Journal of Building Performance Simulation

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tbps20

Trade-off decisions in a novel deep reinforcement learning for energy savings in HVAC systems

Suroor M. Dawood, Alireza Hatami & Raad Z. Homod

To cite this article: Suroor M. Dawood, Alireza Hatami & Raad Z. Homod (2022) Trade-off decisions in a novel deep reinforcement learning for energy savings in HVAC systems, Journal of Building Performance Simulation, 15:6, 809-831, DOI: 10.1080/19401493.2022.2099465

To link to this article: https://doi.org/10.1080/19401493.2022.2099465

Published online: 04 Aug 2022.

JOURNAL OF BUILDING PERFORMANCE SIMULATION
2022, VOL. 15, NO. 6, 809–831
https://doi.org/10.1080/19401493.2022.2099465

Trade-off decisions in a novel deep reinforcement learning for energy savings in HVAC systems

Suroor M. Dawood[a,b], Alireza Hatami [a] and Raad Z. Homod [c]

[a]Dept. of Elec. Eng., Bu-Ali Sina University, Mahdieh Street, Hamedan 65178-38695, Iran; [b]Dept. of Chem. Eng. and Oil Refining, Basrah University for Oil and Gas, Basrah 61004, Iraq; [c]Dept. of Oil and Gas Eng., Basrah University for Oil and Gas, Basrah 61004, Iraq

ABSTRACT

This paper presents Model-based Reinforcement Learning (MB-RL) techniques to control the indoor air temperature, and CO_2 concentration level, and minimize the energy consumption of the heating, ventilating, and air conditioning (HVAC) systems, simultaneously. For this purpose, a trade-off is made between maintaining indoor comfort levels and minimizing energy consumption. The control of the HVAC system is performed using the Deterministic Policy RL (DP-RL) method. Moreover, the nonlinear autoregressive exogenous neural network (NARX-NN) is employed as an approximation function with DP-RL method to provide a hybrid DP-NARX-RL controller. By applying the DP-RL and DP-NARX-RL controllers to the HVAC system of a typical building, parameters such as the indoor comfort levels, the electrical power, and energy consumed, and the energy costs at various pricing schemes are evaluated for two case studies. In both cases, the results show the better performance of DP-NARX-RL compared to DP-RL, RL, and PID controllers.

ARTICLE HISTORY
Received 1 November 2021
Accepted 30 June 2022

KEYWORDS
Heating, ventilating, and air conditioning (HVAC) systems; Energy saving; Reinforcement learning; Nonlinear autoregressive exogenous neural network (NARX-NN)

1. Introduction

Buildings are one of the largest consumers of energy in the world. They absorb 40% of total energy consumption and generate 1/3 of total CO_2 emissions (Homod, Gaied, et al. 2020; Homod, Togun, et al. 2020). About half of the buildings' energy is allocated to heating, ventilation, and air conditioning systems (HVAC systems) (Homod et al. 2021). Many studies have been conducted to reduce energy consumption to minimize greenhouse gas emissions. In building sections, much attention has been paid to energy management of the HVAC systems by controlling or scheduling them (Yoon, Kang, and Moon 2020).

Some classic approaches, such as feedback and rule-based control strategies are used for controlling the HVAC systems but these methods are inefficient, as the set-point cannot always be properly tracked and in the long run, the flexibility and efficiency are lost (Chen, Cai, and Bergés 2019).

Model predictive control (MPC) control strategy addresses these disadvantages via iterative optimization of an objective function over a planning horizon time. But its adoption is restricted because for controlling the HVAC systems, an accurate model is needed and typically the buildings are heterogeneous and have different layouts. These characteristics limit the scalability of MPC approach. In addition, the model quality of MPC is assessed by prediction error, which may lead to good or bad control performance (Zhao et al. 2021).

The machine learning (ML) approach has been adopted by many researchers for optimal control of HVAC systems (Alawadi et al. 2020). Based on the learning process, ML methods can be divided into four groups: supervised, unsupervised, semi-supervised, and Reinforcement Learning (RL) (Moubayed et al. 2018). RL is a learning-based approach that responds to different situations by interacting with the agent and its environment through learning a control policy (Sutton and Barto 2018). The RL agent goal is to find an appropriate action model that would maximize/minimize the total cumulative returned reward from its environment. RL method can be divided into model-free (MF) and Model-based (MB). In MF-RL control methods, the agent uses the trial-and-error process to learn the optimal policy using its own experiences and actions without the need of the system model. But it needs a long time and a considerable amount of training data to achieve an acceptable control performance (Wang and Hong 2020). MB-RL depends on the environment transition model to make actions. This model consists of terminal states, reward, and state transition functions. MB-RL method uses the log data to update the environment model. This makes the agents freely interact with their environment and achieve bet-

CONTACT Alireza Hatami hatamisharif@basu.ac.ir

ter control efficiency with less training data and consumed time (Polydoros and Nalpantidis 2017). RL methods have gained significant success in managing and controlling the HVAC systems (Sutton and Barto 2018). Many optimization hybrid RL methods have been applied to the HVAC systems to minimize the energy consumption or cost, with/without considering the demand response while keeping the thermal comfort levels at the specified levels (Wang and Hong 2020; Polydoros and Nalpantidis 2017; Du et al. 2021; Gao, Li, and Wen 2020; Azuatalam et al. 2020; Kurte et al. 2020; Wei, Ren, and Zhu 2019; Zhang, Chong et al. 2019; Gao, Li, and Wen 2019; Zhang, Chong et al. 2018; Vázquez-Canteli et al. 2018; Zou, Yu, and Ergan 2019; Wang, Velswamy, and Huang 2017; Yu et al. 2020; Yuan et al. 2020; Hao, Gao, and Zhang 2020; Sangi and Müller 2018; Ahn and Park 2019; Zhang, Kuppannagiri et al. 2019; Marantos et al. 2018). However, few studies have focused on keeping CO_2 levels below a specified value to maintain comfort levels (Table 1).

Deep reinforcement learning (DRL) has been applied as a hybrid of model-free RL and feed-forward neural network (NN) (Du et al. 2021; Gao, Li, and Wen 2020; Azuatalam et al. 2020; Kurte et al. 2020; Wei, Ren, and Zhu 2019; Zhang, Chong, et al. 2019; Gao, Li, and Wen 2019; Zhang et al. 2018; Vázquez-Canteli et al. 2018) or recurrent NN (Zou, Yu, and Ergan 2019; Wang, Velswamy, and Huang 2017) for optimal control of different types of buildings. The multi-agent system (MAS) DRL has been applied to the HVAC systems (Yu et al. 2020). RL combined with MPC, and the rule-based algorithm have been presented by Chen, Cai, and Bergés (2019) and Yuan et al. (2020), respectively. A hybrid agent-based RL and MPC have been proposed by Hao, Gao, and Zhang (2020) and Sangi and Müller (2018), respectively. To learn the dynamics of the HVAC systems, Deep Q-network (DQN) for model-free RL, NN for model-based RL, and Neural Fitted Q-iteration for RL have been applied for scheduling the HVAC systems by Ahn and Park (2019); Zhang, Kupannagiri, et al. (2019); Marantos et al. (2018), respectively. An event-triggered compatible off-policy deterministic-actor-critic method hybrid with a Q-learning approach has been introduced by Hosseinloo et al. (2020). A hybrid model-based DRL has been applied to the HVAC systems by Zhao et al. (2021). Table 1 summarizes and compares the main features of this paper, and the hybrid RL-based methods used to control the HVAC systems.

This paper addresses some of the lacks in the literature, such as maintaining the CO_2 concentration level and indoor air temperature as the occupants' comfort level and at the same time minimizing the building electrical energy consumption of the HVAC system, simultaneously. It presents Model-based Reinforcement Learning (MB-RL) techniques to control the indoor air temperature, and CO_2 concentration level, and minimize the electrical energy consumed by the HVAC systems, simultaneously. In fact, a trade-off is made between keeping the indoor comfort levels within the acceptable ranges and minimizing the electrical energy consumption of the HVAC systems.

For this purpose, the proposed approaches are developed to be applicable to control HVAC systems, lighting, windows, and process ventilation air systems, to respond to the requirements of the smart buildings. The occupants' comfort levels are evaluated regarding the indoor air temperature and CO_2 concentration level. The temperature signal is sensed and evaluated by building a mathematical model. In contrast, the indication of CO_2 concentration level is numerically modeled by the Lagrangian approach. Both two signals are fed directly to the proposed control algorithms to obtain the state space. Two control algorithms based on RL have been proposed; one uses the deterministic policy (DP) algorithm with a model-based RL, denoted by the DP-RL method. The DP-RL method realizes an accurate control execution. It can be trained as a precise model with a finite amount of data without the need for trial-and-error parameter adjustment. The other uses the nonlinear autoregressive exogenous (external inputs) neural network (NARX-NN) combined with deterministic policy and a model-based RL. The second is deep reinforcement learning (DRL), indicated by the DP-NARX-RL method. By implementing the second online control strategy on the building system, electrical energy cost and consumption are reduced over a planning horizon (for example, 24 h). In contrast, the system's occupant thermal and air quality comfort are improved.

After developing the control approaches, two case studies are analyzed to evaluate the performance of the proposed controllers. In *case study 1*, the DP-RL and DP-NARX-RL controllers are used to control the HVAC system of a typical building in Basra, Iraq. Then some parameters such as the indoor comfort levels, the electrical power and energy consumed by the HVAC system, and the energy costs at various pricing schemes are assessed for a planning horizon of a day. The results show the better performance of the DP-NARX-RL method than other approaches. In *case study 2*, a multi-chiller HVAC system of Basra International Airport, Iraq is controlled by the DP-NARX-RL and PID approaches. Then the performance of controllers is assessed in terms of indoor conditions stability, reducing the electrical power and energy consumed, and energy costs with different pricing schemes. The results show the better performance of DP-NARX-RL than benchmark.

Table 1. Summary of the related works compared with this work.

Ref.	Year	Building	HVAC	OC	DR	CO_2	Compared with	Software used	Run time and/or episodes
Du et al. (2021)	2021	Multi-zone residential	√	√	✗	✗	DQN and baseline	Python+ TensorFlow	300 episodes, 10-days
Zhao et al. (2021)	2021	–	√	√	✗	✗	MPC and model-free RL	Australian energy market operator	2880 iterations
Gao, Li, and Wen (2020)	2020	Laboratory	√	√	✗	✗	Support vector machine, Linear-, Ensemble- and Gaussian process – regression	TRNSYS+ MATLAB+ PyTorch + MySQL	300 episodes
Yu et al. (2020)	2020	Multi-zone commercial	√	√	✗	√	Rule-based (RB) and heuristic	EnergyPlus+ Python	20000 episodes, t_{train}: 13 hrs.
Azuatalam et al. (2020)	2020	Commercial	√	√	√	✗	RL+ Downward-DR, Baseline, RL+ Upward-DR	EnergyPlus+ Python+ BCVTB	10 hrs. for 1000 episodes
Yuan et al. (2020)	2020	Single- and multi-zone commercial	√	√	✗	✗	RB and PID	TRNSYS+ MATLAB+ Python	(2–7) Years
Kurte et al. (2020)	2020	Single- and two-zone residential	√	√	✗	✗	Baseline	Home energy management systems software	50 episodes every episode had 61-days, $t_{test} = 15$ min
Hao, Gao, and Zhang (2020)	2020	Commercial recreation center	√	√	✗	✗	Multi-agent game theory	Equest	10000 Iteration
Hosseinloo et al. (2020)	2020	Single-zone residential	√	√	✗	✗	Classic RL	EnergyPlus	10-days
Wei, Ren, and Zhu (2019)	2019	Single and multiple commercial distributed mixed-use	√	√	✗	✗	RB	EnergyPlus+ BCVTB	100 episodes
Chen, Cai, and Bergés (2019)	2019	Conference room and office	√	√	✗	✗	Baseline *P*- And existing *P*-controllers	EnergyPlus+ PyTorch (Python)	Three-weeks
Zhang, Chong et al. (2019)	2019	Office	√	✗	√	✗	RB	EnergyPlus+ BCVTB + Python (OpenAI gym)	One episode lasts from Jan 1st to Mar 31th, 3-months of observed data, 5 min.
Zou, Yu, and Ergan (2019)	2019	Office	√	√	✗	✗	RB	EnergyPlus	31 hrs., 200 episodes
Gao, Li, and Wen (2019)	2019	Laboratory	√	√	✗	✗	Q Learning and DQN	TRNSYS+MySQL+ PyTorch	1440 min 300–450 episodes
Ahn and Park (2019)	2019	Office	√	✗	✗	√	Baseline	EnergyPlus	14-days, 1000 episodes
Zhang, Kupannagiri et al. (2019)	2019	Two-room data center	√	√	✗	✗	Baseline and model-free RL	EnergyPlus	3–14 days 30–120 episodes
Marantos et al. (2018)	2018	Single-zone residential	√	√	✗	✗	RB	EnergyPlus+ BCVTB + MATLAB	90-days
Zhang, Chong et al. (2018)	2018	Office	√	√	✗	✗	-	BuildSimHub+ EnergyPlus	10 hrs.
Vázquez-Canteli et al. (2018)	2018	Multi-zone residential	√	✗	√	✗	RB+Multi-agent batch RL	CitySim+ TensorFlow+ Python (Keras)	$t_{simulation}$ for hot weather: 122-days, $t_{simulation}$ for case study: 13 min. and 21sec.
Sangi and Müller (2018)	2018	Energy research center	√	√	✗	✗	Mode-based and classic agent-based	Software-in-the-loop simulation	28-day
Wang, Velswamy, and Huang (2017)	2017	Office	√	√	✗	✗	Ideal PMV and variable control	EnergyPlus+ BCVTB + MATLAB	5–days
This paper		Residential Airport	√	√	✗	√	DP-RL and RL PID	MATLAB	15 min. and 1000 episodes 4 min. and 1000 episodes

Key: **Building**: Type of building used, **HVAC**: Paper objective for minimizing energy consumption or energy cost, **OC**: Occupants' thermal comfort levels, **DR**: Demand response, **CO_2**: CO_2 concentration, Compared with: Compared with other controllers, √: a feature existing, **✗**: a feature lacking.

The main contributions of this paper are as follows:

1 Modifying a building model by adding a CO_2 sensor to evaluate the CO_2 concentration level and indoor temperature as the occupants' comfort level.
2 Developing two model-based RL control approaches using the deterministic policy (DP) algorithm, denoted by the DP-RL and DP-NARX-RL; the last approach uses the nonlinear autoregressive exogenous neural network (NARX-NN) combined with the deterministic policy.
3 Minimizing the electrical energy consumption by applying the proposed approaches to the HVAC systems while the system's occupant thermal and air quality comfort are maintained within the acceptable ranges for a short-term planning horizon.
4 Comparing the performance of the DP-NARX-RL, with RL, DP-RL, and PID for controlling an HVAC system and a multi-chiller system in two case studies. The results show the superiority of the DP-NARX-RL compared to other approaches in different aspects.

2. Problem formulation

2.1. The HVAC system model description

In most traditional HVAC systems, mathematical models are based on the indoor temperature and, in some cases, relative temperature and humidity (Homod et al. 2011; Homod et al. 2010). In this paper, a modified model for the HVAC system defined by Hosseinloo et al. (2020) is systematized to be the indoor air temperature and CO_2 concentration level as its outputs that can be considered as controlled variables. Figure 1 shows the block diagram of the proposed model which consists of the main subsystems: the air handling unit (heat exchanger), the CO_2 sensor, and the indoor conditioned space model. This model is a multi-input/multi-output (MIMO) system with two outputs and seven inputs. The system inputs are: position damper of the fresh and return air ($D_{rf}(t)$ and $D_{rr}(t)$, close/open windows/doors (WiON(t)), turn On/Off lights (LiON(t)), chilled water valve position (ch(t)), outdoor temperature ($T_o(t)$), and air supply temperature ($T_s(t)$)). The system outputs are the essential comfort conditions in the building: indoor air temperature ($T_r(t)$) and indoor air quality level (CO_2 concentration(t)). The geometry of the building used in this model is similar to that one used by Homod, Sahari, and Almurib (2014). In the following, the subsystem models are described.

2.1.1. The air handling unit (AHU)

The air processing unit (cooling /heating system) model is achieved by changing the temperature ratio in the control volume of the air transfer unit and its heat exchanger. It can be obtained using the first law of thermodynamics and energy conservation law in Laplace transform as follows (Homod et al. 2011):

$$T_s(s) = \frac{D_{rr}(s)\,T_r(s) + D_{rf}(s)T_o(s)}{(\tau_1 s + 1)} + \frac{ch(s)\,cp_w \Delta T_w}{m_a^{\bullet} cp_a(\tau_1 s + 1)} \quad (1)$$

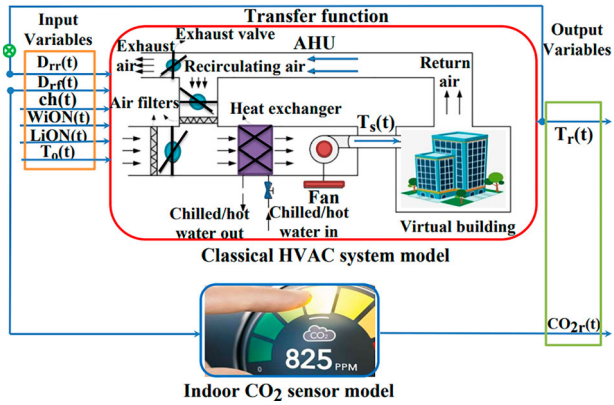


Figure 1. The modified HVAC system block diagram.

2.1.2. The CO_2 concentration sensor model

The mass conservation law has been employed for evaluating the indoor CO_2 concentration level, which is highly related to the indoor air quality index (IAQ). Assuming that the outside CO_2 concentration does not change (600 ppm [Baghaee and Ulusoy 2018]), the CO_2 emission can be divided into two parts: the first part is for the CO_2 produced by the inhabitants, and indoor appliances make the second part.

Lagrange polynomials modelling (Chapra and Canale 2015) (using the Laplace transform) has been employed to describe this sub-model based on the mass conservation law as shown in Eqns. (2) and (3).

$$CO_{2indoor}(s) = \frac{v_r CO_{2out}\,D_{rr}(s)\,F}{v_r^{\bullet}\,(\tau_2 s + 1)} + \frac{v_r\,CO_{2gen_l}(s)}{v_r^{\bullet}\,(\tau_2 s + 1)} \quad (2)$$

$$CO_{2gen_l}(t) = \prod_{\substack{J=0 \\ J \neq l}}^{Z} \left(\frac{t - t_J}{t_l - t_J}\right) f(t_l) \quad (3)$$

2.1.3. Indoor conditioned space model

The temperature ratio efficiency can be studied by applying the energy and mass conservation laws to the conditioned space control volume. Using the thermal balance equations on the conditioned space, the thermal load components can be given as (Homod et al. 2011).

$$T_r(s) = \frac{m_{as} cp_a T_s(s)}{\left(\frac{KA}{\Delta x} + 2m_{as} cp_a\right)(\tau_3 s + 1)} + \frac{m_{av} cp_a T_o(s) * WiON(s)}{\left(\frac{KA}{\Delta x} + 2m_{as} cp_a\right)(\tau_3 s + 1)}$$

$$+\frac{KAT_o(s)(1+0.6)}{\Delta x\left(\frac{KA}{\Delta x}+2m_{as}cp_a\right)(\tau_3 s+1)}+\frac{40\,LiON(s)}{\left(\frac{KA}{\Delta x}+2m_{as}cp_a\right)(\tau_3 s+1)} \quad (4)$$

Table 2 describes the parameters and variables used in HVAC system modeling and the related building in Basra, Iraq which is shown in Figure 2 (Pita and Stevenson 1998). Hereinafter, this case study is called *case study 1*. A more detailed description of the physical behaviour of the system can be found in Appendix A (Homod et al. 2011). It is necessary to mention that in the building model, the latent heat is not taken into consideration because the controller deals with indoor temperature and carbon dioxide, so its presence has no effect on the efficiency of the controller, which is the main objective of this study. While the rest of the thermal loads, when added to the model calculations, will lead to a reduction of the response time, which will increase the time learning of the agent, therefore it was estimated and incorporated with the indoor heating load, as a function of time and this does not affect the test of the controller. Meanwhile, the building model used for controllers design is not a reduced-order model. In other words, first, the equations of the complete model (heat exchanger, building model, … .) are transferred from time space to Laplace

Table 2. Description of HVAC system model parameters and variables (Homod et al. 2011).

Component	Description	Value	Parameter/ Variable	Unit
$m_a^\bullet = m_{as} = m_{av}$	The mass flow rate of outside, ventilation, and supply air at time t	0.84	Par.	(*kg*/ sec .)
$ch(t)$	The mass flow rate of chilled water at time t	[0 1]	Var.	(*kg*/ sec .)
$D_{rr}(t)$	Damper ratios for return air at time t	[0.25 0.75]	Var.	%
$D_{rf}(t)$	Damper ratios for fresh air at time t	[0.25 0.75]	Var.	%
$WiON\,(s)$	Open/close windows at time t	0 or 1	Var.	–
$LiON\,(s)$	On/off lights at time t	0 or 1	Var.	–
$T_r(t)$	Room temperature at time t	[16 30]	Var.	○C
$T_o(t)$	Outdoor temperature at time t	[20 36]	Var.	○C
$T_s(t)$	Supply air temperature at time t	–	Var.	○C
S	Laplace variable	–	Var.	–
$f(t_0), f(t_1), \ldots f(t_l)$	Indoor CO_2 concentration at time t	[550 1000]	Var.	ppm
F	Volumetric air flow rate	0.437	Par.	m^3/Sec.
Z	Zth order version of Lagrange	4	Par.	–
$v_r^\bullet$	The volume rate of the room	0.626	Par.	m^3/Sec.
v_r	The volume of the building	616	Par.	m^3
cp_a	Specific heat of air	1.005	Par.	$J/kg.\circ C$
K	Conductivity	0.7	Par.	–
Δx	Thickness	0.4	Par.	M
A	Surface area	173.6	Par.	m^2
ΔT_w	The difference between temperatures of water output and input	5	Par.	○C
cp_w	Specific heat of water	4200	Par.	$J/kg.\circ C$
τ_1	Time delay for the heat exchanger	4.7	Par.	Sec.
τ_2	Time delay for the CO_2 sensor	985.6	Par.	Sec.
τ_3	Time delay for the conditioned space	381.58	Par.	Sec.
$t_0 - t_l$	Time	[0 24]	Var.	Hours
T	Time	[0 24]	Var.	Hours
CO_2out	Outdoor CO_2 concentration	600	Par.	ppm
ρ	The air density	1.228	Par.	kg/m^3
A_1	The cross-sectional area	220	Par.	m^2
h	High of building	2.8	Par.	M
n	Number of air replaced times	4	Par.	1/sec
M_{He}	The mass of the heat exchanger	10	Par.	Kg
cp_{He}	The specific heat of the heat exchanger	0.4	Par.	$J/kg.^oC$
T_{win}	The inlet water temperature to the heat exchanger	10	Par.	^oC
T_{wout}	The outlet water temperature from heat exchanger	5	Par.	^oC
cp_w	The specific heat of the water	4200	Par.	$J/kg.^oC$
W_l	Power of Light	60	Par.	watt
$v_f^\bullet$	Volume rate of air filtration	–	Par.	m^3/sec.
$v_w^\bullet$	Volume rate of air comes from windows	–	Par.	m^3/sec.
$v_d^\bullet$	Volume rate of air comes from doors	–	Par.	m^3/sec.
$v_v^\bullet$	Volume rate of air ventilation	–	Par.	m^3/sec.
ACH	Average change per hr.	0.625	Par.	m^3/sec.
$C_p^\bullet$	CO_2 generation rate of a person	$4.9 \times 10{-}6$	Par.	m^3/sec.
n_p	Number of occupants	7	Par.	Person
FP	Fixed pricing	0.12	Par.	$/kWh
TOU	Time-of-use pricing	–	Par.	$/kWh
RTP	Real-time pricing	–	Par.	$/kWh

Note: (1) – Var. (Par.) denotes a variable (parameter), (2) $v_f^\bullet + v_w^\bullet + v_d^\bullet + v_{ven}^\bullet = ACH$ (Pita and Stevenson 1998).

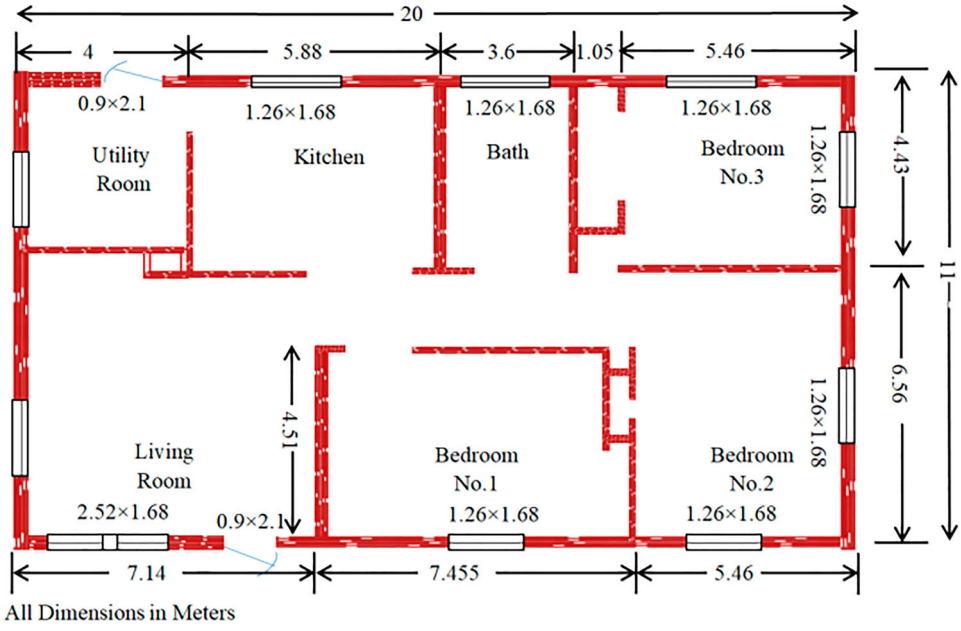


Figure 2. The plan of building used for simulation (case study 1) (Pita and Stevenson 1998).

space and then the simulations are carried out in MATLAB software.

3. The controllers' design

3.1. Model-based RL controller framework

RL is the body of algorithms that enables the agents to learn based on the reward mechanism. The agent interacts with its environment by observing a present state and taking action. After executing this action, the agent will receive a scalar reward. The agent aims to learn a mapping between the environmental conditions and the agent's actions for getting the preferred reward. Therefore, the central part of the RL controller is determining the state-actions space for both MDP and reward function (Noel and Pandian 2014), as shown in Figure 2.

3.1.1. States

State means the available information that is appropriate for decision-making by the RL agent. In this paper, three system states, as shown in Figure 3, are selected as follows: $T_r(t)$ = state (1), $T_o(t)$ = state (2), and indoor $CO_2(t)$ = state (3). Each state is discretized into a finite number due to the discretization of the agent's actions so that it is straightforward to converge optimal policy by the DP-RL algorithm. After that, the generalization capability of NARX-NN has been used to estimate the optimal control action for continuous states based on information provided on discretized states. Therefore, this reduces the discretization errors and obtains good performance. The

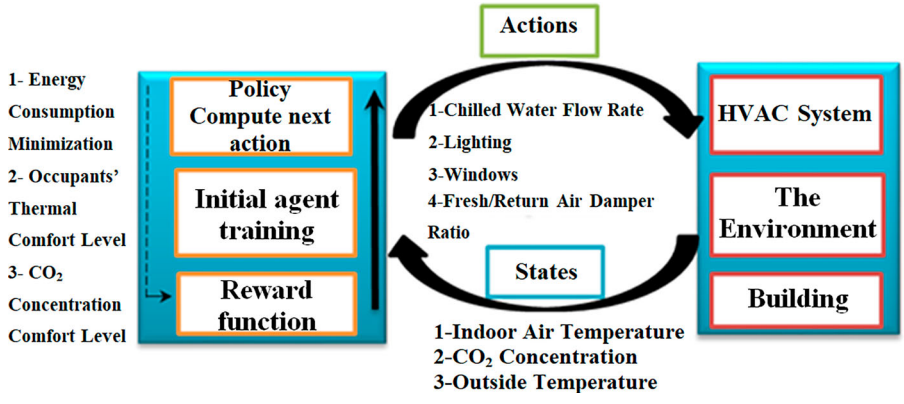


Figure 3. RL Based HVAC system key parts.

set of the discretized states, S, can be given as: $S = [T_r(t), To(t), CO_2(t)^T]$.

3.1.2. Actions

As shown in Figure 3, to control the HVAC system, four actions (action space) are considered as follows: Chilled water flow rate (ch(t)), Fresh/air damper ratio control ($D_{rf}(t)$), On/Off lighting (LiON(t)), and Open/Close windows (WiON(t)). So, the actions' vector variables (A) are: $A = [ch(t), D_{rf}(t), LiON(t), WiON(t)^T]$.

3.1.3. Reward

The reward function measures the success/failure of an agent's action for a specified state. As shown in Figure 3, three goals for controlling the HVAC system are considered: electrical energy consumption minimization, indoor air temperature control, and the CO_2 concentration level control. The reward function makes a trade-off between these goals.

Therefore, this research uses the violation as a punishment in the agent's reward function to deal with these constraints. The energy consumption reward part is considered negative because more energy consumption should punish the RL agent. It is designed to be an exponential (exp) function practiced on (ch(t)) calculations that means the importance of HVAC system On/Off switching. The reward parts for occupants' thermal comfort levels are designed to penalize $T_r(t)$ and $CO_2(t)$. The overall agent's reward is negative to reach a terminal state as quickly as possible. Therefore, the reward function can be given as:

$$R = -\exp(ch(t)) - \Omega \times \left[\left| \frac{2T_r(t) - \overline{T}_{r-des}(t) - \underline{T}_{r-des}(t)}{2} \right| + D_{rf}(t) \times \left| \frac{2CO_2(t) - \overline{CO}_{2-des}(t) - \underline{CO}_{2-des}(t)}{2} \right| \right]^2 \quad (5)$$

One can notice that the lighting action's impact does not appear in the reward function. Still, it can be realized on the building indoor temperature and the lighting action's impact included in the chilled water flow rate effect that appears in the reward function.

3.1.4. Value function

The value function denotes the value of a given state for an agent to be in it. The state-value function mainly characterizes the expected return from a specific state. The agent will obtain the discounted rewards-sum of progressively discounted rewards of states after executing a policy that starts from the initial state S(0) until it reaches the desired state S.

The relationship used for the state-V function for the given policy π can be expressed by the Bellman's equation as follows (Perera and Kamalaruban 2021):

$$V^{\pi}(s) = R(s, \pi(s)) + \gamma \sum P_{ss'} V^{\pi}(s') \quad (6)$$

The discount factor, $\gamma \in (0, 1)$, is used to adjust the achieved rewards during the convergence process. The policy describes the agent's method, and a mapping between S and A. $P_{ss'}$ is the state transition probability from state s to state s'. The optimal V is achieved by the agent that follows the optimal policy and is given as follows.

$$V^*(s) = \max_{\pi} V^{\pi}(s) \quad (7)$$

The optimal policy tends to maximize the total expected reward and can be expressed as follows:

$$\pi^*(s) = \arg\max_{a1 \in A} [R(s,a) + \gamma \sum P_{ss'} V^*(s')] \quad (8)$$

The RL agent uses the value iteration algorithm to discover the optimal control deterministic policy (DP). Then the indoor service optimal scheduling will be obtained based on this DP. The characteristics and factors employed in control algorithm implementation are shown in Table 3.

3.2. DP-RL method

The scheduling problem aims to adjust the building indoor services, which are comprised of four actions mentioned above along the day to minimize the electricity energy consumption.

Stochastic and deterministic are two different policy algorithms used in RL (Gao, Li, and Wen 2019; Hunt et al. 2016). In stochastic policy, both state-action spaces are integrated, but the DP type only combines the state space. RL-agent policy algorithm proceeds by sampling the DP and adjusting its parameters to obtain optimal control for building indoor services (Rijal, Humphreys, and Nicol 2018). After computing the optimal value function, using the value iteration algorithm, and improving the estimated value function using Bellman's equation, a DP algorithm is applied to calculate the optimal action-spaces scheduling and update the policy value parameters. In other words, a continuous set of action spaces can be used via a DP algorithm for optimal scheduling of any building indoor services within a short time. Consequently, the computational costs and time are reduced. Figure 4 explains the core of the DP-RL agent-environment interaction steps. From this figure, it can be seen that in each episode, the agent observes the state (*s*) and then chooses an action (*a*) using the deterministic policy (DP) algorithm. Then, the DP-RL agent selects an

Table 3. Description of parameters and variables used in controller design.

Component	Description	Value	Parameters/ Variables	Unit
γ	The discount factor	0.99	Par.	–
Ω	A trade-off factor between the energy (electricity) consumption of the reward part and occupants' comfort levels part	0.98	Par.	–
$\underline{T}_{r-des}(t)$ and $\overline{T}_{r-des}(t)$	The low and high desired indoor air temperature at time t	[20 24]	Var.	℃
$\underline{CO}_{2-des}(t)$ and $\overline{CO}_{2-des}(t)$	The low and high desired indoor CO_2 at time t	[750 850]	Var.	ppm.
R(s,a)	Reward function	–	Var.	–
$\gamma V^{\pi}(s')$	The future discounted rewards' summation	–	Var.	–
$V^{\pi}(s)$	Value function	–	Var.	–
$V^{*}(s)$	Optimal V	–	Var.	–
$\pi^{*}(s)$	Best policy	–	Var.	–

Note: – Var. (Par.) denotes a variable (parameter).

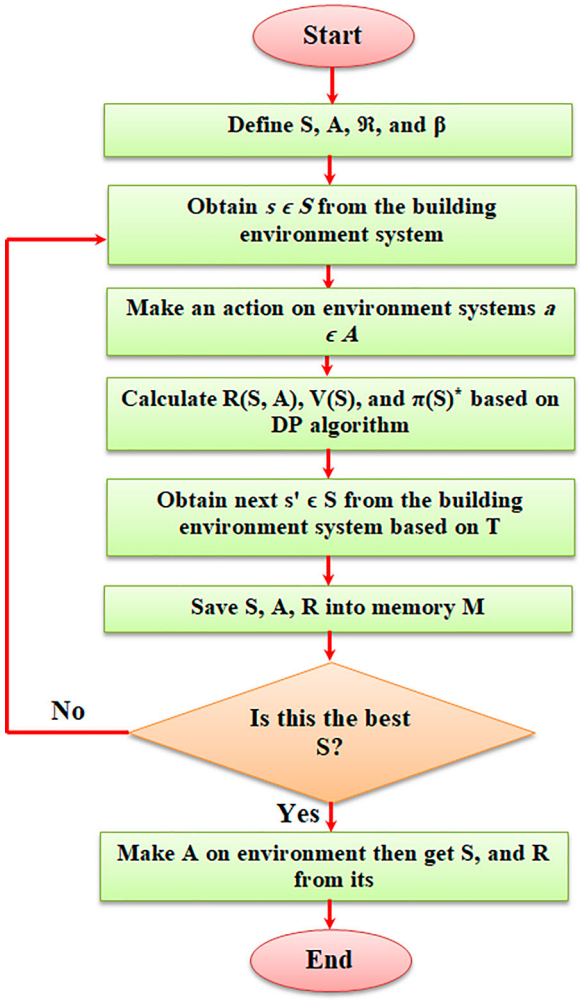


Figure 4. DP-RL algorithm flowchart.

action whose current V-value is the maximum. After acting, the DP-RL agent receives an immediate reward R(S, A), observes the following state (s'), updates the V-value, and calculates the optimal value using Eqns. (6, 7, and 8). This approach is carried out until the best state-space S is obtained. If this condition is not met, the DP-RL agent will repeat the above procedures for the next episode. Figure 5 shows the overall structure of DP- RL algorithm to control the HVAC system.

3.3. The DP-NARX-RL method

In this section, the structure of DP-NARX-RL method is described. The NARX-NN is based on the nonlinear autoregressive with exogenous inputs model. This recurrent dynamic, which consists of several layers and feedback, effectively handles continuous state-action spaces for the long-horizon forecast. The NARX-NN acts a nonlinear form recognition tool to decrease the errors of the DP-RL method (Raptodimos and Lazakis 2020). This type of NNs has been chosen because its learning process is more effective, generalizes better, and converges faster than the other NNs (Carbonera et al. 2021). In addition, in these NNs, the external inputs features improve the accuracy of the DP-NARX-RL controller (Ruiz et al. 2016).

The series-parallel neural network type is selected for NARX-NN. Here the amount of future output is predicted based on the present and past values of the inputs and the actual past values of the outputs. As shown in Figure 6, three high-dimensional input states, $T_r(t)$, To(t), $CO_2(t)$, and four specific control actions, ch(t), $D_{rf}(t)$, LiON(t), WiON(t), have been introduced as the inputs of the NARX-NN. Then, the optimal V has been estimated by the NARX-NN.

Based on the trial and error process, the number of hidden layers, neurons, and delays of the NARX-NN are selected. It has two hidden layers with ten neurons in each layer with two delay parameters. In addition, the learning rate is fixed to be 0.05. Therefore as shown in Figure 6, the NARX-NN structure consists of four layers: one input, one output, and two hidden layers with two delays of the target actions and two input delays of the external input states. The tangent sigmoid function is used as an activation function of the hidden layers' neurons and four linear neurons are used in the output layer.

The DP-NARX-RL controller can evaluate several control actions and then pick the best control actions for the next step, to satisfy the indoor air and thermal comfort levels with minimum electrical energy consumption.

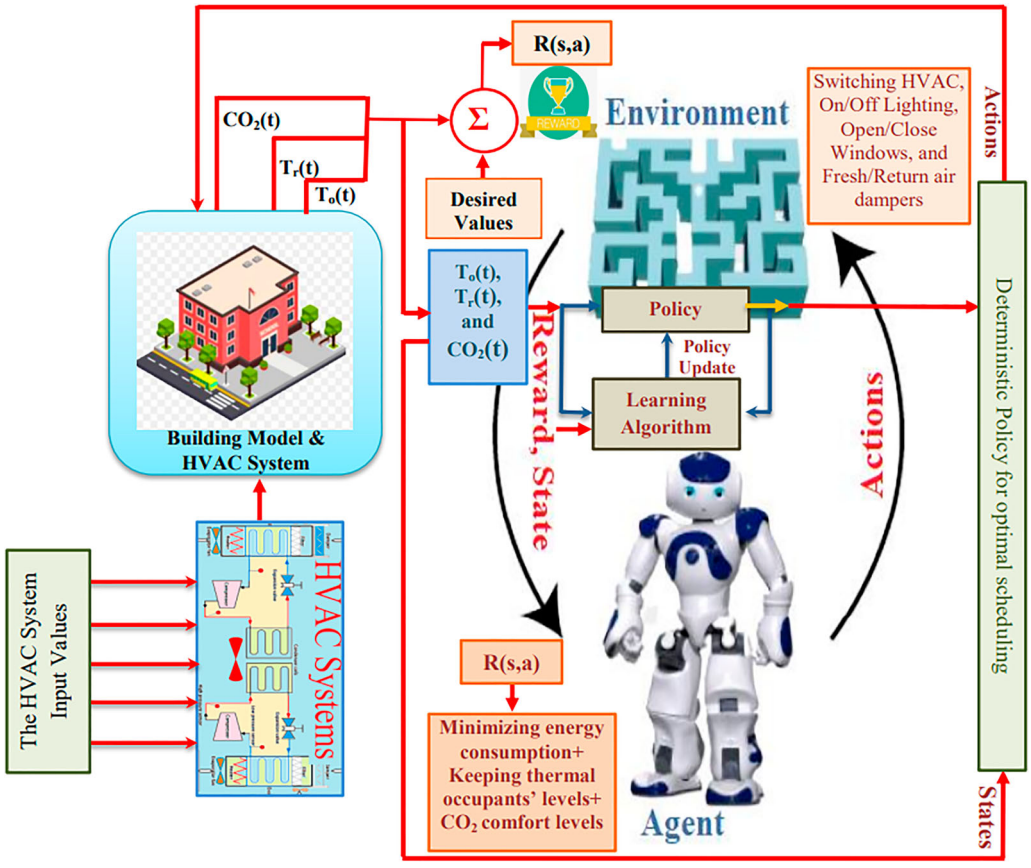


Figure 5. Overall structure of DP- RL algorithm for controlling the HVAC system.

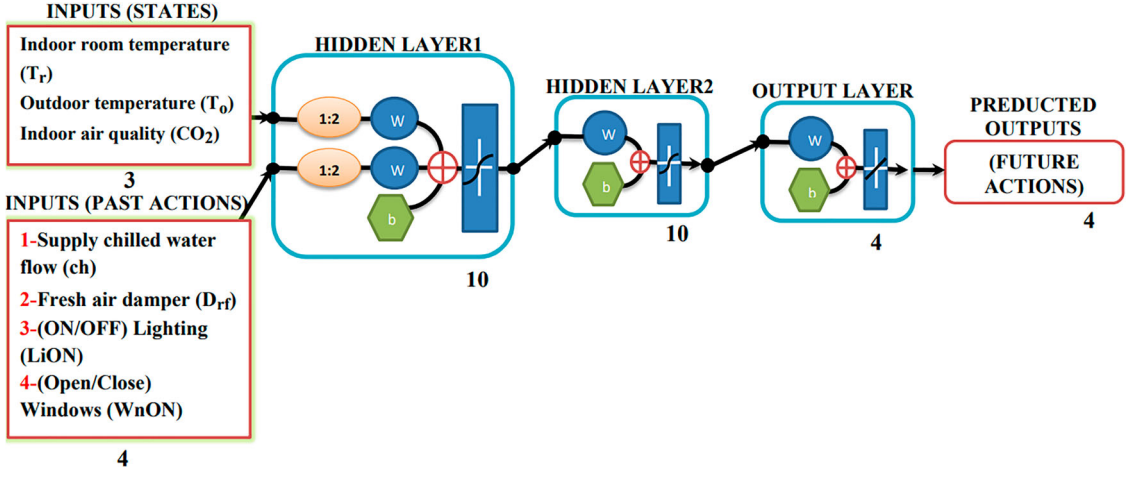


Figure 6. The series-parallel NARX-NN structure.

The DP-NARX-RL algorithm conducts the best control continuous actions in every control cycle by setting the HVAC system appropriate actuators. Simultaneously, the controller obtains the next HVAC system continuous state-space based on the control actions that have been performed. Indeed, in the DP-NARX-RL method, first, the optimal V is obtained by the NARX-NN, and then the best selected actions are computed by DP-RL. Table 4, shows the pseudo-code of DP-NARX-RL method. By implementation of this pseudo-code, the optimal scheduling of the HVAC can be determined.

The general algorithm for the DP-NRAX-RL controller connected to the HVAC system is depicted in Figure 7.

4. Results and discussions

In sections 4.1., 4.2., 4.3., and 4.4., *the case study 1* are analyzed; and section 4.5. describes the results for *case study 2*.

Table 4. The pseudo-code of the DP-NARX-RL method.

1.) *Define S, A, γ, Ω, and R (S, A).*
S← current states ($T_r(t)$, $T_o(t)$, and $CO_2(t)$)
A← set of possible actions ($ch(t)$, $D_{rf}(t)$, $LiON(t)$, and $WiON(t)$)
2.) *For each state S, set the initial guess for the policy of each action* $\pi = 0 and V^{\pi}(s) = 0$
3.) ***For** runs = 1–5*
4.) *Repeat for every discretized S and A.*
5.) *Repeat* $V^{\pi}(s) = R(s, \pi(s)) + \gamma \sum P_{ss'} V^{\pi}(s')$ and calculate $V^*(s')$
6.) *Based on DP- scheduling, repeat for each state* $\pi^*(s) = \arg\max_{a1 \in A}[R(s,a) + \gamma \sum P_{ss'} V^*(s')]$
End For
7.) *Use S and π^* to train the NARX-NN for starting the deep learning of the DP-NARX-RL method.*
8.) *Let $V^{*'}$ (S) be the approximation to V*(S) computed by the NARX-NN.*
9.) *Do the following for continuous control.*
a) Obtain $S_{current}$ from HVAC system environment
b) Calculate A' for $S_{current} = \pi^{*'}(S_{current})$, Updating the old policy of the agent with a new one
c) Set the HVAC system at $A'(S_{current})$
d) Go to a)

4.1. Evaluation of DP-RL controller performance

In this section, first, the performance of DP-RL algorithm applied to the HVAC system (case study 1) is evaluated. The simulations for the DP-RL algorithm have been carried out in MATLAB software. Reasonably, the value function and the best policy are dependent on state variables, $T_r(t)$ and $T_o(t)$, and $CO_2(t)$. However, the state $CO_2(t)$ has very little effect on value function and the best policy. Therefore, for simplicity the effects of $CO_2(t)$ on value function and best policy are not considered. Figure 8 shows the optimal value function versus $T_r(t)$ and $T_o(t)$ states. As shown in Figure 8, the surface of the optimal value function is smooth (Noel and Pandian 2014); therefore, the RL technique via DP policy selects more acceptable action values with flexibility.

By applying the DP-RL method to the HVAC system, the simulation results for controlling the indoor air temperature are shown in Figure 9. In Figure 9, the

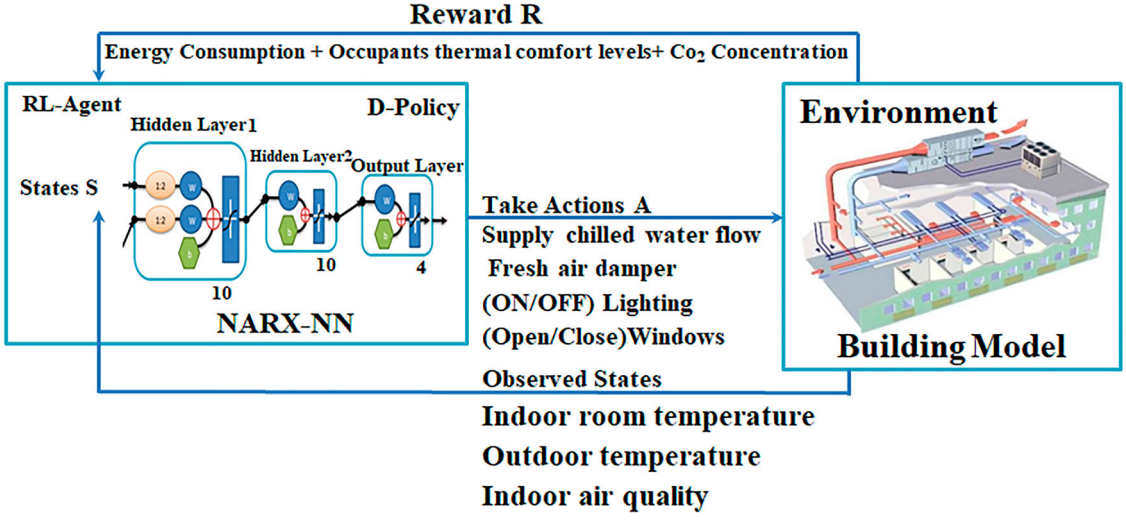


Figure 7. DP-NARX-RL controller diagram.

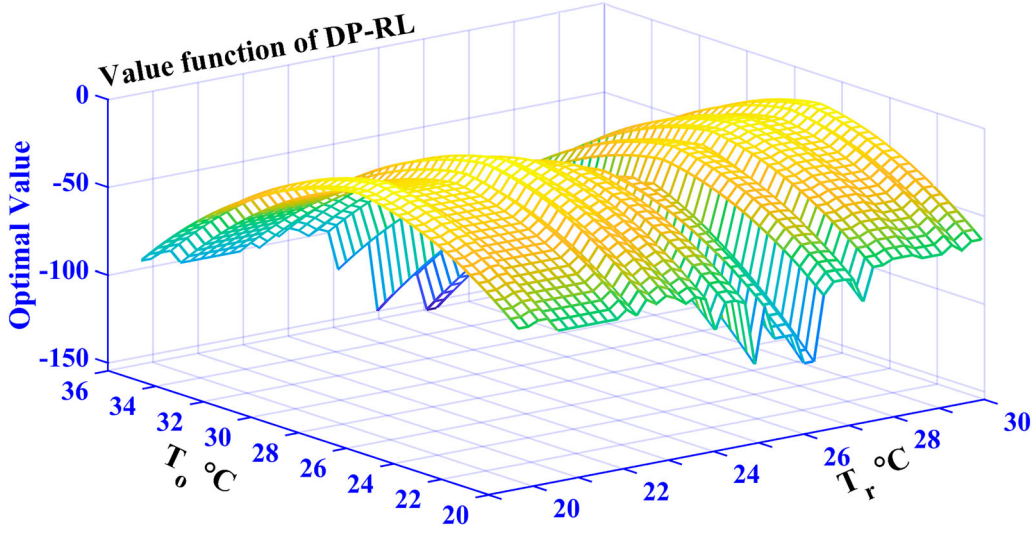


Figure 8. Optimal V learned by the DP-RL method.

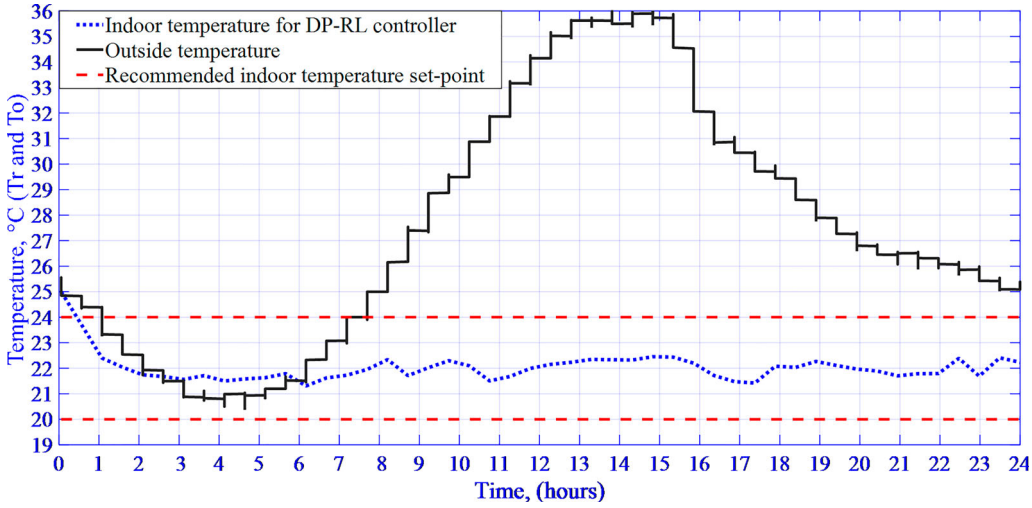
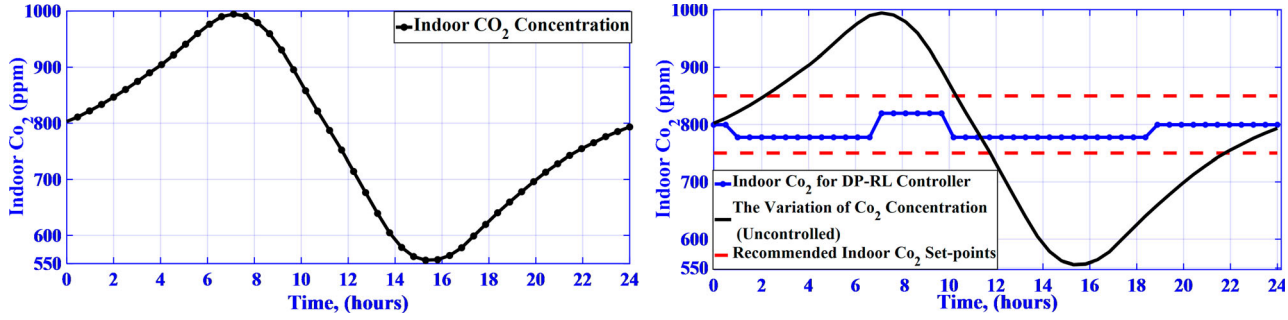


Figure 9. Indoor air temperature controlled by DP-RL. (a) Uncontrolled CO_2 concentration level (b) CO_2 concentration level controlled by the DP-RL



(a) Uncontrolled CO_2 concentration level **(b)** CO_2 concentration level controlled by the DP-RL

Figure 10. The variation of indoor CO_2 concentration vs. time.

outdoor temperature, and the up/down set points temperatures are also depicted. As shown in Figure 9, the DP-RL technique maintains the indoor temperature within the desired ranges.

CO_2 concentration level has been used for presenting the building IAQ occupants' comfort levels. Indoor acceptable CO_2 range varies between [550 and 1000] ppm, which is affected by indoor occupants' number and spent time (Baghaee and Ulusoy 2018; Wang, Wang, and Yang 2012). The ventilation upper limit corresponds to the AHU maximum performance, while the lower limit depends on the zoning occupancy (Ryzhov et al. 2019). Several time points have been used to illustrate CO_2 concentration level variations as shown in Figure 10(a). From 00:00 AM to 7:00 AM, the indoor CO_2 concentration exhibited an increasing trend (reaching the maximum concentration) because of people indoors such that it increases from 8:00 ppm to 10:00 ppm. From 7:00 AM to 3:30 PM, the indoor people started to leave the building, and the CO_2 level dropped rapidly to the minimum value (about 550 ppm). From 3:30 PM to 00:00 AM, the occupants entered the building, and the CO_2 increased again to around 800 ppm. In Figure 10(a), the CO_2 concentration dynamic curve is depicted for seven occupants (Yuan et al. 2021). It is reasonable that if the number of people or their spent time change, the indoor CO_2 concentration is also changed but the general shape of this figure is reserved.

Figure 10(b) shows the CO_2 concentration level controlled by the DP-RL in a day. As shown in Figure 10(b), the CO_2 concentration level is adjusted within the allowable range by controlling the HVAC system using the DP-RL technique.

4.2. Evaluation of DP-NARX-RL controller performance

This section shows the simulation results for HVAC control using the hybrid DP-NARX-RL method. The Bayesian regularization back-propagation algorithm has been chosen

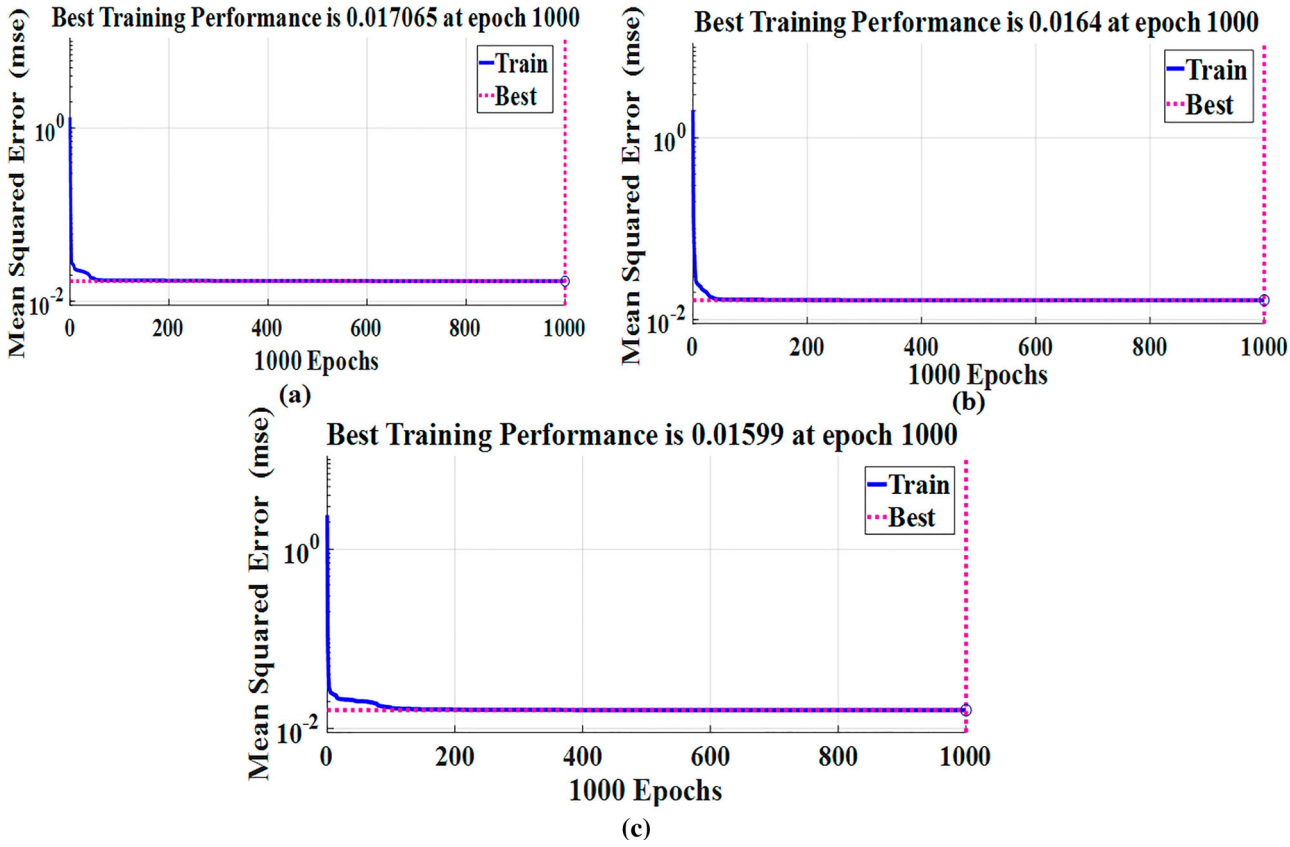


Figure 11. MSE vs. epoch index in NARX-NN training with 50% training data set (top left), 60% training dataset (top right), and 70% training dataset (bottom).

Table 5. MSE of different data set split scenarios for NARX-NN training.

Data split% (Training set: Test+ Validation sets)	MSE
50:50	0.0171
60:40	0.0164
70:30	0.01599

for NARX-NN training. This algorithm minimizes the mean square error and calculates the correct mixture to create a perfect network.

After conducting the training process, the NARX-NN tracks the output. Figure 11(a–c) show the NARX-NN output tracking with different mean squared errors (MSE) versus the number of epochs for (training data, test data + validation data) = (50%, 50%), (60%, 40%), and (70%, 30%), respectively. The MSE for different data set splits is given in Table 5. As shown in Table 5, the MSE value is the minimum for (training data, test data + validation data) = (70%, 30%). Therefore, this combination is selected for the NARX-NN training.

The discretization errors can be decreased using the NARX-NN capacity to predict the correct control target actions for high-dimensions input states through the available information. Figure 11(c) displays that the performance of NARX-NN has been improved with a high level of accuracy during the training process. During the training NARX-NN and for (training data, test data + validation data) = (70%, 30%), the gradient error variation, μ value, and validation checks are displayed in Figure 12. Gradient (4.4771e-5) characterizes the tangent slope of a function graph and indicates how much change occurs in the error rate; 'μ' is the back-propagation index for the NARX-NN. A validation check is used to terminate the NARX-NN learning by updating each iteration.

Table 6 describes the parameters used for NARX-NN training. In this table, 70% dataset was used for the training process, 15% data was used for testing the results, and 15% dataset was used for the validation process.

Figure 13 shows the optimal value function versus $T_r(t)$ and $T_o(t)$ states using the DP-NARX-RL method. As shown

Table 6. Description of NARX-NN training parameters.

Description	Value
The maximum number of epochs	1000
Performance goal	0.0000001
Learning rate	0.05
Training, validation, test set ratio	70%–15%–15%
Performance function	MSE
Training algorithm	Trainbr

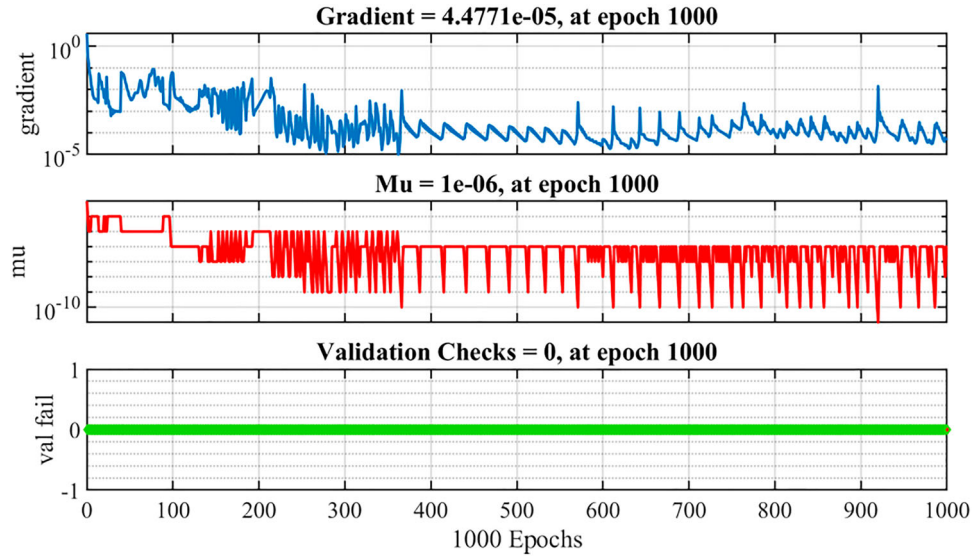


Figure 12. Variation of gradient error and validation checks using NARX-NN.

in Figure 13, the surface of the optimal value function is smoother as compared with the optimal value function of the DP-RL method that is shown in Figure 8 (Noel and Pandian 2014). This feature creates excellent HVAC system behaviour.

The smoother surface of optimal V reduces the oscillations of actions executed by the DP-NARX-RL method. This minimizes the chattering effects on the solenoid valve and air dampers and consequently leading the agent to achieve its goal as faster as possible.

In DP-NARX-RL method, the simulation results for the set of action spaces are shown in Figure 14(a–d). Figure 14(a) represents supplied chilled water valve position. According to the desired states, it is the first control action that adjusted by the DP-NARX-RL method and changed directly based on the temperature. Figure 14(b) shows the position damper for the fresh air. It maintains the measured airflow rate within its desired set point level and provides good air quality inside the building. Figure 14(c) shows the control action of the DP-NARX-RL method for lighting on/off. The control action of the DP-NARX-RL method for opening/closing windows is shown in Figure 14(d).

The occupants' comfort levels of opening/closing windows have been adjusted by the DP-NARX-RL method for increasing/decreasing in the $T_r(t)$ and associated with raising/decreasing the $T_o(t)$. If the $T_o(t)$ falls below the upper set-point value, the window must be opened and vice versa. The on/off optimal lighting control has been provided and regulated by the agent according to the occupants' behaviour setting time (optionally) using the DP-NARX-RL strategy.

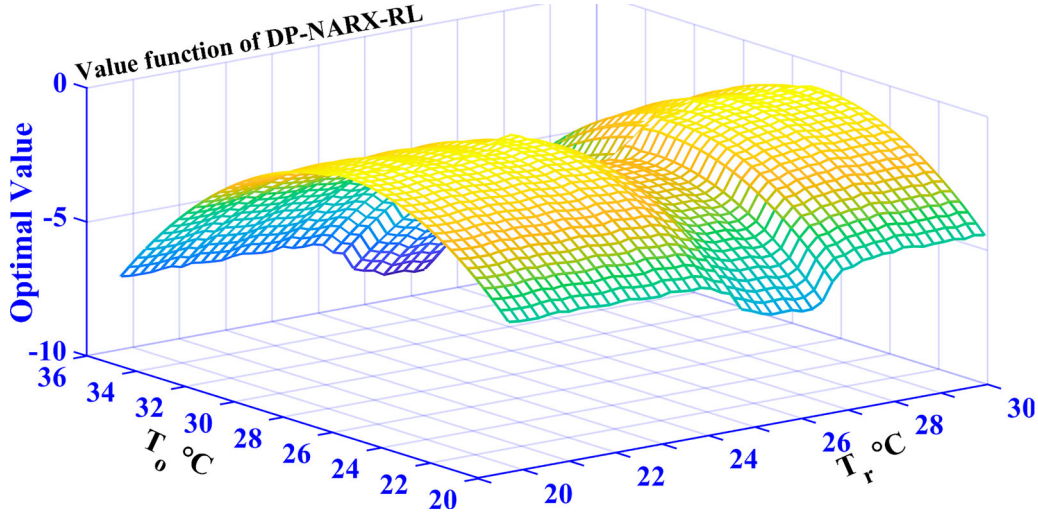


Figure 13. Optimal value function learned by the hybrid DP-NARX-RL method.

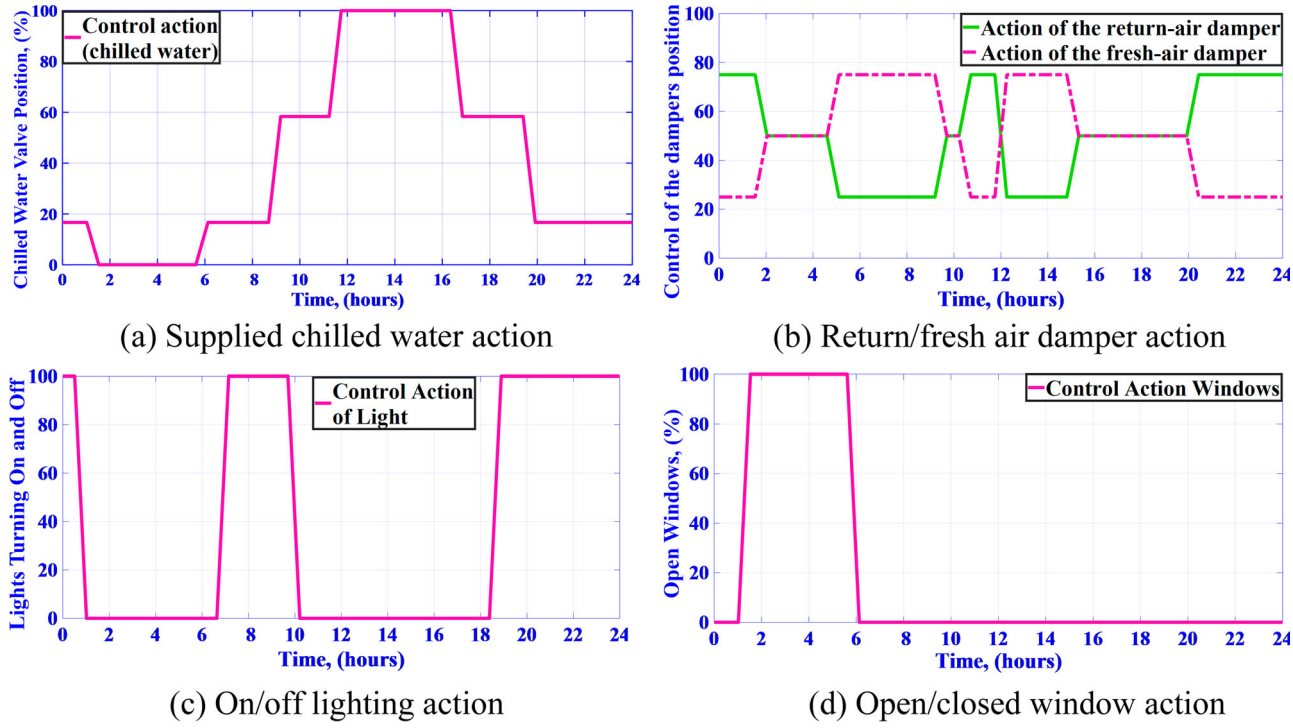


Figure 14. The set of actions designed by the DP-NARX-RL method

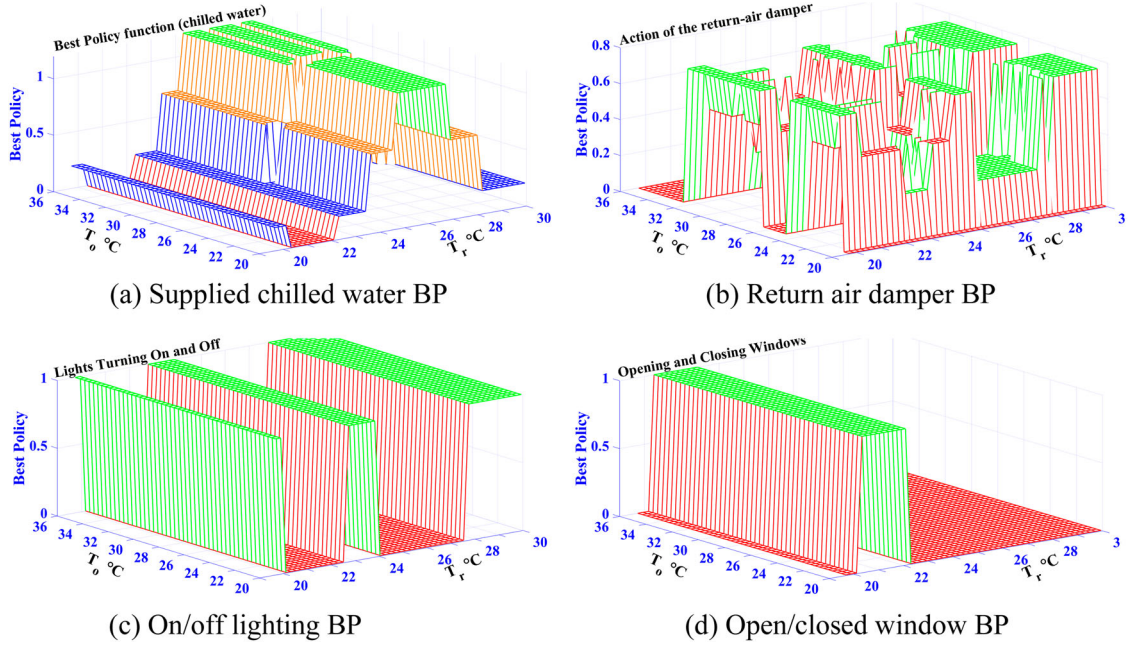


Figure 15. The set of the best policy (BP) functions learned by the DP-NARX-RL method.

The policy functions for the agent actions are optimized by the DP-NARX-RL learning, as shown in Figure 15. Figures 15(a–d) show the best policy for the supplied chilled water, the damper position of the return air action, lighting on/off, and opening/closing windows, respectively.

As mentioned above, These best policy functions can be presented based on three state variables: $T_r(t)$, $T_o(t)$, and $Co_2(t)$, but clearly, the most effective states on the environmental performance of thermal power are $T_r(t)$ and $T_o(t)$.

By applying the DP-NARX-RL method, the simulation results for controlling the indoor temperature and CO_2 concentration level are shown in Figure 16 and Figure 17, respectively. In Figure 16, the outdoor temperature, the indoor temperature controlled by the DP-NARX-RL

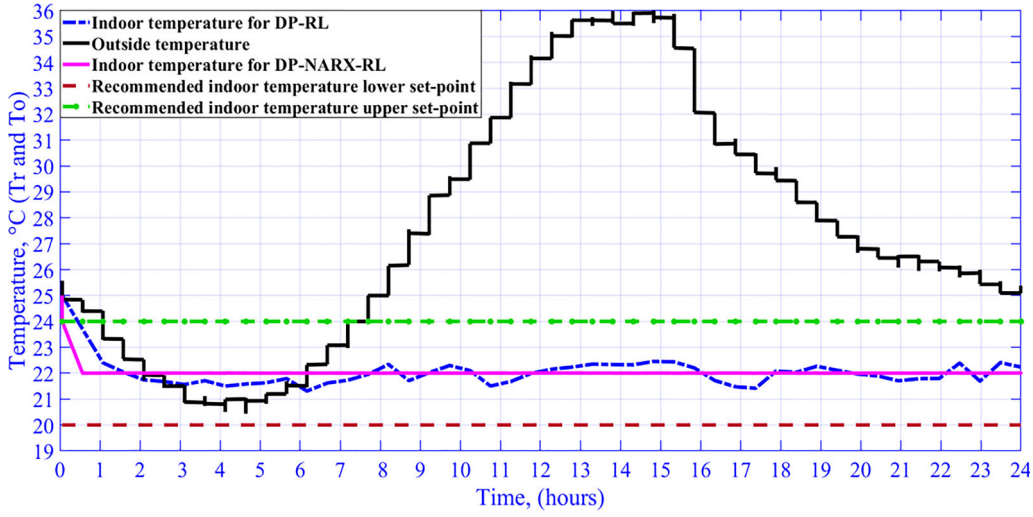


Figure 16. Indoor temperature controlled by the DP-RL and DP-NARX-RL methods.

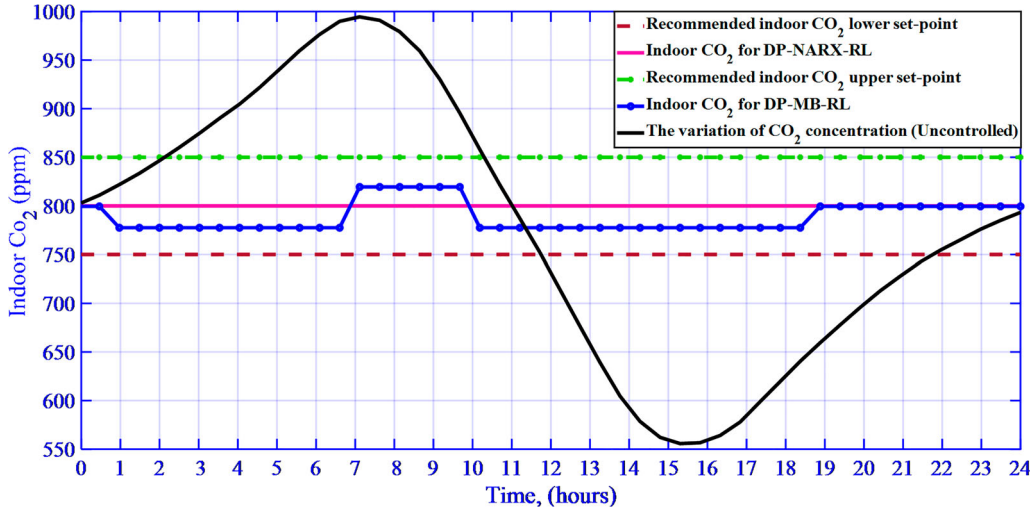


Figure 17. The CO_2 Concentration level controlled by the DP-RL and DP-NARX-RL methods.

method, and the up/down set points temperatures are depicted. As shown in Figure 16, the controller maintains the indoor temperature within desired comfort levels very well. The indoor temperature has low oscillations around the desired set-point values compared with those under the DP-RL controller.

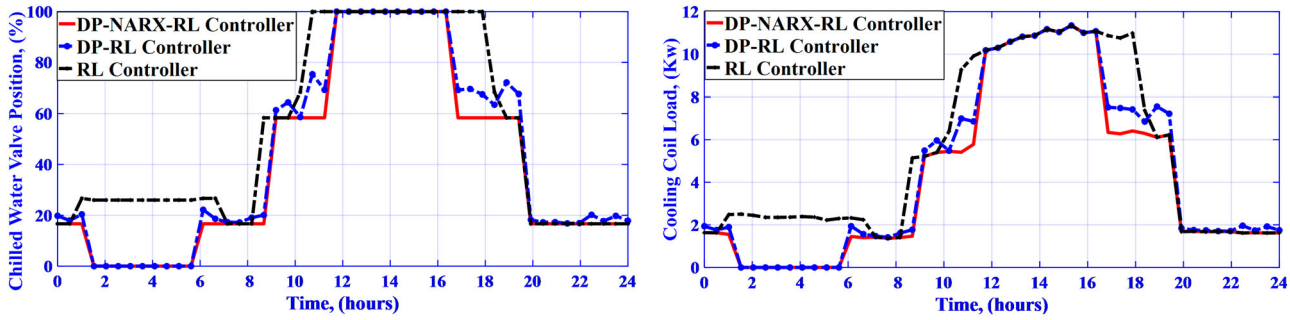
Furthermore, the DP-NARX-RL method has a better performance than the DP-RL controller. Figure 17 depicts the concentration CO_2 level controlled by the DP-NARX-RL method. As shown in Figure 17, the CO_2 concentration level is controlled within the allowable range.

4.3. Energy saving evaluation

To evaluate the energy efficiency and performance of the proposed controllers, the electrical energy consumed by the HVAC systems for 24 hrs is computed. The performance of DP-RL and DP-NARX-RL controllers is compared with the RL controller, although the RL controller is out of the scope of this paper. To calculate the electrical energy consumption of the HVAC system, the actions are investigated for the opening position of the cooling coil valves, the return and fresh air dampers, windows, and on/off lighting.

The agent signal of the chilled water valve position is employed for computing the flow rate of chilled water to obtain energy cooling load, as shown in Figure 18(a). The flow rate of chilled water is used for calculation of the cooling load displayed in Figure 18(b).

In this study, the electrical loads for all of the fans and pumps are also under uncontrolled parameters of the agent, so they were not taken into consideration. Furthermore, such loads are constant in a specific building, it can typically account for around 1% of energy usage in HVAC systems (Ding, Du, and Cerpa 2019). As for the



(a) Action of the supplied chilled water (b) Cooling coil load

Figure 18. The agent signal positions for computing the energy consumption of the HVAC system under the DP-NARX-RL, DP-RL, and RL controllers.

Table 7. Summary of the cooling load results for energy consumption calculations.

Component	Description	Value	Parameters/ Variables	Unit
∆m CW	The incremental amount in the steady-state	0.1	Par.	(kg/s)
∆CL	Cooling load incremental	0.166667	Par.	(kW)

electricity consumption of the chiller, it has been implied as a function of the control signal of the chilled water valve position.

Iterative approaches in MATLAB code have been used to calculate the cooling coil load (energy usage) by implementing the chilled water temperature difference ∆T between the inlet and outlet of the cooling coil. Furthermore, the data in Table 1 and Table 7 are used to calculate the energy profile, as depicted in Figure 18(b).

Figure 19 shows the cumulative electrical energy consumption for 24 hrs for three methods: RL, DP-RL, and DP-NARX-RL. By applying the controllers RL, DP-RL, and DP-NARX-RL, the energy consumed by the HVAC systems are 127, 108, and 101 kWh/day, respectively. The calculation results demonstrate their ability to distinguish differences in the energy consumption for the same building under the three different controllers (RL, DP-RL, and DP-NARX-RL). As shown in Figure 19, when the DP-NARX-RL method is applied, the building electrical energy consumption for a day is reduced by 20.5% compared with the RL method.

The reduction of energy consumption in the building used the DP-NARX-RL algorithm is due to the temperature dropping at night. Therefore, the deterministic policy exploits this opportunity to open windows and turn off indoor/outdoor nighttime running lights.

Usually, the energy-saving of different deep RL and DDP Gradient-RL controllers is within the range (5-15) % compared to existing control methods (Du et al. 2021; Ding, Du, and Cerpa 2019; Lissa et al. 2021; Zhang, Zhang, and Lam 2018). The energy-saving of the DP-NARX-RL controller is 20.5% compared with RL controller. Therefore, this result is remarkable compared with (Du et al.

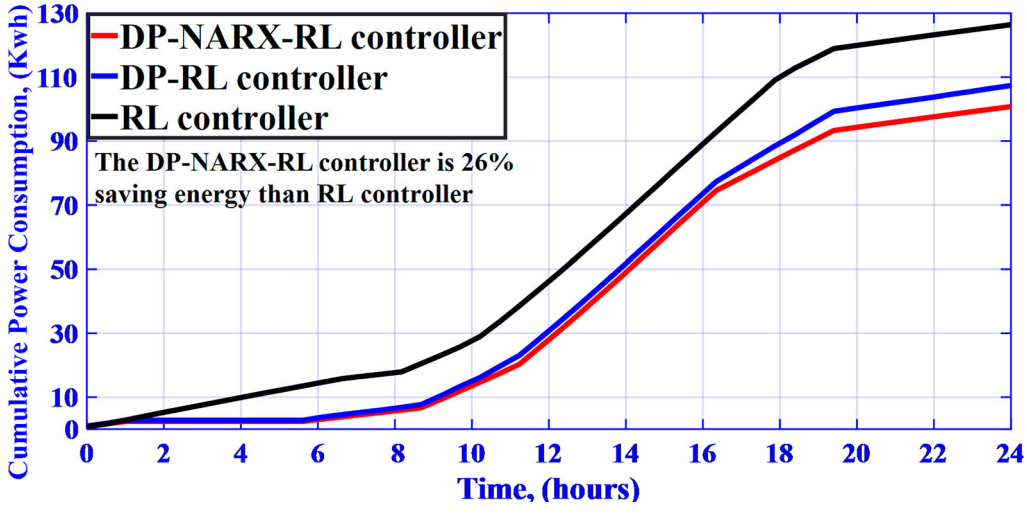


Figure 19. Energy consumed of the HVAC system using DP-NARX-RL, DP-RL, and RL controllers.

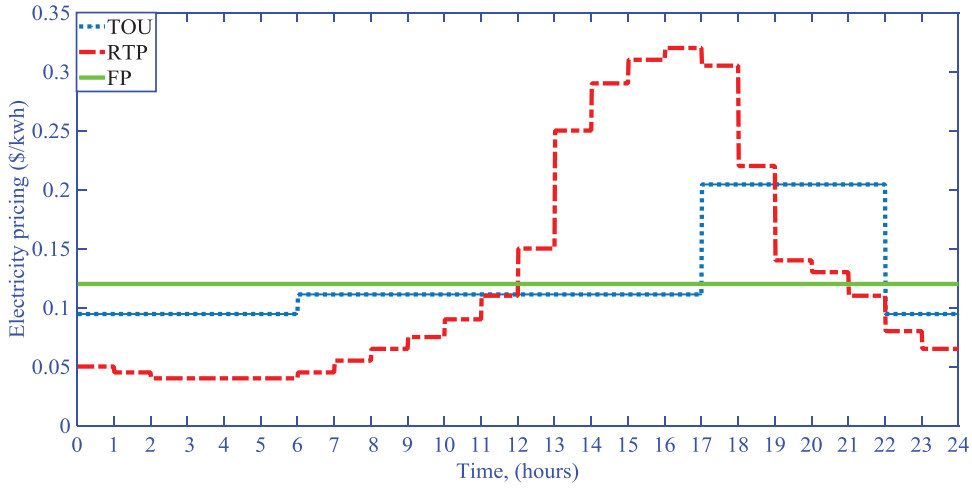


Figure 20. Electricity pricing schemes.

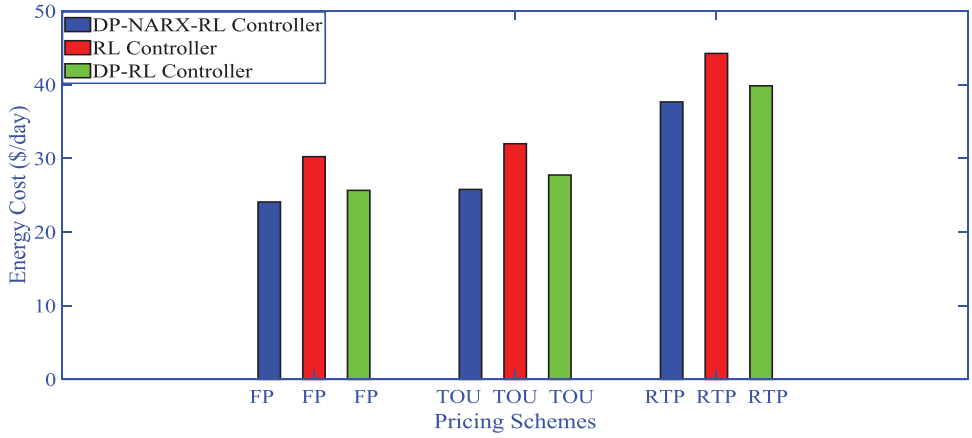


Figure 21. Comparison of energy costs for different controllers and pricing schemes.

2021; Ding, Du, and Cerpa 2019; Lissa et al. 2021; Zhang, Zhang, and Lam 2018).

It is necessary to mention that although the energy savings of the DP-NARX-RL method is 6.5% compared to the DP-RL approach. Still, the performance of the DP-NARX-RL is better than that of the DP-RL method concerning indoor temperature and air quality oscillations. The smoother surface value function under the proposed DP-NARX-RL controller shows that the better optimization can be achieved.

4.4. Energy cost evaluation

To evaluate the performance of different controllers, three electricity pricing schemes, fixed pricing (FP), real-time pricing (RTP), and time-of-use pricing (TOU) are used for cost analysis. Figure 20 shows electricity pricing for a typical day (Talebi and Hatami 2020). By applying the DP-NARAX-RL, DP-RL, and RL controllers to the of HVAC system, the energy costs are calculated based on the electricity pricing ($/kWh) and the cooling coil loads (kW) and depicted in Figure 21.

As shown in Figure 21, the DP-NARAX-RL method has a better performance compared to other controllers as it runs the HVAC system with lower cost for FP, TOU, and RTP schemes. For example as shown in Table 8, the proposed controller has a reduction of 14.1%, 19.4%, and 20.3% in energy cost compared to RL controller for RTP, TOU and FP schemes, respectively.

4.5. Comparing the performance of the DP-NARX-RL controller and PID controller

The performance of the DP-NARX-RL approach and the PID controller reported by Hussein, Ateeq, and Homod (2022) is compared by controlling a multi-chiller of Basra International Airport, Iraq for a day (*case study 2*). The results of both controllers from different aspects are explained as follows.

Figure 22 shows the performance of PID and DP-NARX-RL controllers for controlling indoor air temperature.

Table 8. Energy cost comparison for different controllers.

		Energy Cost ($/day)		
Controllers		RL	DP-RL	DP-NARX-RL
Pricing schemes ($/kWh)	FP	30.24 (100%)	25.67 (84.9%)	24.09 (79.7%)
	TOU	32 (100%)	27.75 (86.7%)	25.79 (80.6%)
	RTP	44.26 (100%)	39.86 (90%)	37.68 (85.1%)

From Figure 22, it is evident that both controllers keep the indoor air temperature within the acceptable range but the DP-NARX-RL controller has a better performance in terms of indoor air temperature low fluctuations. Still, indoor air temperature controlled by the DP-NARX-RL controller is more stable than the benchmark controller.

Figure 23 shows the cooling coil loads of multi HVAC systems for different controllers. As shown in Figure 23, the DP-NARX-RL controller has a better performance than other controllers reported by Hussein, Ateeq, and Homod (2022) because the proposed controller absorbs less power than other controllers.

The energy-saving evaluations of controllers are depicted in Figure 24 for a multi-chiller system. Figure 24 represents the cumulative energy consumptions of a day for the DP-NARX-RL and other controllers reported by Hussein, Ateeq, and Homod (2022). As can be seen in Figure 24, the DP-NARX-RL controller has a better performance because it consumes 38% less energy than PID controller.

Figure 25 illustrates the energy cost of a day for different controllers in various pricing schemes. The energy costs are calculated based on the cooling coil loads (Figure 23) and electricity pricing schemes (Figure 20). As shown in Figure 25, compared with the PID controller, the DP-NARX-RL controller has reduced the energy cost by 38%, 30%, and 15% for FP, TOU, and RTP schemes, respectively.

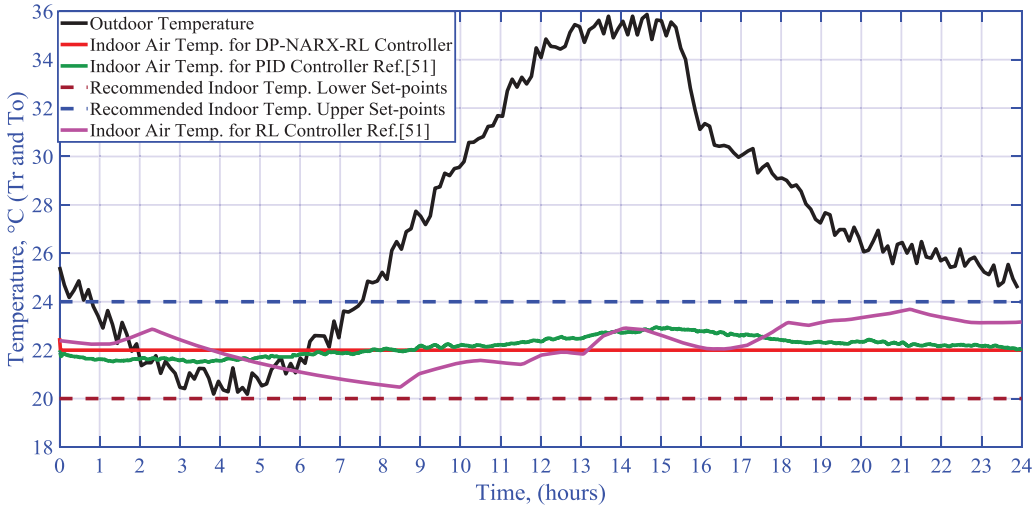


Figure 22. The performance of DP-NARX-RL and PID controllers applied to a multi HVAC system of the Basra International Airport.

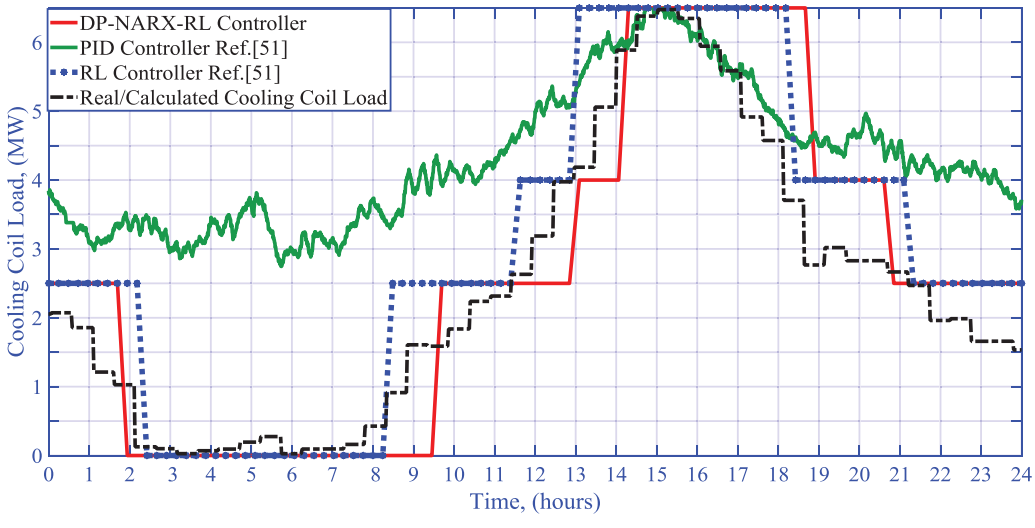


Figure 23. Comparison of cooling coil loads for a multi HVAC system by applying different controllers.

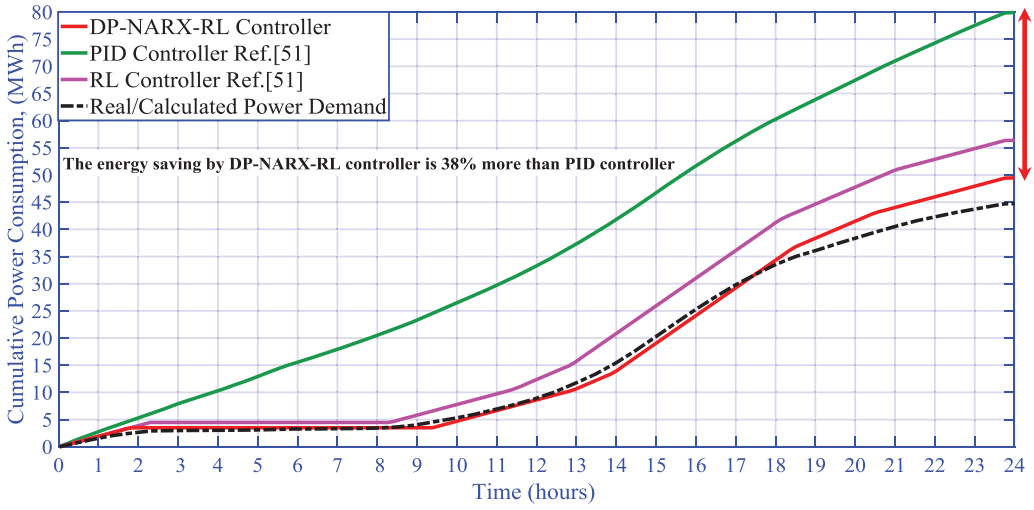


Figure 24. Cumulative energy consumption for different controllers.

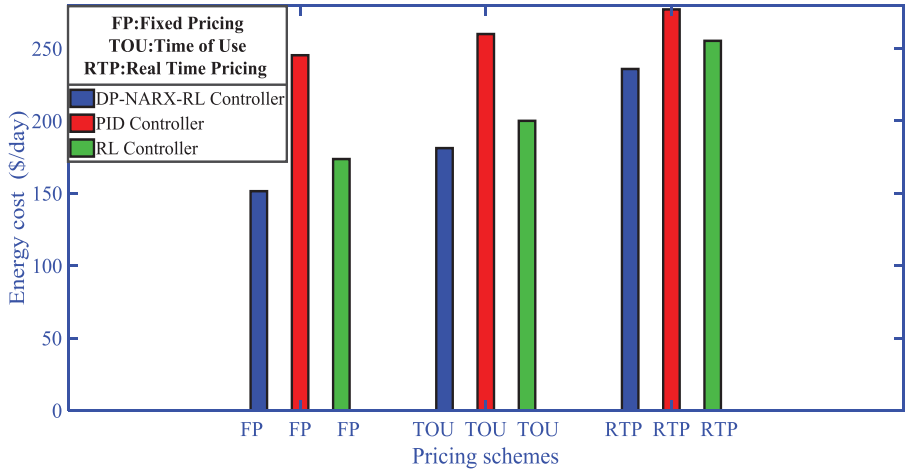


Figure 25. The energy cost for different controllers in various pricing schemes.

5. Conclusion

This paper presented new control approaches using the model-based reinforcement learning (RL) to minimize electrical energy consumption and maintain the occupants' comfort levels within the desired ranges, respectively. The indoor temperature and CO_2 concentration level were considered as the occupants' thermal comfort levels. To control the CO_2 concentration level, the classical building model was modified. The Lagrange polynomials were used for CO_2 concentration levels modeling. Based on model-based RL, two controllers were presented. The first was a deterministic policy (DP) with a model-based RL (DP-RL) and the other employed a hybrid of the nonlinear autoregressive exogenous neural network (NARX-NN) and an DP-RL method (DP-NARX-RL). By selecting appropriate components for reward functions, a trade-off was made between maintaining the indoor comfort levels and minimizing the electrical energy consumption of the HVAC system. To evaluate the performance of controllers, simulations of two case studies were conducted in MATLAB environment. In case study 1, the HVAC system of a building in Basra, Iraq was analyzed from different aspects. The results show that both approaches, DP-RL, and DP-NARX-RL, have kept the indoor comfort levels within the desired ranges, but the DP-NARX-RL provided a more stable indoor air temperature. At the same time, the daily energy consumption of the DP-NARX-RL method had a reduction of 20.5% and 6.5% compared to RL and DP-RL, respectively. In addition, the energy cost of DP-NARX-RL method has reduced considerably with respect to RL and DP-RL methods and for different pricing schemes. As a result, the DP-NARX-RL method has a better performance compared to DP-RL from different aspects. In case study 2, a multi-chiller system of Basra International Airport, Iraq was controlled by DP-NARX-RL and PID methods and the results were assessed from different

aspects such as thermal comfort conditions, energy consumption, and energy cost for various pricing schemes. The results show that DP-NARX-RL has a saving energy of 38% compared to the benchmark controller. At the same time, the energy cost for DP-NARX-RL is lower than PID method for different pricing options such as Fixed pricing, Time-of-use pricing, and real-time pricing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Alireza Hatami http://orcid.org/0000-0002-0370-3903
Raad Z. Homod http://orcid.org/0000-0002-4161-7539

References

Ahn, K. U., and C. S. Park. 2019. "Application of Deep Q-Networks for Model-Free Optimal Control Balancing Between Different HVAC Systems." *Science and Technology for the Built Environment* 26 (1): 61–74. doi:10.1080/23744731.2019.1680234.

Alawadi, S., D. Mera, M. Fernández-Delgado, F. Alkhabbas, C. M. Olsson, and P. Davidsson. 2020. "A Comparison of Machine Learning Algorithms for Forecasting Indoor Temperature in Smart Buildings." *Energy Systems*, 1–17. doi:10.1007/s12667-020-00376-x.

Azuatalam, D., W. L. Lee, F. de Nijs, and A. Liebman. 2020. "Reinforcement Learning for Whole-Building HVAC Control and Demand Response." *Energy and AI* 2: 1–21. doi:10.1016/j.egyai.2020.100020.

Baghaee, S., and I. Ulusoy. 2018. "User Comfort and Energy Efficiency in HVAC Systems by Q-Learning". In *2018 26th Signal Processing and Communications Applications Conference (SIU)*, 1–4. IEEE. doi:10.1109/SIU.2018.8404287.

Carbonera, L. F. B., D. Pinheiro Bernardon, D. de Castro Karnikowski, and F. Alberto Farret. 2021. "The Nonlinear Autoregressive Network with Exogenous Inputs (NARX) Neural Network to Damp Power System Oscillations." *International Transactions on Electrical Energy Systems* 31 (1): e12538.

Chapra, S. C., and R. P. Canale. 2015. *Numerical Methods for Engineers*. 7th Ed. New York, NY: McGraw Hill.

Chen, B., Z. Cai, and M. Bergés. 2019. "Gnu-RL: A Precocial Reinforcement Learning Solution for Building HVAC Control Using a Differentiable MPC Policy". *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, pp. 316–325. doi:10.1145/3360322.3360849.

Ding, X., W. Du, and A. Cerpa. 2019. "Octopus: Deep reinforcement learning for holistic smart building control". In *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, 326–335. doi:10.1145/3360322.3360857.

Du, Y., H. Zandi, O. Kotevska, K. Kurte, J. Munk, K. Amasyali, and F. Li. 2021. "Intelligent Multi-Zone Residential HVAC Control Strategy Based on Deep Reinforcement Learning." *Applied Energy* 281: 116117. doi:10.1016/j.apenergy.2020.116117.

Gao, G., J. Li, and Y. Wen. 2019. "*Energy-Efficient Thermal Comfort Control in Smart Buildings Via Deep Reinforcement Learning*". arXiv preprint arXiv:1901.04693.

Gao, G., J. Li, and Y. Wen. 2020. "Deep Comfort: Energy-Efficient Thermal Comfort Control in Buildings via Reinforcement Learning." *IEEE Internet of Things*, 1–13. doi:10.1109/JIOT.2020.2992117.

Hao, J., D. W. Gao, and J. J. Zhang. 2020. "Reinforcement Learning for Building Energy Optimization Through Controlling of Central HVAC System." *IEEE Journal of Power and Energy* 7: 320–328. doi:10.1109/OAJPE.2020.3023916.

Homod, R. Z., A. Almusaed, A. Almssad, M. K. Jaafar, M. Goodarzi, and K. S. Sahari. 2021. "Effect of Different Building Envelope Materials on Thermal Comfort and air-Conditioning Energy Savings: A Case Study in Basra City, Iraq." *Energy Storage* 34: 101975. doi:10.1016/j.est.2020.101975.

Homod, R. Z., K. S. Gaeid, S. M. Dawood, A. Hatami, and K. S. Sahari. 2020. "Evaluation of Energy-Saving Potential for Optimal Time Response of HVAC Control System in Smart Buildings." *Applied Energy* 271: 115255. doi:10.1016/j.apenergy.2020.115255.

Homod, R. Z., K. S. M. Sahari, and H. A. Almurib. 2014. "Energy Saving by Integrated Control of Natural Ventilation and HVAC Systems Using Model Guide for Comparison." *Renewable Energy* 71: 639–650.

Homod, R. Z., K. S. M. Sahari, H. A. Almurib, and F. H. Nagi. 2011. "Double Cooling Coil Model for Nonlinear HVAC System Using RLF Method." *Energy and Buildings* 43 (9): 2043–2054. doi:10.1016/j.enbuild.2011.03.023.

Homod, R. Z., K. S. M. Sahari, H. A. Mohamed, and F. Nagi. 2010. "Hybrid PID-Cascade Control for HVAC System." *International Journal of Systems Control* 1 (4): 170–175.

Homod, R. Z., H. Togun, H. J. Abd, and K. S. Sahari. 2020. "A Novel Hybrid Modelling Structure Fabricated by Using Takagi-Sugeno Fuzzy to Forecast HVAC Systems Energy Demand in Real-Time for Basra City." *Sustainable Cities and Society* 56: 102091. doi:10.1016/j.scs.2020.102091.

Hosseinloo, A. H., A. Ryzhov, A. Bischi, H. Ouerdane, K. Turitsyn, and M. A. Dahleh. 2020. "Data-driven Control of Microclimate in Buildings: An Event-Triggered Reinforcement Learning Approach." *Applied Energy* 277: 115451. doi:10.1016/j.apenergy.2020.115451.

Hunt, J. J., A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra. 2016. "Continuous learning control with deep reinforcement". *International Conference on Learning Representations (ICLR)*.

Hussein, A. L., A. A. Ateeq, and Z. R. Homod. 2022. "Energy Saving by Reinforcement Learning for Multi-Chillers of HVAC Systems". *Proceedings of 2nd International Multi-Disciplinary Conference Theme: Integrated Sciences and Technologies*, IMDC-IST 2021, 7–9 September 2021, Sakarya, Turkey. EAI. doi:10.4108/eai.7-9-2021.2315301.

Kurte, K., J. Munk, O. Kotevska, K. Amasyali, R. Smith, E. McKee, and H. Zandi. 2020. "Evaluating the Adaptability of Reinforcement Learning Based HVAC Control for Residential Houses." *Sustainability* 12 (18): 7727. doi:10.3390/su12187727.

Lissa, P., C. Deane, M. Schukat, F. Seri, M. Keane, and E. Barrett. 2021. "Deep Reinforcement Learning for Home Energy Management System Control. Energy and AI", 3, 100043. doi:10.1016/j.egyai.2020.100043.

Marantos, C., C. P. Lamprakos, V. Tsoutsouras, K. Siozios, and D. Soudris. 2018. "Towards Plug & Play Smart Thermostats Inspired by Reinforcement Learning". *Proceedings of the Workshop on Intelligent Embedded Systems Architectures and Applications*, 39–44. doi:10.1145/3285017.3285024.

Moubayed, A., M. Injadat, A. B. Nassif, H. Lutfiyya, and A. Shami. 2018. "E-learning: Challenges and Research Opportunities Using Machine Learning & Data Analytics." *IEEE Access* 6: 39117–39138. doi:10.1109/ACCESS.2018.2851790.

Noel, M. M., and B. J. Pandian. 2014. "Control of a Nonlinear Liquid Level System Using a new Artificial Neural Network-Based Reinforcement Learning Approach." *Applied Soft Computing* 23: 444–451. doi:10.1016/j.asoc.2014.06.037.

Perera, A. T. D., and P. Kamalaruban. 2021. "Applications of Reinforcement Learning in Energy Systems." *Renewable and Sustainable Energy Reviews* 137: 110618.

Pita, E. G., and S. Stevenson. 1998. *Air Conditioning Principles and Systems Book*. Columbus, OH: Prentice Hall.

Polydoros, A. S., and L. Nalpantidis. 2017. "Survey of Model-Based Reinforcement Learning: Applications on Robotics." *Journal of Intelligent & Robotic Systems* 86 (2): 153–173. doi:10.1007/s10846-017-0468-y.

Raptodimos, Y., and I. Lazakis. 2020. "Application of NARX Neural Network for Predicting Marine Engine Performance Parameters." *Ships and Offshore Structures* 15 (4): 443–452. doi:10.1080/17445302.2019.1661619.

Rijal, H. B., M. A. Humphreys, and J. F. Nicol. 2018. "Development of a Window Opening Algorithm Based on Adaptive Thermal Comfort to Predict Occupant Behavior in Japanese Dwellings." *Japan Architectural Review* 1 (3): 310–321. doi:10.1002/2475-8876.12043.

Ruiz, L. G. B., M. P. Cuéllar, M. D. Calvo-Flores, and M. D. C. P. Jiménez. 2016. "An Application of Nonlinear Autoregressive Neural Networks to Predict Energy Consumption in Public Buildings." *Energies* 9 (9): 684.

Ryzhov, A., H. Ouerdane, E. Gryazina, A. Bischi, and K. Turitsyn. 2019. "Model Predictive Control of Indoor Microclimate: Existing Building Stock Comfort Improvement." *Energy Conversion and Management* 179: 219–228. doi:10.1016/j.enconman.2018.10.046.

Sangi, R., and D. Müller. 2018. "A Novel Hybrid Agent-Based Model Predictive Control for Advanced Building Energy Systems." *Energy Conversion and Management* 178: 415–427. doi:10.1016/j.enconman.2018.08.111.

Sutton, R. S., and A. G. Barto. 2018. *Reinforcement Learning: An Introduction*. 2nd ed. Cambridge, MA: MIT Press.

Talebi, A., and A. Hatami. 2020. "Online Fuzzy Control of HVAC Systems Considering Demand Response and Users' Comfort." *Energy Sources, Part B: Economics Planning, and Policy* 15 (7-9): 403–422. doi:10.1080/15567249.2020.1825557.

Vázquez-Canteli, J. R., S. Ulyanin, J. Kämpf, and Z. Nagy. 2018. "Fusing TensorFlow with Building Energy Simulation for Intelligent Energy Management in Smart Cities." *Sustainable Cities and Society* 45: 243–257. doi:10.1016/j.scs.2018.11.021.

Wang, Z., and T. Hong. 2020. "Reinforcement Learning for Building Controls: The Opportunities and Challenges." *Applied Energy* 269: 115036. doi:10.1016/j.apenergy.2020.115036.

Wang, Y., K. Velswamy, and B. Huang. 2017. "A Long-Short Term Memory Recurrent Neural Network-Based Reinforcement Learning Controller for Office Heating Ventilation and air Conditioning Systems." *Processes* 5 (3): 46. doi:10.3390/pr5030046.

Wang, L., Z. Wang, and R. Yang. 2012. "Intelligent Multi-Agent Control System for Energy and Comfort Management in Smart and Sustainable Buildings." *IEEE Transactions on Smart Grid* 3 (2): 605–617. doi:10.1109/TSG.2011.2178044.

Wei, T., S. Ren, and Q. Zhu. 2019. "Deep Reinforcement Learning for Joint Datacenter and HVAC Load Control in Distributed Mixed-use Buildings." *IEEE Transactions on Sustainable Computing*, 1–16. doi:10.1109/TSUSC.2019.2910533.

Yoon, A. Y., H. K. Kang, and S. I. Moon. 2020. "Optimal Price Based Demand Response of HVAC Systems in Commercial Buildings Considering Peak Load Reduction." *Energies* 13 (4): 862. doi:10.3390/en13040862.

Yu, L., Y. Sun, Z. Xu, C. Shen, D. Yue, T. Jiang, and X. Guan. 2020. "Multi-agent Deep Reinforcement Learning for HVAC Control in Commercial Buildings." *IEEE Transactions on Smart Grid* 2: 1–14. doi:10.1109/TSG.2020.3011739.

Yuan, Z., Y. Huang, X. Lu, J. Huang, Q. Liu, G. Qi, and Z. Cao. 2021. "Measurement of CO_2 by Wavelength Modulated Reinjection Off-Axis Integrated Cavity Output Spectroscopy at 2 μm." *Atmosphere* 12 (10): 1247. doi:10.3390/atmos12101247.

Yuan, X., Y. Pan, J. Yang, W. Wang, and Z. Huang. 2020. "Study on the Application of Reinforcement Learning in the Operation Optimization of HVAC System." *Building Simulation*, 1–13. doi:10.1007/s12273-020-0602-9.

Zhang, Z., A. Chong, Y. Pan, C. Zhang, and K. P. Lam. 2019. "Whole Building Energy Model for HVAC Optimal Control: A Practical Framework Based on Deep Reinforcement Learning." *Energy and Buildings* 199: 472–490. doi:10.1016/j.enbuild.2019.07.029.

Zhang, Z., A. Chong, Y. Pan, C. Zhang, S. Lu, and K. P. Lam. 2018. "A Deep Reinforcement Learning Approach to Using Whole-Building Energy Model for HVAC Optimal Control". *Building Performance Analysis Conference and SimBuild* (Vol. 3, pp. 22–23).

Zhang, C., S. R. Kuppannagari, R. Kannan, and V. K. Prasanna. 2019. "Building HVAC Scheduling Using Reinforcement Learning Via Neural Network-Based Model Approximation". *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, 287–296. doi:10.1145/3360322.3360861.

Zhang, Z., C. Zhang, and K. P. Lam. 2018. "A deep reinforcement learning method for model-based optimal control of HVAC systems". doi:10.14305/ibpc.2018.ec-1.01.

Zhao, H., J. Zhao, T. Shu, and Z. Pan. 2021. "Hybrid-model-based Deep Reinforcement Learning for Heating, Ventilation and air-Conditioning Control." *Frontiers in Energy Research* 8: 412. doi:10.3389/fenrg.2020.610518.

Zou, Z., X. Yu, and S. Ergan. 2019. "Towards Optimal Control of air Handling Units Using Deep Reinforcement Learning and Recurrent Neural Network." *Building and Environment* 168: 106535. doi:10.1016/j.buildenv.2019.106535.

Appendix A

The AHU model

Using Eqn. (A1), the AHU control volume can determine the latent energy absorbed by the cooling coil based on the energy flow.

$$M_{He}cp_{He}\frac{dT_s(t)}{dt} = m_a^{\bullet}cp_aT_m(t) - m_a^{\bullet}cp_aT_s(t) + m_w^{\bullet}kcp_wT_{win} - m_w^{\bullet}kcp_wT_{wout} \quad \text{(A1)}$$

$$T_{wout} - T_{win} = \Delta T = 5 - 10 = -5$$

Where $T_m(t)$ is the temperature of the mixing air at time t (^oC), T_{wout}&T_{win} are water out and in heat exchanger temperature(^oC), $m_w^\bullet = ch\,(t)$ and $m_a^\bullet$ are the mass flow rate of the chilled water and outside air at time $t(kg/\sec.)$.

Applying the Laplace transform to both sides of the equation, using zero initial conditions, and rearranging the equation, provides the transfer function of the overall description of physical behaviour for the heat exchanger temperature ratio of out supply air as follows:

$$T_s(s) = \frac{T_m(s)}{(\tau_1 s + 1)} + \frac{ch(s)\,cp_w \Delta T_w}{m_a^\bullet cp_a(\tau_1 s + 1)} \tag{A2}$$

$$m_a^\bullet = \frac{A * h * \rho * \#of.air.replaced.times(=4)}{3600} = 0.84\,kg/\sec.$$

$$\tau_1 = \frac{M_{He} cp_{He}}{m_a^\bullet cp_a} = 4.7382\ \sec.$$

$$Where: D_{rr}(s) = 1 - D_{rf}(s),$$

$$and\ T_m(s) = D_{rr}(s)T_r(s) + D_{rf}(s)T_o(s)$$

$$T_s(s) = \frac{D_{rr}(s)T_r(s) + D_{rf}(s)T_o(s)}{(\tau_1 s + 1)} + \frac{ch(s)\,cp_w \Delta T_w}{m_a^\bullet cp_a(\tau_1 s + 1)} \tag{A3}$$

$$T_s(s) = \frac{ch(s)\ * 4200 * -5}{0.84 * 1.005(4.7s + 1)} + \frac{D_{rr}(s) * T_r(s)}{(4.7s + 1)} + \frac{D_{rf}(s)T_o(s)}{(4.7s + 1)}$$

CO_2 sensor model

Applying mass conservation law on the control volume for the components of conditioning space to get general formula as following:

$$\text{rate of } CO_2 \text{ change} = \text{rate of } CO_2 \text{ transfer} + \text{rate of } CO_2 \text{ generation}$$

$$\frac{dCO_{2r}(t)}{dt} = \sum\nolimits_i input\,CO_2 rate - \sum\nolimits_e output\,CO_{2out} rate + \sum\nolimits_{gen.} CO_2\ generation\ rate \tag{A4}$$

The mass balance of conditioned space is given by:

$$\frac{dM_r CO_{2r}(t)}{dt} = v_f^\bullet CO_{2out} + m^\bullet CO_{2gen}(t) - \frac{ppm CO_{2r}(t) v_{ro}^\bullet}{v_{ro}}$$

$$\frac{dM_r CO_{2r}(t)}{dt} + \frac{ppm CO_{2r}(t) v_{ro}^\bullet}{v_{ro}} = v_f^\bullet CO_{2out} + m^\bullet CO_{2gen}(t), M_r = ppm \tag{A5}$$

The authors get the transfer function below by applying the Laplace transform to both sides of Eq. (A5), supposing zero initial conditions, and rearranging the expression:

$$ppm\,CO_{2r}(s)\frac{v_r^\bullet}{v_r}(\tau_2 s + 1) = v_f^\bullet\,CO_{2out} + CO_{2gen}(s)$$

$$\therefore ppm\,CO_{2r}(s) = \frac{v_r CO_{2out}\,D_{rr}\,F}{v_r^\bullet\,(\tau_2 s + 1)} + \frac{v_r\,CO_{2gen}(s)}{v_r^\bullet\,(\tau_2 s + 1)} \tag{A6}$$

Substituting $v_r^\bullet = v_f^\bullet + v_w^\bullet + v_d^\bullet + v_{ven}^\bullet = 0.625\frac{m^3}{s} = ACH$, $v_r = 616 m^3$ in Eq. (A6)

$$ppm\,CO_{2r}(s) = \frac{616\,CO_{2out} v_f^\bullet}{0.625\,(\tau_2 s + 1)} + \frac{616 CO_{2gen}(s)}{0.625\,(\tau_2 s + 1)}$$

Where:

$v_f^\bullet$: Volume rate of air filtration (m^3/sec.)
$v_w^\bullet$: Volume rate of air comes from windows (m^3/sec.)
$v_d^\bullet$: Volume rate of air comes from doors (m^3/sec.)
$v_v^\bullet$: Volume rate of air ventilation (m^3/sec.)
ACH = Average change per hr.

$$ppm\,CO_{2r}(s) = \frac{v_r CO_{2out} D_{rr}(s) F}{v_r^\bullet\,(\tau_2 s + 1)} + \frac{v_r\,CO_{2gen}(s)}{v_r^\bullet\,(\tau_2 s + 1)} \tag{A7}$$

$$CO_{2gen_Z}(t) = \sum_{l=0}^{Z} W_l(t)\,f(t_l)$$

Where:

$$CO_{2gen_l}(t) = \prod_{J=0}^{Z}\left(\frac{t - t_J}{t_l - t_J}\right) f(t_l) \tag{A8}$$

$$ppm\,CO_{2r}(s) = \frac{616\,CO_{2out} D_{rr}(s) F}{0.625\,(\tau_2 s + 1)} + \frac{616\,CO_{2gen}(s)}{0.625\,(\tau_2 s + 1)},$$

$$F = 0.7 v_r^\bullet = 0.437\frac{m^3}{s}$$

Sub. Eq. (A8) in Eq. (A7) after taking Laplace transform and sub. $CO_{2out} = 600$ ppm, $\tau_2 = \frac{v_r}{v_r^\bullet} = 985.6$ sec.

$$ppm\,CO_{2r}(s) = \frac{616 \times 600 \times 0.437 \times D_{rr}(s)}{0.625\,(\tau_2 s + 1)} + \frac{616 \times CO_{2gen_l}(s)}{0.625\,(\tau_2 s + 1)}$$

Indoor conditioned space model

By applying the energy and mass conservation laws to the conditioned space control volume, the temperature ratio efficiency has been studied.

Heat balance in the conditioned building is given by:

$$Q_r^\bullet = Q_s^\bullet + Q_{wd/dr}^\bullet + Q_{wal}^\bullet + Q_{cel}^\bullet + Q_l^\bullet \tag{A9}$$

Where: $Q_r^\bullet = M_{ar} cp_{ar}\frac{dT_r(t)}{dt}$, $Q_s^\bullet = \sum_j m_{as} cp_a(T_s(t) - T_r(t))$, $Q_{wd/dr}^\bullet = \sum_j m_{av} cp_a(T_o(t) - T_r(t))$, $Q_{wal}^\bullet = \sum_j \frac{KA}{\Delta x}(T_o(t) - T_r(t))$, $Q_{cel}^\bullet = 0.6 Q_{wal}^\bullet$, and $Q_l^\bullet = 40 * LiON(t)$.

$$m_{as} = m_{av} = m_a^\bullet$$

The simplified transfer function of the complete physical behaviour description for the conditioned space temperature ratio of room/out air is given by:

$$M_{ar} cp_a \frac{dT_r(t)}{dt} = \sum\nolimits_j m_{as} cp_a(T_s(t) - T_r(t)) + \sum\nolimits_j m_{av} cp_a(T_o(t) - T_r(t)) * WiON(t) + \sum\nolimits_j \frac{KA}{\Delta x}(T_o(t) - T_r(t)) + 0.6 * \sum\nolimits_j \frac{KA}{\Delta x}(T_o(t) - T_r(t)) + 40 * LiON(t) \tag{A10}$$

Taking Laplace transform and simplifying Eq. (A10)

$$T_r(s) = \frac{m_{as} cp_a T_s(s)}{\left(\frac{KA}{\Delta x} + 2 m_{as} cp_a\right)(\tau_3 s + 1)} + \frac{m_{av} cp_a T_o(s) * WiON\ (s)}{\left(\frac{KA}{\Delta x} + 2 m_{as} cp_a\right)(\tau_3 s + 1)}$$

$$+\frac{KAT_o(s)(1+0.6)}{\Delta x\left(\frac{KA}{\Delta x}+2m_{as}cp_a\right)(\tau_3 s+1)} + \frac{40*LiON\ (s)}{\left(\frac{KA}{\Delta x}+2m_{as}cp_a\right)(\tau_3 s+1)} \tag{A11}$$

$$\tau_3 = \frac{M_{ar}cp_a}{\frac{KA}{\Delta x}+2m_{as}cp_a} \rightarrow \tau_3 = \frac{Ab*h*\rho*cp_a}{\frac{KA}{\Delta x}+2m_{as}cp_a} = 381.5791\ \text{sec}.$$

$$T_r(s) = \frac{0.84*1005*T_s(s)}{1992*(381.5791s+1)} + \frac{0.84*1005*T_o(s)*WiON(s)}{1992*(381.5791s+1)} + \frac{(304+(304*0.6))*T_o(s)}{1992*(381.5791s+1)} + \frac{40*LiON(s)}{1992*(381.5791s+1)}$$