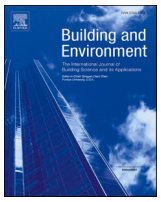
Building and Environment 210 (2022) 108704

Contents lists available at ScienceDirect

Building and Environment

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



Development of an adaptive artificial neural network model and optimal control algorithm for a data center cyber–physical system

Young Jae Choi, Bo Rang Park, Ji Yeon Hyun, Jin Woo Moon *

School of Architecture and Building Science, Chung-Ang University, 84, Heukseok-ro, Dongjak-gu, Seoul, 06974, Republic of Korea

ARTICLE INFO

Keywords:
Cyber-physical system
Predictive control
Adaptive control
Artificial neural network
Optimal control algorithm
Data center

ABSTRACT

This study aimed to develop an adaptive artificial neural network model (AAM) for the prediction of the rack inlet temperature and cooling system energy, and the optimal control algorithm for cooling system of a containment-type data center. A cyber-physical system (CPS) framework, that incorporated the AAM and control algorithm, was also proposed for the precise control of the data center cooling system. To develop the AAM model and control algorithm, mathematical modeling of a reference physical model was conducted, and training data were acquired from this model. The performance of the proposed AAM and control algorithm was then compared with that of a non-adaptive ANN model (NAAM) in terms of prediction accuracy and control stability. The analysis results indicated that the optimal control algorithm with the AAM exhibited superior prediction accuracy and control stability than the algorithm with the NAAM. In particular, for the AAM-based algorithm under conditions representing a novel data center environment, the root mean square error (RMSE) and coefficient of variation of the RMSE (CV(RMSE)) for the predicted and actual values were 0.22 °C and 1.02%, respectively, for the inlet rack temperature and 0.19 kW and 0.76% for the cooling system energy. The control was also stable, with an MAE of 0.08 °C and a maximum error of 1.17 °C. Based on this analysis, a CPS-based control strategy incorporating an ANN-based optimal control algorithm is expected to be an effective energy efficiency solution for existing data center without changing IT equipment or cooling systems.

1. Introduction

1.1. Background

Data-based industries, such as artificial intelligence, autonomous driving, and cloud services, require high-capacity storage devices and high-performance computing systems. In 2020, 97% of the world's data went through data centers and, with an average annual data growth of 40%, the demand for data centers is expected to continue to increase [1]. In 2019, data centers consumed 200 TWh of energy [2], accounting for about 1% of global energy consumption and 0.3% of carbon dioxide emissions [2,3], and this energy consumption has doubled every four years over the past decade [4]. As such, it is predicted that about 321 TWh of energy will be used by data centers in 2020 [5]. Given this rapid rise in energy consumption, it is important to improve the energy efficiency of data centers to prevent even more rapid increases. Therefore, efforts to reduce data center energy consumption are essential.

Data centers use energy for various purposes, including their cooling systems, information technology (IT) equipment, and power equipment. Cooling systems and IT equipment consume the largest proportion of energy at data centers (50% and 26%, respectively) [6,7]. The energy efficiency of a data center can be monitored via its power usage effectiveness (PUE), which is the ratio of its overall energy consumption to that of the IT equipment. The closer the PUE is to 1, the higher is the data center efficiency. Thus, to achieve a low PUE, energy savings in the cooling system are essential. To reduce the cooling system energy, various state-of-the-art technologies are being applied to newly constructed data centers. However, increasing the energy efficiency of existing data centers is difficult because data center characteristics are difficult to remodel. Therefore, a simple and highly applicable energy saving method that requires minimal hardware changes is needed.

1.2. Related studies

1.2.1. Energy saving methods

Various strategies have been proposed to reduce the energy consumption of data center cooling systems, and they can generally be divided into hardware or software approaches. Hardware-based strategies include the efficient arrangement of IT equipment and cooling

* Corresponding author.
E-mail addresses: chldudwo13@cau.ac.kr (Y.J. Choi), pbrighten@cau.ac.kr (B.R. Park), jiyuneco@cau.ac.kr (J.Y. Hyun), gilerbert73@cau.ac.kr (J.W. Moon).

https://doi.org/10.1016/j.buildenv.2021.108704
Received 29 September 2021; Received in revised form 17 December 2021; Accepted 20 December 2021
Available online 23 December 2021

Nomenclature

A	area (m^2)
C	heat capacity (kcal/hr°C)
Capacity	capacity of the chiller (kW)
COP	coefficient of performance
c_p	specific heat at constant pressure (kcal/kg°C)
ε	effectiveness of heat exchanger
m	flow rate of heat transfer medium (kg/hr)
NTU	number of transfer units
P	electricity energy usage (kW)
PLR	partial load ratio
Q	heat transfer rate (kcal/hr)
T	temperature (°C)
U	coefficient of over-all heat transmission (kcal/mhr°C)
y	output value of input data
$\widehat{y}$	output of predictive model
$\overline{y}$	average of output

Superscripts

i, k	index of number

Subscripts

air	air as heat transfer medium
CW	indication of chilled water
CRAH	indication of computer room air handler
HAC	hot aisle containment
IT	indication of IT equipment
in	inlet
max	maximum
min	minimum
n	index of number
(n)	present timestep
(n-1)	past timestep
out	outlet
pred	predicted value
rack	indication of rack
set	setpoint
tran	actual transferred heat

systems [8,9], use of containment systems [10–12], introduction of high-efficiency IT equipment and cooling systems [4,13–16], improvement of airflow distribution systems [17–19], and outdoor air utilization using economizer mode [20–22]. Software-based approaches include efficient IT resource distribution [23–25], IT resource and cooling system scheduling [26–28], and optimal system control [29–31]. Both types of strategies effectively conserve energy. However, hardware-based strategies can only be applied to new or remodeled data centers, and once IT equipment is installed, it is difficult to change its location. Additionally, since the coefficient of performance (COP) of a cooling system varies according to the surrounding climate and operates within a specified cooling capacity, a software-based approach is required for optimal operation.

1.2.2. Conventional control methods

Software-based approaches generally focus on the optimal operation of installed systems. ON/OFF and PID control are commonly employed for controlling cooling systems [32], but these methods are designed to correct the temperature based on the feedback process when it deviates from the setpoint temperature. Consequently, fluctuations of about 1–2 °C occur [33] because the deadband establishes upper and lower boundaries of ±1 °C at the setpoint temperature for the ON/OFF system and the controller operates in proportion to the generated error for PID control. Fluctuations exceeding the recommended ranges, such as 18–27 °C from American Society of Heating, Refrigerating and Air-conditioning Engineers (ASHRAE) Technical Committee 9.9 class A1 [34], can increase server failure rates.

1.2.3. Model predictive control and predictive models

The abovementioned problems can be prevented using model predictive control (MPC) [2]. MPC is used to predict future conditions by sampling data at set intervals, causing the establishment of optimal control methods through the optimization processes. Optimal control minimizes the fluctuations and enables stable control. Several studies have proven that MPC is superior to existing ON/OFF, PI, and PID control in terms of minimizing the energy consumption and enabling stable control [35–37]. The most important part in developing MPC is the predictive model because MPC performs optimal control based on the prediction results. Thus, many studies have been conducted to develop accurate predictive models for data centers.

Parolini et al. [39] conducted mathematical modeling of a data center. The computational and thermal networks were modeled separately, and the energy consumption of the cooling system was defined as a function of the inlet and outlet of the computer room air conditioner (CRAC). It was assumed that the cooling tower was composed of nodes of CRACs. MPC simulations were conducted at the data-center level. As a result, MPC exhibited superior PUE compared to non-MPC approaches. Ogura et al. [40] also conducted mathematical modeling in which the server outlet temperature was defined using the server inlet temperature, server heat generation, and server energy consumption as independent variables, while COP regression was used for the cooling system. They subsequently reported that MPC minimized the energy consumption by optimizing server resource allocation.

Computational fluid dynamics (CFD) modeling has also been employed to provide more accurate airflow and heat distribution analysis. Marchall and Bemis [41] presented a CFD-based data center modeling method. The modeling of CRACs, fans, UPS, and raised floors was achieved using numerical methods, and summarized the factors that need to be considered in the modeling process. Additionally, Gao et al. [42] demonstrated the superiority of cold aisle containment by conducting airflow analysis using CFD modeling, while Watson and Venkiteswaran [43] used CFD simulations to demonstrate that a water-cooling system possesses a superior cooling capacity compared to that of an air-cooling system.

Both mathematical and CFD models presented high accuracy for the target model. However, although mathematical models operating within MPC can effectively reduce the energy consumption, considering all the relevant variables during modeling is difficult. An approximative model including mathematical modeling can be used to simulate the heat exchange process or energy consumption of a target data center, but it cannot easily reflect system aging and changes to the data center environment. Till date, CFD models have mainly focused on airflow analysis, heat exchange process analysis, and simulations of various environments rather than MPC or real-time control. This may be due to the fundamental characteristics of CFD modeling, which heavily depends on the boundary conditions that are input into the model; the results vary according to the size of the mesh and the calculation. Moreover, objectively assessing a model's accuracy is difficult because no criteria exist for comparing accuracy unless actual environmental data are provided and modifying the model using feedback is challenging when applying real-time control. Additionally, the more complex the model, the slower is its operation for iterative optimization; in other words, for real-time applications, high-performance computers are required [44].

Data-driven methods can overcome these disadvantages and allow

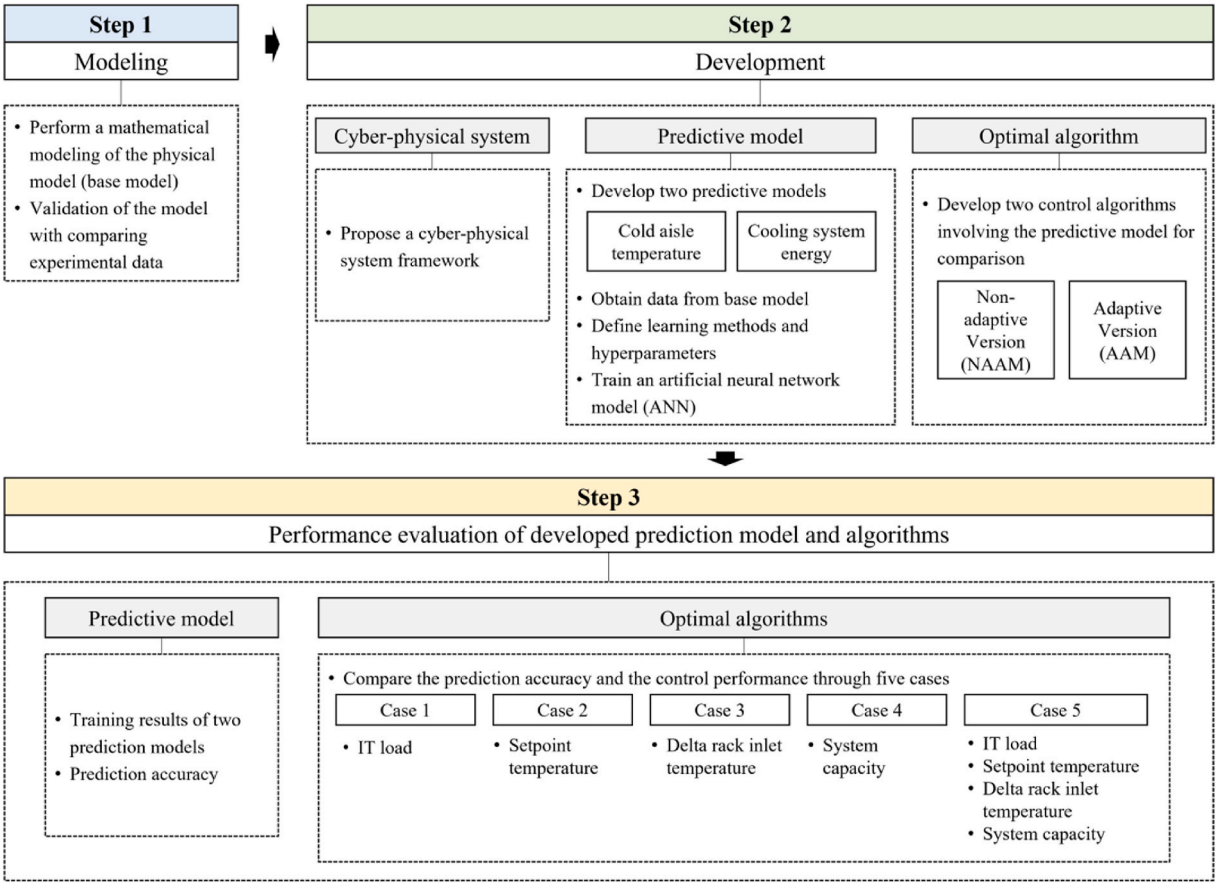


Fig. 1. Study method.

for relatively easy modeling. In this approach, specialized knowledge is not required because the correlations between the input and output are determined based on simulations and field data. To date, data-driven models have primarily employed statistical analysis and machine learning [45]. For example, Lei and Masanet [46] performed thermodynamic-based PUE modeling using statistical analysis to predict the PUE of hyper-scale data centers, while Saiyad et al. [44] predicted server inlet and outlet temperatures using an artificial neural network (ANN), with a prediction accuracy in terms of the mean squared error (MSE) of 0.58 at the rack level. Song et al. [47] also developed a model to predict the average airflow rates and the temperature discharged from tiles using an ANN. By applying this model to a genetic algorithm (GA), optimal control variables were derived by inversely predicting the operating conditions to satisfy a specific output variable. Athavale et al. [45] developed four machine learning models, ANN, Support vector regression, Gaussian process regression, and Property organization, to predict the rack inlet air temperature, and CFD simulations were conducted to compare the prediction performance. Asgari et al. [48] developed a gray-box model comprising an ANN that predicts the pressure to reflect the thermodynamic principles or system's operating state. Most of these studies adopted ANN models because they afford strong predictive performances for nonlinear relationships and the inputs and outputs can be freely configured.

Although previous studies have proven the superiority of data-driven models, they exhibit performance difficulties when extended to sites differing from the target building or when the performance changes due to system aging [31]. To solve this problem, model calibration and optimization processes are required, which increases maintenance costs and can cause a lack of control stability if model calibration is not regularly conducted. Adaptation and self-tuning technologies allow the prediction models to adapt to environmental changes by modifying the hyperparameters affecting the performance of the prediction model based on the difference between the predicted and actual values (i.e., the error). The development of MPC with adaptation has been investigated in various fields [49–52], but few studies have focused on the control of data centers [31].

1.3. Objectives

According to the previous research analysis, three main issues were derived to be solved: (1) A simple but accurate predictive model is required for model predictive control. (2) For the applicability and scalability of MPC, the predictive model must be able to adapt to the applied system. (3) In order to provide a stable thermal environment to the data center, an advanced platform for operating MPC is required.

Therefore, the purpose of this study is to develop an adaptive ANN predictive model (AAM) and an optimal control algorithm for the data center cooling system as well as proposing an advanced operation platform based on the cyber-physical system (CPS). The AAM model predicts the rack inlet temperature and the cooling system energy usage of the applied data center and is then included in the optimal control algorithm with a real-time training function that allows it to adapt to the environment. The application of the AAM and optimal control algorithm at the containment level in a data center can reduce the additional processes required for local optimization by removing individual calibration processes and can conserve energy by efficiently utilizing resources at the global level, such as chillers and cooling towers. Furthermore, if scalability is possible by diversifying the input data, optimal control can be achieved even when applied to various heat source systems.

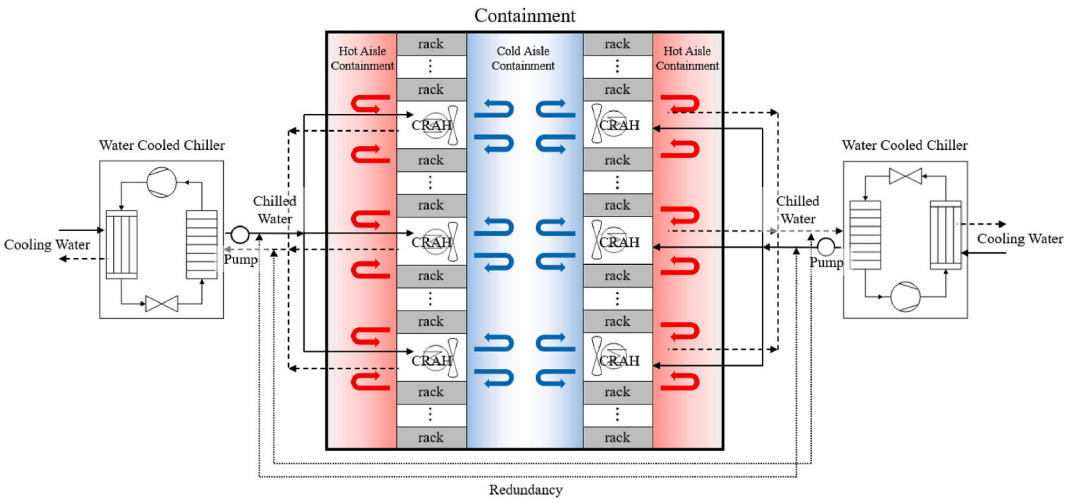


Fig. 2. Reference model of a containment-type data center.

Fig. 3. Images of the containment system, CRAH, and water-cooled chiller in the reference containment data center.

2. Methodology

2.1. Research method

Fig. 1 summarizes the study method. In Step 1, a base model was developed through mathematical modeling to represent an actual containment data center. The base model was utilized for data acquisition and evaluating the performance of AAM and the optimal control algorithm. Step 2 describes the development processes of the CPS framework, AAM, and optimal algorithms. Two predictive models were developed that predict the cold aisle temperature and cooling system energy. Additionally, two optimal algorithms were developed to represent non-adaptive ANN model (NAAM) and AAM. NAAM was introduced to compare the adaptive performance of the two algorithms. In Step 3, the performance of the predictive model and algorithm was evaluated. The prediction accuracy and control performance of NAAM and AAM were compared by conducting simulations for five cases representing changes in the data center environment. The entire process, including development and performance evaluation, was performed in Python.

2.2. Description and modeling of the physical system

Figs. 2 and 3 present a schematic diagram and images, respectively, of the containment-type data center that is used as a testbed in the present study. Both the cold aisle containment (CAC) and hot aisle containment (HAC) are blocked by the containment system sharing CAC in the center. Each row comprises racks and computer room air handers (CRAHs), which are placed between the racks in an in-row arrangement. Racks are arranged with up to 18 EA and three CRAHs per row.

A CRAH discharges cold air from each row to the CAC system. The cold air is drawn in through the front of the rack to cool the server and expelled to the back of the rack. The discharged air is returned to the CRAH, cooled by a cooling coil, and discharged back to the CAC system. The temperature of the air supplied by the CRAH is controlled by the flow rate of the chilled water in the cooling coil, with the chilled water transferring heat to the evaporator in the water-cooled chiller. An inverter compressor in the water-cooled chiller responds to partial loads so that chilled water is provided at a constant temperature.

Mathematical modeling was conducted based on the specification tables for the equipment installed in the physical model and a few associated assumptions. The main assumptions are as follows: (1) the input power to the IT equipment can be converted to the amount of the heat generated and (2) the air in the CAC and HAC is assumed to be well-mixed. These assumptions facilitate the calculations without compromising the reliability of the mathematical model. For accurate modeling of the thermal environment, the mathematical model consisted of three main sections: air–server heat exchange in the rack, air–water heat exchange in the cooling coil, and the energy consumption of fans, pumps, and chillers.

The air–server heat exchange section of the model assumed that the air removed the entire load generated by the server. In the actual

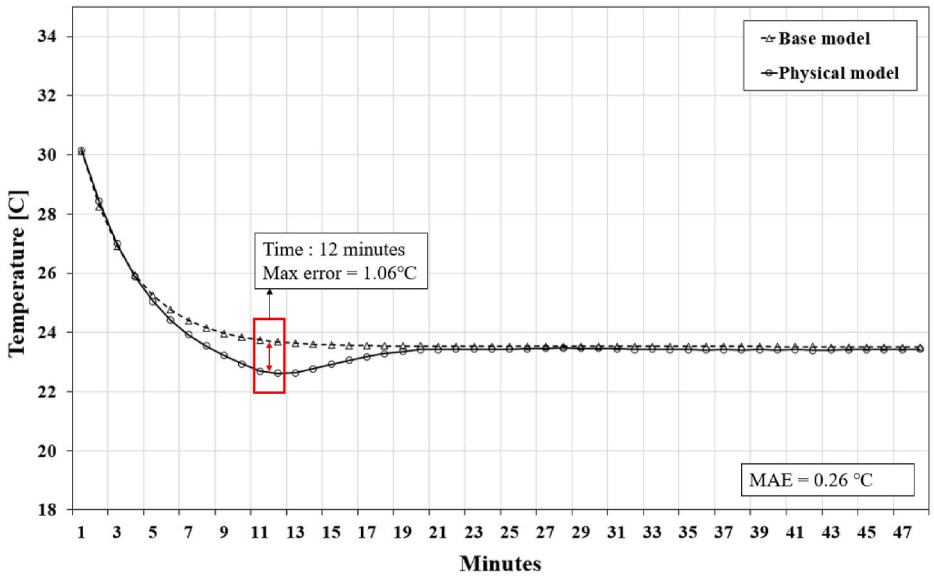


Fig. 4. Validation results for the base and physical models.

physical model, differential pressure control was used to maintain the CAC at a higher pressure than the HAC, and the amount of air discharged from the CRAH was determined based on this. However, in the present study, it was assumed that the flow rate of the discharged air was equal to that of the returned air. The delta temperature between the CAC and HAC systems was set at 6 °C based on the average of the measured data. Due to these assumptions, the airflow was proportional to the generated IT load and was controlled to maintain the delta temperature between the CAC and HAC systems at 6 °C. Equations (1) and (2) describe this relationship:

$$Q_{rack} = Q_{rack,air} = m_{rack,air} \bullet c_{p,air} \bullet \left(T_{rack,out} - T_{rack,in}\right) \tag{1}$$

$$m_{rack,air} = Q_{rack} \,/\, \left\{c_{p,air} \bullet \left(T_{rack,out} - T_{rack,in}\right)\right\} \tag{2}$$

where Q_{rack} [kcal/hr] is the load generated by a rack, $Q_{rack,air}$ is the load in the air leaving the rack, $m_{rack,air}$ and $c_{p,air}$ are the airflow rate of each rack [kg/hr] and the specific heat of air [kcal/kg•°C], respectively, and $T_{rack,out}$ and $T_{rack,in}$ [°C] are the outlet and inlet air temperature, respectively, of the rack.

CRAHs in each row must supply 50% of the CAC airflow rate to maintain the pressure between HAC and CAC. Therefore, the airflow rate of the returning CRAH air is calculated as a third of the total airflow for the HAC systems, as defined in Equation (3):

$$m^{i}_{CRAH} = \frac{1}{3} m_{HAC} = \frac{1}{3} \sum_{k=1}^{n} m^{k}_{rack} \tag{3}$$

where m^{k}_{rack} [kg/hr] is the airflow rate for the k^{th} rack, m_{HAC} [kg/hr] is the airflow rate for the HAC systems, m^{i}_{CRAH} [kg/hr] is the airflow rate for the i^{th} CRAH, and n is number of racks in each row.

The water–air heat exchange process in the cooling coil was modeled based on the effectiveness−number of transfer units (ε-NTU) method [54], which includes information on the heat transfer rate and heat exchange area of the heat exchanger, which is U•A = 630 kcal/h•°C according to the on-site system capacity. Here, effectiveness refers to the heat exchange efficiency and is defined as the heat capacity of the NTU and heat transfer medium. The mathematical relationship between effectiveness and the NTU is defined as follows in Equations (4)–(6):

$$C = \frac{C_{min}}{C_{max}} \tag{4}$$

$$NTU = \frac{U \bullet A}{C_{min}} \tag{5}$$

$$\varepsilon = \frac{1 - \exp[-NTU \bullet (1 + C)]}{1 - C \bullet \exp[-NTU \bullet (1 - C)]} \tag{6}$$

where C_{min} and C_{max} are the heat capacity of air and chilled water, respectively, C is the heat capacity ratio between C_{min} and C_{max}, NTU is the number of transfer units, U [kcal/hr•m^2•°C] is the heat transfer coefficient of the cooling coil, A [m^2] is the heat transfer area of the cooling coil, and ε is the effectiveness of the heat exchanger. In the heat exchanger, the actual heat transfer rate is equal to the maximum heat transfer rate multiplied by the effectiveness. Using the actual heat transfer rate, the air outlet temperature and the chilled water outlet temperature can be calculated as a result of the heat exchange, and these are substituted into the CRAH supply air temperature and chilled water return temperature in the subsequent time step in the simulation. This process can be summarized as Equations (7)–(10):

$$Q_{max} = C_{min} \bullet \left(T_{rack,out} - T_{CW,out}\right) \tag{7}$$

$$Q_{tran} = \varepsilon \bullet Q_{max} = \varepsilon \bullet C_{min} \bullet \left(T_{rack,out} - T_{CW,out}\right) \tag{8}$$

$$T_{rack,in} = T_{rack,out} - \frac{Q_{tran}}{C_{min}} \tag{9}$$

$$T_{CW,out} = T_{CW,in} - \frac{Q_{tran}}{C_{max}} \tag{10}$$

where Q_{max} [kcal/hr] is the maximum heat transfer rate, Q_{tran} [kcal/hr] is the actual heat transfer rate, and $T_{CW,in}$ and $T_{CW,out}$ [°C] are the inlet and outlet chilled water temperature, respectively.

The energy-consuming elements in this cooling system are the CRAH fans, circulating pumps, and chillers. Each CRAH has four fans, and one pump and one chiller are installed per row. The energy consumption of each component is calculated using a regression equation. The part load ratio (PLR) for the fans is taken from ASHRAE Standard 90.1 Appendix G, and the energy consumption regression equation for the pump was calculated according to the chilled water mass flow rate based on the information from the product datasheet. The energy consumption of the chiller was calculated by referring to the energy input ratio graph for the PLR [55]. The regression equation for each element is defined in Equations (11)–(13):

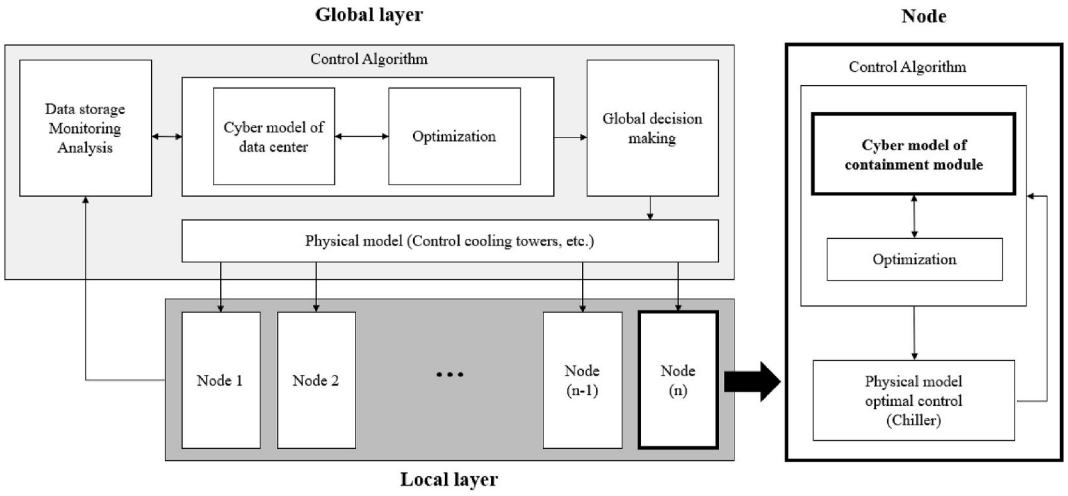


Fig. 5. Conceptual diagram of data center CPS.

$$P_{fan} = 0.51 \cdot \left(0.0013 + 0.1470 \cdot \mathrm{PLR}_{\mathrm{fan}} + 0.9506 \cdot \mathrm{PLR}_{\mathrm{fan}}^2 - 0.0998 \cdot \mathrm{PLR}_{\mathrm{fan}}^3\right) \tag{11}$$

$$P_{pump} = 0.15 \cdot \mathrm{m}_{\mathrm{CW}} + 3 \tag{12}$$

$$P_{chiller} = \frac{Capacity}{COP} \cdot \left(0.8531 \cdot \mathrm{PLR}_{\mathrm{chiller}}^2 - 0.1659 \cdot \mathrm{PLR}_{\mathrm{chiller}} + 0.3057\right) \tag{13}$$

where P_{fan} [kW] is the power of the fan, PLR_{fan} is the PLR of the fan, P_{pump} is the power of the pump, m_{CW} is the mass flow rate of chilled water [kg/hr], $P_{chiller}$ [kW] is the power of the chiller, *Capacity* is the capacity of the chiller [kW], *COP* is the coefficient of performance, and $PLR_{chiller}$ is the PLR of the chiller. PLR_{fan} is defined as the ratio of the current airflow rate to the maximum airflow rate, and $PLR_{chiller}$ is defined as the ratio of the capacity to the amount of heat removed from the returned chilled water.

To validate the base model, experimental data were acquired from the referenced physical model in Daejeon, South Korea, with the containment system located in the data center building. The computer room holding the containment system is maintained at a constant temperature throughout the year and operates without interruption. A scenario-based experiment was performed to determine the suitability and stability of the containment cooling system. Three CRAHs were operated with an IT load of 72 kW, and the initial CAC temperature was 30.12 °C. Chilled water at 8.5 °C was constantly supplied from the chiller at 9400–9900 kg/h, and the temperature of the returned chilled water was 12.7 °C. The temperature convergence point for the CAC under these conditions was identified, and experimental data were obtained when the CAC temperature was maintained at a steady state.

The base model was run under the same environmental conditions as those for the experiments conducted in the referenced physical model; a 50-min simulation was conducted. The CAC temperature convergence processes for the experimental data and simulation results were compared (Fig. 4). A maximum error of 1.06 °C occurred after 12 min, which is the transient response period; then, the error reached <0.5 °C after 16 min, and both models attained nearly 23.5 °C after 20 min. The mean absolute error (MAE) for all data was 0.26 °C, representing a relatively low error, even when accounting for the transient response period. Therefore, the base model was highly similar to the referenced physical model and the heat exchange environment could be simulated in a stable manner.

2.3. CPS framework

The technical problems associated with the practical application of MPC in the field can be solved by applying cyber–physical systems (CPSs), which have recently attracted attention as a method for overcoming the problems associated with high-performance computing resources and network deployment. A CPS is a holistic communication network comprising either hardware or software components for facilitating information exchange (e.g., sensors, actuators, and energy consumption) between the cyber and physical systems [38]. Data storage and computation are handled by a server affording a large storage capacity and high computational speeds. Thus, the prediction accuracy and control stability of the cyber system are further emphasized.

Real-time data transmission and high-performance computing resources available in a CPS allow for rapid data processing and learning times in the real-time training of an AAM. This enables AI-based control, which is difficult to achieve with existing control boards due to performance limitations, such as insufficient memory and slow operating speeds. Moreover, real-time training facilitates rapid adaptation to a target environment. These technical advantages can be realized by installing a minimum of IT equipment and sensors without changing the existing servers and cooling system. Despite the data center characteristics that are difficult to remodel, the energy efficiency can be improved by maximizing the utilization of the provided resources by applying CPS.

Various types of CPS framework have been presented, but few cases have been applied to the control of data center cooling systems. In this section, based on the distributed framework system (DFS) method proposed by Villalonga et al. [53], a CPS framework suitable for containment data centers is proposed. The DFS method comprising global and local layers allows the application of CPS to hierarchical systems, such as the cooling system of a data center. Fig. 5 presents the CPS framework. The global layer is applied at the data center level, increasing the energy efficiency of the entire data center via schedule optimization, facility management (e.g., cooling towers), and server provisioning. The local layers are applied to individual containers as nodes, which acquire data from the sensors installed within each containment system and immediately transmit them to the AAM. The AAM predicts the future rack inlet temperature based on the transmitted data, and the predicted result is used to derive the optimal mass flow rate of the chilled water. Moreover, the acquired data are transmitted to the global layer and used for global decision making.

2.4. Development of the AAM

An ANN is a supervised machine learning tool that consists of neurons or nodes that mimic human brain neurons and is employed to learn the relationship between given input and output data [56]. In general, an ANN consists of an input layer, hidden layers, and output layer. The

Table 1
Variable range for data generation.

Variable	Range	Interval Scale	Unit
P_{IT}	50–80	5	[kW]
$T_{rack,out}$	24–33	1	[°C]
m_{CW}	500–40,000	500	[kg/hr]

input and output layers each contain the same number of neurons as the number of input and output variables, while the hidden layers have an arbitrary number of layers and neurons. Each neuron multiplies the input value by a particular weight and produces an output value through an applied activation function. This output is then fed to the next layer as an input. This relationship is defined in Equation (14).

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{14}$$

where x_i is the input data at the node, ω_i and b are the weight and bias allocated to the neuron, respectively, f is the activation function that is adjusted to the neurons, and y is the output value. The present study employed the Rectified Linear Unit (ReLU) as the activation function. The ReLU is commonly used in ANNs because it offers fast learning, low computational costs, and simple implementation. If the operation value of the neuron is ≤ 0, 0 is used as the output, and if it is > 0, the output value is calculated using a straight line with a slope of 1.

The Levenberg–Marquardt (LM) algorithm was used for ANN training. The backpropagation-based LM algorithm minimizes the cost by modifying the weights of the neurons based on iterative learning. If the error between the cost and the target value is large, the solution can be obtained using the gradient descent method, and if the error is small, the solution converges to the minimum cost using the Gauss–Newton method. In this case, the cost function represents the error function between the predicted value of the input data and the answer. In this study, the MSE was set as the cost function:

$$MSE = \frac{1}{n} \bullet \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{15}$$

Where n is number of data, y_i and $\hat{y}_i$ are the output and predicted value, respectively, of the ith training epoch.

To ensure the accuracy of the predictions, it is important to use input data that is highly correlated with the target output. Reducing the dimensionality of the input data is also essential for real-time prediction and training because computational speeds must be considered [47]. In the present study, two versions of the AAM were developed to predict either the rack inlet temperature or the cooling system energy based on the same input variables. The predicted rack inlet temperature can be directly used by the control system, while the predicted cooling system energy assists in decision-making regarding the optimal chilled water mass flow rate derived from the control algorithm. Therefore, environmental and system variables directly related to the rack inlet temperature and cooling system energy were employed as input variables for the AAM which are able to be obtained during the real-time operations.

Variable selection was based on Equations (1)–(10). The rack inlet temperature was determined by $T_{rack,out}$, Q_{tran}, and C_{min}. Here, Q_{tran} was determined by ε and Q_{max}. Because $T_{CW,out}$ was constant at 8.5 °C, Q_{tran} is greatly affected by ε, which is a function of NTU and C. Because these values are determined by C_{min} and C_{max}, the airflow rate and chilled water mass flow rate have a fundamental effect on the rack inlet

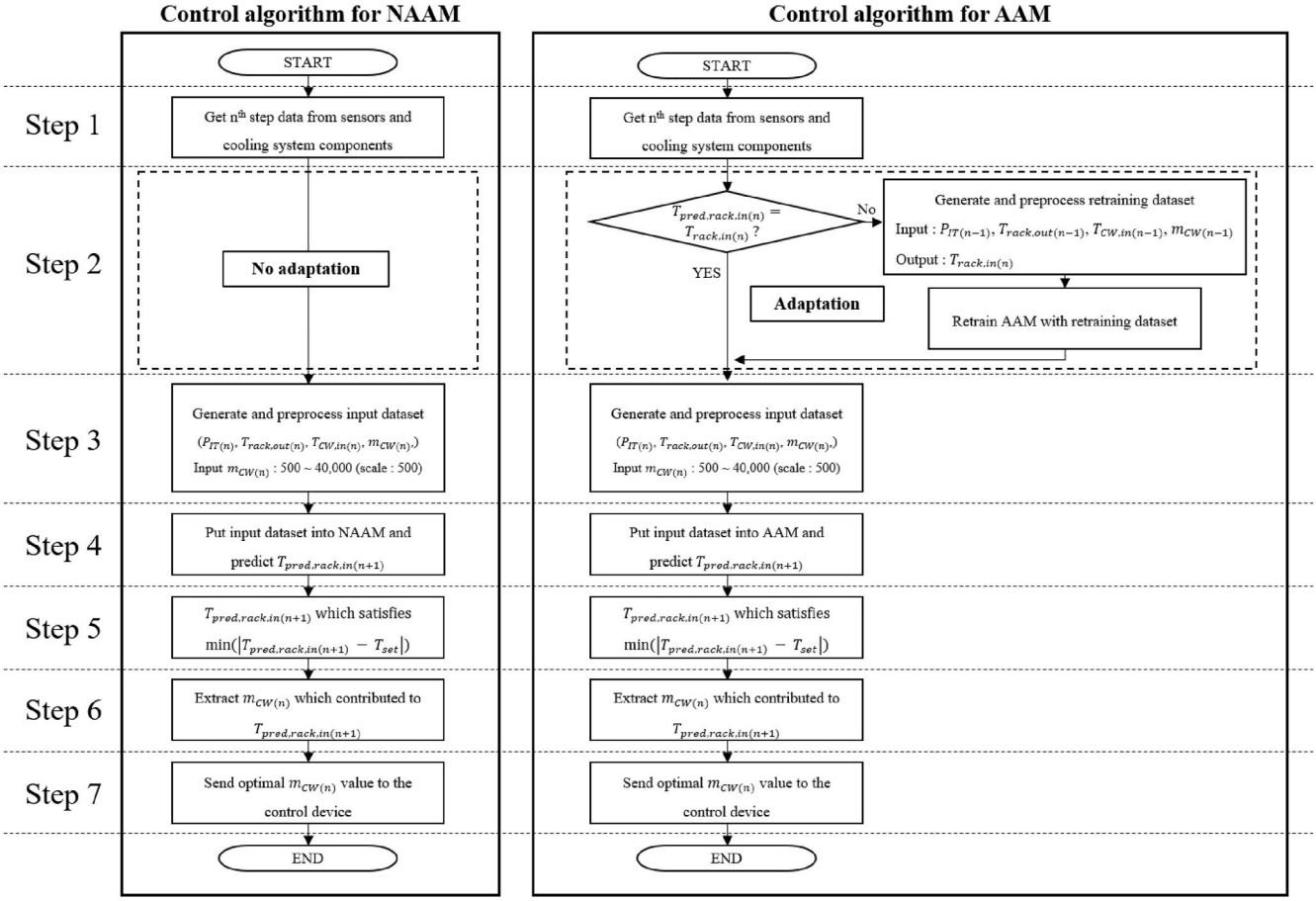


Fig. 6. Optimal control algorithms for the NAAM and the AAM.

temperature. For the CRAH fans, the PLR was calculated based on the change in the airflow rate caused by the change in the IT load (P_{IT}), and the pump power was calculated according to m_{CW}, meaning that it shares the same variable used to calculate the rack inlet temperature. Becausem_{CW}, $T_{CW,out}$, and $T_{CW,in}$ determine the chiller PLR, $T_{rack,out}$, P_{IT}, m_{CW}, and $T_{CW,in}$ were used as the input data.

The dataset for ANN training was generated using the base model. To generate the data, one of P_{IT}, $T_{rack,out}$, and m_{CW} was varied following a set scale and the other two variables were fixed. As a result, 5600 data points were acquired. Table 1 shows the control range for each variable. The acquired data were pre-processed using the min–max normalization method and split into training data (60%), validation data (10%), and test data (10%) for ANN learning.

ANN learning requires initial hyperparameters, such as the number of hidden layers and neurons and the learning rate, to be set in advance, but no specific guidelines exist for this. Therefore, herein, the optimal combination of hyperparameters was determined by applying Bayesian optimization during the ANN learning stage. The Bayesian optimization includes a surrogate model and an acquisition function. The surrogate model performs a probabilistic estimation of the objective function in a black-box manner based on the processed data, and a Gaussian process is used as the probabilistic model. Thereafter, the acquisition function recommends the subsequent combination of input values based on the objective function from the surrogate model, while the optimizer searches for the optimal hyperparameter combination by repeating this process. The search range for the hidden layers, nodes, and learning rate was set as 1–5, 1–32, and 1e-2, 1e-3, or 1e-4, respectively.

2.5. Optimal control algorithm

The optimal control algorithm was applied to the individual containment systems at the local layer level. The algorithm included the AAM and optimization process and controlled the CAC temperature by determining the optimal mass flow rate for the chilled water. This reduced the unnecessary use of cold water, thus conserving energy within the cooling system. This section outlines the two control algorithms used for a comparison of the performance of the NAAM and AAM.

The two control algorithms were divided into seven steps (Fig. 6). The details of each step are as follows:

1. Step 1: Data for the nth time step were acquired from the sensors and cooling system. The variables were the IT load, inlet and outlet rack temperature, inlet and outlet chilled water temperature, chilled water mass flow rate, fan power, pump power, and chiller power. These data were transmitted and saved to a local server and the global layer.
2. Step 2: Adaptation to the environment was implemented with the AAM but not with NAAM. AAM was retrained by the control algorithm using real-time online training. During the real-time training, the weights and bias of AAM were continuously updated at each time step. The rack inlet temperature ($T_{pred,rack,in(n)}$) predicted at the $n-1$st time step and the nth rack inlet temperature acquired at the nth time step ($T_{rack,in(n)}$) were compared. If the two values were identical, the algorithm proceeded to Step 3; if the two values differed, a dataset for retraining was created. The retraining dataset comprised $P_{IT(n-1)}$, $T_{rack,out(n-1)}$, $T_{CW,in(n-1)}$, and $m_{CW(n-1)}$; it was used for prediction at the n-1st time step and $T_{rack,in(n)}$ for the output data. This process updated the weights and bias of the AAM by learning that the predicted value for the current environment of the n-1st input data was $T_{rack,in(n)}$ and not $T_{pred,rack,in(n)}$.
3. Step 3: An input dataset for predictions was created. $P_{IT(n)}$, $T_{rack,out(n)}$, and $T_{CW,in(n)}$ were extracted from the data acquired in Step 1. For the extracted data, $m_{CW(n)}$ in the 500–40,000 kg/h (scale: 500 kg/h) range was appended to yield 79 data points.
4. Step 4: By feeding input data into the NAAM or AAM, $T_{pred,rack,in(n+1)}$ values for each $m_{CW(n)}$ for the current environment were obtained.
5. Step 5: The 79 predicted values were compared with the setpoint temperature (T_{set}) to determine the final $T_{pred,rack,in(n+1)}$, for which the error between the two values was minimized.
6. Step 6: The $m_{CW(n)}$ value to be used for the control was determined by extracting input $m_{CW(n)}$, which was appended to the input data used to predict the determined $T_{pred,rack,in(n)}$ in Step 5. The determined $m_{CW(n)}$ was the optimal mass flow rate for the chilled water that satisfied T_{set} in the NAAM or AAM criteria.
7. Step 7: Control was executed by transmitting commands to the physical model to operate the cooling system at the determined optimal chilled water mass flow rate.

Table 2
Simulation cases for evaluating the performance of the control algorithm.

Case	Variable	Setting
Base model	IT load	60 kW
	Setpoint temperature	20 °C
	Rack inlet temperature	+0 °C
	CRAH capacity	24 kW
	Chiller capacity	120 kW
Case 1	IT load	60 kW, 65 kW, 70 kW
Case 2	Setpoint temperature	20 °C, 21 °C, 22 °C
Case 3	Delta rack inlet temperature	+1 °C, +2 °C, +3 °C
Case 4	CRAH capacity	CRAH capacity: from 24 kW to 48 kW
	Chiller capacity	Chiller capacity: from 120 kW 240 kW
Case 5	IT load	120 kW, 125 kW, 130 kW
	Setpoint temperature	22 °C, 21 °C, 20 °C
	Delta rack inlet temperature	+2 °C
	CRAH capacity	48 kW
	Chiller capacity	240 kW

The performance of the two algorithms was evaluated by applying five cases to the base model to evaluate the prediction accuracy, adaptability, and control stability according to the application of NAAM and AAM. Each case represented an environmental change. The variables subject to change were the IT load, setpoint temperature, delta rack inlet temperature, and cooling system capacity. In the final case, all of the modified variables were collectively applied to the base model, simulating the application of the AAM to a new environment. Table 2 summarizes the initial values for the base model and details for the five cases.

In Cases 1 and 2, the IT load and setpoint temperature were changed, respectively. These represent situations in which the IT load changes over time and the setpoint temperature is adjusted in the base model. The IT load was increased from 60 kW to 70 kW in 5 kW increments, while the setpoint temperature was increased from 20 °C to 22 °C in 1 °C increments.

Case 3 represents changes in the physical environment of a containment data center. In general, the air supply for the CRAH is introduced into the rack at a higher temperature due to heat exchange with the surrounding environment while entering the server. The important effect of the ambient environment on the rack inlet temperature has been reported by Sun et al. [57]. In particular, the surrounding environment, including the air distribution system, size of the containment area, the IT load, and the location of the cooling system, affects the rack inlet temperature. Thus, the delta rack inlet temperature was adjusted in the base model by +1 °C, +2 °C, and +3 °C. Case 4 doubled the cooling capacity of the CRAH and the chiller in the base model. Even within the same data center, the load differs depending on the number of racks installed inside the containment system, and a cooling system with a different capacity can be applied for this purpose.

In Case 5, the variables that were adjusted for Cases 1–4 were all included together to represent the use of the AAM in a new containment environment. The IT load was increased from 125 kW to 130 kW and the setpoint temperature was set at 22 °C, 21 °C, and 20 °C in order. The

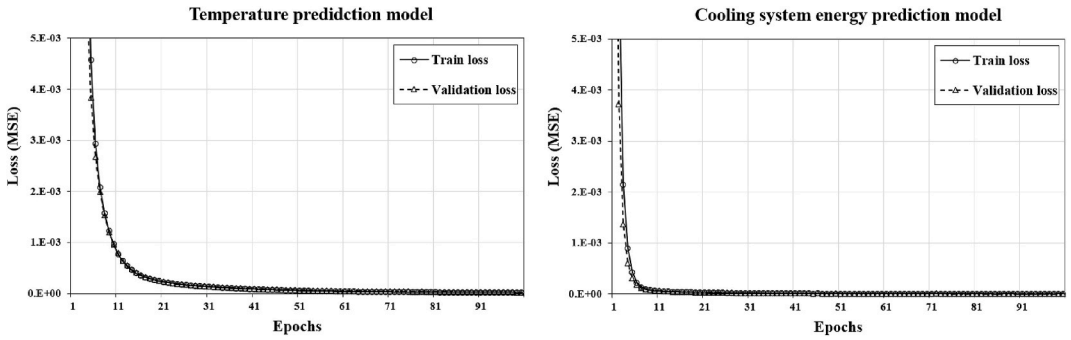


Fig. 7. Training stability of the AAM for temperature and cooling system energy predictions.

Table 3
Structure and accuracy of the optimized AAM.

Parameters	Temperature prediction	Cooling system energy prediction
Optimization Method	Bayesian Optimization	Bayesian Optimization
Structure and number of neurons	Input layer (4) Hidden layer 1 (31) Hidden layer 2 (31) Output layer (1)	Input layer (4) Hidden layer 1 (21) Hidden layer 2 (13) Output layer (1)
Performances	$R^2 = 0.9988$ RMSE = 0.0598 °C CV(RMSE) = 0.27%	$R^2 = 0.9999$ RMSE = 0.0237 °C CV(RMSE) = 0.14%

delta rack inlet temperature was set at +2 °C, and the CRAH and chiller capacity had the same values as in Case 4.

3. Results and discussion

3.1. Training results for the optimized AAM

The AAM was developed to predict the rack inlet temperature and the cooling system energy using the same input dataset. Fig. 7 presents the trend in the MSE across the training epochs for the prediction of these variables. The MSE for the temperature prediction version of the AAM using the training data converged to 1.8852e-5, while that using the validation data converged to 2.5772e-05. The MSE for the cooling system energy prediction version of the AAM using the training and validation data converged to 5.2385e-6 and 5.4657e-6, respectively. In both versions of the model, overfitting did not occur, and the learning was stable.

The prediction accuracy of the AAM was evaluated using R^2, the root mean square error (RMSE) and the coefficient of variation of the RMSE (CV(RMSE)) based on the actual and predicted values for the test data. R^2 is calculated based on a scatter graph that indicates how linear the relationship between the actual and predicted values is. The closer R^2 is to 1, the higher the correlation between the two values. RMSE is an index that takes the square root of MSE and converts the error index into units and scales similar to the actual values. The closer the RMSE is to 0, the higher the prediction accuracy of the model. CV(RMSE) is an index that calculates the coefficient of variation by dividing the RMSE by the average of the actual values and is expressed in % by reflecting the difference from the average value. The closer the CV(RMSE) is to 0%, the higher the accuracy, with ASHRAE Guideline 14 recommending a CV (RMSE) of <25% for building energy predictions [58]. The mathematical definition of each indicator is defined in Equations (16)–(18):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y-\hat{y})^2}{\sum_{i=1}^{n}(y-\bar{y})^2} \tag{16}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y-\hat{y})^2}{n}} \tag{17}$$

$$CV(RMSE) = \frac{1}{\bar{y}}\sqrt{\frac{\sum_{i=1}^{n}(y-\hat{y})^2}{n}} \times 100 \tag{18}$$

where y is the actual value, $\hat{y}$ is the value predicted by the model, $\bar{y}$ is the average of the actual values, and n is the number of data points.

Table 3 summarizes the structure and prediction accuracy of the two versions of the AAM following Bayesian optimization. The AAM for the prediction of the rack inlet temperature had two hidden layers and 31 neurons in each layer, which led to a high prediction accuracy of R^2 = 0.9988, RMSE = 0.0598 °C, and CV(RMSE) = 0.27%, which was close to 0%. The AAM for the prediction of the cooling system energy had two hidden layers, one with 21 neurons and the other with 13. The R^2, RMSE, and CV(RMSE) were 0.9999, 0.0237 °C, and 0.14%, respectively, representing high prediction accuracy.

Fig. 8 presents the predicted values for the test data from the two

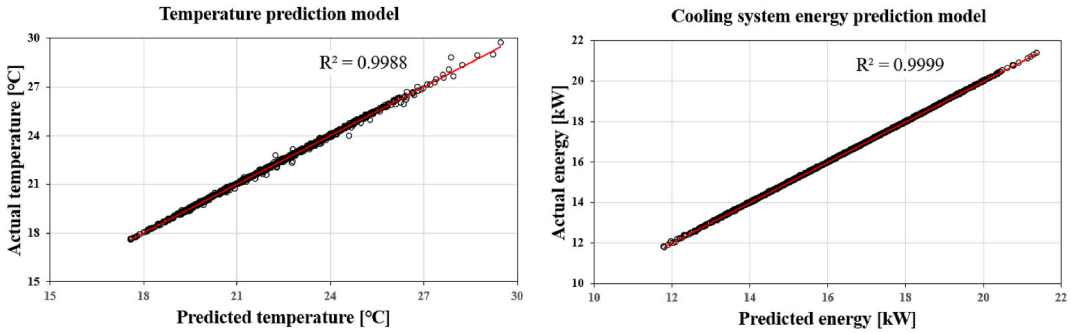


Fig. 8. R^2 for the AAM predictions.

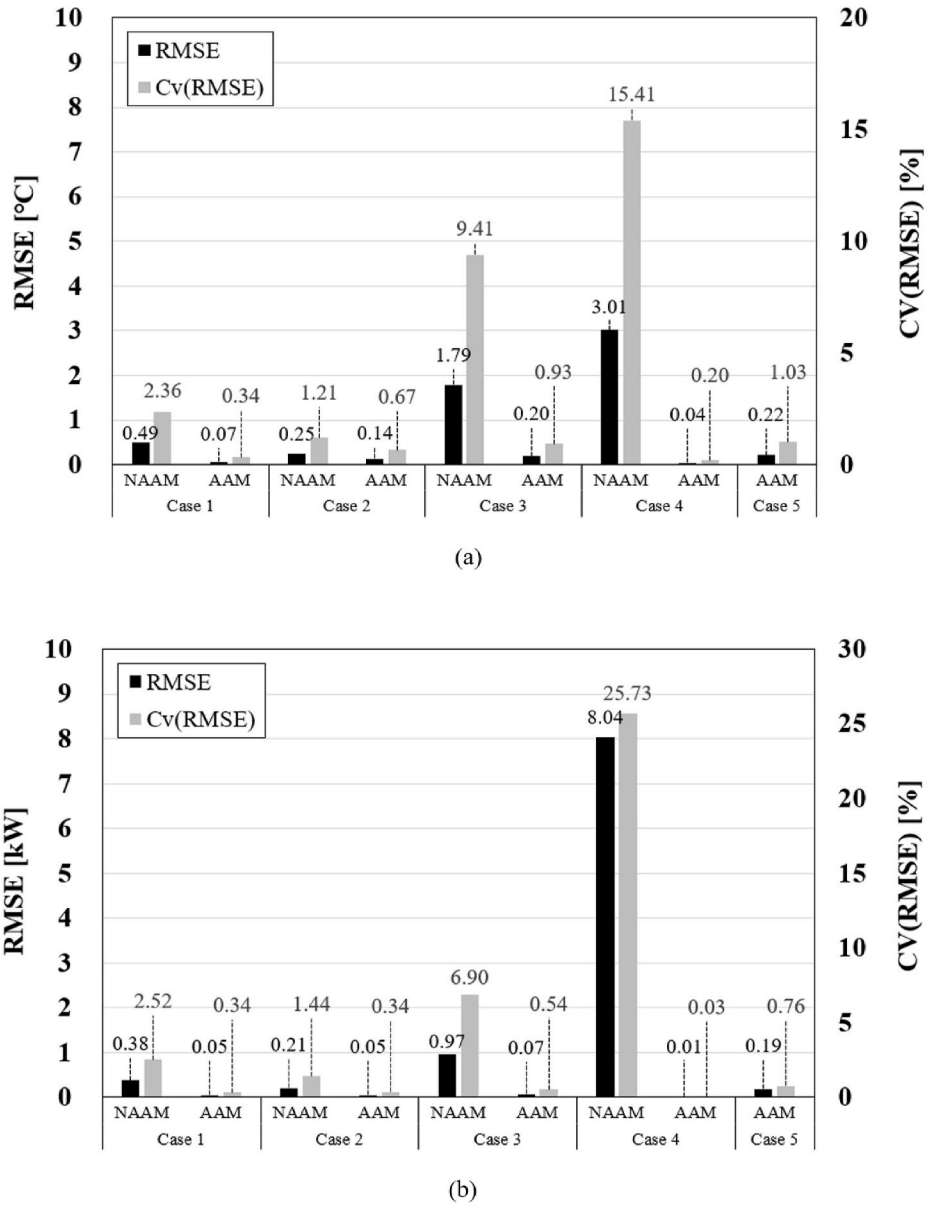


Fig. 9. Prediction accuracy of the NAAM and AAM: (a) inlet rack temperature, and (b) cooling system energy.

versions of the AAM compared with the actual data. The data distribution was close to the $y = x$ function, which means that the prediction was stable without any significant outliers. Overall, both versions of the AAM had a high prediction accuracy and produced stable predictions.

3.2. Performance evaluation of the optimal control algorithm

To test the performance of the optimal control algorithm using the AAM, the prediction of the inlet rack temperature and the cooling system energy was conducted using both an NAAM without a real-time training function and the AAM, which allows for real-time training. The prediction accuracy, adaptability, and control stability of the algorithms for these models in response to changes in the operating environment, physical environment, and system capacity of a containment-type data center were tested using simulations of five cases as previously described in Table 2.

The prediction accuracy of the NAAM and AAM was evaluated using RMSE and CV(RMSE), which were calculated as $T_{pred,rack,in(n)}$ and $T_{rack,in(n)}$. The control stability was evaluated using the mean absolute error (MAE) and the maximum error for $T_{rack,in(n)}$ and $T_{set(n)}$ at the nth time step. The MAE is defined in Equation (19):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|T_{set(n)} - T_{rack,in(n)}\right| \tag{19}$$

3.2.1. Prediction accuracy of the NAAM and AAM

The inlet rack temperature and the cooling system energy were predicted at every time step; the resulting prediction accuracy for the NAAM and AAM for each case is presented in Fig. 9. In terms of the accuracy of the inlet rack temperature prediction for the NAAM, Case 2 had the lowest RMSE (0.25 °C) and CV(RMSE) (1.21%), while Case 4 had the weakest predictive performance, with an RMSE of 3.01 °C and a CV(RMSE) of 15.41%. In Cases 1 and 2, the operating environment was modified by adjusting the IT load and setpoint temperature, respectively. In contrast, Cases 3 and 4 introduced a change to the physical environment, and the prediction accuracy was relatively low compared to Cases 1 and 2. The RMSE was more than 1 °C, which means that the average error was high or that a particularly large error occurred at some point in the simulation. In particular, the CV(RMSE) of Cases 3 and

Table 4
Control performance of NAAM and AAM.

Case	Model	Control performance		Cooling system energy
		MAE	Max. error	Average
		[°C]	[°C]	[kW]
Case 1	NAAM	0.72	2.84	15.13
	AAM	0.04	0.17	13.70
Case 2	NAAM	0.56	2.70	14.50
	AAM	0.05	1.02	13.89
Case 3	NAAM	1.95	2.81	14.10
	AAM	0.09	1.28	12.57
Case 4	NAAM	1.47	1.48	31.24
	AAM	0.02	0.30	27.15
Case 5	AAM	0.08	1.17	25.42

4 was 9.41% and 15.41%, respectively, thus there was a significant difference between the average of the predicted value and the actual value.

The prediction accuracy for the cooling system energy using the NAAM was similar to that for the rack inlet temperature. The RMSE and CV(RMSE) for Cases 1 (0.49 kW and 2.36%, respectively) and 2 (0.21 kW and 1.44%, respectively) represented satisfactory prediction accuracy. The RMSE for Case 3 was 0.97 kW and the CV(RMSE) was 6.90%, which was relatively high, while for Case 4, the RMSE was 8.04 kW and the CV(RMSE) was 25.73%, indicating a significant error in the prediction accuracy.

Overall, the AAM exhibited a stronger prediction accuracy than did the NAAM for both the rack inlet temperature and the cooling system energy. For Cases 1 to 4, in which a single variable was changed, the highest temperature prediction accuracy was for Case 3, with an RMSE of 0.20 °C and a CV(RMSE) of 0.93%. Case 5, which included changes to all environmental variables, had an RMSE of 0.22 °C and a CV(RMSE) of 1.03%. This high prediction accuracy indicates that the prediction performance of AAM was improved by the real-time training process.

The version of the AAM for cooling system energy produced more accurate predictions than did that for the rack inlet temperature. For Cases 1 to 4, Case 3 had the highest RMSE and CV(RMSE) at 0.07 kW and 0.54%, respectively, thus representing the lowest predictive performance, but the error was still considered low. Case 5 also had a high prediction accuracy, with an RMSE of 0.19 kW and a CV(RMSE) of 0.76%.

Overall, the prediction performance for the rack inlet temperature and the cooling system energy improved by 79% and 89%, respectively, with the addition of real-time training to the base model. This means that AAM was able to adapt to changes in the operating environment, physical environment, and the capacity of the cooling system using real-

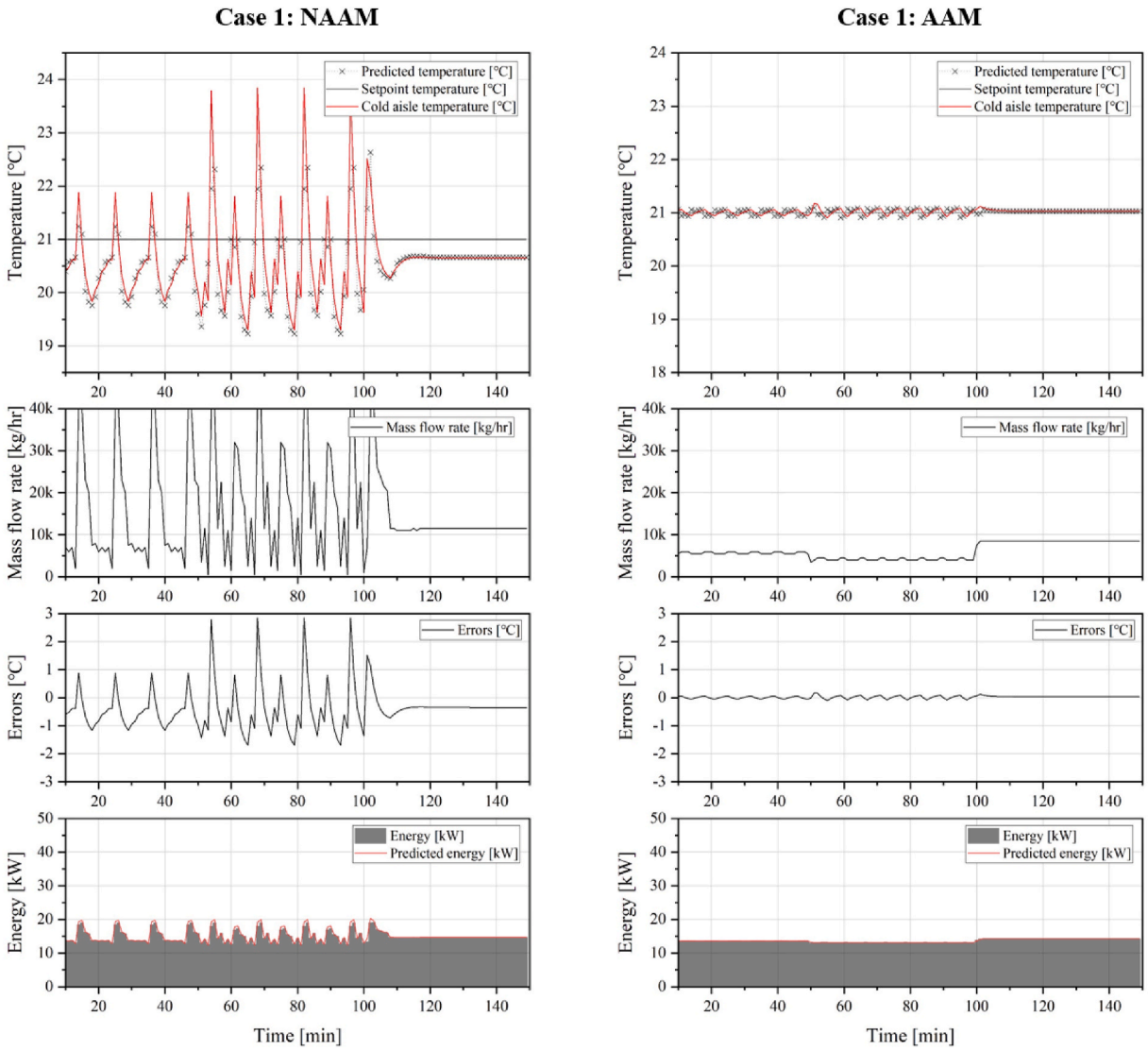


Fig. 10. Case 1 simulation results.

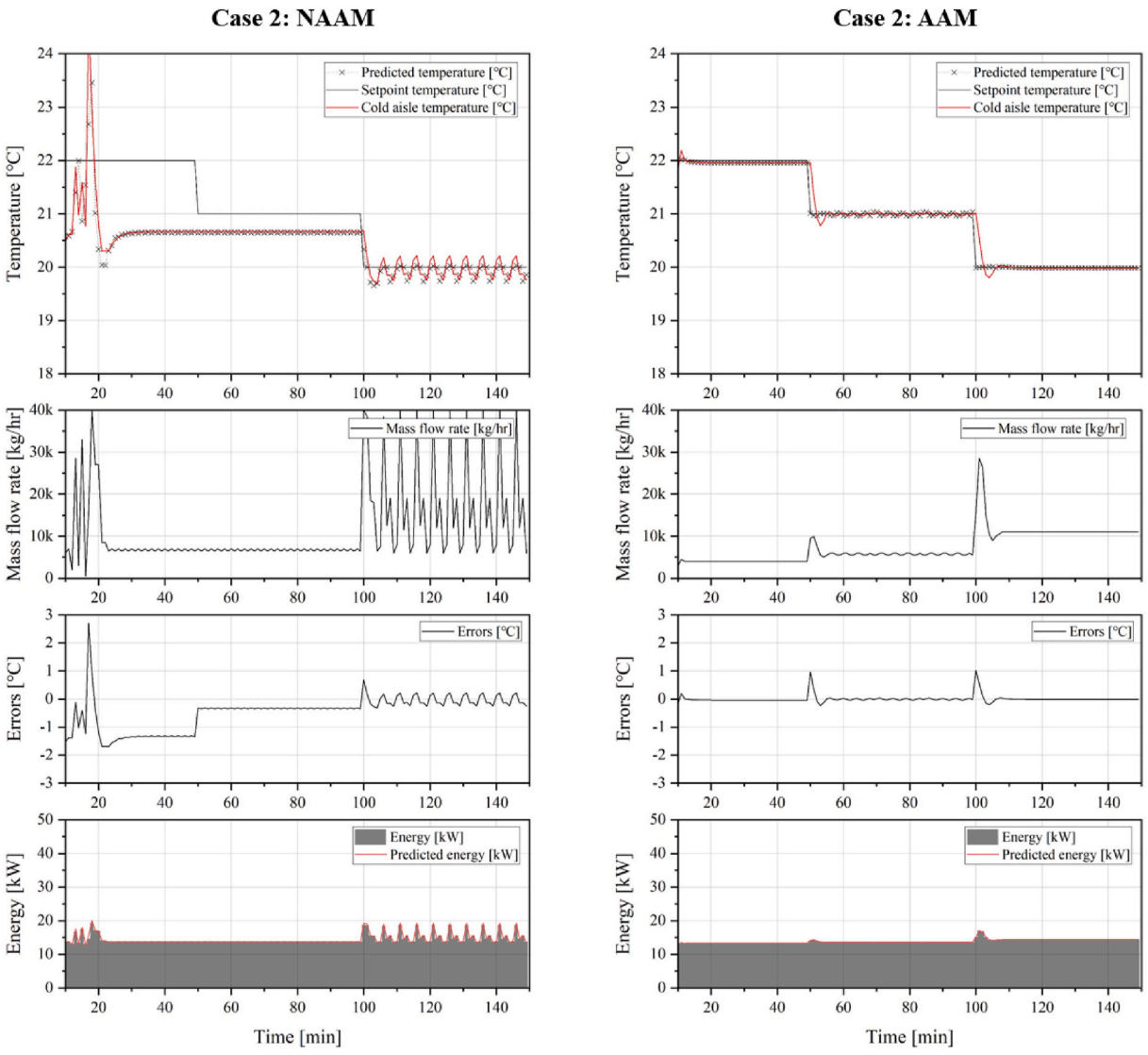


Fig. 11. Case 2 simulation results.

time training. Therefore, when an AAM is employed within the control algorithm, optimal control using accurate environmental predictions is possible.

3.2.2. *Adaptability and energy performance of the control algorithm*

The control algorithm is heavily reliant on the prediction accuracy of the AAM (or the NAAM) because it calculates the optimal mass flow rate of the chilled water based on these predictions. However, a high prediction accuracy does not always guarantee control stability. For example, when the inlet rack temperature is predicted for all possible chilled water mass flow rates but does not converge to the setpoint temperature, the inlet rack temperature is controlled by the mass flow rate of the chilled water entered as the proximity value. If the control system uses the corresponding flow rate and the resulting rack inlet temperature is similar to the predicted nearest value, then the prediction accuracy will be high, but the control performance will be poor. Therefore, this section analyzes the rack inlet temperature, predicted rack inlet temperature, the error between the setpoint temperature and the rack inlet temperature, the cooling system energy, and the predicted cooling system energy to evaluate the adaptability and stability of the optimal control algorithm and calculate energy consumption based on the optimal control configuration. The MAE, maximum error, and average cooling system energy for the NAAM and AAM for each case are summarized in Table 4.

In Case 1, in which the IT load was changed, the rack inlet temperature and cooling system energy predicted by the NAAM and AAM were accurate. However, the MAE was 0.72 °C and the maximum error was 2.84 °C, suggesting that, while the predictions were accurate, the control was not optimal. Fig. 10 shows that the error for the NAAM did not converge to 0, while the rack inlet temperature did not converge to the setpoint temperature. In particular, at an IT load of 65 kW (50–100 min), the maximum error occurred, and the control was very unstable. The predicted rack inlet temperature was similar to the actual rack inlet temperature, meaning that the temperature prediction was accurate; however, the optimal chilled water mass flow rate for convergence to the setpoint temperature could not be derived. The cooling system energy was also high during this period with unstable control. On the other hand, with the use of the AAM, the rack inlet temperature converged in a stable manner to the setpoint temperature, and the error also converged to 0 °C. The average cooling system energy consumption using the NAAM and AAM was 15.13 kW and 13.70 kW, respectively. The average energy consumption using the AAM was consequently 0.44% lower than that using the NAAM.

Case 2 investigated the change in the setpoint temperature. As in Case 1, both the NAAM and AAM had a high prediction accuracy. However, the MAE for the NAAM was 0.56 °C and the maximum error

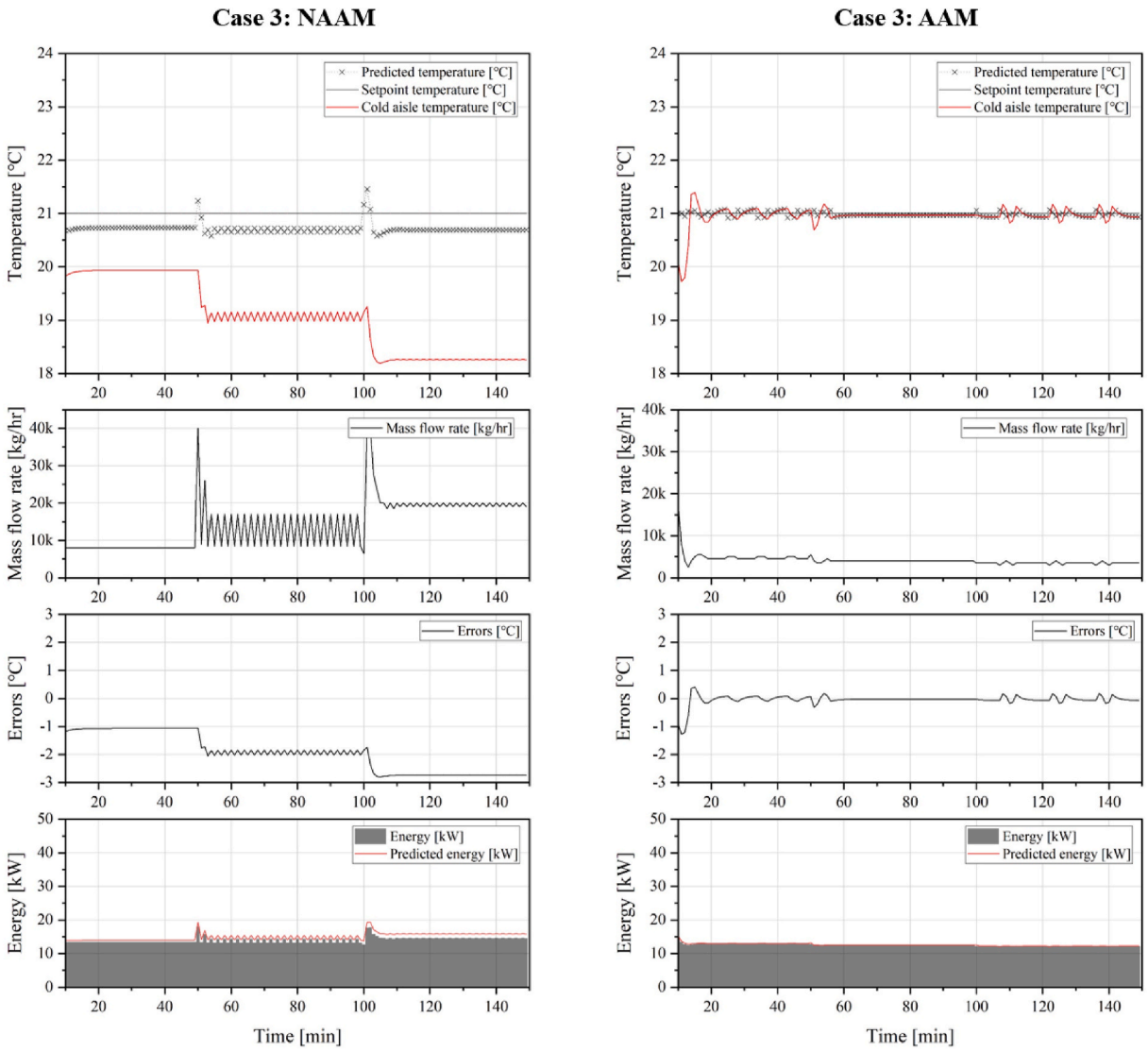


Fig. 12. Case 3 simulation results.

was 2.70 °C, indicating unstable control. In contrast, the MAE of the AAM was 0.05 °C and the maximum error was 1.02 °C. Fig. 11 shows that temperature control using the NAAM failed during the period in which the setpoint temperature was 22 °C and 21 °C and fluctuated within the range of ±0.5 °C when it was 20 °C. This contrasts with the stable control observed with the use of the AAM due to its real-time training function. Temporary undershooting occurred when the setpoint temperature was changed, but the error converged to 0 within 10 min because the mass flow rate of chilled water was stably supplied.

The average cooling system energy using the NAAM was 14.50 kW and that using the AAM was 13.89 kW, a 4.21% reduction. Making an objective comparison is difficult due to the control failure of the NAAM control algorithm. However, when calculating the average energy consumption for 100–150 min, compared to NAAM application case of 14.25 kW, the control algorithm with AAM afforded a lower energy consumption of 13.36 kW and the area below the setpoint temperature was smaller. There is no guarantee that reducing the temperature fluctuations and performing stable control through optimal control inevitably saves energy. However, this result signifies that energy saving is possible when optimal control with AAM is applied.

In Cases 1 and 2, even though only the operating environment was changed in the base model, the NAAM could not achieve stable control. This suggests that control may fail if the direction of the prediction does not match the direction of the system, even if the ANN is trained by extracting data from a single model and the trained ANN is applied to the same model. Because it is difficult to obtain data for all situations within the target system, real-time training and optimization of the applied system are required even for an optimized ANN.

Case 3 represents a change in the physical environment within the containment data center. The prediction accuracy of the NAAM was low, while the AAM exhibited good prediction performance (Fig. 12). The MAE for the NAAM was 1.95 °C, errors occurred in all periods, and the maximum error was 2.81 °C. Errors close to the maximum error also continued to occur after 100 min. The rack inlet temperature did not converge to the setpoint temperature, with predicted values that deviated greatly from the actual values. This is because the NAAM predicted that the rack inlet temperature for the input chilled water mass flow rate would be close to 21 °C during the control process, but in the physical model, the inlet rack temperature was higher. This indicates that errors are unavoidable if there is no retraining process in the cyber model. Therefore, even though the RMSE and CV(RMSE) were not very high, the control continued to fail, and the error increased as the delta rack inlet temperature increased. On the other hand, AAM converged to the setpoint temperature despite the change in the delta inlet temperature. During the initial 10–20 min, a transient response that included a maximum error of 1.28 °C was observed, but errors of less than 0.5 °C

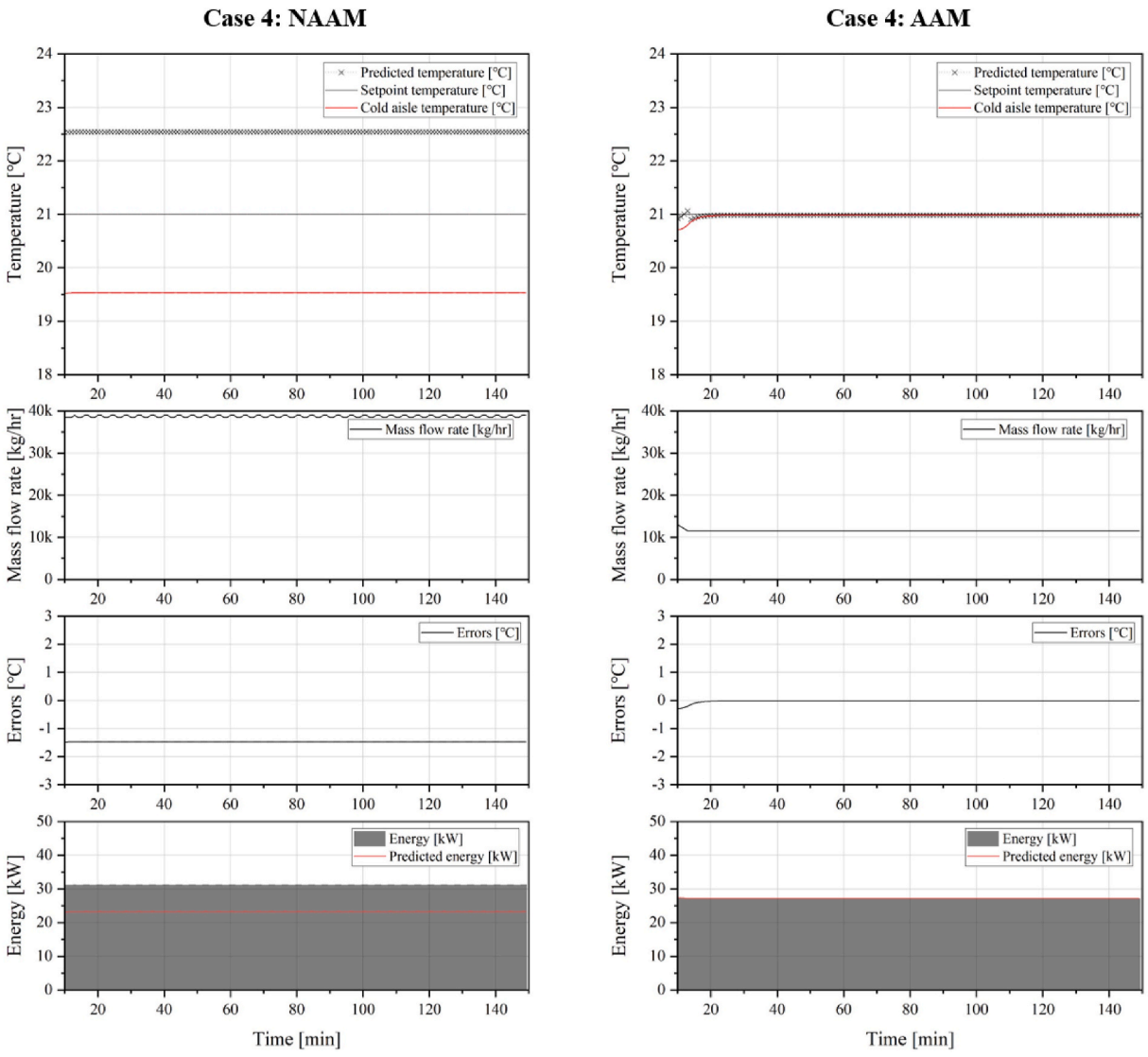


Fig. 13. Case 4 simulation results.

were recorded after this, and a low MAE of 0.09 °C was observed. The difference in chilled water mass flow rate between the two cases indicates that the NAAM derived excessive flow rate which led to the overcooling. The average cooling system energy for the NAAM and AAM was 14.10 kW and 12.57 kW, respectively, representing a 12.17% decrease with the use of the AAM.

In Case 4, the capacity of the cooling systems was changed. The prediction accuracy using the NAAM was very low, with an MAE of 1.47 °C and a maximum error of 1.48 °C for the rack inlet temperature, indicating that a consistent error occurred over the entire operating period as well as the flow rate of chilled water which presented constant statement during the control (Fig. 13). The cooling system energy also had a constant error between the predicted and actual values. In contrast, the MAE for the rack inlet temperature was 0.02 °C and the maximum error was 0.30 °C when using the AAM, indicating that the system was able to adapt well to a change in the system capacity via optimal control. The average cooling system energy was 31.24 kW using the NAAM and 27.15 kW using AAM, a 15.06% difference.

In Case 5, all of the variables that were changed individually in the previous four cases were changed in order to understand the performance of the AAM in a completely new environment. As shown in Fig. 14, a transient response in the temperature occurred during the period in which the IT load and setpoint temperature were changed, but it converged to the setpoint temperature within 10 min. The mass flow rate graph shows how optimal control was implemented in response to partial load and setpoint temperature changes. Though the control was unstable during the first 10 min, most errors were generally within ±0.5 °C. The maximum error (1.17 °C) occurred after 100 min when the IT load was changed to 130 kW and the setpoint temperature was changed to 20 °C. This represented a transient response that was resolved by convergence to the setpoint temperature after slight undershooting. The MAE was low at 0.08 °C even though many variables had been changed. The average cooling system energy was 25.42 kW.

The simulation results showed that the adaptability and control stability of the optimal control algorithm when using the AAM were superior to those with the use of the NAAM. For both prediction models, larger errors occurred when the physical environment was changed compared with the operating environment; however, when the AAM was employed, the error gradually converged to 0 due to the real-time training and the cooling system energy was reduced. Moreover, for all time steps, the real-time training took less than a few seconds, which has little effect on the control delay.

For a more objective comparison, the result was compared with that of the Adaptive Predictive Controller (APC) proposed by Martínez-García et al. [31]. The APC performance was evaluated for the setpoint temperature variation scenario, and the result was expressed using MSE

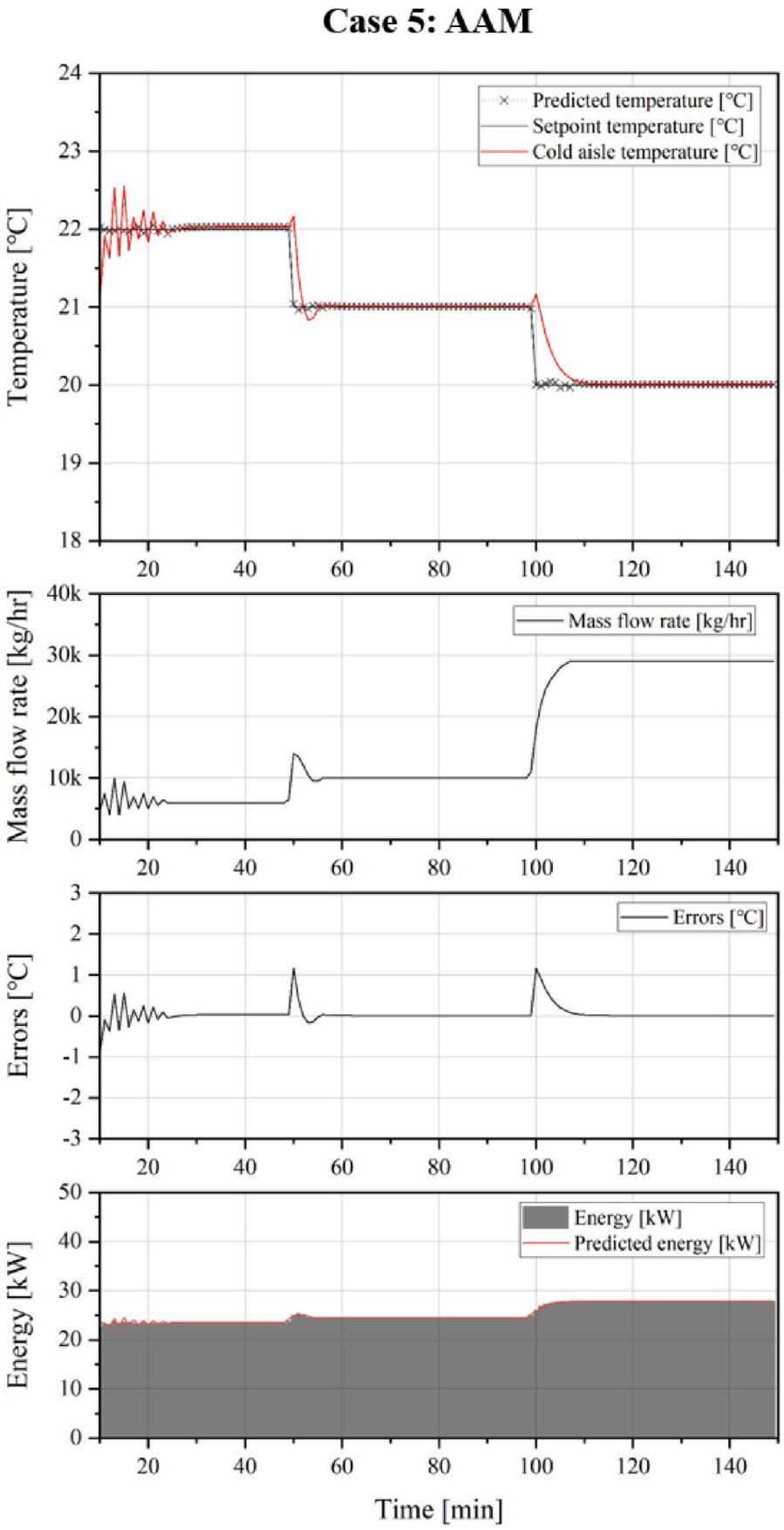


Fig. 14. Case 5 simulation results.

within the best performance of 0.0814 °C. For the comparison, the MSE for Case 5 was calculated to be 0.05 °C. Although the experimental scenarios of the two studies were not completely identical and the fundamentals of the adaptive controller were different, the performance of the proposed AAM with optimal control algorithm was slightly better than that of APC, despite the application of more environmental changes.

These results indicate that the environment can be optimally controlled by applying the AAM and optimal control algorithm to various sites based on proposed CPS framework without requiring a separate calibration process. Furthermore, the energy efficiency can be increased without changing the existing IT equipment and cooling system by optimally employing the provided resources, and the application of individual CPS to the containment is expected lower the overall data center PUE.

4. Conclusion

Recently, the demand to reduce data center energy consumption has been increasing. New data centers incorporating the latest technology can achieve low PUE, but to improve the energy efficiency of an existing data center, an energy saving method that is simple to apply and utilizes the given resources as efficiently as possible is required. To solve these problems, this study presented a CPS framework for conserving energy in a data center cooling system and developed an AAM and optimal control algorithm, which are essential components of the framework.

A mathematical model was developed based on the target data center for data acquisition and performance evaluation of the prediction model and optimal control algorithm. Two predictive models were developed by ANN that predicted the cold aisle temperature and cooling system energy. Bayesian optimization was employed during the training process to optimize the number of hidden layers and neurons. AAM was employed as the predictive model and used with the optimal control algorithm. The optimal control algorithm was developed to perform real-time AAM training and determine the optimal flow rate of chilled water. The real-time training updates the weights and bias of the AAM at each time step, and optimal control variable is derived based on the minimum error between the predicted cold aisle temperature and set-point temperature. For the prediction and control performance evaluation of the optimal control algorithm with AAM, simulations were performed for five cases and compared with the control algorithm with NAAM. Each case represents changes in the operating environment and cooling system. The main conclusions are as follows:

1. A CPS framework comprising AAM, and optimal control algorithm was proposed. The CPS was designed to be easily applied to the various conditions of containment-type data centers; it improved the energy efficiency by optimally utilizing the given resources at the containment level.
2. For the cold aisle temperature predictive model, the structure was optimized as two hidden layers with 31 neurons each and the CV (RMSE) was calculated as 0.27%. The cooling system energy predictive model afforded CV(RMSE) of 0.14%, and the structure was determined as two hidden layers comprising 21 and 13 neurons. Both models afforded high prediction accuracies and involved to the optimal control algorithm as AAM.
3. The performance evaluation results showed that the RMSE and CV (RMSE) of the optimal control algorithm with AAM were superior to those of the control algorithm with NAAM; additionally, the MAE and maximum error were also lower for the optimal control algorithm AAM than that for the control algorithm with NAAM, confirming the control stability of the algorithm. Furthermore, the cooling system energy consumption was reduced due to the more stable control using AAM. Especially, despite the simultaneous application of all the environmental changes (Case 5), the optimal control algorithm with AAM presented high prediction accuracy and control performance due to the adaptation to the new environment. The RMSE and CV(RMSE) for the temperature prediction were 0.22 °C and 1.02%, respectively, while that for the cooling system energy was 0.19 kW and 0.76%, respectively. For the control stability, MAE was calculated as 0.08 °C, while the maximum error was 1.17 °C.

The analyzed results revealed that, when an ANN model capable of real-time training is applied to a CPS, it is possible to rapidly adapt to the environmental changes in a stable manner. Moreover, more energy can be saved with more stable control. Therefore, if the proposed CPS framework comprising the AAM and optimal control algorithm is employed to the existing containment-type data centers, energy consumption can be reduced without changing the IT equipment and cooling system. Furthermore, the application of the CPS framework to individual containments can improve the PUE of the entire data center.

In future research, the application of a CPS to an actual containment data center is required, and technical issues such as data communication and time delays that occur in the process of employing an AAM and an optimal control algorithm need to be solved. Furthermore, if the AAM and optimal control algorithm are improved using continuous field data feedback and their scalability is improved, they could be extended to data centers with other heat sources, and they can be used to assist managers in decision-making through accurate environmental predictions.

CRediT authorship contribution statement

Young Jae Choi: Writing – original draft, Project administration, Methodology. **Bo Rang Park:** Investigation. **Ji Yeon Hyun:** Visualization. **Jin Woo Moon:** Writing – review & editing.

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.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT, MOE; No. 2019M3E7A1113095) and by the Chung-Ang University Graduate Research Scholarship in 2020.

References

[1] Y. Liu, X. Wei, J. Xiao, Z. Liu, Y. Xu, Y. Tian, Energy consumption and emission mitigation prediction based on data center traffic and PUE for global data centers, Glob. Energy Interconnect. 3 (3) (2020) 272–282.
[2] E. Masanet, A. Shehabi, N. Lei, S. Smith, J. Koomey, Recalibrating global data center energy-use estimates 367 (6481) (2020) 984–986.
[3] N. Jones, How to stop data centres from gobbling up the world's electricity, Nature 561 (7722) (2018) 163–166.
[4] J. Cho, J. Woo, Development and experimental study of an independent row-based cooling system for improving thermal performance of a data center, Appl. Therm. Eng. 169 (2020) 114857.
[5] M. Koot, F. Wijnhoven, Usage impact on data center electricity needs: a system dynamic forecasting model, Appl. Energy 291 (2021) 116798.
[6] M. Dayarathna, Y. Wen, R. Fan, Data center energy consumption modeling: a survey, IEEE Commun. Surv. Tutorials 18 (1) (2016) 732–794.
[7] X. Zhang, T. Lindberg, N. Xiong, V. Vyatkin, A. Mousavi, Cooling energy consumption investigation of data center IT room with vertical placed server, Energy Proc. 105 (2017) 2047–2052.
[8] A.M. Abbas, A.S. Huzayyin, T.A. Mouneer, S.A. Nada, Thermal management and performance enhancement of data centers architectures using aligned/staggered in-row cooling arrangements, Case Stud. Therm. Eng. 24 (2021) 100884.
[9] J. Cho, Y. Kim, Development of modular air containment system: thermal performance optimization of row-based cooling for high-density data centers, Energy 231 (2021) 120838.

[10] S.A. Nada, K.E. Elfeky, Experimental investigations of thermal managements solutions in data centers buildings for different arrangements of cold aisles containments, J. Build. Eng. 5 (2016) 41–49.
[11] B. Zhan, S. Shao, M. Lin, H. Zhang, C. Tian, Y. Zhou, Experimental investigation on ducted hot aisle containment system for racks cooling of data center, Int. J. Refrig. 127 (2021) 137–147.
[12] M. Tatchell-Evans, N. Kapur, J. Summers, H. Thompson, D. Oldham, An experimental and theoretical investigation of the extent of bypass air within data centres employing aisle containment, and its impact on power consumption, Appl. Energy 186 (2017) 457–469.
[13] C. Jin, X. Bai, C. Yang, W. Mao, X. Xu, A review of power consumption models of servers in data centers, Appl. Energy 265 (2020) 114806.
[14] A.M. Abbas, A.S. Huzayyin, T.A. Mouneer, S.A. Nada, Effect of data center servers' power density on the decision of using in-row cooling or perimeter cooling, Alex. Eng. J. 60 (4) (2021) 3855–3867.
[15] M.H. Jahangir, R. Mokhtari, S.A. Mousavi, Performance evaluation and financial analysis of applying hybrid renewable systems in cooling unit of data centers – a case study, Sustain. Energy Technol. Assessments 46 (2021) 101220.
[16] H. Cheung, S. Wang, Reliability and availability assessment and enhancement of water-cooled multi-chiller cooling systems for data centers, Reliab. Eng. Syst. Saf. 191 (2019) 106573.
[17] X. Yuan, X. Xu, J. Liu, Y. Pan, R. Kosonen, Y. Gao, Improvement in airflow and temperature distribution with an in-rack UFAD system at a high-density data center, Build. Environ. 168 (2020) 106495.
[18] X. Yuan, Y. Wang, J. Liu, X. Xu, X. Yuan, Experimental and numerical study of airflow distribution optimisation in high-density data centre with flexible baffles, Build. Environ. 140 (2018) 128–139.
[19] J. Cho, J. Yang, W. Park, Evaluation of air distribution system's airflow performance for cooling energy savings in high-density data centers, Energy Build. 68 (2014) 270–279.
[20] A.J. Díaz, R. Cáceres, R. Torres, J.M. Cardemil, L. Silva-Llanca, Effect of climate conditions on the thermodynamic performance of a data center cooling system under water-side economization, Energy Build. 208 (2020) 109634.
[21] M. Deymi-Dashtebayaz, S. Valipour Namanlo, A. Arabkoohsar, Simultaneous use of air-side and water-side economizers with the air source heat pump in a data center for cooling and heating production, Appl. Therm. Eng. 161 (2019) 114133.
[22] S.-W. Ham, M.-H. Kim, B.-N. Choi, J.-W. Jeong, Energy saving potential of various air-side economizers in a modular data center, Appl. Energy 138 (2015) 258–275.
[23] Q. Fang, J. Wang, Q. Gong, Thermal-aware server provisioning with switched MPC for HPC data Centers**This work was supported in part by the National Science Foundation of China under grant No. 51406136, and the International Science and technology Cooperation program of China under grant No. 2011DFG13020, IFAC-PapersOnLine 49 (18) (2016) 766–771.
[24] K. Li, Optimal power allocation among multiple heterogeneous servers in a data center, Sustain. Comput. Inform. Syst. 2 (1) (2012) 13–22.
[25] B. Yi, X. Wang, M. Huang, Y. Zhao, Novel resource allocation mechanism for SDN-based data center networks, J. Netw. Comput. Appl. 155 (2020) 102554.
[26] Q. Zhang, C. Tang, T. Bai, Z. Meng, Y. Zhan, J. Niu, M.J. Deen, A two-layer optimal scheduling framework for energy savings in a data center for Cyber–Physical–Social Systems, J. Syst. Architect. 116 (2021) 102050.
[27] B. Qureshi, Profile-based power-aware workflow scheduling framework for energy-efficient data centers, Future Generat. Comput. Syst. 94 (2019) 453–467.
[28] M. Sharma, R. Garg, An artificial neural network based approach for energy efficient task scheduling in cloud data centers, Sustain. Comput. Inform. Syst. 26 (2020) 100373.
[29] E. Terzi, L. Fagiano, M. Farina, R. Scattolini, Structured modelling from data and optimal control of the cooling system of a large business center, J. Build. Eng. 28 (2020) 101043.
[30] G. Mohsenian, S. Khalili, M. Tradat, Y. Manaserh, S. Rangarajan, A. Desu, D. Thakur, K. Nemati, K. Ghose, B. Sammakia, A novel integrated fuzzy control system toward automated local airflow management in data centers, Control Eng. Pract. 112 (2021) 104833.
[31] F. Martínez-García, G. Badawy, M. Kheradmandi, D.G. Down, Adaptive Predictive Control of a data center cooling unit, Control Eng. Pract. 107 (2021) 104674.
[32] A. Afram, F. Janabi-Sharifi, Theory and applications of HVAC control systems – a review of model predictive control (MPC), Build. Environ. 72 (2014) 343–355.
[33] K. Dunlap, N. Rasmussen, Choosing between Room, Row, and Rack-Based Cooling for Data Centers, vol. 130, Schneider Electric White Paper, 2012.
[34] ASHRAE, Data center power equipment thermal guidelines and best practices, in: ASHRAE TC9.9, 2016.
[35] J.B. Petersen, J.D. Bendtsen, J. Stoustrup, Nonlinear Model Predictive Control for Energy Efficient Cooling in Shopping Center HVAC, 2019 IEEE Conference on Control Technology and Applications, CCTA, 2019, pp. 611–616.
[36] Y. Ma, F. Borrelli, B. Hencey, B. Coffey, S. Bengea, P. Haves, Model predictive control for the operation of building cooling systems, IEEE Trans. Control Syst. Technol. 20 (3) (2012) 796–803.
[37] N. Lazic, T. Lu, C. Boutilier, M. Ryu, E.J. Wong, B. Roy, G. Imwalle, Data center cooling using model-predictive control, in: Proceedings of the Thirty-second Conference on Neural Information Processing Systems, 2018, pp. 3818–3827 (NeurIPS-18).
[38] J. Pang, N. Zhang, Q. Xiao, F. Qi, X. Xue, A new intelligent and data-driven product quality control system of industrial valve manufacturing process in CPS, Comput. Commun. 175 (2021) 25–34.
[39] L. Parolini, B. Sinopoli, B.H. Krogh, Z. Wang, A cyber–physical systems approach to data center modeling and control for energy efficiency, Proc. IEEE 100 (1) (2012) 254–268.
[40] M. Ogura, J. Wan, S.J.I.-P. Kasahara, Model predictive control for energy-efficient operations of data centers with cold aisle containments 51 (20) (2018) 209–214.
[41] L. Marshall, P.J. Bemis, Using CFD for Data Center Design and Analysis, 2011.
[42] C. Gao, Z. Yu, J. Wu, Investigation of airflow pattern of a typical data center by CFD simulation, Energy Proc. 78 (2015) 2687–2693.
[43] B. Watson, V.K.J.E.P. Venkiteswaran, Universal cooling of data centres: a CFD analysis 142 (2017) 2711–2720.
[44] A. Saiyad, A. Patel, Y. Fulpagare, A.J. Bhargav, Predictive modeling of thermal parameters inside the raised floor plenum data center using Artificial Neural Networks 42 (2021) 102397.
[45] J. Athavale, M. Yoda, Y. Joshi, Comparison of data driven modeling approaches for temperature prediction in data centers, Int. J. Heat Mass Tran. 135 (2019) 1039–1052.
[46] N. Lei, E. Masanet, Statistical analysis for predicting location-specific data center PUE and its improvement potential, Energy 201 (2020) 117556.
[47] Z. Song, B.T. Murray, B. Sammakia, Airflow and temperature distribution optimization in data centers using artificial neural networks, Int. J. Heat Mass Tran. 64 (2013) 80–90.
[48] S. Asgari, S. MirhoseiniNejad, H. Moazamigoodarzi, R. Gupta, R. Zheng, I.K. Puri, A gray-box model for real-time transient temperature predictions in data centers, Appl. Therm. Eng. 185 (2021) 116319.
[49] R.J. Hedjar, Adaptive neural network model predictive control, Inform. Contr. 9 (3) (2013) 1245–1257.
[50] V. Adetola, D. DeHaan, M. Guay, Adaptive model predictive control for constrained nonlinear systems, Syst. Control Lett. 58 (5) (2009) 320–326.
[51] T.A.N. Heirung, B.E. Ydstie, B. Foss, Dual adaptive model predictive control, Automatica 80 (2017) 340–348.
[52] S. Yang, M.P. Wan, W. Chen, B.F. Ng, S. Dubey, Experiment study of machine-learning-based approximate model predictive control for energy-efficient building control, Appl. Energy 288 (2021) 116648.
[53] A. Villalonga, E. Negri, G. Biscardo, F. Castano, R.E. Haber, L. Fumagalli, M. Macchi, A decision-making framework for dynamic scheduling of cyber-physical production systems based on digital twins, Annu. Rev. Control 51 (2021) 357–373.
[54] H. Navarro, L.J. Cabezas-Gómez, Effectiveness-NTU computation with a mathematical model for cross-flow heat exchangers 24 (2007) 509–521.
[55] B. Seo, Y.B. Yoon, B.H. Yu, S. Cho, K.H. Lee, Comparative analysis of cooling energy performance between water-cooled VRF and conventional AHU systems in a commercial building, Appl. Therm. Eng. 170 (2020) 114992.
[56] A.J. Zilouchian, Fundamentals of neural networks, 2001, pp. 1–5, 1.
[57] K. Sun, N. Luo, X. Luo, T.J.E. Hong, Buildings, Prototype energy models for data centers 231 (2021) 110603.
[58] ASHRAE, Measurement of Energy, Demand, and water savings, in: ASHRAE Guideline, 2014, 14-2014.