Construction of operational control rules for an earth-to-air heat exchanger through transfer reinforcement learning

Yuki Adachi [1], Yasuyuki Shiraishi[1]

1 The University of Kitakyushu
1-1 Hibikino, Wkamatsu, Kitakyusyu, Fukuoka
Japan 808-0135

ABSTRACT

In recent years, earth-to-air heat exchanger (EAHE) systems, which is a method of pre-cooling and pre-heating outdoor air with earth-to-air heat, have been attracting attention as one of the technologies to achieve ZEB. However, at the operational phase, in order to achieve both energy saving and suppression of dew condensation control, EAHE control methods such as the timing or amount of outdoor air introduction have not been established. Recently, research on operational control by reinforcement learning (RL) has become popular and has attracted attention in the field of air conditioning control. RL is effective even in cases where future states are difficult to predict, such as EAHE. In previous studies, the unsteady CFD analysis method proposed by the authors made it possible to evaluate the annual energy savings and dew condensation in EAHE in detail. In addition, it was clarified that the RL, which uses the same CFD method as a simulator, can establish a control law that achieves both energy-saving effects and prevent indoor air pollution by suppressing dew condensation. On the other hand, RL requires a huge number of trials to construct the control law.

Therefore, the purpose of this study is to improve the learning speed and control performance. First, we adopt transfer learning (TL), which reuses a model pre-trained in RL for training in a new environment. Next, we verify the effectiveness of using this transfer reinforcement learning (TRL) as a control method for EAHE. The result showed that TRL achieved better control performance and faster learning speed than RL. In addition, it was suggested that EAHE with insufficient actual measurement data may be efficiently controlled from the first year of operation by directly using the control law established in advance. It was confirmed that RL performs well in terms of energy efficiency and air quality maintenance.

KEYWORDS

Earth-to-air heat exchanger system, CFD, Transfer Reinforced Learning, Ventilation, Outdoor air load, Condensation, Indoor air quality

1 INTRODUCTION

One of the fundamental technologies for developing zero-energy buildings is earth-to-air heat exchanger (EAHE) systems. An EAHE is a passive component of a heating, ventilation, and air conditioning (HVAC) system that utilizes the large heat capacity of the soil to pre-cool the outdoor air (OA) in the summer and pre-heat it in the winter. Further, it can reduce the heat loads of OA for air handling units or fresh air handling units (FAHU) by introducing pre-heated or pre-cooled OA through this system. The operational control of such a system is limited to control based on schedules, sequential disturbances, and internal system conditions. [1)] We have been using reinforcement learning (RL) to develop a control law for an EAHE system (underground pit system) to achieve energy-saving effects and prevent indoor air pollution by suppressing dew condensation. [2)] However, RL has the problem of requiring a huge number of trials for learning convergence. For this reason, previous studies have reduced the computational load on the environment side of the RL (e.g., by reducing the number of meshes

in computational fluid dynamics [CFD] analysis), thereby enabling a large number of training cycles. Another solution to this problem that has been attracting attention in recent years is transfer learning (TL), which enables faster learning and improved learning performance by reusing previously learned models for training in a new environment. [3] However, few studies have applied TL to RL, and important details have yet to be clarified, such as the area (range) in which TL is effective.

The objective of this study was to construct an efficient operational control law for soil heat exchange systems using RL that utilizes TLs to speed up and improve the learning performance of RL. We do this for an EAHE, and after adapting TL to RL, we conduct a case study on the transition target to verify the effectiveness of transition reinforcement learning (TRL) as an operational control method.

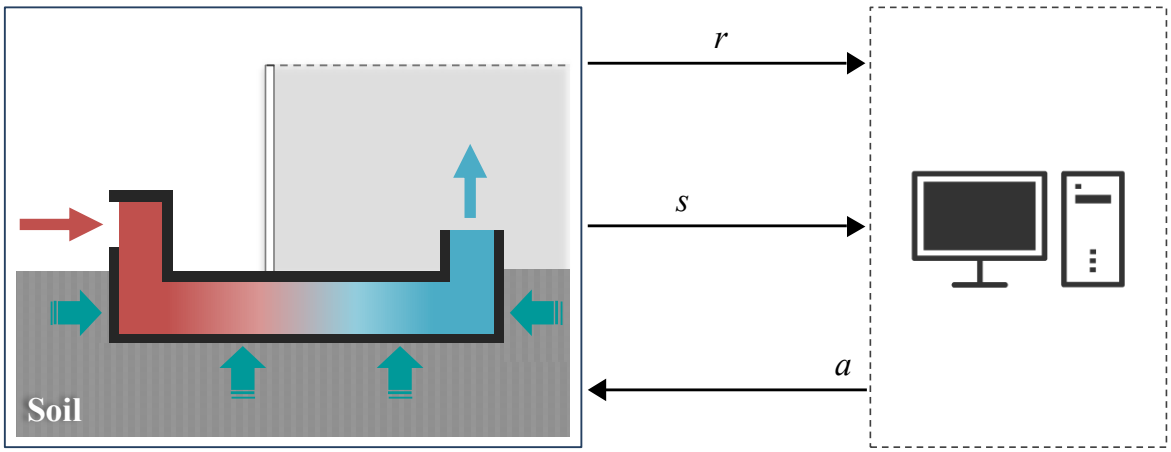


Figure 1 Diagram of RL process with EAHE as example

2 REINFORCEMENT LEARNING

2.1 Reinforcement Learning Overview

As shown in Figure 1, the RL is composed of the environment and the agents that control its operational decisions, with a reciprocal relationship between them. In RL, at a certain state $\boldsymbol{s_t}$ on the environment side, the agent manipulates the environment based on the action $\boldsymbol{a_t}$, the output by the agent, and the next state $\boldsymbol{s_{t+1}}$, resulting from the transition and the immediate reward $\boldsymbol{r_{t+1}}$ obtained at the transition destination (an example is an evaluation value, such as how much energy was saved), is passed from the environment to the agent. By repeating this through trial and error for the state $\boldsymbol{s_t}$, RL learns to output an action $\boldsymbol{a_t}$ that maximizes the sum of the immediate rewards for a certain period of time.

2.2 Reinforcement Learning Problem Setup

The definitions of states $\boldsymbol{s}$, action $\boldsymbol{a}$, and immediate reward $\boldsymbol{r}$ are given in Table 1 as the problem set for RL. In addition to weather conditions, such as the outdoor air temperature and absolute humidity, five types of states $\boldsymbol{s}$ were used as information in the system: the condensation area ratio and surface temperatures at two representative points in the system (near the inlet and outlet ports). Action $\boldsymbol{a}$ was set to five discrete values of "MD_o / MD_e OA damper opening," that is to say the "outdoor air intake through the system," as shown in Figure 2. In the subject system, the outdoor air conditioner was assumed to have a constant air volume of 8,100 m^3/h (CAV) during air intake hours. Since the control objective was to ensure energy-saving performance and to suppress condensation inside the system, the immediate reward $\boldsymbol{r}$

was defined as two kinds of rewards: the amount of heat processed by the external controller r_1 and the condensation area ratio r_2.

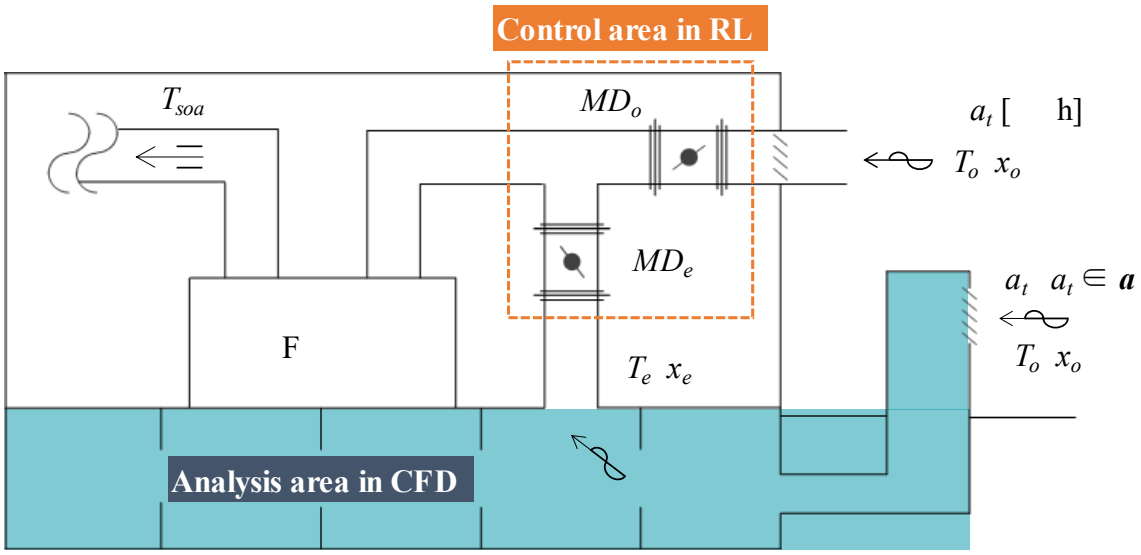


Figure 2 Diagram of EAHE

Table 1: Definition of state, action, and reward

State s	Outdoor air temperature/Outdoor air absolute humidity / Condensation area ratio / Surface temperature (2 points)
Action a	{0 m³/h, 2,025 m³/h, 4,050 m³/h, 6,075 m³/h, 8,100 m³/h}注1)
Reward r	$r = r_1 \times w_1 + r_2 \times w_2$ ($r_1 = 0.3$, $r_2 = 0.7$)
r_1	$r_1 = clip\left(\frac{Q_{FAHU}}{Q_{std}}, -1.0, 0\right)$ $T_{OA} = \frac{T_e \cdot a + T_o(max(a) - a)}{max(a)}$ $\Delta T = \begin{cases} T_{OA} - 22 & (if\ winter\ and\ T_{OA} < 22) \\ 26 - T_{OA} & (if\ summer\ and\ T_{OA} > 26) \\ T_{OA} - 20 & (if\ spring\ /\ fall\ and\ T_{OA} < 20) \\ 28 - T_{OA} & (if\ spring\ /\ fall\ and\ T_{OA} > 28) \\ 0 & (else) \end{cases}$ $Q_{FAHU} = C_p \times \rho \times \frac{max(a)}{3{,}600} \times \Delta T$
r_2	$r_2 = clip\left(\frac{C_{area}}{C_{std}}, -1.0, 0\right)$ $C_{area} = \frac{S_c}{S_e} \times 100$

Q_{std}: Standard value for r_1 [W], Q_{FAHU}: Heat load of FAHU [W], T_{OA}: h F U [℃] T_e: Outlet temp. of h [℃] T_o: Outdoor air [℃] ΔT: Difference in the blow- h F U [℃] C_p: specific heat capacity (=1.007) [kJ/(kg・K)], ρ: Air density (=1.206) [kg/m³], C_{area}: Condensation area ratio [%], S_e: Total surface area in the EAHE [m²], S_c: Condensation area in the EAHE [m²]

3. TRANSFER LEARNING

3.1 Transfer Learning Overview

TL is a framework in which the knowledge learned by the source task agent is reused by the target task agent. After incorporating TL in RL, the RL agent learns and acquires measures in the source task. Then in the target task, which is the same or a similar environment, the measures

acquired in the source task are reused, enabling faster learning and improved learning performance in the target task.

3.2 Transition Reinforcement Learning with Deep Learning

An effective method for policy reuse is the TL method using deep learning (DL). DL is a machine-learning method that uses a multi-layered neural network (NN), which is a mathematical model that mimics the network structure of neurons in the brain. NNs consist of an input layer, an intermediate layer, and an output layer, and the relationship between inputs and outputs in each layer is given by Equation (1).

$$\boldsymbol{y} = f(\boldsymbol{x}\boldsymbol{W} + \boldsymbol{b}) \qquad (1)$$

Here, $\boldsymbol{y}$ is the output vector, $\boldsymbol{x}$ is the input vector, $\boldsymbol{b}$ is the bias vector, $\boldsymbol{W}$ is the weight matrix, and f is the activation function. In an NN, the output of the L-1 layer is the input of the L layer. That is, each layer computes Equation (1) independently, so the NN can extract and combine layers. In TL using DL, this feature is utilized to reuse the NN model learned in the source task for training the NN in the target task. Also, by changing the number of layers to be extracted and the positions to be combined, it is possible to respond flexibly according to the target task.

The steps involved in the TRL implemented in this paper, shown in Figure 3, are as follows. First, the optimal measures (pre-trained Q-Network) in the source model (e.g., a simple CFD model of an underground pit) are learned using Deep Q-Network (DQN). Second, all or partial layers are extracted from the pre-trained Q-Network. Next, the layers extracted from the pre-trained Q-Network are combined with newly added layers to train the target model (e.g., a CFD model of a real underground pit). Then, the weights of all or some of the layers extracted from the pre-trained Q-Network are fixed (optional). Finally, training (re-training) is performed for all the layers that were combined or only the layers that were added. Thus, by reusing the Q-Network acquired in the source model and changing the transition rate (number of layers extracted from the pre-trained Q-Network) and layers to be fixed, we expect to improve initial performance, convergence (reduction in the amount of training data), and learning performance, as shown in Figure 3.

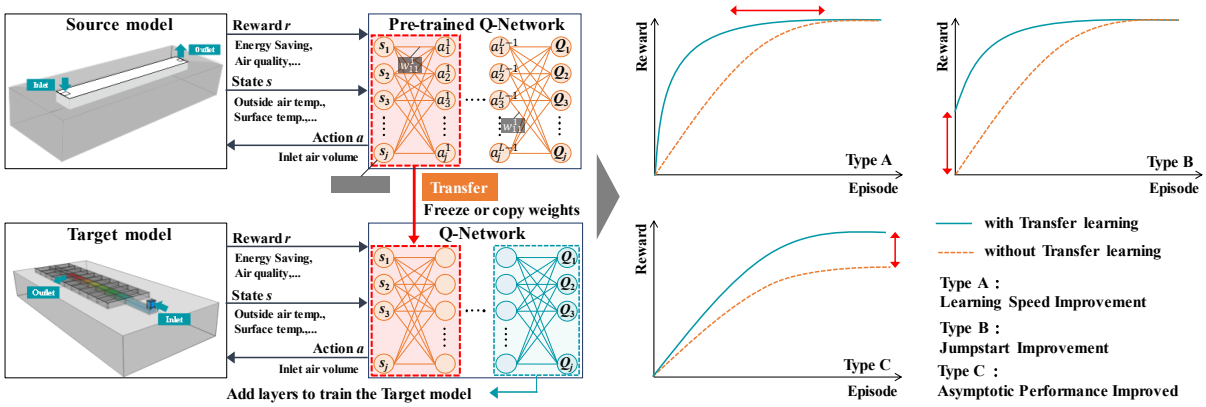


Figure 3 Diagram of Transfer Reinforcement Learning (Image of using DQN as reinforcement learning)

4. Analysis Condition

4.1 Target Building and CFD

EAHE system (underground pit system) is installed in a medium-sized office in Fukuoka Prefecture, Japan (Table 2). Outdoor air introduced into the underground pit through the inlet

protruding outdoors exchanges heat with the soil (concrete) for approximately 70 m to the outlet of the underground pit. The introduced outdoor air is pre-cooled and pre-heated and supplied to the outdoor air conditioner, contributing to energy savings in the outdoor air conditioner. Table 3 shows the CFD analysis conditions and Figures 4 and 5 show the analytical models. The outdoor air introduction period was from 9:00 to 18:00 daily. During the analysis, the amount of outdoor air introduced through the system was switched every hour according to the operational values output by reinforcement learning. CFD was employed as a simulator in the reinforcement learning environment, and the computational load reduction method of unsteady CFD, which assumes a fixed flow field, was used as the solution method.[4)]

Table 2: Brief Description of the Building

Location	Fukuoka
Use	Accommodations and Research Facilities
Structure	RC
Number of Stairs	1F-4F
Year Completed	July 2008
Extended Bed Area	5,498m^2
Underground Pit (W x H)	× 5 – 1.7m
Underground Pit (Length)	76.8m

Table 3: CFD conditions

Condition	Method/Parameter
Calculation period	1/1~12/31 (Approached period:1 year)
Time interval	3,600s
Domain	40.4m(X)×13.4m(Y)×6.9m(Z)
Mesh	Source model:6,422 (26(x)×19 (y)×13(z)) Target model:92,610 (70(x)×49 (y)×27(z))
Turbulence model, Scheme	Standard k-ε model, 1st-order upwind scheme for advection term, SIMPLE algorithm
Inflow boundary	U_{in} : a_t , T_o : Outdoor air temperature[5)] [℃] x_o : Outdoor air absolute humidity[5)] [k k ’] k_{in} = 3/2(U_{in}× 5 ^{2} , ε_{in} = $C\mu^{3/4}$・$k_{in}^{3/2}$/l_{in}
Initial temperature	Results of the 3D pre-analysis of this system controlled by schedule
Wall boundary	Velocity and Temperature: General logarithmic function Humidity: L ’ h x = 67
Upper side Boundary of the pit	: 22 26 ° Heat transfer coefficient: 23.0 W/(m^2·K)
Ground surface boundary	: [℃] Convective heat transfer coefficient: 17.9 W/m^{2}K

U_{in}: Inlet velocity [m/s], l_{in}: Length scale(=1.0m), k_{in}: Turbulence kinetic energy [m^2/s^2],
ε_{in}: Dissipation rate of k_{in} [m^2/s^3], C_μ: Model constant (=0.09) [-]

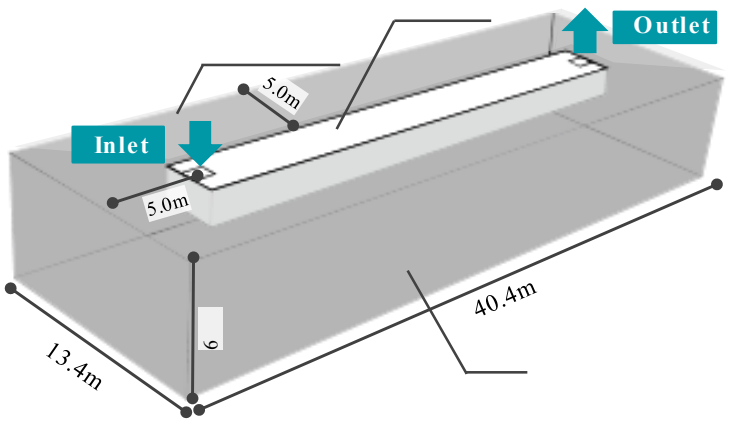


Figure 4 CFD model of EAHE system (Straight)

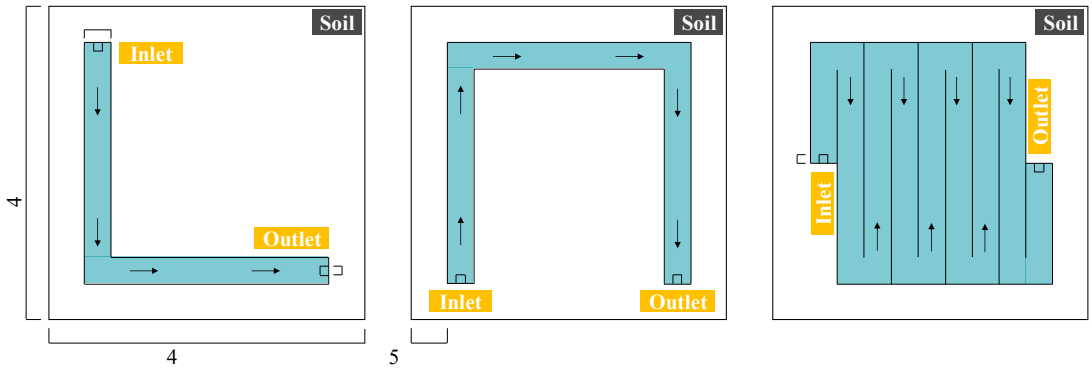


Figure 5 CFD model of EAHE system (**Left**: L, **Mid**: Corridor, **Right**: Meandering)

4.2 RL

The RL conditions are given in Table 4. In this study, DQN was used to implement TRL with DL. In an environment where actions are discrete, the Q function can be expressed without deepening the middle layer of the NN. Therefore, the middle layer for both the source and target 4 64 × 64 To speed up the learning process, the learning rate was set to 0.001 in the target model to efficiently learn the measures acquired in the source model. The ε-greedy method was used as the action s h ε = 5 h search in the target model while using the measures acquired in the source model from the initial stage of learning.

Table 4 Conditions of reinforcement learning

Algorithm / Episode	Deep Q-Network (DQN)[6] / 200
Discount factory	0.99
Exploration rate ε	ource: Linear schedule, $\varepsilon_0 = 1.0$, $\varepsilon_N = 0.02$ Target: Linear schedule, $\varepsilon_0 = 0.5$, $\varepsilon_N = 0.02$
Learning rate η	Source: 0.0005, Target: 0.001
Replay memory buffer	49,984
Q-Network / Batch size	5×64×64×2 5 2

N: Number of episode (≦200) [-]

4.3 TL

The similarity of the source and target is important when implementing TL. In this study, the effectiveness of TRL was verified by being conducting on various targets with measures learned with the same source model. The analysis cases are given in Table 5. The source model for each case was a straight-type CFD model. In Case 1 (transition of action), TRL from two types of airflow to five types of airflow was implemented. In the Case 2 series (shape transition), TRLs were performed from a rectilinear to an L-shape, a corridor, meandering, and for real buildings. In the Case 3 series (weather transition), TRL was conducted from Kitakyushu (warmer climate) to Fukuoka (warmer climate) and Akita (colder climate). Regarding the transition rate, the first two layers extracted from the Q-Network of the source model were combined with the new output layer because the number of nodes in the output layer was different from that in Case 1, which is a transition of action. Additionally, Cases 2 and 3 were assumed to be all layers. In all cases, the weights of the first half of the layers were fixed, and only the second half of the layers were trained. In Cases 2-4, the target model was an EAHE (straight) installed in a real building, and an analysis was conducted to directly apply the control laws constructed in the

source model to the operational control of the building to study the versatility of RL. For comparison, the analysis also included random control of RL and the outdoor air intake.

Table 5 Conditions of reinforcement learning

CASE	Source	Target	Object
1	2	5	action
2-1	Straight	L	Shape
2-2	Straight	Corridor	
2-3	Straight	Meandering	
2-4	Straight	Actual tunnel	
3-1	Kitakyushu	Fukuoka	Weather
3-2	Kitakyushu	Akita	

5. RESULTS

5.1 Progress of RL and TRL

Figure 6 shows the episode reduction rate and the number of episodes required for convergence of the RL and TRL studies for each case, and Figure 7 shows the progress of the studies from Cases 2-1 to 2-3. Figure 6 shows that Case 3-1 (Kitakyusyu to Fukuoka) resulted in the fastest learning speed and a 90-episode reduction. This is presumably because both the source and target models were straight EAHE systems, and the climates of Kitakyushu and Fukuoka are very similar, having the highest similarity between the source and target among all cases. For the Case 2 series (shape transition), more than 50 episodes were reduced in Case 2-1 (straight to L) from Figure 7. TRL also remained higher than RL with respect to total rewards. In Cases 2-2 (straight to corridor) and 2-3 (straight to meandering), the total reward remained high from the early stages of learning, which is presumably the result of efficiently utilizing the source model's measures from the early stages of learning. Cases 2-2 and 2-3 achieved an increase in learning speed of 40 episodes and 20 episodes, respectively. As in Case 2-1, TRLs showed a high sum of rewards, suggesting that they improved learning performance. This is presumably because, as in Case 3-1, the L-type had the highest similarity to the straight type. In all cases, however, TRL achieved episode reduction.

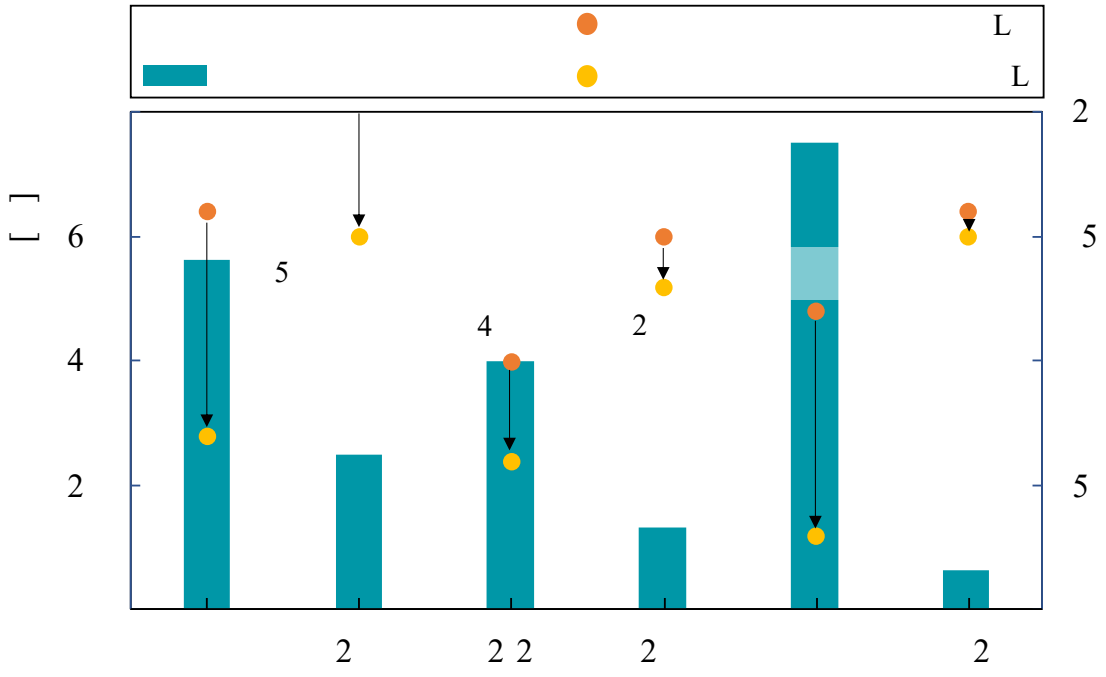
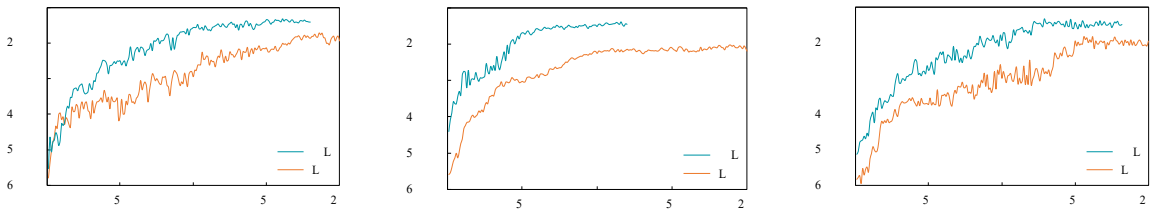


Figure 6 Episode reduction ratio / Episode of convergence



CASE2-1 (L type) CASE2-2 (Corridor type) CASE2-3 (Meandering type)

Figure 7 Total reward each episode (CASE2)

5.2 Comparison with RL and Random

Figures 8 and 9 show the heat rate and condensation area ratio of the external controller when controlled by Case 2-1, RL, and random, which are the fastest learning methods in Case 2. The condensation area percentages for Case 2-4 are also shown in Figure 10. For Case 2-1, the annual heat rates for the external air conditioner process were 128.8 GJ, 128.9 GJ, and 139.4 GJ for TRL, RL, and random, respectively, indicating that control with TRL and RL had high energy-saving performance. As for the condensation suppression effect, it was confirmed that TRL mitigated the condensation situation compared with RL and random, especially during the summer season. This confirms that TRL improved control performance. Finally, for Case 2-4, Figure 10 shows that the direct use of the source model's control law in the EAHE of a real building resulted in partial suppression of condensation compared with the random case, but the control performance was inferior to the case in which TRL was implemented. The annual heat output of the external controller was 211.9 GJ and 216.7 GJ for the random and source policies, respectively, with random being slightly higher. This may be due to the fact that the reward design prioritized the suppression of condensation. The results suggest that by reviewing the RL parameters and reward design, it is possible to obtain sufficient control performance for practical use without conducting new training.

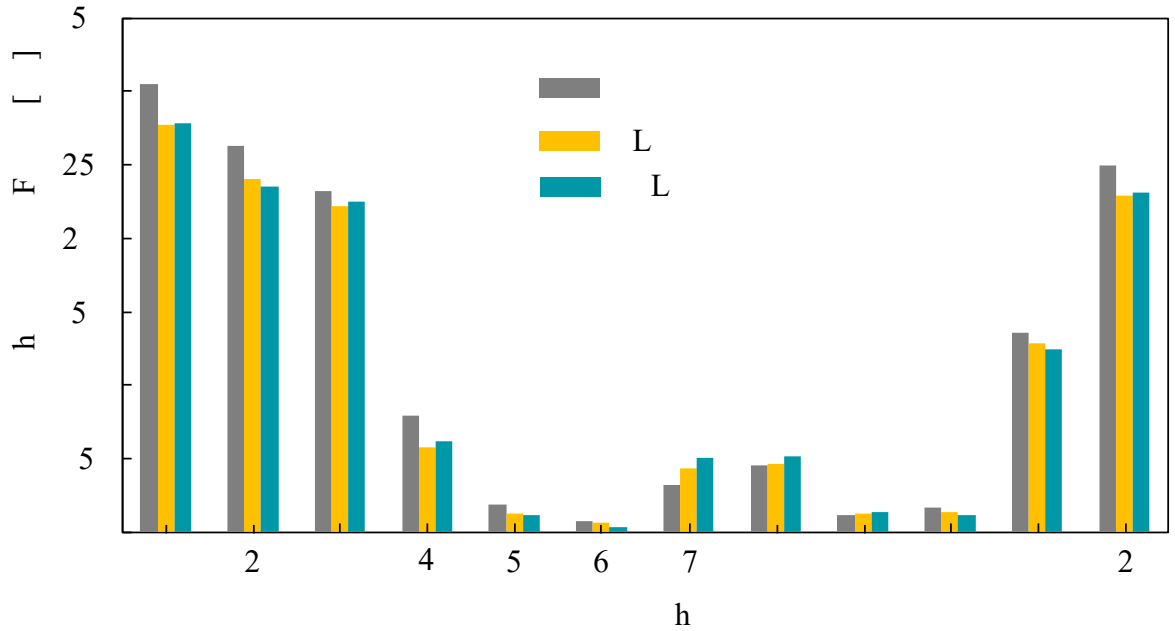


Figure 8 Monthly accumulated heat load of FAHU (CASE2-1)

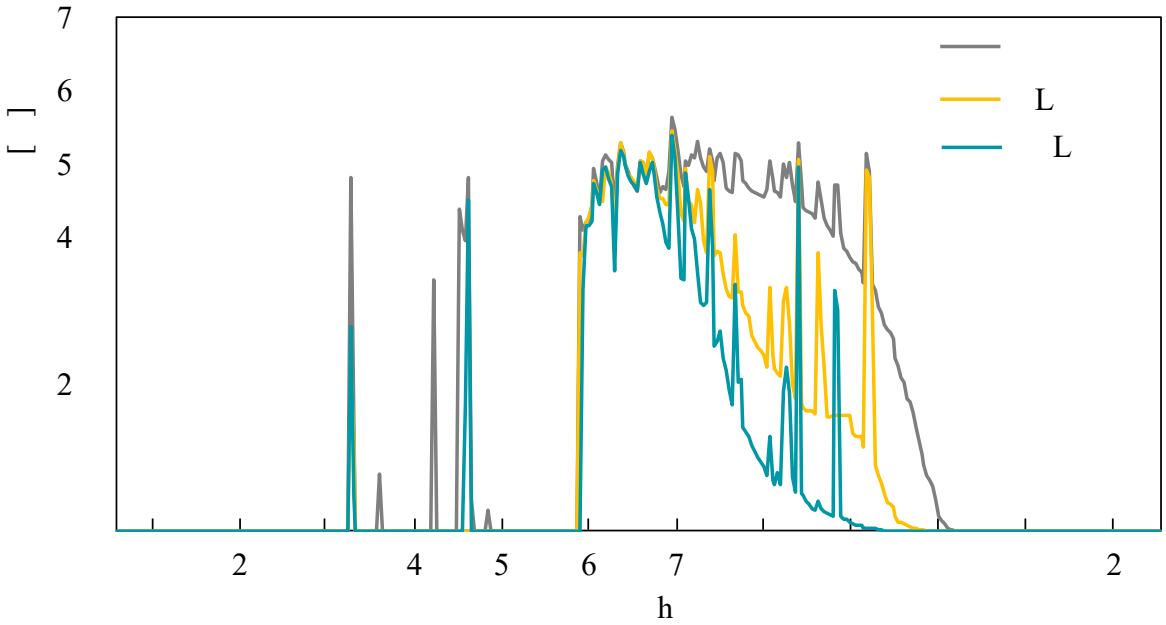


Figure 9 Condensation area ratio (CASE2-1)

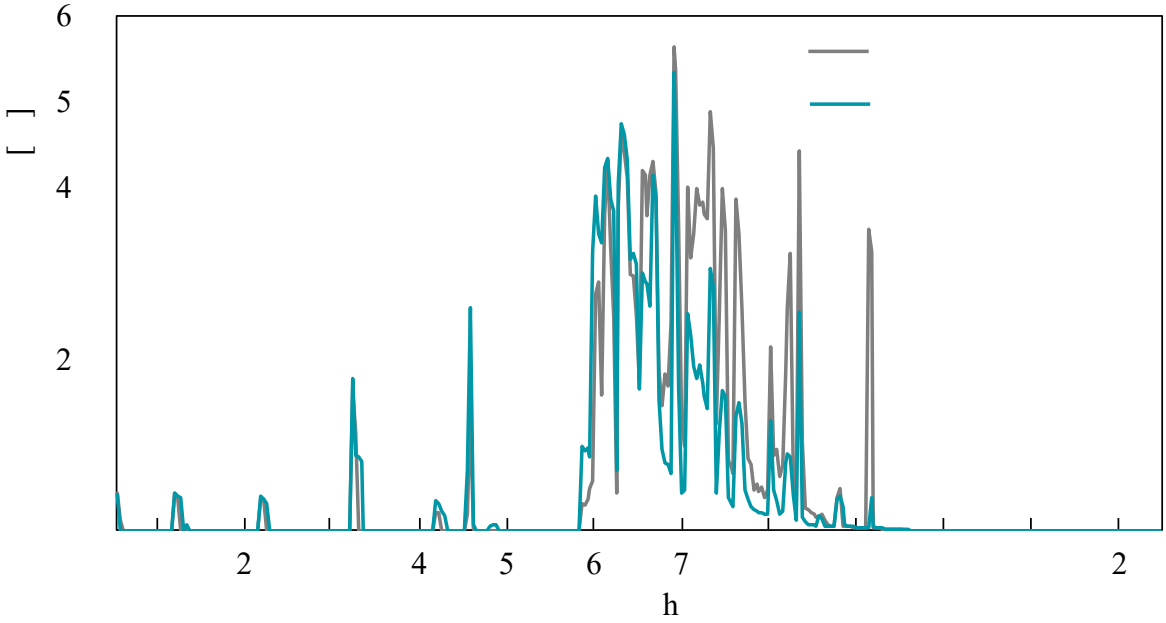


Figure 10: Condensation area ratio (CASE2-4)

6. CONCLUSIONS

By utilizing learned measures in the construction of new control laws (TRL), we were able to achieve better control performances of energy-savings and suppressing dew condensation and a faster learning speed than conventional RL. Our results also suggest that EAHEs with insufficient measured data can be efficiently controlled from the first year of operation by directly using the pre-constructed control law.

7. ACKNOWLEDGMENT

This work was supported by JSPSKAKENHI Grant Numbers 19H02301 and 20KK0102.

8. REFERENCES

1) Suzuki, H et al. (2021). Verification of energy and environmental performance of urban eco-campus, Part 6 Analysis of cool & heat pit energy savings in ZEB Ready buildings, Heating, Air-Conditioning and Sanitary Engineers of Japan, J-89, pp.357-360,2021.9 (in Japanese).
2) Motoyama, Y. (2021). Proposal for an optimal operation method for earth-to-air heat exchanger systems using reinforcement learning, Construction of control law by coupled analysis of Deep Q Network and CFD, Architectural Institute of Japan Kyusyu Chapter Architectural Research Meeting, No.60, pp.113-116, 2021.3 (in Japanese).
3) G Pinto et al. (2022). Transfer learning for smart buildings, A critical review of algorithms, applications and future perspectives, Advances in Applied Energy, Vol. 5, 2022.2,100084.
4) Tomoda, K. (2016). Annual prediction method for the thermal performance of earth-to-air heat exchanger by CFD analysis where calculation loads were reduced, Evaluation method of cooling and heating effect of an earth-to-air heat exchanger (Part 2), The Architectural , 81(722), pp.393-401, 2016.4.(in Japanese).
5) Expanded AMeDAS Weather Data (2000). The Architectural Institute of Japan.
6) Mnih et al. (2015). Human-level control through deep reinforcement learning. Nature Vol.518, pp.529-533, 2015.2.

9. APPENDIX

Table 6 : List of abbreviations

RL	Reinforcement Learning
TL	Transfer Learning
TRL	Transfer Reinforcement Learning
AHU	Air Handling Units
FAHU	Fresh Air Handling Units
DL	Deep Learning
NN	Neural Network
DQN	Deep Q-Network
CFD	Computational Fluid Dynamics