to investigate a novel optimal control method for a cooling water system.

1.1. Optimal control of cooling water systems in buildings

Optimal control methods, can be classified into model-based control, model-free control, hybrid control and performance map-

* Corresponding author.
E-mail address: zhengwei_li@tongji.edu.cn (Z. Li).

https://doi.org/10.1016/j.enbuild.2020.110055
0378-7788/

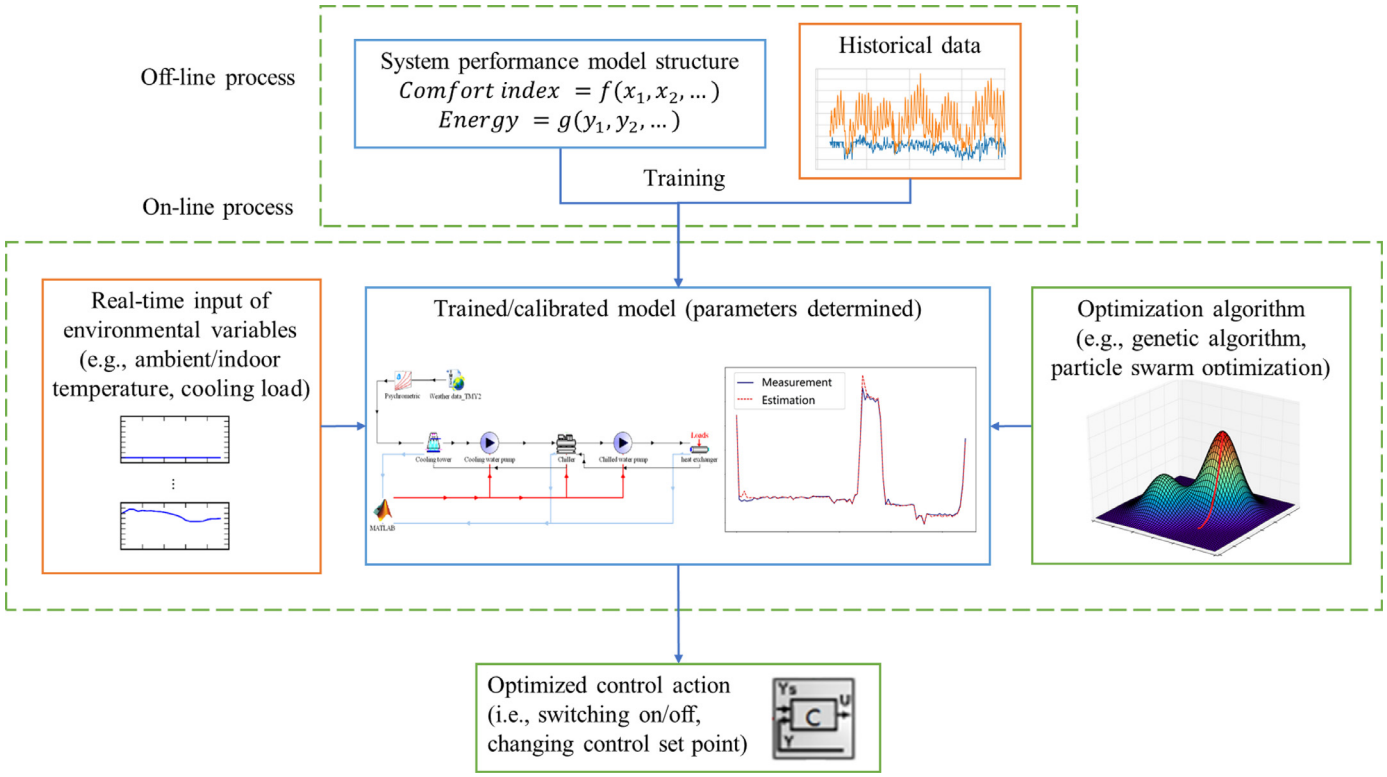


Fig. 1. Typical workflow of model-based control methods.

Nomenclature

Cooling water system variables, Units

T_{chws}	Chilled water temperature (outlet of chillers), $^{\circ}C$
T_{chwr}	Chilled water temperature (inlet of chillers), $^{\circ}C$
T_{cwr}	Cooling water temperature (inlet of chillers), $^{\circ}C$
T_{cws}	Cooling water temperature (outlet of chillers), $^{\circ}C$
F_{chw}	Chilled water flowrate, m^3/h
F_{cw}	Cooling water flowrate, m^3/h
$F_{cw,\,t}$	Cooling water flowrate through a cooling tower, m^3/h
$F_{cw,\,c}$	Cooling water flowrate through a chiller, m^3/h
C_p	Specific thermal capacity of water, kJ $\cdot$ kg^{-1} $\cdot$ K^{-1}
$T_{chws,\,set}$	T_{chws} set point of a chiller, $^{\circ}C$
T_{wet}	Ambient wet bulb temperature, $^{\circ}C$
f_{pump}	Pump operating frequency, Hz
f_{tower}	Cooling tower fan operating frequency, Hz
P	Electrical power, kW

Reinforcement learning variables

s	Last state (environment) of the learning agent
s'	Current state of the learning agent
a	Last action (operating frequency) taken by the learning agent
a'	A potential action that learning agent may execute in the current state
r	Reward to the agent for the last action at last state
Q	Q-value in Q-table

Accuracy indices

MAPE	Mean absolute percentage error
CV(RMSE)	Coefficient of variation of root-mean-square error

COP	Coefficient of performance
PLR	Partial load ratio
CC	Chiller cooling capacity, kW
CL	Cooling load, kW
RL	Reinforcement learning
AHU	Air handling unit
HVAC	Heating, ventilation and air-conditioning

based control [8]. Among these four optimal control methods, model-based control is most frequently investigated. As illustrated in Fig. 1, the typical workflow of model-based control consists of several steps: (1) Establish an accurate system model with historical data or other prepared information; (2) acquire real-time monitored data; (3) use the optimization algorithm to search the optimal control action based on the prediction of the system model.

Typical model-based control methods for building cooling water systems are reviewed as following. They are in accordance to the workflow in Fig. 1.

Ma and Wang [9] developed a fault-tolerant optimal control method for condenser cooling systems. The method is composed of a model-based optimal control process and an online fault detection scheme. This method adopted simplified equipment models to model the performance of chillers and cooling towers. With system cooling load CL_{system}, and the ambient wet-bulb temperature T_{wet} as the real-time environmental data, the optimal control actions (number of operating cooling towers, and the set point of T_{cwr}) are determined by a hybrid quick search method to maximize the overall system coefficient of performance (SCOP).

Yao et al. [10] built empirical performance models of the chillers, pumps and cooling towers with field test data to predict the system energy consumption under different operation conditions, including different chilled water flowrates F_{chw}, cooling water flowrates F_{cw} and T_{chws} values of the chillers. To enhance the system COP, these variables are optimized given the uncontrollable

environmental variables, namely, the ambient wet-bulb temperature T_{wet}, and T_{chwr} of each operating chiller. In other studies, Yao et al. [11,12] adopted state-space method to establish the model of HVAC system and refrigeration system. In doing so, each component of the system is modeled with matrixes, which could benefit the integration of component models.

Huang et al. [13] proposed a cooling tower control strategy for legacy chiller plants composed of multiple chillers, multiple cooling towers, primary chilled water pumps of constant speed, and condenser water pumps of constant speed. As a model predictive control method, this strategy adopted a novel model to predict chiller power and cooling tower power based on the predicted T_{wet} and system cooling load [14]. The set point of T_{cwr} is optimized to minimize the total power of chillers and cooling towers.

Wang et al. [15] proposed an event-driven optimization method for the building HVAC system. Unlike conventional optimization, which is triggered by a time point, this method is triggered by certain events (chiller on/off and chiller PLR change by 7%) within a predefined event space. The HVAC system model established by Wang [16] was adopted as the system performance model. The set points of four variables are optimized to reduce the system operation cost: the cooling water supply temperature (T_{cws}) from the chiller(s), the chilled water supply temperature (T_{chws}) from the chiller(s), T_{chw} from the heat exchanger(s), and the supply air (SA) temperature of the air handling units (AHU).

As for a model-based control method, the basic system model along with the parameters, and the real-time monitored data are usually defined as necessary preconditions for the applied system to provide (typically from manuals or historical data). Hence, the applicability of a model-based control method highly depends on the difficulty satisfying these requirements. These two essential preconditions of the selected studies are listed in Table 1.

The defects of model-based control are discussed in Section 1.2. And model-free control studies are reviewed in Section 1.3, together with an introduction to reinforcement learning.

1.2. Defects of model-based control

As introduced in Section 1.1, the energy conservation performance of model-based optimal control methods has been investigated in a number of studies. The defects of model-based control are also quite obvious:

1.2.1. Dependence of a priori knowledge

As listed in Table 1, most model-based control methods are based on multi-variate models, the parameters of which need to be determined by regression of field test data or historical operational data. Field test data requires manual labor and measurement instruments; historical data requires sufficient sensors implemented in the targeted system. These cost and requirements can affect the applicability of model-based control methods. Moreover, the number of required real-time monitored variables determines how many sensors must be implemented in the targeted system during the optimized operation, which influences the practicability of a control method, too.

1.2.2. Risk of the model uncertainty

Model uncertainties could strongly affect the performance of the model-based controller. Zhu et al. [23] argued that the uncertainties of model error could be classified into two types: model structure error and model parameter error. Model structure error is the result of model simplification. Model parameter error is mainly caused by faulty historical data; historical data is usually adopted in regression to determine the parameters/coefficients in the equipment model [24,25]. Moreover, even if the initial model

Table 1
Characteristics of selected model-based optimal control methods.

Reference	Equipment models	Definition of model parameters	Required real-time monitored data
[17–20]	COP-PLR model: $P_{chiller} = f(PLR,\ PLR^2, PLR^3)$	Equipment manual	System cooling load CL_{system}
[9]	Simplified physical model: $P_{chiller} = f(F_{chw}, F_{cw}, CL_{chiller})$ $F_{air} = f(UA, F_{cw}, T_{cws}, T_{wet})$ $P_{tower} = f(F_{air}^3,\ F_{air}^2,\ F_{air})$	Identification with historical data	CL_{system}, and the ambient wet-bulb temperature T_{wet}
[10]	$P_{chiller} = f(T_{condensing},\ T_{evaporating})$ Empirical model: $P_{pump} = f(rt^4, n^3, n^2, n)$ $T_{cwr} = f(T_{wet}^2, T_{wet})$	Regression with field test data	T_{chwr}, T_{wet}
[21]	Air-cooled chiller Empirical model: $P_{chiller} = f(T_{chwr}, T_{air}, F_{chw},\ PLR)$	Manufacturer's data	T_{chwr}, T_{air}, F_{chw}, CL_{system}
[15,16,22]	$P_{chiller} = f(T_{chws,set}, T_{cwr}, CL_{chiller})$ Simplified physical model: $F_{air} = f(UA, F_{cw}, T_{wet})$ $P_{tower} = 0.6F_{air}^3$	Regression based on historical operational data	Each chiller's T_{chwr}, T_{chws}, F_{chw}, T_{cwr} and on/off status
[13,14]	$P_{chiller} = f(T_{cwr}, \overrightarrow{S_{chiller}},\ \overrightarrow{CL_{system}^{pre}})$ Novel model: $P_{tower} = f(T_{wet}^{pre}, T_{cwr,set}, T_{cws}, \overrightarrow{S_{tower}})$ $T_{cwr} = f(T_{wet}^{pre}, T_{cwr,set}, T_{cws}, S_{tower})$	Regression based on historical data	Equipment operating status, water temperature in chiller condenser and evaporator, etc.

is perfectly accurate, its accuracy might decrease due to unavoidable system degradation.

Sun et al. [26] analyzed the calculation error of the chiller capacity resulting from model simplification and evaluated the accuracy of the calculation result using the confidence level. When the calculation result exceeds the confidence interval, it should be revised. Li et al. [27] investigated the model prediction error of the chiller cooling capacity and observed the chiller capacity calculation results from the simplified model and that the sophisticated model differed by approximately 4%.

1.3. Application of reinforcement learning (RL) in building system control

Reinforcement learning (RL), focuses on design of a learning agent that adapts its own actions based on an environmental reward to achieve a predefined goal (e.g., acquire the most reward) [28]. Actions taken by the agent depend on the agent's own experience accumulated during the game instead of a priori knowledge. In the building control field, research has been conducted to apply RL to the model-free control.

De Gracia et al. [29] developed a model-free control method based on RL techniques to optimize the open-close action of a phase change material (PCM) ventilated facade [30]. In this study, the temperature of the PCM is taken as the state of the RL agent, and the reward is the thermal energy obtained by the facade minus its own electrical energy consumption.

Valladares et al. [31] adopted deep reinforcement learning method to optimize the operation of AHUs and ventilation fans. In their study, a comprehensive objective function composed of four variables (Predicted mean vote (PMV) index, CO_2 concentration, AHU power and ventilation fan power) with four weight factors is proposed and used as the reward. Indoor temperature set point and the on-off signal of ventilation fan are selected as control actions. And the state of the system contains a group of quantities including indoor temperature, ambient temperature, CO_2 concentration, PMV index, ambient humidity, average radiant temperature and occupancy amount in the controlled room. After the training with a ten-year dataset, the RL agent could save 4–5% of the system energy with accepted PMV values and CO_2 concentration.

Zou et al. [32] implemented deep deterministic policy gradient to optimize the operation of the AHU. Two virtual environments were established using long-short-term-memory (LSTM) networks based on two-year measured operational data to approximate the real HVAC operations. The LSTM built with the first-year data was used to train the RL agents, and the other LSTM built with the second-year data was used to test the control performance of the trained agents. In their study, the state is defined with a combination of multiple parameters including environmental parameters and system operation parameters; the reward is composed of predicted percentage of discomfort (PPD) and system energy consumption; and the action is defined as the combination of four sub-actions: damper position, heating valve status, fan speed, and liquid solenoid status. The test results indicate that the weights of PPD and energy consumption could influence the control performance of the RL agents. After proper tuning of weight factors, the agents could save 27–30% of AHU energy comparing to the actual energy consumption, while maintaining the PPD at 10%.

Henze et al. [33] adopted the Q-learning method to optimize the charge-discharge action of energy storage equipment and investigated methods for choosing the state variables (i.e., state-of-charge, cooling loads, and real-time pricing rates) when the reward variable varies from the time-of-use utility rates to real-time pricing utility rates. In that study, the controller agent was established based only on a statistical summary of plant operation, without any prediction or system models.

1.4. Motivation and structure of this research

As discussed in Section 1.2, the greatest defect in model-based control is that it highly depends on accurate system models and sufficient sensors, which are difficult to provide for certain buildings. To tackle the optimal control problem in these buildings, a model-free control method based on RL techniques is proposed in this paper for the control of building cooling water systems.

Section 2 presents the control methodology. Section 3 discusses a measured-data based simulation case study that compares the energy-conservation performance of the model-based control method, basic control method, local feedback control method and the proposed model-free control method. The simulation results are illustrated and analyzed in Sections 4, and 5 concludes this paper.

2. Methodology

2.1. Overview

Applied RL algorithm: Q-learning, which is a classical RL method based on the Q-table. The mechanism of Q-learning makes it easy to converge and easy to be realized in programming [28]. Compared with deep reinforcement learning techniques which are usually based on networks, Q-learning is more feasible in engineering practice. In the HVAC domain, studies have been carried out to investigate the performance of Q-learning in controlling energy storage equipment [33–35], lighting [36],and ventilation systems [37].

2.1.1. Applied condition

The proposed method is for a cooling water system composed of identical units (i.e., identical chillers, identical cooling towers, and identical cooling water pumps). Also, the chilled water system corresponding to the targeted cooling water system should be a decoupled system (constant primary chilled water flowrate and variable secondary chilled water flowrate) [38]. An example of the investigated system's layout is illustrated in Fig. 3.

2.2.2. Optimization objective

The optimization objective is the system COP, which is calculated by Eq. (1).

$$System\ COP = \frac{CL_{system}}{P_{chillers} + P_{cwps} + P_{towers}} \tag{1}$$

where CL_{system} is the system cooling load (kW), $P_{chillers}$ is the total power of all chillers (kW), P_{cwps} is the total power of all cooling water pumps (kW), and P_{towers} is the total power of all cooling towers (kW). The total energy consumption of the chillers, cooling water pumps and cooling towers is referred to as the "system energy consumption" in this paper.

2.2.3. Requirements on prepared knowledge and real-time data inputs

The proposed method needs prepared information prior to application, i.e., system layout, history weather data and equipment characteristics on name plates. No models needed. And the required real-time data includes system cooling load, ambient wet bulb temperature, and equipment power.

2.2.4. Optimized variable (control action)

Operating frequencies of cooling tower fans and cooling water pumps.

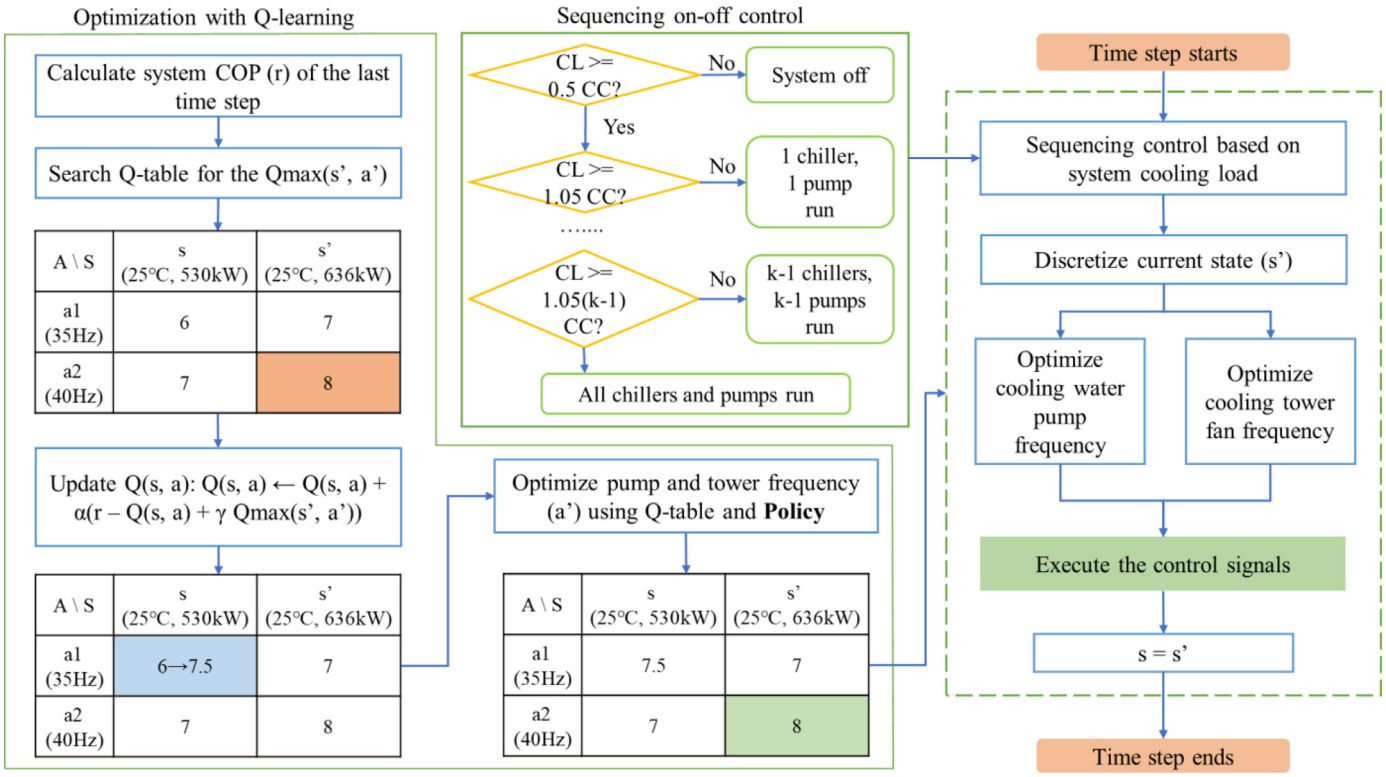


Fig. 2. Workflow of the proposed method (the states, actions and Q values in this figure are merely shown as an example).

2.2.5. Optimization interval

Additionally, the optimization interval should be set between twenty to sixty minutes when applying the proposed method because of two reasons: (1) the action taken by an untrained controller is stochastic, thus shorter optimization interval may result in oscillating behavior; (2) the RL-based controller needs accurate environmental feedback (i.e., the reward) to evaluate the last taken action, and it takes time for the system to stabilize after each control action.

As shown in Fig. 2, the workflow of the proposed control method is composed of several steps:

A At the beginning of each time step, the amount of equipment that must operate is determined using the sequencing on-off control method, which is discussed in Section 2.2.
B The real-time measured data (wet bulb temperature and system cooling load) of the cooling water system are discretized to the state (s') to match the structure of Q-tables. This step is demonstrated in Section 2.3.
C The frequency of the cooling tower fans is optimized using the Q-learning method: (1) Search the tower Q-table for the maximum Q-value in the current state (s'); (2) update the Q-value of the last state (s) using updating formula (Eq. (2)) with the system COP (Eq. (1)); and (3) determine the optimal frequencies for the cooling tower fans using a certain optimization policy and the updated tower Q-table. Details are given in Section 2.4.
D Optimize the frequency of running cooling water pumps in the same manner as in Step C. Note, Steps C and D are in parallel, the order of these two steps does not matter.
E Execute the on-off control signals and optimal frequency control signals on the system.
F Record the current state (s') as s, because the current time step is ending, and the current state will be regarded as the last state in the next time step.

2.2. Sequencing on-off control

As introduced in Section 2.1, in the beginning of each time step, the on-off status of each equipment should be controlled before the optimization of frequencies. The following rules are adopted to determine proper on-off signals.

(1) The entire central chilled water system remains in the off condition if the system cooling load is less than 50% of one chiller's rated cooling capacity [17–19,38,39].
(2) While the cooling system is operating, all cooling towers are run to optimize the energy consumption tradeoff between the chillers and cooling towers, according to the studies of Braun and Hartman [7,40].
(3) The number of chillers required is determined based on the system cooling load. When the system cooling load exceeds 105% of the current system cooling capacity, one additional chiller is switched on [27,41,42]. The "k" in Fig. 2 represents the total number of chillers in the applied system.
(4) The on-off status of the cooling water pumps and primary chilled water pumps are coordinated to the status of the chillers. In other words, when chiller 1 is switched on/off, cooling water pump 1 and primary chilled water pump 1 are switched on/off [43].
(5) The frequency of each operating primary chilled water pump is maintained at a nominal value because the targeted system is a decoupled system, the primary chilled water flowrate of which remains constant to protect the chillers [18,44].
(6) For each running chiller, the set point of chilled water supply temperature ($T_{chws,\ set}$) is held equal to its nominal T_{chws}.

2.3. Configuration and initialization of two Q-tables

In this method, the frequencies of cooling water pumps and cooling tower fans are optimized by two RL agents using the Q-

Table 2
Format of the Q-table.

A\S	$T_{wet,\ 1}$, 0.5 CC	$T_{wet,\ 1}$, 0.6 CC	.....	$T_{wet,\ l}$, (k-0.1) CC	$T_{wet,\ l}$, k CC
f_1	Q (s_1, a_1)	Q (s_2, a_1)		Q (s_{n-1}, a_1)	Q (s_n, a_1)
f_2	Q (s_1, a_1)	Q (s_2, a_2)		Q (s_{n-1}, a_2)	Q (s_n, a_2)
			.....		
f_m	Q (s_1, a_m)	Q (s_2, a_m)	.....	Q (s_{n-1}, a_m)	Q (s_n, a_m)

learning method. And the tower Q-table and pump Q-table should be configured and initialized in the following way.

2.3.1. State

The state in this study is defined as the combination of the discretized ambient wet bulb temperature (T_{wet}) and the discretized system cooling load (CL_{system}) because (1) these two variables are not influenced by the system operation; (2) the system cooling load is an essential variable in the operation of the central chilled water system [17,18,20,27,41,45,46]; (3) T_{wet} could significantly influence the cooling capacity of cooling towers, which is important to the performance of the entire central chilled water system [7,10,47,48].

These two continuous variables are discretized as follows: (1) T_{wet} is discretized to an integer (e.g., 24 °C, 25 °C), and the range of discretized T_{wet} (i.e., 24–28 °C, 23–29 °C, or else) should be specified according to the historical weather data of the city in which the applied system is located; (2) The system cooling load should be discretized according to the cooling capacity (CC) of each chiller of the applied system, and if the system includes k chillers altogether, then the system cooling load should be discretized to 0.5, 0.6, ..., k CC (e.g., when the measured cooling load is 0.57 CC, then it should be discretized to 0.6 CC); (3) the real time values of these two discretized variables are taken as the real time state, e.g., (24 °C, 530 kW). All combinations of these two discretized variables are taken as the state space (i.e., Q-table columns).

2.3.2. Reward

The system COP calculated by Eq. (1) is regarded as the reward, same as the optimization objective.

2.3.3. Q-value

The meaning of the Q-value in this method is the system COP, in accordance with the optimization objective. The Q-values should be initialized 15–20% higher than the nominal system COP to encourage the RL agents to search for optimal control actions.

2.3.3. Action

The frequency set point of a pump or a cooling tower fan is taken as the action in this method. The values of the actions should be limited within a reasonable range considering capacity allowance, according to the system manager and equipment manual, to protect the hardware devices (e.g., the frequency of a variable-speed pump is typically limited above 20 Hz). The precision of frequency is 1 Hz in this method.

The abovementioned discretization precisions of the states and actions are defined with consideration of the exploration cost. Higher precision results in a larger Q-table (i.e., larger action space and larger state space), which could benefit the control performance of a well-trained model-free controller, because (1) the identification of the state would be more accurate; (2) the control action would be more precise. But a larger Q-table also requires a longer period and more data to train the controller by updating Q values in the table [28].

Table 2 is an example of the Q-table in this study, where l is the number of possible T_{wet} values, m is the number of available actions, and n is the number of states. Two RL agents optimize the cooling water pumps and cooling towers respectively and each agent uses one Q-table.

2.4. Decision-making and Q-table updating

During the "game", a Q-learning agent must accumulate experience by updating the Q-table. The updating principle of the Q-learning agents is described by Eq. (2) [28]:

$$Q(s,\ a)\ \leftarrow\ Q(s,\ a)+\alpha\left[r+\gamma\max_{a'}Q\left(s',a'\right)-Q(s,\ a)\right] \tag{2}$$

where $Q(s,a)$ is the Q-value corresponding to the last state (s) and last action (a), r is the reward resulting from action (a), α is the learning rate (which is defined 0.9 in this study to accelerate the agents' learning), and γ represents the impact of future reward on the decision of the current action. In this study, γ is set to 0.01 because the agent action does not affect the next state, which means that in every time step, the pump agent and tower agent only need to focus on how to maximize the current reward instead of the total reward over the long term. $\max_{a'}Q(s',a')$ is the max Q-value at state (s') according to the current Q-table. In the proposed method, the agents' Q-tables should be updated for the entire system life to cope with continuing system degradation [23].

In every time step, the agents must determine the next action based on the Q-table and a certain policy. In this study, a modified version of the ε-greedy policy [28] is developed for the agents to determine the next actions. The original ε-greedy policy is described by Eq. (3):

$$\pi(a|s)=\begin{cases}1-\varepsilon+\dfrac{\varepsilon}{m} & if\ a=arg\max_a Q(s,a)\\ \dfrac{\varepsilon}{m} & if\ a\neq arg\max_a Q(s,a)\end{cases} \tag{3}$$

where ε is the predefined parameter that determines the balance between exploration and exploitation, m is the number of practical actions at state (s), and $\pi(a|s)$ is the probability that a certain action is chosen at state (s). Eq. (3) means that in each time step, the probability that an agent chooses the known best action is $1-\varepsilon+\frac{\varepsilon}{m}$, and the probability that the agent chooses another action is $\frac{\varepsilon}{m}$.

In the ε-greedy policy, the ε value does not change with the passage of time, which means that the balance between exploration and exploitation is never changed during the game [28]. To improve this mechanism, a modified version of the ε-greedy policy is developed herein. The modified version of the ε-greedy policy in this study is described by Eq. (4):

$$\pi(a|s)=\begin{cases}\dfrac{10\times\frac{q}{p}}{1+10\times\frac{q}{p}}+\dfrac{Q(s,\ a)}{\sum Q(s,\ a)}\times\dfrac{1}{1+10\times\frac{q}{p}} & if\ a=arg\max_a Q(s,a)\\ \dfrac{Q(s,\ a)}{\sum Q(s,\ a)}\times\dfrac{1}{1+10\times\frac{q}{p}} & if\ a\neq arg\max_a Q(s,a)\end{cases} \tag{4}$$

where q is the number of passed time steps, and p is a predefined parameter representing the initial period of agents' learning, which should be defined based on the length of a cooling season and the optimization interval of the applied system. For instance, for a system which is (1) used ten hours every day; (2) optimized hourly; (3) located in a city with four-month long cooling seasons, it is recommended to set p by 4 $month$ × 30 day × 10 $hour$ × 1 $optimization\ per\ hour$ = 1200. In this study, p is defined as 2208 because the length of the case study is three months, and the case system operates 24/7 with hourly optimization (24×92×1 = 2208). Eq. (4) is explained in the following way.

At the start of system operation, the agent lacks experience. In this situation, the probability that an action is selected is

Table 3
Cooling water system characteristics (nominal system COP = 5.83).

Equipment	Number	Characteristics
Screw chiller	2	Nominal: COP = 6.639, cooling capacity = 1060 kW, power = 159.7 kW, variable speed, cooling water flow rate = 195 m³/h, chilled water flow rate = 131 m³/h, cooling water temperature = 30.5/35.5 °C, chilled water temperature = 10/17 °C
Chilled water pump	2	Nominal: power = 12.0 kW, flowrate = 150 m³/h, head: 24 m, variable speed
Cooling water pump	2	Nominal: power = 14.7 kW, flowrate = 240 m³/h, head: 20 m, variable speed
Cooling tower	2	Nominal: power = 7.5 kW, flowrate = 260 m³/h, variable speed

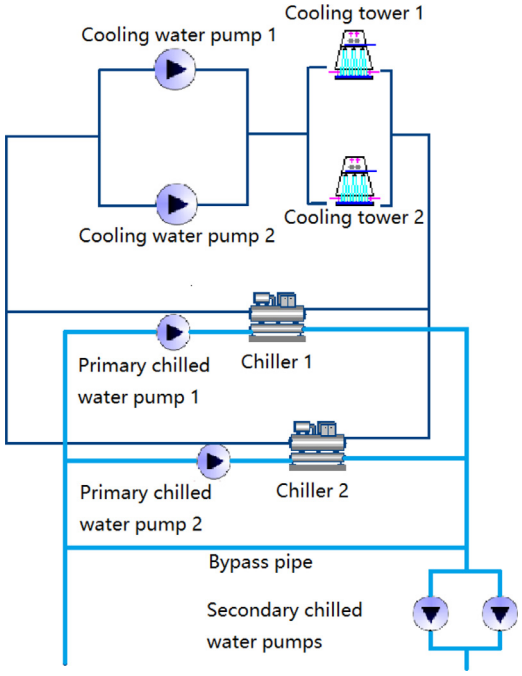


Fig. 3. Layout of the central chilled-water system.

Fig. 4. Photo of the real system (Chiller 2).

nearly proportional to its Q-value. And there is a small bonus (bonus $= \frac{10\times\frac{q}{p}}{1+10\times\frac{q}{p}}$) on the probability that the optimal action is selected. In other word, exploration is encouraged.

As time passes (q increases), the agent becomes more willing to choose the action with the maximum Q-value (bonus increased) while other actions can still be chosen, and the probability of each nonoptimal action to be selected is proportional to its Q-value. In brief, exploitation is more heavily encouraged as time passes.

When q reaches p, the bonus of the optimal action is $\frac{10}{11}$, which means that the probability that the optimal action is selected is approximately 10 times the summed probability of the other actions. Meanwhile, the initial learning process is finished. The exploration of the agent will continue after the initial learning period, with a continuous reduction of exploring probability (i.e., continuous increase of the bonus).

3. Measured data-based simulation case study

3.1. Case system

A real HVAC system in a metro station in Guangzhou city is adopted as the case system. The equipment characteristics provided by the manufacturer are listed in Table 3. The two chillers are identical, as are the other three types of appliances. The layout of the case system is illustrated in Figs. 3, and 4 show a photo of the real system. The chilled water system in the case is a decoupled system, as introduced in Section 2.1. Measured weather data (Fig. 5) and measured system cooling load data (Fig. 6) are adopted as the simulation input. Measured operational data (Fig. 7) of the real system is used to establish the simulation model.

The control strategy of the real case system includes: (1) the temperature of supplied chilled water at the header pipe is controlled at 10 °C. The number of operating chillers is adjusted manually by the management engineer based on weather and passenger flow; (2) the number of running cooling water pumps and chilled water pumps is equal to the number of operating chillers; (3) the frequencies of operating cooling water pumps are adjusted equally to keep the ΔT_{cw} (the difference between T_{cwr} and T_{cws}) at 3.3 °C (the frequency is restrained within 35–50 Hz); (4) the frequency of the cooling tower fan is adjusted to maintain the approach (the difference between T_{cwr} and T_{wet}) at 2.5 °C (the frequency is restrained within 30–50 Hz); (5) the number of operating cooling towers is adjusted manually by the management engineer, most time both cooling towers operates simultaneously; (6) primary chilled water pumps typically operate at nominal frequency. Note, the set point values (3.3 °C and 2.5 °C) are determined empirically by the management engineer of the real case system.

The control strategy above is also reflected by the measured data in Fig. 7: (1) ΔT_{cw} does not change much with time or system working condition; (2) T_{chws} is maintained at approximately 10 °C; (3) F_{chw} does not evidently vary with time, while F_{cw} is on the contrary; (4) T_{cwr} is basically stable at 29 °C because T_{wet} is quite stable during the investigated period (Fig. 5)

As is shown in Figs. 6(b) and 7, the measured data contains missing values and measurement faults. Hence the measured data needs pre-processing before being used for the model setup. Fig. 6(b) shows the distribution of the missing data in the cooling load measurement. In this case study, missing data of cooling load is filled with interpolation. Specially, when both sides of a missing series are zero values, this missing series is filled with zeros; in other words, the system is considered off during this period.

And for the operational data used for model establishment (Section 3.2), the pre-processing is realized in three steps: (1) abandon data items including missing values; (2) select data items

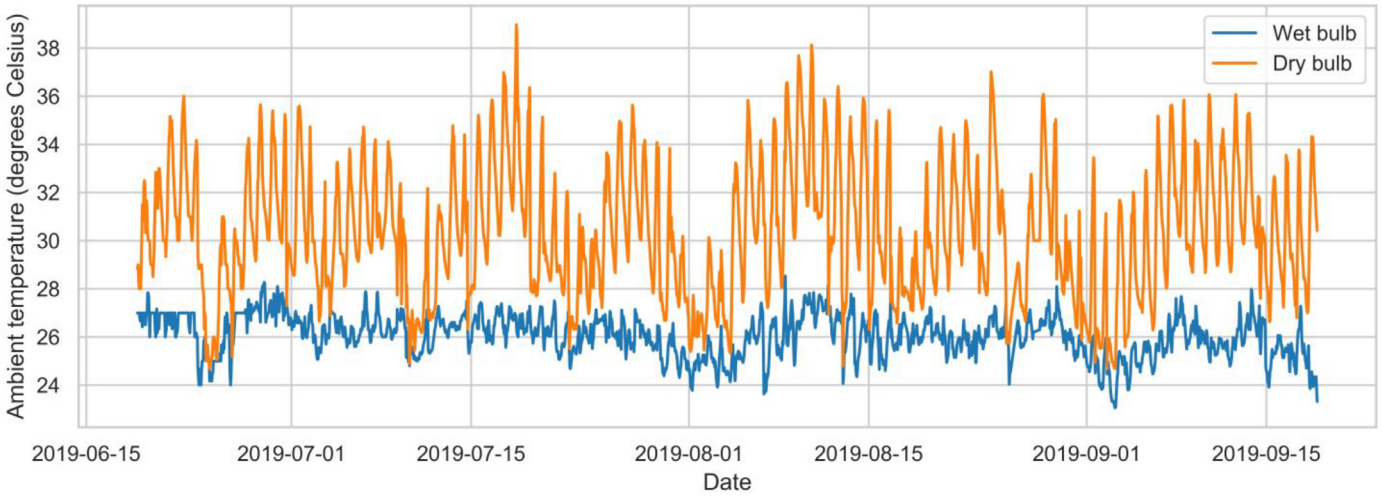


Fig. 5. Measured weather data (hourly), (a). Measured data after interpolation, (b). Missing data in the cooling load measurement (blanks represent missing data).

Table 4
Thresholds for the pre-processing of training data.

Equipment	Thresholds
Chiller	$80 \le F_{chw} \le 160, and\, 3 \le \mathrm{COP} \le 10, and\, F_{cw,\,c} \ge 150$
Cooling water pump	$35 \le f_{pump} \le 50, and\, 50 \le F_{cw}$
Cooling tower	$30 \le f_{tower} \le 50, and\, 50 \le F_{cw,\,t} \le 260$

with thresholds defined in Table 4 for each equipment to be modeled. This step is intended to drop the data which may be affected by measurement faults or unstable operation; (3) use Hoteling's T-square test (95% confidence level) to remove the extreme outliers [49].

3.2. Simulation system model based on measured data

The hourly simulation case study is conducted on the Python 3.6 platform, and the simulation system model is built according to the layout and characteristics of the real case system. The simulation process Fig. 8) imitates TRNSYS [50]. In each time step, the system receives inputs and begins calculation, the system variable values are updated by one equipment model, and the values are subsequently circulated in the entire system, from equipment to equipment. The iteration does not stop until the variable values converge. In this study, Eqs. (8)–((11) compose the chiller model; Eqs. (6) and (7) compose the pump model; and Eqs. (5) and (6) compose the cooling tower model. As Fig. 8 shows, the variable values are circulated in these three equipment models until T_{cwr} converges (i.e., the difference of T_{cwr} between current iteration circle and the last iteration circle is less than 0.2 °C). If the T_{cwr} does not converge within 50 iteration circles, then the iteration will be stopped and the result of the last circle will be adopted.

In detail, the chiller COP and cooling tower outlet water temperature (T_{cwr}) are simulated by random forest, a classical regression model proposed by Breiman [51–53]. In this study, random forest regressors are trained and validated by the measured operational data (from 19th June to 18th September) of the real system. The power of the pumps and fans is modeled with the conventional frequency-power formula Eq. (6)). The coefficients in Eq. (6) are determined by regression with measured data. Eqs. (9)–((11) are conventional equations used to calculate the other intermediate variables. For Eq. (7), the flowrate through a cooling water pump is calculated simply with similarity because the pipeline resistance does not change substantially in a cooling water system.

$$T_{cwr} = Random\ forest\ regressor\ (T_{cws},\ f_{tower},\ T_{wet},\ F_{cw,t}) \tag{5}$$

$$P = af^2 + bf + c \tag{6}$$

$$F_{cw} = \frac{f_{pump}}{f_{nominal}} \times F_{nominal} \tag{7}$$

$$COP_{chiller} = Random\ forest\ regressor\ (CL,\ T_{cwr},\ T_{chwr},\ T_{chws},\ F_{cw,c}) \tag{8}$$

$$P_{chiller} = \frac{CL}{COP_{chiller}} \tag{9}$$

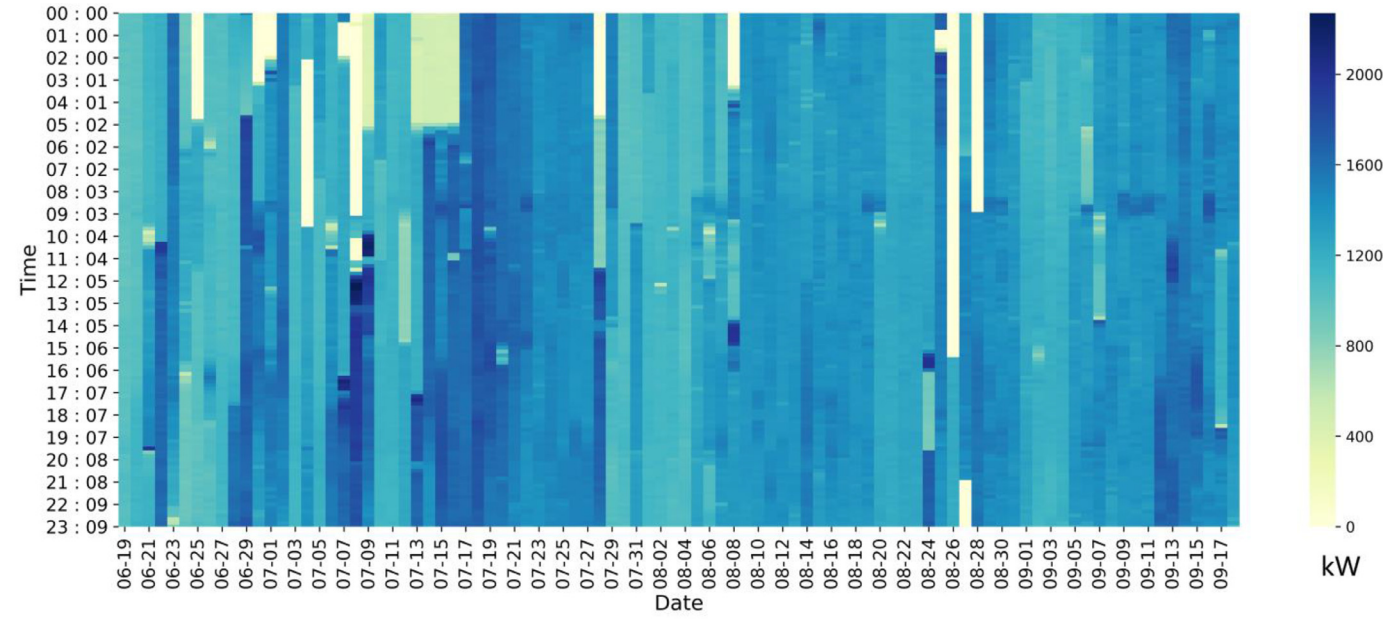
$$T_{cws} = T_{cwr} + (P_{chiller} + CL) \div \left(\frac{C_p \times F_{cw,c} \times \rho}{3600s/h}\right) \tag{10}$$

$$\begin{cases} T_{chws} = max\left[T_{chws,\ set},\ T'_{chwr} - CC/\left(\dfrac{C_p \times F_{chw} \times \rho}{3600s/h}\right)\right] \\ T_{chwr} = T_{chws} + CL/\left(\dfrac{C_p \times F_{chw} \times \rho}{3600s/h}\right) \end{cases} \tag{11}$$

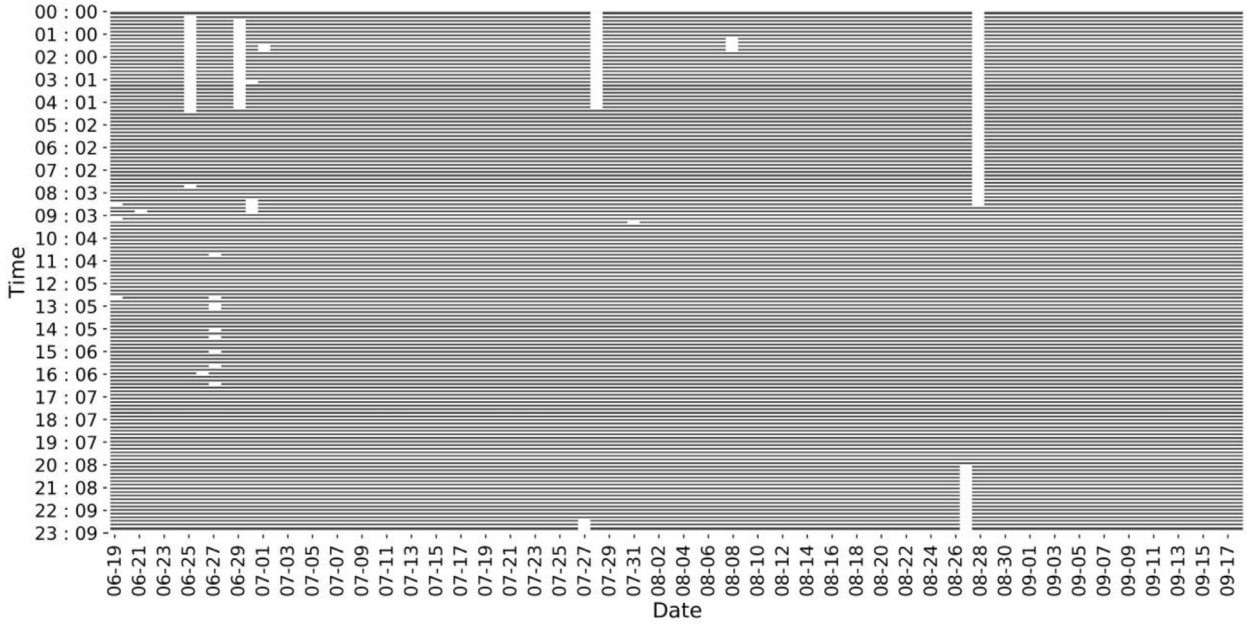
where CL is the cooling load on the chiller (kW), CC is the nominal cooling capacity of the chiller (kW), T_{cwr} is the temperature of the cooling water returning to the chillers from the cooling towers (°C), T_{chwr} is the temperature of the chilled water returning to the chillers (°C), T'_{chwr} is T_{chwr} of last time step, T_{chws} is the temperature of the chilled water leaving the chillers (°C), $F_{cw,\,c}$ is the cooling water flowrate through the chiller condenser (m³/h), $F_{cw,\,t}$ is the cooling water flowrate through a cooling tower (m³/h), C_p is the specific heat capacity of water = 4.2 kJ/(kg•K), ρ is the water density = 1000 kg/m³, F_{chw} is the chilled water flowrate through the chiller evaporator (m³/h), P is the electrical power (kW), f is the frequency (Hz), a, b, and c are coefficients to be determined, $F_{nominal}$ is the nominal flowrate of a cooling water pump (m³/h), F_{cw} is the flowrate through a cooling water pump (m³/h), and $P_{chiller}$ is the chiller power (kW).

The coefficient of variation of the root-mean-square error (CV(RMSE), Eq. (12)) and the mean absolute percentage error (MAPE, Eq. (13)) are adopted as the error indices to assess the accuracy of the abovementioned equipment models [54–56]. The accuracy of the models is illustrated in Fig. 9 and Table 5.

$$CV(RMSE) = \frac{\sqrt{n\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}}{\sum_{i=1}^{n} y_i} \tag{12}$$



(a). Measured data after interpolation



(b). Missing data in the cooling load measurement (blanks represent missing data)

Fig. 6. Measured system cooling load data (hourly, kW).

Table 5
Error index values of equipment models and the system model.

	Chiller COP model		T_{cwr} model of cooling tower		Cooling tower fan power model	Cooling water pump power model	Overall system model	
	Train set	Test set	Trains set	Test set			T_{cwr}	T_{cws}
MAPE	0.60%	1.75%	0.24%	0.67%	1.58%	2.69%	1.39%	1.43%
CV(RMSE)	0.90%	2.59%	0.36%	1.01%	2.31%	3.43%	2.71%	2.62%

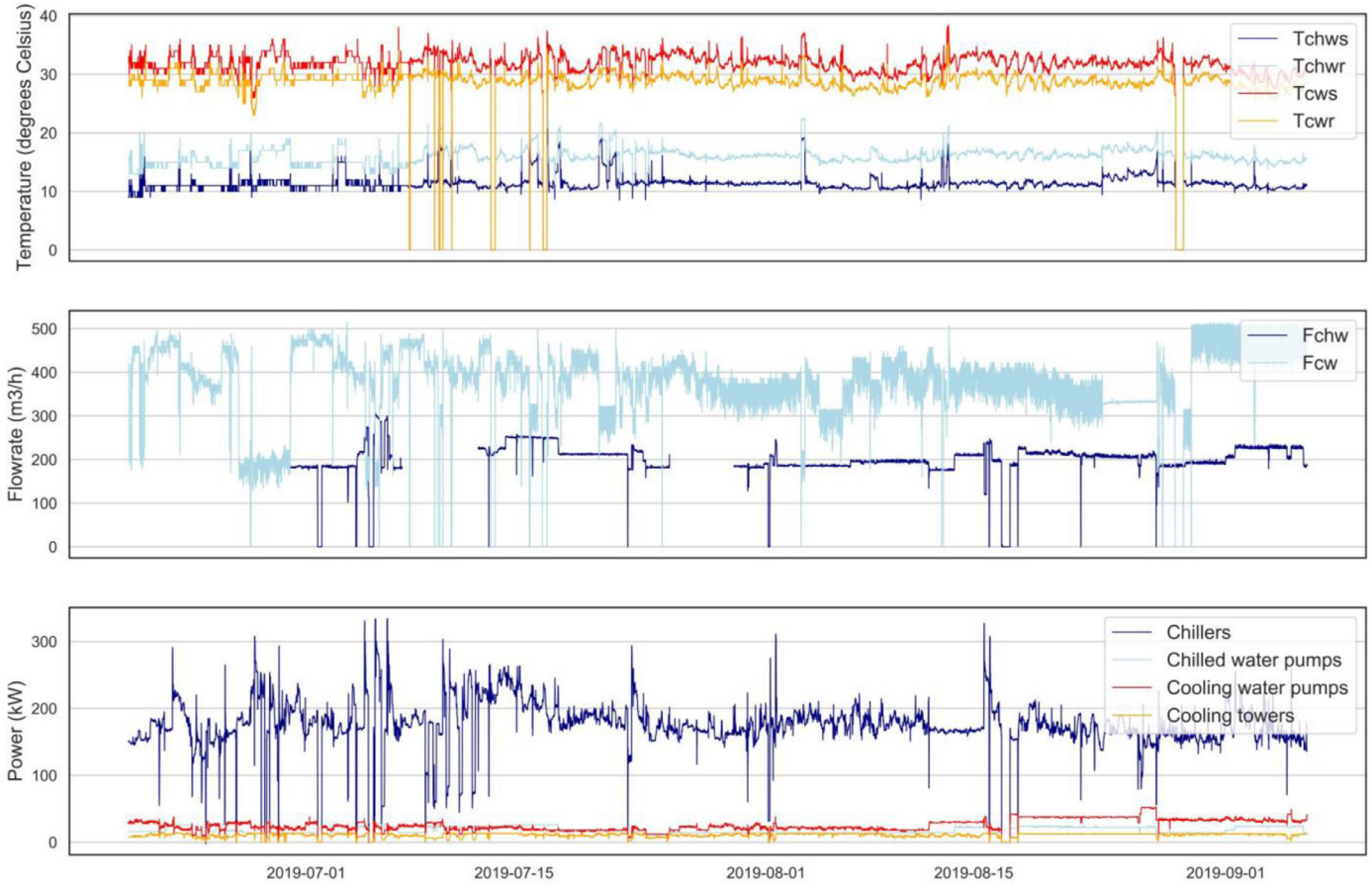


Fig. 7. Measured operational data of the real system, (all variables in this figure are of the header pipe, e.g., F_{cw} is the total flowrate of cooling water in the system).

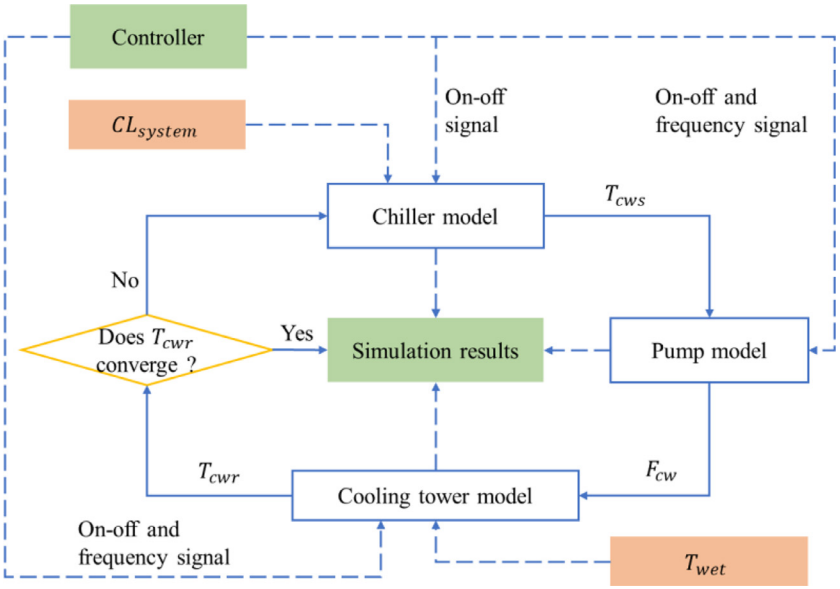
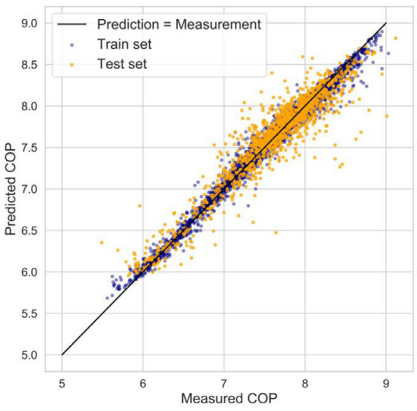


Fig. 8. Simulation process at one time step (solid lines represent the iteration loop of the simulation process, and dashed lines represent the input and output procedure of the iteration loop).

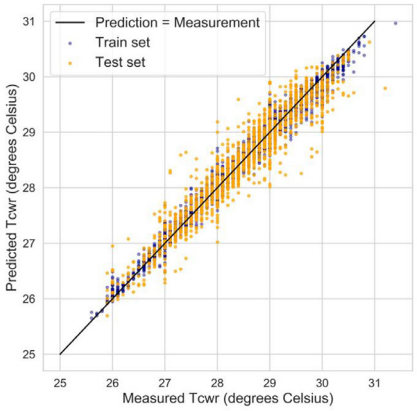
$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y}_i}{y_i}\right| \times 100\% \quad (13)$$

where n is the number of data points, y_i is the ith measured value, and $\widehat{y}_i$ is the ith predicted (simulated) value.

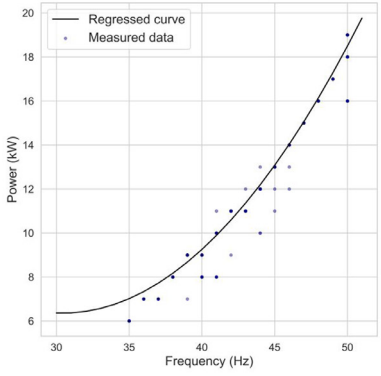
The error index values in Table 5 are all less than 5% [57], which verifies the accuracy of equipment models. It should be noted that the datasets of COP and T_{cwr} models are split randomly and independently, into a training set (80%) and a test set (20%) (the random seed of the COP model dataset is different from that of the T_{cwr} model dataset), because the regressors used to simulate COP



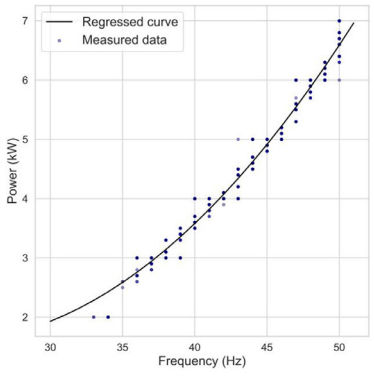
(a) Prediction accuracy of chiller COP



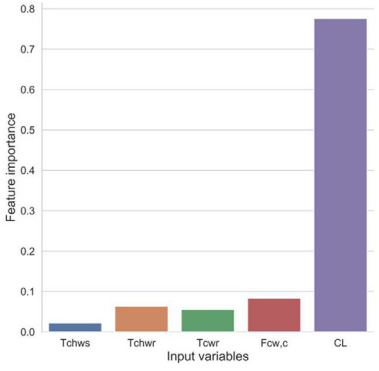
(b) Prediction accuracy of cooling tower outlet water temperature (T_{cwr})



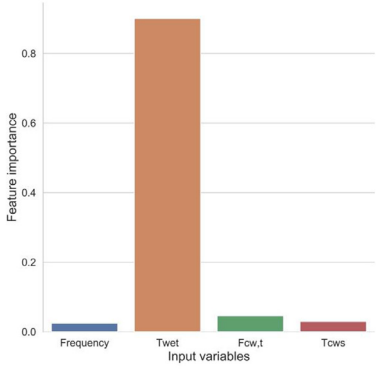
(c) Prediction accuracy of cooling water pump power



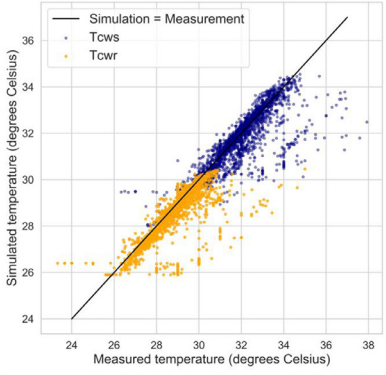
(d) Prediction accuracy of cooling tower fan power



(e) Feature importance (Gini importance [53, 58]) of chiller COP model



(f) Feature importance (Gini importance [53, 58]) of T_{cwr} model



(g) Simulation accuracy of the whole system model on the header pipe cooling water temperatures

Fig. 9. Accuracy validation of equipment models and the system model [[58]].

and T_{cwr} are black-box models, and their generalization must be validated by test sets, which are not used in model training. However, the datasets of the pump power and fan power are not split because the model of the pump and fan power is a conventional white-box model for which the accuracy has been validated, and thus, there is no need to validate its generalization [52].

Moreover, a validation simulation is conducted to verify the accuracy performance of the overall system model. This validation simulation is based on measured cooling load data, measured weather data and especially measured control signals of the real system. This work is intended to restore the real operation history of the case system (from 19th June to 18th September), and T_{cwr} and T_{cws} are selected as two variables to evaluate the model accuracy. Validation results are illustrated in Fig. 9(g) and Table 5.

Additionally, an important feature of the random forest is that it is not able to extrapolate, which could result in inaccurate output when the input combination is "special" compared to the training data. As for this case study, the variety of the training data is important to this issue. The more variable the training data is, the more robust the trained regressor is. Figs. 17 and 18 are attached in the appendix to show the distribution and co-distribution of the pre-processed training data. And these two figures indicate that the variety of the training data is acceptable because (1) correlations among variables are poor, except for the Frequency-Power relationship and the $T_{wet} - T_{cwr}$ relationship; (2) the distributional width of each variable is not evidently restricted except for the T_{chws}, but it should not be a concern because T_{chws} is constantly set to 10 °C (within the range of the training data) in the following simulation case study.

Note that the system performance model is established only for simulation of system operation. The knowledge of the system performance is not embedded in the proposed model-free controller. Only the system layout, weather data, and information on the equipment name plates (Table 3) are embedded in the model-free controller prior to simulated operation.

3.3. Realization details of proposed method and three comparative control methods

In this study, three other control methods are simulated together with the proposed method to validate the performance of the proposed method. The sequencing on-off control process of these four control methods are the same as in Section 2.2, and the frequencies of the cooling water pumps and cooling tower fans are controlled in different ways by four controllers. The frequencies of all running chilled water pumps are 50 Hz. The set point of the chilled water supply temperature ($T_{chws,\,set}$) is 10 °C (nominal value) under four controllers. Because the two cooling water pumps are identical, their frequencies are set equal when they are run simultaneously. The frequencies of two cooling tower fans are set equally as well.

As mentioned in Section 2.3 and according to the management engineer or equipment manuals, the frequencies of the variable speed appliances should be limited to protect the hardware. In this simulation, the frequency of a cooling water pump is limited within 35–50 Hz, and the frequency of a cooling tower fan is limited within 30–50 Hz. Three comparative methods are described below.

3.3.1. Basic control method

The frequencies of all running cooling tower fans and cooling water pumps are set at 50 Hz. This control method is considered as the baseline control method in this study.

Table 6
Initial Q-table (cooling tower) in the case study.

A\S	24 °C, 530 kW	.....	24 °C, 2120 Kw	.....	28 °C, 2120 kW
30 Hz	7		7		7
31 Hz	7		7		7
			.....		
50 Hz	7	.....	7	.....	7

3.3.2. Local feedback control

The control strategy of the real system which is introduced in Section 3.1 is partly adopted herein as a comparative control method. Specifically, frequencies of cooling water pumps and cooling towers are adjusted as introduced in Section 3.1 in every simulation time step. While the sequencing on-off control process is the same as in Section 2.2, which differs from the real system.

3.3.3. Model-based control

Accurate performance models of all equipment (i.e., Eqs. (5)–(11)) are embedded in the model-based controller prior to the simulated operation. In every time step (i.e., every hour) of the simulation, the model-based controller goes through all of the potential operation plans (frequencies of pumps and fans) and predicts the corresponding energy consumption of each operation plan with the simulation process described in Section 3.2. The operation plan with the maximum system COP (Eq. (1)) is selected and executed on the system model. Note that the traversal step of the operation plans is 1 Hz, which means in every time step, the model-based controller attempts (pump: 35 Hz, tower: 30 Hz), (pump: 36 Hz, tower: 30 Hz), ..., (pump: 50 Hz, tower: 50 Hz) to find the optimal pair of frequency sets.

In this case study, model-based control, local feedback control and model-free control are referred to "variable-speed control", in contrast to the basic constant-speed control. Note, technically the local feedback controller is also free of models, but the term of "model-free controller" in this case study is only referred to the proposed model-free controller.

For the Q-table in the case study, the states are specified according to the measured weather data and system rated cooling capacity. The real time wet bulb temperature is discretized to (24, 25, 26, 27, 28 °C). The discretized real time cooling load is (530, 636, ..., 2014, 2120 kW). The action options are restricted as mentioned. The initial values of the Q-table are set to 7, which is slightly higher than the nominal system COP. Specifically, the Q-table of the cooling tower in the case study is initialized as in Table 6, the shape of which is 80 states×21 actions. The Q-table of the cooling water pump is similar to Table 6, but actions range from 35 Hz to 50 Hz, the shape of pump Q-table is 80 states×16 actions.

4. Results and discussion

The operation of the case system from 19th June to 18th September is simulated on an hourly basis under the supervision of four control methods. The simulation results are discussed from several aspects: energy consumption, learning process of the model-free controller, control actions taken by different controllers, system water temperature, randomness of the model-free controller, performance evolution of the model-free controller in a longer period. Because the decision-making procedure of the model-free controller contains uncertainty and randomness, the simulation under the model-free control is conducted five times independently, details are given in Section 4.5.

Table 7
Energy consumption of case system under four control methods.

	Direction	Cooling tower energy consumption (kWh)	Cooling water pump energy consumption (kWh)	Chiller energy consumption (kWh)	System energy consumption (kWh)
Real system measurement		23 260	54 634	380 379	458 273
Basic control	Forward/Reverse	28 235	72 964	362 995	464 194
Local feedback control	Forward/Reverse	18 814	48 316	361 974	428 924
Model-based control	Forward/Reverse	8 722	28 151	360 772	397 645
Model-free control	Forward	13 068	34 551	362 231	409 850
(five- round average)	Reverse	12 939	35 084	361 979	410 003

4.1. Energy consumption

The three-month energy consumptions of the system under four control methods are listed in Table 7. Compared with the basic controller, the local feedback controller can conserve 7% of system energy (summary of chillers, cooling water pumps and cooling towers), the model-free controller can conserve 11% of the system energy, and the model-based controller can conserve 14% of the system energy.

Table 7 indicates that when controlled by the basic control method, the case system requires the most energy, mainly due to the unnecessary use of cooling water pumps and cooling towers. Local feedback control performs better than basic control in terms of energy conservation because this control method adjusts the frequencies of the pumps and tower fans to maintain ΔT_{cw} and the cooling tower approach at predefined set points. However, the energy conservation capability of the variable speed pumps and fans is not fully utilized under this control method because (1) the set points of ΔT_{cw} and the cooling tower approach are predefined at a low level by the management engineer of the actual system, and (2) these set points are constant instead of adaptive during system operation.

Moreover, Table 7 shows that the chiller energy consumption is slightly influenced by the control method. That is because according to Fig. 9(e), chiller COP is slightly influenced by T_{cwr} and $F_{cw,\,c}$, both of which vary with the control logic (Fig. 15).

The energy conservation performance of model-free control is better than that of local feedback control but worse than that of model-based control because (1) compared with local feedback control, the model-free controller continues to learn and evolve by searching for the optimal set points; (2) unlike the model-based controller, the model-free controller is not embedded with the historical operation data and equipment performance models, indicating that the model-free controller must learn from square one to build its own experience, which leads to the gap between the model-free controller and model-based controller.

Fig. 10(a) illustrates the daily energy conservation amounts (scatters) of the three variable-speed control strategies compared with the basic control. The scatters indicate that in the beginning of operation, the energy saving rates of all three variable speed controllers are unstable because the cooling load data is variable in the beginning. On the contrary, the scatters after 60 days are closer to regressed lines because the working condition of the system is more stable in this period.

Moreover, Fig. 10(a) shows that the model-free controller outperforms the local feedback controller in the very beginning, when the model-free controller is not sufficiently trained. That is because in this study, the set points of two controlled variables (ΔT_{cw} and cooling tower approach) are set at a low level, which makes the local feedback controller in this study tends to keep equipment operating at high frequencies (Fig. 14(c)). Hence, in many situations (especially when the partial load ratio of running chiller(s) is high), the control signals determined by the local feedback controller are close to the ones by the basic controller. Besides, according to Fig. 7, the real system controlled by a real local feedback controller also kept the F_{cw} at a high level on both sides of the investigated period. The abovementioned is why an untrained model-free controller could outperform the local feedback controller in the beginning.

Three regression lines represent the daily energy conservation trends of three variable-speed control methods, which all show upward trends because the system cooling load is slightly decreased after 30th July, and partial load working conditions are more suitable for variable-speed control methods in conservation of energy in terms of pumps and tower fans.

In order to better expose the learning effectiveness of the proposed model-free control method, simulations are conducted once more in the reversed direction, from 18th September to 19th June, under the supervision of four controllers. Still, five independent simulations are conducted for the model-free controller in this reversed direction. The simulation results of basic controller, local feedback controller and model-based controller do not change with the simulation direction because (1) the system model is basically independent of time sequence; (2) the control logic of these three controllers are deterministic. As is shown in Fig. 10(b), without the benefit of the input data, the regression line of the proposed method is nearly horizontal, unlike the other two variable speed controllers which go downwards due to the input data. This could verify the effectiveness of the reinforcement learning process.

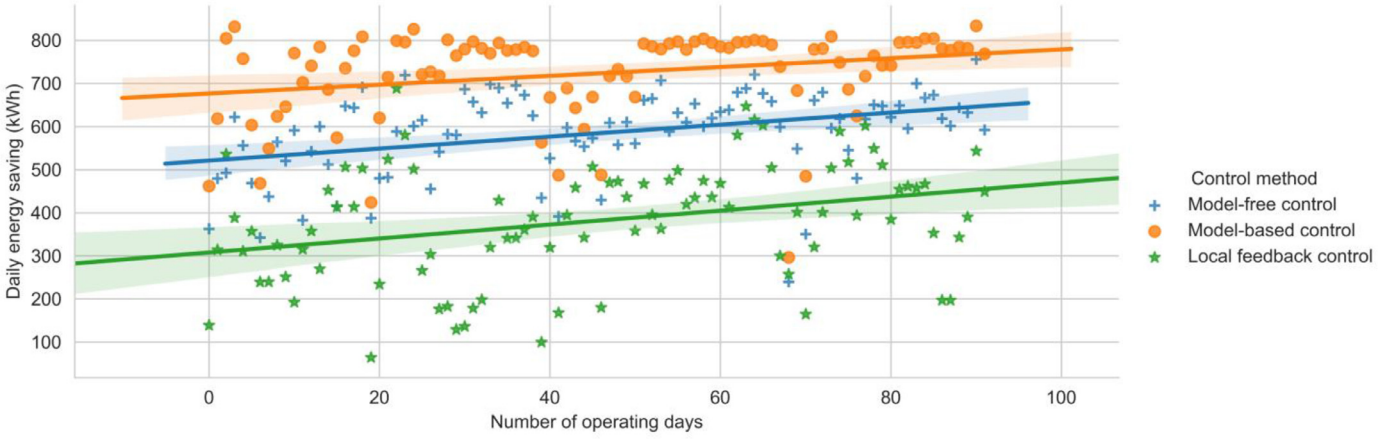
Parameters of regression lines are listed in Table 8. Slope values could prove the improvement of the model-free controller due to the learning process. Meanwhile, the standard deviations of slopes suggest that the regression confidence of the model-free control result is the best, which is also verified by the shade areas in Fig. 10 (95% confidence intervals).

4.2. Learning process of the model-free controller

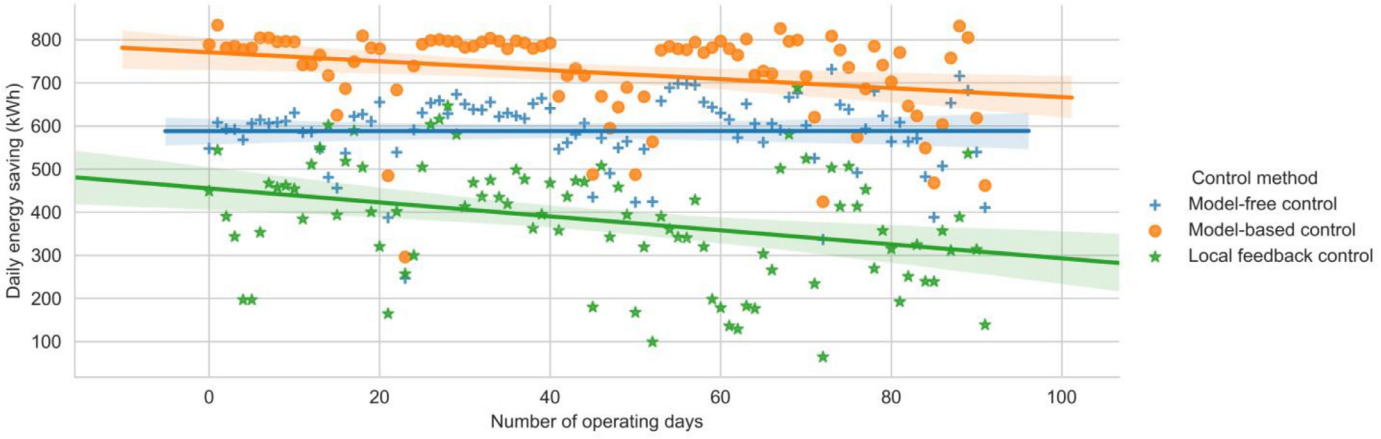
As is shown in Fig. 11, the reward (i.e., the comprehensive system COP) of the model-free controller is accumulated almost linearly during the simulated operation. That is because (1) the comprehensive system COP is mainly determined by the chiller COP, which is typically within the range of 7–9 according to Fig. 18; (2) the chiller COP is not evidently influenced by the operation of the cooling water system (Fig. 9(e)). Moreover, Fig. 11 indicate that the direction of learning (from heating period to cooling period or on the contrary) does not substantially influence the learning effectiveness, which is also reflected in Table 7. Additionally, the reverse line is slightly above the forward line because the cooling load after 30th July is lower, and the case system is more efficient at partial-load working conditions.

Because the learning process is not evidently influenced by the learning direction, the following content including Sections 4.2–4.4 will be discussed merely based on the results of simulations in forward direction.

Fig. 12 shows trends of three variables: accumulated number of updated pump Q-table entries, accumulated number of updated



(a). Forward direction simulation (from 19th June to 18th September)



(b). Reverse direction simulation (from 18th September to 19th June)

Fig. 10. Daily energy saving (kWh) (shades around the regression lines are 95% confidence intervals for the regression estimate).

Table 8
Parameter of regression lines.

	Forward simulation		Reverse simulation	
	Slope value	Standard deviation of the slope	Slope value	Standard deviation of the slope
Local feedback control	1.620	0.514	−1.620	0.514
Model-free control (five- round average)	1.389	0.364	0.004	0.338
Model-based control	1.038	0.399	−1.038	0.399

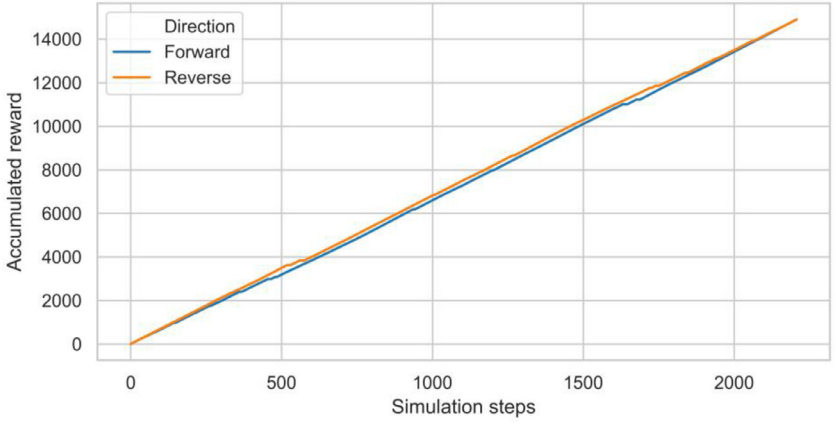


Fig. 11. Accumulated rewards of learning processes in two directions.

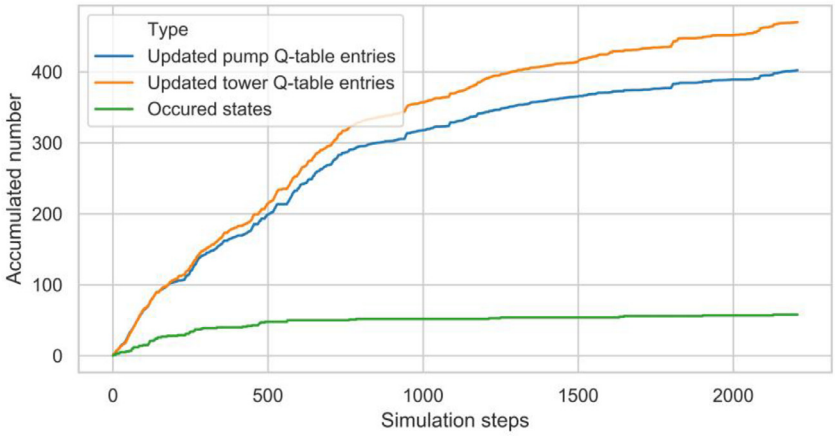


Fig. 12. Accumulated number of Q-table entries which are updated at least once; accumulated number of occurred different states.

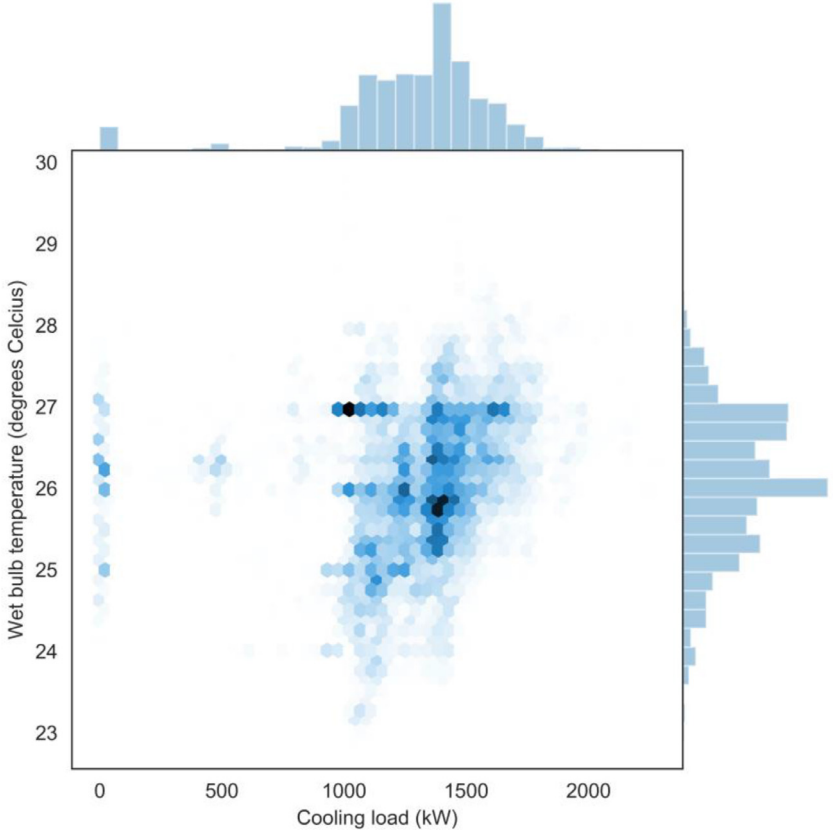


Fig. 13. Co-distribution of the system cooling load and T_{wet}.

tower Q-table entries, and accumulated number of occurred states. These three lines indicate that (1) almost all possible states have occurred at least once in the first month of system operation (0–700 steps); (2) the exploration of two RL agents is most evident in the first month of learning; (3) the accumulated number of updated Q-table entries stabilizes between 400 and 500 for both pump Q-table and tower Q-table, which is far from the total entry amount of the defined Q-table (tower Q-table: 1280, pump Q-table: 1680). This gap exists because the state is normally distributed as Fig. 13 shows, and the control actions at rare states could not be sufficiently explored without a long period. But that would not affect the performance of the proposed controller because only the control decision at typical, regular states really matters to the controller's performance (Pareto's law), and the learning at these states is always quickly accomplished because these states occur frequently.

4.3. Frequency set points and system COP under three variable-speed controllers

Fig. 14 shows the optimal frequency set points and maximum system COP values under three variable-speed controllers. It should be noted that during the simulation, some states (such as T_{wet} = 24 °C and system cooling load = 2120 kW) did not occur even once. Thus, in Fig. 14, the values at these states are blank. Additionally, the results of the model-free controller represent the policy after the learning of the first cooling season (from 19th June to 18th September). The following are inferred from this figure.

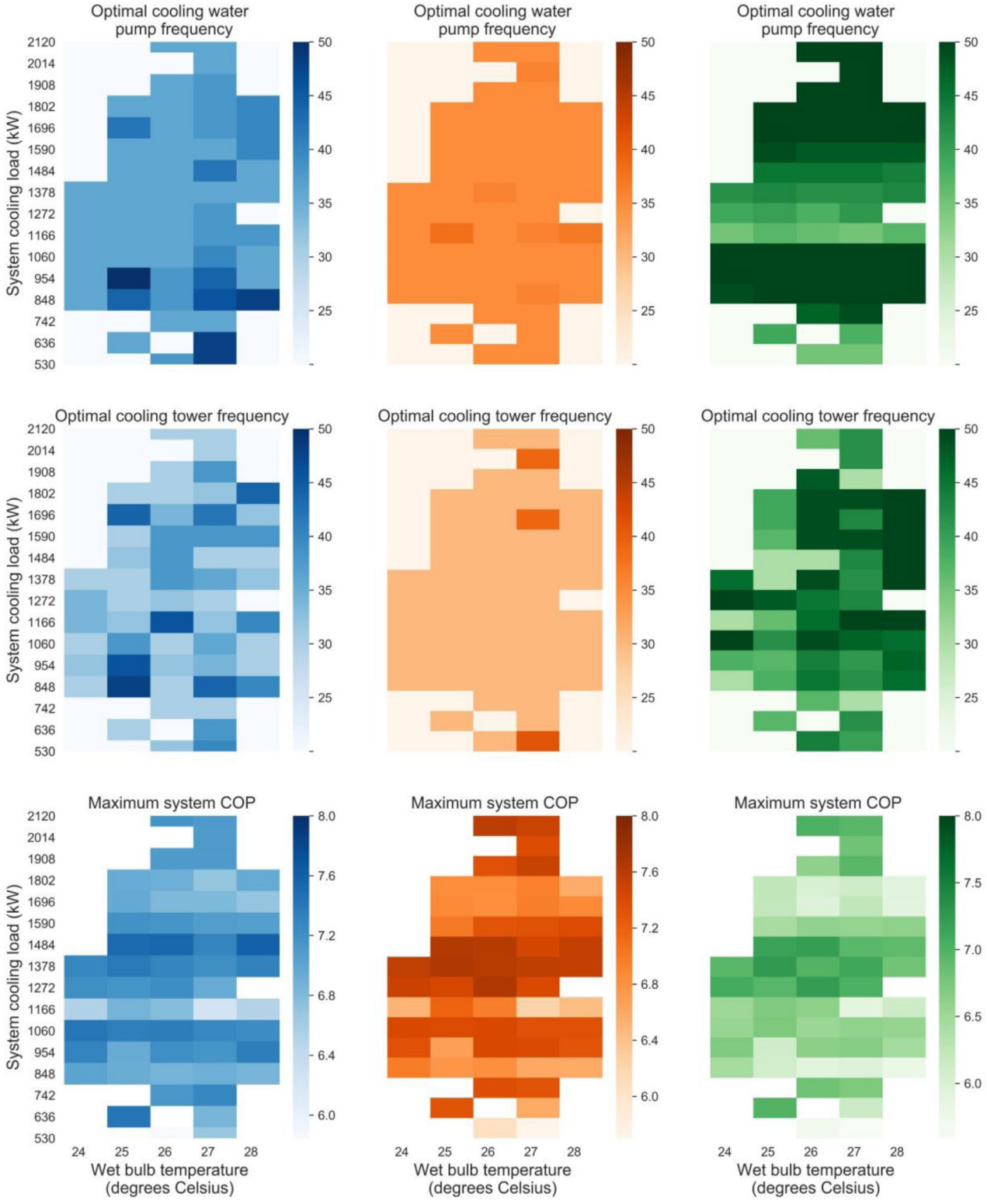


Fig. 14. Frequency set points and system COP of three variable-speed control methods, (a) Model-free control, (b) Model-based control, (c) Local feedback control.

Compared with model-free control and local feedback control, model-based control sets lower frequencies for the pumps and cooling tower fans. The system COP values of the model-based controller are higher than those of the other two. The excellent performance of the model-based controller is due to three aspects: (1) it was embedded with accurate performance models for all equipment, which means that it could accurately predict the outcome energy consumption caused by each control action; (2) the model-based controller is designed to search for the frequency set points to achieve the maximum system COP in each time step; and (3)

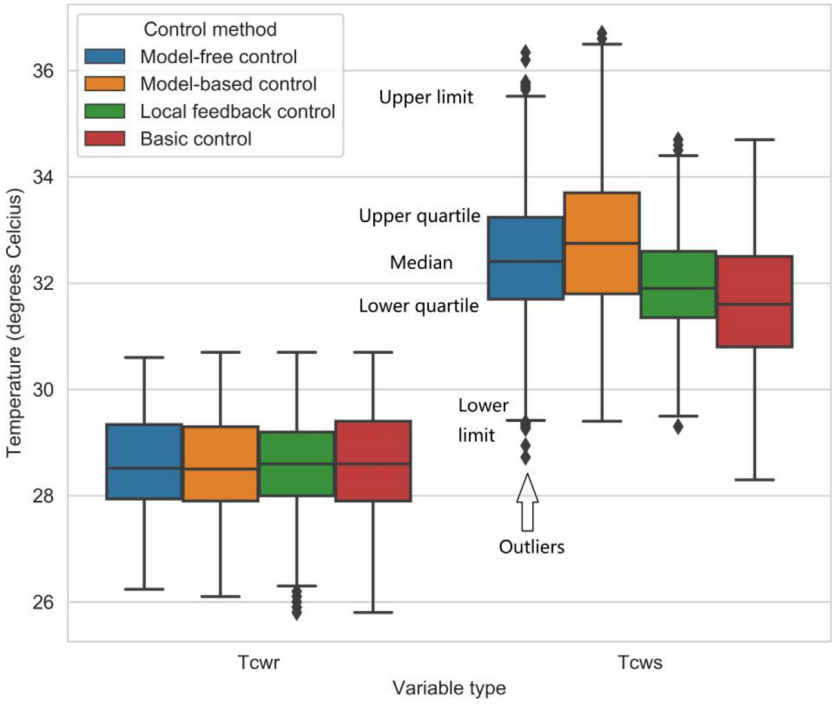


Fig. 15. Distributions of system water temperatures under four controllers (upper limit is the maximum among the data points which are smaller than upper quartile plus 1.5 interquartile range; lower limit is the minimum among the data points which are larger than the lower quartile minus 1.5 interquartile range; and outliers are the data out of upper and lower limits).

Table 3 suggests that the cooling water pumps and cooling towers of the case system are actually oversized for the case chillers, thus revealing additional potential for energy conservation [5].

Under local feedback control, the cooling water pump frequency is highly correlated with the system cooling load because the heat rejected by the chillers must be transferred by the cooling water pumps to the ambient environment. The higher the cooling load, the more heat that must be rejected. To maintain ΔT_{cw} at the constant set point, the local feedback controller must increase the cooling water flowrate to cope with the increasing system cooling load.

The heatmaps of the maximum system COP are similar under three variable-speed controllers. The peak system COP appears when T_{wet} is approximately 25 °C and the system cooling load is approximately 1484 kW for the following reasons: (1) The chillers in the case system are both screw chillers, and they reach peak COP when the PLR (partial load ratio) is approximately 70–80% [59]. When the system cooling load is approximately 1484 kW, the cooling load on each chiller is 742 kW, and the PLRs on both chillers are 70%, close to the optimal PLR, which leads to a high system COP value; (2) The lower the value of T_{wet}, the easier it is for this system to reject heat to the outdoor environment [60].

4.4. System water temperature distribution under four control methods

The simulation results of the system water temperature are illustrated in Fig. 15, which indicates the following.

The system T_{cwr} does not change substantially with the controllers, because T_{cwr} is primarily influenced by T_{wet}, which is shown in Fig. 9(f). Under basic control, the values of ΔT_{cw} are generally smaller than those under variable-speed control because the frequency of the cooling water pumps remains unchanged (50 Hz) under basic control, which leads to high flowrate and low delta-T in the cooling water loop [61].

Together with the energy consumption results in Section 4.1, it is noted that ΔT_{cw} is positively correlated with the energy conservation performance of a control method because the pump frequency and pump power are negatively correlated with the ΔT_{cw} [61]. When the system is operating at partial load conditions, reducing the pump frequency could significantly reduce the pump power with a limited sacrifice in chiller COP. In brief, when chiller operation is not optimized, a controller that better optimizes pump operation is more likely to conserve more energy for the entire system.

4.5. Randomness of the proposed model-free controller

As is shown in Eq. (4), the optimization policy of the proposed model-free controller contains uncertainty and randomness. Thus, in this study, the simulations under the model-free control are all repeated five times independently to cope with this issue. The energy consumption results are listed in Table 9. The standard deviations of chiller energy, pump energy, cooling tower energy and system energy are all less than 2.5% of the corresponding average values. The diversity of cooling tower energy is the largest because the frequency of cooling tower fans is adjustable within 30–50 Hz which is wider than the 35–50 Hz of cooling water pumps. And the diversity of chiller energy is not evident because as is mentioned in Section 4.1, the control on the cooling water system does not evidently influence the chiller COP.

4.6. Performance evolution of the proposed model-free controller in longer period

To better investigate the evolution of the proposed model-free controller, five independent rounds of ten-episode simulation are conducted. A ten-episode simulation is realized by continuously simulating the system operation under the model-free control for ten times of the period from 19th June to 18th September, end to

Table 9
Energy consumption of five independent runs.

Direction of simulation		Cooling tower energy consumption (kWh)	Cooling water pump energy consumption (kWh)	Chiller energy consumption (kWh)	System energy consumption (kWh)
Forward	Round 1	13 269	35 179	361 972	410 420
	Round 2	12 746	33 932	362 421	409 099
	Round 3	12 698	34 383	362 100	409 181
	Round 4	13 207	34 668	362 362	410 237
	Round 5	13 419	34 592	362 301	410 312
	Average	13 068	34 551	362 231	409 850
	Standard deviation	325	453	188	651
Reverse	Round 1	12 905	35 032	362 056	409 993
	Round 2	12 644	34 807	361 903	409 354
	Round 3	12 958	35 007	362 138	410 103
	Round 4	13 159	35 059	362 055	410 273
	Round 5	13 031	35 517	361 744	410 292
	Average	12 939	35 084	361 979	410 003
	Standard deviation	191	261	157	383

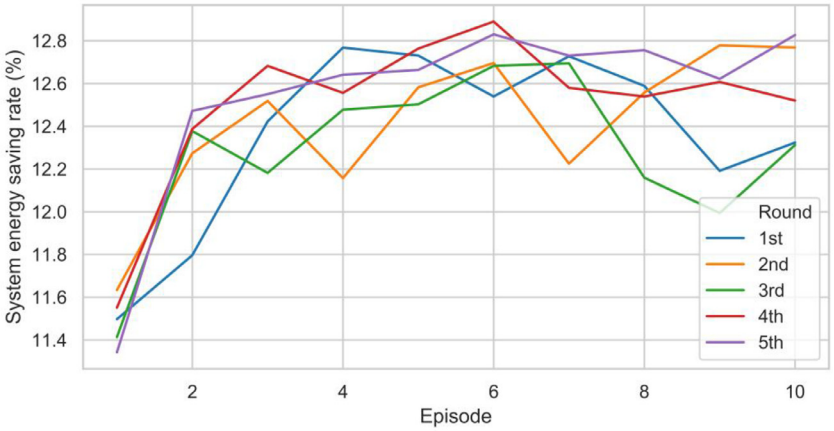


Fig. 16. Evolution of the energy saving performance in ten episodes.

end. As is shown in Fig. 16, the energy saving rate of the model-free controller improves evidently in the first two episodes; afterwards the performance gets stabilized between 12% and 13%. Two reasons could account for this: (1) the p value is defined as the number of total simulation steps in one episode, which results in more exploitation instead of exploration after the first episode; (2) the cooling load data and weather data do not change with episodes, which limits the learning of the controller after the first episode.

5. Conclusion and future work

5.1. Conclusion

A model-free optimal control method based on reinforcement learning is proposed in this paper to control the building cooling water system. In the proposed method, discretized wet bulb temperature and system cooling load are the states, the frequencies of the fans and pumps are the actions, and the rewards are the comprehensive COP of the chillers, cooling water pumps, cooling towers. A measured data-based simulation is conducted under the supervision of four types of controllers: basic controller, local feedback controller, model-based controller, and proposed model-free controller. The three-month simulation results indicate that the model-free controller was able to function and evolve simultaneously during the system operation period.

Compared with the basic controller, the model-free controller could conserve 11% of the system energy, which is more than that of the local feedback controller at 7% but less than that of the model-based controller at 14%. Although the energy conservation performance of the model-free controller is inferior to that of the model-based controller, the model-free controller requires less a priori knowledge and sensors to function, which makes it more applicable in the engineering practice.

For a central chilled water system with a scale is similar to that of the case system, three month's learning in the cooling season is sufficient to develop a model-free controller with acceptable energy conservation performance. In this case study, simulations are conducted on a laptop with 8G RAM, i7-8650U CPU. The hourly optimization by the model-free controller takes less than one second, which is sufficient for engineering practice, and the model-based controller takes ten seconds for one optimization.

Finally, the proposed model-free method is not intended to replace the model-based method, but to offer an alternative to the buildings whose accurate system performance models are not accessible due to the lack of data or sensors.

5.2. Future work

The optimization of chillers and chilled water pumps are not included in this study. Because the energy-saving performance of the optimal chiller loading has been validated in many studies, it is promising and meaningful to apply model-free control to optimization of the operation of chillers and even the entire central chilled water system. Additionally, in this paper, the proposed method is only validated on the system composed of identical cooling units (identical chillers, identical cooling water pumps and identical cooling towers). The performance of the proposed method

on systems with multiple sized units is worth further investigation. Furthermore, the optimization of cooling water pumps and cooling towers are performed independently in this study, thus a better cooperation mechanism is worth further investigation

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.

CRediT authorship contribution statement

Shunian Qiu: Conceptualization, Methodology, Software, Writing - original draft, Investigation. **Zhenhai Li:** Conceptualization, Supervision. **Zhengwei Li:** Conceptualization, Project administration, Writing - review & editing, Resources. **Jiajie Li:** Methodology, Investigation. **Shengping Long:** Supervision. **Xiaoping Li:** Project administration.

Acknowledgement

The funding agency is 'Ministry of Science and Technology of the People's Republic of China', under the program 'National Key R&D Program of China', with a grant number '2017YFC0704200'.

Appendix

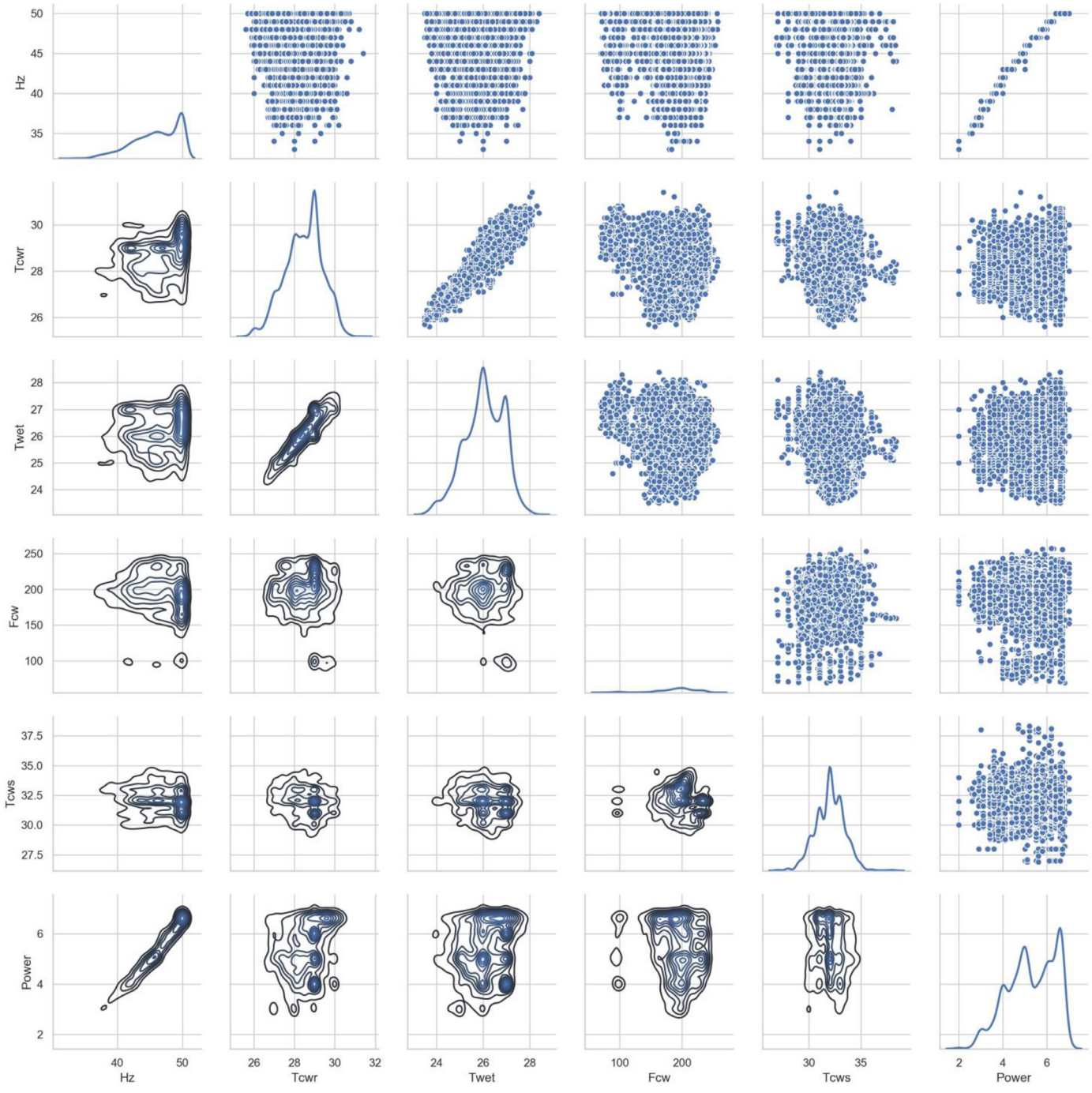


Fig. 17. Distributions and co-distributions of the pre-processed training data of the cooling tower T_{cwr} model (Hz refers to the fan frequency).

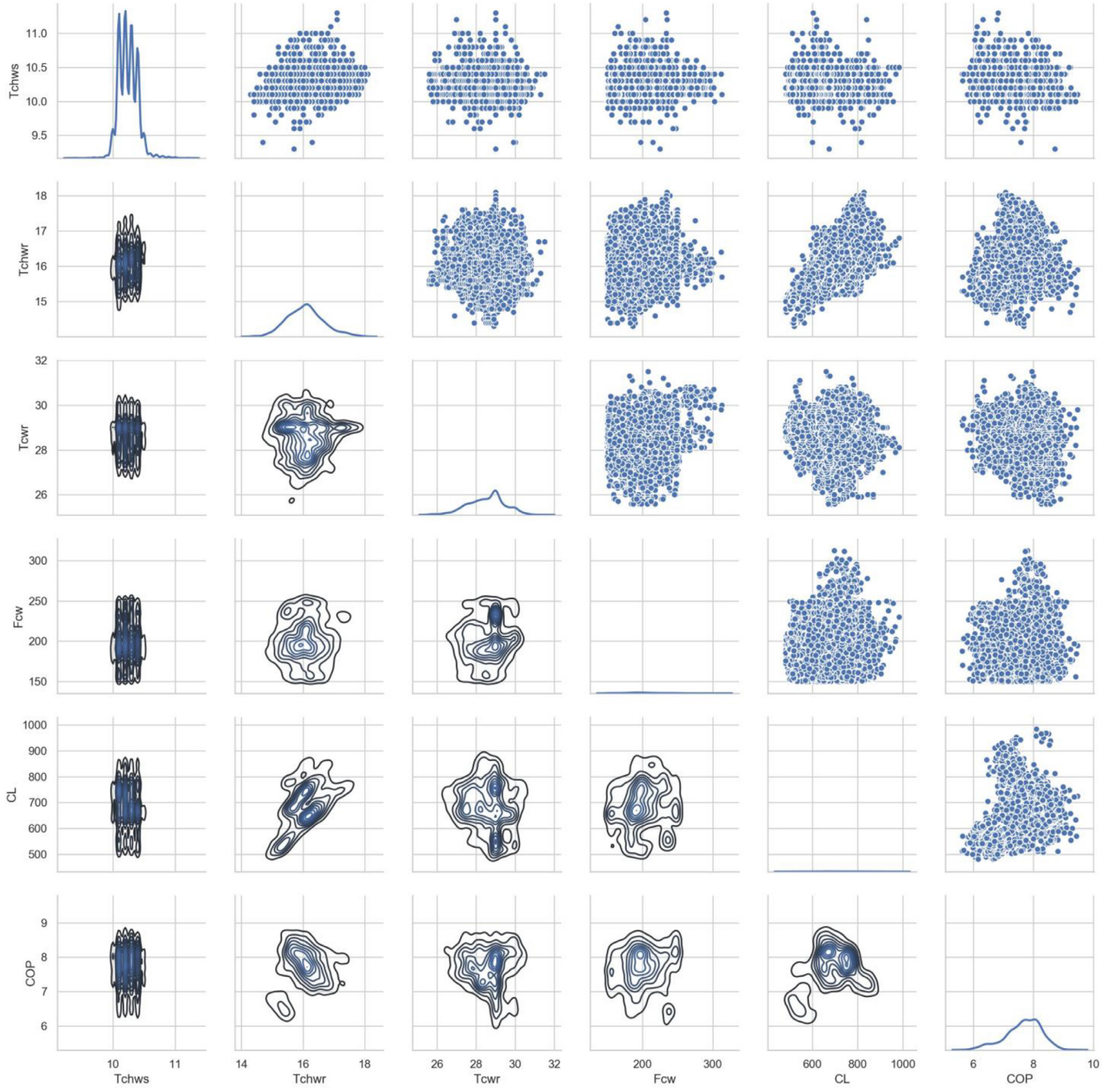


Fig. 18. Distributions and co-distributions of the pre-processed training data of the chiller COP model.

References

[1] L. Pérez-Lombard, J. Ortiz, C. Pout, A review on buildings energy consumption information, Energy Build. 40 (3) (2008) 394–398.

[2] W. Li, P. Xu, X. Lu, H. Wang, Z. Pang, Electricity demand response in China: status, feasible market schemes and pilots, Energy 114 (2016) 981–994.

[3] J. Hou, P. Xu, X. Lu, Z. Pang, Y. Chu, G. Huang, Implementation of expansion planning in existing district energy system: a case study in China, Appl. Energy 211 (2018) 269–281.

[4] Z. Pang, Z. O'Neill, Y. Li, F. Niu, The role of sensitivity analysis in the building performance analysis: a critical review, Energy Build. 209 (2020) 109659.

[5] S.T. Taylor, Fundamentals of Design and Control of Central Chilled-Water Plants, ASHRAE learning institute, 2017.

[6] D.J. Swider, A comparison of empirically based steady-state models for vapor–compression liquid chillers, Appl. Therm. Eng. 23 (5) (2003) 539–556.

[7] J.E. Braun, G.T. Diderrich, Near-optimal control of cooling towers for chilled-water systems, ASHRAE Trans. 96 (1990) 2.

[8] Shengwei Wang, Hvac, Z.J., and Research, R., Supervisory and optimal control of building HVAC systems: a review. 2008.14(1): p. 3–32.

[9] Z. Ma, S. Wang, Online fault detection and robust control of condenser cooling water systems in building central chiller plants, Energy Build. 43 (1) (2011) 153–165.

[10] Y. Yao, Z. Lian, Z. Hou, X. Zhou, Optimal operation of a large cooling system based on an empirical model, Appl. Therm. Eng. 24 (16) (2004) 2303–2321.

[11] Y. Yao, J. Chen, J. Feng, S. Wang, Modular modeling of air-conditioning system with state-space method and graph theory, Int. J. Refrig. 99 (2019) 9–23.

[12] Y. Yao, W. Wang, M. Huang, A state-space dynamic model for vapor compression refrigeration system based on moving-boundary formulation, Int. J. Refrig. 60 (2015) 174–189.

[13] S. Huang, W. Zuo, M.D. Sohn, Improved cooling tower control of legacy chiller plants by optimizing the condenser water set point, Build. Environ. 111 (2017) 33–46.

[14] S. Huang, W. Zuo, M.D. Sohn, A Bayesian network model for predicting the

cooling load of educational facilities, the ASHRAE and IBPSA-USA SimBuild 2016: Building Performance Modeling Conference, 2016.
[15] J. Wang, G. Huang, Y. Sun, X. Liu, Event-driven optimization of complex HVAC systems, Energy Build. 133 (2016) 79–87.
[16] S. Wang, Dynamic simulation of a building central chilling system and evaluation of EMCS on-line control strategies, Build. Environ. 33 (1) (1998) 1–20.
[17] W. Lee, L. Lin, Optimal chiller loading by particle swarm algorithm for reducing energy consumption, Appl. Therm. Eng. 29 (8) (2009) 1730–1734.
[18] Y.C. Chang, J.K. Lin, M.H. Chuang, Optimal chiller loading by genetic algorithm for reducing energy consumption, Energy Build. 37 (2) (2005) 147–155.
[19] A.J. Ardakani, F.F. Ardakani, S.H. Hosseinian, A novel approach for optimal chiller loading using particle swarm optimization, Energy Build. 40 (12) (2008) 2177–2187.
[20] M.H. Sulaiman, M.I.M. Rashid, M.R. Mohamed, O. Aliman, H. Daniyal, An application of cuckoo search algorithm for solving optimal chiller loading problem for energy conservation, Appl. Mech. Mater. 793 (2015) 500–504.
[21] A. Beghi, L. Cecchinato, M. Rampazzo, A multi-phase genetic algorithm for the efficient management of multi-chiller systems, Energy Convers. Manage. 52 (3) (2011) 1650–1661.
[22] Y. Sun, G. Huang, Z. Li, S. Wang, Multiplexed optimization for complex air conditioning systems, Build. Environ. 65 (65) (2013) 99–108.
[23] N. Zhu, K. Shan, S. Wang, Y. Sun, An optimal control strategy with enhanced robustness for air-conditioning systems considering model and measurement uncertainties, Energy Build. 67 (4) (2013) 540–550.
[24] J. Cui, S. Wang, A model-based online fault detection and diagnosis strategy for centrifugal chiller systems, Int. J. Therm. Sci. 44 (10) (2005) 986–999.
[25] Y. Fu, Z. Li, F. Feng, P. Xu, Data-quality detection and recovery for building energy management and control systems: case study on submetering, Hvac & R Res. 22 (6) (2016) 798–809.
[26] Y. Sun, S. Wang, G. Huang, Chiller sequencing control with enhanced robustness for energy efficient operation, Energy Build. 41 (11) (2009) 1246–1255.
[27] Z. Li, G. Huang, Y. Sun, Stochastic chiller sequencing control, Energy Build. 84 (84) (2014) 203–213.
[28] R.S. Sutton, A.G. Barto, F. Bach, Reinforcement Learning: an Introduction, Bradford Book, 2018.
[29] A. de Gracia, C. Fernández, A. Castell, C. Mateu, L.F. Cabeza, Control of a PCM ventilated facade using reinforcement learning techniques, Energy Build. 106 (2015) 234–242.
[30] Pang, Z., Xu, P., Lu, X., Qiu, S., Chen, L., and Hou, J., Evaluation of the performance of a new solar ventilated window: modeling and experimental verification. J. Renew. Sustain. Energy2017(9): p. 1–16.
[31] W. Valladares, M. Galindo, J. Gutiérrez, W.-.C. Wu, K.-.K. Liao, J.-.C. Liao, K.-.C. Lu, C.-.C. Wang, Energy optimization associated with thermal comfort and indoor air control via a deep reinforcement learning algorithm, Build. Environ. 155 (2019) 105–117.
[32] Z. Zou, X. Yu, S. Ergan, Towards optimal control of air handling units using deep reinforcement learning and recurrent neural network, Build. Environ. 168 (2020) 106535.
[33] G.P. Henze, J. Schoenmann, Evaluation of reinforcement learning control for thermal energy storage systems, HVAC & R Res. 9 (3) (2003) 259–275.
[34] S. Liu, G.P. Henze, Experimental analysis of simulated reinforcement learning control for active and passive building thermal storage inventory. Part 2: results and analysis, Energy Build. 38 (2) (2006) 148–161.
[35] S. Liu, G.P. Henze, Experimental analysis of simulated reinforcement learning control for active and passive building thermal storage inventory: part 1. Theoretical foundation, Energy Build. 38 (2) (2006) 142–147.
[36] Cheng, Z., Zhao, Q., Wang, F., Jiang, Y., Xia, L., and Ding, J., Satisfaction based Q-learning for integrated lighting and blind control. Energy Build. 127(sep.): p. 43–55.
[37] Chen, Y., Norford, L.K., Samuelson, H.W., and Malkawi, A., Optimal control of HVAC and window systems for natural ventilation through reinforcement learning. Energy Build.: p. S0378778818302184.
[38] Y.C. Chang, A novel energy conservation method—optimal chiller loading, Electric Power Syst. Res. 69 (2) (2004) 221–226.
[39] Y.C. Chang, F.A. Lin, C.H. Lin, Optimal chiller sequencing by branch and bound method for saving energy, Energy Convers. Manage. 46 (13) (2005) 2158–2172.
[40] T. Hartman, All-Variable speed centrifugal chiller plants, ASHRAE J. (9) (2001).
[41] S. Qiu, F. Feng, W. Zhang, Z. Li, Z. Li, Stochastic optimized chiller operation strategy based on multi-objective optimization considering measurement uncertainty, Energy Build. 195 (2019) 149–160.
[42] Y. Liao, Y. Sun, G. Huang, Robustness analysis of chiller sequencing control, Energy Convers. Manage. 103 (2015) 180–190.
[43] S. Qiu, F. Feng, Z. Li, G. Yang, P. Xu, Z. Li, Data mining based framework to identify rule based operation strategies for buildings with power metering system, Build. Simul. (2018).
[44] ASHRAE, Liquid Chilling System, in ASHRAE Systems and Equipment Handbook, ASHRAE learning institution, 2000.
[45] C.W. Chen, Y.C. Chang, W.T. Liao, C.W. Lee, Application of genetic programming method combined with neural network in HVAC optimal operation, Appl. Mech. Mater. 548–549 (2014) 1030–1034.
[46] Z. Liu, H. Tan, D. Luo, G. Yu, J. Li, Z. Li, Optimal chiller sequencing control in an office building considering the variation of chiller maximum cooling capacity, Energy Build. 140 (2017) 430–442.
[47] J. Liao, X. Xie, H. Nemer, D.E. Claridge, C.H. Culp, A simplified methodology to optimize the cooling tower approach temperature control schedule in a cooling system, Energy Convers. Manage. 199 (2019) 111950.
[48] Y. Yao, J. Chen, Global optimization of a central air-conditioning system using decomposition–coordination method, Energy Build. 42 (5) (2010) 570–583.
[49] H. Hotelling, The generalization of student's ratio, Annals Math. Stat. 2 (3) (1930).
[50] A.S. KLEIN, TRNSYS-A Transient System Simulation Program, University of Wisconsin-Madison, 1988.
[51] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
[52] A. Géron, Hands-on Machine Learning With Scikit-Learn and Tensorflow, O'REILLY, 2018.
[53] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: machine learning in python, J. Mach. Learn. Res. 1 (12) (2011) 2825–2830.
[54] Z. Pang, P. Xu, Z. O'Neill, J. Gu, S. Qiu, X. Lu, X. Li, Application of mobile positioning occupancy data for building energy simulation: an engineering case study, Build. Environ. 141 (2018) 1–15.
[55] J. Gu, P. Xu, Z. Pang, Y. Chen, Y. Ji, Z. Chen, Extracting typical occupancy data of different buildings from mobile positioning data, Energy Build. (2018).
[56] S. Qiu, Z. Li, Z. Pang, W. Zhang, Z. Li, A quick auto-calibration approach based on normative energy models, Energy Build. 172 (2018) 35–46.
[57] ASHRAE Standards Committee, ASHRAE Guideline 14, Measurement of Energy and Demand Savings, ASHRAE, Atlanta, 2002.
[58] sklearn.tree.DecisionTreeRegressor. 2019; Available from: https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegressor.html?highlight=decision%20tree#sklearn.tree.DecisionTreeRegressor.feature_importances_.
[59] L.D.D. Harvey, A handbook on low-energy buildings and district-energy systems: fundamentals, Techniques and Examples, English Edition, Routledge, 2012.
[60] J.F. Kreider, Heating and Cooling of Buildings : Design for Efficiency, McGraw-Hill, 1994.
[61] D.C. Gao, S. Wang, K. Shan, C. Yan, A system-level fault detection and diagnosis method for low delta-T syndrome in the complex HVAC systems, Appl. Energy 164 (2016) 1028–1038.