

Energy Saving by Reinforcement Learning for Multi-Chillers of HVAC Systems

Lubna A. Hussein¹, Adnan A. Ateeq², Raad Z. Homod³

{lubna.ali15783@gmail.com¹, adnan_ateeq@stu.edu.iq², raadahmood@yahoo.com³}

Engineering Technical College of Basra, Southern Technical University, Iraq^{1,2}
Oil and Gas Engineering Department, Basra University for Oil and Gas, Iraq³

Abstract. This paper presents a method for controlling and operating a multi-chillers system: (1) Model-based control approach was used by MATLAB/SIMULINK to model a building containing two non-identical chillers depending on thermal loads. (2) ON/OFF all chillers alternately using the model reinforcement learning controller (RL-control) to select the appropriate chiller for the building conditioning process. The results were in terms of energy efficiency and performance of the enhanced learning control for the chiller, and a control unit signal (PID) was applied to make a comparison with the signals of energy, power, and temperatures. After comparison, it was found that the energy saving through the proposed controller is 45% of the traditional (PID) strategy, where can the proposed strategy control for the chiller appropriate for the building's conditioning process.

Keywords: Multi-chillers system, Reinforcement Learning, Model-based control, Chilled water system.

1 Introduction

The Heating, Ventilating, and Air Conditioning (HVAC) system is a vast field. In general, the categories of building service and (HVAC) systems make up the main sources of energy use in buildings, and it is estimated that HVAC is responsible for 50% of the building's total energy consumption. Buildings are responsible for a considerable amount of energy use as the global building strip currently consumes about 40% of the overall energy produced [1][2][3]. In the HVAC system same, the most energy-consumption portion is the chiller unit system, that made up of chillers, pumps, cooling towers, and some other equipment. The chiller which is widely used to provide the cooling load is one of the main energy-consuming equipment, the energy consumption reduction of the multi-chillers system is of great importance in the HVAC system [4]. Reinforcement learning (RL) was initially demonstrated as a practical approach to control HVAC systems [5]. System selection depends on the building configuration, the environmental conditions, which HVAC design should consider different agents such as load variations, operating precision, Energy preservation, and total cost, and flexibility, etc. while selecting the equipment, and considered some standards that can be such as environment alteration (e.g., temperature, humidity, and pressure). All water systems are suitable for low-cost centralized HVAC systems in high-rise commercial applications, and it is the ideal solution to replace all air systems in high-rise buildings and is more suitable in buildings requiring individual control [6][7].

Liu and Henze (2006) have been proposed Simulated RL, a hybrid control planner that combines aspects of model-based optimum control with model-free learning control — is used

to control the building. The feasibility performance testing demonstrated that employing both active and negative thermal storage inventories, the hybrid control strategy can supply reliable control. Using the measured data, the hybrid control technique realized an 8.3 % cost provides during the basis state [8]. Dalamagkidis et al. (2007) the RL controller is created and simulated in the MATLAB/SIMULINK environment [9]. Fan et al. (2011) proposed an optimal control strategy for a multi-chiller system based on the probability density distribution of cooling load ratio,. At the same time, unnecessary chillers switch on/off behaviour can be prevented [10]. Huang et al. (2015) suggested chiller scheduling based on cooling demands Their goal was to control the on/off the multi-chiller systems to reduce the total energy consumption when compared with two existent sequence control methods—the optimal load distribution technique and the critical points reset strategy based on cooling capacity [11]. Azuatalam et al. (2020) proposed a new RL architecture for the control HVAC system with leveraging its demand response possible, design an active RL controller for a whole building model in compared to a handcrafted baseline controller, simulation results show that using RL for plain HVAC process will result in maximum weekly energy decreasing of up to 22% [12]. Yao et al. (2004) a mathematical model of optimal running for the refrigeration system are developed, the results showed that by applying the optimal model to the refrigeration system, energy savings of up to 10% are possible [13]. Qiu et al. (2020) this approach is based on Q-learning, a classic RL method, and it proposes a model-free optimal chiller loading method for optimizing chiller operating by altering the set points of the chilled water exit temperature. According to the simulation results, the proposed method can save 4.36 % of chiller energy during the first refrigerating season when compared to the baseline control [14]. Qiu et al. (2020) a model-free optimal control method depends on RL is proposed to control building refrigerating water, [15]. Homod et al. (2021) the HVAC system established was adopted as the system performance model; it provided dynamic system modelling, simulation, and monitoring cooling system performance. The test results show that significant energy saving through reducing the system operation cost and that by improving the four variables: the cooling water supply temperature (T_{cws}) from the chillers, the chilled water supply temperature (T_{chw}) from the chillers, (T_{chw}) from the heat exchangers, and the supply air temperature of the air handling units (AHU) [16]. The main objectives were to study the operating strategies of chillers alternately as follows:

1. A model-based control approach was used by MATLAB/Simulink to model a building containing two non-identical chillers depending on thermal loads.
2. ON/OFF for all chillers alternately using the model reinforcement learning controller (RL-control) to select the appropriate chiller for the building conditioning process.

HVAC system modelling is necessary to study and organize energy-consuming and kind of internal environment [17]. In general, three types of modelling are used for HVAC systems:

- a) The black box models: The black-box model concept is an appropriate transport function model to the input/output real model data to produce polynomials coefficient, and relevance between the input and output variables is identified using mathematical techniques. Use of the black-box model is at times preferred because it is simple to construct and there is no requirement to know the system's internal structure [18][19]. Black-box models are simple to build and computationally efficient [20].
- b) The white box models: The system models are created utilizing the governing laws and extensive understanding of the underlying process in this type of approach,

which is known as a physics-based method. Mathematical models are commonly used to represent the processing signal in the realm of HVAC system modelling. This signal, however, is formed using conservation rules such as components, mass, momentum, and energy balance [18][19][20].

- c) The Gray box models: Compared to black-box models and white-box models, these models have good precision and better generalization capabilities, but they are also the hardest to build. For some (HVAC) subsystems, the underlying physical phenomenon may be difficult to represent [17][18][19].

2 Chillers

Chillers are important commercial components as they expand the cooling capabilities of (HVAC) systems. When choosing a chiller, you must correspond to the requirements for the cooling temperature and how efficient they are. For large complexes of buildings used large chiller or multiple chillers to produce chilled water that can be provided to meet the air-conditioning demands of the different buildings [21][22]. And to meet different load requirements, air-conditioned buildings contain several centralized refrigeration systems for multi chillers. Different types of chillers are used in (HVAC) systems, the chiller may be the best choice for one application, but it may be too expensive or inconvenient for another [22][23].

Air-cooled chillers (Direct chillers): These chillers remove heat from the refrigerant by using fans that push air through tubes to condense the hot refrigerant. Air is a poor conductor of heat as compared to water, so air-cooled chillers are larger and less effective. Air-cooled devices have a higher compressor ratio, which means they use less energy to cool per watt [24][25].

Water-cooled chillers (Indirect chillers): This type of chiller removes heat utilizing pumps that send water through the condenser to recycle the water from the cooling tower to condense the refrigerant. They work well in medium and large commercial HVAC installations because they aren't affected by the ambient temperature. These chillers are more efficient because of water's high heat capacity but cost more to install and maintain [25].

In HVAC plants of medium-high refrigeration capacity, multiple chiller systems (MCSs) are often adopted, these systems consistency of multiple chillers, as the arrangement of chillers are differentiated through performance. The series arrangement is better because it can transfer heat, but it is not with a better efficiency than Parallel. Every chiller can be operated at various capacities to meet various refrigeration demands [26][27].

One of the most important problems with MCSs is to save more energy by reducing the total electrical energy consumption of the chillers, since the electrical consumption of MCSs increases significantly when the chillers are poorly handled; improving the efficiency of MCSs may result in substantial energy savings. The amount of energy saved by converting from a single chiller to a multi-chiller cooling plant varies depending on the type of building and environment [28].

The use of multiple chillers compared to a single chiller provides some advantages such as operational flexibility, standby capacity, maintenance is less disruptive, as well as using more than one chiller is better economical and reducing energy cost in the case of partial loading. Also, the coefficient of performance (COP) is better and will give high efficiency, as if we use one chiller and with high capacity, its cost will be high as well, and because these systems

depend on many factors including (capacity, partial load efficiency of each chiller, selection of refrigeration unit, refrigeration circuit design, type and number of compressors), etc [21].

3 Reinforcement Learning (RL)

Reinforcement learning (RL) is a branch of machine learning that studies how software agents should behave in each environment to maximize the concept of cumulative reward. Along with supervised and unsupervised learning, the RL is one of three fundamental machine learning paradigms. In the RL domain, we will only provide our algorithms with a (reward function), which refers to a learning factor when it is performing well, and when it is performing poorly, the learning algorithm's task is to know how to choose actions over time to obtain big rewards. The goal of "RL" is to guide the agent to determine what action to take that maximizes (or minimizes) the sum of all RL signals (the numerical reward) or punishment, it receives over time, called the total expected reward [29].

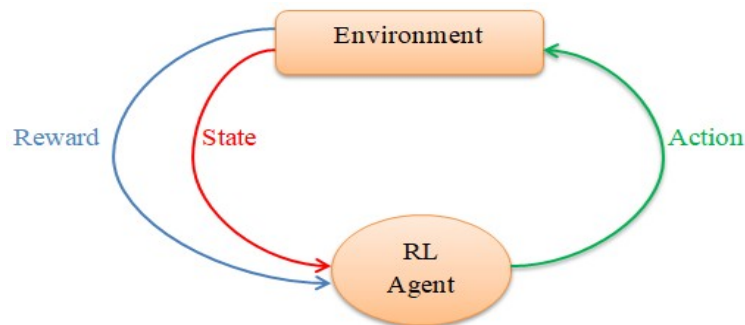


Fig. 1. RL Agent-Environment Interaction.

According to Fig. 1. If the agent performs one of many possible acts on the environment, it arrives in one of many possible states and is rewarded or punished for that specific action. This reward is intended to provide feedback to the agent about how well it is performing at each decision time step [12].

Markov Decision Processes (MDPs) are a fundamental formalism for reinforcement learning (RL). The method of sequential agent-environment interactions including state, actions, and rewards in the RL trouble can be expressed as an MDP since it includes sequential decision making, the problem consists of sequential decisions that an action (A) must be taken in each state (S) that is by the agent [30][31].

The major components for the RL system are agent, state, action, policy, reward signal, value function, and model. Agent: In RL, an agent learns to accomplish an objective by reacting with the environment repeatedly. State (s): (RL) concentrated on learning how good it is for the agent to be in each state in the long run. State indicates the current position returned by the environment. Action (a): actions are the Agent's method for reacting to and varying its environment, and thus transporting between states. Any action the Agent takes results in a reward from the environment [30][32]. Policy (π): a policy is the learning agent's way of behaviour at a specific time. Simply the policy maps something that perceived states of the environment to actions to be taken when in those states. The policy is the essence of the "RL" agent in the mean that it is a strategy that applies by the agent to determine the next action

based on the current state. Finding the optimal policy is the basic object of the (RL) process [31]. Reward signal (R): the reward signal (R) refers to how good or poor an occurrence is, and it determines the problem's target, which is for the agent to increase the total reward obtained. The better action is chosen by values of states or actions because the higher value gets about the biggest amount of reward in the long run. The learning agent will then be able to optimize the total reward it gets. Depending on the action taken, the agent receives a positive or negative reward [30][31]. Value function (V): The value function is an expectation of the future overall rewards, specifies which is good in the long run. The state-value function $V^\pi(s)$ is the predicted return when beginning from a state "s" [31].

Model-based methods, such as dynamic programming, require a model of the environment, whereas model-free methods are learning without one. Model-based methods study the transmission and reward models from reaction to the environment and then utilize the learned model to compute the optimal policy through value iteration.. Model-free methods optimize the value function instantly from the observed experiment and do not depend on the transition and reward models, value functions are learned by trial and error [30].

4 Description of the Building

The Basra International Airport building model was chosen to study the operating strategies of chillers alternately. This building includes 2-chillers "York" and the type of gas used is R-134a. Fig. 2, photo of the real system (chiller 1) in Basra International Airport. The basic information about the building and chiller system is listed in Table 1:

Table 1. Building and Chillers System Data.

Building location	Basra, Iraq
Cooling system	Water-cooled
Pump system	The primary pump, the secondary pump
Chillers	2 units
Setpoint of temperature	5.6°C
Compressor	Centrifugal Chiller
Capacity	1900 ton
Cooling towers	6 units



Fig. 2. Photo of the Real System (Chiller 1) in Basra International Airport.

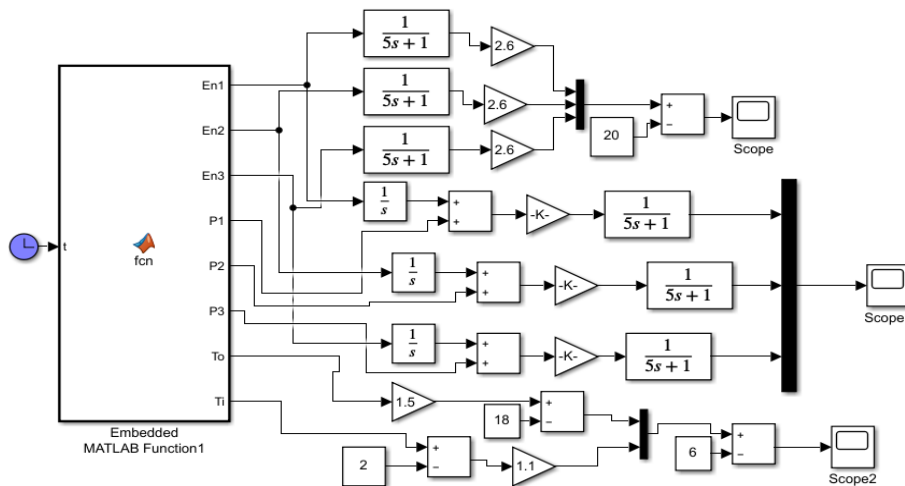


Fig. 3. Simulation of the Cooling System.

4.1. Building structure, simulation of cooling system

The proposed model was built to determine the suitable chiller for a building adaptation process using a model-based control approach. The major advantage of choosing model-based learning. As shown in Fig. 3, the model of the cooling system for an airport building by using MATLAB / SIMULINK. This model represents the products of the solutions (white box), the mathematical model. First, the thermal loads of the airport building are calculated. By calculating these loads, we find the signals of energy, power, and temperatures. The method recommended is for a chilled water system with non-identical units (i.e., non-identical chillers). The chilled water system should also be disconnected (constant primary chilled

water flow rate and variable secondary chilled water flow rate). This study depended on the computation of heating and cooling loads for opaque surfaces [33].

4.1.1. Opaque surfaces

The rate of energy stored can be calculated by applying the basic theory of energy conservation to the HVAC process [33]:

$$\frac{dE_s}{dt} = \sum_i \dot{E}_{in} - \sum_i \dot{E}_{out} \quad (1)$$

Where denotes the rate of change in the system's total stored energy, and , denotes the rate of net energy transfer. The left-hand side of Eq. 2 represents the cumulative rate of energy of opaque surfaces, through the control volume from inside to outside [33].

$$M_{wl} c p_{wl} \frac{dT_{wl,t}}{dt} = \sum_i \dot{Q}_{opq_{in}} - \sum_i \dot{Q}_{opq_{out}} \quad (2)$$

is the heat capacitance (J/K), and and are the heat gains. The surface area , and the surface cooling factor determine how much heat enters) through the control volume across opaque surfaces. Thus, it is possible to write it mathematically as follows [34]:

$$\dot{Q}_{opq_{in}} = A_{w_j} \times C F_{opq_j} \quad (3)$$

$$C F_{opq} = U(O F_t \Delta t + O F_b + O F_r D R) \quad (4)$$

Where represents the building -factor, W/(m².K), represents the cooling design temperature differential (°C), , , and represent the opaque surface cooling factors, and represents the cooling daily range (K) [33].

$$\dot{Q}_{opq_{in}} = \sum_j A_{w_j} U_j O F_t (T_{wl_{ou}} - T_{wl_{in}}) + \sum_j A_{w_j} U_j O F_b + \sum_j A_{w_j} U_j O F_r D R \quad (5)$$

The heat transfer due to convection into the conditioned area is represented by the second part in Eq. 2. For convection heat transmission, Newton's law of cooling is as follows [35]:

$$\dot{Q}_{opq_{out}} = \sum_j A_{w_j} h_{i_j} (T_{wl_{in}} - T_r) \quad (6)$$

As a result, using Eq. 2, apply the RL approach to the entire building. The time-dependent thermal balance equation can be explained using the empirical formula [36]:

$$M_{wl} c p_{wl} \frac{T_{wl_{out,t}} - T_{wl_{in,t}}}{\Delta t} = \sum_j A_{w_j} U_j O F_t (T_{wl_{out,t}} - T_{wl_{in,t}}) + \sum_j A_{w_j} U_j O F_b + \sum_j A_{w_j} U_j O F_r D R - \sum_j A_{w_j} h_{i_j} (T_{wl_{in,t}} - T_{r,t}) \quad (7)$$

The empirical formula given in Eq. 7, the temperature varies for the opaque surfaces of the building, given by a linear combination of and . This method works well for thin, homogenous building materials, but for thick, heavy materials, the equation shows a variance in mistakes, and the only way to eliminate these errors is to offer a larger number of nodes in

the equation, such as $T_{w_{l_{in},t}}$, $T_{w_{l_{out},t}}$, $T_{r,t}$, ..., the temperature curve must be expressed as a linear set of functions to be approximated to $T_{w_{l_{in},t}}$, $T_{w_{l_{out},t}}$, $T_{r,t}$, ..., [33].

$$M_{w_{l_{cpwl}}} \frac{dT_{w_{l_{in},t}}}{dt} = \sum_j A_{w_j} U_j OF_t (T_{w_{l_{out},t}} - T_{w_{l_{in},t}}) + \sum_j A_{w_j} U_j OF_b + \sum_j A_{w_j} U_j OF_r DR - \sum_j A_{w_j} h_{i_j} (T_{w_{l_{in},t}} - T_{r,t}) \quad (8)$$

We take the Laplace transformation for both sides of Eq. 8 and assume initial conditions are zero, and to simplify the expression further it is possible to use the transfer function in the spatial domain (s-domain) [33].

$$G_{1,11} = k_1/(\tau_5 s + 1), T_{w_{l_{in}}}(s) = [G_{1,11} \quad G_{1,12} \quad G_{1,13}] \begin{bmatrix} T_o(s) \\ k_2 \\ T_r(s) \end{bmatrix} \quad (9)$$

Where, $G_{1,12} = 1/(\tau_5 s + 1)$, $G_{1,13} = k_3/(\tau_5 s + 1)$, $\tau_5 = M_{w_{l_{cpwl}}}/(\sum_j A_{w_j} U_j OF_t + \sum_j A_{w_j} h_{i_j})$,

$$k_1 = \sum_j A_{w_j} U_j OF_t / (\sum_j A_{w_j} U_j OF_t + \sum_j A_{w_j} h_{i_j}),$$

$$k_2 = (\sum_j A_{w_j} U_j OF_b + \sum_j A_{w_j} U_j OF_r DR) / (\sum_j A_{w_j} U_j OF_t + \sum_j A_{w_j} h_{i_j}),$$

$$k_3 = \sum_j A_{w_j} h_{i_j} / (\sum_j A_{w_j} U_j OF_t + \sum_j A_{w_j} h_{i_j})$$

In Eq. 9, Outdoor temperature (T_o), thermal resistance and solar radiation incident (k_2), and room temperature (T_r) are the inputs for the opaque internal temperature surface ($T_{w_{l_{in}}}$), ($G_{1,11}$) is the function of thermal resistance and external temperature ($^{\circ}C$), and ($G_{1,13}$) is the function of thermal resistance

4.1.2. Transparent Fenestration Surfaces

This area's heat transport differs from that of opaque surfaces. We can utilize it in Eq. 2. To calculate the heat gain, the following factors are used: area and surface cooling factor [35].

$$\dot{Q}_{fen} = \sum_j A_{fen_j} CF_{fen_j} \quad (10)$$

Where \dot{Q}_{fen} is given by $CF_{fen} = u_{NFRC}(\Delta t - 0.46DR) + PXI \times SHGC \times IAC \times FF_s$ is the fenestration cooling load (W), A_{fen} is fenestration area (m²), u_{NFRC} is the surface refrigeration factor (W/m²), u_{NFRC} is the fenestration NFRC heat U-factor (W/(m².K)), NFRC is the National Fenestration Rating Council, and Δt is the refrigeration design difference in temperature (K), DR stands for cooling daily range (K), PXI stands for peak exterior irradiance (W/m²), SHGC stands for fenestration rated, IAC stands for internal shading attenuation coefficient, and FF_s [34]. The following formula is used to calculate PXI:

$$PXI = T_x E_t \quad (\text{Unshaded fenestration}) \quad (11)$$

$$PXI = T_x [E_d + (1 - F_{shd})ED] \quad (\text{Shaded fenestration}) \quad (12)$$

Where PXI is the peak exterior irradiance (W/m²), T_x , E_t , E_d , and ED are the peak total, diffuse, and direct irradiance (W/m²), F_{shd} is the transmission of exterior attachment, and ϕ is the fraction [33].

The values, T_x , E_t , and E_d are based on two surface conditions: horizontal surfaces and vertical surfaces.

For horizontal surfaces [33]:

$$E_t = 952 + 6.49L - 0.166L^2, E_d = \min(E_t, 170), \text{ and } ED = E_t - E_d \quad (13)$$

$$\phi = \left| \frac{\psi}{180} \right| \quad (\text{normalized exposure, } 0 - 1)$$

For vertical surfaces [33]:

$$E_t = 453.4 + 1341\phi - 5279\phi^3 + 3260\phi^4 - 34.09\phi L + 0.2643\phi L^2 - 12.83L - 0.8425L^2 + \left[\frac{0.9835L^2}{\phi+1} \right] \quad (14)$$

$$E_d = \min \left(E_t, 357 - 86.98\phi^2 + 1.764\phi L - \frac{108.4\sqrt{L}}{\phi+1} \right) \quad (15)$$

Where L is the site latitude, and °N is the exposure.

For any fenestration shadowed by nearby for peak hours, the fraction can be taken as 1. An estimated equation provides simple overhang shading [33]:

$$F_{shd} = \min \left[1, \max \left(0, \frac{SLF \times D_{oh} - X_{oh}}{h} \right) \right] \quad (16)$$

Where is the shade line factor, is the depth of overhang (from the plane of fenestration) (m), is the vertical distance from the top of the fenestration to the overhang (m), and is the height of the fenestration (m).

The following is how values are calculated [33]:

$$IAC = 1 + F_{cl}(IAC_{cl} - 1) \quad (17)$$

Where the interior attenuation coefficient, is the shade fraction, and is the interior attenuation coeff. [33].

The heat gain through a fenestration is given as [34].

$$\dot{Q}_{fen} = \sum_j A_{fen_j} u_{NFRC_j} (T_o - T_{gin}) - \sum_j A_{fen_j} u_{NFRC_j} \times 0.46DR + \sum_j A_{fen_j} PXI_j \times SHGC_j \times IAC_j \times FF_{s_j} \quad (18)$$

The transfer function can be determined using the same approach as the opaque surfaces after obtaining the heat transmitted into the control volume. The outcome is the glass's internal temperature, which is defined as [33]:

$$T_{gin}(s) = \begin{bmatrix} G_{1.14} & G_{1.15} & G_{1.16} \end{bmatrix} \begin{bmatrix} T_o(s) \\ T_r(s) \\ f_{DR} \end{bmatrix} \quad (19)$$

Outdoor temperature, indoor temperature, and conditioned place location are used as inputs, and the result is inside glass temperature.

$$G_{1.14} = R_g f_1 / (f_1 R_g + 1)(\tau_g S + 1).$$

$$G_{1.15} = 1 / (f_1 R_g + 1)(\tau_g S + 1),$$

$$G_{1.16} = -R_g / (f_1 R_g + 1)(\tau_g S + 1), \quad \tau_g = Ca_g R_g / (f_1 R_g + 1)$$

$$f_{DR} = \sum_j A_{fen_j} u_{NFRC_j} \times 0.46DR$$

$$R_g = (1 / \sum_j A_{fen_j} h_{ij}), \quad f_1 = \sum_j A_{fen_j} u_{NFRC_j} (W/k) \quad [33].$$

4.1.3. Slab floors

The slab floor is a significant thermal capacitance storage capacity. It is the greatest and also most complex to calculate the building's several components. Eq. 2, energy conservation law can be rewritten like this [35]:

$$M_{slab} c_{p_{slab}} \frac{dT_{slab,t}}{dt} = \sum_i \dot{Q}_{slab_{in}} - \sum_i \dot{Q}_{slab_{out}}$$

(20)

Where $\dot{Q}_{slab_{in}}$ and $\dot{Q}_{slab_{out}}$ are the heat gain and loss of the slab, respectively (W) and C_{slab} is the heat capacitance of slab (J/K).

According to [37] heat gain is largely through the perimeter, rather than the floor and earth [38].

Overall heat loss/gain is more almost related to the length of the perimeter than to the area of the floor, and it may be computed by the following equation for both unheated and heated slab floors [33].

$$\dot{Q}_{slab_{in}} = f_t P (T_{slab_{in}} - T_o) \quad (21)$$

Where $\dot{Q}_{slab_{in}}$ is the heat gain through slab floors (W), $\dot{Q}_{slab_{out}}$ is the heat loss of the perimeter, $W/(m.K)$, P is the perimeter or exposed edge of the floor (m), $T_{slab_{in}}$ is the inside slab floor temperature, and T_o is the outdoor temp. [33].

The ASHRAE organization calculated the heat loss from a concrete slab floor using the same procedures as for opaque and fenestration surfaces [39]. As a result, the heat production of the control volume is as shown in Eq. 22.

$$\dot{Q}_{slab_{out}} = \sum_j A_{slab_j} C_{f_{slab_j}} \quad (22)$$

Where A_{slab_j} is the area of the slab (m^2), $C_{f_{slab_j}} = (1.9 - 1.4h_{srf})$ is cooling factor (W/m^2), given by

$$h_{srf} = 1/(R_{cvr} + 0.12)$$

Where R_{cvr} is the effective surface conductivity, which includes slab covering substance resistance, like carpet (Representative values can be found in [40]).

Eqs. 21 and 22 are substituted into Eq. 20 to produce the slab floor transfer function, and then after simplifying the expression, Laplace transformation is used on both sides of the resultant equation [34].

As illustrated below, the slab floors insides are slabbed floors area (A_{slab}) and outside temperature (T_o), while the slab floors outcomes are internal slab floors temperature ($T_{slab_{in}}$) [33].

$$T_{slab_{in}}(S) = [G_{1.17} \quad G_{1.18}] \begin{bmatrix} A_{slab} \\ T_o \end{bmatrix} \quad (23)$$

Where $G_{1.17} = (1.9 - 1.4h_{srf})/(\tau_{slab}S + 1)$, $G_{1.18} = f_t P/(\tau_{slab}S + 1)$, $\tau_{slab} = C_{slab}/f_t P$,

$C_{slab} = \sum_i M_{slab_i} c_{p_{slab_i}}$, is the heat capacitance of slab floors (J/k) [33].

4.2. Building System Control of the Reinforcement Learning (RL) Process

Reinforcement Learning (RL) is a framework in which an agent interacts with its environment and learns the best set of behaviours to achieve a goal, which is represented by a policy. RL focuses on the learning agent that adjusts its behaviours in response to the environment's reward to achieve a specified objective, such as getting the maximum amount of reward [41]. Instead of a priori knowledge, the agent's actions are based on his or her own experience gained during the game. Research has been done in the field of building control to apply RL to model-based control [15].

In this research, the environment is the model for the airport building described previously. The agent's learning goal is expressed through a reward, which is a scalar feedback value that

quantifies how the agent is acting about the learning objective. A MDP is characterized by a tuple (S, A, R, P, γ) , is used to formalize the agent-environment interaction. [41] stated S is the set of states, A is the set of actions, $r: S \times A \rightarrow \mathbb{R}$ is the reward function, P is a state transition probability matrix; the transition function specifies the probability of transition from the state (s) to state on acting (a) and γ is a discount factor, $\gamma \in [0, 1]$ [41]. The state-value function is the predicted discounted return when beginning from a state " s " [41].

$$(24)$$

Where " R_t " represents the total rewards received during the period " t ", " r_{t+k} " represents the signal of an instant return, and (γ) represents the discount rate, calculated inside the iniquality to check that (R_t) is limited [12].

$$(25)$$

$$(26)$$

The (RL) agent uses an optimal strategy to discover the ideal state-value function. It may be written as the sum of all state-values in Eqs. 27, 28, and 29. A set of equations known as the "Bellman Optimality Equations" can resolve into instant return and discounted future values [12].

$$(27)$$

$$(28)$$

$$(29)$$

The agent optimizes its policy $\pi: S \rightarrow P(A)$, which describes its behaviour in the environment, where $P(A)$ is a probability distribution among the global set of actions, over a discrete sequence of time steps. The goal of the policy improvement technique of optimal policy (π^*) for better action that maximizes the projected discounted return [42].

In our problem, (energy, temperature) are state components, and the agent chooses an action from a finite set of feasible actions that represent the chillers' capacity at each time step (0, 25, 40, and 65). As a result, the state of the environment is changed to s_{t+1} , and the agent receives a reward. These incentives have a real-world value and can be good or negative, but they must be constrained. The agent can detect all conceivable states of the environment. According to the Markov transition model, the likelihood of transitioning from state (s) to state (s') after acting an A is solely determined by the current state, not by the history of previous states. Based on specific needs represented by the reinforcement function, a policy for an agent of the building management system to work on the HVAC system is obtained in these trials. Each episode is divided into numerous time steps t , with the building management system deciding on an action to adjust the temperature inside the building based on its policy at each step [43]. The goal of applying the RL algorithm to the HVAC management system is to conserve energy while maintaining comfort, the reward function's learning aim must account for a trade-off between the two components of comfort and main energy consumed.

4.3. Chiller Operation

The HVAC system considers for more than 50% of a building's total energy usage, chilled water chillers represent 30-40% of the energy used for this system. As a result, chillers must be run properly to increase energy efficiency. The chiller operates based on the refrigeration load for the building (i.e., appropriately switching chillers ON/OFF) [44]. Furthermore, most chillers' operation strategies necessitate reliable real-time monitored cooling water system data. Chiller operation techniques are occasionally designed to improve chiller COP to reduce energy consumption and improve chiller efficiency. It is possible to clarify the following chiller operation strategies:

1. Chilled water is turned ON/OFF based on measured cooling load data.
2. The cooling limit of the chiller exceeds the refrigeration load by a large quantity, and based on the building cooling load, the chillers are turned ON/OFF.

The flowchart in Fig. 4, is based on the case presented that consists of a large chiller with a capacity of (4.0 Mw) and a small chiller with a capacity of (2.5 Mw), with the chiller capacity dictated by the system load. In other cases, the chiller's ON/OFF alternately is to increase the partial load and improve energy efficiency. The structure with two non-identical chillers is depicted in the diagram. When one or more chillers are in use, another chiller should not be turned ON/OFF unless load exceeds the system's cooling load.

- a) If the cooling load of the building is lower than the limit of the chillers, all chillers are switched off.
- b) The small chiller is operated with a capacity (2.5MW) if the load exceeds the capacity of this chiller.
- c) If the load exceeds the limit of the small chiller, it will be off and the large radiator with a capacity (4.0MW) is turned on.
- d) If the load exceeds the limit of the large chiller, all chillers will be operated when the building load is at its peak.

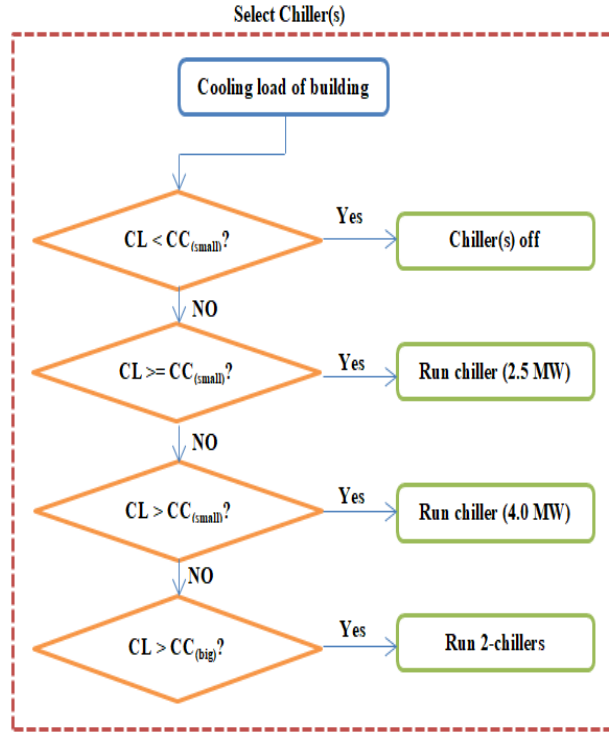


Fig. 4. Workflow of the proposed method.

4.4. Chiller sequencing control

The goal of chiller sequencing control is to save enough cooling to maintain building indoor thermal comfort while maintaining chiller plant high energy efficiency.

Fig. 5, depicts the chiller sequencing control principle, with the x-axis representing the building's cooling load (CL), given by:

$$(30)$$

Where CL is the calculated cooling load (kW), c_p is the specific thermal capacity of water (J/kg.°C), \dot{m} is the measured chilled water flow rate (kg/s), T_r and T_s are the chilled water return and supply temperatures (°C) respectively.

Each chiller's ON/OFF status should be regulated at the start of each time step. To identify suitable ON/OFF signals, the following rules are used:

- If the system cooling load is less than 1 chiller's rated cooling capability, the entire central chilled water system is shut off.
- The cooling load on the system determines the number of chillers necessary; when the cooling load on the system exceeds the current cooling capacity, an extra chiller is activated.

- c) The ON/OFF case of the main chilled water pump is synchronized with the condition of the chillers. When the chiller is turned ON/OFF, the principal chilled water pump is also turned ON/OFF.

The first phase entails selecting the chiller(s) based on the chillers' cooling capacity (CC) and cooling load. In this method, the chiller partial load rate (PLR) is maximized; the COP of centrifugal chillers is also tuned because the PLR of centrifugal chillers increases with the COP.

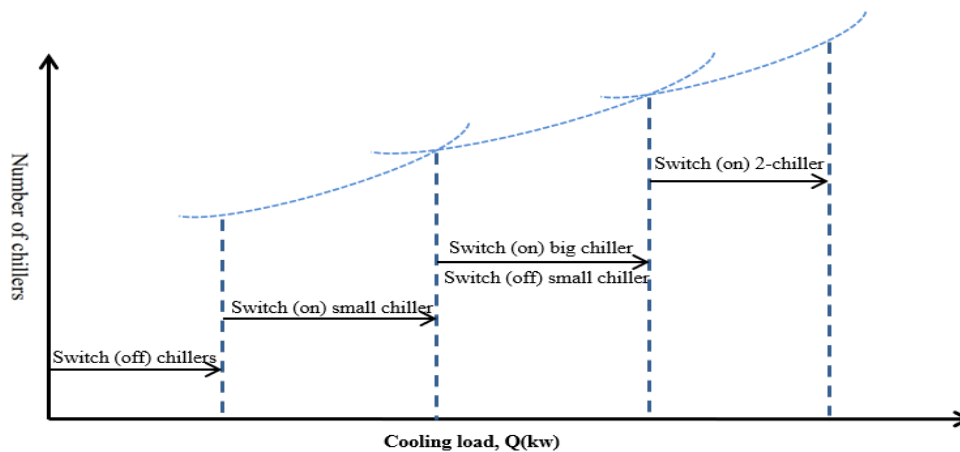


Fig. 5. The basic principle of chiller sequencing control

The buildings inside temperature setpoints from 20 °C to 24 °C in the second step [44]. After we built the control model-based for the airport building, which is the product of the equations of the mathematical model, we found the thermal load signal for a whole day (24 hours) for the multiple chillers. We take the signal for this load and build an RL-control on, it as it learns the ON/OFF process for each chiller, the energy and temperature indication will be (state), and the action represents the capacity of the two chillers (2.5 MW, 4.0 MW) respectively. If the building load exceeds the capacity of the first chiller (2.5 MW), then RL-control will operate the second chiller (4.0 MW) instead of the first chiller, and when the building's load exceeds the capacity of the second chiller at the peak, RL-control will operate the two chillers together and thus the capacity of the two chillers is (6.5 MW). The reason for our use of more than one chiller is to conserve the largest amount of energy possible and reduce the energy consumed, while when operating a single chiller and with a high capacity while the building load is low, it will be exposed to faults or damage.

5 Results and Discussion

The HVAC system is simulated under the supervision of a mathematical model, and the performance of the RL-controller and the PID controller is compared by turning the mismatched chillers ON/OFF every 30 minutes for 24 hours.

In Fig. 3, a full-day Airport system operation is simulated every half-hour for various chillers under the management of model-based control. The model results were discussed in terms of

energy consumption, the controller's learning process, control actions, and system water temperature, as well as the load equations for opaque, transparent surfaces, and slab floors, which were used to compute the building cooling load for 24 hours using model-based control. These calculated cooling load results were to know the temperatures, power, and energy inside the building.

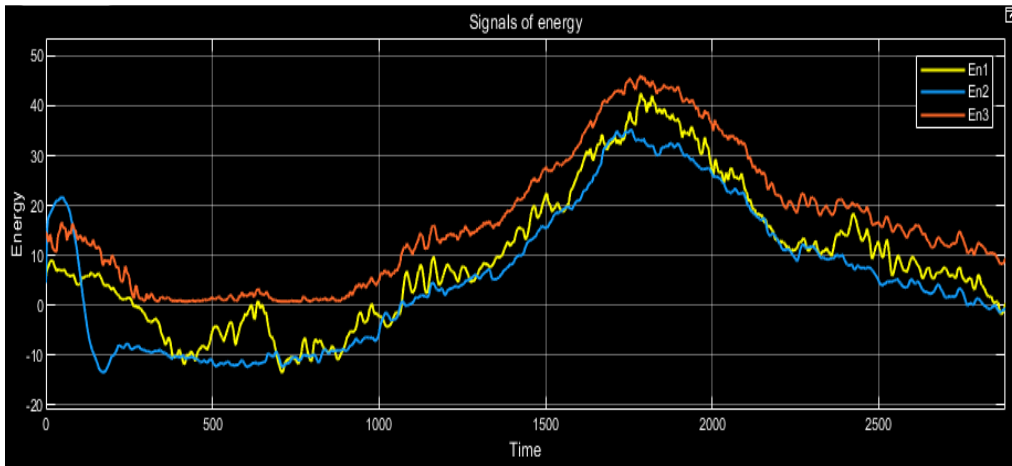


Fig. 6. Signals of Energy.

Figs. 6, 7, and 8 represent the signals (energy, power, indoor and outdoor temperatures) coming out of the Simulink, changing with time, for the model of chillers with capacities (2.5 MW, 4.0 MW, and 6.5MW) respectively.

Where Fig. 6, represents the energy signals and it is dependent on the cooling coil, which is the thermal loads of the airport building, where we see these loads and how they work with temperature and time and will be within 24 hours. As Fig. 7, it represents the power signals of the HVAC system, which is the electrical power that the system dissipates over time. Fig. 8, represents the temperatures of the building, where we note that the indoor temp. is less than the outdoor temperature, and in this case, the HVAC system will not work and the difference in temperatures depends on the insulation of the building, walls, and others.

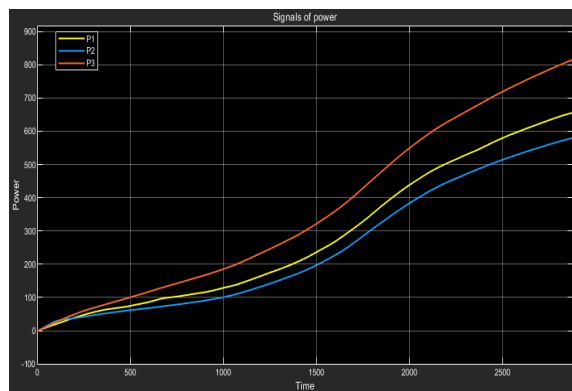


Fig. 7. Signals of Power.

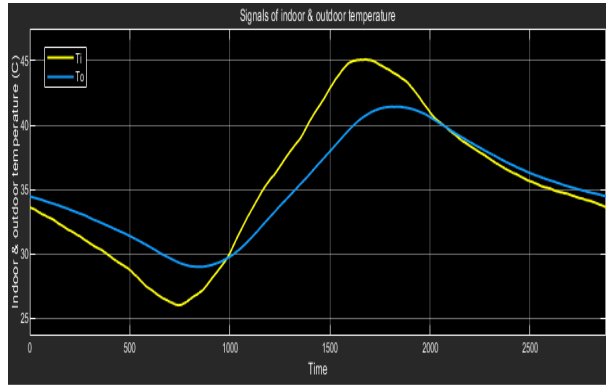


Fig. 8. Signals of Indoor & Outdoor Temperature.

Fig. 9, illustrates the cases of reinforcement learning for indoor and outdoor temperatures (°C) with time (hours). The black line represents the temperature outside the building and the red line is the temperature inside the building, which is within the limits of thermal comfort, which is between 20 °C to 24 °C. The chiller plant must remove the cooling load accumulated throughout the night as well as the potential heat of the air conditioning system at the start of the day, resulting in a tremendous load met by chillers. The chillers should be turned on/off at the maximum energy-efficient point, which is usually less than 100 percent, according to the chiller sequencing control. When the cooling load is plainly changing, this is fair for chiller energy efficiency. In comparison with Fig. 10, where the reinforcement learning of the building's internal temperature using multiple HVAC systems, the traditional building control proportional integral derivative (PID) was used for the HVAC system and it was within the comfort limits of the building, but it was not stable compared with the reinforcement learning when the ON/OFF control.

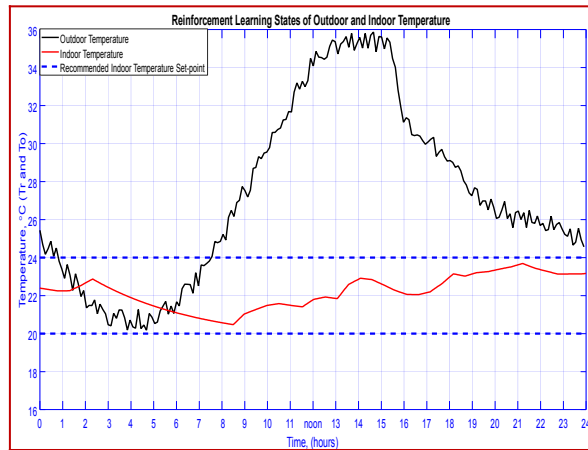


Fig. 9. RL States of Outdoor & Indoor Temperature.

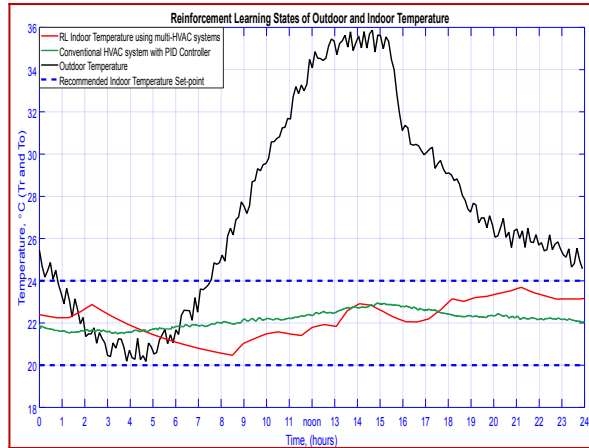


Fig. 10. RL States of Outdoor & Indoor Temperature with PID Controller

As shown in Fig. 11, the performance comparison between the three control strategies for the response step, is the reinforcement learning for the energy-based multi-HVAC system, the traditional HVAC system with the controller (PID) in addition to the real or calculated response to the building strength. After comparing the performance, we found that saving energy through the proposed controller is (45%) from the traditional strategy.

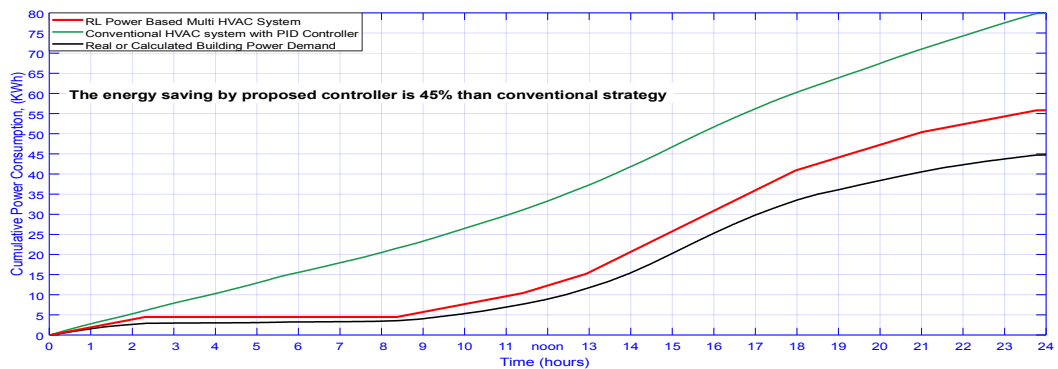


Fig. 11. Performance Comparison between Three Controllers' Strategies for the Step Response

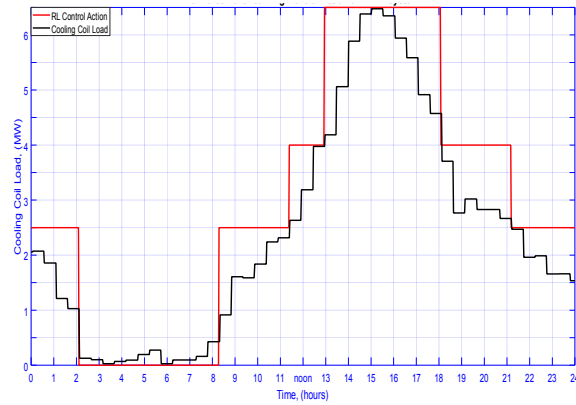


Fig. 12. Reinforcement Learning Control Action of the HVAC system.

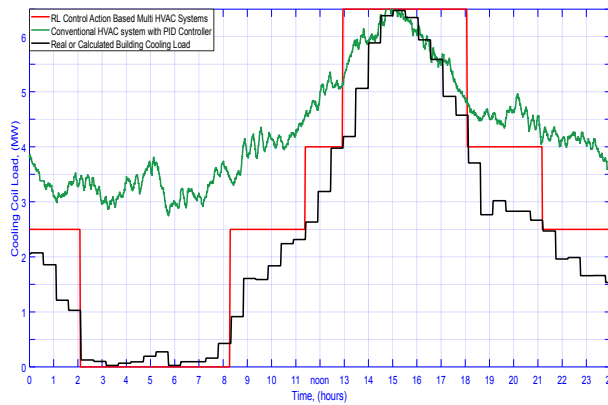


Fig. 13. Reinforcement Learning Control Action of Multi HVAC System.

In Fig. 12, illustrating the HVAC system reinforcement learning (RL) control action for the cooling load profile (MW) and time (hours), use the reinforcement learning to perform the control with the building cooling load. This action features temperature control by adjusting the flow rate of the cooling coil water related to indoor/ outdoor temperature indication. The law of control is also based on the "ON/OFF" algorithm. While in Fig. 13, represents a control based RL for multiple HVAC systems with a conventional "PID" control strategy for a real or computed building cooling load. HVAC systems varies the cooling profile (energy consumption) and the reinforcement learning of the control procedure is calculated utilizing iterative techniques in the "MATLAB code". During the day, energy consumption increases, this indicates an increase in the number of chillers by 100%, especially at peak to improving energy efficiency.

Fig. 14, shows the value function for energy saving in the case of multi-chillers system operation. The RL control algorithm was trained using iterators in the MATLAB code for states and actions to the maximum value using the reward function to build the value function by applying the Bellman Eq. 30. After optimizing the value function, we chose the best value to guarantee the control policy. In addition, when agent rewarded after taking action for each time step of control depends on the energy and temperature, if a low outdoor temp. is maintained and to avoid punishing the agent for taking heavy action, there is no need to turn

on the chiller because it is considered energy consumption. For Fig. 15, consider the existing analysis which represents the best function of the control policy for actions (0, 2.5, 4.0, and 6.5) to find the best suitable chiller for the building conditioning process and the best possible efficiency. The best action applied to operate the chiller suitable for the building cooling load is the best policy than calculating the maximum value function from the Bellman equation. Therefore, the optimum energy efficiency for cooling the building is obtained when both small and large chillers are operated.

In conclusion, the reinforcement learning control technique suggested in this work has the potential to conserve more energy. Furthermore, the proposed strategy's energy-saving effectiveness can be increased while maintaining internal thermal comfort. As a result, the proposed technique can slightly shorten the operating duration of the tiny chiller by removing certain superfluous starting activities.

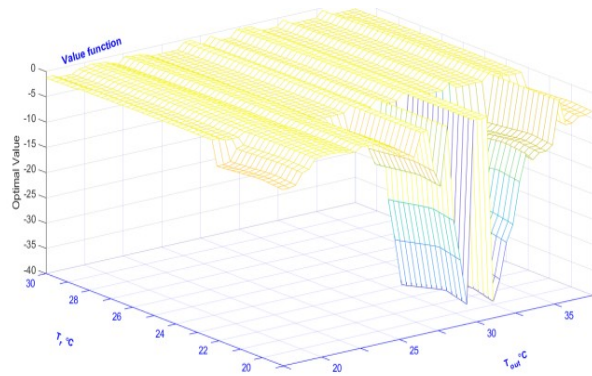


Fig. 14. Optimal value function.

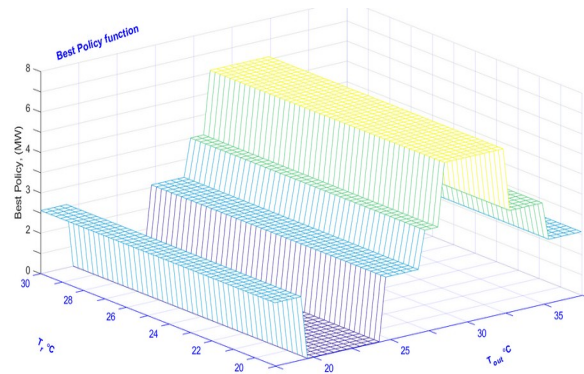


Fig. 15. Best policy function.

6 Conclusion

This study demonstrated a modelling technique for HVAC systems using the MATLAB/Simulink platform, as well as the use of the RL-controller strategy to rise the energy efficiency of cooling load. To control the multi-chillers system, an ideal model-based control strategy has been developed. In this method, energy and temperature are the states, chillers capacities are the actions, and the reward is energy savings, chilled water pumps, and chillers. A model-based simulation is performed under the supervision of a white box (mathematical model). The simulation results for 24 hours indicate that the proposed energy and power controller (RL-Controller) can save 45% of the energy compared to the traditional PID controller. This limited application in an airport terminal model cooling plant using non-identical chillers; it takes a whole day to simulate the operation of the whole system which is controlled every half hour by the proposed reinforcement learning control strategy. Where the reinforcement learning control (RL-control) system for multiple chillers can improve system stability and energy efficiency. Hence, it is concluded the method can lower the energy consumption of the building, which is acceptable for engineering applications.

Acknowledgment

I would like to express my thanks and appreciation to the Basra International Airport, which supported the research and assisted me in obtaining the required information.

References

- [1] Homod, R.Z., Gaeid, K.S., Dawood, S.M., Hatami, A. and Sahari, K.S., 2020. Evaluation of energy-saving potential for optimal time response of HVAC control system in smart buildings. *Applied Energy*, 271, p.115255.
- [2] Homod, R. Z., Almusaed, A., Almssad, A., Jaafar, M. K., Goodarzi, M., & Sahari, K. S. M. (2020). Effect of different building envelope materials on thermal comfort and air-conditioning energy savings: A case study in Basra city, Iraq. *Journal of Energy Storage*, 101975.
- [3] Homod, R.Z., Togun, H., Abd, H.J. and Sahari, K.S., 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, p.102091.
- [4] Dong He, Qingyu Xiong, Xuyang Zhang, Yunchuang Dai and Ziyang Jiang, A Decentralized, Flat-Structured Control System for Chiller Plants (2019).
- [5] Bingqing Chen, Zicheng Cai and Mario Bergés, Gnu-RL: A Precocial Reinforcement Learning Solution for Building HVAC Control Using a Differentiable MPC Policy, Nov. (2019).
- [6] Homod, R.Z., 2018. Analysis and optimization of HVAC control systems based on energy and performance considerations for smart buildings. *Renewable Energy*, 126, pp.49-64.
- [7] Poe, W. A., & Mokhtab, S. (2017). *Process Control. Modeling, Control, and Optimization of Natural Gas Processing Plants*, 97–172.
- [8] Liu, S., & Henze, G. P. (2006). Experimental analysis of simulated reinforcement learning control for active and passive building thermal storage inventory. *Energy and Buildings*, 38(2), 148-161.
- [9] Dalamagkidis, K., Kolokotsa, D., Kalaitzakis, K., & Stavrakakis, G. S. (2007). Reinforcement learning for energy conservation and comfort in buildings. *Building and Environment*, 42(7), 2686–2698.
- [10] Fan, B., Jin, X., & Du, Z. (2011). Optimal control strategies for multi-chiller system based on probability density distribution of cooling load ratio. *Energy and Buildings*, 43(10), 2813–2821.
- [11] Sen Huang, Wangda Zuo, Michael D. Sohn (2015). A NEW METHOD FOR THE OPTIMAL CHILLER SEQUENCING CONTROL, 14th Conference of International Building Performance Simulation Association, Hyderabad, India, Dec. 7-9, 2015, 316-323.
- [12] Azuatalam, D., Lee, W.-L., de Nijs, F., & Liebman, A. (2020). Reinforcement learning for whole-building HVAC control and demand response. *Energy and AI*, 2, 100020.

- [13] Yao, Y., Lian, Z., Hou, Z., & Zhou, X. (2004). Optimal operation of a large cooling system based on an empirical model. *Applied Thermal Engineering*, 24(16), 2303–2321.
- [14] Qiu, S., Li, Z., Li, Z., & Zhang, X. (2020). Model-free optimal chiller loading method based on Q-learning. *Science and Technology for the Built Environment*, 1–22.
- [15] Qiu, S., Li, Z., Li, Z., Li, J., Long, S., & Li, X. (2020). Model-free control method based on reinforcement learning for building cooling water systems: Validation by measured data-based simulation. *Energy and Buildings*, 218, 110055.
- [16] Homod, R.Z., Almusaed, A., Almssad, A., Jaafar, M.K., Goodarzi, M. and Sahari, K.S., 2021. Effect of different building envelope materials on thermal comfort and air-conditioning energy savings: A case study in Basra city, Iraq. *Journal of Energy Storage*, 34, p.101975.
- [17] Afroz, Z., Shafiullah, G., Urmee, T., & Higgins, G. (2018). Modeling techniques used in building HVAC control systems: A review. *Renewable and Sustainable Energy Reviews*, 83, 64–84.
- [18] Afram, A., & Janabi-Sharifi, F. (2014). Review of modeling methods for HVAC systems. *Applied Thermal Engineering*, 67(1-2), 507–519.
- [19] Homod, R. Z. (2013). Review on the HVAC System Modeling Types and the Shortcomings of Their Application. *Journal of Energy*, 2013, 1–10.
- [20] Homod, R.Z., Sahari, K.S.M., Almurib, H.A. and Nagi, F.H., 2012. RLF and TS fuzzy model identification of indoor thermal comfort based on PMV/PPD. *Building and Environment*, 49, pp.141-153.
- [21] Poe, W. A., & Mokhatab, S. (2017). *Process Control. Modeling, Control, and Optimization of Natural Gas Processing Plants*, 97–172.
- [22] Krarti, M., 2011. *Energy Audit for Building Systems: An Engineering Approach*, 2nd ed. CRC Press, Francis, and Taylor, Boca Raton, FL, 600 pp.
- [23] Rao, R. V. (2015). Optimization of Multiple Chiller Systems Using TLBO Algorithm. *Teaching Learning Based Optimization Algorithm*, 115–128.
- [24] A. Bhatia (2012). HVAC Equipments and Systems, PDHonline Course M322 (7 PDH).
- [25] Carrier Corporation (2005). *Commercial HVAC Chiller Equipment*.
- [26] Beghi, A., Cecchinato, L., Cosi, G., & Rampazzo, M. (2010). Two-layer control of multi-chiller systems. 2010 IEEE International Conference on Control Applications.
- [27] Ardakani, A. J., Ardakani, F. F., & Hosseinian, S. H. (2008). A novel approach for optimal chiller loading using particle swarm optimization. *Energy and Buildings*, 40(12), 2177–2187.
- [28] Askarzadeh, A., & Coelho, L. dos S. (2015). Using two improved particle swarm optimization variants for optimization of daily electrical power consumption in multi-chiller systems. *Applied Thermal Engineering*, 89, 640–646.
- [29] Diniz, A. A. R., Pires, P. R. M., de Melo, J. D., Neto, A. D. D., Filho, A. J. J. L., & Kanazava, S. M. (2010). Reinforcement Learning for Controlling a Coupled Tank System Based on the Scheduling of Different Controllers. 2010 Eleventh Brazilian Symposium on Neural Networks.
- [30] Miyoung Han (2018). *Reinforcement Learning Approaches in Dynamic Environments*.
- [31] Ahmad Hammoudeh (2018). *A Concise Introduction to Reinforcement Learning*.
- [32] B. Abdulhai, R. Pringle, and G.J. Karakoulas (2001). *Reinforcement Learning for ITS: Introduction and a Case Study on Adaptive Traffic Signal Control*.
- [33] Homod, R.Z., Sahari, K.S.M., Almurib, H.A. and Nagi, F.H., 2011. Double cooling coil model for non-linear HVAC system using RLF method. *Energy and buildings*, 43(9), pp.2043-2054.
- [34] Homod, R.Z. and Sahari, K.S.M., 2013. Energy savings by smart utilization of mechanical and natural ventilation for hybrid residential building model in passive climate. *Energy and Buildings*, 60, pp.310-329.
- [35] Homod, R.Z., Sahari, K.S.M. and Almurib, H.A., 2014. Energy saving by integrated control of natural ventilation and HVAC systems using model guide for comparison. *Renewable Energy*, 71, pp.639-650.
- [36] Homod, R.Z., Sahari, K.S.M., Almurib, H.A. and Nagi, F.H., 2012. Gradient auto-tuned Takagi–Sugeno fuzzy forward control of a HVAC system using predicted mean vote index. *Energy and Buildings*, 49, pp.254-267.

- [37] F.S. Wang, Mathematical modeling and computer simulation of insulation systems in below grade applications, in: ASHRAE/DOE Conference on Thermal Performance of the Exterior Envelopes of Buildings, Orlando, FL, 1979.
- [38] T.P. Bligh, P. Shipp, G. Meixel, Energy comparisons and where to insulate earth sheltered buildings and basements, Earth covered settlements, U.S. Department of Energy Conference, Fort Worth, TX, 1978, K. Elissa, Title of paper if known, unpublished.
- [39] ASHRAE, Residential cooling and heating load calculations, handbook— fundamentals, chp. 17, in: J. Clerk Maxwell (Ed.), American Society of Heating, Refrigerating, and Air-Conditioning Engineers, A Treatise on Electricity and Magnetism, vol. 2, 3rd ed., Clarendon, Oxford, 2009[1892], pp. 68–73.
- [40] ASHRAE, HVAC Systems and Equipment' Handbook—Fundamentals, American Society of Heating, Refrigerating, and Air-Conditioning Engineers, 2008 (Chp. 6).
- [41] Sutton, R. S. and A. G. Barto (1998). Introduction to Reinforcement Learning, Volume 135. MIT press Cambridge.
- [42] Mastropietro, A., Castiglione, F., Ballesio, S., Enrico Fabrizio, E. (2019). Reinforcement Learning Control Algorithm for HVAC Retrofitting: Application to a Supermarket Building Model by Dynamic Simulation. Conference: Building Simulation 2019.
- [43] Fazenda, P., V., Veeramachaneni, K., Lima, P., O'Reilly, U. (2014). Using Reinforcement Learning to Optimize Occupant Comfort and Energy Usage in HVAC Systems. Journal of Ambient Intelligence and Smart Environments · January 2014.
- [44] Qiu, S., Feng, F., Zhang, W., Li, Z., & Li, Z. (2019). Stochastic optimized chiller operation strategy based on multi-objective optimization considering measurement uncertainty. Energy and Buildings, 195, 149–160.