Tikrit Journal of Engineering Sciences (2025); 32(2): 1614
DOI: http://doi.org/10.25130/tjes.32.2.25

IRAQI Academic Scientific Journals

ISSN: 1813-162X (Print); 2312-7589 (Online)
Tikrit Journal of Engineering Sciences
available online at: http://www.tj-es.com

TJES Tikrit Journal of Engineering Sciences

Optimizing HVAC&R System Efficiency and Comfort Levels Using Machine Learning-Based Control Methods

Suroor M. Dawood *[a], Raad Z. Homod [b], Alireza Hatami [c]
[a] Department of Chemical and Petroleum Refining Engineering, College of Oil and Gas Engineering, Basra University for Oil and Gas, Basrah, Iraq.
[b] Department of Oil and Gas, College of Oil and Gas Engineering, Basra University for Oil and Gas, Basrah, Iraq.
[c] Department of Electrical Engineering, Faculty of Engineering, Bu-Ali Sina University, Hamedan, Iran.

Keywords:
Deterministic policy; Energy saving; HVAC&R system; Machine learning; Model-based reinforcement learning.

Highlights:
- *Optimally controlling the HVAC&R as an MDP for energy savings, IAQ, and temperature violations.*
- *Applying DP-MB-RL for HVAC&R controlling achieved energy minimization of up to 15% compared to the MB-RL controller.*
- *DP-MB-RL also saved up to 21% energy compared to the TSF controller.*

ARTICLE INFO
Article history:

Received	01 Sep. 2023
Received in revised form	10 July 2024
Accepted	24 July 2024
Final Proofreading	22 Aug. 2024
Available online	31 May 2025

Citation: Dawood SM, Homod RZ, Hatami A. **Optimizing HVAC&R System Efficiency and Comfort Levels Using Machine Learning-Based Control Methods**. *Tikrit Journal of Engineering Sciences* 2025; **32**(2): 1614.
http://doi.org/10.25130/tjes.32.2.25

*Corresponding author:

Suroor M. Dawood

Department of Chemical and Petroleum Refining Engineering, College of Oil and Gas Engineering, Basra University for Oil and Gas, Basrah, Iraq.

Abstract: The Heating, Ventilation, Air Conditioning, and Refrigeration (HVAC&R) system is a complex, nonlinear behavior with a high uncertainty control system that equips the thermal comfort desired but consumes significant electrical energy and costs in different types of buildings, such as residential, commercial, and industrial. This paper introduces a new approach for online controlling of HVAC&R systems using model-based reinforcement learning (MB-RL) style to diminish energy usage and energy cost, maintain the occupants' comfort levels by controlling the buildings' indoor temperature, and maintain the desired carbon dioxide levels simultaneously. For this purpose, a new model based on energy and mass conservation laws is presented to model the dynamic variations of temperature and CO2 concentration levels. The HVAC&R system control trouble is defined as a specific Markov Decision Processes (MDPs) model. The reward function balances the ability to increase energy conservation while preserving the interior comfort requirements of occupants. Employing the deterministic policy algorithm (DP), the proposed methodology can manage the dimensionality curse problem due to increased state-action space. Then, it overcomes the nonlinearity and the control system uncertainty. The MB-RL algorithm, which uses a unique DP called DP-MB-RL, can select the best decisions instead of stochastic policy to reduce the calculation time. A real case, a building in Basra City, Iraq, is simulated using MATLAB software. Devoting the MB-RL and DP-MB-RL techniques to online control of an HVAC&R system, the simulation results for both methods are provided. For instance, the parameters, like electrical power, internal comfort levels, energy consumed, and energy cost at different pricing schemes, such as fixed pricing (FP), time-of-use (TOU), and real-time pricing (RTP), are assessed. The results indicated that the suggested DP-MB-RL methodology had better indoor thermal and air quality satisfaction levels, energy-saving (more than 15%), and reduced the cost of electricity by more than 15%, 13%, and 10% for FP, TOU, and RTP pricing schemes, respectively, compared to the benchmark MB-RL style controller. The DP-MB-RL controller also performed better than the Takagi-Sugeno Fuzzy (TSF) controller for the same building, saving more than 21% energy.

تحسين كفاءة ومستويات الراحة لنظام التدفئة والتهوية وتكييف الهواء باستخدام أساليب التحكم القائمة على التعلم الآلي

سرور مؤيد داود[١]، رعد زعلان حمود[٢]، علي رضا حاتمي[٣]
[١] قسم الهندسة الكيميائية وتكرير النفط/ كلية هندسة النفط والغاز / جامعة البصرة للنفط والغاز / البصرة – العراق.
[٢] قسم هندسة النفط والغاز/ كلية هندسة النفط والغاز / جامعة البصرة للنفط والغاز / البصرة – العراق.
[٣] قسم الهندسة الكهربائية/ كلية الهندسة / جامعة بو علي سينا / إيران.

الخلاصة

يعد نظام التدفئة والتهوية وتكييف الهواء والتبريد (HVAC&R) نظاماً ذو سلوك معقد وغير خطي مع تحكم صعب لوجود عوامل عالية الريبة لتوفير الراحة الحرارية المطلوبة. علمًا انه يستهلك طاقة كهربائية وتكاليف كبيرة في أنواع مختلفة من المباني مثل السكنية والتجارية والصناعية. تقدم هذه الورقة نهجًا جديدًا للتحكم عن بعد في أنظمة التدفئة والتهوية وتكييف الهواء والتبريد باستخدام أسلوب التعلم المعزز القائم على النموذج (MB-RL) لتقليل استخدام الطاقة وتكلفتها، وللحفاظ على مستويات راحة الساكنين من خلال التحكم في درجة الحرارة الداخلية للمباني والحفاظ على درجة الحرارة وثاني أكسيد الكربون ضمن الحدود المطلوبة في آن واحد. ولهذا الغرض، تم تقديم نموذج جديد يعتمد على قوانين الحفاظ على الطاقة والكتلة لنمذجة التغيرات الديناميكية في درجات الحرارة ومستويات تركيز ثاني أكسيد الكربون. حيث تم تعريف مسألة التحكم في نظام التدفئة والتهوية وتكييف الهواء والتبريد (HVAC&R) على أنها نموذج محدد لعمليات اتخاذ قرار ماركوف (MDPs). تعمل وظيفة المكافأة على موازنة القدرة على زيادة الحفاظ على الطاقة مع الحفاظ على متطلبات الراحة الداخلية للساكنين. باستخدام خوارزمية السياسة الحتمية (DP)، يمكن للمنهجية المقترحة إدارة مشكلة لعنة الأبعاد بسبب زيادة مساحة عمل الحالة، ومن ثم التغلب على اللاخطية وعدم اليقين في نظام التحكم. يمكن لخوارزمية MB-RL، التي تستخدم DP وتسمى DP-MB-RL، تحديد أفضل القرارات بدلاً من السياسة العشوائية لتقليل وقت الحسابات. في الحالة الواقعية، تمت محاكاة مبنى في مدينة البصرة، العراق، باستخدام برنامج MATLAB. من خلال تخصيص كل من تقنيات MB-RL وDP-MB-RL للتحكم عن بعد في نظام HVAC&R، وبعد استخراج نتائج المحاكاة لكلا الطريقتين، تم تقييم معلمات مثل الطاقة الكهربائية، ومستويات الراحة الداخلية، والطاقة المستهلكة، وتكلفة الطاقة استنادا لأنظمة تسعير مختلفة مثل التسعير الثابت (FP)، ووقت الاستخدام (TOU)، والتسعير في الوقت الفعلي (RTP). تشير النتائج إلى أن منهجية DP-MB-RL المقترحة تتمتع بمستويات رضا أفضل لدرجات الحرارة الداخلية وجودة الهواء، وتوفير الطاقة (أكثر من ١٥٪)، وتقليل تكلفة الكهرباء بأكثر من ١٥٪، و١٣٪، و١٠٪ لمخططات تسعير FP وTOU وRTP، على التوالي، مقارنةً بوحدة التحكم في نمط MB-RL القياسية. تعمل وحدة التحكم DP-MB-RL أيضًا بشكل أفضل من وحدة التحكم Takagi-Sugeno Fuzzy (TSF) لنفس المبنى، من خلال توفير طاقة أكبر بنسبة ٢١٪.

الكلمات الدالة: السياسة الحتمية، توفير الطاقة، نظام التدفئة والتهوية وتكييف الهواء والتبريد، التعلم الآلي، التعلم التعزيزي القائم على النموذج.

1.INTRODUCTION

In recent years, due to environmental issues, optimizing electrical energy consumption from different aspects, particularly demand response programs, has gained more attention [1, 2]. Most of the people's time is spent in buildings, causing about 40% of total energy usage and one-third of the greenhouse gas (GHG) emissions in the world [3, 4]. The Heating, Ventilation, Air Conditioning, and Refrigeration (HVAC&R) and lighting systems consume more than half of the electricity in commercial and 40% of residential buildings [5, 6]. World energy demand rise is estimated to reach 30% in 2040 [7]. Therefore, improving efficiency on the above side can decrease energy consumption and CO_2 emission. These are the most critical factors affecting studies' motivation to develop an HVAC&R system for optimal energy management. Numerous studies have been done to reduce buildings' energy usage and achieve the users' thermal comfort by controlling the HVAC systems [8-10]. Some researchers have also considered demand response programs [11-13]. Several studies have demonstrated that machine learning (ML) methods can be used successfully to control HVAC systems [14, 15]. There are four ML categories of approaches: reinforcement learning (RL), supervised, semi-supervised, and unsupervised [16]. Recently, RL has gained ground because it enhances performance and energy management with accurate control for all building types [11]. RL can be applied as model-based (MB), called MB-RL, [8, 17, 18] or model-free (MF), called MF-RL [3, 11, 19, 20]. In an MF-RL method, the training operation takes a lot of time and requires a large volume of data [4]. In addition, the algorithm is trained in simulation environments before being used in real ones. The MF-RL methods have been presented in [19, 21], which use the Q-learning function for the HVAC&R system control. In [22], using the linear RL in energy saving of the building has been discussed. A neural-fitted RL technique has been proposed to get the desired temperature thresholds [23]. The articles [24-26] presented an MF batch RL method applied to high-dimensional state-action spaces, but the batch update algorithm requires a high computational cost. For enhancement of its performance, the MF-RL method has been combined with a rule-based controller [19] and with a model predictive controller (MPC) [27]. MF-RL has also been combined with neural networks (N.Ns) [13, 17, 28, 29] to obtain a Deep RL (DRL), involving the cost and efficiency of the learning process as the main challenges. Polydoros and Nalpantidis [30], a comprehensive and detailed survey has been presented on applying ML and DRL methods used for the energy management of different systems. Polydoros and Nalpantidis [30] indicated the high usage, importance, and considerable capability of ML and DRL methods for analyzing energy management systems problems. These methods are becoming valuable for numerous applications as they have played an important role in recognizing subtle structures of high-dimension data sets [31]. On the other side, the learning cost and efficiency are the primary difficulties of the DRL controllers to practice [8]. In Ref.[32], an MF-RL controller was developed to observe the stochastic behavior of occupants and PV power production while minimizing energy consumption, ensuring tenants' comfort levels and water hygiene. The results showed that the suggested framework successfully learned and predicted its aim by reducing energy consumption without violating hygiene and comfort. Based on [33], hybrid and DL-based models provide the highest score for robustness in terms of energy consumption prediction. In [34], the authors assessed four RL methods for continuous control of an open-source environment: Twin Delayed Deep Deterministic Policy Gradients (TD_3), Proximal Policy Optimization (PPO), Soft Actor-Critic (SAC), and Trust Region Policy Optimization (TRPO). The results confirmed the controllers' performance in terms of thermal stability, 10% more energy savings, and data efficiency than the baseline control method. MB-RL method has been successfully used to keep the thermal comfort level conditions and save 45% energy for multi-chiller HVAC&R systems in the Basra airport environment compared with the traditional PID controller for a typical day [35]. In [36], a deep MB-RL controller was proposed, which uses the nonlinear autoregressive exogenous NN (NARX-NN) as an approximation function to form a hybrid DP-NARX-RL controller. The results demonstrated the improved performance of the DP-NARX-RL controller compared to some controllers in terms of maintaining the comfort levels of the building occupants, reduction in the electric energy usage, energy costs, and training time for various pricing schemes. The authors of [37] proposed a system layout for a thermal energy comfort control system that maximizes comfort levels for tenants and simultaneously minimizes consumed energy, considering the people of different ages by using different regression ML algorithms, like Support Vector Machine (SVM), Multiple Linear (ML), Decision Tree (DT), and Random Forest (RF). The results showed that the SVM performancd better than the others due to its smallest evaluation error and more flexibility. However, it needs a large amount of database and a long time to improve accuracy. Jiang et al. [38] evolved a Deep Q-Network (DQN) by defining a single-zone building environment as a partially observable MDP with a reward function by a trade-off between minimizing energy cost and discomfort levels of punishment. The results showed the outperforming of DQN against the rule-based control by saving up to 6% and 8% energy costs with and without demand charges, respectively. Ref. [39] reduced the energy usage of HVAC&R simultaneously with improving thermal energy comfort limits in smart buildings using Deep Deterministic Policy Gradients (DDPG), while it requires a lot of time to converge into a fixed policy in the HVAC&R system control problem. Ref.[40] showed a day-ahead economic dispatch model used for water-cooled multi-chiller and ice storage unit systems' co-optimization to save total energy using GAMS (Generalized Algebraic Mathematical Modeling Language System). The results demonstrated that applying short-term scheduling to the total plant reduces energy consumption remarkably. In an MB-RL algorithm, the environmental behavior is known for the RL agent (controller). To resolve the above issues, the hybrid MB-DRL approach has been proposed for the commercial multizone building control problem [8]. The findings indicated that the suggested algorithm increased training proficiency and reduced learning cost periods compared to classical DDPG. MB-RL methods use their previously learned dynamics models to generate or schedule new training sets. Despite less training data, the MB-RL method has a high efficiency, expressed in [41]. The authors of [17] proposed an MB-RL method that learned the HVAC system dynamics using an N.N. and reduced the training data significantly compared to the MF-RL technique. In [18], the MB-RL method was used to control multizone buildings, where the training data was reduced by 10.52x to obtain performance comparable with the MF-RL method. In summary, in an MF-RL method, the characteristics of the environment are unknown for the RL agent (treated as a black box), and the agent learns its optimal behavior through a tedious trial and error style [42]. Therefore, MF-RL strategies require large amounts of operating data to converge in the HVAC system to enhance the users' comfort levels. However, collecting and providing such data is complicated and time-consuming in a real-world system [17]. Therefore, in this paper, the MB-RL method is adopted. In this article, an integrated white-box model for the HVAC&R system is presented, wherein the internal heat and CO_2 concentration levels are modeled. The developed model's derivative relations are based on energy and mass conservation laws. Meanwhile, the CO_2 level is represented by a Lagrange polynomial model. Then, using the developed model, the process control of the HVAC&R system is introduced as a Markov Decision Process (MDP) by defining the collections of states and actions, and the reward function. By adopting the MDP as a mathematical framework for describing the environment, an RL method based on the developed model, called MB-RL, is introduced for online controlling the HVAC&R systems. In the MB-RL method, the agent faces high-dimensional state-action spaces in its learning process, leading to highly time-consuming, probably diverging, and undesired final results. To solve this problem, the MB-RL method uses a deterministic policy (DP) in its learning process called DP-MB-RL. The DP learns the optimal policy by selecting the best future decisions. Indeed, it is a function mapping the conditions of the environment to the group of selected actions. In the presented approach, the reward function consists of three components: the first and second components penalize the agent if the interior heat and the carbon dioxide (CO_2) concentration limits are outside allowable values, respectively. By the third component in the reward function, as the energy consumption is increased, the penalty of agents is increased exponentially. Since the energy consumption is not directly visible in the thermal model, the chilled water flow has been used in the reward function as an index for energy consumption. Adjusting the coefficients of different components in the reward function allows a trade-off between the thermal comforts and energy usage of an HVAC&R system. To evaluate the proposed approach, a real case, a building in Basra City, Iraq, has been analyzed. Both approaches, MB-RL and DP-MB-RL were used to control the HVAC&R system. The results demonstrated the superior performance of the suggested strategy compared to MB-RL. Meanwhile, the energy consumption in a day has been computed. The results showed a 15.03% reduction in energy consumption of the suggested approach compared to the MB-RL method.

In summary, the following are the present paper's main contributions:

- Formulating the HVAC&R control issue as an exact MDP where the reward trades off minimizing energy consumption, CO_2 concentration, and temperature violations.
- Proposing a DP-MB-RL method for online control of HVAC&R systems where the DP algorithm can avoid the cumbersome dimensionality curse due to high action-state spaces.
- A real-case residential building has been simulated using MATLAB software. The simulation results for DP-MB-RL and the MB-RL methods were provided and compared. The results showed a) the energy usage during the day was decreased by 15.03% by applying the proposed approach compared to the MB-RL method; b) the proposed controller had a saving of 15.10%, 13.3%, and 10% in electricity prices compared to benchmark controller for fixed pricing (FP), time-of-use (TOU), and real-time pricing (RTP) pricing schemes, respectively c) the introduced approach had improved performance for providing comfort levels compared with the MB-RL method.

The rest of this manuscript is constructed as follows: Section Two addresses the problem definition and system representation. Section Three presents and analyzes the simulation results. Finally, Section Four addresses the conclusions.

2.PROBLEM DEFINITION AND SYSTEM REPRESENTATION

The HVAC system model was developed based on thermodynamic principles. In this model, cooling was provided by a primary cooling coil (air-water heat exchanger) situated at a central air handling unit (AHU). The AHU supplies air into the conditioned space through the fresh/return air dampers employed to regulate the provided air supply flow rate. The HVAC&R system control problem has been formalized as MDPs. In this framework, a DP-MB-RL has been proposed for controlling the HVAC&R system to diminish the overall use of energy while preserving the users' indoor thermal and air quality within the desired levels over time. Figure 1 illustrates the relationships between the environment, including the thermal design of a building, the HVAC&R system, the CO_2 concentration sensor, and the agent, i.e., the DP-MB-RL controller. In the following, each section is described in detail.

2.1.The Integrated Model of the HVAC System

In classical research, the temperature with or without humidity was modeled and controlled by an HVAC system [43, 44]. Some other researchers considered humidity and temperature continuous states, while CO_2 concentration was a discrete state [45, 46]. In this section, an integrated model for the HVAC system is presented in which the dynamic variations of the temperature and CO_2 concentration levels are modeled. Figure 2 demonstrates the block diagrams of subsystems for the planned model. As exposed in Fig. 2, the proposed model comprised submodels, such as heat exchanger, building fixture, and CO_2 sensor. The input signals of the model were the chilled water flow, damper ratios for returned air/fresh air, outside temperature, open/closed windows and doors, on/off lights, time, and supply temperature. The DP-MB-RL agent's learning process adjusted the values of these variables until they reached an acceptable level. The indoor conditions (concentration of carbon dioxide and room temperature) were the outputs of the HVAC&R system model. The values of parameters and the allowable values of variables for the HVAC&R system model are described in Table 1.

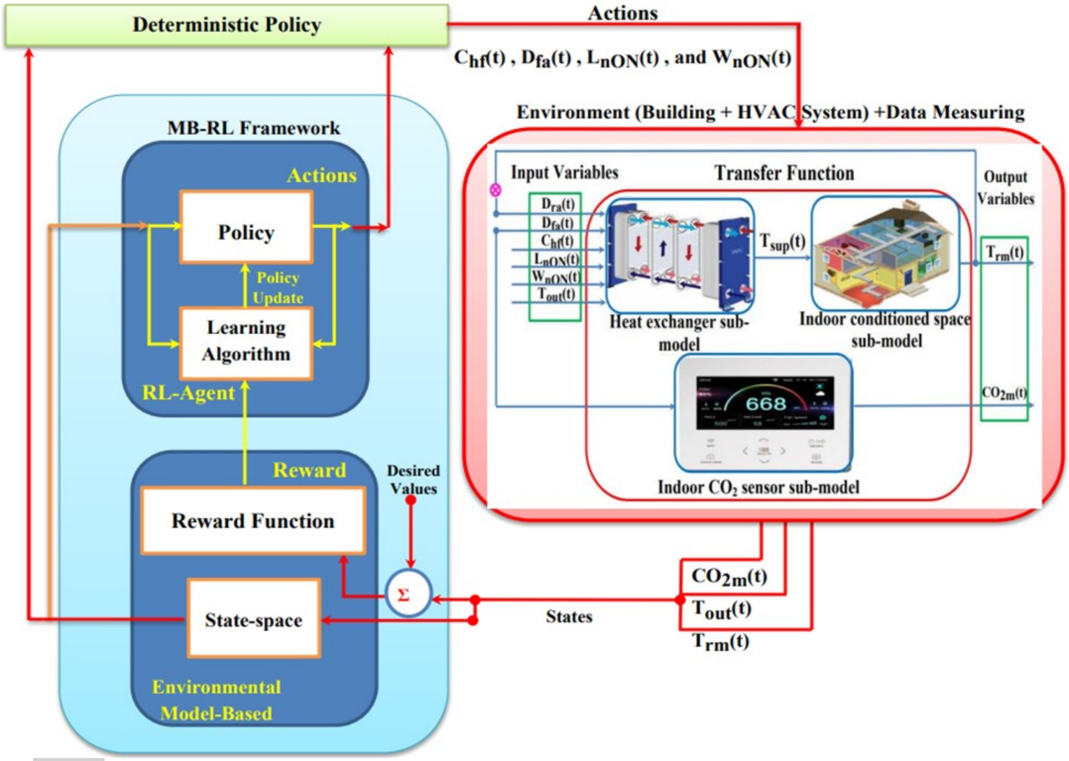


Fig. 1 Overall Block Diagram of the DP-MB-RL Algorithm with the HVAC&R System.

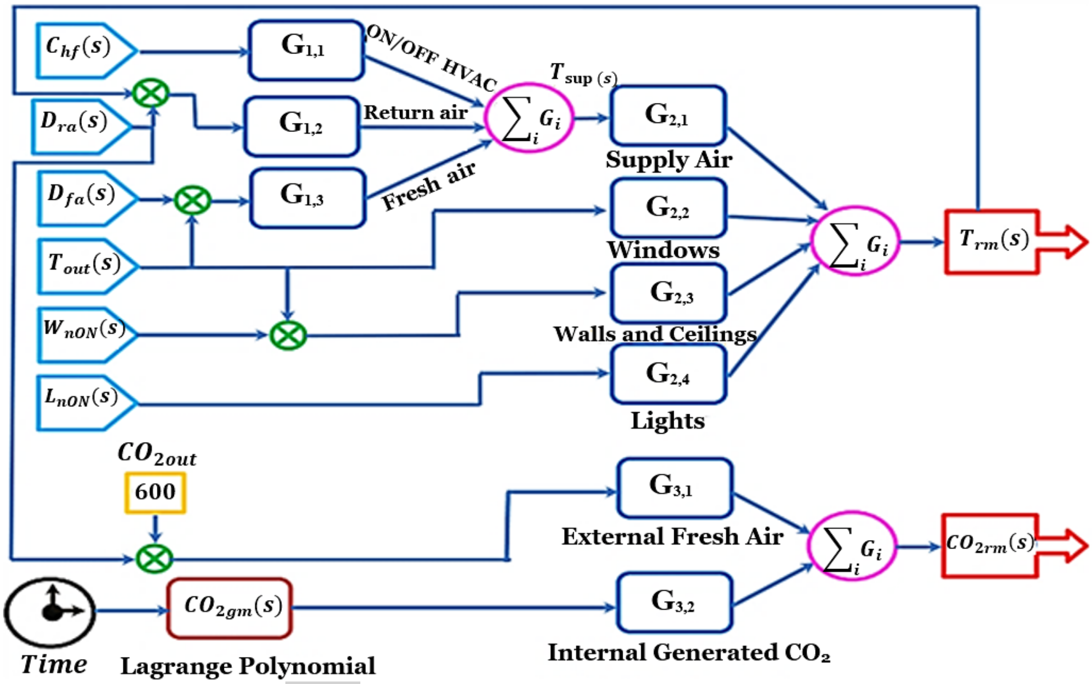


Fig. 2 HVAC&R Subsystems Block Diagram.

2.1.1.The Modeling of Heat Exchanger (Cooling Coil)

The control volume of a heat exchanger can be implemented to get the transfer function of supply temperature using the energy conservation law and the first law of thermodynamics, Eq. (1) [43, 47], which is transferred from the Time-domain into the S-domain, as shown in Eq. (2).

$$M_{He}cp_{He}\frac{dT_{sup}(t)}{dt} = \dot{m}_{ar}(t)cp_{ar}T_{m}(t) - \dot{m}_{ar}(t)cp_{ar}T_{sup}(t) + \dot{m}_{wr}(t)cp_{w}T_{wi}(t) - \dot{m}_{wr}(t)cp_{w}T_{wo}(t) \qquad (1)$$

$$T_{sup\,(s)} = \frac{D_{ra}(s)\times T_{rm}(s)}{(\tau_h s+1)} + \frac{D_{fa}(s)\times T_{out}(s)}{(\tau_h s+1)} + \frac{C_{hf}(s)\times \Delta T_{wio}(s)\times cp_{wo}}{\dot{m}_{ar}(s)\times cp_{ar}\times(\tau_h s+1)} \qquad (2)$$

Table 1 The Parameters and the Allowable Values of Variables for the HVAC System Model [43, 47, 48].

Component	Value	Component	Value
$\dot{m}_{ar}(t) = \dot{m}_{asr}(t) = \dot{m}_{avr}(t)$	0.84	$\Delta T_{wio}(t)$	5
$C_{hf}(t)$	[0 1]	cp_{wo}	4200
$D_{ra}(t)$	[0.25 0.75]	τ_h	4.7
$D_{fa}(t)$	[0.25 0.75]	τ_b	381.58
$W_{nON}(t)$	0 or 1	τ_c	985.6
$L_{nON}(t)$	0 or 1	$t_0 - t_i$	[0 24]
$T_{rm}(t)$	[16 30]	Co_{2out}	600
$T_{out}(t)$	[20 36]	F_{am}	0.437
T_{sup}	[12 15]	$\dot{v}_r(t)$	0.626
$G(t_i)$	[550 1000]	v_{room}	616
cp_{ar}	1.005	Δx_b	0.4
K	0.7	A_b	173.6

where M_{He}(kg) is the heat transfer unit mass, $cp_{He}(J/kg.°C)$ is the specific heat of the cooling coil, $T_{Wi}(t)$(°C) and $T_{Wo}(t)$(°C) are water in/out temperatures of the heat exchanger, $T_m(t)$(°C) and T_{sup}(°C) are the supply air and mixing temperatures at time t, $\dot{m}_{wr}(t) = C_{hf}(t)(kg/sec.)$ is the mass flow of chilled water at time t, $D_{ra}(t)$ and $D_{fa}(t)$(%) are the fresh and return air ratios via damper at t, $T_{rm}(t)$(°C) is the room heat at time t, $T_{out}(t)$(°C) is the external heat at t, $\Delta T_{wio}(t)$(°C) is the difference of water's output and input temperatures, cp_{wo} and $cp_{ar}(J/kg.°C)$ are the specific heat of air and water, $\dot{m}_{ar}(t)(kg)$ is the mass flow rate of outdoor air at time t, and $\tau_h(sec.)$ is the time delay for the cooling coil.

2.1.2. Building Model

By utilizing the mass and energy conservation laws in the control volume of a building structure, the changes in room temperature over time t can be given as in Eq. (3) [43]. It depends on thermal load components, such as walls, doors, lights, windows, and ceilings.

$$T_{rm}(s) = \frac{\dot{m}_{asr}(s)cp_{ar}T_{sup}(s)}{\left(\frac{KA_b}{\Delta x_b}+2\dot{m}_{asr}(s)cp_{ar}\right)(\tau_b s+1)} + \frac{\dot{m}_{avr}(s)cp_{ar}T_{out}(s)\times W_{nON}(s)}{\left(\frac{KA_b}{\Delta x_b}+2\dot{m}_{asr}(s)cp_{a}\right)(\tau_b s+1)} + \frac{KA_bT_{out}(s)(1+0.6)}{\Delta x_b\left(\frac{KA_b}{\Delta x_b}+2\dot{m}_{asr}(s)cp_{ar}\right)(\tau_b s+1)} + \frac{40\times L_{nON}(s)}{\left(\frac{KA_b}{\Delta x_b}+2\dot{m}_{asr}(s)cp_{ar}\right)(\tau_b s+1)} \quad (3)$$

where $\dot{m}_{avr}(t)$ and $\dot{m}_{asr}(t)(kg)$ are the mass flow rate of ventilation and supply air at t, $\tau_b(sec.)$ is the time delay for the air-conditioned area. K is the conductivity, $\Delta x_b(m)$ is the thickness, and $A_b(m^2)$ is the surface area. The $W_{nON}(t)$ and $L_{nON}(t)$ are the open/close windows and on/off lighting at time t. Figure 3 depicts the residential building with an overall area of 220 m² used in an analytical case study [49].

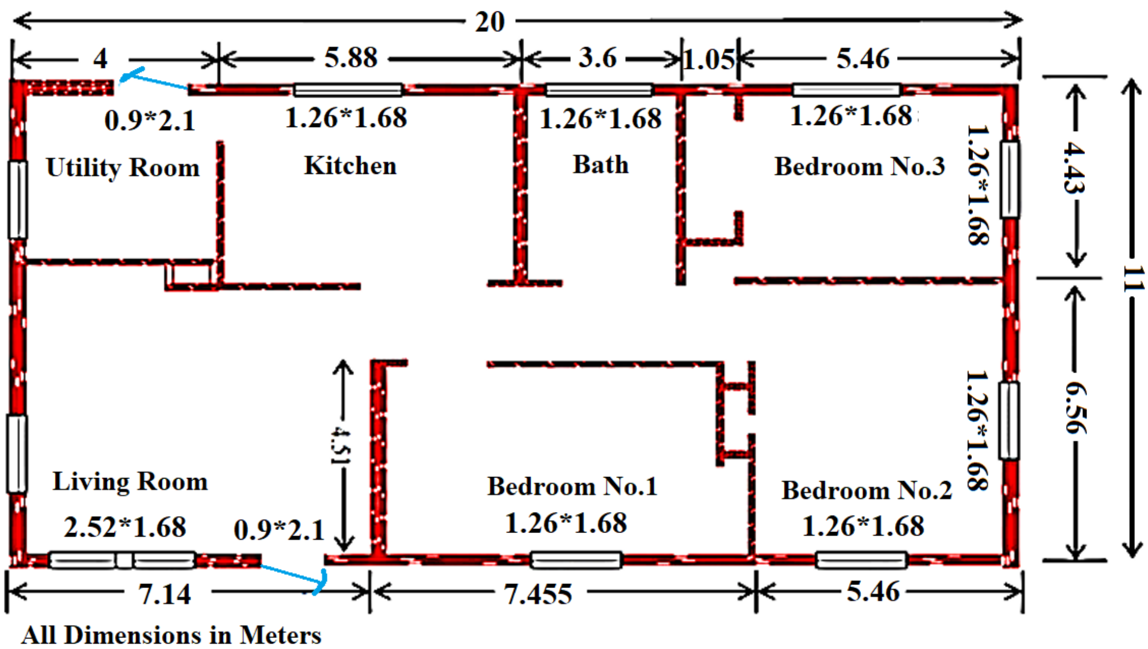


Fig. 3 The Building's Geometry Selected.

2.1.3. CO_2 Concentration Level Model

Some researchers have reported that in some cases, the interior air can be more seriously polluted than outside air [50, 51]. Given the assumption that the outdoor CO_2 concentration is constant (600ppm [52]), the indoor CO2 emissions comprise the building tenants' CO2 quantity and the CO2 released from the indoor appliances. Based on the first law of energy and using the ordinary differential equations in mixing flow [53], the CO2 generation level can be determined (Eqs. (4) and (5)). In steady-state conditions, the CO2 generation level can be described by a Lagrange polynomial model for a given time horizon. A detailed description of the physical behavior for the two output values (IAQ and heat) is given by combining the above three sub-model equations, as presented in Appendix A [36].

$$CO_{2rm}(s) = \frac{v_{room} CO_{2out} D_{ra}(s) F_{am}}{\dot{v}_r(s)(\tau_c s+1)} + \frac{v_{room} CO_{2gm}(s)}{\dot{v}_r(s)(\tau_c s+1)} \quad (4)$$

$$CO_{2gm}(t) = \prod_{\substack{j=0 \\ j \neq i}}^{X} \left(\frac{t - t_j}{t_i - t_j} \right) G(t_i) \quad (5)$$

where $CO_{2gm}(t)(ppm)$ is the indoor generated CO_2 concentration level, $CO_{2out}(ppm)$ is the outside carbon dioxide concentration, $F_{am}\left(m^3/sec.\right)$ is the volumetric airflow rate, $\dot{v}_r(t)(m^3/sec.)$ is the volume rate of the room, $v_{room}(m^3)$ is the volume of the building, $\tau_c(sec.)$ is the time delay for the CO_2 sensor, $t - t_i(hours)$ is the time, and $G(t_i)(ppm)$ is the indoor CO_2 concentration at time t.

2.2. Problem Formulation and MB-RL Controller Design Architecture

The main components of the online MB-RL control method can be defined as a tuple (S, A, β, $\rho_{ss'}$, and ℜ). S and A are the groups of states and actions, respectively. β is the discount factor used to discount the value of future rewards. $\rho_{ss'}$ and ℜ are the matrix of state-to-state transition probability and reward functions, respectively. In a series of episodes, the MB-RL agent (controller/decision-maker) communicates with its environment. Each episode starts with the RL agent in state Sin and ends once the agent makes the best decisions. The agent picks an action a ∊ A at state s ∊ S after observing the state. Consequently, the instant reward R is received by the RL agent (Fig. 4). The main target of the RL agent is to optimize the overall predicted reward obtained over time [54].

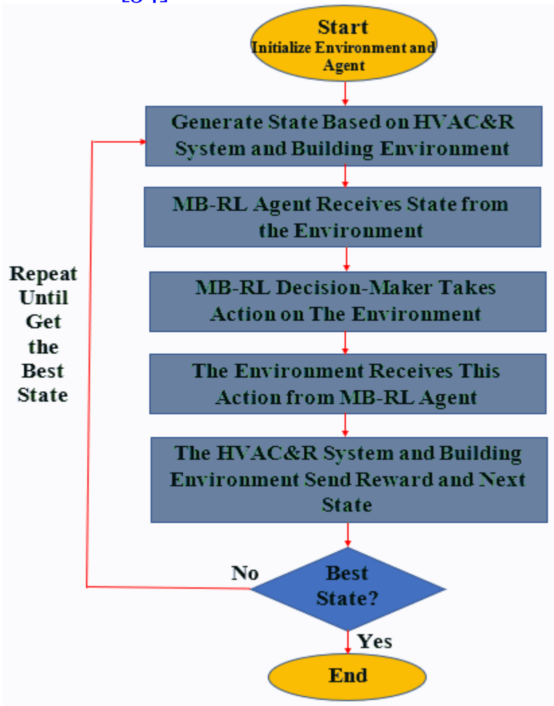


Fig. 4 MB-RL Algorithm's Flowchart.

2.2.1. State–Action Space

The states are the mathematical representation of the environment that is important and useful in decision-making. In this work, three states have been considered: $T_{rm}(t)$ =state (1), $T_{out}(t)$=state (2), and $CO_{2rm}(t)$=state (3), as exposed in Fig. 5. The boundaries of the indoor occupants' comfort levels are included in the state-space values to prevent excessive energy consumption. Actions are the decisions made by the MB-RL agent to control its environment. The set of selected actions, A, contains four control factors: chilled water flow valve position for the heat transfer unit, ON-AND-OFF lights, open/close windows/doors for ventilation purposes, and position damper actuator of the fresh/returned air, respectively. Therefore, the outputs of controller action are $A=[C_{hf}(t), L_{nON}(t), W_{nON}(t), D_{fa}(t)]^T$. To reduce the disturbance effects on the MB-RL controlled state-space values, which are constantly altering in line with the dynamic cooling load, the agent's action-space values are adjusted for each time slot.

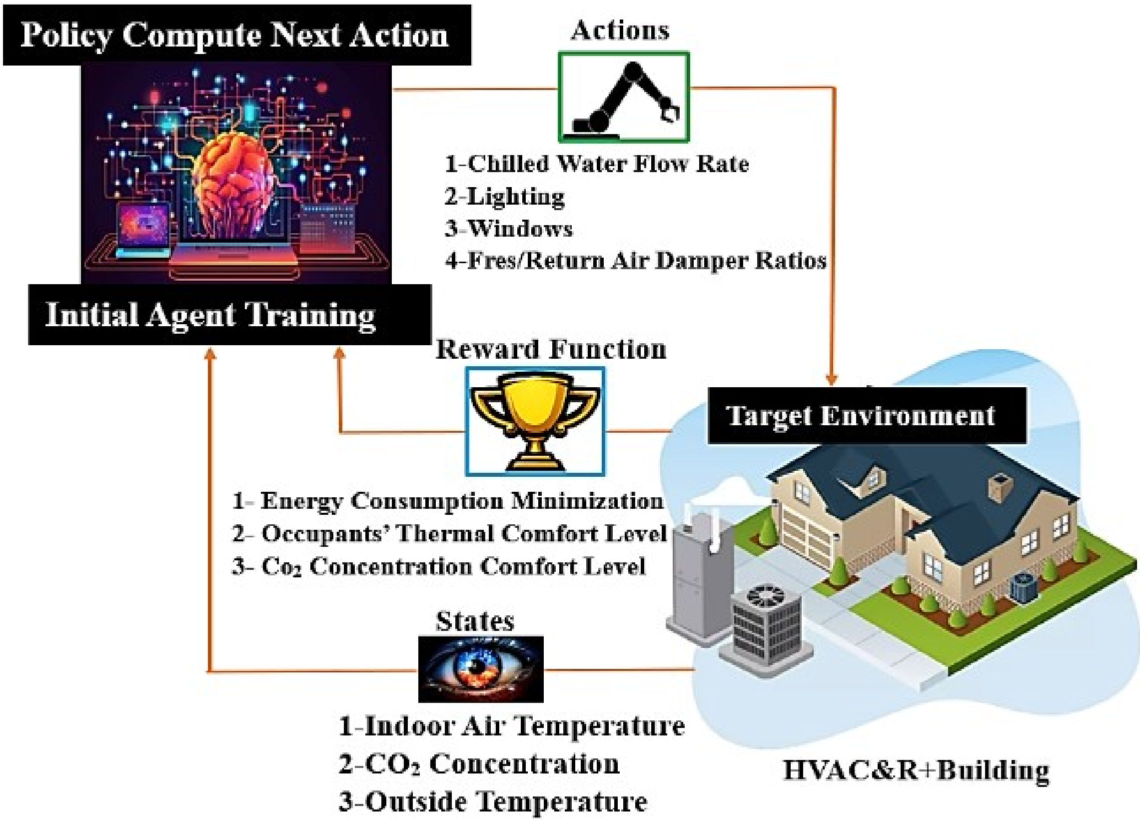


Fig. 5 Schematic Diagram of the MB-RL Controller.

2.2.2.Reward-Function

The reward function (ℜ) estimates the instant rewards of making an action at a specific state. In this work, the designed MB-RL agent's ℜ consisted of three parts, as shown in Eq. (6), including punishment for the HVAC system energy consumption ($X(t)=exp\left(C_{hf}(t)\right)$) and penalties for indoor air quality (IAQ) ($Y(t)=\left|\frac{2CO_{2m}(t)-\overline{CO}_{2m-des}(t)-\underline{CO}_{2m-des}(t)}{2}\right|$) and residents' thermal discomfort levels ($Z(t)=\left|\frac{2T_{rm}(t)-\overline{T}_{rm-des}(t)-\underline{T}_{rm-des}(t)}{2}\right|$). The agent must be penalized if the HVAC system consumes more electricity or the tenants are dissatisfied with the building's air quality and temperature conditions.

$$\boldsymbol{\Re = -\delta \times \left[Z(t) + D_{fa}(t) \times Y(t)\right]^2 - X(t)} \quad \textbf{(6)}$$

The exponential function (exp ($C_{hf}(t)$)) indicates the importance of the On-and-Off switching of the HVAC&R system. In summary, ℜ has been used as the agent's guideline based on energy savings and internal occupants' comfort levels to get the optimal value function by Bellman's equation. By adjusting the coefficient δ in Eq. (6), a trade-off between energy consumption and thermal comfort conditions can be made.

2.2.3.Discount Parameter, Value, and Policy Functions

The value function (V) is composed of the accumulative rewards of several future steps that the RL agent will take based on implementing a fixed policy that starts with S(0) and continues until the end, i.e., S (desired) [6]. Utilizing Bellman's equation, V can be expressed as in Eq. (7). In this equation, β ∈ [0, 1) ensures that the summation of all discounted future rewards is always a finite number and prevents it from reaching infinity.

$$\boldsymbol{V^{\pi}(s) = \Re(s,\pi(s)) + \beta \sum \rho_{ss'} V^{\pi}(s')} \quad \textbf{(7)}$$

The MB-RL agent that follows the optimal policy can achieve the optimal V, calculated from Eq. (8).

$$\boldsymbol{V^{*}(s) = max_{\pi}[\Re(s,\pi(s)) + \beta \sum \rho_{ss'} V^{\pi}(s')]} \quad \textbf{(8)}$$

The best policy has been described as one that significantly improves V for any state s and can be calculated utilizing the formula below.

$$\boldsymbol{\pi^{*}(a/s) = argmax_{a\in A} \sum \rho_{ss'} V^{*}(s')} \quad \textbf{(9)}$$

Depending on Eq. (9) and the current state, the MB-RL agent chooses the actions used to manipulate the HVAC system's inputs. The optimal V and the best policy can be computed using two algorithms: value and policy iterations. In the present paper, the optimal value-iteration was applied, and DP was used for optimal scheduling of the indoor building services. The parameters applied in the DP-MB-RL controller are indexed in Table 2.

Table 2 Descriptions of the Proposed Controllers' Parameters.

Par.	Definition	Value	Unit
β	The discount index	0.990	-
δ	A trade-off between the energy-saving of reward's part and residents' comfort condition part	0.980	-
$\underline{T}_{rm-des} and \overline{T}_{rm-des}$	The desired boundaries of indoor heat	[20 24]	°C
$\underline{CO}_{2m-des} and \overline{CO}_{2m-des}$	The desired boundaries of internal CO_2	[750 850]	Ppm
ℜ(s,a)	Reward	---	-
βVπ(s')	The summation of discounted future rewards	---	-
Vπ(s)	Value-function	---	-
V*(s)	Optimal V-value	---	-
π*(a/s)	Optimal policy	---	-

Note: --- Signifies a variable

2.3.Deterministic Policy for the MB-RL Algorithm

Based on the MDP model, two approaches have been adopted to specify an appropriate action-selection strategy. Typically, these approaches involve stochastic and deterministic policy functions. The significant difference between these two algorithms can be expressed as the stochastic policies are integrated over state-action spaces, while deterministic policies only incorporate over state-spaces. Therefore, the stochastic policy requires more testing samples to compute the state-action space function [55, 56]. In summary, for the stochastic policy, every state in the state-space has a probabilistic distribution of action in that state. The DP describes the behavior that realizes the maximum anticipated reward over time and at any number of episodes [36]. The decision-maker manipulates the action space values to track desired indoor conditions (temperature and air quality) while optimizing performance to maximize energy-saving. The MB-RL controller's policy samples the DP and sets its parameters to achieve the best scheduling, as illustrated in Fig. 6. Meanwhile, the pseudo-code, as exposed in Table 3, is applied as the agent's guide to follow internal conditions changes. In other words, after computing the optimal V for the MB-RL control technique using the value-iteration process and optimizing the V-value calculated by Bellman's equation, a DP technique is used to obtain the optimum action space and update the policy-function factors. In summary, the DP maps every state in states set to a particular action in actions set, i.e., π(s)=a. The agent detects the reward function and then enters the next state to store the information in its memory M. This method is used periodically until the optimum state-space S is found. If this criterion is not satisfied, the DP-MB-RL agent will go to the next episode and repeat the above procedures.

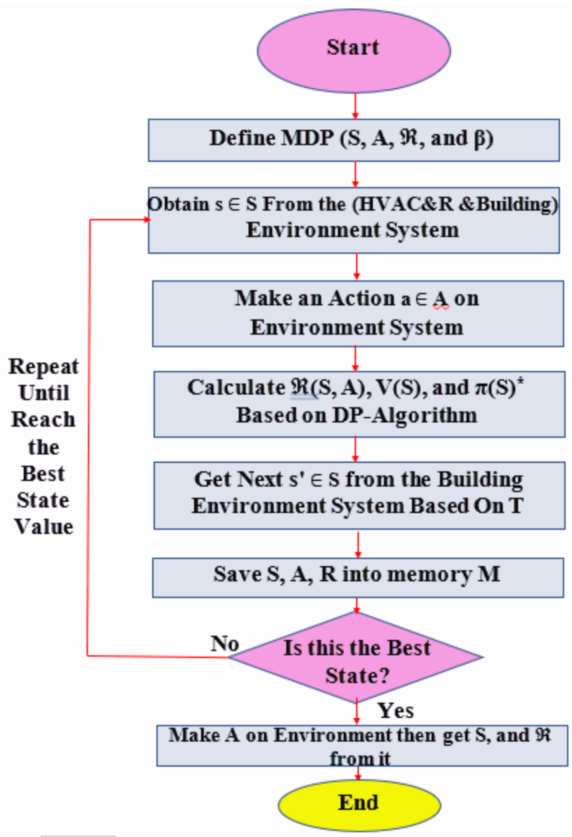


Fig. 6 DP-MB-RL Algorithm's Flowchart.

3.PERFORMANCE EVALUATION

In this section, the performance of the suggested DP-MB-RL controller applied to the HVAC&R system has been evaluated and compared with that of the benchmark controllers. The simulation results for all controllers have been carried out in MATLAB software. The main evaluation parameters are V-value, thermal and IAQ levels, and finally, the energy and cost savings.

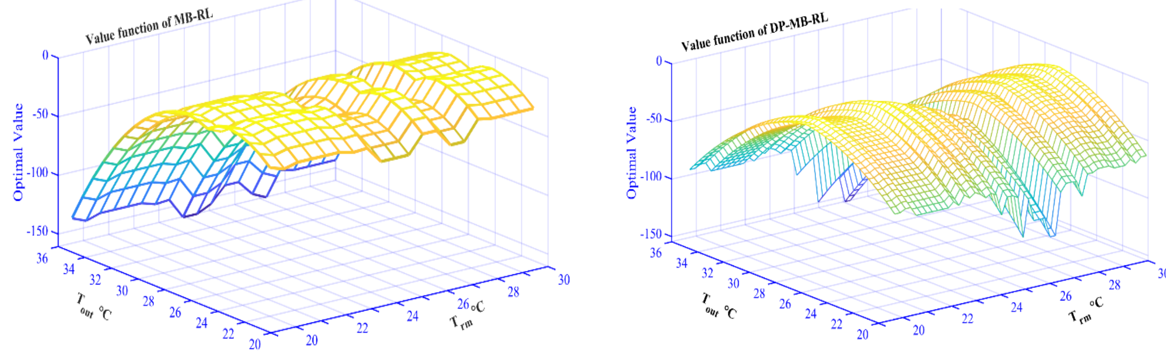
Table 3 The DP-MB-RL Algorithm Pseudo-Code.

1.) procedure MDP (S, A, β, $\rho_{ss'}$, δ, and ℜ)
S→ [$T_{rm}(t)$, $T_{out}(t)$, $CO_{2m}(t)$]

A→ [$C_{hf}(t)$, $D_{fa}(t)$, $L_{nON}(t)$, $W_{nON}(t)$]
ℜ (S, A) → Eq. 5
2.) For each element in state-space S, set the initial value for π(S) of each element in V(S) and A
3.) For I= 1 to 5
4.) Repeat for each element of S and A
5.) Repeat for Vπ(S) and then V*(S) → Eqs. 6 & 7.
6.) Based on the D.P.- method, for each state, find π*(S) =A
End For
7.) Make the following steps for continuing control
  a) Get S_{cnt} from the HVAC&R system
  b) Compute a'(S_{cnt}) from π*'(S_{cnt})
  c) Set the HVAC&R at a'(S_{cnt})
  d) Go to the first step a)
End procedure

3.1.The Performance of the DP-MB-RL Controller

This section covers and evaluates the different features of two controllers, MB-RL and DP-MB-RL. As shown in Fig. 3, a building is considered for an analytical case study. This real case is in Basrah City, Iraq. The controllers have been applied to the HVAC&R system to provide thermal comfort conditions and maximize energy saving. Figures 7 (a) and (b) show the simulation results for optimal V of MB-RL and DP-MB-RL controllers, respectively. These values were calculated using the value iteration technique. As shown in Fig. 7 (a), the optimal V of the DP-MB-RL controller has a smoother surface than the optimal V of the MB-RL controller, meaning that the DP-MB-RL agent had selected more appropriate action space values with greater consistency than MB-RL [57]. It is necessary to mention that three states, $T_{rm}(t)$, $T_{out}(t)$, and $CO_{2rm}(t)$, affect the optimal V. However, the state $CO_{2rm}(t)$ has less effect on optimal V than $T_{rm}(t)$ and $T_{out}(t)$. To avoid unnecessary complexity, the $CO_{2rm}(t)$ effect has not been considered in calculating the optimal V. The smoother surface of the optimal V, as shown in Fig. 7 (b), reduced the period of the oscillations in the actions-space chosen by the DP-MB-RL agent and provided accurate sequential decisions. As a result, the solenoid valve and air dampers' chattering effects were minimized, reflecting the steady state of the required indoor comfort ranges in the building and allowing the agent to achieve its purpose as quickly as possible.



(a) V of MB-RL **(b) V of DP-MB-RL**

Fig. 7 The Controllers' Optimal V-value.

After optimizing the V function, the DP-MB-RL agent chooses the best action-space values to warrant the maximum adaptation of the control policy. For each control time step, after performing actions, the reward received by this agent depends on the energy and violations of both temperature and CO2 concentration. The performances of DP-MB-RL and MB-RL methods for controlling the interior heat at each hour of the day have been evaluated, as shown in Fig. 8. In this figure, the outdoor temperature and the minimum and maximum acceptable indoor temperatures are also given. As shown in Fig. 8, the DP-MB-RL and MB-RL controllers kept the building residents' thermal comfort levels within the required bounds (20 °C to 24 °C). However, the adjusted set points for the DP-MB-RL method are more rapid, have lower oscillations, and are closer to the mean temperature, i.e., the mean value of minimum and maximum acceptable indoor temperatures, compared to the MB-RL method. Therefore, compared to MB-RL, the DP-MB-RL approach performs better at controlling the degree of interior temperature. In Fig. 9, the performances of DP-MB-RL and MB-RL methods are evaluated for controlling the CO2 concentration level at each hour of the day. As shown in Fig. 9, the IAQ for MB-RL and DP-MB-RL methods has been managed to meet the desired satisfaction levels, as determined by CO_2 concentration level. An acceptable CO_2 concentration level inside the room (the black line in Fig. 9) has been continuously represented for 24 hrs., using the Lagrange polynomial model. Indoor-acceptable CO_2 ranges between 550 and 1000 parts per million, which is profoundly affected by indoor personnel's consumed time [52, 58]. Numerous time points were selected to show carbon dioxide changes in concentration. Firstly, from midnight to 7:00 AM, the internal CO_2 concentration displayed an increasing tendency (the highest level from 800-1000 ppm) due to indoor residents. Then, between 7:00 AM and 3:30 PM, the people inside began leaving the place, and the CO_2 level decreased quickly to the smallest value (i.e., 550 ppm). Between 3:30 PM and midnight, the occupants started to enter the house and the carbon dioxide increased to maximum value [59]. The desired CO_2 concentration range was chosen between 750 and 850 ppm based on [58] since any value outside this range harms the occupants' health. The CO_2 concentrations were monitored, and those greater than 850 ppm were avoided. Figure 9 shows that the DP-MB-RL agent offers better stability and faster response performances than the MB-RL controller.

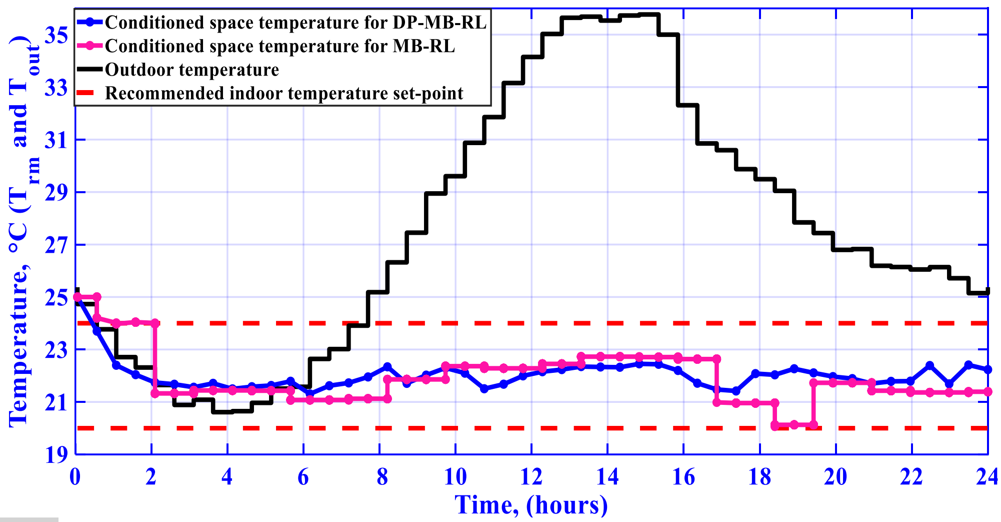


Fig. 8 The Comparison of HVAC System Thermal Response for MB-RL and DP-MB-RL.

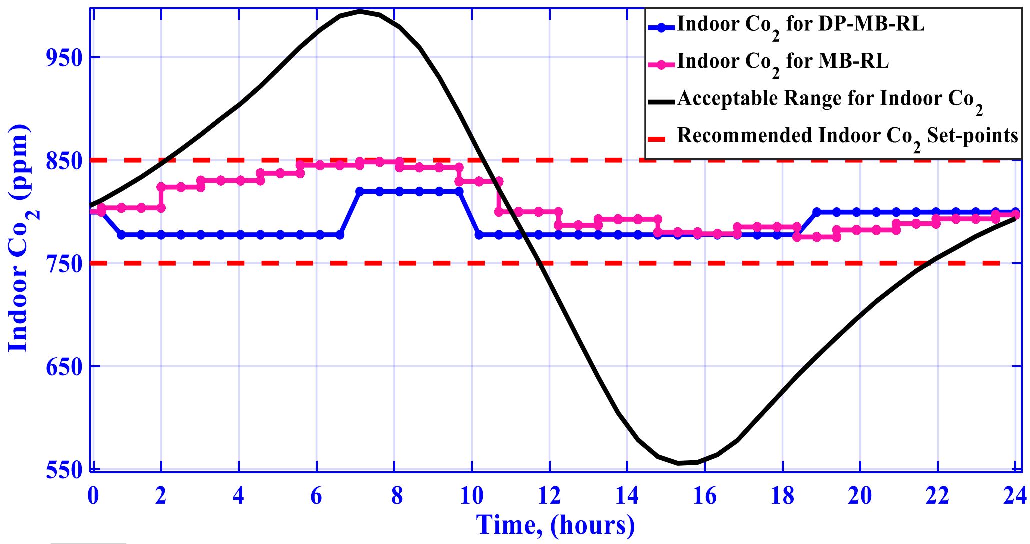


Fig. 9 HVAC System CO2 Concentration-Response for MB-RL and DP-MB-RL.

3.2. Evaluation of Energy Savings and Energy Costs

This section uses the HVAC system's energy usage to evaluate the DP-MB-RL and MB-RL controllers' efficiency throughout the day. It is necessary to mention that the electricity consumed by an HVAC system is directly relative to the cooling coil valve position. The chilled water flow rates are controlled by this valve. The position of this valve is controlled by DP-MB-RL and MB-RL agents. Figure 10 shows the position of the chilled water valve for both agents. As illustrated in this figure, the $C_{hf}(t)$ action is characterized by temperature control via regulating the flow rate of this cooling coil water based on $T_{rm}(t)$ and $T_{out}(t)$. When $T_{out}(t)$ is low and between the desired set points of $T_{rm}(t)$, the DP-MB-RL exploits this chance to open windows for the building ventilation process while switching off the lights. Therefore, to avoid the DP-MB-RL agent punishment, the HVAC&R system is switched off by closing the chilled water flow rate valve to save more energy than that without using the DP algorithm. Figure 11 represents the application of mass and energy conservation principles to the heat exchanger's control volume in an HVAC&R system to create a comprehensive energy equilibrium for this sub-system, as given by Eqs. (1) and (2). This control volume has been used to compute the energy usage of the HVAC&R system for a day. By specifying the position of the chilled water valve during the planning horizon for both controllers, as shown in Fig. 10, considering the control volume of the heat exchanger shown in Fig. 11, and using the relations of the heat exchanger model (Eqs. (1) and (2)), the cooling coil load can be determined, as shown in Fig. 12. To calculate the overall system's energy of the cooling coil load, iterative approaches have been applied to solve the related equations. Figure 12 summarizes the cooling coil load results for energy variation in the building for two controllers. This result shows the electrical power consumption (Kw) of an HVAC system by applying both controllers for 24 hrs. The power consumption increases, especially at peak times, to maintain the occupants' comfort levels in acceptable ranges. However, the energy expended by the HVAC&R system controlled by DP-MB-RL is lower than that of the MB-RL controller as it has a shorter duration of maximum power absorbed by the plant.

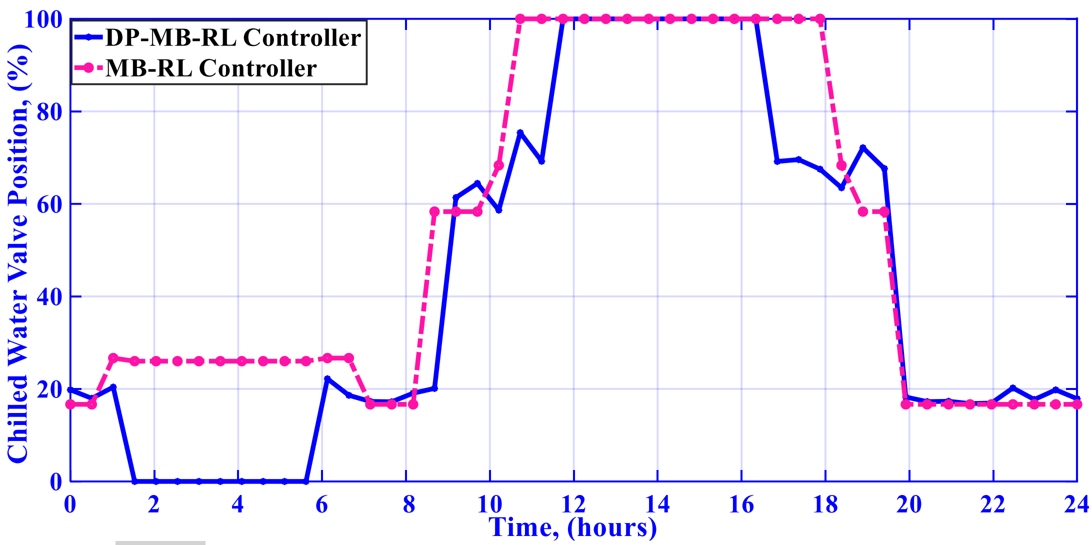


Fig. 10 Action of the Supplied Chilled Water for MB-RL and DP-MB-RL.

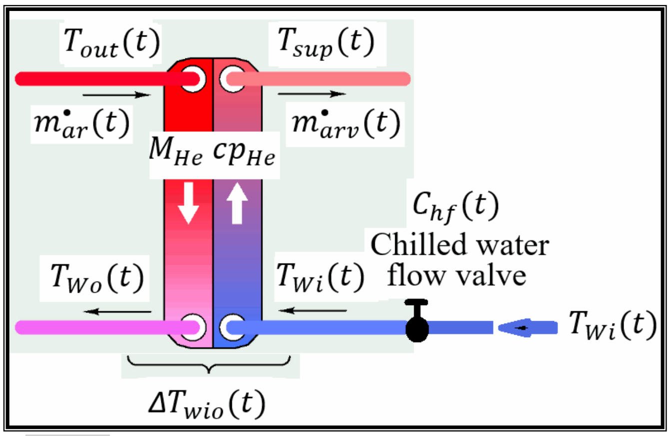


Fig. 11 Thermal Variation Through Heat Exchanger.

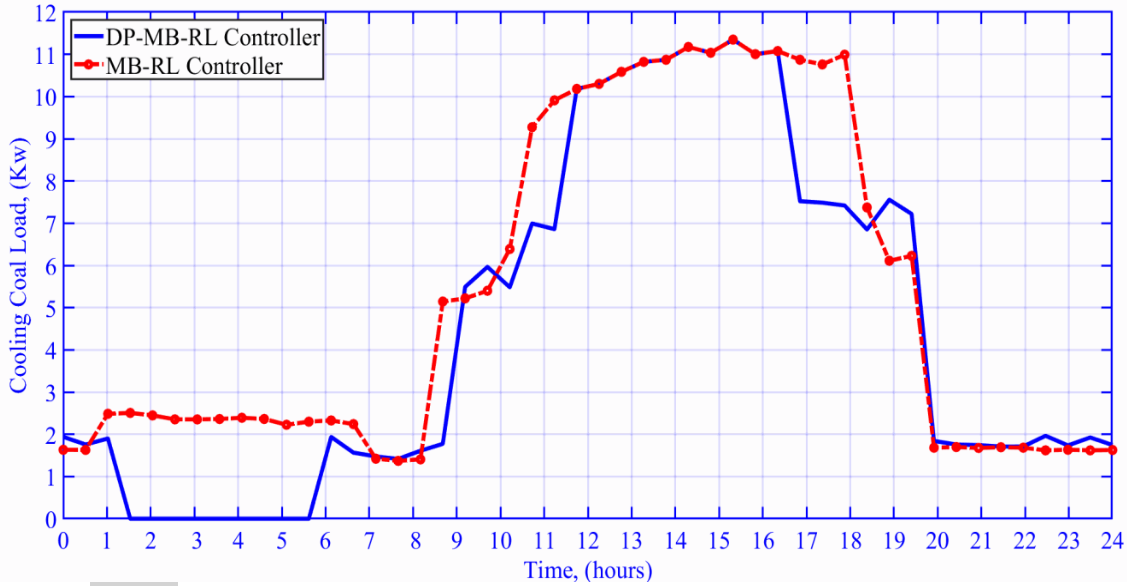


Fig. 12 HVAC System Cooling Coil Load for DP-MB-RL and MB-RL Controllers.

The HVAC&R energy usage during 24 hrs. period, can be seen in Fig. 13 for both controllers. Specifically, Fig. 13 illustrates the energy consumed for cooling the building in which the HVAC system is controlled by both controllers. As exposed in Fig. 13, using the DP-MB-RL and MB-RL controllers, 107.4 kWh/d and 126.4 kWh/d of energy, respectively, were used to cool the building for a day. Due to this fact, the proposed (DP-MB-RL) controller achieved the work's primary goal, which is more energy-saving. The system's energy efficiency has been calculated to be higher by 15.03% for this controller than the MB-RL strategy. As the temperature drops less than the upper level of the desired temperature at night, the DP takes advantage of this feature to open windows and allows the ventilation process into indoor space. Furthermore, at late time of night, the occupants do not require the extra lighting so that it will switch off indoor/outdoor nighttime running lights. The above actions reduce system energy consumption using the DP-MB-RL controller. Also, in this study, the cost-saving performance of both controllers was achieved under different electricity pricing schemes using FP, RTP, and TOU schemes. Figure 14 displays electricity pricing for an average day [60]. The HVAC&R system's energy expenses are analyzed by implementing the recommended and benchmark controllers, as shown in Fig. 15, depending on the cooling coil loads (kW) depicted in Fig. 12 and the electricity pricing ($/kWh). As illustrated in Fig. 15, the proposed method outperforms the MB-RL method since it uses less energy cost to run the HVAC&R system for the three pricing schemes. For the RTP, TOU, and FP schemes, the recommended controller reduced energy costs by 10%, 13.3%, and 15.1%, respectively, compared to the benchmark controller.

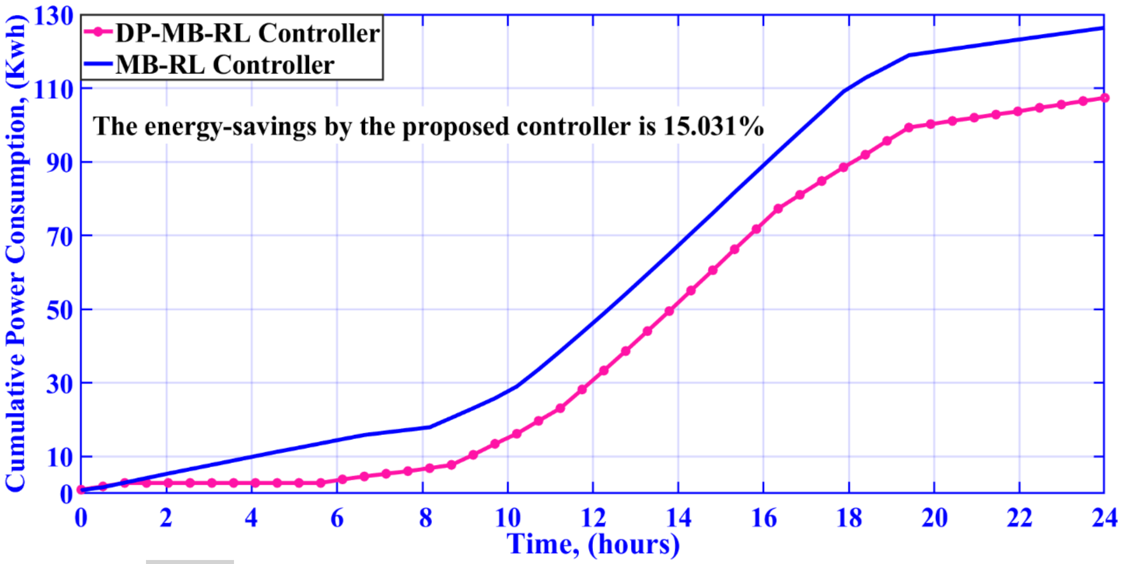


Fig. 13 Energy Consumed by the Building Model for Both Controllers.

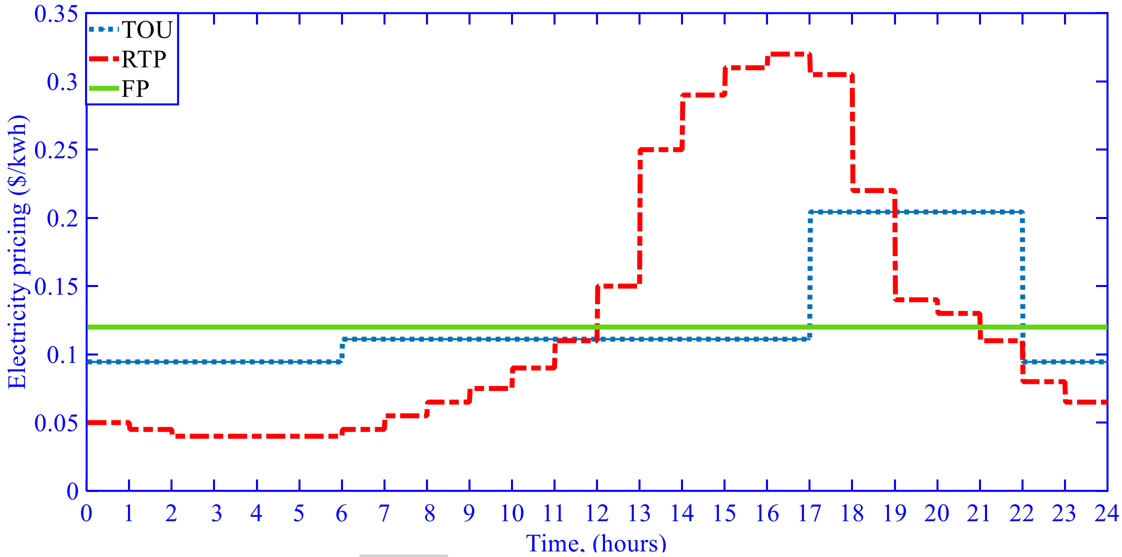


Fig. 14 Electricity Market Schemes.

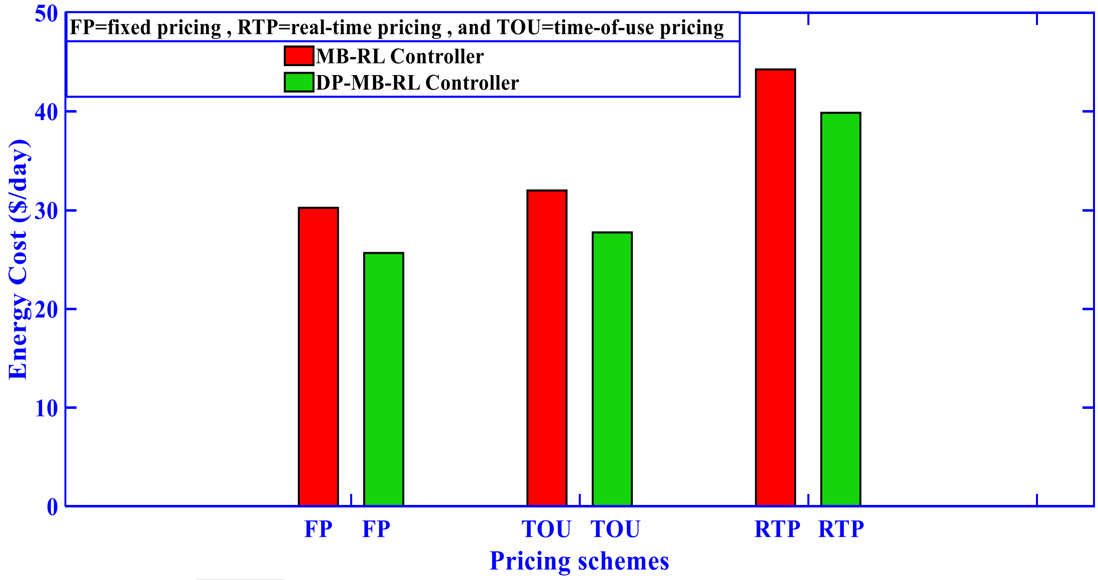


Fig. 15 Comparison of the Two Controllers' Electricity Costs.

3.3. Validation and Verification of the Proposed Controllers

For validation and verification of the proposed controllers' performances, the optimal method was used to determine the acceptable thermal circumstances inside the building based on the outdoor temperature. This method has been recommended by [59]. It is also applied in [60], [6]. The present study's control system performance has been compared with ASHRAE standard 55 suggested ranges for interior temperatures. Where ASHRAE standard 55 states the criterion for accepted operative temperature ($T_{opr}(t)$) limits into the air-conditioned areas [61], and it can be calculated using Eq. (10). Wherever, $T_{opr}(t)$ is the average of the inside air heat $T_{rm}(t)$ and the mean radiant heat $T_{rd}(t)$, respectively. $T_{opr}(t)$ can be calculated using the relationship below with acceptable accuracy [49]:

$$T_{opr}(t) = \frac{T_{rm}(t)\text{-}T_{rd}(t)}{2} \qquad (10)$$

As shown in Fig. 16, 90% acceptability limits were used for higher thermal comfort levels. obviously, the proposed controllers confirmed very good satisfaction as the indoor operative temperature within and interconnects with the ASHRAE standard recommended area [61, 62], as exposed in Fig. 17.

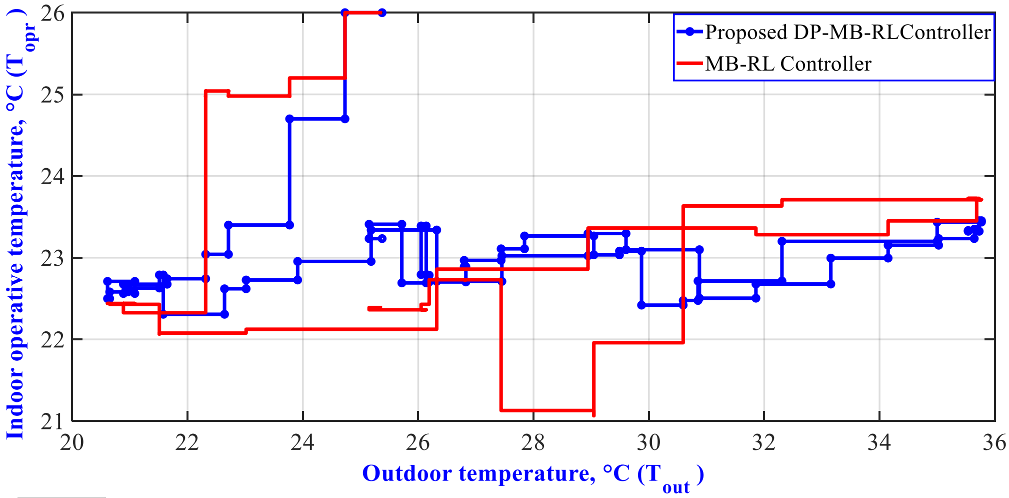


Fig. 16 Indoor Air Operative Temperature for MB-RL and DP-MB-RL Controllers.

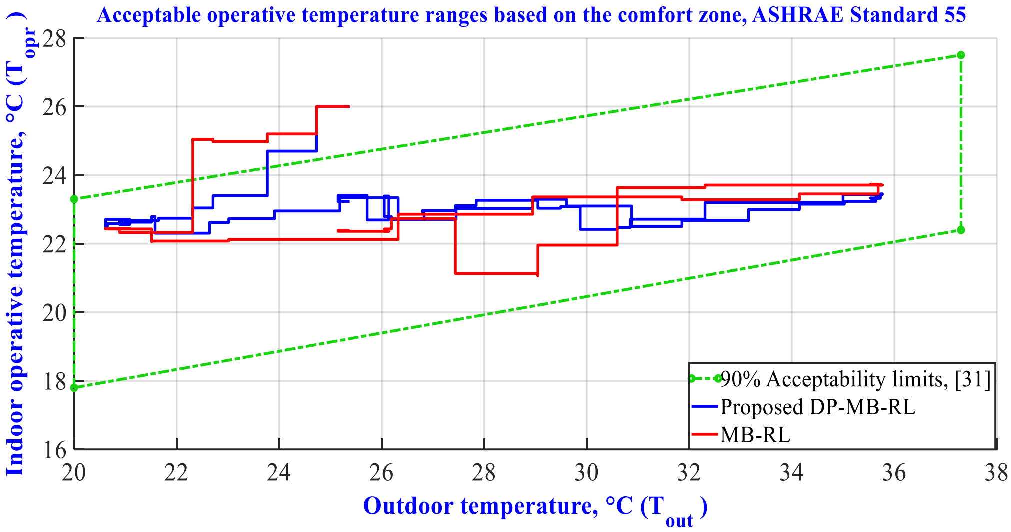


Fig. 17 Comparison of Indoor Air Operative Temperature for Two Controllers with ASHRAE Standard Acceptable Levels.

Also, the energy efficiency assessments of three MB-RL and Takagi-Sugeno Fuzzy (TSF) are represented in Fig. 18 for an HVAC system to verify the controllers' performance. This figure depicts the collective energy usage over 24 hrs for the MB-RL, DP-MB-RL control methods, and TSF controller reported in [47]. The DP-MB-RL agent performs better than the TSF controller for the same building, as shown in Fig. 18, by saving 21% more energy.

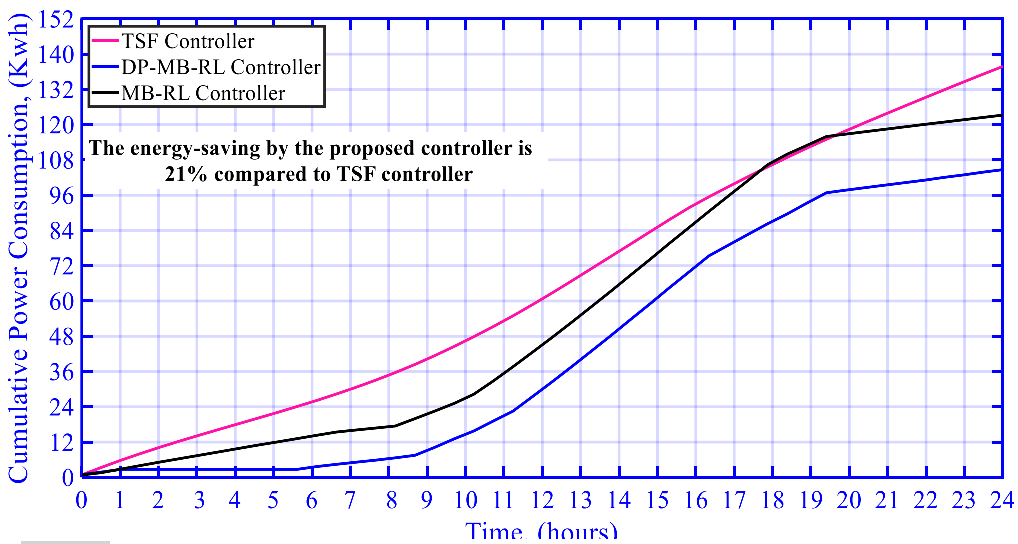


Fig. 18 Comparison of Energy Consumed by the Building for Three Controllers.

4.CONCLUSIONS AND RECOMMENDATIONS

This study provides the HVAC&R system best control via energy consumption minimization with maintaining indoor thermal and air quality simultaneously with minimizing energy costs for different electricity pricing schemes. First, a simple HVAC&R system thermodynamic model was designed and verified with two of the most significant terms for occupants' comfort levels: First) indoor air temperature, and second) CO_2 concentration level. For controlling the developed HVAC&R model, this research used two control methods: online traditional MB-RL and DP-MB-RL. Using the MB-RL algorithm makes the agent easily interact with its environment to increase control effectiveness simultaneously with less data and time and without the tedious trial-and-error process. To overcome the substantial increase in training data of the HVAC&R system control, a DP algorithm was employed to provide a DP-MB-RL control method for selecting the best actions within the MB-RL method. The simulation results revealed the superiority of high-dimensional and nonlinear HVAC&R control with no additional calculations, reducing the cost and time of the computations. Where the DP-MB-RL controller had to preserve indoor temperature tightly, IAQ with energy-saving was calculated to be higher by 15.03% and 21% than the MB-RL and TSF controllers, respectively. In addition, the energy cost for DP-MB-RL was cheaper than the MB-RL approach. To provide more stable indoor comfort levels, increase daily energy and cost savings, and further reduce calculation time, optimizing the proposed DP-MB-RL control method using deep learning methodologies is recommended. Also, in the future, this work can be applied to multizone HVAC&R systems or other types of buildings. The control method can also be expanded into a multi-agent system deep learning control methodology.

ACKNOWLEDGEMENTS

The authors are grateful for the financial support towards this research by the Department of Chemical and Petroleum Refining Engineering, College of Oil and Gas Engineering, Basra University for Oil and Gas. Department Research Grant (PGRG) No.153/HK/ (2024-3-26)/pg.43.

NOMENCLATURE

A_b	Surface area, m^2
cp_{He}	Specific heat of the cooling coil, J/(kg °C)
cp_{ar}	Specific heat of air, J/(kg °C)
cp_{wo}	Specific heat of water, J/(kg °C)
CO_{2gm}	Indoor generated CO_2 concentration level, *Ppm*
CO_{2out}	Outside carbon dioxide concentration, *Ppm*
$\underline{CO}_{2m-des}$ and $\overline{CO}_{2m-des}$	Desired boundaries of internal CO_2, *Ppm*
D_{ra} &D_{fa}	Fresh and return air ratios via damper
F_{am}	Volumetric airflow rate, $\left(m^3/_{sec.}\right)$
$G(t_i)$	Indoor CO_2 concentration at time t, *Ppm*
K	Conductivity
L_{nON}	On/off lighting
M_{He}	Heat transfer unit mass, kg
$m^{\bullet}_{wr} = C_{hf}$	Mass flow rate of chilled water, kg/(sec.)
$m^{\bullet}_{ar}$	Mass flow rate of outdoor air, kg/(sec.)
$m^{\bullet}_{avr}$	Mass flow rate of ventilation air, kg/(sec.)
$m^{\bullet}_{asr}$	Mass flow rate of supply air, kg/(sec.)
T_{Wi} and T_{Wo}	Water in/out temperatures of the heat exchanger, °C
T_m	Mixing temperature, °C
T_{sup}	Supply air temperature, °C
T_{rm}	Room heat, °C
T_{out}	External heat, °C
τ_h	Time delay for the cooling coil, sec.
τ_b	Time delay for the air-conditioned area, sec.
τ_c	Time delay for the CO_2 sensor, sec.
$t - t_i$	Time, *hrs*.
$\underline{T}_{rm-des}$ and $\overline{T}_{rm-des}$	Desired boundaries of indoor heat, °C
ΔT_{wio}	Difference of water's output and input temperatures, °C
T_{opr}	Operative temperature, °C
T_{rd}	Mean radiant heat, °C
$v^{\bullet}_r$	Volume rate of the room, $m^3/_{sec.}$
v_{room}	Volume of the building, m^3
$V\pi(s)$	Value-function
$V^*(s)$	Optimal V-value
Δx_b	Thickness, m
W_{nON}	Open/close windows
Greek symbols	
β	Discount index
δ	A trade-off between the energy-saving of reward's part and residents' comfort condition part.
$\Re(s,a)$	Reward
$\beta V\pi(s')$	The summation of discounted future rewards
$\pi^*(a/s)$	Optimal policy
Subscripts	
am	airflow
ar	air
asr	supply air
b	base
des	desired
fa and ra	fresh and return air
gm	generated
He	Heat exchanger
m	mixing
out	outside
opr	operative
rm	room
rd	radiant
sup	supply
avr	ventilation air
wr	water
Wi and Wo	water in/out

REFERENCES

[1] Chen Y, Norford LK, Samuelson HW, Malkawi A. **Optimal Control of HVAC and Window Systems for Natural Ventilation Through Reinforcement Learning.** *Energy and Buildings* 2018; **169**:195-205.

[2] Homod RZ, Togun H, Abd HJ, Sahari KSM. **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* 2020; **56**(June 2019):102091.

[3] Du Y, Zandi H, Kotevska O, Kurte K, Munk J, Kadir A, Evan M, Fangxing L. **Intelligent Multi-Zone Residential HVAC Control Strategy Based on Deep Reinforcement Learning.** *Applied Energy* 2021; **281**(November 2020):116117.

[4] Homod RZ, Almusaed A, Almssad A, Jaafar MK, Goodarzi M, Sahari KS. **Effect of Different Building Envelope Materials on Thermal Comfort and Air-Conditioning Energy Savings: A Case Study in Basra City, Iraq.** *Energy Storage* 2021; **34**:101975.

[5] Homod RZ, Gaeid KS, Dawood SM, Hatami A, Sahari KS. **Evaluation of Energy-Saving Potential for Optimal Time Response of HVAC Control System in Smart Buildings.** *Applied Energy* 2020; **271**(August):115255.

[6] Zhao H, Zhao J, Shu T, Pan Z. **Hybrid-Model-Based Deep Reinforcement Learning for Heating, Ventilation, and Air-Conditioning Control.** *Frontiers in Energy Research* 2021; **8**:412.

[7] Kim NK, Shim MH, Won D. **Building Energy Management Strategy Using an HVAC System and Energy Storage System.** *Energies* 2018; **11**(10):2690.

[8] Zhang Z, Chong A, Pan Y, Zhang C, Lam KP. **Whole Building Energy Model for HVAC Optimal Control: A Practical Framework Based on Deep Reinforcement Learning.** *Energy and Buildings* 2019; **199**:472-490.

[9] Kurte K, Munk J, Kotevska O, Amasyali K, Smith R, Mckee E, et al. **Evaluating the Adaptability of Reinforcement Learning Based HVAC Control for Residential Houses.** *Sustainability* 2020; **12**(18):1-38.

[10] Vázquez-canteli J, Ulyanin S, Kämpf J, Nagy Z. **Fusing TensorFlow with Building Energy Simulation for Intelligent Energy Management in Smart Cities.** *Sustainable Cities and Society* 2018; **45**:243-257.

[11] Azuatalam D, Lee W, Nijs F De, Liebman A. **Reinforcement Learning for Whole-Building HVAC Control and Demand Response.** *Energy and AI* 2020; **2**:100020.

[12] Bragagnolo SN, Schierloh RM, Vega JR, Vaschetti JC. **Demand Response Strategy Applied to Planning the Operation of an Air Conditioning System. Application to a Medical Center.** *Journal of Building Engineering* 2022; **57**:104927.

[13] Kang J, Weng S, Li Y, Ma T. **Study of Building Demand Response Method Based on Indoor Temperature Setpoint Control of VRV Air Conditioning.** *Buildings* 2022; **12**(4):415.

[14] Ahmad MW, Mourshed M, Yuce B, Rezgui Y, Chen Y, Norford LK, et al. **Computational Intelligence Techniques for HVAC Systems: A Review.** *Building Simulation* 2016; **9**(4):359-398.

[15] Seyedzadeh S, Rahimian FP, Glesk I, Roper M. **Machine Learning for Estimation of Building Energy Consumption and Performance: A Review.** *Visualization in Engineering* 2018; **6**(1):1-20.

[16] Moubayed A, Injadat M, Nassif AB, Lutfiyya H, Shami A. **E-Learning: Challenges and Research Opportunities Using Machine Learning Data Analytics.** *IEEE Access* 2018; **6**:39117-39138.

[17] Zhang C, Kuppannagari SR, Kannan R, Prasanna VK. **Building HVAC Scheduling Using Reinforcement Learning via Neural Network-Based Model Approximation.** *The 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, New York, USA, 2019; 287-296.

[18] Ding X, Du W, Cerpa AE. **MB2C: Model-Based Deep Reinforcement Learning for Multi-Zone Building Control.** *The 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, New York, USA, 2020; 50-59.

[19] Ahn KU, Park CS. **Application of Deep Q-Networks for Model-Free Optimal Control Balancing Between Different HVAC Systems.** *Science and Technology for the Built Environment* 2020; **26**(1):61-74.

[20] Yuan X, Pan Y, Yang J, Wang W, Huang Z. **Study on the Application of Reinforcement Learning in the Operation Optimization of HVAC System.** *Building Simulation* 2020; **14**:75-87.

[21] Qiu S, Li Z, Li Z, Zhang X. **Model-Free Optimal Chiller Loading Method Based on Q-Learning.** *Science and Technology for the Built Environment* 2020; **26**(8):1100-1116.

[22] Dalamagkidis K, Kolokotsa D, Kalaitzakis K, Stavrakakis GS. **Reinforcement Learning for Energy Conservation**

and Comfort in Buildings. *Building and Environment* 2007; **42**(7):2686-2698.

[23] Fazenda P, Veeramachaneni K, Lima P, Reilly UO. **Using Reinforcement Learning to Optimize Occupant Comfort and Energy Usage in HVAC Systems.** *Ambient Intelligence and Smart Environments* 2014; **6**(6):675-690.

[24] Ruelens F, Claessens BJ, Vandael S, Schutter B De, Member S, Babuška R, Belmans R. **Residential Demand Response of Thermostatically Controlled Loads Using Batch Reinforcement Learning.** *IEEE Transactions on Smart Grid* 2016; **8**(5):2149-2159.

[25] Ruelens F, Iacovella S, Claessens BJ, Belmans R. **Learning Agent for a Heat-Pump Thermostat with a Set-Back Strategy Using Model-Free Reinforcement Learning.** *Energies* 2015; **8**(8):8300-8318.

[26] Vázquez-Canteli J, Kämpf J, Nagy Z. **Balancing Comfort and Energy Consumption of a Heat Pump Using Batch Reinforcement Learning with Fitted Q-Iteration.** *Energy Procedia* 2017; **122**:415-420.

[27] Chen B, Cai Z, Bergés M. **Gnu-RL: A Precocial Reinforcement Learning Solution for Building HVAC Control Using a Differentiable MPC Policy.** *The 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, New York, USA, 2019; 316-325.

[28] Wang Y, Velswamy K, Huang B. **A Long-Short Term Memory Recurrent Neural Network-Based Reinforcement Learning Controller for Office Heating Ventilation and Air Conditioning Systems.** *Processes* 2017; **5**(3):46.

[29] Zhang Z, Chong A, Pan Y, Zhang C, Lu S, Lam KP. **A Deep Reinforcement Learning Approach to Using Whole Building Energy Model for HVAC Optimal Control.** *Building Performance Analysis Conference and Simbuild*, ASHRAE and IBPSA-USA, Chicago, 2018; 22-23.

[30] Polydoros A, Nalpantidis L. **Survey of Model-Based Reinforcement Learning: Applications on Robotics.** *Intelligent & Robotic Systems* 2017; **86**(2):153-173.

[31] Forootan MM, Larki I, Zahedi R, Ahmadi A. **Machine Learning and Deep Learning in Energy Systems: A Review.** *Sustainability* 2022; **14**(8):4832.

[32] Heidari A, Mar F, Khovalyg D. **Reinforcement Learning for Occupant-Centric Operation of Residential Energy System: Evaluating the Adaptation Potential to the Unusual Occupant's Behavior During COVID-19 Pandemic.** *CLIMA 2022 Conference The 14th REHVA HVAC World Congress*, Rotterdam, 2022; 1-7.

[33] Ardabili S, Abdolalizadeh L, Mako C, Torok B, Mosavi A. **Systematic Review of Deep Learning and Machine Learning for Building Energy.** *arXiv preprint arXiv:2202.12269* 2022; **10**(March):1-19.

[34] Biemann M, Scheller F, Liu X, Huang L. **Experimental Evaluation of Model-Free Reinforcement Learning Algorithms for Continuous HVAC Control.** *Applied Energy* 2021; **298**(May):117164.

[35] Hussein LA, Ateeq AA, Homod RZ. **Energy Saving by Reinforcement Learning for Multi-Chillers of HVAC Systems.** *2nd International Multi-Disciplinary Conference Theme: Integrated Sciences and Technologies, IMDC-IST 2021*, Sakarya, Turkey, 2021; 118.

[36] Dawood SM, Hatami A, Homod RZ. **Trade-Off Decisions in a Novel Deep Reinforcement Learning for Energy Savings in HVAC Systems.** *Journal of Building Performance Simulation* 2022; **15**(6):809-831.

[37] Gao G, Li J, Wen Y. **Energy-Efficient Thermal Comfort Control in Smart Buildings via Deep Reinforcement Learning.** *arXiv preprint arXiv:1901.04693* 2019; 1-11.

[38] Jiang Z, Risbeck MJ, Ramamurti V, Murugesan S, Amores J, Zhang C, et al. **Building HVAC Control with Reinforcement Learning for Reduction of Energy Cost and Demand Charge.** *Energy and Buildings* 2021; **239**:110833.

[39] Abdulgader M, Lashhab F. **Energy-Efficient Thermal Comfort Control in Smart Buildings.** *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*, NV, USA, 2021; 0022-0026.

[40] Jabari F, Mohammadi-ivatloo B. **Short-Term Co-Optimization of Multi-Chiller Plants and Ice Storage System.** *2018 Smart Grid Conference (SGC)*, 2018; 1-6.

[41] Ma G, Wang Z, Yuan X, Zhou F. **Improving Model-Based Deep Reinforcement Learning with Learning Degree Networks and Its Application in Robot Control.**

Journal of Robotics 2022; **2022**(1):7169594.
[42] Dazeley R, Vamplew P, Cruz F. **Explainable Reinforcement Learning for Broad-XAI: A Conceptual Framework and Survey.** *Neural Computing and Applications* 2023; **35**(23):16893-16916.
[43] Homod RZ, Sahari KSM, Almurib HAF, Nagi FH. **Double Cooling Coil Model for Non-Linear HVAC System Using RLF Method.** *Energy and Buildings* 2011; **43**(9):2043-2054.
[44] Homod RZ, Mahlia TMI, Mohamed HAF. **PID-Cascade for HVAC System Control.** *The Second International Conference on Control, Instrumentation and Mechatronic Engineering (CIM09)*, Malacca, Malaysia, 2009; 598-603.
[45] Chiang ML, Li-Chen F. **Hybrid System Based Adaptive Control for the Nonlinear HVAC System.** *Proceeding of the Conference on American Control*, Minneapolis, MN, USA, 2006; 5324-5329.
[46] Chiang ML, Li-Chen F. **Adaptive Control of Switched Systems with Application to HVAC System.** *IEEE International Conference on Control Applications*, Singapore, 2007; 367-372.
[47] Dawood SM, Hatami A, Homod RZ. **HVAC System Modeling and Control Methods: A Review and Case Study.** *Journal of Energy Management and Technology* 2022; **6**(4):217-231.
[48] ASHRAE. **Ventilation for Acceptable Indoor Air Quality: ASHRAE Standard 62.** Atlanta, USA: American Society of Heating, Refrigerating and Air-Conditioning Engineers; 2001. Available from: https://www.ashrae.org/technical-resources/bookstore/standards-62-1-62-2
[49] Homod RZ, Salleh K, Sahari M, Almurib HAF. **Energy Saving by Integrated Control of Natural Ventilation and HVAC Systems Using Model Guide for Comparison.** *Renewable Energy* 2014; **71**:639-650.
[50] Amouei A, Aghalari Z, Zarei A. **Evaluating the Relationships Between Air Pollution and Environmental Parameters with Sick Building Syndrome in Schools of Northern Iran.** *Indoor and Built Environment* 2019; **28**(10):1422-1430.
[51] Wang H, Xie L, Liu S. **A Model-Based Control of CO2 Concentration in Multi-Zone ACB Air-Conditioning Systems.** *2016 12th IEEE International Conference on Control and Automation (ICCA)*, Kathmandu, Nepal, 2016; 467-472.
[52] Baghaee S, Ulusoy I. **User Comfort and Energy Efficiency in HVAC Systems by Q-Learning.** *2018 26th Signal Processing and Communications Applications Conference (SIU)*, Izmir, Turkey, 2018; 1-4.
[53] Chapra SC, Canale RP. **Numerical Methods for Engineers.** 6th ed. New York, USA: McGraw-Hill; 2011.
[54] Ayoub A, Jia Z, Szepesv C, Lin W. **Model-Based Reinforcement Learning with Value-Targeted Regression.** *37th International Conference on Machine Learning*, 2020; 463-474.
[55] Silver D, Lever G, Heess N, Degris T, Wierstra D, Riedmiller M. **Deterministic Policy Gradient Algorithms.** *International Conference on Machine Learning*, Beijing, China, 2014; 387-395.
[56] Rijal HB, Humphreys MA, Nicol JF. **Development of a Window Opening Algorithm Based on Adaptive Thermal Comfort to Predict Occupant Behavior in Japanese Dwellings.** *Japan Architectural Review* 2018; **1**(3):310-321.
[57] Noel MM, Pandian BJ. **Control of a Nonlinear Liquid Level System Using a New Artificial Neural Network-Based Reinforcement Learning Approach.** *Applied Soft Computing Journal* 2014; **23**(October):444-451.
[58] Wang L, Wang Z, Yang R. **Intelligent Multiagent Control System for Energy and Comfort Management in Smart and Sustainable Buildings.** *IEEE Transactions on Smart Grid* 2012; **3**(2):605-617.
[59] Yuan Z, Huang Y, Lu X, Huang J, Liu Q, Qi G, Cao Z. **Measurement of CO2 by Wavelength Modulated Reinjection Off-Axis Integrated Cavity Output Spectroscopy at 2 μm.** *Atmosphere (Basel)* 2021; **12**(10):1247.
[60] Talebi A, Hatami A. **Online Fuzzy Control of HVAC Systems Considering Demand Response and Users' Comfort.** *Energy Sources, Part B: Economics, Planning, and Policy* 2020; **15**(7-9):403-422.
[61] Turner SC, et al. **ANSI/ASHRAE Standard 55-2010, Thermal Environmental Conditions for Human Occupancy.** Atlanta, GA: American Society of Heating, Refrigerating and Air-Conditioning Engineers; 2011. Available from: www.ashrae.org
[62] ASHRAE. **Standard 55-2004. Thermal Environmental Conditions for Human Occupancy.** Atlanta, USA:

American Society of Heating, Refrigerating and Air-Conditioning Engineers; 2004. Available from: https://webstore.ansi.org/standards/ashrae/ansiashrae552004