

# Lawrence Berkeley National Laboratory

## LBL Publications

### Title

Deep reinforcement learning control for co-optimizing energy consumption, thermal comfort, and indoor air quality in an office building

### Permalink

<https://escholarship.org/uc/item/826584wd>

### Journal

Applied Energy, 377

### ISSN

0306-2619

### Authors

Guo, Fangzhou  
Ham, Sang woo  
Kim, Donghun  
[et al.](#)

### Publication Date

2025

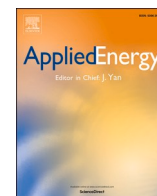
### DOI

10.1016/j.apenergy.2024.124467

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial License, available at <https://creativecommons.org/licenses/by-nc/4.0/>

Peer reviewed



# Deep reinforcement learning control for co-optimizing energy consumption, thermal comfort, and indoor air quality in an office building

Fangzhou Guo<sup>a,\*</sup>, Sang woo Ham<sup>a</sup>, Donghun Kim<sup>a</sup>, Hyeun Jun Moon<sup>b</sup>

<sup>a</sup> Lawrence Berkeley National Laboratory, Berkeley, CA, USA

<sup>b</sup> Department of Architectural Engineering, Dankook University, Yongin, Republic of Korea

## HIGHLIGHTS

- A Deep reinforcement learning (DRL) algorithm is proposed to control HVAC systems.
- A hybrid model is developed to simulate dynamics of the building and HVAC system.
- DRL can co-optimize energy, thermal comfort, CO<sub>2</sub>, and air quality simultaneously.
- How decisions are made by DRL under each circumstance is thoroughly analyzed.
- Performance of DRL, MPC, and rule-based controllers are compared.

## ARTICLE INFO

### Keywords:

Deep reinforcement learning  
Deep deterministic policy gradient  
Smart building  
Air conditioning  
Energy efficiency  
Thermal comfort  
Indoor air quality

## ABSTRACT

With the recent demand for decarbonization and energy efficiency, advanced HVAC control using Deep Reinforcement Learning (DRL) becomes a promising solution. Due to its flexible structures, DRL has been successful in energy reduction for many HVAC systems. However, only a few researches applied DRL agents to manage the entire central HVAC system and control multiple components in both the water loop and the air loop, owing to its complex system structures. Moreover, those researches have not extended their applications by incorporating the indoor air quality, especially both CO<sub>2</sub> and PM2.5 concentrations, on top of energy saving and thermal comfort, as achieving those objectives simultaneously can cause multiple control conflicts. What's more, DRL agents are usually trained on the simulation environment before deployment, so another challenge is to develop an accurate but relatively simple simulator. Therefore, we propose a DRL algorithm for a central HVAC system to co-optimize energy consumption, thermal comfort, indoor CO<sub>2</sub> level, and indoor PM2.5 level in an office building. To train the controller, we also developed a hybrid simulator that decoupled the complex system into multiple simulation models, which are calibrated separately using laboratory test data. The hybrid simulator combined the dynamics of the HVAC system, the building envelope, as well as moisture, CO<sub>2</sub>, and particulate matter transfer. Three control algorithms (rule-based, MPC, and DRL) are developed, and their performances are evaluated on the hybrid simulator environment with a realistic scenario (i.e., with stochastic noises). The test results showed that, the DRL controller can save 21.4 % of energy compared to a rule-based controller, and has improved thermal comfort, reduced indoor CO<sub>2</sub> concentration. The MPC controller showed an 18.6 % energy saving compared to the DRL controller, mainly due to savings from comfort and indoor air quality boundary violations caused by unmeasured disturbances, and it also highlights computational challenges in real-time control due to non-linear optimization. Finally, we provide the practical considerations for designing and implementing the DRL and MPC controllers based on their respective pros and cons.

\* Corresponding author.

E-mail address: [fzguo@lbl.gov](mailto:fzguo@lbl.gov) (F. Guo).

<https://doi.org/10.1016/j.apenergy.2024.124467>

Received 8 May 2024; Received in revised form 20 July 2024; Accepted 6 September 2024

Available online 15 September 2024

0306-2619/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

| Nomenclatures    |                                                              | Subscriptions |                      |
|------------------|--------------------------------------------------------------|---------------|----------------------|
| $C$              | concentration [ppm, $\mu\text{g}/\text{m}^3$ ]               | chw           | chilled water        |
| $D$              | damper opening (from 0 to 100%)                              | chws          | chilled water supply |
| $E$              | energy consumption [kWh]                                     | chwr          | chilled water return |
| $P_{\text{air}}$ | air purifier ON/OFF status                                   | cv            | cooling coil valve   |
| $T$              | temperature [ $^{\circ}\text{C}$ ]                           | ea            | exhaust air          |
| $\dot{V}$        | volume flow rate [ $\text{m}^3/\text{s}$ ]                   | oa            | outdoor air          |
| $f$              | fan speed (from 0 to 100 %)                                  | ra            | return air           |
| $q$              | heat flux [ $\text{W}/\text{m}^2$ ]                          | sa            | supply air           |
| $u$              | valve opening (from 0 to 100%)                               | za            | zone air             |
| $\omega$         | humidity ratio [ $\text{kg H}_2\text{O}/\text{kg dry air}$ ] | sp            | setpoint             |
| $\phi$           | relative humidity (from 0 to 100%)                           | sol           | solar                |

## 1. Introduction

The building sector accounts for approximately 74 % of electricity consumption and 40 % of primary energy consumption in the U.S. Especially, the heating, ventilation, and air conditioning (HVAC) systems could be responsible for 35 % of primary energy usage [1]. In the past few decades, researchers have put considerable effort into new technologies to reduce building energy consumption, such as upgrading the building envelope insulation, increasing the HVAC product efficiency, and adopting advanced control algorithms.

The building automation system is an effective tool that can be used to implement advanced control methods to manage the HVAC equipment and achieve optimal efficiency at the system level. The building control methods can be divided into local control and supervisory control [2]. Local control methods focus on immediate, real-time control at the component level or subsystem level with simple and predefined rules, such as ON/OFF control of a cooling system or proportional-integral-derivative (PID) control of a thermostat to maintain the zone air temperature. They usually operate independently and are reactive to local conditions. After fine-tuning, these methods may have near-optimal energy efficiency for certain subsystems. However, they are difficult to achieve optimality for the operation of the overall HVAC system, particularly for the goal of co-optimizing the energy consumption, thermal comfort, and indoor air quality (IAQ) in the entire building, as the goals may conflict with each other in the control. For example, when the economizer mode is enabled, the outdoor damper increases its opening and brings more fresh air, which reduces the indoor  $\text{CO}_2$  level and saves some energy. However, if the outdoor air pollution is severe, it will also bring in more polluted air, adversely affecting the control of indoor air quality. On the other hand, supervisory control operates at a higher hierarchical level, coordinating and optimizing the performance of multiple local controllers and subsystems to achieve overall system-wide objectives. Two typical examples are model predictive control (MPC) and reinforcement learning (RL) control. They seek to choose the optimal control action systematically through either minimizing a loss function or maximizing a reward.

Due to the stricter legislation on greenhouse gas emissions and the promotion of smart and resilient buildings, MPC and RL control have gained increasingly more attention. MPC develops control-oriented models to compute the influence of future control actions on the building dynamics and searches for the optimal action in each time interval, which has been used for energy/cost saving, load flexibility, IAQ improvement,  $\text{CO}_2$  emission reduction, etc. [3–5]. Several MPC research showed successful field demonstrations for relatively simpler systems such as small and medium commercial buildings with packaged HVAC systems [6,7]. However, the main challenges of MPC are the labor-intensive modeling for control-oriented models (e.g., reduced-order gray-box model) and computational complexity in the iterative online optimization [8], especially for large commercial buildings with

complex building automation systems. Also, due to the nonlinear dynamic behavior and the complex correlations between the systems such as building thermal mass, HVAC performances with respect to the local controllers and weather conditions, and renewable resources, optimizing the supervisory control of the system often results in a non-linear optimization problem whose performance depends on the tuning parameters of solvers and initializations, even if the control-oriented building and HVAC models are solid [9]. In comparison, most RL control methods can be model-free. They learn the control policy directly from empirical observations using the reward signal and can be pre-trained offline before deploying in real buildings. A lot of research papers applied RL for controlling HVAC systems, energy storage systems, PV panels, windows, and others in the building to minimize energy consumption/cost and maximize occupant comfort after training the RL control agent in a simulation environment [10–13].

In practical applications, RL agents can learn to improve their performance continuously through interacting with the environment and requires minimal human intervention. Hence, they can adapt to the changing environments by themselves. Also, they are effective for solving complex problems, especially co-optimization problems with multiple control objectives. What's more, they have the potential for transfer learning [14,15], which means applying the learned policies to new, but similar tasks or environments. The deployment of RL agents also faces some challenges, such as data efficiency, reward shaping, and control safety [8,16]. They usually require a lot of data or a large number of interactions with the environment for learning, and the learning process is computationally expensive. Also, the reward design in RL is very challenging and requires specialized expertise. Moreover, the control safety issue cannot be ignored, which includes the risks in exploration, convergence problems in learning, and limited generalization ability to unseen environments.

### 1.1. Literature review

Table 1 lists related research papers that applied RL controllers in building energy systems. In each paper, the RL controller is described by its control objectives, models of the virtual environment, and the components it controls. Examples of control objectives are energy consumption of the HVAC system, thermal comfort of occupants, and indoor air quality related to  $\text{CO}_2$ ,  $\text{PM}_{2.5}$ , and  $\text{PM}_{10}$  concentration. Also, a virtual environment is often developed to train and evaluate the RL agent before deploying it into the real-world environment. The virtual environment usually includes an HVAC system model, a building envelope model, and other models to describe the dynamics of indoor air quality. Those models can either be developed using building simulation software (e.g., EnergyPlus, TRNSYS, Modelica), or be described by gray-box models (e.g., resistance-capacitance building envelope model), black-box regression models, or white-box models (e.g., ordinary differential equations (ODE), detailed physics-based models). Finally, the

**Table 1**

Literature review: RL applications on building energy systems.

| Reference                          | Control objectives                                                                                                                     | Simulation environment                                                                                                                                     | Control components/variables                                                                                                                                                  |
|------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Ahn and Park (2020) [17]           | <ul style="list-style-type: none"> <li>Energy consumption</li> <li>CO<sub>2</sub> concentration</li> </ul>                             | <ul style="list-style-type: none"> <li>HVAC system (EnergyPlus)</li> <li>Building envelope (EnergyPlus)</li> <li>CO<sub>2</sub> (EnergyPlus)</li> </ul>    | <ul style="list-style-type: none"> <li>AHU: damper</li> <li>Chiller: chilled water temperature setpoint</li> <li>Cooling tower: cooling water temperature setpoint</li> </ul> |
| An et al. (2020,2021,2022) [18–20] | <ul style="list-style-type: none"> <li>Energy consumption</li> <li>Zone temperature</li> <li>PM2.5concentration</li> </ul>             | <ul style="list-style-type: none"> <li>PM2.5model (ODE)</li> </ul>                                                                                         | <ul style="list-style-type: none"> <li>Air purifier</li> <li>Window</li> <li>Air conditioner</li> </ul>                                                                       |
| Biemann et al. (2021) [21]         | <ul style="list-style-type: none"> <li>Energy consumption</li> <li>Zone temperature</li> </ul>                                         | <ul style="list-style-type: none"> <li>HVAC system (EnergyPlus)</li> <li>Building envelope (EnergyPlus)</li> </ul>                                         | <ul style="list-style-type: none"> <li>AHU: fan</li> <li>Thermostat: temperature setpoint</li> </ul>                                                                          |
| Blad et al. (2022) [22]            | <ul style="list-style-type: none"> <li>Energy cost</li> <li>Zone temperature</li> </ul>                                                | <ul style="list-style-type: none"> <li>HVAC system (Modelica)</li> <li>Building envelope (Modelica)</li> </ul>                                             | <ul style="list-style-type: none"> <li>Air-to-water heat pump: supply water temperature setpoint, valve</li> </ul>                                                            |
| Deng et al. (2022) [23]            | <ul style="list-style-type: none"> <li>Energy consumption</li> <li>Thermal comfort</li> </ul>                                          | <ul style="list-style-type: none"> <li>HVAC system (EnergyPlus)</li> <li>Building envelope (EnergyPlus)</li> </ul>                                         | <ul style="list-style-type: none"> <li>Thermostat: temperature setpoint</li> </ul>                                                                                            |
| Du et al. (2021) [24,25]           | <ul style="list-style-type: none"> <li>Energy consumption</li> <li>Zone temperature</li> </ul>                                         | <ul style="list-style-type: none"> <li>Building envelope (gray-box)</li> </ul>                                                                             | <ul style="list-style-type: none"> <li>Thermostat: temperature setpoint</li> </ul>                                                                                            |
| Gao et al. (2020) [26]             | <ul style="list-style-type: none"> <li>Energy consumption</li> <li>Thermal comfort</li> </ul>                                          | <ul style="list-style-type: none"> <li>HVAC system (TRNSYS)</li> <li>Building envelope (TRNSYS)</li> </ul>                                                 | <ul style="list-style-type: none"> <li>Thermostat: temperature and humidity setpoint</li> </ul>                                                                               |
| Lu et al. (2023) [27]              | <ul style="list-style-type: none"> <li>Energy consumption</li> <li>Zone temperature</li> </ul>                                         | <ul style="list-style-type: none"> <li>HVAC system (Modelica and regression)</li> <li>Building envelope (Modelica)</li> </ul>                              | <ul style="list-style-type: none"> <li>AHU: supply air temperature and static pressure setpoint</li> </ul>                                                                    |
| Shang et al. (2023) [28]           | <ul style="list-style-type: none"> <li>Energy consumption</li> <li>PM2.5concentration</li> </ul>                                       | <ul style="list-style-type: none"> <li>PM2.5model (ODE)</li> </ul>                                                                                         | <ul style="list-style-type: none"> <li>Air purifier</li> </ul>                                                                                                                |
| Shen et al. (2022) [29]            | <ul style="list-style-type: none"> <li>Energy cost</li> <li>Thermal comfort</li> </ul>                                                 | <ul style="list-style-type: none"> <li>Distributed energy resources (physics-based)</li> </ul>                                                             | <ul style="list-style-type: none"> <li>AHU: fan</li> <li>Battery: state of charge</li> </ul>                                                                                  |
| Touzani et al. (2021) [30]         | <ul style="list-style-type: none"> <li>Energy consumption and cost</li> <li>Zone temperature</li> <li>Battery charge limits</li> </ul> | <ul style="list-style-type: none"> <li>HVAC system (EnergyPlus)</li> <li>Building envelope (EnergyPlus)</li> <li>Battery, PV panels (Modelica)</li> </ul>  | <ul style="list-style-type: none"> <li>Thermostat: temperature setpoint</li> <li>Battery: charge/discharge setpoint</li> </ul>                                                |
| Xia et al. (2023) [31]             | <ul style="list-style-type: none"> <li>Energy consumption</li> <li>Thermal comfort</li> <li>PM2.5and PM10 concentration</li> </ul>     | <ul style="list-style-type: none"> <li>HVAC system (regression)</li> <li>Room temperature (regression)</li> <li>Indoor air quality (regression)</li> </ul> | <ul style="list-style-type: none"> <li>Thermostat: temperature setpoint</li> <li>Light: lighting relaxation</li> </ul>                                                        |
| Yang et al. (2021) [32]            | <ul style="list-style-type: none"> <li>Energy cost</li> <li>Zone temperature</li> <li>CO<sub>2</sub> concentration</li> </ul>          | <ul style="list-style-type: none"> <li>Building envelope (ODE)</li> <li>CO<sub>2</sub> model (ODE)</li> </ul>                                              | <ul style="list-style-type: none"> <li>Heat pump: power input</li> <li>Ventilation system: fan</li> </ul>                                                                     |
| Yoon and Moon (2019) [33]          | <ul style="list-style-type: none"> <li>Energy consumption</li> <li>Thermal comfort</li> </ul>                                          | <ul style="list-style-type: none"> <li>HVAC system (EnergyPlus)</li> <li>Building envelope (EnergyPlus)</li> </ul>                                         | <ul style="list-style-type: none"> <li>Variable refrigerant flow system: temperature setpoint, fan</li> <li>Humidifier</li> </ul>                                             |
| Zhang and Zhao (2022) [34]         | <ul style="list-style-type: none"> <li>Energy consumption</li> <li>Zone temperature</li> </ul>                                         | <ul style="list-style-type: none"> <li>HVAC system (Modelica)</li> <li>Building envelope (Modelica)</li> </ul>                                             | <ul style="list-style-type: none"> <li>Chiller: chilled water pump</li> </ul>                                                                                                 |

1. ODE: ordinary differential equation.

RL agents can be used to control different components of the HVAC system, such as the condenser, chiller, air handling unit (AHU), and terminal/thermostat.

### 1.2. Research gaps and contributions

According to Table 1, studies that simultaneously consider energy consumption, thermal comfort, CO<sub>2</sub>, and PM2.5/PM10 concentration as control objectives are limited. Some studies included either the CO<sub>2</sub> level or the PM2.5 level. For example, Ahn and Park [17] used RL to control an HVAC system with multiple chillers and AHUs. The study considered the indoor air quality in terms of the CO<sub>2</sub> concentration and controlled both the AHU and the chiller. However, the RL algorithm has a large state space which requires significant computational resources for training, and can only output several discrete action values for each control variable. Besides, thermal comfort is not considered in its control objectives. Xia et al. [31] proposed a human-centered recommender system to co-optimize energy, comfort, and air quality based on the feedback of occupants. But the research aimed at achieving energy efficiency by providing actionable recommendations to occupants instead of optimizing the operation of the HVAC system. Yang et al. [32] used RL to control a smart home with a residential heat pump and a ventilation system. The study considered CO<sub>2</sub> concentration and controlled the power of the heat pump and the volume flow rate of the ventilation

system. However, because the RL controller is targeted for a residential system, it will not achieve the optimal energy performance on a central HVAC system with chilled water systems. Additionally, some studies focused specifically on indoor PM2.5 concentrations. For example, Shang et al. [28] applied RL to control the mode of an air purifier to provide better indoor air quality. The study did not involve any cooling/heating devices. In [18–20], the authors controlled the window and air purifier for a safe indoor PM2.5 level. The researches were conducted in a residential home with a portable air conditioner in ON/OFF control, much different from the operation of central HVAC systems.

Though limited research co-optimized energy, comfort, and indoor air quality (CO<sub>2</sub> and PM2.5/PM10), all of these objectives need to be properly controlled in a building, especially large buildings with central HVAC systems. The task is difficult because the co-optimization of these objectives could result in conflicting decisions. As an example, consider the control of the outdoor damper. A larger opening of the outdoor damper is always beneficial for lowering the indoor CO<sub>2</sub> concentration, but if the outside is heavily polluted, it brings in more harmful particles that are detrimental to human health. Also, from the perspective of energy savings, the optimal damper opening depends on the outdoor conditions. If the outside is hot and humid, then bringing in more outdoor air will significantly increase the cooling load. Another example is to achieve energy savings by balancing the supply fan speed and chilled water temperature. A higher chilled water temperature usually leads to



**Table 2**  
Classification of RL algorithms.

| RL algorithm                              | On-policy / off-policy | Online / offline learning | Model-based / model-free     |
|-------------------------------------------|------------------------|---------------------------|------------------------------|
| Dyna-Q                                    | Off-policy             | Online                    | Model-based                  |
| Deep Q-Network (DQN)                      | Off-policy             | Online                    | Model-free (value-based)     |
| REINFORCE                                 | On-policy              | Online                    | Model-free (policy-gradient) |
| Proximal Policy Optimization (PPO)        | On-policy              | Online                    | Model-free (policy-gradient) |
| Trust Region Policy Optimization (TRPO)   | On-policy              | Online                    | Model-free (policy-gradient) |
| Soft Actor-Critic (SAC)                   | Off-policy             | Online                    | Model-free (actor-critic)    |
| Advantage Actor-Critic (A2C)              | On-policy              | Online                    | Model-free (actor-critic)    |
| Asynchronous Advantage Actor-Critic (A3C) | On-policy              | Online                    | Model-free (actor-critic)    |
| Deep Deterministic Policy Gradient (DDPG) | Off-policy             | Online                    | Model-free (actor-critic)    |
| Twin Delayed DDPG (TD3)                   | Off-policy             | Online                    | Model-free (actor-critic)    |

lower chiller power, but the supply air fan speed has to be increased to compensate for the reduced amount of heat exchange in the cooling coil, resulting in higher fan power. Those issues of conflicting decisions in the optimization are not well-addressed in the literature.

Additionally, since training the RL controller directly in the real-world environment is risky, a common approach is to develop a simulation model for training. A lot of studies used EnergyPlus to develop the HVAC system model. However, EnergyPlus has a few drawbacks in the training and evaluation process of RL controllers for complex HVAC systems. First, it relies on predefined HVAC components and systems (i. e., templates), which limits its flexibility in designing a customized model. Second, it uses a procedural, component-based modeling approach, increasing the complexity of interactions between components. Also, the procedural approach may involve simplified assumptions that limit the accuracy and detail of the dynamic simulation. Third, implementing complex control strategies in EnergyPlus models is difficult, especially a strategy with both supervisory-level and local-level control. On the other hand, some studies used Modelica as a modeling tool. Modelica has several advantages in modeling the HVAC control systems. For example, it is very flexible in model design and allows for the integration of mechanical and control systems in a single environment. Also, it uses acausal modeling, which simplifies the modeling process, enhances the flexibility of component interactions, and increases the simulation efficiency. Moreover, Modelica is well-suited for developing and integrating advanced control algorithms, especially RL, allowing for detailed studies of system dynamics and control strategies. However, the building envelope modeling in Modelica language is still a progressing area, and including both the envelope and HVAC system is difficult to calibrate, since there are many unmeasured disturbances in the building envelopes. A potential solution to this problem could be designing a hybrid simulation model that combines the HVAC model developed by Modelica and other models (e.g., the building envelope, indoor air quality) developed by physics-based or data-driven methods.

Based on the above research gaps, this paper makes the following contributions:

- The optimal balance between energy, temperature, humidity, CO<sub>2</sub>, and PM2.5 levels in the building is explored by designing a deep reinforcement learning (DRL) controller for a central HVAC system. The DRL controller is used to find an optimal solution to resolve the conflicts between those objectives.
- A complicated hybrid simulation model is developed for building dynamics including the HVAC system, the building envelope thermodynamics, indoor humidity ratio changes, and indoor CO<sub>2</sub> and PM2.5 variations.
- Additionally, the performance of the DRL controller is compared with two benchmarks, including a rule-based controller and an MPC controller, to justify its advantages in the co-optimization.

The rest of this paper is organized as follows. Section 2 introduces the methodology of the deep deterministic policy gradient (DDPG) algorithm, a type of DRL controller suitable for the co-optimization of

building HVAC systems. Then, Section 3 details the hybrid simulation model and the simulation setups for the DRL controller. Finally, Section 4 shows the performance of the DRL controller in comparison with the benchmarks, and also provides insights into the behavior of the DRL controller.

## 2. Methodology

This section introduces the methodology of RL algorithms. Section 2.1 discusses its several classifications with a couple of important terminologies. Then, Section 2.2 details the deep deterministic policy gradient (DDPG) algorithm, which is adopted in this study for building control.

### 2.1. Overview of reinforcement learning (RL)

Reinforcement learning (RL) is a branch of machine learning techniques that trains an intelligent agent to learn the optimal control policy in a dynamic environment with a trial-and-error approach. Generally, at the current state ( $s_t$ ) of the environment, the agent tries different control actions ( $a_t$ ) from a defined action space and receives a reward ( $r_t$ ) and the next state ( $s_{t+1}$ ) from the environment, and the target of the agent is to find the optimal policy to maximize the cumulative reward of each state. When deep learning models (artificial neural networks, ANN) are used in the development of RL, the algorithms are usually referred to as deep reinforcement learning (DRL).

RL algorithms can be categorized in multiple ways, such as 1) on-policy and off-policy, 2) online learning and offline learning, 3) model-based and model-free [16]. On-policy algorithms update themselves based on the feedback of the current policy, while off-policy algorithms make updates using data stored in the memory collected by other policies as well. Online learning refers to the process where the agent learns and updates its policy or value functions continuously as it interacts with the environment in real-time, while in offline learning, the agent learns from a fixed dataset of experiences which are not updated in real-time. Model-based algorithms learn a model of the environment by interacting with it and plans are taken based on the model, while model-free algorithms directly learn from the rewards of its actions. Table 2 shows the classification of RL algorithms.

Because model-based algorithms can simulate multiple future scenarios and make plans using the learned model, they are more sample efficient, learn the optimal policies faster, and can foresee and avoid risky actions by predicting their outcomes. However, their performance rely on the accuracy of the model, and incorrect predictions of the environment can lead to suboptimal policies. Because developing an accurate model of the environment can be complicated, those algorithms often suffer from intensive computation in the process of model training. On the other side, model-free algorithms learn policies or value functions directly from interactions with the environment, which are often more straightforward, simpler to implement, and thus can be applied to a wide range of problems. Although they also have some drawbacks, such as less sample efficient, training more slowly and

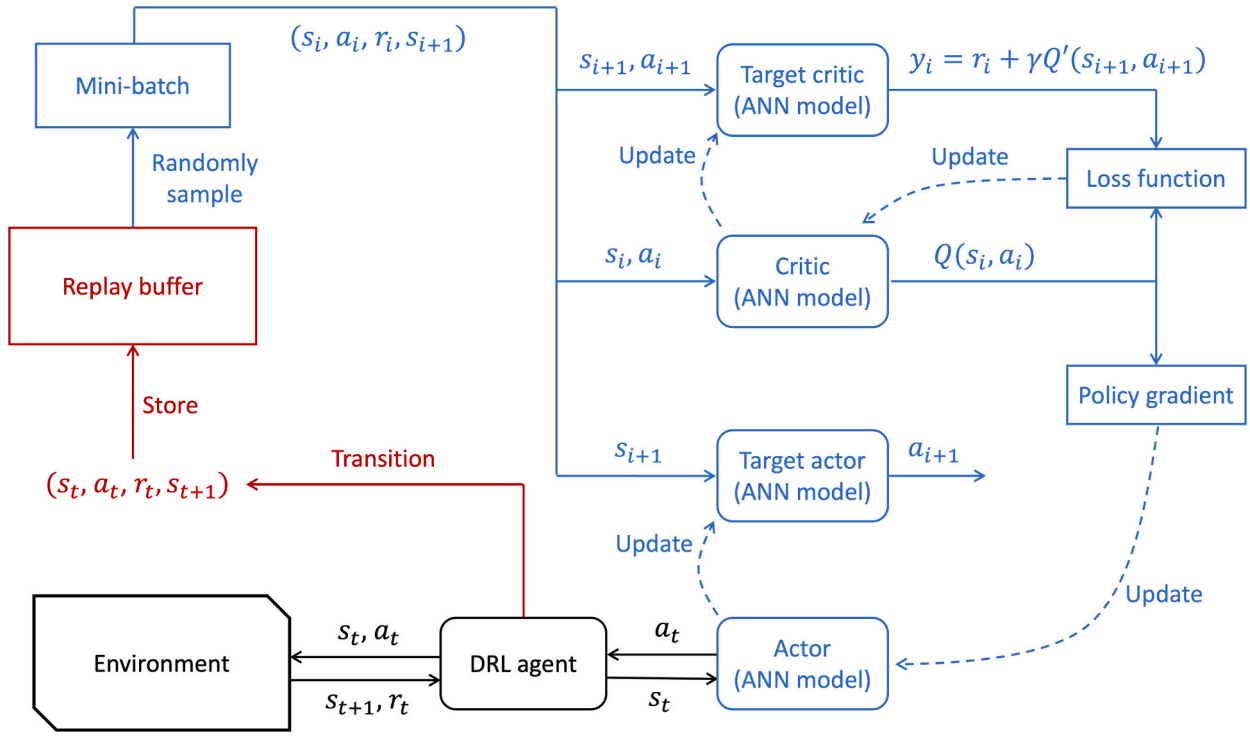


Fig. 1. Structure of the DDPG algorithm.

requiring a large amount of data, they are increasingly popular in control of many research areas.

Model-free algorithms can be further categorized into value-based, policy-gradient, and actor-critic algorithms [16]. Value-based algorithms (e.g., Q-learning) aim at learning the action-value function  $Q_\pi(s, a)$ , also called Q-function, which is defined as the expected cumulative reward for the agent starting from state  $s$ , first taking action  $a$ , and then following a policy  $\pi$ :

$$Q_\pi(s, a) = \mathbb{E}_\pi \left[ \sum_{k=0}^T \gamma^k r_{t+k+1} | S_t = s, A_t = a \right] \quad (1)$$

where  $T$  is the final step and  $\gamma$  is a discount factor. According to the Bellman optimality equation, the optimal Q-function can be written as:

$$Q_\pi^*(s_t, a_t) = r(s_t, a_t) + \gamma \max_{a_{t+1}} Q_\pi^*(s_{t+1}, a_{t+1}) \quad (2)$$

The Q-function usually takes  $s_t$  as input and outputs the Q-values for every possible  $a_t$ . In training, parameters ( $\theta^Q$ ) in the Q-function can be updated by:

$$L(\theta^Q) = L \left( r(s_t, a_t) + \gamma \max_a Q_\pi(s_{t+1}, a), Q_\pi(s_t, a_t) \right) \quad (3)$$

$$\theta^Q \leftarrow \theta^Q - \alpha \nabla_{\theta} L(\theta) \quad (4)$$

where  $L()$  is the loss function (e.g., mean squared error loss) and  $\alpha$  is the learning rate. The optimal control action at state  $s_t$  can be calculated by  $\arg\max_{a_t} Q(s_t, a_t)$ . Because the Q-values for every possible  $a_t$  need to be calculated, Q-learning is only applicable to problems with small and discrete action spaces. A typical Q-learning algorithm is Deep Q-Network (DQN) [35], which uses an ANN model to represent the Q-function.

In comparison, policy-gradient algorithms learn to approximate an optimal policy  $\pi^*$ , a policy function  $\pi_\theta(s_t)$  that maps  $s_t$  to  $a_t$ , and are applicable to continuous action spaces. In training, an estimate of the policy gradient is first computed and then stochastic gradient ascent is

used to optimize the parameters ( $\theta^\pi$ ) in the policy function:

$$J(\theta^\pi) = \mathbb{E}_{\pi_\theta} \left[ \sum_{k=0}^T \gamma^k r_{t+k+1} \right] \quad (5)$$

$$\theta^\pi \leftarrow \theta^\pi + \alpha \nabla_{\theta} J(\theta^\pi) \quad (6)$$

The actor-critic algorithms are a hybrid of the value-based and policy-gradient algorithms, such as Advantage Actor-Critic (A2C), Asynchronous Advantage Actor-Critic (A3C) [36], and Deep Deterministic Policy Gradient (DDPG) [37]. They are all well-suited for environments with continuous action spaces. Also, they are more sample-efficient compared to policy-gradient algorithms such as Proximal Policy Optimization (PPO). The next subsection will detail DDPG, a typical actor-critic algorithm.

## 2.2. Introduction to the deep deterministic policy gradient (DDPG) algorithm

Deep deterministic policy gradient (DDPG) is a model-free off-policy actor-critic DRL algorithm [37]. The algorithm concurrently learns an actor function and a critic function both described by deep learning models. The actor function maps a state ( $s_t$ ) to a deterministic action ( $a_t$ ), and the critic function maps  $s_t$  and  $a_t$  to a Q-value. Fig. 1 together with Algorithm 1 details the training process. The actor function interacts with the environment and the obtained  $(s_t, a_t, r_t, s_{t+1})$  pairs are stored in the replay buffer. For the actor to better explore the environment, noise generated by an Ornstein-Uhlenbeck (OU) process is added to its control actions. In each step of training, a mini-batch of data is randomly sampled from the replay buffer and used to train the ANN models. The critic function is trained by minimizing the mean squared error loss between  $y_i$  and  $Q(s_i, a_i)$  through batch gradient descent, and the actor function is trained by maximizing the Q-value predicted by the critic through batch gradient ascent. To improve the stability in training, DDPG introduces two target functions which are slowly updated by the actor and critic.

## Algorithm 1. The DDPG algorithm [37].

---

Randomly initialize critic  $Q(s_t, a_t|\theta^Q)$  and actor  $\pi(s_t|\theta^\pi)$  with weights  $\theta^Q$  and  $\theta^\pi$ .  
Initialize target network  $Q'$  and  $\pi'$  with weights  $\theta^{Q'} \leftarrow \theta^Q$  and  $\theta^{\pi'} \leftarrow \theta^\pi$ .  
Initialize replay buffer  $R$ .  
**for** episode = 1, ..., M **do**  
    Initialize an OU process  $\mathcal{N}_t$  for action exploration  
    Set initial observation state  $s_1$   
    **for** time step  $t = 1, \dots, T$  **do**  
        Obtain action  $a_t = \pi(s_t|\theta^\pi) + \mathcal{N}_t$   
        Execute action  $a_t$  and the hybrid model returns reward  $r_t$  and new state  $s_{t+1}$   
        Store transition  $(s_t, a_t, r_t, s_{t+1})$  in  $R$   
        Randomly sample a mini-batch of  $N$  transitions  $(s_i, a_i, r_i, s_{i+1})$  from  $R$   
        Compute  $y_i = r_i + \gamma Q'(s_{i+1}, \pi'(s_{i+1}|\theta^{\pi'}))|\theta^{Q'}$   
        Update  $\theta^Q$  by minimizing the loss:  $L = \frac{1}{N} \sum_{i=1}^N (y_i - Q(s_i, a_i|\theta^Q))^2$   
        Update  $\theta^\pi$  using the sampled policy gradient:  

$$\nabla_{\theta^\pi} J \approx \frac{1}{N} \sum_{i=1}^N \nabla_a Q(s, a|\theta^Q)|_{s=s_i, a=\pi(s_i)} \nabla_{\theta^\pi} \pi(s|\theta^\pi)|_{s=s_i}$$
  
        Update the target networks:  $\theta^{Q'} \leftarrow \tau \theta^Q + (1 - \tau) \theta^{Q'}$ ,  $\theta^{\pi'} \leftarrow \tau \theta^\pi + (1 - \tau) \theta^{\pi'}$   
    **end for**  
**end for**

---

### 3. Simulation setup

A central HVAC system usually has multiple continuous control variables in the chiller, AHU, and thermostats, which is difficult to tackle with value-based algorithms since they are only suitable for small and discrete action spaces. However, DDPG applies very well to problems featuring continuous action spaces. It has been applied to building energy systems in a few studies which shows its robustness, economy, and efficiency [24,26,30]. Therefore, this study chose DDPG as the DRL algorithm.

The DRL interacts with a hybrid simulation model to learn the optimal control actions. This step is typically essential before deploying the control algorithm in actual buildings, as a pre-trained controller can avoid sending poor or unreasonable commands to the real HVAC system in the beginning. As compared to real buildings in practical applications,

the simulation model has a higher degree of flexibility. Multiple scenarios can be simulated to study the system's behavior under different conditions. Also, hyperparameter tuning of the DRL algorithm is easier in simulation, as several DDPG algorithms with different settings can be trained at the same time. However, the behavior of real buildings and equipment is usually more complex and there are disturbances from multiple sources (e.g., change of indoor occupancy level, change of lighting and equipment load, disturbances of outdoor CO<sub>2</sub> level), making the controller harder to achieve optimal performance. In the simulation, the multiple sources of disturbances are modeled by random noises (detailed in Section 3.2).

This section first describes the control target and the control problem. Then, the hybrid simulation model is introduced, and the design of the DRL controller is detailed in every aspect, including the selection of state and actions, construction of the reward function, tuning the hyperparameters, etc. Finally, a supervisory rule-based controller and an MPC controller are introduced as benchmarks for comparison with the DRL controller. The simulation is implemented in Python, and the Keras package [38] is used for developing the DDPG algorithm.

#### 3.1. Description of the control problem

The simulation model is developed based on a real test cell located in FLEXLAB (shown in Fig. 2), an experimental test facility at Lawrence Berkeley National Lab. The test cell is arranged as an office room with a size of 56 m<sup>2</sup> (9.14 m × 6.09 m × 4.88 m). The eastern and western walls are interior walls connecting to other cells, and the northern and southern walls are exterior walls. The southern wall has a big window that receives solar irradiation during the daytime. The room is separated into 6 cubicles, with a desk, a chair, equipment, and a mannequin put in each cubicle. All equipment, mannequins, and lights are programmed as internal heat sources to create precise schedules of power outputs (see Fig. 7). An air purifier is placed in the room to provide fresh air when necessary.



Fig. 2. An internal view of the FLEXLAB test cell.

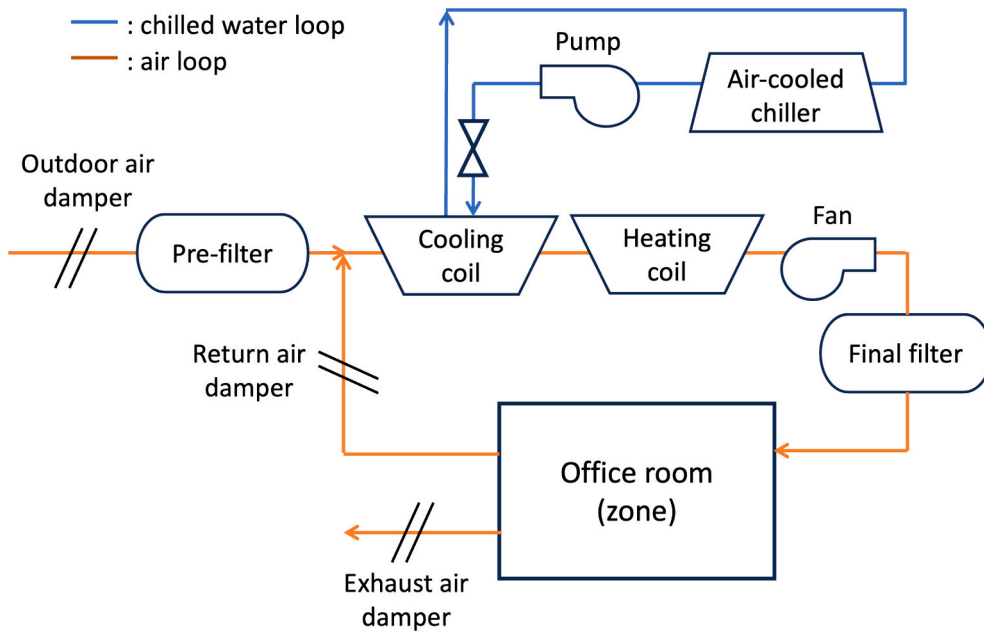


Fig. 3. Schematic diagram of the HVAC system of the test cell.

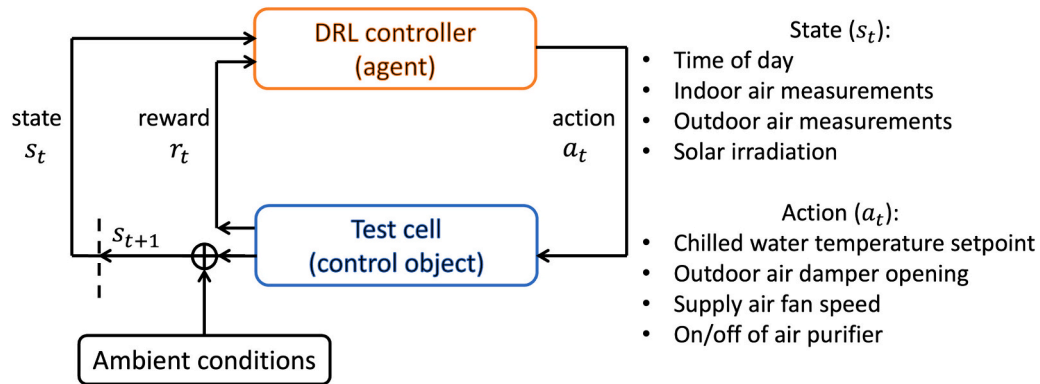


Fig. 4. Schematic diagram of the DRL-based control strategy.

Fig. 3 shows the HVAC system of the test cell, which consists of a chilled water loop<sup>1</sup> and an AHU. The air-cooled chiller is variable-speed. The chiller and the chilled water pump are both controlled by the difference between the chilled water supply temperature and its setpoint, which is set by a supervisory controller. In the AHU, the outdoor air damper opening and the supply air fan speed are both controlled by the supervisory controller. The return air and exhaust air dampers are interlocked with the outdoor air damper. The supply air temperature setpoint is adjusted by the difference between the zone (room) air temperature and its setpoint. In this study, the zone air temperature setpoint is set at 24 °C during the daytime and 29 °C during the nighttime, since 24 °C is generally regarded as a comfortable room temperature for most occupants (see Section 3.3.2 for the analysis of thermal comfort). Then, the supply air temperature is controlled by the cooling coil valve to match the setpoint. To remove the particulate matter entering the zone, a pre-filter is installed in the outdoor air duct and a final filter is installed in the supply air duct.

<sup>1</sup> The real HVAC system of the test cell has both primary and secondary chilled water loops, but in this simulation, we simplified it into one chilled water loop.

The test cell is controlled by the proposed DRL-based control strategy, as shown in Fig. 4. The control inputs (also known as the states for the DRL controller) include the outdoor conditions and measurements of sensors in the HVAC system and the zone. The control outputs (also known as the actions for the DRL controller) are the outdoor air damper opening, the supply air fan speed, the chilled water supply temperature setpoint, and ON/OFF of the air purifier. A detailed discussion of the state and action variables is in Section 3.3.1.

### 3.2. Development of the hybrid simulation model

A hybrid simulation model is developed based on the control problem described in the last section. As shown in Fig. 5, the simulation model receives the current state ( $s_t$ ) and the control action ( $a_t$ ), and predicts the next state ( $s_{t+1}$ ) along with other physical parameters from which reward ( $r_t$ ) is calculated. The hybrid model is comprised of five dynamic models, namely an HVAC system model, a building envelope model, a humidity ratio model, a CO<sub>2</sub> model, and a PM2.5 model. The HVAC system is modeled by the first principles of each individual component, and the variables are represented as continuous functions of time. The other four models represent the behavior of the office room, which are all lumped models that assume the system's behavior can be

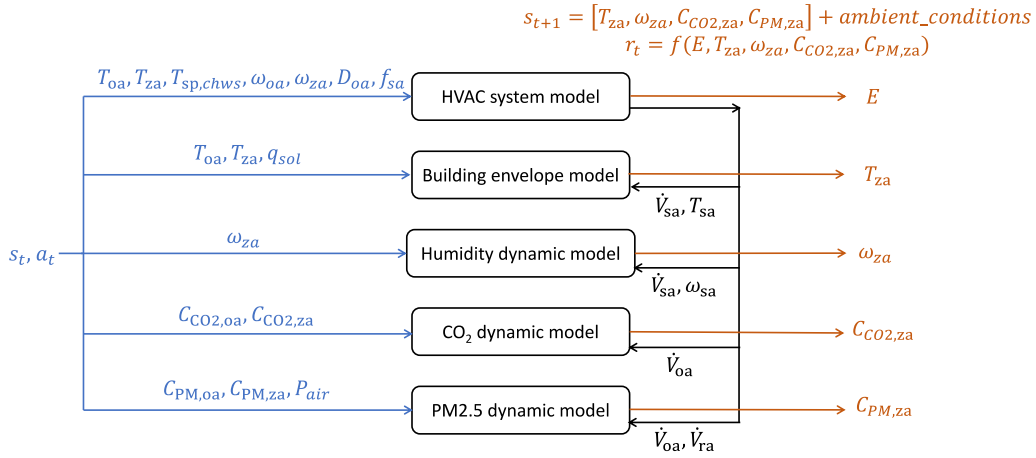


Fig. 5. Data-flow diagram of the hybrid simulation model.

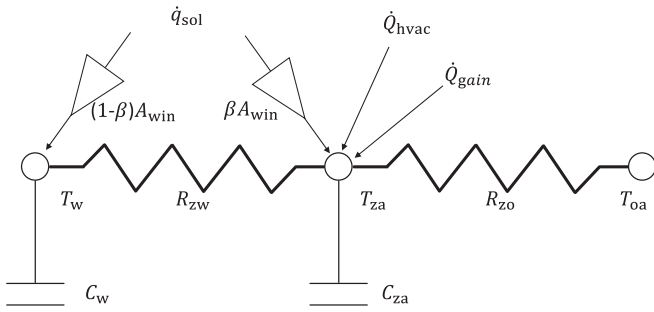


Fig. 6. Structure of the RC model for the FLEXLAB test cell.

described by considering only a few key components, neglecting spatial or temporal variations within the system. For example, in the building envelope model, the thermal mass and thermal conductivity are lumped into discrete elements. This section discusses the development of each model.

### 3.2.1. HVAC system model

A dynamic model for the HVAC system shown in Fig. 3 is developed by the Modelica simulation tool with the Buildings library [39]. In Modelica, components of the chiller, such as compressors, heat exchangers, fans, and sensors, are represented as individual modules or classes. These modules encapsulate the equations describing the behavior of each component, based on fundamental physical principles. Finally, connections are specified between the components and the individual modules are assembled into a cohesive system model. The

Modelica model is exported as a Functional Mock-up Unit (FMU) [40]. In this way, the FMU can be executed in the Python environment using the PyFMI package [41].

### 3.2.2. Building envelope model

The building envelope model simulates the dynamics of sensible heat transfer between each component in the building and its ambient environment. This study developed a 2-resistance-2-capacitance (2R2C) model for the test cell. The structure of the RC model is shown in Fig. 6 with its state parameters derived in Eq. (7) ~ (8). The model has been calibrated using real experimental data.

$$C_{za} \frac{dT_{za}}{dt} = \frac{T_w - T_{za}}{R_{zw}} + \frac{T_{oa} - T_{za}}{R_{zo}} + \dot{q}_{sol} \beta A_{win} + \dot{Q}_{hvac} + \dot{Q}_{gain} \quad (7)$$

$$C_w \frac{dT_w}{dt} = \frac{T_{za} - T_w}{R_{zw}} + \dot{q}_{sol} (1 - \beta) A_{win} \quad (8)$$

$$\dot{Q}_{hvac} = c_p \rho_a \dot{V}_{sa} (T_{sa} - T_{za}) \quad (9)$$

Table 3

Simulation settings of the office room.

| Parameter                                | Value                                                |
|------------------------------------------|------------------------------------------------------|
| Room area                                | 56 m <sup>2</sup>                                    |
| Number of occupants                      | 6                                                    |
| Lighting full load                       | 465 W                                                |
| Equipment plugs full load                | 865 W                                                |
| Occupant sensible load                   | 120 W/person [42]                                    |
| Occupant moisture generation rate        | 2.43 × 10 <sup>-5</sup> kg/s/person [43]             |
| Occupant CO <sub>2</sub> generation rate | 5.2 × 10 <sup>-6</sup> m <sup>3</sup> /s/person [44] |

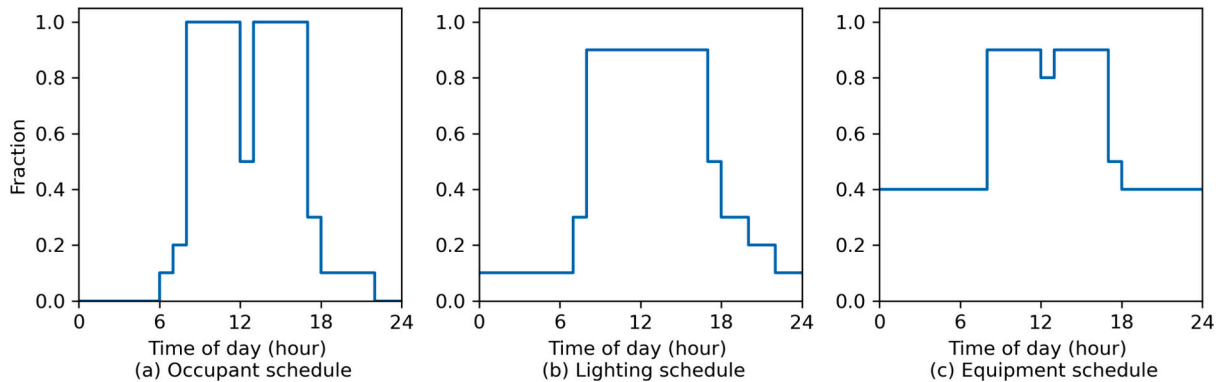


Fig. 7. Weekday occupant, lighting, and equipment schedules in the office room [42].



**Table 4**

Simulation settings regarding the PM2.5 dynamic model.

| Parameter                                       | Value                          | Note                                                                 |
|-------------------------------------------------|--------------------------------|----------------------------------------------------------------------|
| Efficiency of the pre-filter ( $\eta_1$ )       | 0.07 [47]                      | MERV 6 filter                                                        |
| Efficiency of the final filter ( $\eta_2$ )     | 0.26 [47]                      | MERV 8 filter                                                        |
| Particle deposition rate ( $k_d$ )              | $1.94 \times 10^{-4}$ 1/s [46] |                                                                      |
| Clean air delivery rate (CADR)                  | 0.181 m <sup>3</sup> /s        | Data acquired from an air purifier product on the market             |
| Indoor PM2.5 generation rate ( $\dot{S}_{PM}$ ) | 0                              | Assuming particulate matter generated from an office room is minimal |

**Table 5**

Range of each action variable.

| Control variable                                           | Daytime, occupied (6:00–20:00) | Nighttime, unoccupied (20:00–6:00) |
|------------------------------------------------------------|--------------------------------|------------------------------------|
| Chilled water supply temperature setpoint ( $T_{sp,chw}$ ) | 5 °C ~ 15 °C [48]              | 15 °C                              |
| Outdoor air damper opening ( $D_{oa}$ )                    | 20 % ~ 100 %                   | 0–30 %                             |
| Supply air fan speed ( $f_{sa}$ )                          | 10 % ~ 100 %                   | 10 %                               |
| Air purifier operating condition ( $P_{air}$ )             | ON/OFF                         | ON/OFF                             |

where  $T_{za}$  is the zone air temperature [°C],  $T_w$  is the building thermal mass temperature [°C],  $T_{oa}$  is the outdoor air temperature [°C],  $\dot{q}_{sol}$  is the incident solar radiation on the window [W/m<sup>2</sup>],  $A_{win}$  is the effective window size considering a fraction of radiation is reflected [m<sup>2</sup>],  $\beta$  is the fraction of incident solar radiation absorbed by the zone air,  $\dot{Q}_{hvac}$  is the sensible heat transferred into the zone by mechanical cooling, and  $\dot{Q}_{gain}$  is the internal sensible heat generation including the equipment plug load, the occupant load, and the lighting load. This study adopts schedules of the occupants, equipment, and lighting from the U.S. Department of Energy reference building EnergyPlus models [42], as shown in Fig. 7. In practice, if the occupancy and the plug load cannot be monitored, they can be estimated from the daily work schedules. Other important simulation settings of the office room are shown in Table 3.

Additionally, to emulate the real working environment in offices, the simulation settings of the office room are not fixed values. We also added Gaussian noises to the occupant, lighting, and equipment schedules:

$$x \leftarrow x + \varphi \quad (10)$$

where  $x$  can be either the heat load, moisture, or CO<sub>2</sub> generation rate, and  $\varphi$  is a random variable that follows the Gaussian distribution,  $\varphi \sim N(1, 0.1^2)$ .

### 3.2.3. Dynamic model of zone air humidity ratio

Based on the moisture mass balance, an ordinary differential equation can be used to describe the variation of the zone air humidity ratio [45]:

$$\rho_a V \frac{d\omega_{za}(t)}{dt} = \rho_a \dot{V}_{sa}(\omega_{sa}(t) - \omega_{za}(t)) + \dot{S}_w(t) \quad (11)$$

where  $\rho_a$  is the air density [kg/m<sup>3</sup>],  $V$  is the zone air volume [m<sup>3</sup>],  $\dot{V}_{sa}$  is the supply air volume flow rate [m<sup>3</sup>/s],  $\omega_{za}$  is the zone air humidity ratio [kg H<sub>2</sub>O/kg dry air],  $\omega_{sa}$  is the supply air humidity ratio [kg H<sub>2</sub>O/kg dry air], and  $\dot{S}_w$  is the indoor moisture generation rate by occupants [kg/s]. The air infiltration, the absorption and desorption of moisture by building interior materials, moisture removal by condensation, and moisture added by evaporation are not considered in the simulation. But to account for their effects, some Gaussian noises are added to  $\dot{S}_w$ :

$$\dot{S}_w \leftarrow \dot{S}_w + \varphi \quad (12)$$

where  $\varphi$  is a random variable that follows the Gaussian distribution,  $\varphi \sim N(1, 0.05^2)$ .

### 3.2.4. Dynamic model of zone air CO<sub>2</sub> level

The dynamic model for the variation of zone air CO<sub>2</sub> level can be expressed by an ordinary differential equation:

$$V \frac{dC_{CO2,za}(t)}{dt} = \dot{V}_{oa}(C_{CO2,oa}(t) - C_{CO2,za}(t)) + 10^6 \bullet \dot{S}_{CO2}(t) \quad (13)$$

where  $V$  is the zone air volume [m<sup>3</sup>],  $C_{CO2,za}$  is the zone air CO<sub>2</sub> concentration [ppm],  $C_{CO2,oa}$  is the outdoor air CO<sub>2</sub> concentration [ppm],  $\dot{V}_{oa}$  is the outdoor air volume flow rate [m<sup>3</sup>/s], and  $\dot{S}_{CO2}$  is the indoor CO<sub>2</sub> generation rate by occupants [m<sup>3</sup>/s].

### 3.2.5. Dynamic model of zone air PM2.5 concentration

Based on the mass balance of indoor particulate matter, the dynamics model for the variation of zone air PM2.5 level can also be expressed by an ordinary differential Eq. [46]:

$$V \frac{dC_{PM,za}(t)}{dt} = \left( \dot{V}_{oa}(1 - \eta_1)(1 - \eta_2) \right) \bullet C_{PM,oa}(t) - \left( \dot{V}_{oa} + \eta_2 \dot{V}_{ra} + k_d V + CADR \right) \bullet C_{PM,za}(t) + \dot{S}_{PM}(t) \quad (14)$$

where  $V$  is the zone air volume [m<sup>3</sup>],  $C_{PM,za}$  is the zone air PM2.5 level [μg/m<sup>3</sup>],  $C_{PM,oa}$  is the outdoor air PM2.5 level [μg/m<sup>3</sup>],  $\dot{V}_{oa}$  is the outdoor air volume flow rate [m<sup>3</sup>/s],  $\dot{V}_{ra}$  is the return air volume flow rate [m<sup>3</sup>/s],  $\eta_1$  is the efficiency of the pre-filter,  $\eta_2$  is the efficiency of the final filter (see Fig. 3 for the locations of the two filters in the air handling unit),  $k_d$  is the particle deposition rate onto indoor surfaces [1/s], CADR is the clean air delivery rate of the air purifier [m<sup>3</sup>/s], and  $\dot{S}_{PM}$  is the indoor PM2.5 generation rate [μg/s]. Settings of those parameters are shown in Table 4.

## 3.3. DRL controller design

### 3.3.1. State and actions

Based on the inputs and outputs of the hybrid model shown in Fig. 5, the state vector consists of 10 variables, namely: the time of day, outdoor air temperature ( $T_{oa}$ ), outdoor air humidity ratio ( $\omega_{oa}$ ), incident solar radiation on the window ( $q_{sol}$ ), outdoor air CO<sub>2</sub> concentration ( $C_{CO2,oa}$ ), outdoor air PM2.5 concentration ( $C_{PM,oa}$ ), zone air temperature ( $T_{za}$ ), zone air humidity ratio ( $\omega_{za}$ ), zone air CO<sub>2</sub> concentration ( $C_{CO2,za}$ ), and zone air PM2.5 concentration ( $C_{PM,za}$ ). The time of day guides the DRL controller to learn the occupant, lighting, and equipment schedules. The other variables are all highly related to the cooling load, thermal comfort, and indoor air quality of the office room.



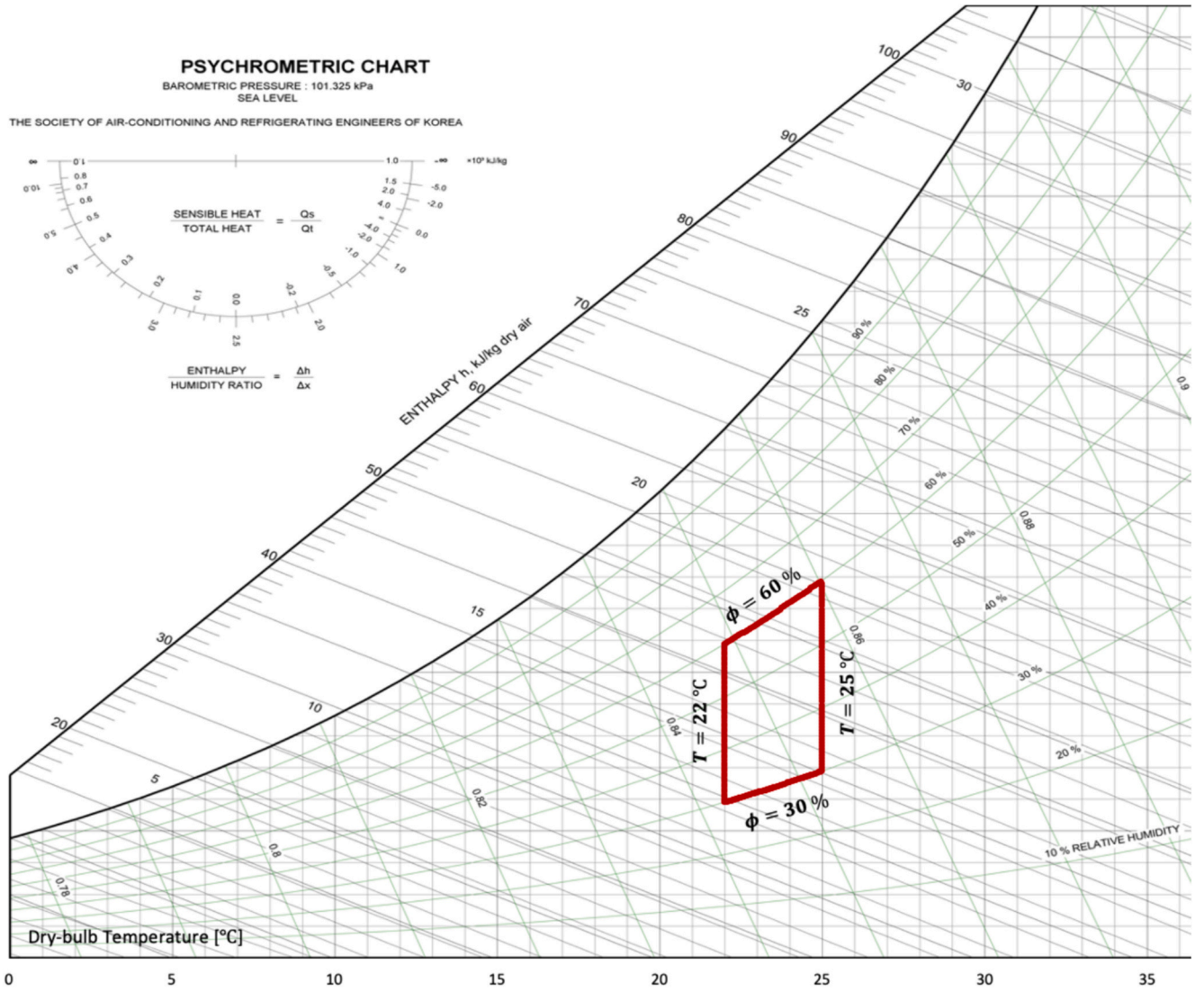


Fig. 8. Thermal comfort zone for occupants inside the office building.

The action vector consists of 4 control variables, which are chilled water supply temperature setpoint ( $T_{sp,chw}$ ), outdoor air damper opening ( $D_{oa}$ ), supply air fan speed ( $f_{sa}$ ), and air purifier ON/OFF ( $P_{air}$ ). They are critical for the control objectives including energy saving, temperature, humidity, and indoor air quality control. A higher  $T_{sp,chw}$  can reduce the energy consumption of the chiller but increase of power of the fan. Also, a very high  $T_{sp,chw}$  could lead to poor humidity control. Therefore, there is a balance between  $T_{sp,chw}$  and  $f_{sa}$ , which can be learned by the DRL controller to save more energy.  $D_{oa}$  is directly related to the ventilation load and indoor air quality, which is an important control variable. Similarly,  $P_{air}$  is important for indoor air quality control. The range of each action variable during the daytime (occupied) and the nighttime (unoccupied) is shown in Table 5. The control time step is 15min.

### 3.3.2. Reward functions

The DRL controller aims at co-optimizing the system energy consumption, indoor thermal comfort, and indoor air quality during the occupied time (6:00–20:00). During the unoccupied time (20:00–6:00), only the energy consumption is attempted to be minimized. For the indoor CO<sub>2</sub> level, generally, a value below 1000 ppm is considered a safe level [49]. For the PM<sub>2.5</sub> level, a new standard proposed by the Environmental Protection Agency in 2023 [50] sets the long-term safety threshold at 10 µg/m<sup>3</sup>. In terms of thermal comfort, a very commonly

used index is the predicted mean vote (PMV) and the predicted percentage of dissatisfied (PPD) proposed by Fanger [51]. The PMV index evaluates human thermal comfort by integrating six parameters: air temperature, air velocity, humidity, mean radiant temperature, metabolic rate, and clothing level. Later, it was modified by a few other researchers to extend its generalizability [52]. A PMV value closer to zero is considered better, and a tolerance of  $\pm 0.5$  around zero is adopted by ASHRAE standard 55 [53], ISO 7730 standard [54], and the EN 15251 standard [55]. PPD is computed based on PMV to represent the percentage of occupants that are dissatisfied with a certain thermal comfort. Fig. 8 shows the zone of thermal comfort adopted in this study on a psychrometric chart, which has dry-bulb temperature between 22 °C and 25 °C and relative humidity between 30 % and 60 %. Because the simulation model does not have a heating device and a humidifier, the bottom and left boundaries are relaxed. That is, we mainly focus on the cooling and dehumidification performance of the HVAC system. Based on the typical clothing level, metabolic rate, and air speed in an office building, the PMV values in this comfort zone range from  $-0.16$  to  $-0.52$ , which has PPD lower than 11 %.

Based on the above standards, the reward function during the occupied time contains 5 components, including the penalty of total energy consumption of the HVAC system, the penalty of violating indoor temperature and relative humidity comfort zone, and the penalty of

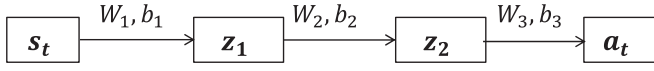


Fig. 9. Architecture of the actor model in DDPG.

indoor CO<sub>2</sub> and PM2.5 concentration exceeding acceptable levels. It can be expressed by:

$$r_t = -\alpha_1 E(t+1) - \alpha_2 f(T_{za}(t+1)) - \alpha_3 f(\phi_{za}(t+1)) - \alpha_4 f(C_{CO2,za}(t+1)) - \alpha_5 f(C_{PM,za}(t+1)) \quad (15)$$

where

$$E(t) = E_{chiller}(t) + E_{pump}(t) + E_{fan}(t) + E_{purifier}(t) \quad (16)$$

$$f(T_{za}(t)) = [T_{za}(t) - T_u]_+ + [T_l - T_{za}(t)]_+ \quad (17)$$

$$f(\phi_{za}(t)) = [\phi_{za}(t) - \phi_u]_+ \quad (18)$$

$$f(C_{CO2,za}(t)) = \begin{cases} 0 & \text{if } C_{CO2,za}(t) < 1000 \text{ ppm} \\ 1 & \text{if } C_{CO2,za}(t) \geq 1000 \text{ ppm} \end{cases} \quad (19)$$

$$f(C_{PM,za}(t)) = \begin{cases} 0 & \text{if } C_{PM,za}(t) < 10 \mu\text{g}/\text{m}^3 \\ 1 & \text{if } C_{PM,za}(t) \geq 10 \mu\text{g}/\text{m}^3 \end{cases} \quad (20)$$

In Eq. (15),  $\alpha_1 \sim \alpha_5$  are critical hyperparameters that determine the relative importance of each penalty component. For instance, if  $\alpha_3$  is very small, then DRL controller may not care about the humidity at all,

and as a result, the indoor relative humidity may be very high. The idea is to minimize energy consumption while putting constraints on the other factors. In other words, a relatively large penalty is given when the thermal comfort is not satisfied or the indoor air quality is not healthy, so as to prevent the controller from violating the thermal comfort constraints and indoor air quality safety level, meanwhile guiding the controller to use the least energy. Based on the magnitude of those physical quantities in different units, a grid search is performed for  $\alpha_2$  and  $\alpha_3$  to select their best values. In this process,  $\alpha_2$  is searched from 1.0 to 5.0 with an increment of 0.5, and  $\alpha_3$  is searched from 2.0 to 10.0 with an increment of 1.0. Finally, they are set as follows:  $\alpha_1 = 1.0$ ,  $\alpha_2 = 2.5$ ,  $\alpha_3 = 5.0$ ,  $\alpha_4 = 1.0$ ,  $\alpha_5 = 1.0$ .

Eq. (16) penalizes the energy consumption of the chiller, chilled water pump, supply air fan, and air purifier. Eq. (17) penalizes the temperature discomfort if the zone air temperature is above  $T_u = 24.5^\circ\text{C}$  or below  $T_l = 22^\circ\text{C}$ . This upper threshold is  $0.5^\circ\text{C}$  above the zone air cooling setpoint considering the deadband in the control system, and the lower threshold is to prevent the room from being overcooled in the economizer mode when the outside is cold. Eq. (18) penalizes the discomfort caused by high relative humidity, where the upper bound  $\phi_u$  is 60 %. Because this HVAC system does not have a humidifier, no penalty is imposed on low relative humidity. When the zone air temperature or humidity rises or drops beyond their respective thresholds, the penalty of thermal comfort will instantly increase to a level much higher than the penalty of energy consumption. In this way, comfort is guaranteed and the energy consumption is minimized. Eq. (19) penalizes the indoor CO<sub>2</sub> level when it exceeds 1000 ppm. In Eq. (20), when the indoor PM2.5 level is above  $10 \mu\text{g}/\text{m}^3$ , a large penalty is imposed.

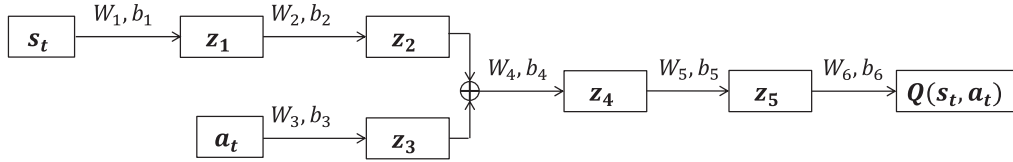


Fig. 10. Architecture of the critic model in DDPG.

Table 6

A list of hyperparameters in training.

| Hyperparameter                                   | Value                | Range in the grid search               |
|--------------------------------------------------|----------------------|----------------------------------------|
| Learning rate of the critic network              | $5 \times 10^{-5}$   | $[2 \times 10^{-5}, 1 \times 10^{-3}]$ |
| Learning rate of the actor network               | $2.5 \times 10^{-5}$ | $[1 \times 10^{-5}, 5 \times 10^{-4}]$ |
| Size of the replay buffer                        | $1.5 \times 10^6$    | —                                      |
| Size of a mini-batch ( $N$ )                     | 128                  | —                                      |
| Discount factor for future rewards ( $\gamma$ )  | 0.98                 | —                                      |
| The rate to update the target network ( $\tau$ ) | 0.005                | —                                      |
| Reversion rate parameter in the OU process       | 0.15                 | —                                      |
| Standard deviation parameter in the OU process   | 0.2                  | —                                      |

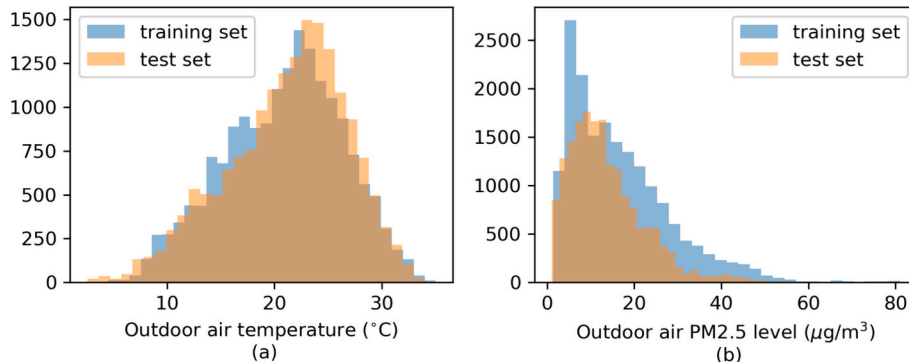


Fig. 11. A comparison of the weather and air quality datasets used in training and testing.

**Table 7**

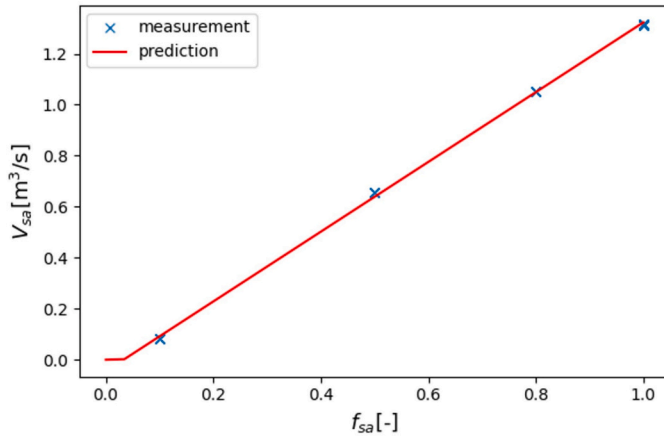
Design of the rule-based controller (RBC).

| Control variable                                           | Daytime,<br>occupied<br>(6:00–20:00)                                                                                                                                                                                                                                                                                      | Nighttime, unoccupied (20:00–6:00) |
|------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------|
| Chilled water supply temperature setpoint ( $T_{sp,chw}$ ) | 7 °C                                                                                                                                                                                                                                                                                                                      | 7 °C                               |
| Supply air fan speed ( $f_{sa}$ )                          | 30 %<br>if <i>economizer</i> == ON:<br>$D_{oa} = 100\%$ or modulating<br>if <i>economizer</i> == OFF:<br>from 6:00 to 8:00:<br>$D_{oa} = 20\%$<br>from 8:00 to 12:00:<br>$D_{oa} = 45\%$<br>from 12:00 to 13:00:<br>$D_{oa} = 40\%$<br>from 13:00 to 17:00:<br>$D_{oa} = 45\%$<br>from 17:00 to 20:00:<br>$D_{oa} = 20\%$ | 10 %                               |
| Outdoor air damper opening ( $D_{oa}$ )                    | if $C_{PM,oa} \geq 10\mu\text{g}/\text{m}^3$ :<br>$P_{air} = \text{ON}$<br>else:<br>$P_{air} = \text{OFF}$                                                                                                                                                                                                                | 0                                  |
| Air purifier operating condition ( $P_{air}$ )             |                                                                                                                                                                                                                                                                                                                           | OFF                                |

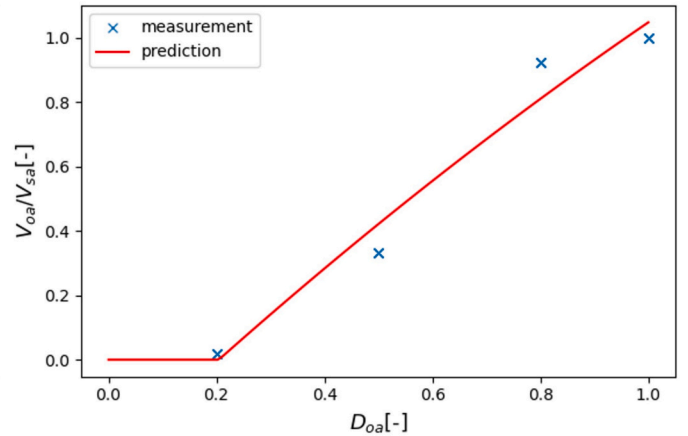
### 3.3.3. Architecture of the critic and actor networks

Figs. 9 and 10 show the architecture of the ANN model for the actor and critic respectively in the DDPG algorithm. The actor model has two fully-connected hidden layers ( $z_1$  and  $z_2$ ), and each layer has 256

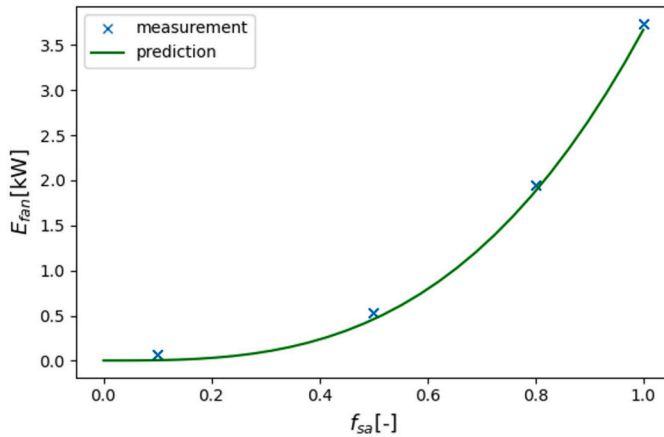
neurons. In the critic model, layer  $z_1$  has 16 neurons,  $z_2$  and  $z_3$  have 32 neurons,  $z_4$  and  $z_5$  have 256 neurons. The Adam optimizer [56] is used in training.



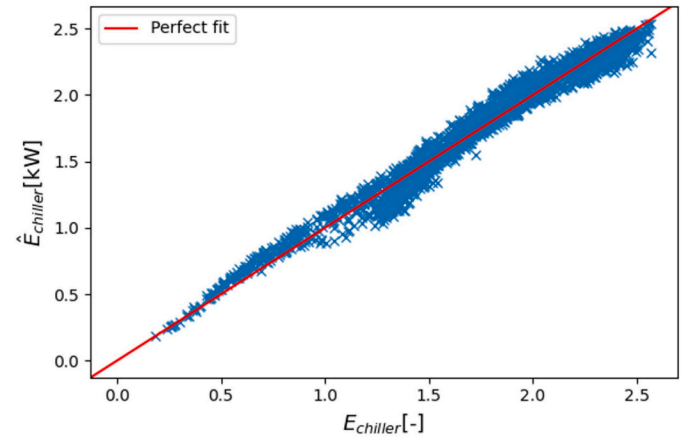
(a) Fan volume flow



(b) OA fraction



(c) Fan power



(d) Chiller prediction

**Fig. 12.** Prediction vs. measurement of HVAC component models.

**Table 8**

Accuracy of the HVAC system model.

| Parameter                        | RMSE                    | MAPE  |
|----------------------------------|-------------------------|-------|
| Chilled water mass flow rate     | 0.003 kg/s              | 1.2 % |
| Chilled water return temperature | 0.64 °C                 | 3.3 % |
| Pump power                       | 0.002 kW                | 1.9 % |
| Fan power                        | 0.006 kW                | 2.1 % |
| Chiller load                     | 0.65 kWh                | 9.9 % |
| Supply air volume flow rate      | 0.014 m <sup>3</sup> /s | 6.9 % |
| Supply air temperature           | 1.0 °C                  | 4.7 % |

### 3.3.4. Hyperparameters tuning

The previous two subsections discussed the hyperparameter settings in the reward function and in the ANN models. However, there are also a few important hyperparameters in the training process of the DRL algorithm (see Algorithm 1). This study referred to the hyperparameter settings in [30] and conducted a grid search for the learning rates since they are the most important hyperparameters. Table 6 lists each hyperparameter and its values set in training.

### 3.3.5. Exogenous inputs in training and testing

The exogenous inputs in the simulation are outdoor weather conditions, outdoor CO<sub>2</sub> concentration, and outdoor PM<sub>2.5</sub> concentration:

- The weather data including outdoor air temperature, humidity, and solar irradiation come from typical meteorological year (TMY) data in Seoul, South Korea. TMY datasets collected by two weather stations are used for training and testing respectively, and their distributions in temperature are shown in Fig. 11a. Note that the TMY datasets are built up by meteorological measurements made at hourly intervals between 1973 and 2014 for the training set, and between 1950 and 2021 for the test set. Because the developed HVAC system model lacks a heating device, only half a year from May to October is simulated in both training and testing.
- The outdoor air CO<sub>2</sub> concentration usually varies between 400–500 ppm [57–59]. This amount of daily and seasonal variation (usually around 50 ppm) is much lower than the safe CO<sub>2</sub> threshold in the room, which equals 1000 ppm as discussed in Section 3.3.2. In other words, the variation of outdoor air CO<sub>2</sub> concentration has very little impact on the control of the indoor CO<sub>2</sub> level, as compared to the CO<sub>2</sub> generated by occupants. Therefore, in this study, the outdoor air CO<sub>2</sub> concentration in Seoul is modeled as a Gaussian distribution with the mean equal to 450 ppm and the standard deviation equal to 10 ppm for both the training and the test set.
- For the outdoor PM<sub>2.5</sub> concentration, the air quality data in Seoul recorded in 2018 is used for training, and the data recorded in 2022 is used for testing. The distributions of the two air quality datasets are shown in Fig. 11b.

## 3.4. Benchmark controller design

To compare with the DRL controller, a supervisory rule-based controller (RBC) and an MPC controller are developed separately to solve the control problem in this study, and their performance is compared with the DRL controller for benchmarking purposes.

### 3.4.1. Supervisory rule-based controller (RBC)

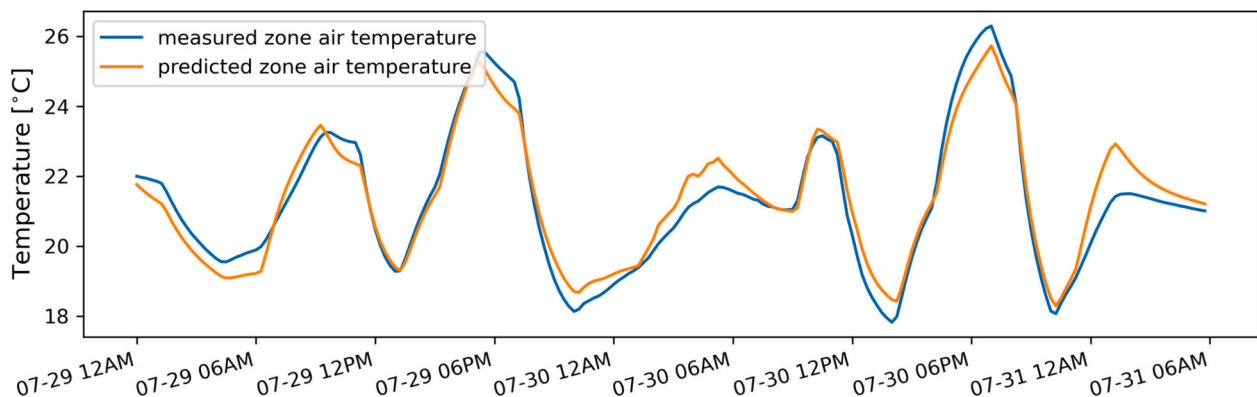
The design of the supervisory rule-based controller (RBC) is shown in Table 7. The chilled water temperature setpoint is fixed at 7 °C during the daytime and the nighttime, which is the empirical industrial practice [60]. The fan speed is fixed at 30 % during the daytime considering the occupancy level and the cooling demand in summer. The outdoor air damper follows the daily occupancy schedule, which has a larger opening when the office has more occupants. If possible, the economizer mode is switched on to help cool the office room using the outside air. The economizer is enabled basically when the outdoor air temperature is lower than the return air temperature and the outdoor air humidity ratio is lower than the dewpoint humidity ratio of the supply air temperature setpoint, but several other rules are added to prevent overcooling. The air purifier is turned on when the outdoor PM<sub>2.5</sub> concentration is greater than 10 µg/m<sup>3</sup>. The control time step is 15min, same as the DRL controller.

### 3.4.2. MPC controller

As discussed in the Introduction, the development of a control-oriented model (e.g., gray-box model) for a large and complicated system is time-consuming and requires the engineer's advanced knowledge. In addition, the direct use of the off-the-shelf building energy models (e.g., EnergyPlus or Modelica) for the MPC is practically limited because optimizing objective function over those models is a non-linear optimization problem, which is typically time-consuming and sensitive to tuning parameters, especially, when the system is complex with a long prediction horizon. Since a thorough comparison between RL and MPC is out of the scope of this study, we developed an MPC controller from a practical perspective considering those limitations for the performance benchmarking purpose. We will also present the practical limitations of the application of the MPC compared to RL controller in the Result section.

The hybrid model structure in Fig. 5 is directly used for the MPC except for the HVAC model. The MPC tries to find optimal trajectory control variables over the prediction horizon to minimize the negative reward (Eq. 21).

$$\min_{T_{sp, chws}(j), D_{oa}(j), f_{ia}(j), P_{air}(j)} \sum_{j=1}^{N_p} -r(t+j) \quad (21)$$



**Fig. 13.** Calibration of the building envelope model: a comparison of the measured zone air temperature data with predictions by the building envelope model.



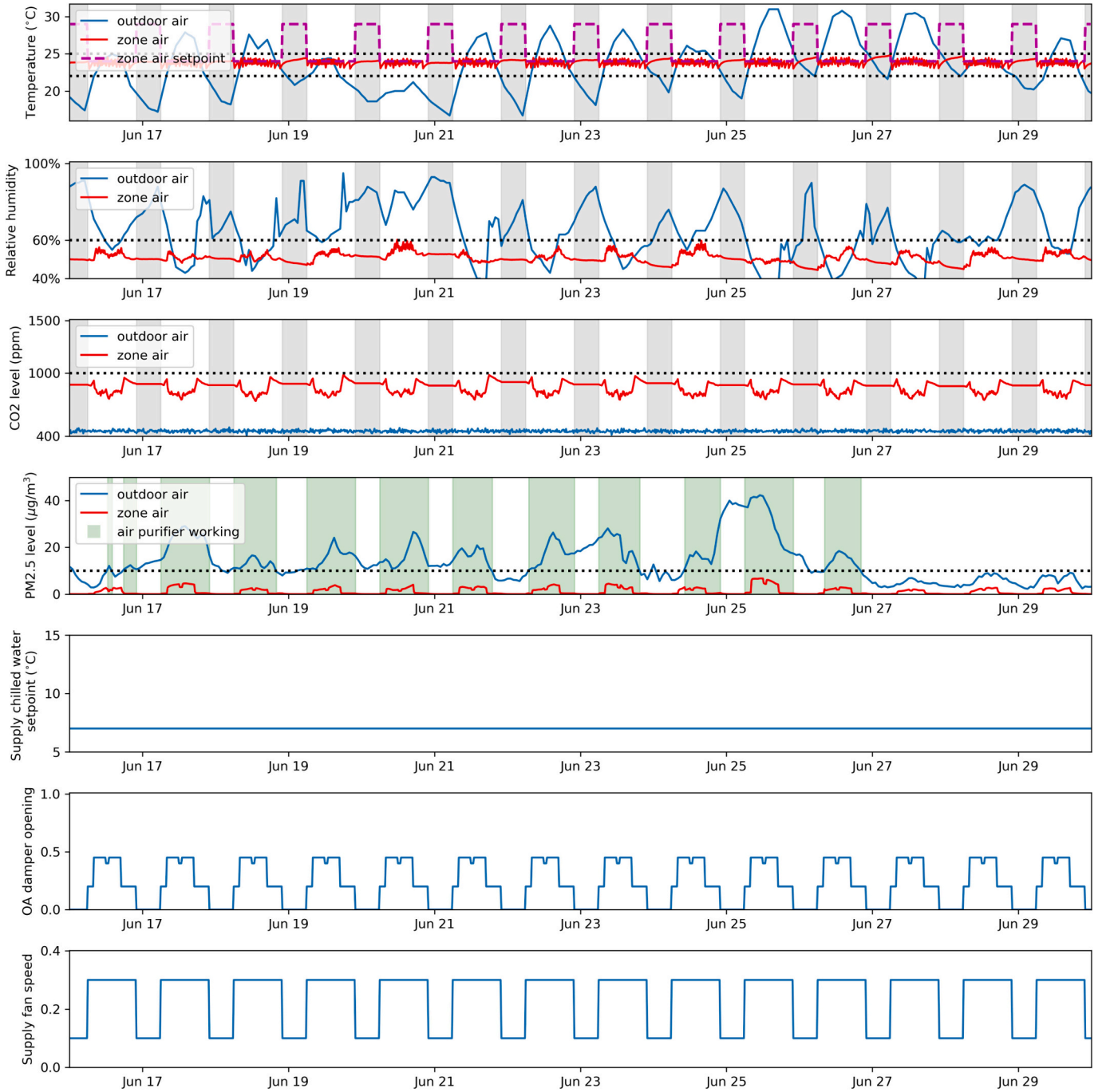


Fig. 14. Performance of the RBC controller on the test set from June 16th to June 30th.

where  $j$  is the prediction control timestep,  $N_p$  is a prediction horizon,  $T_{sp, chws}$ ,  $D_{oa}$ ,  $f_{sa}$ , and  $P_{air}$  are control variables and bounded by the ranges presented in Table 5.

The HVAC model is replaced by a steady-state performance model (Fig. 12) because of the computation time of FMU within the optimization solver and difficulties in state initialization of the Modelica model. Given the control variables (i.e.,  $D_{oa}$  and  $f_{sa}$ ),  $\dot{V}_{sa}$ ,  $\dot{V}_{oa}$ , and  $E_{fan}$  are predicted via Eq. (22) ~ (24), respectively.

$$\dot{V}_{sa} = \max(0, \beta_{0, \dot{V}_{sa}} + \beta_{1, \dot{V}_{sa}} f_{sa}) \quad (22)$$

$$\dot{V}_{oa} = \max(0, (\beta_{0, \dot{V}_{oa}} + \beta_{1, \dot{V}_{oa}} D_{oa}) \hat{\dot{V}}_{sa}) \quad (23)$$

$$E_{fan} = \max(0, \beta_{0, E_{fan}} + \beta_{1, E_{fan}} f_{sa} + \beta_{2, E_{fan}} f_{sa}^2 + \beta_{3, E_{fan}} f_{sa}^3) \quad (24)$$

where  $\beta$ 's are regression parameters.

The chiller energy consumption ( $E_{chiller}$ ) is predicted via EnergyPlus Chiller component model [61]. This model requires the chilled water setpoint, outdoor air temperature, and chiller load as inputs.

In Fig. 5, the FMU simulates the AHU's behavior (damper and cooling coil control) and provides the conditions of the supply air, which are used for the input for the other models. However, this steady-state performance model does not include the control mechanism inside, and the required chiller load at the control time step is iteratively calculated. To be specific, the supply air temperature is regulated by the cooling coil to meet its setpoint (12.5 °C). However, because the



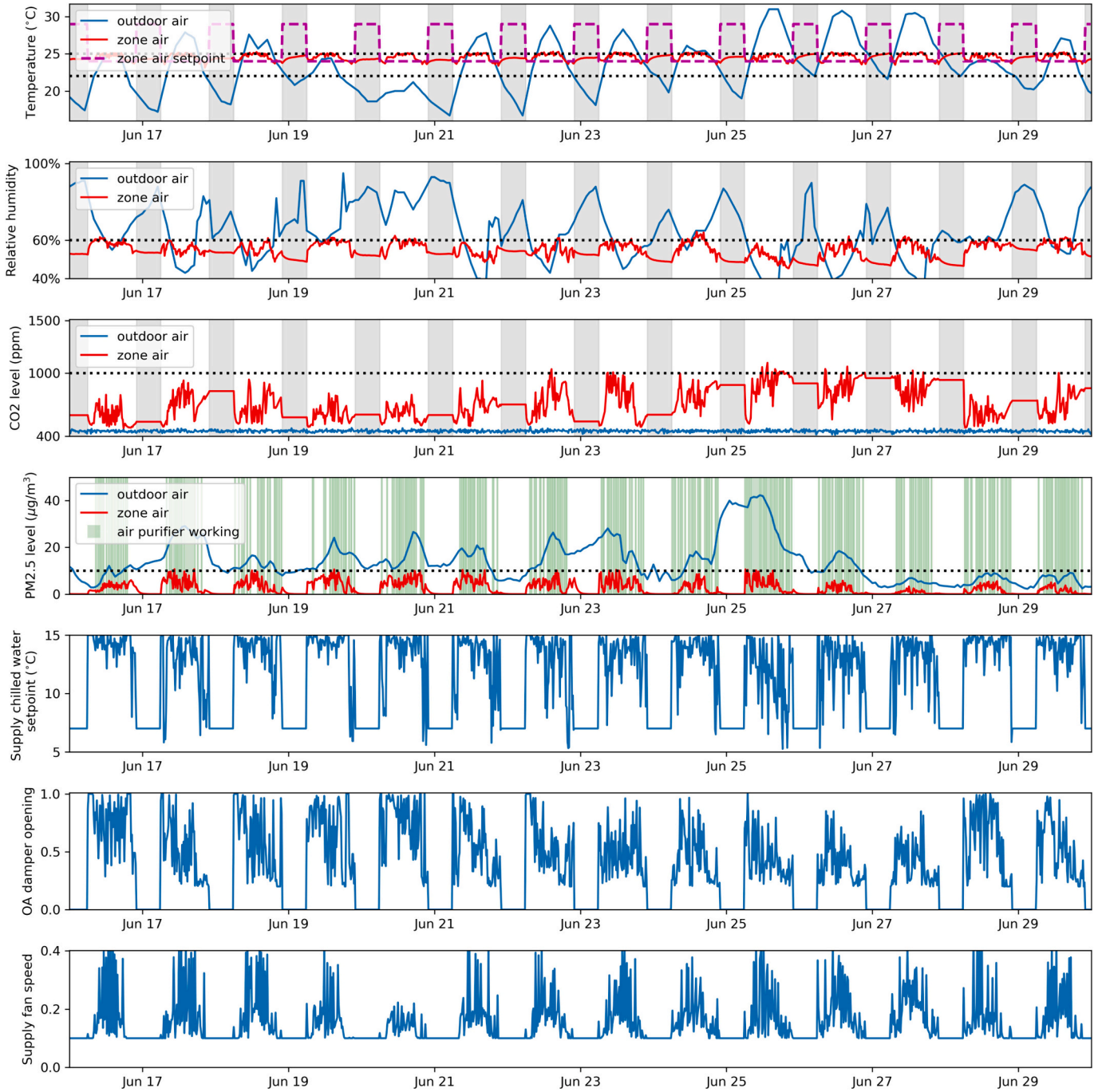


Fig. 15. Performance of the MPC controller on the test set from June 16th to June 30th.

dehumidification is involved, the exact supply air temperature is unknown. The calculation is achieved in the following procedure. If the full chiller load can cool down the mixed air temperature lower than the supply air temperature setpoint (i.e., 12.5 °C) and its dewpoint, it is a dehumidification mode, and the relative humidity of the supply air is assumed to be controlled to 90 %, which is obtained based on the measured data. Except for this situation, the coil is modeled as a sensible cooling-only, and the chiller load is calculated based on the calculated supply air state.

With all the information, the optimization problem is solved every MPC sampling time (15min) via derivative-free optimization (Scipy's differential evolution algorithm [62]). In each optimization step, the initial values of the optimizer were set based on the previous step's solution to reduce the abrupt changes of the control decisions in each

time step. Additionally, the multithread feature (30 workers) of the optimization was enabled to speed up the calculation. Due to the computation time, once the solution converges, it is used for optimal control without further optimization iterations. Specifically, one optimization takes about 100 s, while the MPC sampling time is 900 s. Therefore, performing multiple optimizations would delay the control.

In this formulation, the wall temperature in Eq. (8) is the only hidden state, so a Kalman filter is used to estimate the initial state for each MPC control time for offset-free control [63]. The prediction horizon is set to 1 h due to computation time and the control problem characteristics. More specifically, this control problem does not utilize long-term dynamics (e.g., renewable resources, thermal storage, etc.), so there is no clear benefit of having a long-term prediction horizon except for the thermal energy charge in thermal mass. Finally, during the prediction

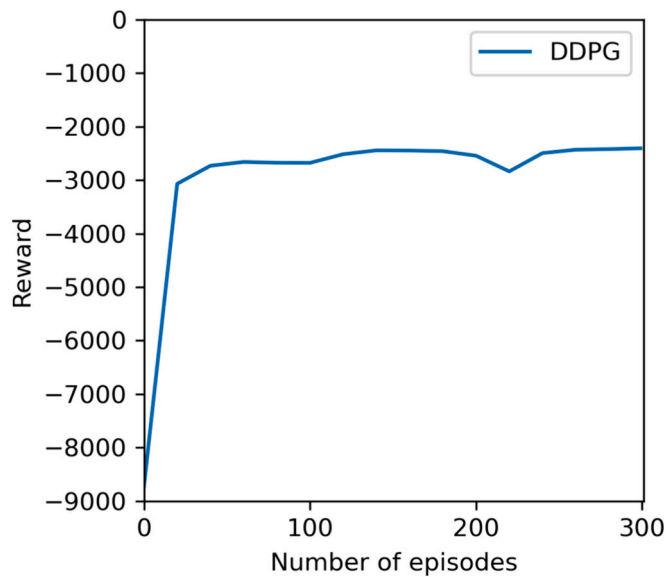


Fig. 16. Change of reward on the test set when DDPG is trained by different numbers of episodes.

period, it is assumed that the internal heat gain disturbances are not predicted (Fig. 7) as this information is typically unavailable in real-world situations. In the discussion section, we will present more details regarding the limitations and challenges of applying MPC for this type of control problem.

#### 4. Results and discussion

The designed DRL controller is trained by interacting with the hybrid simulation model to maximize the reward function, using the exogenous inputs in the training set. After that, its performance is evaluated on the same hybrid simulation model but using the exogenous inputs in the test set. Before discussing the simulation results, this section first displays the calibration error of the hybrid model. Then, the performance of the benchmarks and the DRL controllers are evaluated separately in Section 4.2 and Section 4.3, based on the same exogenous inputs in the test set. After that, Section 4.4 moves a step further towards analyzing the behavior of the DRL controller, helping mechanical engineers understand this black-box model from their domain expertise. Section 4.5 compares the energy consumption and indoor conditions of the office building after applying the DRL controller and the benchmarks. The effectiveness and advantages of the DRL controller are discussed. Lastly, Section 4.6 analyzed some practical considerations of applying MPC and DRL controllers in practice.

##### 4.1. Model calibration

This section shows some calibration results of the hybrid simulation model. Since the HVAC system model and the building envelope model are very important for the simulation, they were both calibrated using real experimental data. For the HVAC system model, Table 8 shows the root mean square error (RMSE) and mean absolute percentage error (MAPE) of several physical parameters in the chilled water loop and the AHU. Because the chiller of the test cell is largely oversized with very low COP, its power is not calibrated in this simulation. From the table, the simulated data are very close to the experimental data, with MAPE lower than 10 %.

Fig. 13 compares the real zone air temperature data (blue curve) with the simulation results of the building envelope model (orange curve) from July 29th to July 31st in 2023. The predicted temperature values follow closely with the experimental data, and the RMSE is 0.536 °C.

##### 4.2. Performance of the benchmark controllers

For comparison, Fig. 14 shows the performance of the rule-based controller in an office room from June 16th to June 30th using the weather data in the test set. Since the comfort and indoor air quality are concerned only during the occupied time, in the first three subplots we covered the unoccupied nighttime in gray color. From the top two subplots, the indoor temperature is well-controlled around the setpoint, which is set at 24 °C during the daytime. Also, the indoor relative humidity is below 60 %. Thus, the thermal comfort is satisfied in the occupied time. From the third subplot, even in the morning and the afternoon when all occupants work in the office, the CO<sub>2</sub> level can be maintained below 1000 ppm. In the fourth subplot, the indoor PM2.5 concentration is below 10 µg/m<sup>3</sup>. In summary, by using a rule-based controller, the thermal comfort and air quality can be well-maintained.

In Fig. 15, the operational characteristics of the MPC controller are presented for the same period and with the same plots as the RBC controller. The zone air temperature is maintained slightly higher than the zone air setpoint but remains near the comfort upper boundary (see Fig. 8). Since the MPC does not account for unmeasured disturbances such as internal heat gains in its prediction, as discussed in Section 3.4.2, its temperature prediction could be lower than the actual temperature during the high cooling demand period. As a result, the system does not provide enough cooling, resulting in a comfort violation.

Similarly, the prediction mismatch could explain the noisy control decisions in the chilled water setpoint and the damper position. Whenever there are non-negligible prediction mismatches, the offset-free MPC (with a Kalman filter) and the non-linear optimization may yield a completely new control trajectory, resulting in abrupt changes in the control decisions. The prediction mismatch can be alleviated by including an additional model for predicting the unmeasured disturbances [64]. Furthermore, some heuristics, such as penalizing abrupt decision changes, can be used but are not included in this research, as they require the control engineer's tuning, which is not a practical control solution. Nevertheless, the MPC controller can achieve the humidity, indoor air quality, and CO<sub>2</sub> requirements with slight comfort violations.

##### 4.3. Performance of the DRL controller

The designed DRL controller is trained for 300 episodes, which takes about 2.5 days on a workstation with CPU and GPU cores. Fig. 16 displays the reward of the DDPG algorithm computed on the test set with an increasing number of training episodes. DDPG is computationally efficient and can find the optimal strategy for the complex control problem with only around 50 episodes. In practical applications, training the DDPG controller up to 100 episodes in the virtual environment is sufficient before deploying it into the real-world environment.

Then, to validate the generalizability of the DDPG controller, its performance at the same period with the benchmarking controllers is shown in Fig. 17. First, the room temperature is maintained very well by the DRL controller, very smooth and close to the setpoint during the daytime. This is because the DRL algorithm tries to identify the optimal chilled water supply temperature and the optimal supply fan speed that can send an appropriate amount of cooling into the room. Eventually, the chiller operates at a lower part load ratio and has fewer ON/OFF cycles, which is beneficial to the chiller and increases its life expectancy. Second, the indoor relative humidity is basically under 60 %, no matter how the outdoor humidity level changes. Third, the indoor CO<sub>2</sub> and PM2.5 levels are maintained below their respective thresholds. The DRL controller sends commands to switch on the air purifier only when it is necessary, i.e., the outdoor PM2.5 level is high but ventilation is required to dilute the indoor CO<sub>2</sub> concentration.

The last two subplots in Fig. 17 show the outdoor air damper opening and the supply air fan speed. They are mainly adjusted based on the time

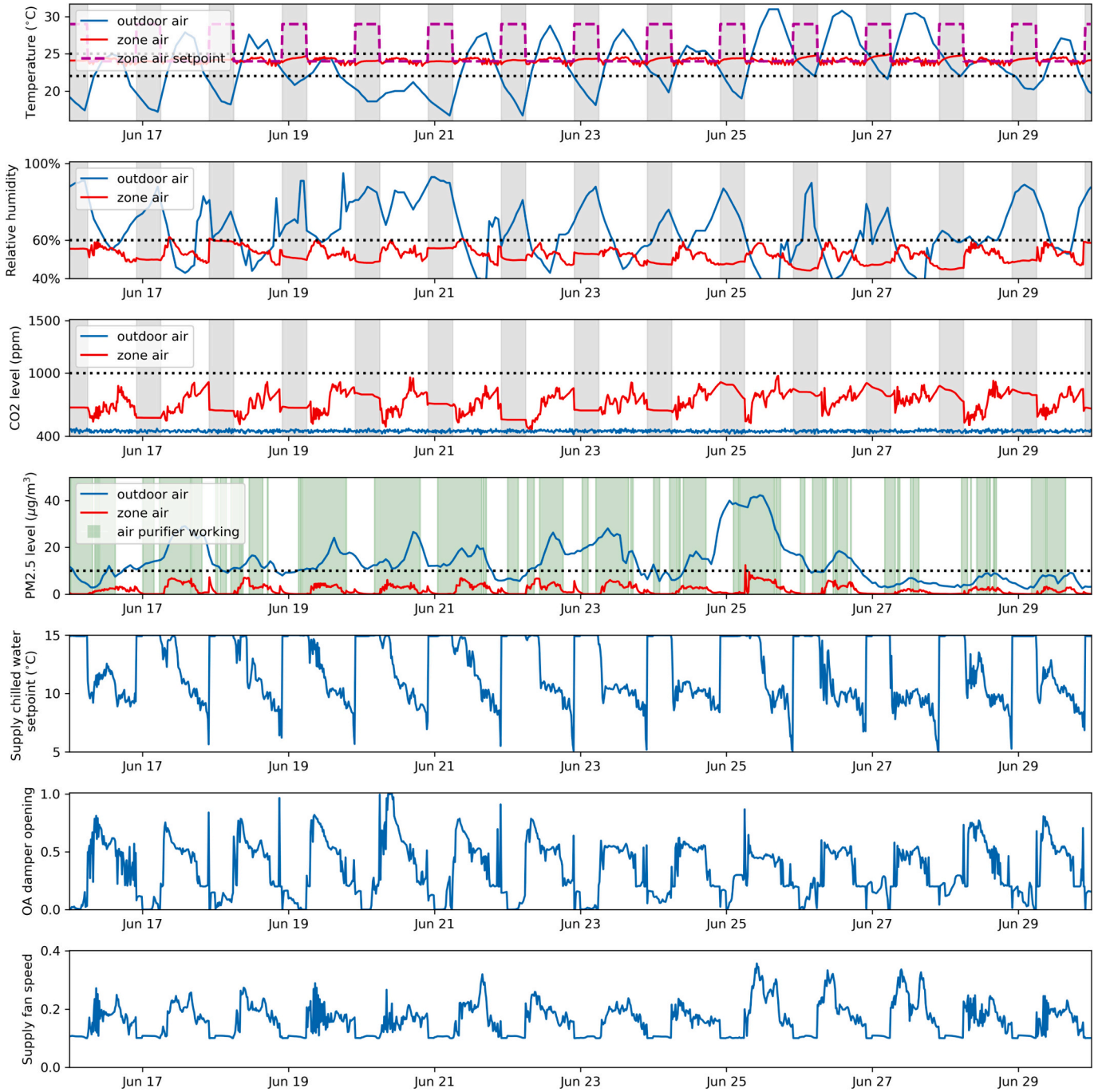


Fig. 17. Performance of the DRL controller on the test set from June 16th to June 30th.

of day, a state variable of the DRL controller. In the morning and the afternoon, the office is full of people and the CO<sub>2</sub> generation rate is high. Hence, the DRL controller learns to increase the damper opening and the fan speed. Conversely, in the early morning, at noon, and in the evening, the occupancy level in the office is low (see Fig. 7), so the controller tries to decrease the damper opening and the fan speed a little bit. But these two action variables are also heavily affected by the outdoor weather conditions, which will be discussed in the next section.

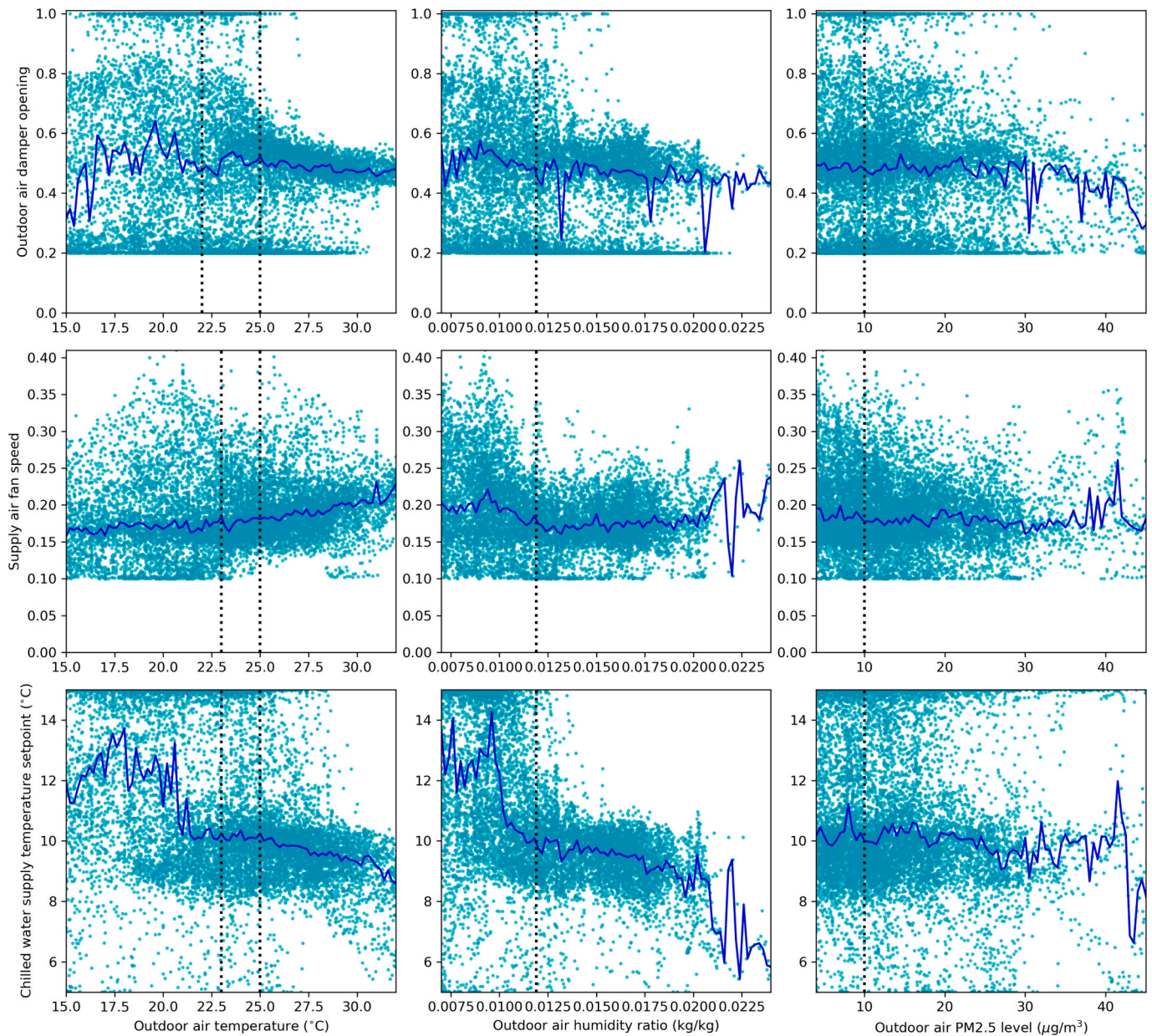
#### 4.4. Analysis of the DRL controller

This section goes one step further to analyze how the DRL controller makes decisions based on given state variables. Because indoor conditions generally have small variations within the comfort and healthy

zone, the action variables are mainly correlated with the outdoor conditions, including the temperature, humidity, and PM<sub>2.5</sub> levels. The outdoor CO<sub>2</sub> level is not considered as an impact factor since it is nearly a constant and the indoor CO<sub>2</sub> generation rate is a more important factor for the damper opening, as discussed in the above section.

After the DRL controller is evaluated on the test set, this study extracts all of the state-action pairs in the daytime and then plots their relationships in Fig. 18. In those plots, the horizontal axes are state variables and the vertical axes are action variables. Each data point is a state-action pair and the blue curves show the median values of an action variable in each bin of the state variable. Since the air purifier ON/OFF control is mostly associated with the outdoor PM<sub>2.5</sub> level, their relationship is then plotted in Fig. 19 with the vertical axis showing the proportion of time that the air purifier is in operation. A few findings





**Fig. 18.** The actions of the DRL controller in response to the states. Each data point is a state-action pair in the occupied time and the curves are the median value in each bin of the states.

from the plots should be addressed:

- *The economizer mode is enabled when the outdoor temperature and the outdoor humidity ratio are both low enough.* Indoor thermal comfort is considered to be satisfied when the temperature is between 22 °C and 25 °C and the relative humidity is below 60 %. Hence, in the reward function, a large penalty is imposed if indoor temperature rises above 24.5 °C or relative humidity rises above 60 %. At this threshold, the humidity ratio is about 0.0115 kg/kg at atmospheric pressure. Therefore, the DRL controller has learned that only when the outdoor temperature and humidity ratio are below these thresholds, the economizer mode is beneficial and can save some energy from mechanical cooling.
- *When the outdoor temperature or humidity ratio is high, the outdoor air damper tries to open as small as possible, just bringing in enough air to maintain an acceptable indoor CO<sub>2</sub> level.* In this scenario, the indoor

CO<sub>2</sub> generation by occupants is the main driving factor for outdoor air damper openings.

- *When the outdoor temperature is low, the supply air fan operates at relatively low speeds to save the fan power.* In this scenario, the zone cooling load and the ventilation cooling load are both very low, so a lower fan speed can satisfy the cooling demand.
- *When the outdoor temperature is high, the supply air fan has to increase its speed to process more cool air.* In this scenario, the zone load is very high and a large amount of supply air is needed to meet the cooling demand in the office room. Also, if the supply air flow rate is low, then the chilled water temperature in the cooling coil has to further decline in order to have enough heat exchange, which would significantly increase the chiller power.
- *The outdoor temperature and humidity ratio have a large influence on the chilled water supply temperature setpoint.* When the outside is hot or humid, the chilled water temperature in the cooling coil has to be reduced to enhance the dehumidification performance. In other

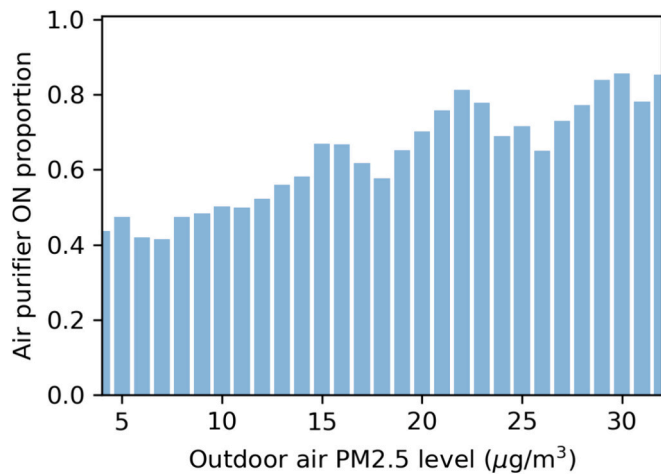


Fig. 19. Relationship between outdoor air PM2.5 level and the proportion of time that the air purifier is turned on.

words, regulation of the chilled water temperature is crucial for maintaining thermal comfort and achieving better humidity control.

- The outdoor air PM2.5 level does not have a large influence on the damper, fan speed, and chilled water temperature. This is because the power of the air purifier is only 42 W, much lower than the chiller and the fan. When the outside is not heavily polluted, switching on the air purifier is a better choice. Since this study simulates the cooling season when severe air pollution rarely occurs, the air purifier itself is capable of maintaining good IAQ. Severe air pollution

usually occurs in winter due to coal burning for central heating, and in that scenario, the outdoor air PM2.5 level might have a larger impact on the control of the AHU.

- The proportion of time that the air purifier is ON is positively correlated with the outdoor PM2.5 level. When the outdoor air has higher pollution, the air purifier should be switched on more frequently to maintain good IAQ.

#### 4.5. Performance comparison between RBC, MPC, and DRL controllers

Fig. 20 compares the indoor temperature, relative humidity, CO<sub>2</sub> level, and PM2.5 level during the daytime between RBC, MPC, and DRL controllers. The following should be noted:

- From Fig. 20(a), the distributions of indoor temperature with RBC and DRL controls are both around 24 °C and rarely rise above 25 °C. The indoor temperature controlled by MPC is a little bit higher because unmeasured disturbances such as internal heat gains are not considered in the models of MPC, causing its predicted cooling load lower than the actual cooling load, as discussed in Section 4.2. Note that the indoor temperature setpoint is fixed at 24 °C for all controllers in the occupied time.
- From Fig. 20(b), the DRL has better control of humidity. Taking the relative humidity at 60 % as the threshold, overshoots of relative humidity happen 10.8 % of the time for the RBC controller and 18.3 % for the MPC controller, but this number is only 1.2 % for the DRL controller.
- From Fig. 20(c), both RBC and DRL controllers are able to maintain the indoor CO<sub>2</sub> level below 1000 ppm by adjusting the outdoor air damper position according to the occupancy schedules in the office.

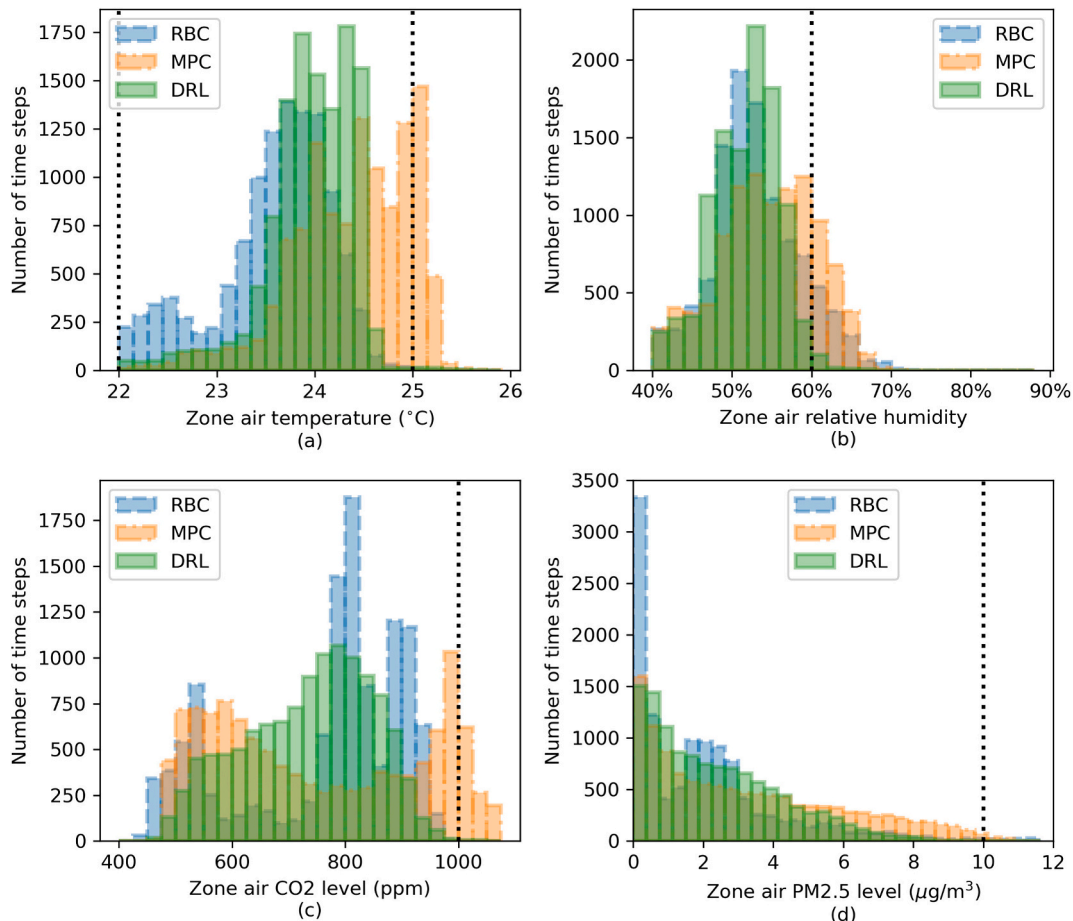


Fig. 20. A comparison of indoor conditions during the daytime between RBC, MPC, and DRL.



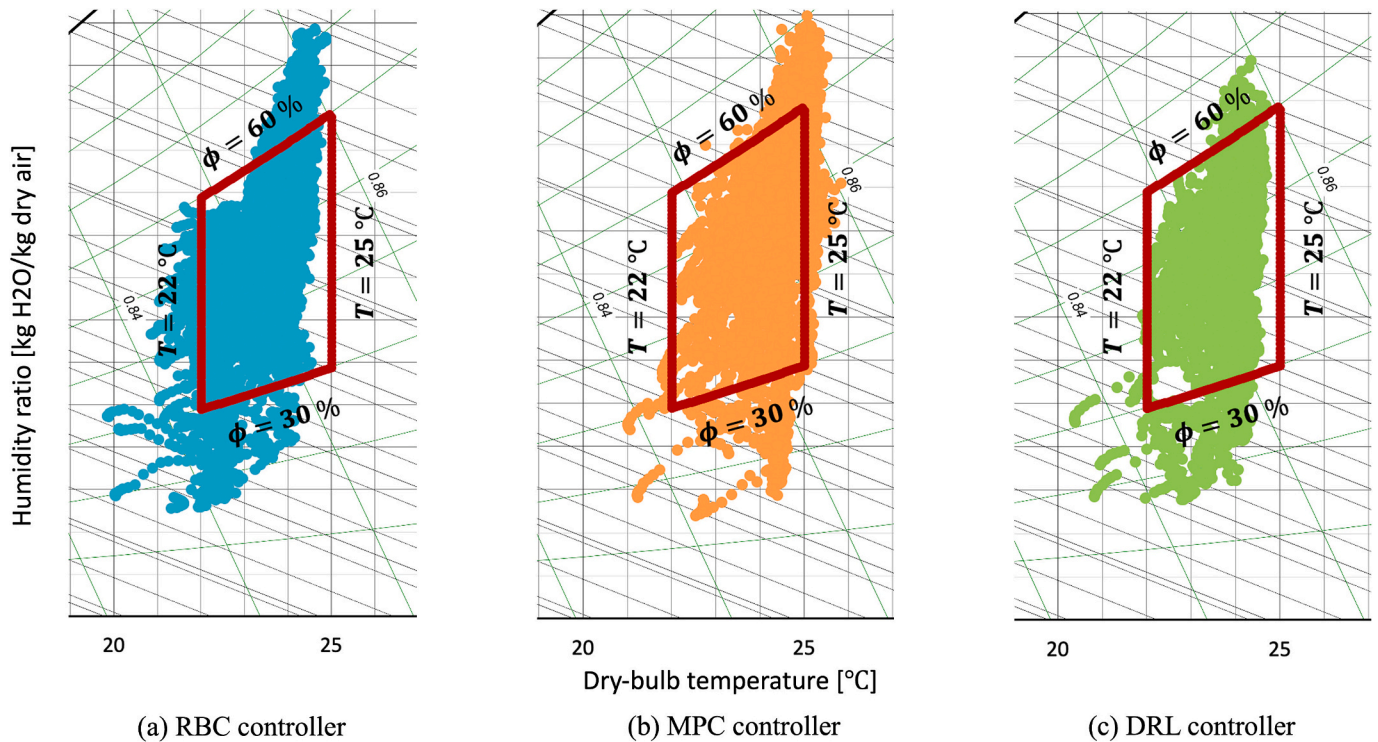


Fig. 21. A comparison of indoor conditions during the daytime between RBC, MPC, and DRL on the psychrometric chart.

With the MPC controller, the indoor CO<sub>2</sub> level may occasionally rise above 1000 ppm, but overall, the level is still acceptable.

- From Fig. 20(d), the MPC and DRL controllers have fewer overshoots of the PM<sub>2.5</sub> level compared to the RBC controller. Taking 10 μg/m<sup>3</sup> as a threshold, overshoots happen 2.5 % of the time for the RBC controller, but 0.7 % of the time for the MPC controller and less than 0.1 % of the time for the DRL controller. DRL has learned to switch on the air purifier when necessary.

Fig. 21 compares the thermal comfort of the office room using these three controllers on a psychrometric chart. For the RBC controller, 82.5 % of data are within the defined thermal comfort zone; for the MPC controller, 69.6 % of data are within the defined thermal comfort zone. In comparison, this number is 95.2 % for the DRL controller. The DRL controller has better temperature and humidity control, with a larger fraction of data in the comfort zone. Occupants will feel more comfortable when the HVAC system is controlled by the DRL algorithm.

Fig. 22 shows the daily energy consumption by the chiller, the chilled water pump, the supply air fan, and the air purifier from May to October, and Table 9 shows the statistics. Compared with the RBC controller, employing the DRL controller saves 21.4 % of energy in this half-year period. The majority of energy consumption comes from the chiller and the fan. Because the DRL controller has learned to increase the chilled water temperature and decrease the fan speed when it is possible, the chiller can operate at a lower part load ratio and consume less energy. Although consequently, the chiller has a longer ON time, which increases the energy consumption of the constant-speed chilled water pump, the energy saved by the chiller is still a lot more than the extra energy consumed by the pump. Compared with the MPC controller, the DRL controller consumes 18.6 % more energy. However, the lower energy consumption of the MPC controller is penalized by the increased occupant discomfort and poorer indoor air quality. Some practical considerations of MPC and DRL controllers are discussed in the next section.

To summarize, the energy savings by adopting the DRL controller come from the following aspects:

- 1) When the cooling demand of the office room is moderate, the chiller operates at a lower part load ratio, the chilled water temperature is higher, and the supply air fan operates at a lower speed. This increases the efficiency of the entire HVAC system. We found that this strategy contributes to most of the energy savings.
- 2) When the outdoor condition is favorable (i.e., cool and less humid), the economizer mode is enabled. This allows for free cooling using the outdoor air.
- 3) In the early morning and the evening when the occupancy level in the office is low, the opening of the outdoor air damper is reduced. This decreases the ventilation cooling load.

#### 4.6. Considerations between MPC and DRL for control implementation from a practical perspective

We propose a DRL controller to co-optimize energy consumption, thermal comfort, CO<sub>2</sub> concentration, and IAQ for the entire HVAC system. While performance benefits are important, it is also crucial to understand the practical considerations of DRL and MPC controllers for this non-linear co-optimization problem. We list these considerations below as references for controller designers, as there is no clear advantage of one over the other.

##### 4.6.1. Computational effort in real-time controller

Although the DRL controller usually requires a high-fidelity hybrid simulator for training, once it is well-trained, it can efficiently provide optimal decisions. On the other hand, the MPC controller needs to run optimization at every control time step, which typically requires significant computation time, making it sometimes unsuitable for real building controllers, especially if the decision space is large and involves non-linear optimization. Considering typical control time steps of 15–60 min (or less), optimization that requires more than 5 min of computational time, including all data processing and state estimation, may not be practically useful. Additionally, reducing the prediction horizon to save computation time may degrade the performance of the MPC controller [65]. Some studies [66,67] have shown efficient non-linear

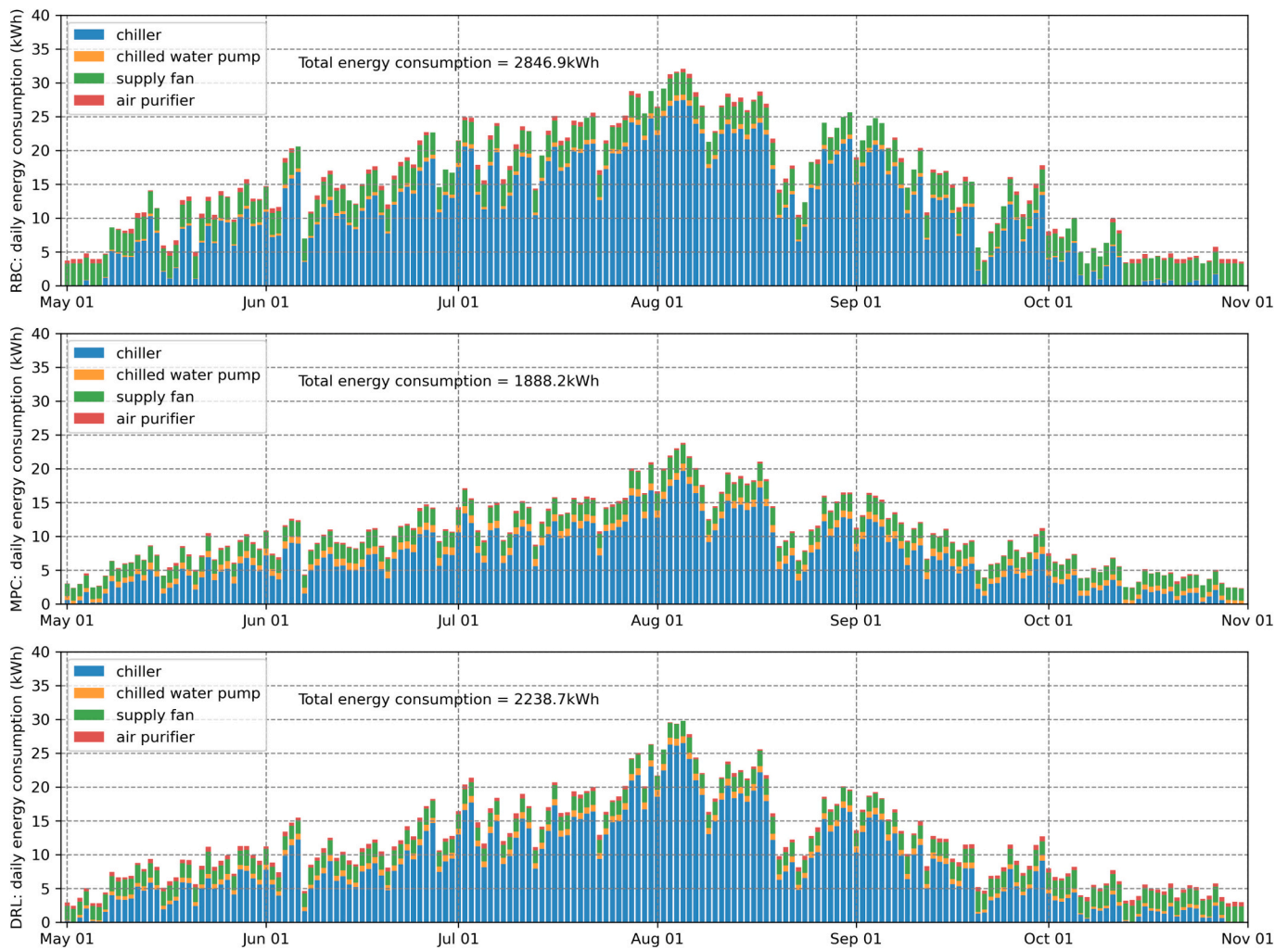


Fig. 22. A comparison of energy consumption between RBC, MPC, and DRL.

Table 9

Comparison of energy consumption by component between RBC, MPC and DRL.

|                    | RBC        | MPC        | DRL        | Change from RBC to DRL | Change from MPC to DRL |
|--------------------|------------|------------|------------|------------------------|------------------------|
| Chiller            | 2096.4 kWh | 1269.0 kWh | 1613.5 kWh | −23.0 %                | +27.1 %                |
| Chilled water pump | 70.6 kWh   | 172.5 kWh  | 131.8 kWh  | +86.7 %                | −23.6 %                |
| Supply fan         | 605.1 kWh  | 403.7 kWh  | 403.7 kWh  | −33.3 %                | 0                      |
| Air purifier       | 74.8 kWh   | 42.9 kWh   | 89.6 kWh   | +19.8 %                | 109.1 %                |
| Total              | 2846.9 kWh | 1888.2 kWh | 2238.7 kWh | −21.4 %                | +18.6 %                |

MPC, but these are still at a research level, requiring advanced control knowledge, toolboxes, and tuning techniques.

#### 4.6.2. Modeling effort

Traditionally, the main advantage of the DRL controller is its ‘model-free’ structure, as having a well-calibrated model for control is a major bottleneck in building control applications. Typically, offline training of DRL requires a high granularity and sophisticated model as an environment, such as EnergyPlus or Modelica, but the choice of model and the efforts of the calibration depends on the complexity of the building. Similarly, the model for an MPC controller depends on the building’s complexity as well. For small and medium-sized buildings, a control-oriented, gray-box model can be easily established. However, for large buildings, the model structure becomes complex and requires significant effort from the modeler if there are several unique building

characteristics and local controllers, such as multi-zone VAV systems with different terminal units, natural ventilation, etc.

#### 4.6.3. Control performance

A common critique of the DRL controller is that its decisions can be inconsistent, non-physical, and non-interpretable. Since it provides an optimal decision for each control time step, the control decisions can change abruptly, resulting in unstable control performance. However, as discussed in Section 4.2, the MPC controller’s performance can also be unstable due to prediction mismatches when there are significant unmeasured disturbances, especially in non-linear optimization. Both instabilities can be mitigated by using heuristics, such as limiting the magnitude of changes in control actions for each time step.

Unmeasured disturbances can degrade the performance of both controllers. The DRL controller could incorporate unmeasured

disturbances in its neural network since it is purely data-driven. In contrast, the MPC controller needs a separate model for unmeasured disturbances, as discussed in Section 4.2. However, significant changes in the dynamics of unmeasured disturbances require adjustments for both controllers. Both controllers need to recalibrate building, HVAC, and unmeasured disturbance models. Although not guaranteed, the online learning feature of the DRL controller may adapt to changes in the dynamics of unmeasured disturbances in real time, which is a significant research topic and beyond the scope of this study.

#### 4.6.4. Price signal flexibility

The MPC controller can be advantageous in some cases. With utility and grid policies frequently changing, price signals for the reward function can vary. While the MPC controller can compute the optimal decision by simply updating the price signals in the objective function, the DRL controller needs to be retrained using a new reward equation with the updated price signal.

## 5. Conclusion

This paper applied deep deterministic policy gradient (DDPG), a DRL algorithm, to co-optimize energy consumption, thermal comfort, CO<sub>2</sub> concentration, and IAQ of an office building. One advantage of using the DDPG algorithm is that it can generate continuous actions for the actuators and is also applicable to problems with relatively large action spaces. In this study, the DRL controller sends commands to control the chilled water temperature, the damper, the fan, and the air purifier directly. The DRL algorithm is trained and tested by using the weather data in Seoul and interacting with a hybrid simulation model. The results showed that, compared to a rule-based controller, the DRL controller can achieve 21.4 % energy savings while maintaining a proper temperature, humidity, and CO<sub>2</sub> and PM2.5 concentration in the office. In other words, a well-trained DRL is able to balance different optimization objectives in the control of building HVAC systems. The DRL controller is also compared to an MPC controller, and practical considerations of the two types of controllers are discussed. Moreover, the hybrid simulation model developed in this research represents the dynamics of the HVAC system, heat transfer, moisture balance, CO<sub>2</sub> dissipation, and particulate transmission in the office room. In the future, the model could be used to evaluate the performance of other artificial intelligence-based control algorithms through building optimization tests. For example, the performance of different DRL algorithms such as PPO, DDPG, and SAC could be compared together to determine the algorithms that are more suitable in the application of building energy systems.

There are also a few limitations of the proposed DRL controller. First, since the controller makes deterministic decisions based on the time of day, it is not able to quickly adapt to the changes in equipment, occupant, and electricity price schedules. If one of those schedules is changed, the DRL algorithm must be retrained. To tackle this problem in future work, the predictions of future schedules could also be considered as inputs to train the DRL algorithm. Second, the performance of the DRL algorithm is very sensitive to the shaping of the reward function, especially in co-optimization problems where several objectives need to be achieved simultaneously. This study determines the weights of each control objective in the reward function by parametric analysis and grid search. Future work can focus on designing more efficient methods to balance the penalties of energy, cost, comfort, and air quality. What's more, in the future, we also hope to add heating devices into the hybrid model and simulate the whole year's performance. Since the air pollution in winter is more severe in Northeast Asia, controlling the concentration of indoor particulate matter meanwhile meeting other control objectives is expected to be more challenging.

## CRedit authorship contribution statement

**Fangzhou Guo:** Writing – original draft, Validation, Methodology, Formal analysis, Conceptualization. **Sang woo Ham:** Writing – review & editing, Formal analysis, Conceptualization. **Donghun Kim:** Project administration, Funding acquisition, Conceptualization. **Hyeun Jun Moon:** Supervision, Resources, Funding acquisition.

## 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.

## Data availability

The authors do not have permission to share data.

## Acknowledgments

This work was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Building Technologies Office, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231, and by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. 20212020800120).

## References

- [1] An assessment of energy technologies and research opportunities. US Department of Energy; 2015.
- [2] Ma Z, Wang S. Supervisory and optimal control of central chiller plants using simplified adaptive models and genetic algorithm. *Appl Energy* 2011;88:198–211. <https://doi.org/10.1016/j.apenergy.2010.07.036>.
- [3] Taheri S, Hosseini P, Razban A. Model predictive control of heating, ventilation, and air conditioning (HVAC) systems: a state-of-the-art review. *J Build Eng* 2022; 60:105067. <https://doi.org/10.1016/j.jobe.2022.105067>.
- [4] Gholamzadehmir M, Del Pero C, Buffa S, Fedrizzi R, Aste N. Adaptive-predictive control strategy for HVAC systems in smart buildings – a review. *Sustain Cities Soc* 2020;63:102480. <https://doi.org/10.1016/j.scs.2020.102480>.
- [5] Dragoña J, Arroyo J, Cupeiro Figueroa I, Blum D, Arendt K, Kim D, et al. All you need to know about model predictive control for buildings. *Annual Rev Control* 2020;50:190–232. <https://doi.org/10.1016/j.arcontrol.2020.09.001>.
- [6] Kim D, Braun JE. Development, implementation and performance of a model predictive controller for packaged air conditioners in small and medium-sized commercial building applications. *Energy Buildings* 2018;178:49–60. <https://doi.org/10.1016/j.enbuild.2018.08.019>.
- [7] Ham SW, Kim D, Barham T, Ramseyer K. The first field application of a low-cost MPC for grid-interactive K-12 schools: lessons-learned and savings assessment. *Energy Buildings* 2023;296:113351. <https://doi.org/10.1016/j.enbuild.2023.113351>.
- [8] Zhang H, Seal S, Wu D, Bouffard F, Boulet B. Building energy management with reinforcement learning and model predictive control: a survey. *IEEE Access* 2022; 10:27853–62. <https://doi.org/10.1109/ACCESS.2022.3156581>.
- [9] Zanetti E, Kim D, Blum D, Scoccia R, Aprile M. Performance comparison of quadratic, nonlinear, and mixed integer nonlinear MPC formulations and solvers on an air source heat pump hydronic floor heating system. *J Build Performance Simulation* 2023;16:144–62. <https://doi.org/10.1080/19401493.2022.2120631>.
- [10] Wang Z, Hong T. Reinforcement learning for building controls: the opportunities and challenges. *Appl Energy* 2020;269:115036. <https://doi.org/10.1016/j.apenergy.2020.115036>.
- [11] Han M, May R, Zhang X, Wang X, Pan S, Yan D, et al. A review of reinforcement learning methodologies for controlling occupant comfort in buildings. *Sustain Cities Soc* 2019;51:101748. <https://doi.org/10.1016/j.scs.2019.101748>.
- [12] Han M, Zhang X, Xu L, May R, Pan S, Wu J. A review of reinforcement learning methodologies on control systems for building energy. *Höskolan Dalarna* 2018.
- [13] Vázquez-Canteli JR, Nagy Z. Reinforcement learning for demand response: a review of algorithms and modeling techniques. *Appl Energy* 2019;235:1072–89. <https://doi.org/10.1016/j.apenergy.2018.11.002>.
- [14] Coraci D, Brandi S, Hong T, Capozzoli A. Online transfer learning strategy for enhancing the scalability and deployment of deep reinforcement learning control in smart buildings. *Appl Energy* 2023;333:120598. <https://doi.org/10.1016/j.apenergy.2022.120598>.
- [15] Coraci D, Brandi S, Hong T, Capozzoli A. An innovative heterogeneous transfer learning framework to enhance the scalability of deep reinforcement learning controllers in buildings with integrated energy systems. *Build Simul* 2024;17: 739–70. <https://doi.org/10.1007/s12273-024-1109-6>.



- [16] Nagy Z, Henze G, Dey S, Arroyo J, Helsen L, Zhang X, et al. Ten questions concerning reinforcement learning for building energy management. *Build Environ* 2023;110435. <https://doi.org/10.1016/j.buildenv.2023.110435>.
- [17] Ahn KU, Park CS. Application of deep Q-networks for model-free optimal control balancing between different HVAC systems. *Sci Technol Built Environ* 2020;26: 61–74. <https://doi.org/10.1080/23744731.2019.1680234>.
- [18] An Y, Xia T, You R, Lai D, Liu J, Chen C. A reinforcement learning approach for control of window behavior to reduce indoor PM2.5 concentrations in naturally ventilated buildings. *Build Environ* 2021;200:107978. <https://doi.org/10.1016/j.buildenv.2021.107978>.
- [19] An Y, Niu Z, Chen C. Smart control of window and air cleaner for mitigating indoor PM2.5 with reduced energy consumption based on deep reinforcement learning. *Build Environ* 2022;224:109583. <https://doi.org/10.1016/j.buildenv.2022.109583>.
- [20] An Y, Chen C. Energy-efficient control of indoor PM2.5 and thermal comfort in a real room using deep reinforcement learning. *Energy Buildings* 2023;295:113340. <https://doi.org/10.1016/j.enbuild.2023.113340>.
- [21] Biemann M, Scheller F, Liu X, Huang L. Experimental evaluation of model-free reinforcement learning algorithms for continuous HVAC control. *Appl Energy* 2021;298:117164. <https://doi.org/10.1016/j.apenergy.2021.117164>.
- [22] Blad C, Bøgh S, Kallesøe CS. Data-driven offline reinforcement learning for HVAC-systems. *Energy* 2022;261:125290. <https://doi.org/10.1016/j.energy.2022.125290>.
- [23] Deng X, Zhang Y, Zhang Y, Qi H. Towards optimal HVAC control in non-stationary building environments combining active change detection and deep reinforcement learning. *Build Environ* 2022;211:108680. <https://doi.org/10.1016/j.buildenv.2021.108680>.
- [24] Du Y, Zandi H, Kotevska O, Kurte K, Munk J, Amasyali K, et al. Intelligent multi-zone residential HVAC control strategy based on deep reinforcement learning. *Appl Energy* 2021;281:116117. <https://doi.org/10.1016/j.apenergy.2020.116117>.
- [25] Du Y, Li F, Munk J, Kurte K, Kotevska O, Amasyali K, et al. Multi-task deep reinforcement learning for intelligent multi-zone residential HVAC control. *Electr Pow Syst Res* 2021;192:106959. <https://doi.org/10.1016/j.epr.2020.106959>.
- [26] Gao G, Li J, Wen Y. DeepComfort: energy-efficient thermal comfort control in buildings via reinforcement learning. *IEEE Internet Things J* 2020;7:8472–84. <https://doi.org/10.1109/JIOT.2020.2992117>.
- [27] Lu X, Fu Y, O'Neill Z. Benchmarking high performance HVAC rule-based controls with advanced intelligent controllers: a case study in a multi-zone system in Modelica. *Energy Buildings* 2023;284:112854. <https://doi.org/10.1016/j.enbuild.2023.112854>.
- [28] Shang W, Liu J, Wang C, Li J, Dai X. Developing smart air purifier control strategies for better IAQ and energy efficiency using reinforcement learning. *Build Environ* 2023;242:110556. <https://doi.org/10.1016/j.buildenv.2023.110556>.
- [29] Shen R, Zhong S, Wen X, An Q, Zheng R, Li Y, et al. Multi-agent deep reinforcement learning optimization framework for building energy system with renewable energy. *Appl Energy* 2022;312:118724. <https://doi.org/10.1016/j.apenergy.2022.118724>.
- [30] Touzani S, Prakash AK, Wang Z, Agarwal S, Pritoni M, Kiran M, et al. Controlling distributed energy resources via deep reinforcement learning for load flexibility and energy efficiency. *Appl Energy* 2021;304:117733. <https://doi.org/10.1016/j.apenergy.2021.117733>.
- [31] Xia S, Wei P, Liu Y, Sonta A, Jiang X. RECA: A Multi-Task Deep Reinforcement Learning-Based Recommender System for Co-Optimizing Energy, Comfort and Air Quality in Commercial Buildings. In: *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, Istanbul Turkey: ACM; 2023. p. 99–109. <https://doi.org/10.1145/3600100.3623735>.
- [32] Yang T, Zhao L, Li W, Wu J, Zomaya AY. Towards healthy and cost-effective indoor environment management in smart homes: a deep reinforcement learning approach. *Appl Energy* 2021;300:117335. <https://doi.org/10.1016/j.apenergy.2021.117335>.
- [33] Yoon YR, Moon HJ. Performance based thermal comfort control (PTCC) using deep reinforcement learning for space cooling. *Energy Buildings* 2019;203:109420. <https://doi.org/10.1016/j.enbuild.2019.109420>.
- [34] Zhang Y, Zhao Q. Energy Saving Algorithm of HVAC System Based on Deep Reinforcement Learning with Modelica Model. In: *2022 41st Chinese control conference (CCC)*. IEEE; 2022. p. 5277–82.
- [35] Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, et al. Playing atari with deep reinforcement learning. *arXiv Preprint arXiv:1312.5602*. 2013.
- [36] Mnih V, Badia AP, Mirza M, Graves A, Harley T, Lillicrap TP, et al. Asynchronous methods for deep reinforcement learning n.d.
- [37] Lillicrap TP, Hunt JJ, Pritzel A, Heess N, Erez T, Tassa Y, et al. *Continuous control with deep reinforcement learning*. 2019.
- [38] Chollet F. Others. Keras. 2015.
- [39] Wetter M, Zuo W, Nouidui TS, Pang X. Modelica buildings library. *J Build Performance Simulation* 2014;7:253–70. <https://doi.org/10.1080/19401493.2013.765506>.
- [40] Nouidui T, Wetter M, Zuo W. Functional mock-up unit for co-simulation import in EnergyPlus. *J Build Performance Simulation* 2014;7:192–202.
- [41] Andersson C, Åkesson J, Führer C. Pyfmi: A python package for simulation of coupled dynamic models with the functional mock-up interface. 2016.
- [42] Deru M, Field K, Studer D, Benne K, Griffith B, Torcellini P, et al. U.S. Department of Energy Commercial Reference Building Models of the National Building Stock. 2011. <https://doi.org/10.2172/1009264>.
- [43] American Society of Heating Refrigerating and Air-Conditioning Engineers. *Ashrae handbook : fundamentals (SI)*. ASHRAE 2001;2001.
- [44] Standard A, others. Standard guide for using indoor carbon dioxide concentrations to evaluate indoor air quality and ventilation. ASTM Standard D6245–12, Standard guide for using indoor carbon dioxide concentrations to evaluate indoor air quality and ventilation. Conshohocken, PA: ASTM International West; 2012.
- [45] El Diasty R, Fazio P, Budaiwi I. Modelling of indoor air humidity: the dynamic behaviour within an enclosure. *Energy Buildings* 1992;19:61–73. [https://doi.org/10.1016/0378-7788\(92\)90036-G](https://doi.org/10.1016/0378-7788(92)90036-G).
- [46] Nazaroff WW. Indoor particle dynamics. *Indoor Air* 2004;14:175–83. <https://doi.org/10.1111/j.1600-0668.2004.00286.x>.
- [47] ZHAO D, Azimi P, Stephens B. Evaluating the Long-Term Health and Economic Impacts of Central Residential Air Filtration for Reducing Premature Mortality Associated with Indoor Fine Particulate Matter (PM 2.5) of Outdoor Origin. *Int J Environ Res Public Health* 2015;12:8448–79. doi:<https://doi.org/10.3390/ijerph120708448>.
- [48] Taylor ST. Optimizing design & control of chilled water plants: part 5: optimized control sequences. *ASHRAE J* 2012;54:56–69.
- [49] Kapsalaki M. ASHRAE position document on indoor carbon dioxide. 2022.
- [50] Abraczinskas M, Meyers J. Matter National Ambient Air Quality Standard: Sloan J; 2023.
- [51] Fanger PO. Thermal comfort. Analysis and applications in environmental engineering. 1970. p. 244.
- [52] Enescu D. A review of thermal comfort models and indicators for indoor environments. *Renew Sustain Energy Rev* 2017;79:1353–79. <https://doi.org/10.1016/j.rser.2017.05.175>.
- [53] ASHRAE. Standard 55–2023; thermal environmental conditions for human occupancy. 2023.
- [54] Standardization IO for ISO 7730. Moderate thermal environments-determination of the PMV and PPD indices and specification of the conditions for thermal comfort. ISO; 1994.
- [55] EN 15251 indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality, thermal environment, lighting and acoustics 2007.
- [56] Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv Preprint arXiv:1412.6980* 2014.
- [57] Gratani L, Varone L. Daily and seasonal variation of CO2 in the city of Rome in relationship with the traffic volume. *Atmos Environ* 2005;39:2619–24. <https://doi.org/10.1016/j.atmosenv.2005.01.013>.
- [58] Xueref-Remy I, Dieudonné E, Vuillemin C, Lopez M, Lac C, Schmidt M, et al. Diurnal, synoptic and seasonal variability of atmospheric CO2 in the Paris megacity area. *Atmos Chem Phys* 2018;18:3335–62. <https://doi.org/10.5194/acp-18-3335-2018>.
- [59] Vogt R, Christen A, Rotach MW, Roth M, Satyanarayana ANV. Temporal dynamics of CO2 fluxes and profiles over a central European city. *Theor Appl Climatol* 2006; 84:117–26. <https://doi.org/10.1007/s00704-005-0149-9>.
- [60] EnergyPlusTM version 23.2.0 documentation - Input Output Reference. 2023.
- [61] Hydeman M, Gillespie Jr KL. Tools and techniques to calibrate electric chiller component models/discussion. *ASHRAE Trans* 2002;108:733.
- [62] Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020;17:261–72. <https://doi.org/10.1038/s41592-019-0686-2>.
- [63] Pannocchia G, Rawlings JB. Disturbance models for offset-free model-predictive control. *AIChE J* 2003;49:426–37. <https://doi.org/10.1002/aic.690490213>.
- [64] Ham SW, Kim D. Hybrid modeling approach for better identification of building thermal network model and improved prediction. 2022.
- [65] Arroyo J, Spiessens F, Helsen L. Comparison of optimal control techniques for building energy management. *Front Built Environ* 2022;8:849754. <https://doi.org/10.3389/fbuil.2022.849754>.
- [66] Michael H Blumand, Wetter D. MPCPy: an Open-Source Software Platform for Model Predictive Control in Buildings. 2017. <https://doi.org/10.26868/25222708.2017.351>.
- [67] Jansen J, Jorissen F, Helsen L. Mixed-integer non-linear model predictive control of district heating networks. *Appl Energy* 2024;361:122874. <https://doi.org/10.1016/j.apenergy.2024.122874>.