https://doi.org/10.1016/j.buildenv.2021.107978

An, Y., Xia, T., You, R., Lai, D., Liu, J., and Chen, C. 2021. A reinforcement learning approach for control of window behavior to reduce indoor $PM_{2.5}$ concentrations in naturally ventilated buildings. ***Building and Environment***, 200: 107978.

A Reinforcement Learning Approach for Control of Window Behavior to Reduce Indoor $PM_{2.5}$ Concentrations in Naturally Ventilated Buildings

Yuting An[1], Tongling Xia[1], Ruoyu You[2], Dayi Lai[3], Junjie Liu[4], Chun Chen[1,5,*]

[1] Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Shatin, N.T. 999077, Hong Kong SAR, China

[2] Department of Building Services Engineering, The Hong Kong Polytechnic University, Kowloon, 999077, Hong Kong SAR, China

[3] School of Design, Shanghai Jiao Tong University, Shanghai 200240, China

[4] Tianjin Key Lab of Indoor Air Environmental Quality Control, School of Environmental Science and Engineering, Tianjin University, Tianjin, 300072, China

[5] Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen 518057, China

[*]Email: chunchen@mae.cuhk.edu.hk

Abstract:

Smart control of window behavior is a means of effectively reducing concentrations of indoor $PM_{2.5}$ (particulate matter with aerodynamic diameter less than 2.5 μm) in naturally ventilated residential buildings without indoor air cleaning devices. This study aimed to develop a reinforcement learning approach to automatically control window behavior in real time for mitigation of indoor $PM_{2.5}$ pollution. The proposed method trains the window controller with the use of a deep Q-network (DQN) in a specific naturally ventilated apartment in the course of a month. The trained controller can then be employed to control window behavior in order to reduce the indoor $PM_{2.5}$ concentrations in that apartment. The required input data for the controller are the real-time indoor and outdoor $PM_{2.5}$ concentrations with a 1-min resolution, which can easily be obtained with low-cost sensors available on the market. A series of simulations were conducted in a virtual typical apartment in Beijing and a real apartment in Tianjin. The results show that, compared with the baseline I/O ratio algorithm, the proposed reinforcement learning window-control algorithm reduced the average indoor $PM_{2.5}$ concentration by 12.80% in a one-year period. Furthermore, the proposed algorithm reduced the indoor $PM_{2.5}$ concentrations in the real apartment by 9.11% when compared with the I/O ratio algorithm and by 7.40% when compared with real window behavior.

Keywords: Reinforcement Learning, Smart Control, $PM_{2.5}$, Natural Ventilation, Artificial Intelligence and Internet of Things (AIoT).

1. Introduction

Epidemiologic evidence has indicated a strong relationship between exposure to $PM_{2.5}$

(particulate matter with aerodynamic diameter less than 2.5 μm) and adverse health effects, including lung cancer [1], respiratory infections [2], stroke [2], chronic obstructive pulmonary disease (COPD) [2], cardiovascular disease [3], asthma [4], thereby substantially reducing life expectancy[5]. Indoor $PM_{2.5}$ originates from outdoor infiltration and indoor emission [6]. Outdoor $PM_{2.5}$ can enter indoor environments through windows under natural ventilation, through fans under mechanical ventilation, or through envelope cracks under infiltration [6–10]. There are also numerous indoor $PM_{2.5}$ sources, such as smoking [11–13], printing [14], cooking [15–17] and other activities [18-19]. Since most people spend 85-90% of their time in indoor environments [20], it is crucial to reduce indoor $PM_{2.5}$ concentrations and the associated health risks.

In naturally ventilated residential buildings, window behavior significantly influences the indoor $PM_{2.5}$ concentrations. Many investigations have addressed the characteristics of window behavior [e.g., 21-28]. For example, Fabi et al. [25] found that environmental factors such as temperature, humidity and noise are the most crucial driving forces for window opening/closing. Andersen et al. [26] reported that indoor CO_2 concentration and outdoor temperature were the most significant factors in window behavior, based on long-term measurements in 15 Danish dwellings. Shi et al [27-28] developed stochastic models for window behavior based on outdoor temperature, relative humidity, wind speed, and outdoor $PM_{2.5}$ concentration. Several studies have focused on the development of window control strategies [29-33]. For example, Stazi et al. [30] developed an adaptive window control algorithm to achieve a low indoor CO_2 level and good thermal comfort in a classroom. Dussault et al. [31] compared the performance of four smart window control strategies in reducing energy consumption while maintaining thermal and visual comfort. These studies have provided great insight into the characteristics and control of window behavior. In addition to thermal comfort, ventilation (indicated by CO_2), and energy consumption, the control of window behavior can also minimize indoor $PM_{2.5}$ concentrations, which has not been well studied.

In naturally ventilated buildings, the closing of windows tends to reduce the entry of $PM_{2.5}$ of outdoor origin [34], whereas opening windows increases the ventilation rate, which is beneficial for diluting $PM_{2.5}$ generated from indoor emissions [35]. However, when both indoor- and outdoor-originating $PM_{2.5}$ contribute significantly to the total indoor $PM_{2.5}$, it is challenging for occupants to determine whether to open or close windows. It should be noted that the optimal operation of windows would minimize indoor $PM_{2.5}$ concentrations. In real applications, occupants can easily obtain the real-time indoor and outdoor $PM_{2.5}$ concentrations with the use of low-cost light-scattering sensors (e.g. [36,37]). However, to the best of our knowledge, there is no existing window control approach that minimizes indoor $PM_{2.5}$ concentrations using only real-time indoor and outdoor $PM_{2.5}$ sensors in naturally ventilated buildings. In China, most of the residential buildings are naturally ventilated. Furthermore, less than 2% of people in China have air cleaners in their homes[38]. Therefore, such a window control approach can benefit a lot of people by reducing their exposure to indoor $PM_{2.5}$ and the associated health risks.

To achieve the mitigation of indoor $PM_{2.5}$ by window control, the traditional closed-loop, model predictive, and rule-based control approaches may be considered. For the closed-loop control, when the indoor $PM_{2.5}$ concentration is higher than the setpoint, the window actuator will act for reducing the concentration. However, since it is unknown to the controller whether the increase in indoor $PM_{2.5}$ concentration is attributed to indoor emission or outdoor infiltration, the controller cannot make the decision on opening or closing the window. Therefore, the closed-loop control may not be applicable in this application. For the model predictive control, it is essential to establish a model with accurate inputs to predict the indoor $PM_{2.5}$ concentration. However, it is challenging to monitor the key inputs such as indoor $PM_{2.5}$ emission rate and air exchange rate in real time. Therefore, the model predictive model may not be suitable for practical applications. For the rule-based control, a typical rule is based on the indoor-to-outdoor $PM_{2.5}$ concentration ratio (I/O ratio), which opens the window when the I/O ratio is larger than 1, while closes the window when the I/O ratio is lower than 1. However, indoor $PM_{2.5}$ emissions may still exist when the I/O ratio is smaller than 1 [8]. Therefore, it is also difficult for the rule-based window control to minimize indoor $PM_{2.5}$ concentrations. Note that window control is a sequential decision-making process. Reinforcement learning (RL), which is a powerful artificial intelligence algorithm, has achieved great success on sequential decision-making problems [39–40]. Therefore, it can be a more suitable approach to control window behavior for minimizing indoor $PM_{2.5}$ concentrations.

Multiple studies have applied reinforcement learning methods in the field of smart buildings [29, 32, 41–46]. For example, Han et al. [29] reported a reinforcement learning method that used Sarsa and Q-learning to improve the comfort of occupants in an office through control of window behavior. Dalamagkidis et al. [44] developed a reinforcement learning controller to improve overall building performance in terms of thermal comfort, indoor CO_2 concentration, and energy consumption. Heo et al. [45] proposed a deep Q-network-based approach to control the mechanical ventilation system of a subway station in real time in order to reduce the energy consumption while maintaining the PM_{10} level. However, in naturally ventilated buildings, few studies, if any, have used the reinforcement learning approach to control window behavior in real time to minimize the indoor $PM_{2.5}$ concentrations.

Therefore, this study aimed to develop a reinforcement learning approach to automatically control window behavior in real time in order to effectively reduce the indoor $PM_{2.5}$ concentrations in naturally ventilated buildings. The proposed method trains the window controller with a deep Q-network (DQN) in a specific naturally ventilated apartment for a one-month period. The trained controller can then be used to control the window behavior to reduce the indoor $PM_{2.5}$ concentrations in that apartment. The required input data for the controller are the real-time indoor and outdoor $PM_{2.5}$ concentrations with a 1-min resolution, which can be easily obtained with the use of low-cost sensors available on the market. This study first demonstrated the proposed reinforcement learning method in a virtual typical apartment with natural ventilation in Beijing. This investigation then applied the proposed approach in a real apartment in Tianjin to demonstrate its feasibility. The proposed reinforcement learning method can facilitate the development of smart window controllers for reducing indoor $PM_{2.5}$ concentrations, which will further support the rapid development of artificial intelligence with internet of things (AIoT) for smart and healthy buildings.

2. Methods

2.1 Control objective and inputs

This study focused on the mitigation of indoor $PM_{2.5}$ pollution in naturally ventilated buildings without indoor $PM_{2.5}$ filtration units. Outdoor $PM_{2.5}$ can enter a building via natural ventilation or infiltration, while indoor sources can also contribute to indoor $PM_{2.5}$. The objective of window behavior control is to minimize the total indoor $PM_{2.5}$ concentration. In practical applications, only the real-time indoor and outdoor $PM_{2.5}$ concentrations, $C_{in}(t)$ and $C_{out}(t)$, can be easily obtained from low-cost light-scattering sensors with careful calibration (e.g.[36]). Therefore, the control inputs in this study were the real-time indoor and outdoor $PM_{2.5}$ concentrations with a 1-min time resolution. The actuator was the window, and this study assumed that there were only two window-related actions, i.e., the window was fully closed or fully opened. With the sensors, cloud server, and window actuator, an AIoT system can be established. The indoor and outdoor $PM_{2.5}$ concentration data recorded by the light-scattering sensors will be sent to the cloud through Wi-Fi. With the input data, the control algorithm operated in the cloud will generate action signals, i.e. the window should be opened or closed. The action signals will then be sent from the cloud to the window actuator through Wi-Fi to open or close the window. The control algorithm was trained based on reinforcement learning, which will be introduced in the following section.

2.2 Control algorithm

2.2.1 Reinforcement learning framework

The basic framework of reinforcement learning consists of an agent and the environment. The agent is trained to act properly through interaction with the environment. After proper training, the agent serves as the control strategy to achieve the control objectives. Three crucial elements of reinforcement learning are the state s, the action a, and the reward function r, and they should be carefully designed. The state is the agent's observation of the environment. The action is the behavior of the agent in each time step. The reward function allows the agent to evaluate the effectiveness of its action. At a given time point t, the agent takes action a_t in accordance with the current observed state s_t and the policy $\pi(a_t|s_t)$. As the state changes from s_t to s_{t+1}, the agent receives the reward r_t and updates the policy with the aim of maximizing the summation of discounted future rewards, G_t, which can be expressed by:

$$G_t = r_{t+1} + \gamma r_{t+2} + \ldots + \gamma^{T-t-1} r_T \qquad (1)$$

where γ is the discount factor which balances the immediate and delayed rewards, and T is the final time of the whole process. Meanwhile, the value function, $Q_\pi(s,a)$, is used to estimate how beneficial it would be to choose a given action a in a given state s. Here, the benefit is defined in terms of the future reward that can be expected [39]:

$$Q_\pi(s,a) = E_\pi[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \,|s_t = s, a_t = a] \qquad (2)$$

where $E_\pi(\cdot)$ represents the expectation value of the expression inside the brackets under a given policy $\pi(a|s)$. The state, action, reward function, environment, and discount factors are modeled as a Markov decision process. To obtain the optimal policy for the process, this study utilized a deep Q-network (DQN) [47], which combines deep learning with reinforcement learning to train the agent through multiple iterations. This technique allows the agent to deal with the continuous state space and learn from the past mistakes.

As shown in Figure 1, the architecture of the DQN includes two neural networks and a memory in order to reduce the correlation of network input data and thus avoid overfitting. The behavior network with parameters w_e makes the decisions, while the target network with parameters w_t is used to optimize the behavior network. In each step of the training process, the state s_t serves as the input to the behavior network, and it chooses an action a_t based on the output $Q_e(s_t, a_t^*)$ with the ε-greedy strategy. Here, a_t^* represents all the actions that could be chosen. Under the ε-greedy strategy, the probability that an action is taken randomly is $1-\varepsilon$, while the probability that an action is taken with the maximum $Q_e(s_t, a_t^*)$ is ε. With the action a_t, the state of the environment changes from s_t to s_{t+1}, and it is returned to the agent together with the reward r_t. The s_{t+1} will then be the input to the behavior network in the next time step. In each step, the (s_t, a_t, r_t, s_{t+1}) is stored in the memory, with a total capacity of N. Every n time steps, a certain number of (s_k, a_k, r_k, s_{k+1}) $(max\{t-N+1, 0\} \leq k \leq t)$ are sampled for training of the agent. For each sampled (s_k, a_k, r_k, s_{k+1}), the s_k and s_{k+1} serve as the inputs to the behavior and target networks, respectively, and the outputs are the $Q_e(s_k, a_k)$ and the maximum of $Q_t(s_{k+1}, a_{k+1}^*)$, respectively. Both of these outputs are utilized to calculate the loss function L with the reward r_k:

$$L = (r_k + \gamma max Q_t(s_{k+1}, a_{k+1}^*) - Q_e(s_k, a_k))^2 \qquad (3)$$

The mini-batch stochastic gradient descent method is used for the loss function summation of the sampled (s_k, a_k, r_k, s_{k+1}) to update the parameters w_e in the behavior network. It should be noted that, every m time steps, the parameters of the behavior network w_e are duplicated to the target network for updating of the parameters w_t. The reinforcement learning framework introduced above was used in the development of the control algorithm in this study. The following section details the window control algorithm that was based on the reinforcement learning framework.

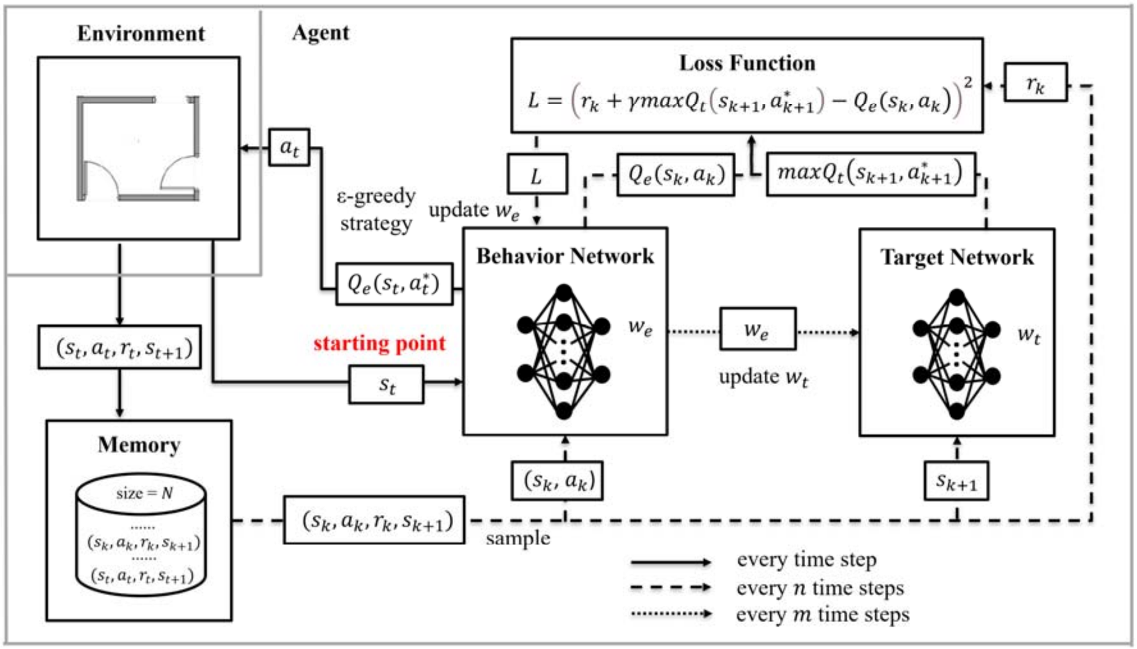

Figure 1**.** The deep Q-network architecture used in this study.

2.2.2 Reinforcement learning approach for window control

The reinforcement learning framework introduced above was applied to the control of window behavior in naturally ventilated buildings for reducing indoor $PM_{2.5}$ concentrations. In each time step, the state s_t consists of the indoor and outdoor $PM_{2.5}$ concentrations, which can be easily obtained by low-cost sensors with a time resolution of 1 min:

$$s_t = [C_{in}(t), C_{out}(t)] \qquad (4)$$

For the action a_t, the window behavior controller is the agent, and thus the decision of window state is:

$$a_t = 0 \text{ or } 1 \qquad (5)$$

where 1 represents the fully open window state, and 0 represents the fully closed window state. When the a_t changes from 0 to 1, the agent opens the window; when the a_t changes from 1 to 0, the agent closes the window. The reward function was designed to effectively reduce the indoor $PM_{2.5}$ concentration in the next time step, which is defined as:

$$r_t = -C_{in}(t + \Delta t) \qquad (6)$$

where Δt represents the time step and is equal to 1 min in this study.

The DQN network is a fully connected 2×8×2 network with an input layer, a hidden layer, and an output layer. For the input layer, since the state space is two-dimensional, the number of nodes was set as 2. For the output layer, since the number of elements in the one-dimensional action space is 2, the number of nodes was set as 2. For the hidden layer, with fewer nodes, the agent would learn less information of the environment. On the other hand, more hidden layers or nodes would result in a time-consuming training process and a higher risk of overfitting. This study tested several combinations and found that one hidden layer with 8 nodes achieved the best performance. The training of the DQN was performed in the given naturally ventilated apartment for a one-month period, which was found to be sufficiently long for obtaining satisfactory results. The well-trained DQN could then be used to control the window behavior in that apartment in order to reduce the indoor $PM_{2.5}$ concentrations.

3. Demonstration in a virtual typical apartment

3.1 Case setup

As a preliminary proof of concept, this study first applied the proposed reinforcement learning method to the living room of a virtual typical apartment in Beijing from January 1 to December 31, 2019. The inputs from the sensors were the time-resolved outdoor and indoor $PM_{2.5}$ concentrations. The hourly outdoor $PM_{2.5}$ concentrations recorded at the 35 weather stations in Beijing in 2019 were retrieved from the official air pollution monitoring website. The retrieved outdoor $PM_{2.5}$ concentrations were averaged and then interpolated into data with an interval of 1 min, serving as the input data from the virtual outdoor $PM_{2.5}$ sensor. For a given apartment, the indoor $PM_{2.5}$ concentrations would be influenced by the window behavior and also by the control algorithm. In real applications, such an influence would be directly reflected in the data measured by the actual indoor $PM_{2.5}$ sensor. However, in the virtual environment tested in this study, the indoor $PM_{2.5}$ concentrations were generated by the particle mass balance model [8]:

$$\frac{dC_{in}(t)}{dt} = \alpha P C_{out}(t) - \alpha C_{in}(t) - \frac{A}{V} v_d C_{in}(t) + \frac{\dot{S}(t)}{V} \qquad (7)$$

where t is the time, α is the air exchange rate, P is the penetration factor, A is the room surface area, V is the volume of the room, v_d is the particle deposition velocity, and $\dot{S}$ is the indoor particle emission rate. Particle resuspension was neglected in this study. It should be noted that Eq. (7) was used only for mimicking the environment. More importantly, the indoor $PM_{2.5}$ concentrations in the next time step, which would be influenced by the control algorithm, were calculated from Eq. (7). And these concentrations served as the input data "measured" by the virtual indoor $PM_{2.5}$ sensor.

Shi et al. [48-49] summarized the typical values of the building parameters in the model, α, P, A/V, and V, for naturally ventilated apartments in Beijing. According to the summarized data, the α when the windows were open and closed was set at 4.38 and 0.21 h^{-1}, respectively; the P when the windows were open and closed was set at 1 and 0.8, respectively; and the A/V and V were set at 1.63 m^{-1} and 60.34 m^3, respectively. The v_d was calculated according to the empirical equation proposed by Liu et al. [50]. The $\dot{S}$ was set separately in the training and testing, so that the robustness of the algorithm could be examined. Note that the P, α, and v_d can vary with indoor/outdoor temperature differential, wind speed, and wind direction. As a preliminary proof of concept, in this virtual environment case, the variations in these parameters were neglected for the sake of simplicity. As a further proof of concept, the real apartment case in Section 4 will further consider the variations in P, α, and v_d according to the real data. Furthermore, the room air was assumed to be well-mixed in this virtual apartment case so that the virtual indoor $PM_{2.5}$ sensor modeled by Eq. (7) could be valid. However, in real applications, the real indoor $PM_{2.5}$ sensor can be installed at any location of interest, which is not constrained to the well-mixed assumption.

3.2 Training

The DQN agent was trained from January 1 to January 29 of 2019. As shown in Table 1, the learning rate, α_{lr}, was set at 0.02 to allow the agent to learn at a moderate speed. The discounted rate, γ, was set at 0.9 to take future rewards into consideration. The ε-greedy rate, ε, was set at 0.999 to avoid taking actions randomly most of the time. The replay memory size and batch size were set at 20,000 and 64, respectively, to allow the agent to remember past mistakes while keeping the training time within a month. The agent learned every 10 time steps, in order to reduce the learning time. The target network parameters were updated every 500 time steps.

Table 1. Hyperparameters for DQN agent training in the virtual typical apartment

Hyperparameters	Value
Network structure	2×8×2
Learning rate (α_{lr})	0.02
Discounted rate (γ)	0.9
ε-greedy rate (ε)	0.999
Replay memory size (N)	20000

Batch size	64
Learning interval (n)	10 time steps
Target network update interval (m)	500 time steps

To mimic indoor $PM_{2.5}$ emissions in the virtual apartment, this study set the indoor activities and $PM_{2.5}$ emissions in January for the training of the DQN as listed in Table 2. It was assumed that the occupants cooked breakfast, lunch, dinner, and a midnight snack, smoked a cigarette, and used the printer twice each day. The $PM_{2.5}$ emission rates from smoking and printing were set according to data measured by Chen et al.[13] and Eggert et al. [51], respectively. The cooking activities were assumed to occur in the kitchen with a range hood. Since the range hood could not completely remove the $PM_{2.5}$ generated from cooking, some $PM_{2.5}$ would enter to the living room from the kitchen, especially when the kitchen door is open [52-53]. This was considered as an equivalent $PM_{2.5}$ emission in the living room due to cooking. The equivalent $PM_{2.5}$ emission rates in the living room due to cooking, $\dot{S}_{cooking,eqivalent}$, can be roughly estimated by:

$$\dot{S}_{cooking,eqivalent} = \beta(1-\eta)\dot{S}_{cooking} \qquad (8)$$

where η is the range hood efficiency, which was set at 58% [16], $\dot{S}_{cooking}$ is the original $PM_{2.5}$ emission rate measured by Chen et al.[16], and β is a coefficient which was roughly set at 0.6, as the field measurements by [53] found that the living room-to-kitchen $PM_{2.5}$ ratio ranged from 0.45 to 0.8 during cooking. Furthermore, for each $PM_{2.5}$ emission, the emission rate was randomly generated within the uncertainty range shown in Table 2.

Table 2. Indoor activities and $PM_{2.5}$ emissions in January 2019 that were set for the training of the DQN agent controller.

Time Period	Occupant behavior	$PM_{2.5}$ emission rate (μg/min)
0:30 – 0:50	Boiling	34.8 ± 29.9 [16]
8:00 – 8:20	Steaming	21.8 ± 15.0 [16]
11:30 – 11:50	Boiling	34.8 ± 29.9 [16]
14:00 – 14:07	Smoking	2250 ± 390 [13]
16:00 – 16:05	Printing	61 [51]
18:00 – 18:10	Deep frying	197.4 ± 64.3 [16]
20:00 – 20:05	Printing	61 [51]

The training lasted for the first 29 days of the year and included 41,760 time steps. In total, the DQN agent learned 4,176 times. The trained DQN was then used to control the window behavior to reduce indoor $PM_{2.5}$ concentrations for the rest of the year, i.e., from February to December.

3.3 Testing of the proposed control algorithm

The trained DQN agent was used to control the window behavior from February 1 to December 31 of 2019, in order to reduce the indoor $PM_{2.5}$ concentrations. This study assumed that the occupants stayed in the living room only in the daytime. Therefore, the testing of the proposed control algorithm was conducted for the time period between 6:00 and 24:00 each day. To mimic the environment and test the robustness of the proposal control algorithm, the indoor activities and $PM_{2.5}$ emissions from February to December, as listed in Table 3, were set differently from those in January. For each $PM_{2.5}$ emission, the emission rate was also randomly generated within the uncertainty range shown in Table 3. It should be noted that the DQN agent did not have information about the indoor $PM_{2.5}$ emissions. The only inputs to the control algorithm were the real outdoor $PM_{2.5}$ concentrations and the indoor $PM_{2.5}$ concentrations "measured" by the virtual sensors.

Table 3. Indoor activities and $PM_{2.5}$ emissions from February to December of 2019 that were set for the testing of the DQN agent controller.

Time period	Source type	$PM_{2.5}$ emission rate (μg/min)
7:30 – 7:50	Steaming	21.8 ± 15.0 [16]
11:30 – 11:40	Deep frying	197.4 ± 64.3 [16]
14:00 – 14:07	Smoking	2250 ± 390 [13]
16:00 – 16:05	Printing	61 [51]
18:00 – 18:20	Boiling	34.8 ± 29.9 [16]
22:00 – 22:04	Printing	61 [51]

For the sake of comparison, this study set a simple baseline control algorithm that was based on the I/O ratio. The algorithm opened the windows when the I/O ratio was greater than 1, while it closed the windows when the I/O ratio was less than 1. The rationale of this control strategy was that, when the I/O ratio is greater than 1, there must be indoor $PM_{2.5}$ emissions that will be diluted by opening the windows. Furthermore, controlling the window behavior according to the I/O ratio cutoff of 1 was straightforward and served as a baseline for comparison.

Figure 2 shows the indoor $PM_{2.5}$ concentrations from 6:00 to 24:00 on March 8 based on the proposed reinforcement learning algorithm and the baseline control algorithm. The outdoor $PM_{2.5}$ concentrations and the window behavior from both control algorithms are also shown in the figure. Noted that the $PM_{2.5}$ emission rates were the same for both algorithms to ensure a fair comparison. In general, the indoor $PM_{2.5}$ concentration from the proposed reinforcement learning algorithm was lower than that from the baseline control algorithm.

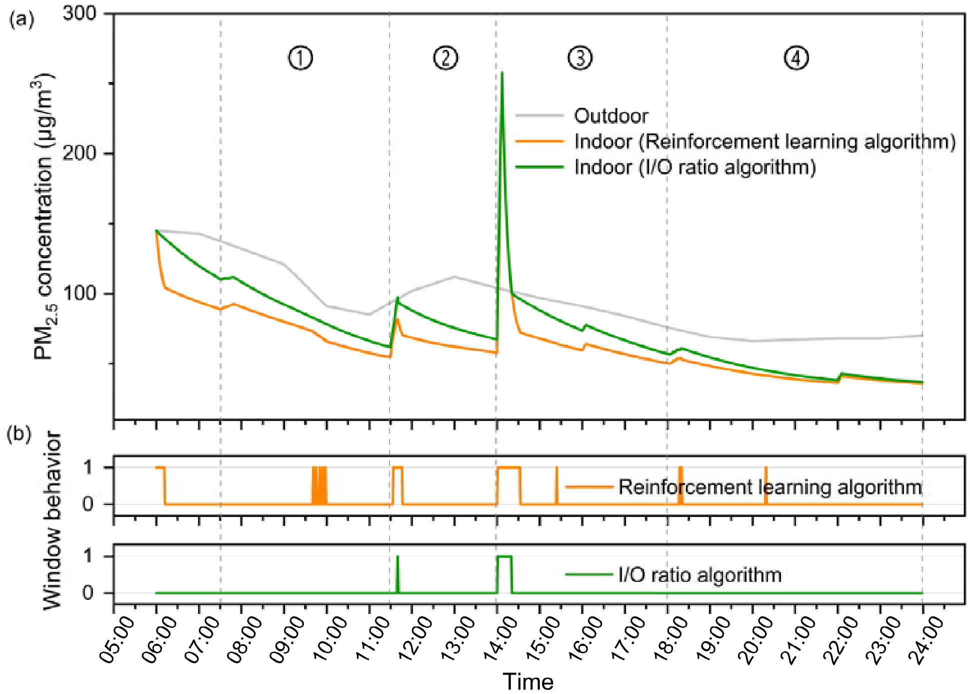

Figure 2. (a) Comparison of indoor $PM_{2.5}$ concentrations controlled by the proposed reinforcement learning and baseline I/O ratio algorithms, and (b) window behavior from both control algorithms in a virtual typical apartment in Beijing on March 8.

To facilitate the detailed analysis, the time period was divided into four periods. For period ③ from 14:00 to 18:00, the indoor $PM_{2.5}$ emission (smoking) started at 14:00. The reinforcement learning algorithm kept the windows open from 14:01 to 14:32, while the I/O ratio algorithm opened the windows only from 14:01 to 14:20. Although the I/O ratio was lower than 1 during the decay process, the smoking emission still contributed to the indoor $PM_{2.5}$ concentration through the process. The additional 12 min of open-window time from the reinforcement learning algorithm effectively reduced the indoor $PM_{2.5}$ concentrations in this time period. A similar phenomenon occurred in period ②.

For periods ① and ④, the reinforcement learning algorithm resulted in lower indoor $PM_{2.5}$ concentrations than the baseline algorithm, through a longer open-window time. However, in period ① between 9:40 and 10:00, although the reinforcement learning algorithm still led to lower indoor $PM_{2.5}$ concentrations than the baseline I/O ratio algorithm, the windows were opened and closed very frequently, mainly because of the rapid drop in outdoor $PM_{2.5}$ concentration. These frequent actions rendered the algorithms impractical for real applications, as will be discussed further in the next section.

Quantitatively, on March 8, 2019, when the reinforcement learning window control algorithm was used, the average indoor $PM_{2.5}$ concentration was 63.13 μg/m^3, while the concentration with the I/O ratio algorithm was 74.10 μg/m^3. Namely, the reinforcement learning algorithm reduced the concentration by 14.80% compared with the I/O ratio algorithm on that day. Moreover, for 99.9 % of the time on that day, the proposed reinforcement learning algorithm outperformed the baseline I/O ratio algorithm in reducing indoor $PM_{2.5}$ concentrations.

Figure 3 shows the average indoor $PM_{2.5}$ concentration from February to December when the reinforcement learning and baseline I/O ratio window control algorithms were used. The fraction of time that the reinforcement learning algorithm outperformed the I/O ratio algorithm is also shown as the rectangles. The results show that the reinforcement learning algorithm effectively lowered the indoor $PM_{2.5}$ concentrations compared with the baseline throughout the year. From February to December, the average indoor $PM_{2.5}$ concentration controlled by the reinforcement learning algorithm was 25.85 μg/m^3, which was 12.80% lower than that controlled by the I/O ratio algorithm (29.64 μg/m^3). Furthermore, in all the months, the indoor $PM_{2.5}$ concentrations with the reinforcement learning algorithm were lower than those with the I/O ratio algorithm for over 98% of the time. Therefore, this virtual typical naturally ventilated apartment case demonstrated the effectiveness of the developed reinforcement learning window behaviour algorithm in reducing indoor $PM_{2.5}$ concentrations. To further demonstrate the feasibility of the proposed reinforcement learning window control approach, this study tested the algorithm in an actual apartment.

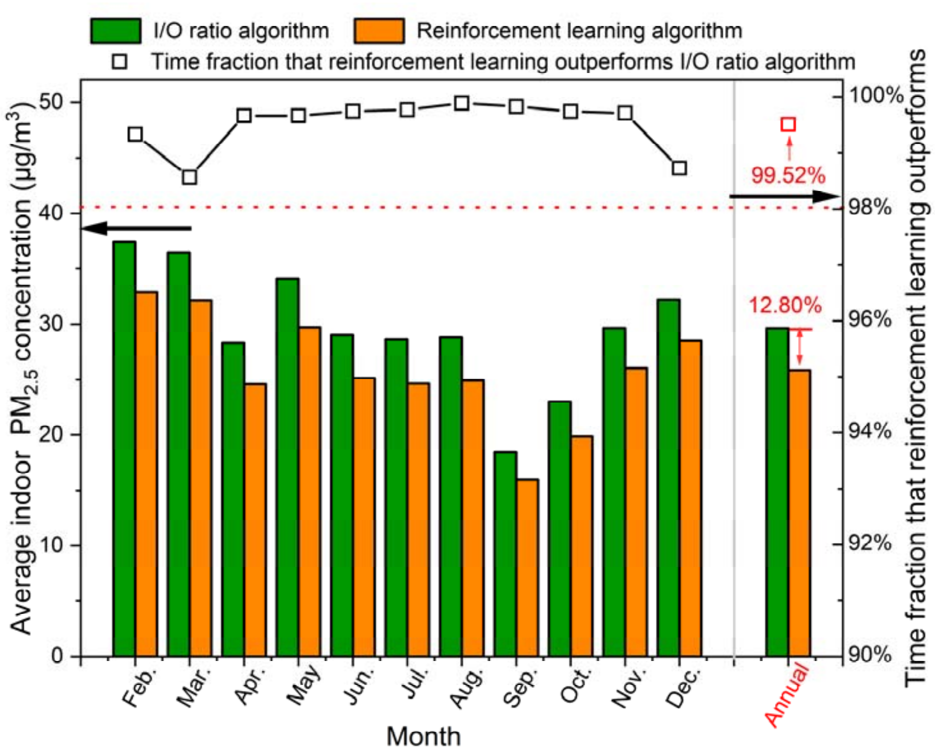

Figure 3. Comparison of the monthly average indoor $PM_{2.5}$ concentration controlled by the reinforcement learning and baseline I/O ratio algorithms (left vertical axis); and the time

fraction that the reinforcement learning algorithm outperformed the baseline I/O ratio algorithm in reducing indoor $PM_{2.5}$ concentration (right vertical axis) in the virtual apartment from February to December of 2019.

4. Demonstration in a real apartment

4.1 Case setup

This study further demonstrated the proposed approach with the use of data measured in the living room of a real apartment in Tianjin from February 1 to March 31 of 2017. The volume of the apartment, V, was 61.22 m^3, and the surface area to volume ratio, A/V, was 1.59 m^{-1}. The inputs to the algorithm were again the time-resolved outdoor and indoor $PM_{2.5}$ concentrations. The outdoor $PM_{2.5}$ concentrations were measured using a light-scattering sensor with a time resolution of 1 min, which had been calibrated by a gravimetric $PM_{2.5}$ instrument [36]. It should be noted that, if the proposed algorithm had been used to control the window behavior during that time period, the indoor $PM_{2.5}$ concentrations would have been altered by the behavior. Therefore, this study used the particle mass balance model introduced in Section 3.1 as a virtual sensor to "measure" the indoor $PM_{2.5}$ concentrations. The main difference from the virtual apartment case was that the mass balance model with the specific parameters of this case mimicked the dynamics of the actual environment.

Based on our previous work [6], the time periods without indoor $PM_{2.5}$ emissions in this apartment were identified. The parameters α, P, and v_d were then estimated by a grid search method, which minimized the root-mean-square error (RMSE) when comparing the model-generated and measured indoor $PM_{2.5}$ concentrations. According to the results, in February and March, the α when the windows were open and closed was 0.96 ± 0.08 and 0.22 ± 0.06 h^{-1}, respectively; the P when the windows were open and closed was 0.88 ± 0.07 and 0.72 ± 0.06, respectively; and the v_d when the windows were open and closed was 0.47 ± 0.11 and 0.19 ± 0.07 $m \cdot h^{-1}$, respectively. When this method is used, the specific parameters, α, P, and v_d, reflect the dynamics of the actual environment. The real time-resolved indoor $PM_{2.5}$ emission rates, $\dot{S}$, were then calculated by the particle mass balance equation with the building parameters (α, P, v_d, A/V, and V). Figure 4 compares the indoor $PM_{2.5}$ concentrations on March 14 from the actual measurements and the virtual indoor $PM_{2.5}$ sensor (i.e., the particle mass balance model with the apartment-specific parameters obtained above). The RMSE and the relative error between the actual measurements and the virtual indoor $PM_{2.5}$ sensor for the two months was 12.17 $\mu g/m^3$ and 6.2%, respectively. Therefore, the virtual indoor $PM_{2.5}$ sensor provided data that was reasonably close to the actual measurements. Since the virtual indoor $PM_{2.5}$ sensor can effectively "measure" the indoor $PM_{2.5}$ concentrations, it was used in both training and testing. Note that the fitted parameters of α, P, v_d, and $\dot{S}$ were only used in the virtual indoor $PM_{2.5}$ sensor. The reinforcement learning algorithm does not train or use these parameters.

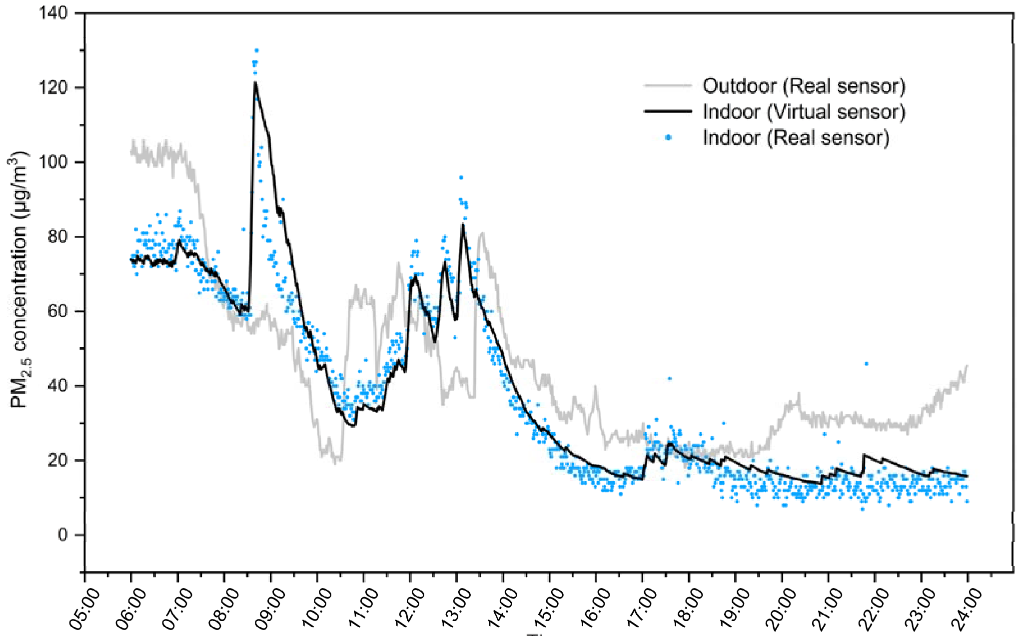

Figure 4. Comparison of indoor $PM_{2.5}$ concentrations from the actual measurements and the virtual indoor $PM_{2.5}$ sensor (i.e., the particle mass balance model with the apartment-specific parameters obtained above) in the real apartment in Tianjin on March 14.

This case study can be regarded as computer experiments. To demonstrate its superiority, the proposed reinforcement learning algorithm should be compared with the I/O ratio algorithm and the real window behavior. The computer experiments can ensure that the environment, which is represented by the particle mass balance model with the fitted parameters obtained above, was exactly the same for the window control algorithms to be compared. However, in real experiments, it would be challenging to have identical apartments for comparing different algorithms. Furthermore, even with the same brand and careful calibration, there would be deviations among the real $PM_{2.5}$ sensors. Therefore, compared with real experiments, the computer experiments can avoid the influence of inconsistency in the apartments and sensor uncertainties, which is beneficial for the proof of concept.

4.2 Training

The DQN agent was trained from February 1 to February 23 of 2017. The hyperparameters were the same as those in Table 1, except for the learning rate α_{lr}, which was set at 0.001. This value was smaller than that in the virtual case because the real apartment case was much more complex than the virtual case and required a slower learning rate to ensure that the main features could be fully learned. The training lasted for 23 days, including 33,120 time steps, and the DQN learned 3,312 times in total. The trained DQN agent was then used to control the window behavior for reducing indoor $PM_{2.5}$ concentrations from March 7 to March 31 of 2017.

4.3 Testing of the proposed control algorithm

To test the proposed control algorithm, this study used the trained DQN agent to control the window behavior from March 7 to March 31 of 2017 in order to lower the indoor $PM_{2.5}$ concentrations. Due to bad weather and unstable power supply, the outdoor $PM_{2.5}$ concentration data were missing in several days during the monitoring. There were in total 20 days with complete input data for the testing. Again, the only inputs to the control algorithm were the outdoor $PM_{2.5}$ concentrations from the outdoor $PM_{2.5}$ sensor and the indoor $PM_{2.5}$ concentrations "measured" by the virtual indoor $PM_{2.5}$ sensor. For the sake of comparison, this study again selected the I/O ratio algorithm as the baseline. Figure 5 shows the indoor $PM_{2.5}$ concentrations from 6:00 to 24:00 on March 14 based on the reinforcement learning window-control algorithm, the baseline I/O ratio window-control algorithm, and the real window behaviors. Note that all the results in Figure 5 were all based on the virtual indoor $PM_{2.5}$ sensor, i.e. the mass balance model with the same fitted parameters. Therefore, the comparison of the reinforcement learning algorithm, I/O ratio algorithm, and real window behavior was based on exactly the same conditions, i.e. the same $PM_{2.5}$ emission, air exchange rate, penetration factor, and deposition velocity at any time point. This comparison method can exclude the influence of the uncertainty of model prediction. The outdoor $PM_{2.5}$ concentrations and window behaviors for all three methods are displayed in Figure 5. In general, the indoor $PM_{2.5}$ concentrations controlled by the proposed reinforcement learning algorithm were lower than those controlled by the baseline I/O ratio algorithm and the real window behaviors. This difference can be attributed to the longer open-window time when there were indoor emissions, or the influence of the emissions on indoor $PM_{2.5}$ concentration may have remained strong. Essentially, the reinforcement learning algorithm better captured the characteristics of the influence of indoor $PM_{2.5}$ emissions and outdoor $PM_{2.5}$ concentration variations on the indoor $PM_{2.5}$ level. However, it should be noted that frequent window opening and closing actions occurred in some time periods even with the reinforcement learning algorithm, which would limit its practical applications.

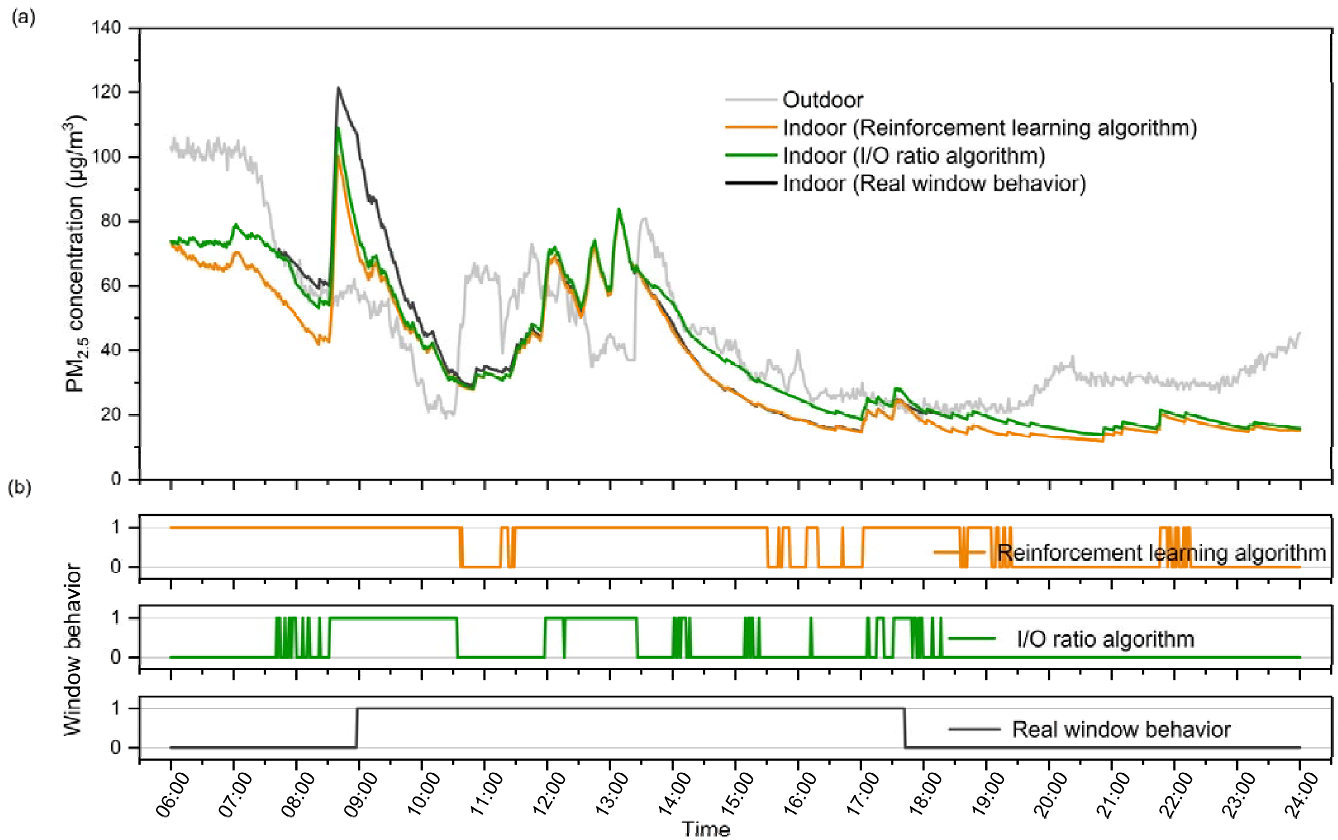

Figure 5. (a) Comparison of indoor $PM_{2.5}$ concentrations controlled by the proposed reinforcement learning algorithm, the baseline I/O ratio algorithm, and real window behavior, and (b) window behavior from the three methods in the real apartment on March 14.

Quantitatively, on March 14, 2017, the average indoor $PM_{2.5}$ concentration with the reinforcement learning window control algorithm was 34.94 μg/m^3, while that with the I/O ratio algorithm was 38.97 μg /m^3, and that with the real window behavior was 39.03 μg/m^3. Namely, the reinforcement learning algorithm reduced the indoor $PM_{2.5}$ concentration by 10.80% compared with the I/O ratio algorithm, and by 10.64% compared with the real window behavior. In addition, the reinforcement learning algorithm outperformed the I/O ratio algorithm for 96.11% of the time, and the real window behavior for 89.82% of the time, in reducing the indoor $PM_{2.5}$ concentrations.

Figure 6 compares the daily average indoor $PM_{2.5}$ concentrations in March 2017 under the reinforcement learning window control algorithm, baseline I/O ratio window control algorithm, and real window behavior. The results demonstrated that the reinforcement learning algorithm mitigated the indoor $PM_{2.5}$ pollution more effectively than the I/O ratio algorithm and real window behavior throughout the month. When the reinforcement learning window control algorithm was used, the total average indoor $PM_{2.5}$ concentration was 47.77 μg/m^3, which was 9.11% lower than that with the I/O ratio algorithm (52.56 μg/m^3), and 7.40% lower than that with the real window behavior (51.59 μg/m^3). Furthermore, in the month as a whole, the reinforcement learning algorithm exhibited better performance more than 85% of the time in reducing the indoor $PM_{2.5}$ concentration, compared with the I/O ratio algorithm and real window behavior. According to the global concentration-mortality relationships for ambient $PM_{2.5}$ based on the Global Burden of Disease studies, the theoretical minimum-risk concentration ranges from 5.8 to 8.0 μg/m^3 [2,54]. In this case, the indoor $PM_{2.5}$ concentration was higher than the theoretical minimum-risk concentration for 93% of the time. Therefore, almost all the reduction in indoor $PM_{2.5}$ from using the proposed window control algorithm can reduce the associated health risks. As a rough estimation, the decrease of indoor $PM_{2.5}$ concentration from 52.56 μg/m^3 to 47.77 μg/m^3 can lower the population total attributable mortality from 97.1 in 100,000 per year to 91.8 in 100,000 per year [54-55] and increase the life expectancy by about 0.27 years [56]. Considering that the proposed window control did not require additional energy or consumables such as filters, the improvement in the indoor $PM_{2.5}$ control was satisfactory. In general, the real naturally ventilated apartment case demonstrated the effectiveness and feasibility of the proposed reinforcement learning window control algorithm in reducing indoor $PM_{2.5}$ concentration.

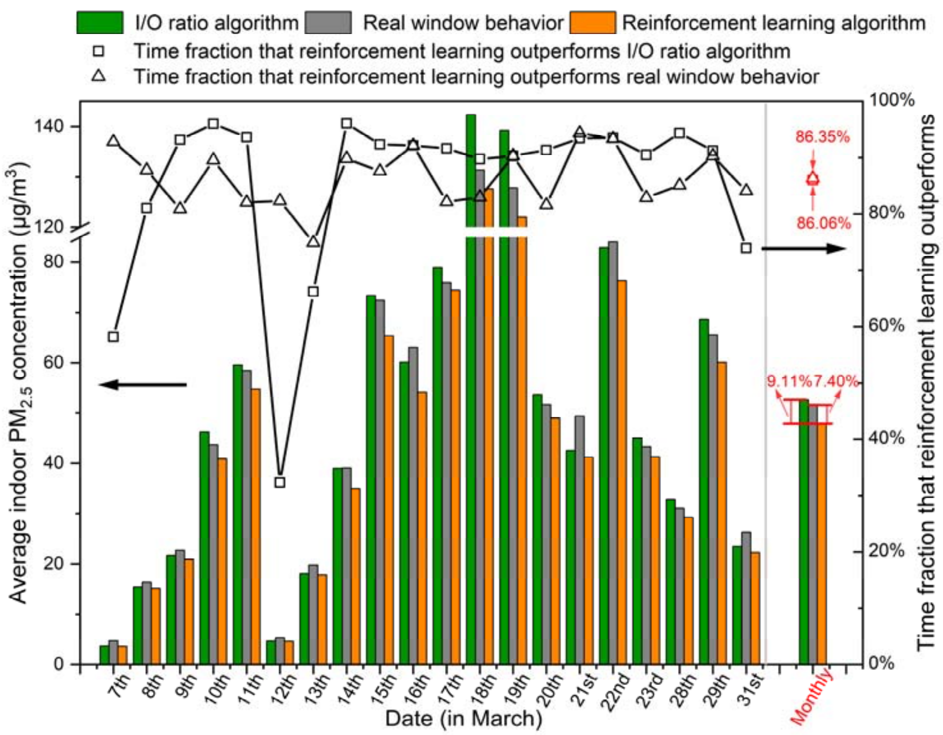

Figure 6. Comparison of the daily average indoor $PM_{2.5}$ concentration under the reinforcement learning window control algorithm, the baseline I/O ratio window control algorithm, and the real window behavior (left vertical axis); and the time fraction that the reinforcement learning algorithm outperformed the baseline I/O ratio algorithm and real window behaviors in reducing indoor $PM_{2.5}$ concentration (right vertical axis) in the real apartment in March 2017.

4.4 Control of window behavior frequency

In both cases, the trained reinforcement learning controller took action frequently in some time periods. Such frequent window opening and closing would be impractical and challenging because it affects the experience of the residents with frequent operation noise and reduces the lifetime of the window actuators. Also, the time for a single window opening/closing action for the actuators available on the market is normally more than 15 seconds, based on the product specifications provided on the largest e-commerce platforms in China. Using an action interval of 1 min would be risky, and the window actions would be too frequent. To make the proposed reinforcement learning algorithm more practical, the minimum window action interval was set at 5 min for the real apartment case.

Figure 7 shows the indoor $PM_{2.5}$ concentrations from 6:00 to 24:00 on March 14 based on the reinforcement learning window-control algorithm with 1-min and 5-min minimum action interval, and I/O ratio algorithm. By setting the minimum action interval at 5 min, the frequent window actions at around 10:40, 11:20, 15:40, 19:00 and 22:00 were successfully avoided without increasing the indoor $PM_{2.5}$ concentrations when compared with the algorithm with 1-min minimum action interval.

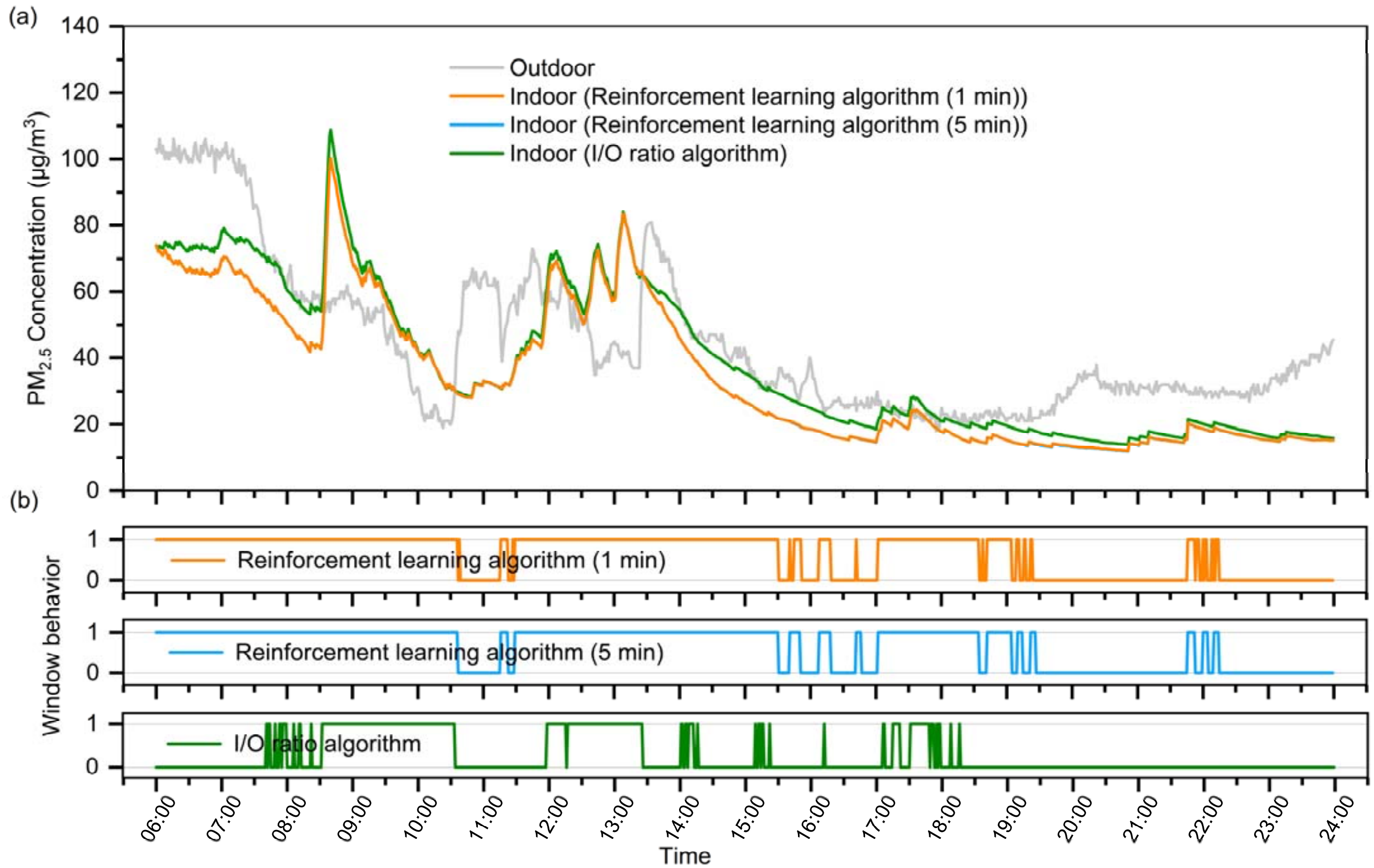

Figure 7. (a) Comparison of indoor $PM_{2.5}$ concentrations controlled by the proposed reinforcement learning algorithm with 1-min and 5-min minimum action interval, and the I/O ratio algorithm, and (b) window behavior from the three methods in the real apartment on March 14.

Figure 8 compares the daily average indoor $PM_{2.5}$ concentration in March 2017 under the reinforcement learning window control algorithm with a 5-min minimum action interval, and the concentration under the original algorithm. The average indoor $PM_{2.5}$ concentration when using the original reinforcement learning algorithm was 47.77 μg/m^3, while that with the 5-min minimum action interval was 47.81 μg/m^3. In general, the additional constraint of a 5-min minimum action interval did not compromise the performance of the reinforcement learning window control algorithm in reducing indoor $PM_{2.5}$ concentrations. Therefore, the improved reinforcement learning window control algorithm can be used in practical applications.

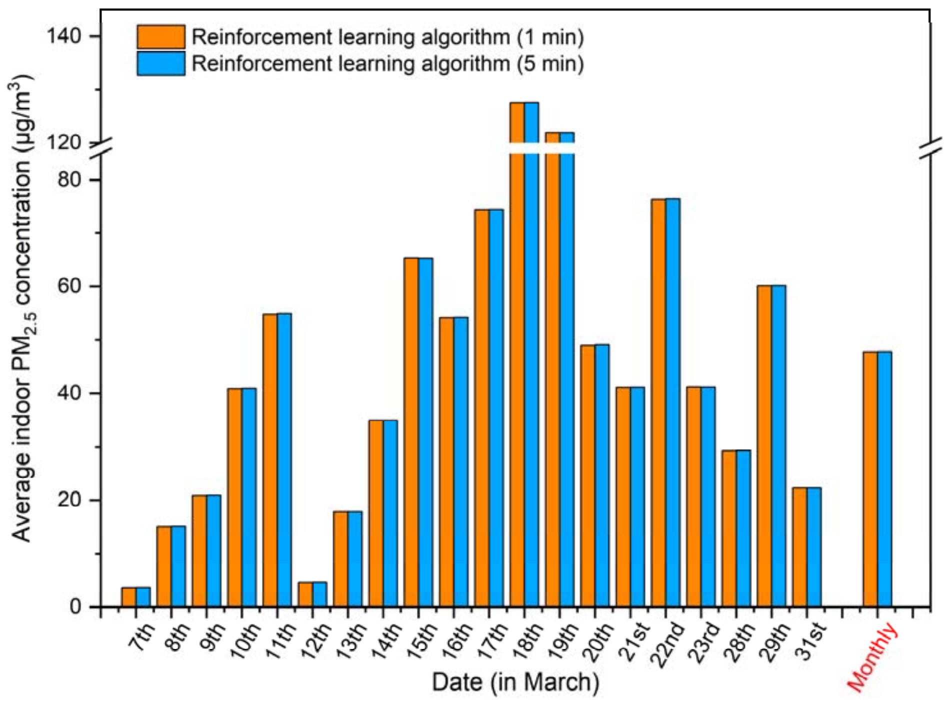

Figure 8. Comparison of the daily average indoor $PM_{2.5}$ concentration under the reinforcement learning window control algorithm with 1-min and 5-min minimum action interval in the real apartment in March 2017.

Another concern would be whether the window control with 5-min minimum action interval can effectively responds to a short but strong indoor $PM_{2.5}$ emission. For example, assume that a window closing action occurs at a certain time point for whatever reason. If the occupant smokes a cigarette 2 min later, although the algorithm would detect the need of opening the window, the window will not be opened immediately due to the 5-min minimum action interval constraint. Therefore, there will be a 3-min gap between the smoking and window opening action, which may influence the effectiveness of the control algorithm. Based on the data in this real apartment, such scenarios did not occur. However, it is still worthwhile to explore such extreme cases to further test the robustness of the algorithm. This study assumed that a 7-min smoking with the $PM_{2.5}$ emission rate of 2250 μg/(m^3·min) [13] occurred at 21:55 on March 14, 2 min after the previous action of closing the window, which corresponded to the extreme scenario mentioned above. Figure 9 shows the indoor $PM_{2.5}$ concentrations from 21:50 to 22:10 on March 14 when using reinforcement learning algorithm with 1-min and 5-min minimum action interval and the I/O ratio algorithm. The reinforcement learning algorithm with 5-min minimum action interval opened the window at 21:58, 2 min later than the other two methods. With this 5-min constraint, the peak indoor $PM_{2.5}$ concentration was slightly higher than that with 1-min minimum action interval, and the difference was only 2.0 μg/m^3. For the whole period from 21:50 to 22:10, the average indoor $PM_{2.5}$ concentrations when using the reinforcement learning algorithm with 5-min minimum action interval was 145.76 μg/m^3, which was close to that without the constraint (144.56 μg/m^3).

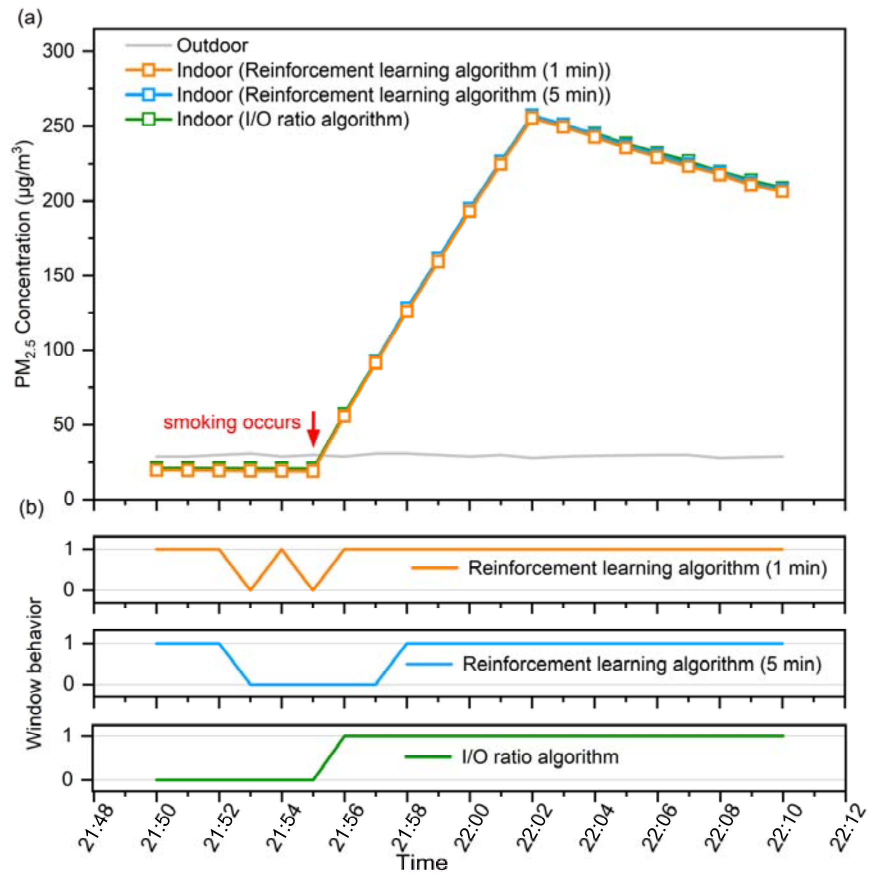

Figure 9. (a) Comparison of indoor $PM_{2.5}$ concentrations controlled by the proposed reinforcement learning algorithm with 1-min and 5-min minimum action interval and the I/O ratio algorithm, and (b) window behavior from the three methods in the real apartment from 21:50 to 22:10 on March 14 (a 7-min smoking occurred at 21:55, 2 min after the previous action of closing the window).

5. Limitations and prospects

There are several limitations to the proposed reinforcement learning window control algorithm, which merits further investigation. First, the proposed algorithm should be implemented more extensively in real apartments and tested with real online control for identification of potential issues in practical applications. Second, in real applications, the window behavior is not limited to fully open or fully closed. Window opening with different opening angles or opening areas may result in different air exchange rates for removing indoor $PM_{2.5}$ from indoor emissions. To take this factor into account, methods such as the deep deterministic policy gradient (DDPG) algorithm could be used. Finally, filtration devices such as air cleaners are widely used to mitigate indoor $PM_{2.5}$ pollution. The reinforcement learning control algorithm should be further developed to include the control of these devices, in order to minimize indoor $PM_{2.5}$ pollution, also minimize the energy consumption of the filtration devices, and extend the lifetime of the devices.

In general, people's perception is more sensitive to temperature and ventilation than $PM_{2.5}$. Therefore, even with a high indoor $PM_{2.5}$ concentration, the occupants might not response to it by opening/closing the window, although the exposure may lead to significant health risks. That is one of the major reasons why an automated control for window is needed to protect people from what they do not perceive but harmful. There are two possible ways to further consider the human-window interaction in the system. A simple way is to design the system with both manual and automated modes. If the occupant wants to take over the control based on his/her perception, the manual mode can be switched on. A better way is to further train the control agent to learn the probability that the occupant will open or close the windows based on temperature, relative humidity, wind speed, etc. If the probability is high, say over 90%, the human perception-based window action will overwrite the action for indoor $PM_{2.5}$ control. Such further development for the window control algorithm would facilitate practical applications.

6. Conclusions

This study proposed a reinforcement learning approach to effectively reduce indoor $PM_{2.5}$ concentrations in any naturally ventilated residential building through automatic control of window behavior. The proposed window control approach can benefit the occupants living in naturally ventilated buildings without air cleaners by reducing their exposure to indoor $PM_{2.5}$ and the associated health risks. The proposed algorithm trained the window controller with a DQN in a specific naturally ventilated apartment in a one-month period. The trained controller can now be used to control window behavior to reduce indoor $PM_{2.5}$ concentrations. The required input data for the controller are the real-time indoor and outdoor $PM_{2.5}$ concentrations with a 1-min resolution, which can easily be obtained with low-cost sensors available on the market. A series of simulations were conducted in a virtual typical apartment in Beijing and a real apartment in Tianjin. Within the scope of this research, the following conclusions can be drawn.

1. The proposed reinforcement learning window control algorithm effectively reduced the indoor $PM_{2.5}$ concentrations in the virtual typical apartment. Compared with the baseline I/O ratio algorithm, the proposed algorithm reduced the average indoor $PM_{2.5}$ concentration by 12.80% in the course of a year.
2. The proposed reinforcement learning window control algorithm reduced the indoor

$PM_{2.5}$ concentrations in the real apartment by 9.11% when compared with the I/O ratio algorithm, and by 7.40% when compared with real window behavior.

3. Adding a 5-min minimum action interval constraint did not compromise the performance of the reinforcement learning window control algorithm in reducing indoor $PM_{2.5}$ concentrations.

Acknowledgement

This work was partially supported by the Early Career Scheme and General Research Fund of Research Grants Council of Hong Kong SAR, China (Grant No. 24208518 and 14204520).

References

[1] Y. Guo, H. Zeng, R. Zheng, S. Li, A.G. Barnett, S. Zhang, X. Zou, R. Huxley, W. Chen, G. Williams, The association between lung cancer incidence and ambient air pollution in China: A spatiotemporal analysis, Environ. Res. 144 (2016) 60–65. https://doi.org/10.1016/J.ENVRES.2015.11.004.

[2] S.S. Lim, T. Vos, A.D. Flaxman, G. Danaei, K. Shibuya, H. Adair-Rohani, M. Amann, H.R. Anderson, K.G. Andrews, M. Aryee, C. Atkinson, L.J. Bacchus, A.N. Bahalim, K. Balakrishnan, J. Balmes, S. Barker-Collo, A. Baxter, M.L. Bell, J.D. Blore, F. Blyth, C. Bonner, G. Borges, R. Bourne, M. Boussinesq, M. Brauer, P. Brooks, N.G. Bruce, B. Brunekreef, C. Bryan-Hancock, C. Bucello, R. Buchbinder, F. Bull, R.T. Burnett, T.E. Byers, B. Calabria, J. Carapetis, E. Carnahan, Z. Chafe, F. Charlson, H. Chen, J.S. Chen, A.T.A. Cheng, J.C. Child, A. Cohen, K.E. Colson, B.C. Cowie, S. Darby, S. Darling, A. Davis, L. Degenhardt, F. Dentener, D.C. Des Jarlais, K. Devries, M. Dherani, E.L. Ding, E.R. Dorsey, T. Driscoll, K. Edmond, S.E. Ali, R.E. Engell, P.J. Erwin, S. Fahimi, G. Falder, F. Farzadfar, A. Ferrari, M.M. Finucane, S. Flaxman, F.G.R. Fowkes, G. Freedman, M.K. Freeman, E. Gakidou, S. Ghosh, E. Giovannucci, G. Gmel, K. Graham, R. Grainger, B. Grant, D. Gunnell, H.R. Gutierrez, W. Hall, H.W. Hoek, A. Hogan, H.D. Hosgood, D. Hoy, H. Hu, B.J. Hubbell, S.J. Hutchings, S.E. Ibeanusi, G.L. Jacklyn, R. Jasrasaria, J.B. Jonas, H. Kan, J.A. Kanis, N. Kassebaum, N. Kawakami, Y.H. Khang, S. Khatibzadeh, J.P. Khoo, C. Kok, F. Laden, R. Lalloo, Q. Lan, T. Lathlean, J.L. Leasher, J. Leigh, Y. Li, J.K. Lin, S.E. Lipshultz, S. London, R. Lozano, Y. Lu, J. Mak, R. Malekzadeh, L. Mallinger, W. Marcenes, L. March, R. Marks, R. Martin, P. McGale, J. McGrath, S. Mehta, G.A. Mensah, T.R. Merriman, R. Micha, C. Michaud, V. Mishra, K.M. Hanafiah, A.A. Mokdad, L. Morawska, D. Mozaffarian, T. Murphy, M. Naghavi, B. Neal, P.K. Nelson, J.M. Nolla, R. Norman, C. Olives, S.B. Omer, J. Orchard, R. Osborne, B. Ostro, A. Page, K.D. Pandey, C.D.H. Parry, E. Passmore, J. Patra, N. Pearce, P.M. Pelizzari, M. Petzold, M.R. Phillips, D. Pope, C.A. Pope, J. Powles, M. Rao, H. Razavi, E.A. Rehfuess, J.T. Rehm, B. Ritz, F.P. Rivara, T. Roberts, C. Robinson, J.A. Rodriguez-Portales, I. Romieu, R. Room, L.C. Rosenfeld, A. Roy, L. Rushton, J.A. Salomon, U. Sampson, L. Sanchez-Riera, E. Sanman, A. Sapkota, S. Seedat, P. Shi, K.

Shield, R. Shivakoti, G.M. Singh, D.A. Sleet, E. Smith, K.R. Smith, N.J.C. Stapelberg, K. Steenland, H. Stöckl, L.J. Stovner, K. Straif, L. Straney, G.D. Thurston, J.H. Tran, R. Van Dingenen, A. Van Donkelaar, J.L. Veerman, L. Vijayakumar, R. Weintraub, M.M. Weissman, R.A. White, H. Whiteford, S.T. Wiersma, J.D. Wilkinson, H.C. Williams, W. Williams, N. Wilson, A.D. Woolf, P. Yip, J.M. Zielinski, A.D. Lopez, C.J.L. Murray, M. Ezzati, A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010, Lancet. 380 (2012) 2224–2260. https://doi.org/10.1016/S0140-6736(12)61766-8.

[3] F. Dominici, R.D. Peng, M.L. Bell, L. Pham, A. McDermott, S.L. Zeger, J.M. Samet, Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases, J. Am. Med. Assoc. 295 (2006) 1127–1134. https://doi.org/10.1001/jama.295.10.1127.

[4] M. Brauer, G. Hoek, H.A. Smit, J.C. de Jongste, J. Gerritsen, D.S. Postma, M. Kerkhof, B. Brunekreef, Air pollution and development of asthma, allergy and infections in a birth cohort, Eur. Respir. J. 29 (2007) 879–888. https://doi.org/10.1183/09031936.00083406.

[5] C.A.P. III, M. Ezzati, D.W. Dockery, Fine particulate air pollution and life expectancies in the United States: The role of influential observations, J. Air Waste Manage. Assoc. 63 (2013) 129–132. https://doi.org/10.1080/10962247.2013.760353.

[6] T. Xia, C. Chen, Differentiating between indoor exposure to $PM_{2.5}$ of indoor and outdoor origin using time-resolved monitoring data, Build. Environ. (2019). https://doi.org/10.1016/j.buildenv.2018.10.046.

[7] C. Chen, B. Zhao, C.J. Weschler, Indoor exposure to "Outdoor PM_{10}": Assessing its influence on the relationship between PM_{10} and short-term mortality in U.S. Cities, Epidemiology. 23 (2012) 870–878. https://doi.org/10.1097/EDE.0b013e31826b800e.

[8] C. Chen, B. Zhao, Review of relationship between indoor and outdoor particles: I/O ratio, infiltration factor and penetration factor, Atmos. Environ. (2011). https://doi.org/10.1016/j.atmosenv.2010.09.048.

[9] T. Xia, Y. Bian, S. Shi, L. Zhang, C. Chen, Influence of nanofiber window screens on indoor $PM_{2.5}$ of outdoor origin and ventilation rate: An experimental and modeling study, Build. Simul. 13 (2020) 873–886. https://doi.org/10.1007/s12273-020-0622-5.

[10] Y.H. Cheng, Measuring indoor particulate matter concentrations and size distributions at different time periods to identify potential sources in an office building in Taipei City, Build. Environ. 123 (2017) 446–457. https://doi.org/10.1016/j.buildenv.2017.07.025.

[11] C. He, L. Morawska, J. Hitchins, D. Gilbert, Contribution from indoor sources to particle number and mass concentrations in residential houses, Atmos. Environ. 38 (2004) 3405–3415. https://doi.org/10.1016/j.atmosenv.2004.03.027.

[12] E. Diaz Lozano Patino, J.A. Siegel, Indoor environmental quality in social housing: A literature review, Build. Environ. 131 (2018) 231–241.

https://doi.org/10.1016/j.buildenv.2018.01.013.

[13] C. Chen, Y. Zhao, B. Zhao, Emission rates of ultrafine and fine particles generated from human smoking of Chinese cigarettes, Atmos. Environ. (2018). https://doi.org/10.1016/j.atmosenv.2018.09.014.

[14] C. He, L. Morawska, L. Taplin, Particle emission characteristics of office printers, Environ. Sci. Technol. 41 (2007) 6039–6045. https://doi.org/10.1021/es063049z.

[15] Y. Zhao, B. Zhao, Reducing human exposure to $PM_{2.5}$ generated while cooking typical Chinese cuisine, Build. Environ. 168 (2020) 106522. https://doi.org/10.1016/j.buildenv.2019.106522.

[16] C. Chen, Y. Zhao, B. Zhao, Emission rates of multiple air pollutants generated from Chinese residential cooking, Environ. Sci. Technol. 52 (2018) 1081–1087. https://doi.org/10.1021/acs.est.7b05600.

[17] Y. Zhao, B. Zhao, Emissions of air pollutants from Chinese cooking: A literature review, Build. Simul. 11 (2018) 977–995. https://doi.org/10.1007/s12273-018-0456-6.

[18] A.C.K. Lai, Y. Tian, J.Y.L. Tsoi, A.R. Ferro, Experimental study of the effect of shoes on particle resuspension from indoor flooring materials, Build. Environ. 118 (2017) 251–258. https://doi.org/10.1016/j.buildenv.2017.02.024.

[19] K.C. Cheng, D. Zheng, A.O. Tetteh, H.K. Park, K.C. Nadeau, L.M. Hildemann, Personal exposure to airborne particulate matter due to residential dryer lint cleaning, Build. Environ. 98 (2016) 145–149. https://doi.org/10.1016/j.buildenv.2016.01.008.

[20] N.E. Klepeis, W.C. Nelson, W.R. Ott, J.P. Robinson, A.M. Tsang, P. Switzer, J. V. Behar, S.C. Hern, W.H. Engelmann, The National Human Activity Pattern Survey (NHAPS): A resource for assessing exposure to environmental pollutants, J. Expo. Anal. Environ. Epidemiol. 11 (2001) 231–252. https://doi.org/10.1038/sj.jea.7500165.

[21] M. Schweiker, F. Haldi, M. Shukuya, D. Robinson, Verification of stochastic models of window opening behaviour for residential buildings, J. Build. Perform. Simul. 5 (2012) 55–74. https://doi.org/10.1080/19401493.2011.567422.

[22] F. Stazi, F. Naspi, M. D'Orazio, Modelling window status in school classrooms. Results from a case study in Italy, Build. Environ. 111 (2017) 24–32. https://doi.org/10.1016/j.buildenv.2016.10.013.

[23] H.B. Rijal, P. Tuohy, M.A. Humphreys, J.F. Nicol, A. Samuel, J. Clarke, Using results from field surveys to predict the effect of open windows on thermal comfort and energy use in buildings, Energy Build. 39 (2007) 823–836. https://doi.org/10.1016/j.enbuild.2007.02.003.

[24] D. Lai, S. Jia, Y. Qi, J. Liu, Window-opening behavior in Chinese residential buildings across different climate zones, Build. Environ. 142 (2018) 234–243. https://doi.org/10.1016/j.buildenv.2018.06.030.

[25] V. Fabi, R.V. Andersen, S. Corgnati, B.W. Olesen, Occupants' window opening

behaviour: A literature review of factors influencing occupant behaviour and models, Build. Environ. 58 (2012) 188–198. https://doi.org/10.1016/j.buildenv.2012.07.009.

[26] R. Andersen, V. Fabi, J. Toftum, S.P. Corgnati, B.W. Olesen, Window opening behaviour modelled from measurements in Danish dwellings, Build. Environ. 69 (2013) 101–113. https://doi.org/10.1016/j.buildenv.2013.07.005.

[27] S. Shi, B. Zhao, Occupants' interactions with windows in 8 residential apartments in Beijing and Nanjing, China, Build. Simul. 9 (2016) 221–231. https://doi.org/10.1007/s12273-015-0266-z.

[28] S. Shi, H. Li, X. Ding, X. Gao, Effects of household features on residential window opening behaviors: A multilevel logistic regression study, Build. Environ. 170 (2020) 106610. https://doi.org/10.1016/j.buildenv.2019.106610.

[29] M. Han, R. May, X. Zhang, X. Wang, S. Pan, Y. Da, Y. Jin, A novel reinforcement learning method for improving occupant comfort via window opening and closing, Sustain. Cities Soc. 61 (2020). https://doi.org/10.1016/j.scs.2020.102247.

[30] F. Stazi, F. Naspi, G. Ulpiani, C. Di Perna, Indoor air quality and thermal comfort optimization in classrooms developing an automatic system for windows opening and closing, Energy Build. 139 (2017) 732–746. https://doi.org/10.1016/j.enbuild.2017.01.017.

[31] J.M. Dussault, M. Sourbron, L. Gosselin, Reduced energy consumption and enhanced comfort with smart windows: Comparison between quasi-optimal, predictive and rule-based control strategies, Energy Build. 127 (2016) 680–691. https://doi.org/10.1016/j.enbuild.2016.06.024.

[32] Y. Chen, L.K. Norford, H.W. Samuelson, A. Malkawi, Optimal control of HVAC and window systems for natural ventilation through reinforcement learning, Energy Build. (2018). https://doi.org/10.1016/j.enbuild.2018.03.051.

[33] R.A. Tanner, G.P. Henze, Stochastic control optimization for a mixed mode building considering occupant window opening behaviour, J. Build. Perform. Simul. 7 (2014) 427–444. https://doi.org/10.1080/19401493.2013.863384.

[34] W.W. Nazaroff, Indoor Partical Dynamics, Indoor Air. 14 (2004) 175–183. https://escholarship.org/uc/item/7sq4x34d#author.

[35] C. Howard-Reed, L.A. Wallace, W.R. Ott, The effect of opening windows on air change rates in two homes, J. Air Waste Manag. Assoc. 52 (2002) 147–159. https://doi.org/10.1080/10473289.2002.10470775.

[36] J. Liu, X. Dai, X. Li, S. Jia, J. Pei, Y. Sun, D. Lai, X. Shen, H. Sun, H. Yin, K. Huang, H. Tan, Y. Gao, Y. Jian, Indoor air quality and occupants' ventilation habits in China: Seasonal measurement and long-term monitoring, Build. Environ. 142 (2018) 119–129. https://doi.org/10.1016/j.buildenv.2018.06.002.

[37] J. Pei, C. Dong, J. Liu, Operating behavior and corresponding performance of portable

air cleaners in residential buildings, China, Build. Environ. 147 (2019) 473–481. https://doi.org/10.1016/j.buildenv.2018.08.009.

[38] Institute Forward Industry Research, Market analysis report on air cleaner industry worldwide. (In Chinese) (2020), https://bg.qianzhan.com/trends/detail/506/200422-7abf8eb0.html.

[39] R.S. Sutton, A.G. Barto, Reinforcement learning: An introduction, MIT press, 2018.

[40] M. Van Otterlo, M. Wiering, Reinforcement learning and markov decision processes, in: Adapt. Learn. Optim., 2012. https://doi.org/10.1007/978-3-642-27645-3_1.

[41] M. Han, R. May, X. Zhang, X. Wang, S. Pan, D. Yan, Y. Jin, L. Xu, A review of reinforcement learning methodologies for controlling occupant comfort in buildings, Sustain. Cities Soc. 51 (2019) 101748. https://doi.org/10.1016/j.scs.2019.101748.

[42] Z. Zhang, A. Chong, Y. Pan, C. Zhang, K.P. Lam, Whole building energy model for HVAC optimal control: A practical framework based on deep reinforcement learning, Energy Build. 199 (2019) 472–490. https://doi.org/10.1016/j.enbuild.2019.07.029.

[43] M. Han, R. May, X. Zhang, X. Wang, S. Pan, D. Yan, Y. Jin, L. Xu, A review of reinforcement learning methodologies for controlling occupant comfort in buildings, Sustain. Cities Soc. (2019). https://doi.org/10.1016/j.scs.2019.101748.

[44] K. Dalamagkidis, D. Kolokotsa, K. Kalaitzakis, G.S. Stavrakakis, Reinforcement learning for energy conservation and comfort in buildings, Build. Environ. 42 (2007) 2686–2698. https://doi.org/10.1016/J.BUILDENV.2006.07.010.

[45] S.K. Heo, K.J. Nam, J. Loy-Benitez, Q. Li, S.C. Lee, C.K. Yoo, A deep reinforcement learning-based autonomous ventilation control system for smart indoor air quality management in a subway station, Energy Build. (2019). https://doi.org/10.1016/j.enbuild.2019.109440.

[46] Z. Zhang, A. Chong, K. Poh Lam, Y. Pan, C. Zhang, S. Lu, A deep reinforcement learning approach to using whole building energy model for HVAC optimal control ASHRAE multidisciplinary task group on occupant behavior in buildings view project International Energy Agency Energy in Buildings and Communities Program, Blog.Nus.Edu.Sg. (2018). https://www.researchgate.net/publication/326711617.

[47] V. Mnih, K. Kavukcuoglu, D. Silver, A.A. Rusu, J. Veness, M.G. Bellemare, A. Graves, M. Riedmiller, A.K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, D. Hassabis, Human-level control through deep reinforcement learning, Nature. 518 (2015) 529–533. https://doi.org/10.1038/nature14236.

[48] S. Shi, C. Chen, B. Zhao, Air infiltration rate distributions of residences in Beijing, Build. Environ. 92 (2015) 528–537. https://doi.org/10.1016/j.buildenv.2015.05.027.

[49] S. Shi, C. Chen, B. Zhao, Modifications of exposure to ambient particulate matter: Tackling bias in using ambient concentration as surrogate with particle infiltration factor

and ambient exposure factor, Environ. Pollut. 220 (2017) 337–347. https://doi.org/10.1016/j.envpol.2016.09.069.

[50] C. Liu, J. Yang, S. Ji, Y. Lu, P. Wu, C. Chen, Influence of natural ventilation rate on indoor $PM_{2.5}$ deposition, Build. Environ. 144 (2018) 357–364. https://doi.org/10.1016/j.buildenv.2018.08.039.

[51] Eggert T. A, Emission of ozone and dust from laser printers, in: Proc. 1990 EPA/AWMA Int. Symp. Meas. Toxic Relat. Air Pollut., 1990.

[52] H. Kim, K. Kang, T. Kim, Measurement of particulate matter ($PM_{2.5}$) and health risk assessment of cooking-generated particles in the kitchen and living rooms of apartment houses, Sustain. 10 (2018). https://doi.org/10.3390/su10030843.

[53] M.P. Wan, C.L. Wu, G.N. Sze To, T.C. Chan, C.Y.H. Chao, Ultrafine particles, and $PM_{2.5}$ generated from cooking in homes, Atmos. Environ. 45 (2011) 6141–6148. https://doi.org/10.1016/j.atmosenv.2011.08.036.

[54] J.S. Apte, J.D. Marshall, A.J. Cohen, M. Brauer, Addressing global mortality from ambient $PM_{2.5}$, Environ. Sci. Technol. 49 (2015) 8057–8066. https://doi.org/10.1021/acs.est.5b01236.

[55] R.T. Burnett, C. Arden Pope, M. Ezzati, C. Olives, S.S. Lim, S. Mehta, H.H. Shin, G. Singh, B. Hubbell, M. Brauer, H. Ross Anderson, K.R. Smith, J.R. Balmes, N.G. Bruce, H. Kan, F. Laden, A. Prüss-Ustün, M.C. Turner, S.M. Gapstur, W.R. Diver, A. Cohen, An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure, Environ. Health Perspect. 122 (2014) 397–403. https://doi.org/10.1289/ehp.1307049.

[56] J.S. Apte, M. Brauer, A.J. Cohen, M. Ezzati, C.A. Pope, Ambient $PM_{2.5}$ reduces global and regional life expectancy, Environ. Sci. Technol. Lett. 5 (2018) 546–551. https://doi.org/10.1021/acs.estlett.8b00360.