Deep Reinforcement Learning-Based Joint Optimization Control of Indoor Temperature and Relative Humidity in Office Buildings

Changcheng Chen [1], Jingjing An [1,*], Chuang Wang [1], Xiaorong Duan [1], Shiyu Lu [2], Hangyu Che [2], Meiwei Qi [3] and Da Yan [4,5,*]

1   School of Environment and Energy Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China
2   Hitachi (China), Ltd., Beijing 100190, China
3   Beijing Tongheng Energy & Environment Technology Institute, Beijing 100085, China
4   Building Energy Research Center, School of Architecture, Tsinghua University, Ministry of Education, Beijing 100084, China
5   Key Laboratory of Eco Planning & Green Building, Tsinghua University, Ministry of Education, Beijing 100084, China
*   Correspondence: anjingjing@bucea.edu.cn (J.A.); yanda@tsinghua.edu.cn (D.Y.)

Abstract: Indoor temperature and relative humidity control in office buildings is crucial, which can affect thermal comfort, work efficiency, and even health of the occupants. In China, fan coil units (FCUs) are widely used as air-conditioning equipment in office buildings. Currently, conventional FCU control methods often ignore the impact of indoor relative humidity on building occupants by focusing only on indoor temperature as a single control object. This study used FCUs with a fresh-air system in an office building in Beijing as the research object and proposed a deep reinforcement learning (RL) control algorithm to adjust the air supply volume for the FCUs. To improve the joint control satisfaction rate of indoor temperature and relative humidity, the proposed RL algorithm adopted the deep Q-network algorithm. To train the RL algorithm, a detailed simulation environment model was established in the Transient System Simulation Tool (TRNSYS), including a building model and FCUs with a fresh-air system model. The simulation environment model can interact with the RL agent in real time through a self-developed TRNSYS–Python co-simulation platform. The RL algorithm was trained, tested, and evaluated based on the simulation environment model. The results indicate that compared with the traditional on/off and rule-based controllers, the RL algorithm proposed in this study can increase the joint control satisfaction rate of indoor temperature and relative humidity by 12.66% and 9.5%, respectively. This study provides preliminary direction for a deep reinforcement learning control strategy for indoor temperature and relative humidity in office building heating, ventilation, and air-conditioning (HVAC) systems.

Keywords: fan coil units; reinforcement learning; DQN algorithm; indoor temperature and relative humidity control; co-simulation

1. Introduction

With a rapidly developing economy, the number of office buildings in China has gradually increased. One study reported that the area of office buildings in China has grown from 1.6 billion to 4.8 billion m^2 in the past two decades [1]. An increase in building area has driven the growth of building energy consumption, and a study reported that the entire lifecycle energy consumption of buildings accounts for approximately 46.5% of the total energy consumption, and the total lifecycle carbon emissions of buildings account for 51.2% of total carbon emissions in China [2]. The total energy consumption and carbon emissions from office buildings are relatively high compared to other building

types [3–5]. This indicates that office buildings are essential for economic development. Furthermore, office buildings are a kind of building where modern people spend much time. Their indoor environmental quality significantly affects the health and work efficiency of users [6]. Mechanical and electrical equipment in office buildings also provides various convenient services for humans, such as air-conditioning equipment, elevators, and lighting. In general, office buildings are among the most important buildings that support economic growth, stimulate investments, and facilitate services in any country. Therefore, research on various aspects of office buildings is crucial for social and economic development, human health, and air quality.

As the location of human activities gradually shifts from outdoors to indoors, humans spend on average 80–90% of their time inside buildings, especially office buildings [7], which increases requirements for the environment and indoor thermal comfort of office buildings. Studies [8,9] have indicated that the office building environment and thermal comfort affect occupant health and work efficiency. With a suitable office environment and thermal comfort, occupants will have a greater sense of well-being, and increase their work efficiency by approximately 15–20%. Therefore, controlling the indoor-air status of office buildings within an appropriate range and improving the indoor environment quality of office buildings have received considerable attention from various scholars worldwide.

The indoor-air environment is mainly affected by outdoor weather, occupant behavior, and energy-using equipment [10]. Creation of the indoor-air environment mainly depends on various air-conditioning equipment. Currently, the number of office buildings in Chinese major cities is increasing, and fan coil units, as a type of air-conditioning equipment, have been widely used in heating, ventilation, and air-conditioning (HVAC) systems of office buildings because of their small size, high flexibility in arrangement, and individual control [11]. Existing research on the control of fan coil units has mainly focused on reducing the fluctuation of indoor temperature to obtain satisfactory indoor thermal comfort. This method of using only indoor temperature as a control object ignores the influence of indoor relative humidity on the human body and thermal comfort evaluation of the indoor environment by different groups of people in different relative humidities. Relative humidity affects human thermal comfort mainly by influencing heat and water–salt metabolism in the human body [12], and different people have different sensitivities to indoor relative humidity. For most people, fluctuations in indoor relative humidity within an appropriate range at the same indoor temperature do not significantly affect their thermal comfort evaluation of indoor environments, whereas, for people with respiratory diseases, differences in relative humidity significantly increase their discomfort and affect their actual thermal comfort evaluation of indoor environment [13,14]. Therefore, it is crucial to develop a control method for fan coil units to jointly control the indoor temperature and relative humidity of office buildings within an appropriate range for the health status of occupants, improvement in occupant work efficiency, and thermal comfort evaluation.

Currently, commonly used control methods for fan coil units usually consider only indoor temperature as the control object, such as on/off control, rule-based control (RBC), and proportional–integral–derivative (PID) control. These control methods are widely used in actual projects owing to their simple deployment. For example, a PID controller was proposed in [15] and its control effect on indoor temperature and relative humidity was tested. Lifei Xu of the Harbin Institute of Technology [16] designed a cascade control system for indoor temperature and relative humidity, and optimized the performance of the controller by self-tuning the parameters of the PID controller using artificial neural networks (ANNs). However, HVAC systems, as a class of highly nonlinear time-varying systems, often have difficulty achieving the desired control effect using conventional control methods [17]. Recently, the application of model predictive control (MPC) to HVAC systems has received considerable attention. The MPC as a supervisory control has better stability and multiobjective rolling optimization, but the operation effect of MPC depends on accurate mathematical models, and requires data information that can accurately reflect changes in indoor and outdoor building parameters [18]. If the difference between the

mathematical model and the actual HVAC system is significant, the control effect of the MPC is difficult to ensure.

With the development of big data technology and artificial intelligence (AI), a machine learning method that is model-free and self-learning has emerged in recent years called reinforcement learning (RL) [19–21]. Some scholars have conducted research on the optimal control of RL algorithms in HVAC systems. Table 1 summarizes the related RL studies for the optimal control of building HVAC systems. Junwei Yan et al. [22] applied the double deep Q network (DQN) algorithm to the energy-saving optimization operation of a central air-conditioning system in an office building in Guangzhou. In the premise of meeting indoor thermal comfort requirements, compared with PID control, this algorithm reduces the total energy consumption of the system by approximately 5.36%. Guangcai Gong et al. [23] applied the DQN algorithm to a variable air volume (VAV) system to save the total system energy and satisfy indoor thermal comfort. They verified that the control effect of the DQN algorithm was superior to RBC in most cases by controlling the setpoint of the air supply temperature and that of the chiller water supply temperature. Ruihua Ding et al. [24] proposed a deep reinforcement learning optimal control method based on expert knowledge to study a water-cooled air-conditioning system in a data center, and compared the method with traditional RBC and PID control to demonstrate that the method can reduce the total system energy consumption, while retaining the cabinet outlet air temperature within a safe range. Yan Du et al. [25,26] applied the deep deterministic policy gradient (DDPG) algorithm to a multizone residential HVAC system to minimize energy consumption costs while maintaining indoor environment thermal comfort. Zhiang Zhang et al. [27] proposed a control method based on the asynchronous advantage actor–critic (A3C) algorithm, which was then deployed in an actual radiant heating system for testing. The results demonstrated that the control method had over 95% probability of saving 16.6% of the heating demand during the deployment period. Marco Biemann et al. [28] evaluated four actor–critic algorithms in a simulated data center environment and demonstrated that all four algorithms could achieve zone temperature maintenance within the desired range while reducing energy consumption by 10% compared to a model-based controller. Guanyu Gao et al. [29] used the DDPG algorithm to regulate a HVAC system to reduce energy consumption while meeting the thermal comfort requirements of occupants. Yiqun Pan et al. [30] used a VAV air-conditioning system for an office building as a case study to validate the optimization performance of an RL controller-based DQN algorithm. They demonstrated that the RL controller is more energy efficient than RBC and PID controllers in meeting indoor temperature requirements. Currently, research on the application of reinforcement learning algorithms in HVAC systems mainly focuses on reducing the total system energy consumption while meeting indoor temperature requirements, and its control objects are mostly various setpoints. This ignores the potential risk of not meeting indoor relative humidity and deviation between setpoints and actual operating conditions.

From the literature review, in the control research of fan coil units, the current common control method considers temperature as a single control object and disregards the effect of indoor relative humidity. On the one hand, owing to the coupling relationship between indoor temperature and relative humidity, it is difficult to regulate fan coil units to maintain both the indoor temperature and relative humidity of office buildings within an appropriate range, and there are still relatively few related studies. On the other hand, as machine learning control methods, RL algorithms have appeared in recent years, and there have been preliminary studies on their application in HVAC systems, but most of these studies focus on reducing system energy consumption in the premise of meeting only indoor temperature requirements, and control objects are often various setpoints. There are few studies on joint control of both indoor temperature and relative humidity in office buildings using RL algorithms. Therefore, it is significant to study the RL algorithm for the joint control of indoor temperature and relative humidity. To solve these problems, this study considers fan coil units with a fresh-air system in an office building in Beijing as the study object, develops a TRNSYS–Python co-simulation platform, and proposes a reinforcement

learning algorithm based on action intervention to regulate the air supply volume for the fan coil units. The study objective is to improve the joint control satisfaction rate of indoor temperature and relative humidity. In summary, this study provides preliminary direction for a deep reinforcement learning control strategy for indoor temperature and relative humidity in office building HVAC systems.

Table 1. Summary of the RL algorithm for optimal control of HVAC system.

Reference	Building	HVAC System	Control Action	Optimization Objective	RL Algorithm
Junwei Yan et al. [22]	Office building	Water-cooled central air-conditioning system	Chilled water outlet temperature and chilled water flow	Energy consumption and indoor air temperature	Double DQN
Xi Fang et al. [23]	Office building	VAV	Air supply temperature setpoint and chilled supply water temperature setpoint	Energy consumption and thermal comfort	DQN
Ruihua Ding et al. [24]	Data center	Water-cooled central air-conditioning system	Chilled water outlet temperature and the pressure difference of chilled water pump	Energy consumption and temperature of air inlet area of cabinet	DQN based on expert knowledge
Yan Du et al. [25,26]	Residential building	Split air conditioner	Zone temperature setpoint	Energy consumption cost and thermal comfort	DDPG
Zhiang Zhang et al. [27]	Office building	Water-based radiant heating system	Supply water temperature setpoint	Energy consumption and thermal comfort	A3C
Marco Biemann et al. [28]	Data center	VAV	Zone temperature setpoint and zone fan mass flow rate	Energy consumption and indoor temperature	Actor–critic
Guanyu Gao et al. [29]	Laboratory	Split air conditioner	Air temperature setpoint and humidity setpoint	Energy consumption and thermal comfort	DDPG
Yiqun Pan et al. [30]	Office building	VAV	Air supply volume	Energy consumption and indoor temperature	DQN

The remaining parts of this article is organized as follows: Section 2 introduces methodology, including overall technical approach, establishment of simulation environment, algorithm principle and design, co-simulation platform operating principle, and algorithm evaluation; Section 3 shows the optimization control results of the controller proposed in this study, which are compared with other controllers. Additionally, the sensitivity analysis results of the DQN algorithm are also shown in this section; lastly, the conclusions and limitations are summarized in Section 4.

2. Methodology

2.1. Overall Technical Approach

In this study, an RL algorithm based on action intervention was proposed to regulate the air supply volume for the fan coil units, commonly used in office buildings in China. This study used a fan coil unit with a fresh-air system in an office building as a case study to validate the optimization control performance of the RL controller proposed in this study. The objective was to improve the joint control satisfaction rate of indoor temperature and relative humidity. The overall technical approach of this study is illustrated in Figure 1, and is divided into four parts:

- **Establish a building virtual simulation environment**. The building and its energy system are modeled in TRNSYS software, which provides an interactive environment for subsequent agent training.
- **Design and deployment of a reinforcement learning algorithm.** To improve the joint control satisfaction rate of indoor temperature and relative humidity, this study designed an RL algorithm with advanced applicability for regulating the air supply volume for the fan coil units. The algorithm was deployed in TensorFlow.
- **Development of the TRNSYS–Python co-simulation platform.** In this study, real-time interactions between TRNSYS and Python were realized using a data transfer method. This method is based on files. A co-simulation platform was developed for RL algorithm testing and evaluation.
- **Algorithm evaluation.** For the joint control effect on indoor temperature and relative humidity, the RL control algorithm proposed in this study was compared with the traditional control method. Subsequently, sensitivity of the RL algorithm was analyzed.
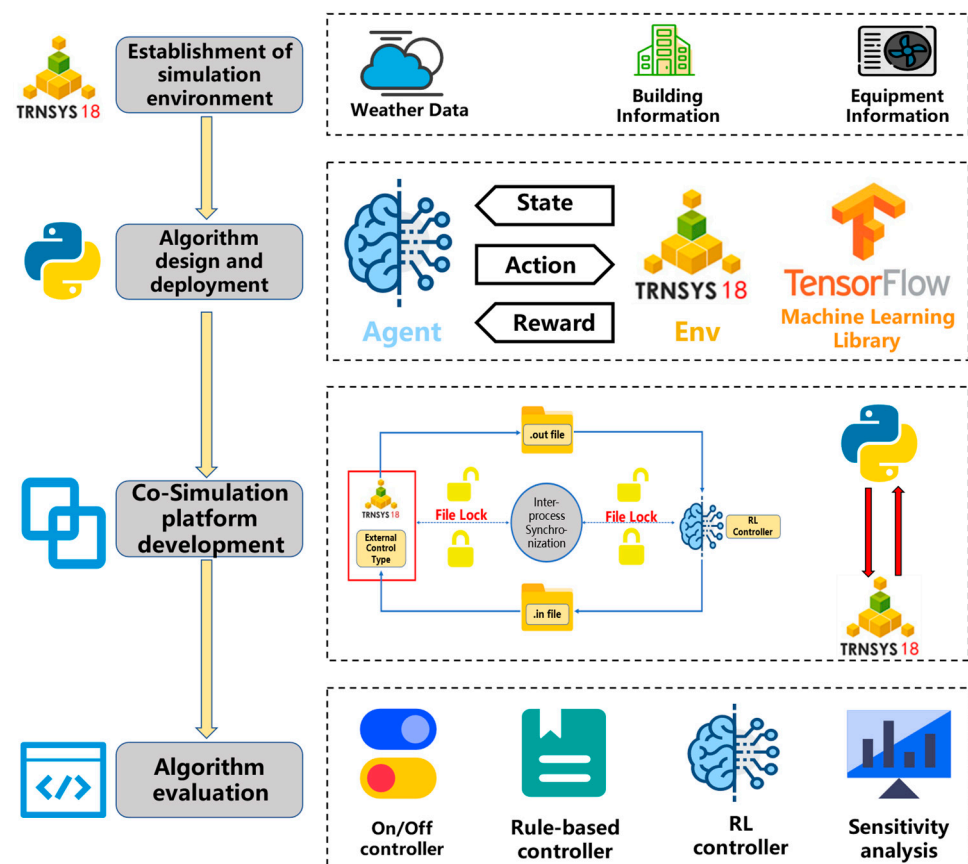

Figure 1. Schematic diagram of the overall technical approach.

2.2. Establishment of Simulation Environment

In this study, the Transient System Simulation Tool (TRNSYS) was used to build a simulation environment. TRNSYS is an extremely flexible graphically based software environment used to simulate the behavior of transient systems. It is a modular system with a large component library, and users can create their own models. The TRNSYS software has been widely used to study the performance simulation of HVAC components and systems. The software is also verified by comparing with the experimental setup. For example, Martinez et al. [31] modeled an air system with a desiccant wheel in TRNSYS, and then designed a test facility to verify the effectiveness of the model.

This study established a simulation environment in TRNSYS based on weather data, building information, and HVAC equipment information collected on-site. The simulation environment is used for subsequent algorithm training, testing, and evaluation.

2.3. Reinforcement Learning Algorithm Design and Deployment

2.3.1. Reinforcement Learning Introduction

In this study, an RL control algorithm based on action intervention was proposed to regulate the air supply volume for the fan coil units to improve the joint control satisfaction rate of indoor temperature and relative humidity in office buildings.

Reinforcement learning is the third basic learning method in machine learning, in addition to supervised and unsupervised learning. Its inspiration comes from behaviorism theory in psychology, which focuses on the idea that organisms constantly interact with the environment to obtain rewards or punishments given by the environment, and then gradually form expectations of rewards and punishments to produce actions that can obtain maximum benefits [32]. Figure 2 shows a schematic of the reinforcement learning algorithm.
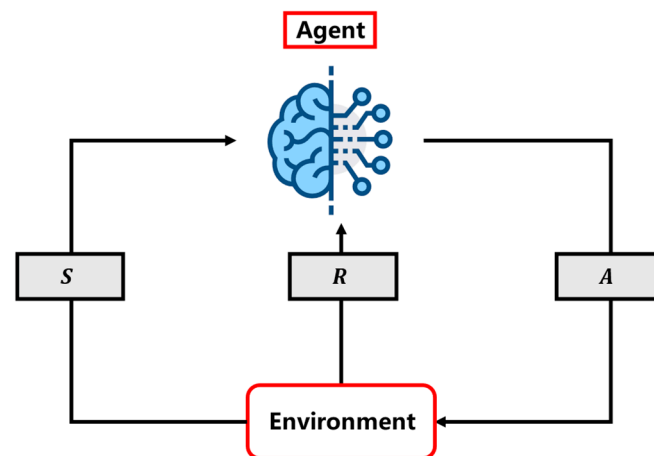

Figure 2. Schematic diagram of reinforcement learning algorithm.

In Figure 2, S stands for the state and observation of the agent, A stands for an action taken by the agent, and R stands for the reward given to the agent by the environment. The specific interaction process follows: at each decision moment t, the agent executes the action a_t, and after a time step Δt, the environment is at moment $t + 1$, and the state changes from s_t to s_{t+1}. The agent observes s_{t+1} and realizes the reward $R(s_t, a_t)$ in this time step, which is fed back by the environment.

The iterative object of the RL algorithm is the maximum expected reward value function Q based on the state–action pair, represented by $Q(s_t, a_t)$, which is the cumulative reward value that the system will obtain when the action a_t is executed in state s_t. Through the continuous interaction between the agent and environment, the Q value is updated by Equation (1).

$$Q^{new}(s_t, a_t) \leftarrow (1 - \alpha) \cdot Q^{old}(s_t, a_t) + \alpha \cdot (R(s_t, a_t) + \gamma \cdot maxQ(s_{t+1}, a)) \tag{1}$$

where α is the learning rate, $\alpha \in (0,1]$. When the learning rate approaches one, the algorithm converges faster, but the risk of oscillation is higher; when the learning rate approaches zero, the algorithm converges slower but the risk of oscillation is lower. Let γ denote the discount factor $\gamma \in [0,1]$, which means the effect of the current action on future long-term rewards. The larger γ is, the more the agent values long-term rewards obtained in the future; conversely, the smaller γ is, the more myopic the agent is regarding the rewards.

In an actual HVAC system, there are various devices and sensors, the dimensions of the state are large, and many states are continuous rather than discrete. Calculating each $Q(s_t, a_t)$ is complicated and inefficient. To solve this problem, a method for estimating the Q value using artificial neural networks (ANNs) was proposed. The input of the ANNs is the state, and its output is Q value for each action. Such RL algorithms equipped with ANNs are called deep reinforcement learning (DRL) algorithms. The deep Q network (DQN) algorithm is a DRL algorithm with two ANNs (i.e., Q-network and target Q-network) and an experience memory. The Q-network must be trained to output the maximum Q-value. The target Q-network does not require training but only serves as a label for the Q-network when it is trained, and its parameters are updated from replicating the Q-network parameters over a fixed time step. The experience memory holds experience generated by the agent interacting with the environment, which is extracted and input into the Q-network as training data when the Q-network is being trained. The specific flow of the DQN algorithm is presented in Algorithm 1.

Algorithm 1: Deep Q Network Algorithm Flow

1: Initialize memory M = [empty set]
2: Initialize Q network with parameters ω
3: Copy Q network and store as $\hat{Q}(\cdot|\hat{\omega})$
4: Initialize control action a and state s_{pre} and s_{cur}
5: **for** m: = 1 to N **do**
6:      Reset the environment to the initial state
7:      **for** t_s: = 0 to L **do**
8:          **if** t_s mod k == 0 **then**
9:              $s_{cur} \leftarrow current\ observation$
10:             $r = reward(s_{pre}, a, s_{cur})$
11:             $M \leftarrow (s_{pre}, a, r, s_{cur})$
12:             $Draw\ mini-batch(s, a, r, s') \leftarrow M$
13:             $Target\ vectors\ v \leftarrow target(s)$
14:             $Train\ Q(\cdot|\omega)\ with\ s, v$
15:             $Every\ d\Delta t_c\ steps,\ \hat{Q}(\cdot|\hat{\omega}) \leftarrow Q(\cdot|\omega)$
16:             $\varepsilon = \max(\varepsilon - \cdot\varepsilon, min\varepsilon)$
17:             $a = \begin{cases} a_i \in A \mid i = random(n),\ probality\ \varepsilon \\ argmaxQ(s_{cur}, a),\ otherwise \end{cases}$
18:             $s_{pre} \leftarrow s_{cur}$
19:         **end if**
20:         Execute action a in the environment
21:     **end for**
22: **end for**

Considering that the data generated by the operation of HVAC system are considerably large and complex, the indoor-air state is a continuous variable rather than discrete. The DQN algorithm was used in this study to solve the optimal control problem of the air supply volume for the fan coil units.

2.3.2. Design of the DQN Algorithm

- Selection of input parameters for the DQN algorithm.

In optimal control strategies based on DRL algorithms, selecting state *S* is important. The more influencing factors the state contains, the more comprehensive the information about the environment the agent receives, and the closer the final learned strategy is to the optimal control strategy. However, an increase in the state dimension leads to a longer training time and a more extensive space for the agent to explore, which increases the risk of failure in agent learning. Therefore, in this study, after numerous experiments, indoor temperature *tem* and indoor relative humidity *RH* are simultaneously selected as inputs for the DQN algorithm after conversion. These experiments mainly consider different combinations of input parameters and whether these input parameters need to be converted. Detailed experiment settings are shown in Table 2, and the conversion formulas are shown in Equations (2) and (3).

$$
tem' = \begin{cases} -1 + \frac{1-(-1)}{T_{upper\ bound} - T_{lower\ bound}} \cdot (tem - T_{lower\ bound}), & if\ T_{lower\ bound} \leq tem \leq T_{upper\ bound} \\ 1 + \left(tem - T_{upper\ bound} \right), & if\ tem > T_{upper\ bound} \\ -1 - \left(T_{lower\ bound} - tem \right), & if\ tem < T_{lower\ bound} \end{cases} \tag{2}
$$

$$
RH' = \begin{cases} -1 + \frac{1-(-1)}{RH_{upper\ bound} - RH_{lower\ bound}} \cdot (RH - RH_{lower\ bound}), & if\ RH_{lower\ bound} \leq RH \leq RH_{upper\ bound} \\ 1 + \left(RH - RH_{upper\ bound} \right)/10, & if\ RH > RH_{upper\ bound} \\ -1 - \left(RH_{lower\ bound} - RH \right)/10, & if\ RH < RH_{lower\ bound} \end{cases} \tag{3}
$$

where *tem* and *RH* denote the temperature and relative humidity before conversion, and *tem′* and *RH′* are the temperature and relative humidity after conversion, respectively. The purpose of Equation (2) is to distribute *tem′* between −1 and 1 when *tem* is between $T_{lower\ bound}$ and $T_{upper\ bound}$. If *tem* is greater than $T_{upper\ bound}$ or less than $T_{lower\ bound}$, *tem′* increases or decreases when the value of *tem* linearly exceeds the boundary. Similarly, the purpose of Equation (3) is to distribute *RH′* between −1 and 1 when *RH* is between $RH_{lower\ bound}$ and $RH_{upper\ bound}$. If *RH* exceeds the upper or lower boundary, *RH′* increases or decreases with the values of *RH* that exceeded the boundary, based on the scale of one tenth. This conversion keeps the scale of *RH′* close to that of *tem′*.

Table 2. Detailed experiment settings.

Potential Input Parameters	Conversion
Indoor temperature	Yes or No
Indoor relative humidity	Yes or No
Indoor humidity ratio	Yes or No
Outdoor temperature	Yes or No
Outdoor relative humidity	Yes or No
Outdoor humidity ratio	Yes or No
Wind velocity	Yes or No
Occupancy	No
Total horizontal radiation	Yes

- Output setting for the DQN algorithm.

The output of the DQN algorithm can be considered a controllable variable in a HVAC system. Based on the purpose of this study, the air supply volume for the fan coil units was selected as the output. The fan coil units used in this study have four levels of air supply volume: off, low, medium, and high, corresponding to 0%, 50%, 75%, and 100% of the rated air volume, respectively. Therefore, action space $A = [a_0, a_1, a_2, a_3] = [0, 50\%, 75\%, 100\%]$.

- Design of the reward function for the DQN algorithm.

In theory, an agent is trained to maximize the cumulative reward value. The design of the reward function determines the time an agent takes to train and whether the training is effective. According to the purpose of this study, the reward function is represented by the negative form of the temperature penalty and relative humidity penalty terms, as shown in Equations (4)–(6).

$$Reward = -k_1 \cdot penalty_{tem} - k_2 \cdot penalty_{RH} \tag{4}$$

$$penalty_{tem} = \begin{cases} 0, & if \ -1 \leq tem' \leq 1 \\ abs(tem') - 1, & else \end{cases} \tag{5}$$

$$penalty_{RH} = \begin{cases} 0, & if \ -1 \leq RH' \leq 1 \\ abs(RH') - 1, & else \end{cases} \tag{6}$$

where k_1 denotes the temperature penalty term coefficient and k_2 denotes the relative humidity penalty term coefficient.

- Exploration and exploitation of the DQN algorithm and hyperparameter setting.

In this study, we selected the ε-greedy exploration strategy to explore more state–action pairs, and the specific process is that in the training phase, a random number is generated at each time step, and if the random number is smaller than ε_i at this time, the agent randomly selects an action; otherwise, the agent selects an action based on the prediction of the Q-network. The formula for ε_i is shown in Equation (7).

$$\varepsilon_i = \varepsilon_0 - \varepsilon_{decay} \cdot step_i \tag{7}$$

where ε_{decay} is the decay coefficient of ε and $step_i$ is the *i*-th time step.

In this study, we intervened in the actions of the agent to avoid meaningless exploration and enhance the utility of the RL controller. Specifically, during the training phase, if the indoor temperature was higher than $T_{upper\ bound} + 2\ °C$, the air supply volume for the fan coil units was 100% of the rated air supply volume, and if the indoor temperature is lower than $T_{lower\ bound} - 2\ °C$, fan coil units were turned off. Such a setting enables the agent to avoid meaningless exploration and reduce the computation cost of learning. During the testing phase, if the indoor temperature was higher than $T_{upper\ bound}$, the fan coil units were turned on to high airflow volume (i.e., 100% of the rated air supply volume), and if the indoor temperature was lower than $T_{lower\ bound}$, the fan coil units were turned off. On the one hand, this setting can prevent the agent from ignoring indoor temperature to obtain appropriate indoor relative humidity; on the other hand, it can also avoid damage to HVAC equipment.

The settings for the other hyperparameters in the DQN algorithm used in this study are listed in Table 3.
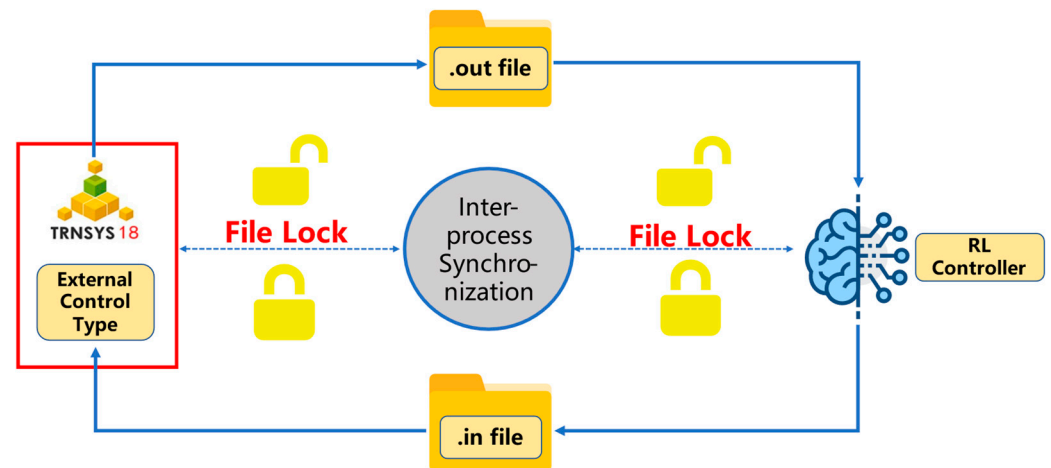
2.4. TRNSYS–Python Co-Simulation Platform Development

The agent must be trained to learn the control strategy. During training, the agent must continuously receive information regarding the environment and output an action to be executed. If an untrained DQN algorithm is deployed in an actual building HVAC system, there is a risk of equipment damage and serious deviation of the indoor air from the comfort range. Therefore, in this study, a virtual simulation environment was built in the TRNSYS software for training the agent, testing, and evaluating the DQN algorithm. An RL controller based on the DQN algorithm was implemented in Python, and the artificial neural networks were built and trained in the free and open-source deep learning library TensorFlow. TensorFlow is an open-source software library developed by Google for various machine learning tasks in perception and language understanding.

Table 3. Hyperparameter settings.

Hyperparameter	Candidate Value	Selected Value
k_1	{1, 2, 5, 10}	5
k_2	-	1
α	{0.001, 0.01, 0.1, 0.3, 0.5, 0.9}	0.01
γ	{0.01, 0.1, 0.3, 0.5, 0.9, 0.99}	0.1
ε_{decay}	{0.0001, 0.001, 0.01}	0.01
ε_0	{0.1, 0.3, 0.5, 0.8, 1}	0.3
Memory size	{32, 64, 128, 256}	256
Batch Size	{32, 64, 128, 256}	256
Target network update	{1 day, 3 days, 5 days, 7 days}	1 day
Number of hidden layers	{1, 2}	1
Number of neural units in hidden layer	{32, 64, 128, 256}	128
Activation function of hidden layer	-	Sigmoid
Activation function of output layer	-	Linear
Number of inputs	-	2
Number of outputs	-	4
Optimizer	-	Adam

To achieve real-time interaction between TRNSYS and the RL controller, we used a file-based data transfer method. Specifically, the RL controller writes a control action to the *.in* file, TRNSYS reads the file and executes the corresponding action, and after reaching the next simulation time step, TRNSYS writes information about the environment to the *.out* file, which is read by the RL controller. The data transfer principle is illustrated in Figure 3.

Figure 3. Schematic diagram of data transfer.

Based on the design of the DQN algorithm and the real-time interaction between the TRNSYS software and the RL controller, the overall architecture of the TRNSYS–Python co-simulation platform proposed in this study is shown in Figure 4.
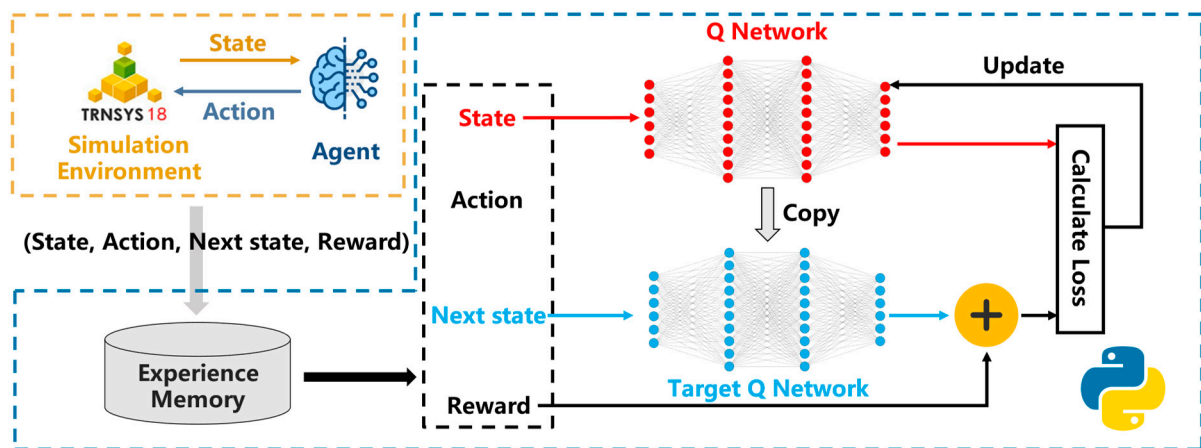
Figure 4. Overall architecture of the TRNSYS–Python co-simulation platform.

2.5. Algorithm Evaluation

2.5.1. Metric for Training Convergence

The agent training must be stopped at the appropriate time. If the training time is overly short, the learning of the agent may be incomplete, and the reliability of the experience learned by the agent may be insufficient. If the training time is overly long, the artificial neural network may fall into the predicament of overfitting. Therefore, it is necessary to set an appropriate metric to determine whether the training of the agent should end. After repeated experiments, we select *Stepwise Average Reward* as the metric for training convergence, as shown in Equation (8).

$$Stepwise\ Average\ Reward = \frac{1}{N}\sum_{i=1}^{N} r_i \tag{8}$$

where r_i denotes the value of the reward in the *i*-th time step and N denotes the number of time steps performed.

2.5.2. Comparison and Evaluation of Control Effects

To verify the effectiveness of the proposed RL controller for the joint control of indoor temperature and relative humidity, we selected on/off and rule-based controllers commonly used in various projects. The specific settings for these controllers are presented in Table 4.

Table 4. Controller settings.

Controller	Indoor Temperature	Air Supply Volume for Fan Coil Units	
On/Off controller	$tem \geq 27\ °C$	I	50% of the rated air volume
		II	75% of the rated air volume
		III	100% of the rated air volume
	$tem \leq 25\ °C$	0	
	Elsewise	Keep the air supply volume constant	
Rule-based controller	$tem \geq 26.7\ °C$	100% of the rated air volume	
	$tem \leq 25.3\ °C$	0	
	Elsewise	75% of the rated air volume	

In this study, we selected the temperature satisfaction rate, relative humidity satisfaction rate, and joint control satisfaction rate of temperature and relative humidity as evaluation indices, which are calculated as shown in Equations (9)–(11).

$$\phi_{tem} = \frac{n_{tem}}{N} \cdot 100\% \tag{9}$$

$$\phi_{RH} = \frac{n_{RH}}{N} \cdot 100\% \tag{10}$$

$$\phi_{tem\&RH} = \frac{n_{tem\&RH}}{N} \cdot 100\% \tag{11}$$

where n_{tem} is the number of indoor temperature points within the upper and lower limits; n_{RH} is the number of indoor relative humidity points within the upper and lower limits, $n_{tem\&RH}$ is the number of both indoor temperature points and relative humidity points within the upper and lower limits, respectively; and N is the total number of points.

2.5.3. Sensitivity Analysis

To evaluate the sensitivity of the DQN algorithm, we analyzed the sensitivity of the key parameters (i.e., learning rate α and discount factor γ) in Equation (1). First, the discount factor γ was fixed, and joint control effects on temperature and relative humidity based on the RL controller with different learning rates were compared. Subsequently, we fixed the learning rate α and compared joint control effects on temperature and relative humidity based on the RL controller with varying discount factors.

3. Case Study

3.1. Case Introduction

The building for the case study is a trade union activity room in an office building in the Haidian District, Beijing, with an area of 116 m^2. Its air-conditioning system comprises fan coil units with a fresh-air system. The geometry of the building was modeled using SketchUp software, as shown in Figure 5. A schematic diagram of the HVAC system operation is shown in Figure 6. The virtual simulation environment of the entire building's HVAC system was built in TRNSYS software, as shown in Figure 7. The RL controller regulates the air supply volume for the fan coil units to improve the joint control satisfaction rate of indoor temperature and relative humidity. The thermodynamic parameters of the office building envelope are listed in Table 5. The setting for the building envelopes is based on actual engineering design drawings. The settings for the environmental thermal disturbances in the office are listed in Table 6. The other HVAC system settings are listed in Table 7. These settings refer to the *Design Standards for Energy Efficiency of Public Buildings* and on-site investigation. It should be noted that the air conditioner is set to be turned on one hour earlier than occupancy. This setting is to ensure that the indoor temperature is within the appropriate range when the staff enters the room, improving the thermal comfort of the staff.
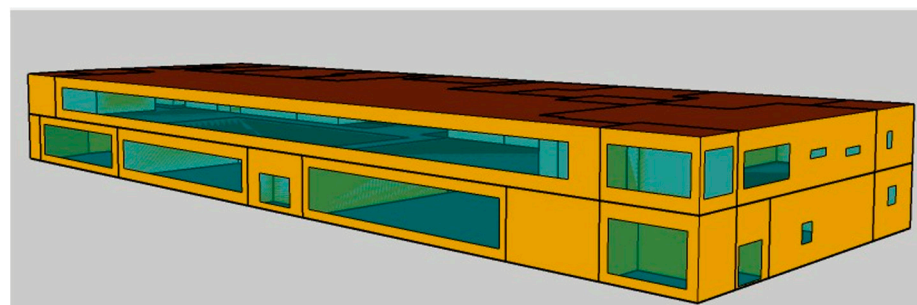

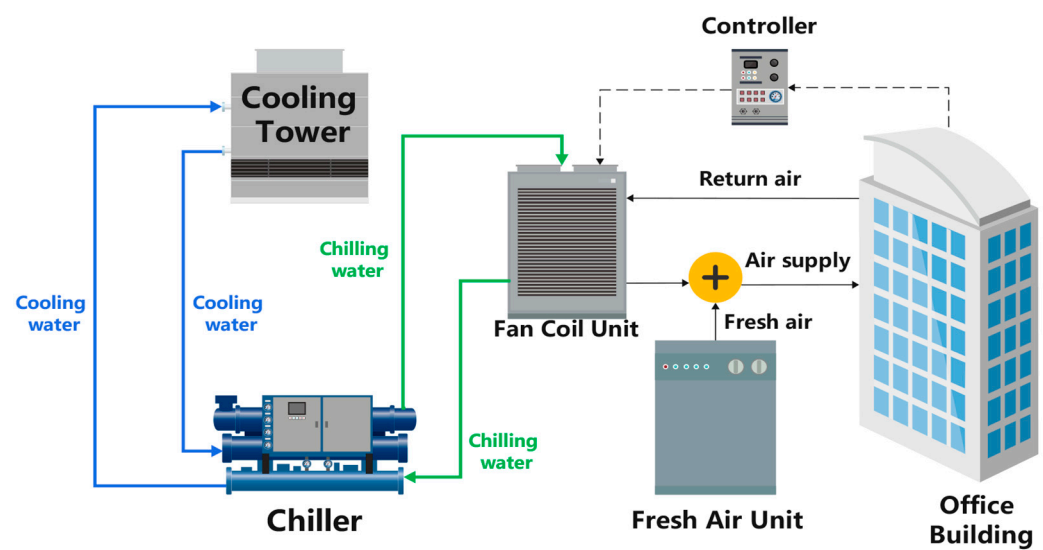
Figure 5. Architectural model in SketchUp.

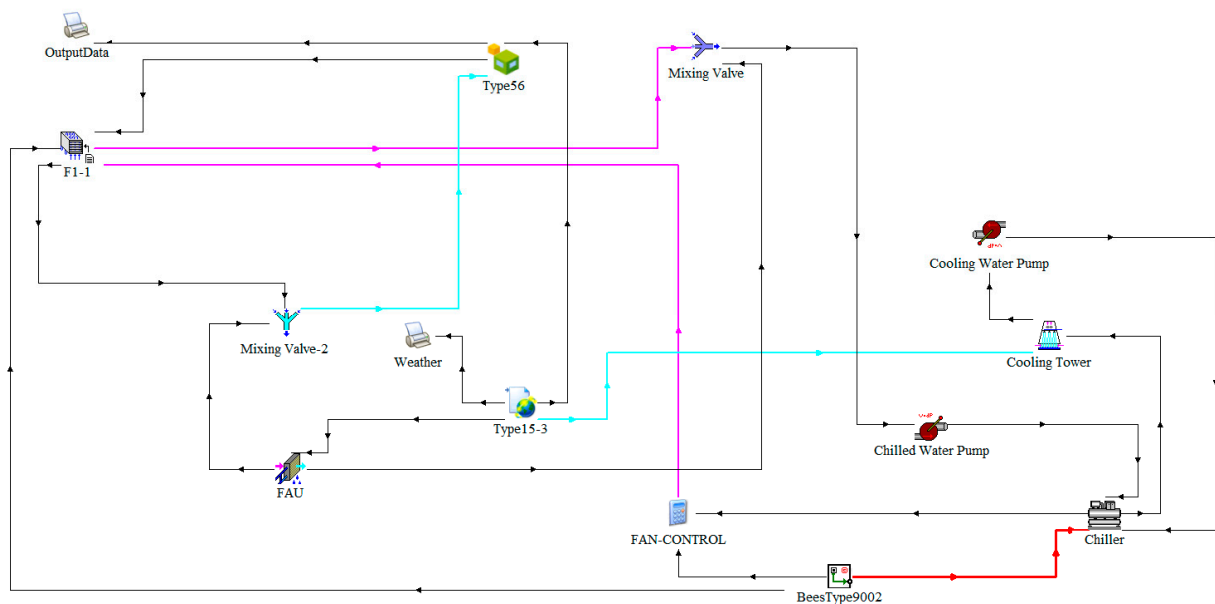**Figure 6.** Schematic diagram of the HVAC system operation.

Figure 7. Schematic diagram of the TRNSYS simulation system.

Table 5. Building envelopes.

	Building Envelope		Heat Transfer Coefficient W/(m²·K)
External wall	Cement slag mortar	20 mm	2.266
	Steel reinforced concrete	370 mm	
	Dali granite basalt	20 mm	
Roof	Cement mortar	20 mm	0.804
	Cellular concrete	200 mm	
	Steel reinforced concrete	130 mm	
	Cement mortar	15 mm	
External window	Glass	6 mm	1.46 (SHGC = 0.52)

Table 6. Thermal disturbances.

Thermal Disturbance		Value
Human body heat generation		66 W/p
Occupant density		0.1 p/m^2
Human body moisture generation		0.109 kg/(h·p)
Light and equipment heat generation		2.586 W/m^2
Occupancy	8:00–20:00	1
	Elsewise	0

Table 7. Settings for the HVAC system.

Setting Items		Value
Fresh-air volume		10% of total air volume
Indoor environmental control objectives	Upper limit of indoor temperature	27 °C
	Lower limit of indoor temperature	25 °C
	Upper limit of indoor relative humidity	60%
	Lower limit of indoor relative humidity	40%
Air conditioning on or off	7:00–20:00	On
	Elsewise	Off
Simulation time step		12 min

3.2. Simulation Results Analysis of the Reinforcement Learning Controller

In this study, we selected 0:00 on July 1 to 0:00 on July 15 as the training period, and the stepwise average reward curve of the training process is shown in Figure 8.
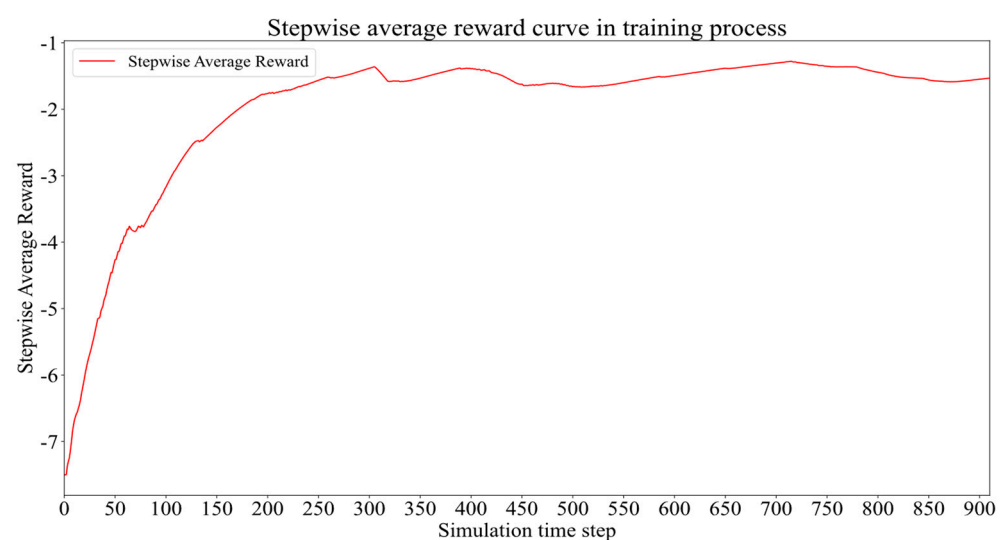
Figure 8. Stepwise average reward curve during training.

As shown in Figure 8, the stepwise average reward climbs rapidly during the first 300 steps when the agent constantly interacts with the environment and learning experiences. After 300 steps, the agent initially completed learning and continued to interact

with the environment and learn more experiences, and the stepwise average reward curve fluctuated within a small range.

The trained model was tested from 0:00 on 1 August to 0:00 on 31 August, and the simulation results during the test period were counted; the results are shown in Figure 9.
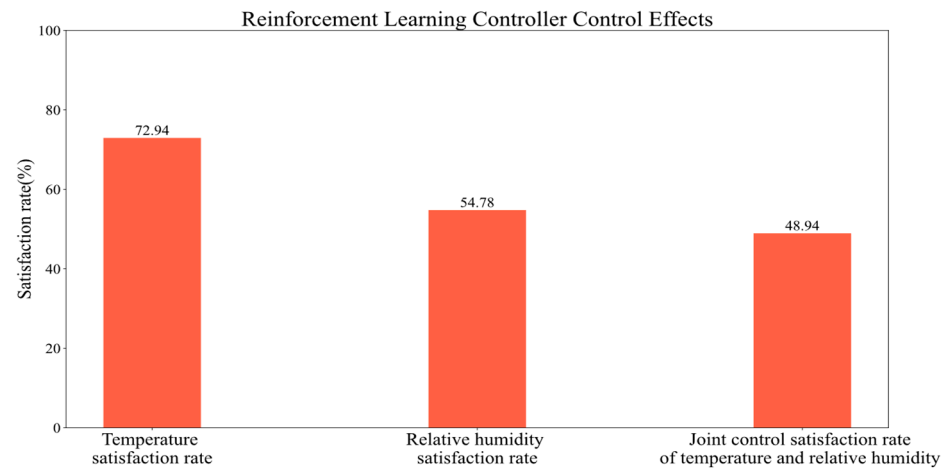
Figure 9. Reinforcement learning controller control effects.

In this study, we selected indoor-air state on a typical day (5 August) for drawing, and the results are shown in Figure 10.
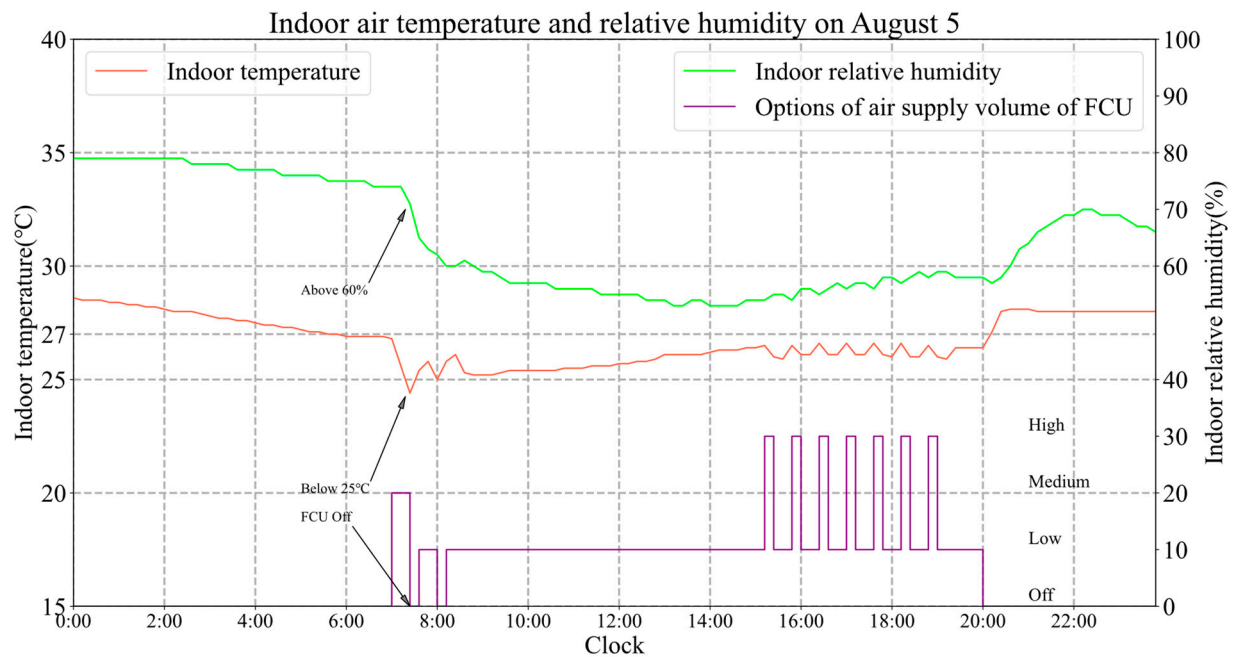
Figure 10. Indoor air temperature and relative humidity on 5 August.

As shown in Figure 10, when the indoor temperature initially deviates from the comfort range and the relative humidity is outside the comfort range, the RL controller proposed in this study takes action to maintain the indoor temperature near the comfort range and to avoid further deviation of the indoor temperature from the comfort range. This ensures the normal operation of the HVAC equipment and avoids damaging the equipment. When the indoor temperature is within the comfortable range, the RL controller can select the air supply volume for the fan coil units to achieve better joint control satisfaction rate of indoor temperature and relative humidity.

3.3. Comparison and Analysis of Simulation Results for Different Controllers

To further verify control the effect of the reinforcement learning controller on the indoor temperature and relative humidity, we selected the on/off and rule-based controllers for simulation comparison. The simulation results of the indoor temperature and relative humidity in different controllers are shown in Figures 11 and 12.
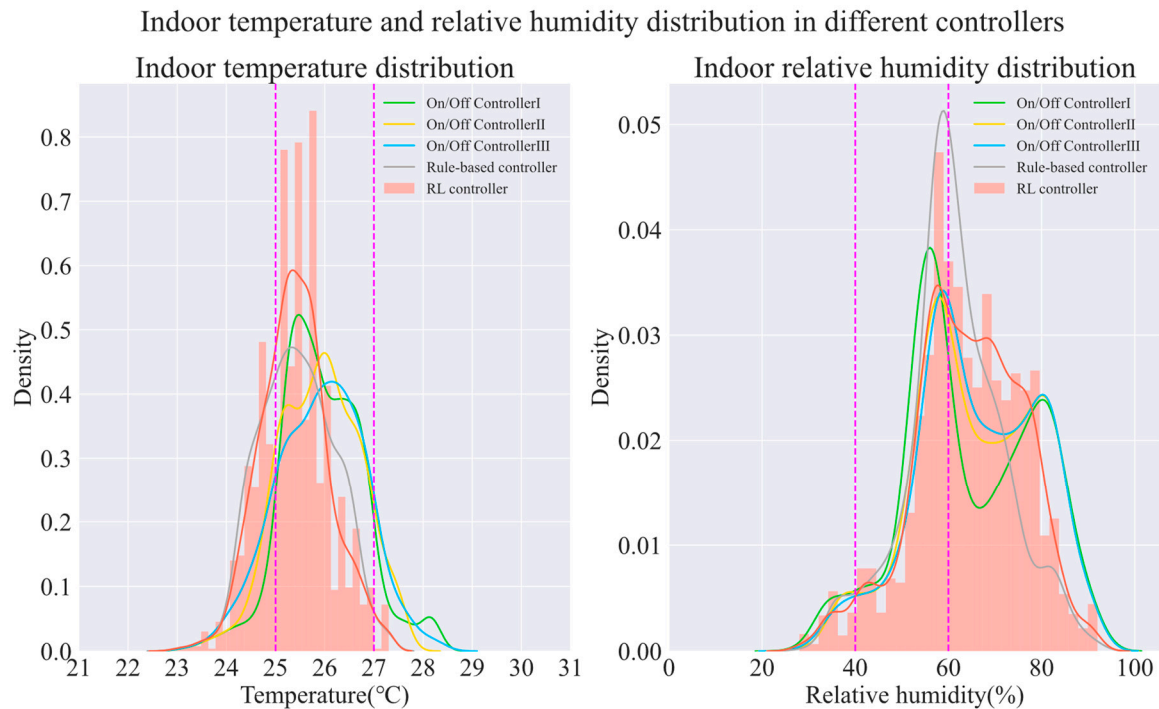
Figure 11. Indoor temperature and relative humidity distribution in different controllers.
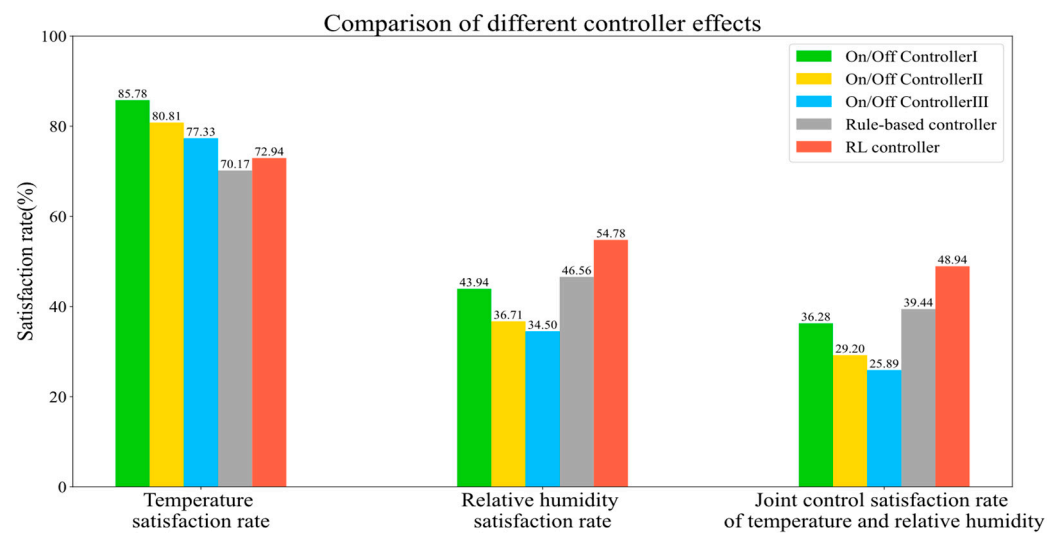
Figure 12. Comparison of different controller effects.

From the temperature distribution shown in Figure 11, the center of distribution of the indoor temperature is more biased toward 25 °C in the RL controller, rule-based controller, and on/off controller I, while the center of distribution of indoor temperature is more concentrated at 26 °C in the on/off controller II and on/off controller III. From the distribution of relative humidity in Figure 11, the distribution of indoor relative humidity deviates to a higher relative humidity in all five controllers. By analyzing the weather data,

it was found that there were more cloudy and rainy days during the test period, and the relative humidity of the outdoor atmosphere was higher during cloudy and rainy days, thus increasing the indoor relative humidity.

As shown in Figure 12, for the indoor temperature satisfaction rate, the on/off controller I has the best control effect, which is 85.78%, and the rule-based controller has the worst effect, which is 70.17%. For the satisfaction rate of indoor relative humidity, the control effect of the RL controller is the best, at 54.78%, and the effect of the on/off controller III is the worst, at 34.50%. For the joint control satisfaction rate of indoor temperature and relative humidity, the control effect of the RL controller is the best, at 48.94%, 9.5% higher than that of the rule-based controller, and 12.66% higher than that of the on/off controller I.

3.4. Sensitivity Analysis

To evaluate the sensitivity of the DQN algorithm, the key parameters (i.e., learning rate α and discount factor γ) in Equation (1) were quantitatively analyzed in this study.

Maintaining the discount factor $\gamma = 0.1$ constant, a result of comparing the joint control effect on indoor temperature and relative humidity by the RL controller at different learning rates is shown in Figure 13.
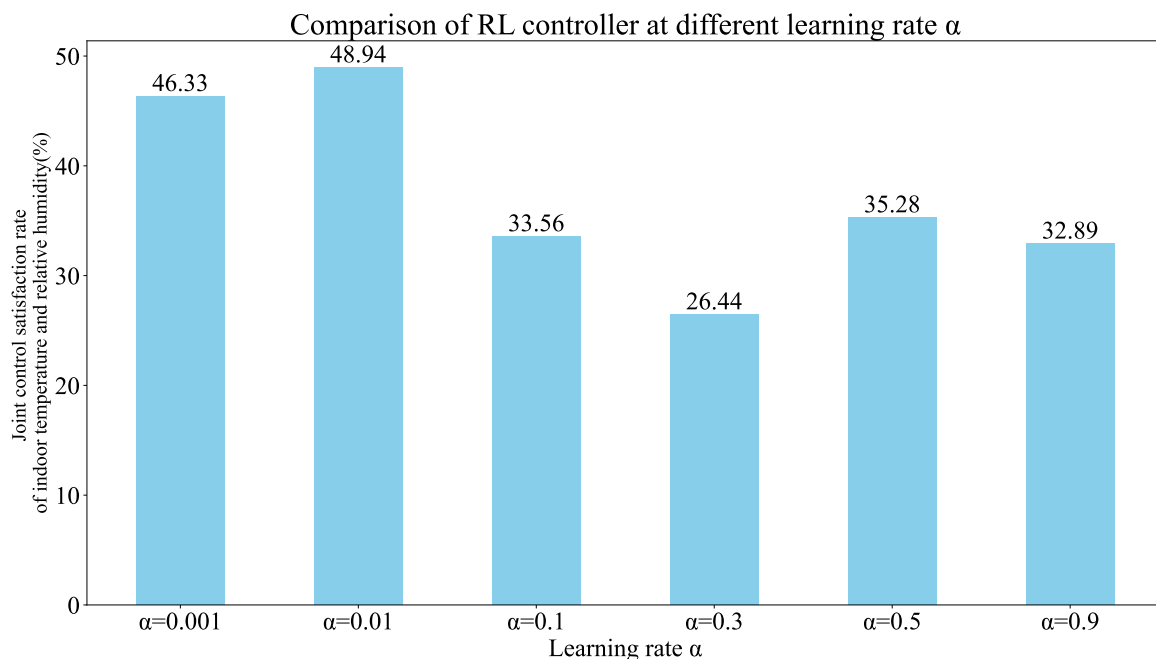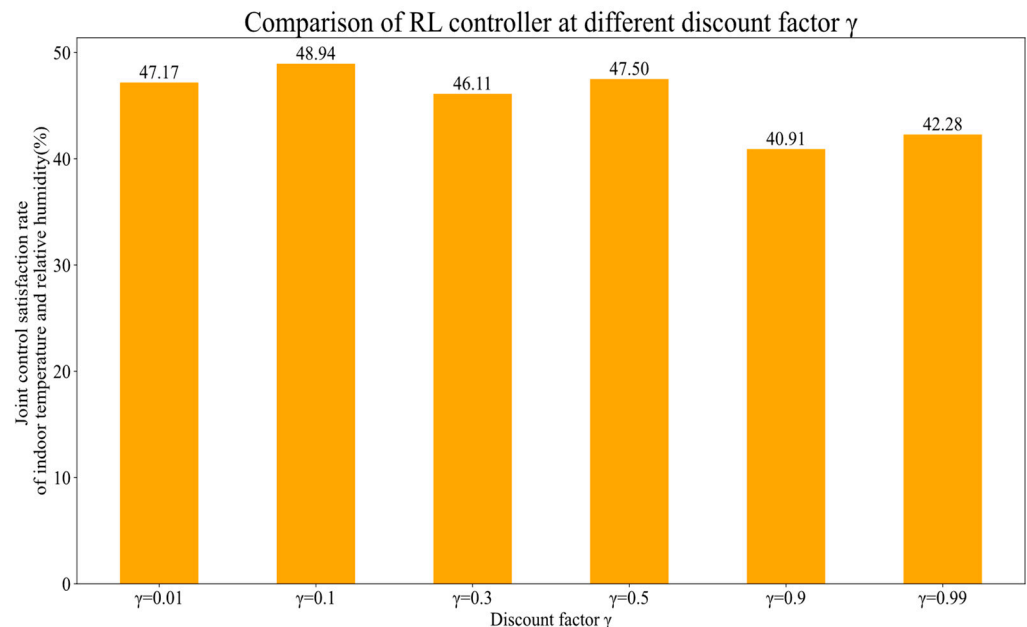

Figure 13. Comparison of the RL controller at different learning rates.

As shown in Figure 13, the joint control effect on indoor temperature and relative humidity by the proposed RL controller is relatively robust in the range of learning rate $\alpha \leq 0.01$. When the learning rate $\alpha \geq 0.01$, the control effect of the controller is reduced and oscillation occurs because with an increase in the learning rate α, the training of the agent oscillates and converges with difficulty.

When the learning rate $\alpha = 0.01$ is constant, the result of comparing the joint control effect on indoor temperature and relative humidity by the RL controller at different discount factors is shown in Figure 14.

As shown in Figure 14, the sensitivity of the proposed RL controller to the discount factor γ is weaker than that of the learning rate α. The overall control effect of the controller is more robust at different discount factors. However, for smaller discount factors $\gamma \leq 0.5$, the controller has a better joint control effect on indoor temperature and relative humidity. This is because the input parameters of the DQN algorithm are the indoor temperature and relative humidity at the same time, and no outdoor weather parameters are introduced. If

we aim to achieve a better control effect, we need the agent to prefer immediate rewards; that is, we need a relatively "short-sighted" agent, so the selection of the discount factor is more suitable for a smaller value.

Figure 14. Comparison of RL controller at different discount factors.

4. Conclusions

In this study, an RL control method based on action intervention was proposed, and its input parameters, reward function, and agent exploration and exploitation mechanism were designed. Subsequently, this study considered fan coil units with a fresh-air system in an office building in Beijing as the research object and developed a TRNSYS–Python co-simulation platform to verify the control effect of the proposed method, and the following conclusions were obtained:

(1) Using file-based data transfer, this study developed a TRNSYS–Python co-simulation platform, which makes it easier to train the agent and test and evaluate the performance of comprehensive RL algorithms in a simulation environment.

(2) The DQN algorithm based on action intervention can reduce the training time computation cost in the training phase and increase the security of algorithm deployment in the testing phase. From the simulation results of this study, the algorithm can achieve a better joint control effect on indoor temperature and relative humidity in office buildings. Specifically, the method can improve the joint control satisfaction rate of indoor temperature and relative humidity by 9.5% and 12.66%, respectively, compared with the traditional rule-based controller and on/off controller I.

(3) The setting of hyperparameters has a relatively significant impact on the control performance of the algorithm, which is robust when the hyperparameters are in an appropriate range. Otherwise, the control effect of the algorithm is reduced, and there is a risk of oscillation.

Therefore, the control method proposed in this study can achieve a better joint control effect on indoor temperature and relative humidity in office buildings. This study provides a new direction for indoor thermal comfort and environment control in office buildings, and has engineering application value.

Deep reinforcement learning for optimization control of an HVAC system is a complicated problem, and some limitations need to be improved and studied in the future. Firstly, the HVAC system and control action in this study are relatively simple, not involving heat and humidity transfer between multizones or excessive control actions. The stronger the coupling and nonlinear relationship between control actions of the HVAC system, the more

potent the RL will be. Applying RL to complicated HVAC systems is challenging work. Furthermore, the RL controller proposed in this study is not deployed in an actual HVAC system. In the future, it will be significant to deploy the RL controller in an actual building and evaluate the practical control effect. Finally, the building in this case study is located in Beijing, China. It is desirable to test the control effect of the proposed RL controller on buildings in different climate zones.

References

1. Tsinghua University Building Energy Conservation Research Center. *China Building Energy Conservation Annual Development Research Report 2022 (Public Building Topics)*; China Building Industry Press: Beijing, China, 2022; pp. 40–83.
2. China Association of Building Energy Efficiency. *China Building Energy Consumption Research Report 2020*; China Building Industry Press: Beijing, China, 2020.
3. Amber, K.; Ahmad, R.; Aslam, M. Intelligent techniques for forecasting electricity consumption of buildings. *Energy* **2018**, *157*, 886–893. [CrossRef]
4. Ye, H.; Hu, X.; Ren, Q.; Lin, T.; Li, X.; Zhang, G.; Shi, L. Effect of urban micro-climatic regulation ability on public building energy usage carbon emission. *Energy Build.* **2017**, *154*, 553–559. [CrossRef]
5. Ye, H.; Ren, Q.; Shi, L.; Song, J.; Hu, X.; Li, X.; Zhang, G.; Lin, T.; Xue, X. The role of climate, construction quality, microclimate, and socio-economic conditions on carbon emissions from office buildings in China. *J. Clean. Prod.* **2018**, *171*, 911–916. [CrossRef]
6. Chan, I.Y.S.; Liu, A.M.M. Effects of neighborhood building density, height, greenspace, and cleanliness on indoor environment and health of building occupants. *Build Environ.* **2018**, *145*, 213–222. [CrossRef] [PubMed]
7. Zhong, L.; Yuan, J.; Fleck, B. Indoor Environmental Quality Evaluation of Lecture Classrooms in an Institutional Building in a Cold Climate. *Sustainability* **2019**, *11*, 6591. [CrossRef]
8. Thayer, J.F.; Verkuil, B.; Brosschot, J.F.; Kampschroer, K.; West, A.; Sterling, C.; Christie, I.C.; Abernethy, D.R.; Sollers, J.J.; Cizza, G.; et al. Effects of the physical work environment on physiological measures of stress. *Eur. J. Cardiovasc. Prev. Rehabil.* **2010**, *17*, 431–439. [CrossRef]
9. Kojima, T.; Sakuma, T.; Nishihara, N.; Hayashi, T.; Munakata, J. Causal Modeling Between Workplace Productivity and Workers' Satisfaction with Various Spaces in Office Buildings. *J. Asian Archit. Build. Eng.* **2018**, *16*, 409–415. [CrossRef]
10. Zhou, X.; Lu, Y.; Hu, S.; Yang, Z.; Yan, D. New perspectives on temporal changes in occupancy characteristics of residential buildings. *J. Build. Eng.* **2023**, *64*, 105590. [CrossRef]
11. Li, X.; Zhao, T.; Zhang, J.; Chen, T. Development of network control platform for energy saving of fan coil units. *J. Build. Eng.* **2017**, *12*, 155–160. [CrossRef]
12. Vellei, M.; Herrera, M.; Fosas, D.; Natarajan, S. The influence of relative humidity on adaptive thermal comfort. *Build. Environ.* **2017**, *124*, 171–185. [CrossRef]
13. Bozic, A.; Kanduc, M. Relative humidity in droplet and airborne transmission of disease. *J. Biol. Phys.* **2021**, *47*, 1–29. [CrossRef] [PubMed]
14. Razjouyan, J.; Lee, H.; Gilligan, B.; Lindberg, C.; Nguyen, H.; Canada, K.; Burton, A.; Sharafkhaneh, A.; Srinivasan, K.; Currim, F.; et al. Wellbuilt for wellbeing: Controlling relative humidity in the workplace matters for our health. *Indoor Air* **2020**, *30*, 167–179. [CrossRef] [PubMed]
15. Ghaddar, D.; Itani, M.; Ghaddar, N.; Ghali, K.; Zeaiter, J. Model-based adaptive controller for personalized ventilation and thermal comfort in naturally ventilated spaces. *Build. Simul.* **2021**, *14*, 1757–1771. [CrossRef]

16. Lifei, X. Simulation Research on Cascade Control of Temperature and Humidity in VAV System. Master's Thesis, Harbin Institute of Technology, Harbin, China, 2006.
17. Han, Z.; Fu, Q.; Chen, J.; Wang, Y.; Lu, Y.; Wu, H.; Gui, H. Deep Forest-Based DQN for Cooling Water System Energy Saving Control in HVAC. *Buildings* **2022**, *12*, 1787. [CrossRef]
18. Afram, A.; Janabi-Sharifi, F.; Fung, A.S.; Raahemifar, K. Artificial neural network (ANN) based model predictive control (MPC) and optimization of HVAC systems: A state of the art review and case study of a residential HVAC system. *Energy Build.* **2017**, *141*, 96–113. [CrossRef]
19. Fan, C.; Yan, D.; Xiao, F.; Li, A.; An, J.; Kang, X. Advanced data analytics for enhancing building performances: From data-driven to big data-driven approaches. *Build. Simul.* **2020**, *14*, 3–24. [CrossRef]
20. Hong, T.; Wang, Z.; Luo, X.; Zhang, W. State-of-the-art on research and applications of machine learning in the building life cycle. *Energy Build.* **2020**, *212*, 109831. [CrossRef]
21. Vázquez-Canteli, J.R.; Nagy, Z. Reinforcement learning for demand response: A review of algorithms and modeling techniques. *Appl. Energy* **2019**, *235*, 1072–1089. [CrossRef]
22. Junwei, Y.; Qi, H.; Xuan, Z. Energy-saving and Optimal Operation of Central Air-conditioning System based on Double-DQN. *J. South China Univ. Technol. (Nat. Sci. Ed.)* **2019**, *47*, 135–144.
23. Fang, X.; Gong, G.; Li, G.; Chun, L.; Peng, P.; Li, W.; Shi, X.; Chen, X. Deep reinforcement learning optimal control strategy for temperature setpoint real-time reset in multi-zone building HVAC system. *Appl. Therm. Eng.* **2022**, *212*, 118552. [CrossRef]
24. Ruihua, D.; Chenggang, C.; Yixuan, W. Air conditioning System Optimization in Data Center Based on Deep Reinforcement Learning. *Cryog./Refrig.* **2022**, *50*, 79–85.
25. Du, Y.; Zandi, H.; Kotevska, O.; Kurte, K.; Munk, J.; Amasyali, K.; McKee, E.; Li, F. Intelligent multi-zone residential HVAC control strategy based on deep reinforcement learning. *Appl. Energy* **2021**, *281*, 116117. [CrossRef]
26. Du, Y.; Li, F.; Munk, J.; Kurte, K.; Kotevska, O.; Amasyali, K.; Zandi, H. Multi-task deep reinforcement learning for intelligent multi-zone residential HVAC control. *Electr. Power Syst. Res.* **2021**, *192*, 106959. [CrossRef]
27. Zhang, Z.; Lam, K.P. Practical implementation and evaluation of deep reinforcement learning control for a radiant heating system. In Proceedings of the Proceedings of the 5th Conference on Systems for Built Environments, Shenzhen, China, 7–8 November 2018; pp. 148–157.
28. 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. [CrossRef]
29. Gao, G.; Li, J.; Wen, Y. DeepComfort: Energy-Efficient Thermal Comfort Control in Buildings Via Reinforcement Learning. *IEEE Internet Things J.* **2020**, *7*, 8472–8484. [CrossRef]
30. Yuan, X.; Pan, Y.; Yang, J.; Wang, W.; Huang, Z. Study on the application of reinforcement learning in the operation optimization of HVAC system. *Build. Simul.* **2020**, *14*, 75–87. [CrossRef]
31. Martínez, P.J.; Llorca, C.; Pla, J.A.; Martínez, P. Experimental Validation of the Simulation Model of a DOAS Equipped with a Desiccant Wheel and a Vapor Compression Refrigeration System. *Energies* **2017**, *10*, 1330. [CrossRef]
32. Zhihua, Z. *Machine Learning*; Tsinghua University Press: Beijing, China, 2016; pp. 371–397.