

Research Papers

Massive energy reduction and storage capacity relative to PCM physical size by integrating deep RL clustering and multi-stage strategies into smart buildings to grid reliability



Raad Z. Homod ^{a,*}, Hayder I. Mohammed ^b, Abdellatif M. Sadeq ^c, Bilal Naji Alhasnawi ^d, Ali Wadi Al-Fatlawi ^e, Ahmed Al-Manea ^d, Omer A. Alawi ^f, Ali Alahmer ^g, Jasim M. Mahdi ^h, Wael Al-Kouz ⁱ, Zaher Mundher Yaseen ^j

^a Department of Oil and Gas Engineering, Basrah University for Oil and Gas, Basra 1004, Iraq

^b Department of Cooling and Air Conditioning Engineering, Imam Ja'afar Al-Sadiq University, Baghdad, Iraq

^c Mechanical and Industrial Engineering Department, College of Engineering, Qatar University, Doha, Qatar

^d Department of Fuel and Energy Techniques Engineering, Petroleum and Energy Engineering Technical College, Al-Furat Al-Awsat Technical University, Kufa, 66001, Iraq

^e Mechanical Engineering Department, University of kufa, Iraq

^f Department of Thermofluids, Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor Bahru, Malaysia

^g Department of Mechanical Engineering, Tuskegee University, Tuskegee, AL 36088, United States of America

^h Department of Energy Engineering, University of Baghdad, Baghdad 10071, Iraq

ⁱ Department of Engineering and Industrial Professions, University of North Alabama, Florence, AL 35632, USA

^j Civil and Environmental Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia

ARTICLE INFO

Keywords:

Optimal tank sequencing control (OTSC)
Phase change material (PCM)
Deep reinforcement clustering for adaptive decision policy (DRCADP)
Double stage of PCM
Multi-stage thermal energy storage (MSTES)
Multi-objective reinforcement learning (MORL)

ABSTRACT

Integrating artificial intelligence (AI) into energy management using phase change materials (PCMs) is a revolutionary approach to improving building energy efficiency. This strategy aims to maximize the coefficient of performance (COP) of chillers to tackle the pressing issues of energy peak demand and increasing costs. In order to address the intrinsic difficulty posed by features opposing each other, such as melting point and latent heat (LH), a multi-stage thermal energy storage (MSTES) system utilizes two different binary composite materials of PCM kinds. This not only increases the capacity for storing energy but also ensures a more evenly distributed demand for energy, hence reducing the cost per kilowatt-hour and relieving strain on power systems. The thermal conductivity is significantly enhanced by segregating the PCM mixture types into four separate tanks. The proposed algorithm of deep reinforcement clustering for adaptive decision policy (DRCADP) utilizes deep reinforcement clustering to optimize the OTSC strategy, which aims to enhance the efficiency of chiller plant operations by optimizing the charge and discharge processes during periods of low cooling demand. To tackle the intricacy of nonlinear multiple variables, agent action rules are divided into clusters, guaranteeing efficient system operation. By implementing a dual stage of eutectic PCM consisting of tetradecane and hexadecane, the size of the thermal energy storage system is significantly decreased. This results in a compact design and a remarkable 32.5 % decrease in energy consumption compared to traditional methods. Additionally, there are cost savings in the construction of tank structures. This comprehensive and cutting-edge approach showcases a possible method for enhancing sustainable and effective energy management in buildings.

* Corresponding author.

E-mail addresses: raadahmood@yahoo.com (R.Z. Homod), hayder.i.mohammad@garmian.edu.krd (H.I. Mohammed), as1004958@qu.edu.qa (A.M. Sadeq), aliwalfatlawi@uokufa.edu.iq (A.W. Al-Fatlawi), Dr.ahmed.almanea@atu.edu.iq (A. Al-Manea), omeralawi@utm.my (O.A. Alawi), aalahmer@tuskegee.edu (A. Alahmer), jasim@siu.edu (J.M. Mahdi), Walkouz@una.edu (W. Al-Kouz).

Abbreviations

AHU	Air handling unit
DRCADP	Deep reinforcement clustering for adaptive decision policy
CMAS	Cooperative multi-agent system
MORL	Multi-objective reinforcement learning
COP	Coefficient of performance
CTES	Cold thermal energy storage
DNN	Deep neural network
EPCMs	Eutectic phase change materials
HVAC	Heating, ventilation, and air conditioning
LH	Latent heat
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MLP	Multilayer perceptron
MxAE	Maximum absolute error
$\dot{m}_w(s)$	The flow rate of chilled water for pre-cooling coil mass (kg/s)
$\dot{m}_{mw}(s)$	The main cooling coil mass flow rate of chilled water (kg/s)
$\dot{m}_{os}(s)$	The fresh air mass flow rate passing through the pre-cooling coil (kg/s)
A_{slab}	The slab floor area of the building conditioned space (m^2)
k_2	The building envelope thermal resistance (W/(m/K))
$\omega_o(s)$	The outdoor air humidity ratio (gm vap/kg dry air)
OPCMs	Organic phase change materials
OTSC	Optimal tank sequencing control
TES	Thermal energy storage
PCM	Phase change material
PEG	Polyethene glycol
PID	Proportional-integral-derivative
PLR	Part load ratio
RAE	Relative fundamental error
RH	Relative humidity
RL	Reinforcement learning
RLF	residential load factor
RMSE	Root mean squared error
SH	Sensible heat
SSE	The sum of squares due to error
MSTES	Multi-stage thermal energy storage
f_4	The internal sensible heat gain (W)
$\dot{Q}_{ig,l}$	The internal latent heat gain (W)
$\dot{m}_r(s)$	The return air mass flow rate (kg/s)
f_{DR}	The location factor of the conditioned space
$T_o(s)$	The outdoor air temperature ($^{\circ}\text{C}$)
$T_r(s)$	The conditioned space temperature ($^{\circ}\text{C}$)
VAV	Variable air volume

1. Introduction

Enhancing energy performance is one of the most important approaches to addressing the issue of rising energy demand and lowering the use of oil and gas and emissions of carbon dioxide [1,2]. The phase change materials (PCMs) have become more popular in smart building applications over the past three decades due to power plant loads realizing the need to spread peak load and flatten the power load curves of their generation demand [3].

Since commercial and residential buildings account for about 40 % of global power use, improving building energy efficiency is crucial for reducing greenhouse gas emissions [4,5]. The chillers in the heating, ventilation, and air conditioning (HVAC) system are the most energy-intensive equipment in buildings [6,7]. Fig. 1 [8] demonstrates that the capacity of chiller systems, together with external factors such as ambient temperature and chilled water supply temperature, significantly influences the coefficient of performance (COP) of chiller plants. Due to the fluctuation of the cooling demand throughout the day, it is impractical to sustain the coefficient of performance (COP) at its optimal level during standard chiller plant operations [9]. Due to lower nighttime temperatures compared to daytime temperatures, as seen in Fig. 1 [10], nighttime is the optimal period for chillers to operate to enhance the coefficient of performance (COP). Furthermore, Fig. 2 demonstrates that the chillers functioned in a dormant state throughout the night, resulting in a diminished coefficient of performance (COP) due to the lower part load ratio (PLR) [11,12]. To achieve the optimal coefficient of performance (COP), it is essential to use this period for cold thermal energy storage (CTES) [13].

Effective dispersion of cooling demand using a thermal energy storage (TES) unit enables the chiller plant to operate at optimal coefficients of performance (COP) [14]. The Thermal Energy Storage (TES) system utilizes sensible or latent heat energy storage devices; the latter necessitates significant energy to alter the phase of the storage medium, resulting in the latent heat (LH) releasing heat much more than the sensible heat (SH) [15]. Phase Change Materials are used to provide elevated storage density and effective solutions that correspond with the high performance of energy management [16]. Water ice is the most often used PCM cold storage material; nevertheless, several cool storage applications cannot utilize water ice owing to its low melting point and

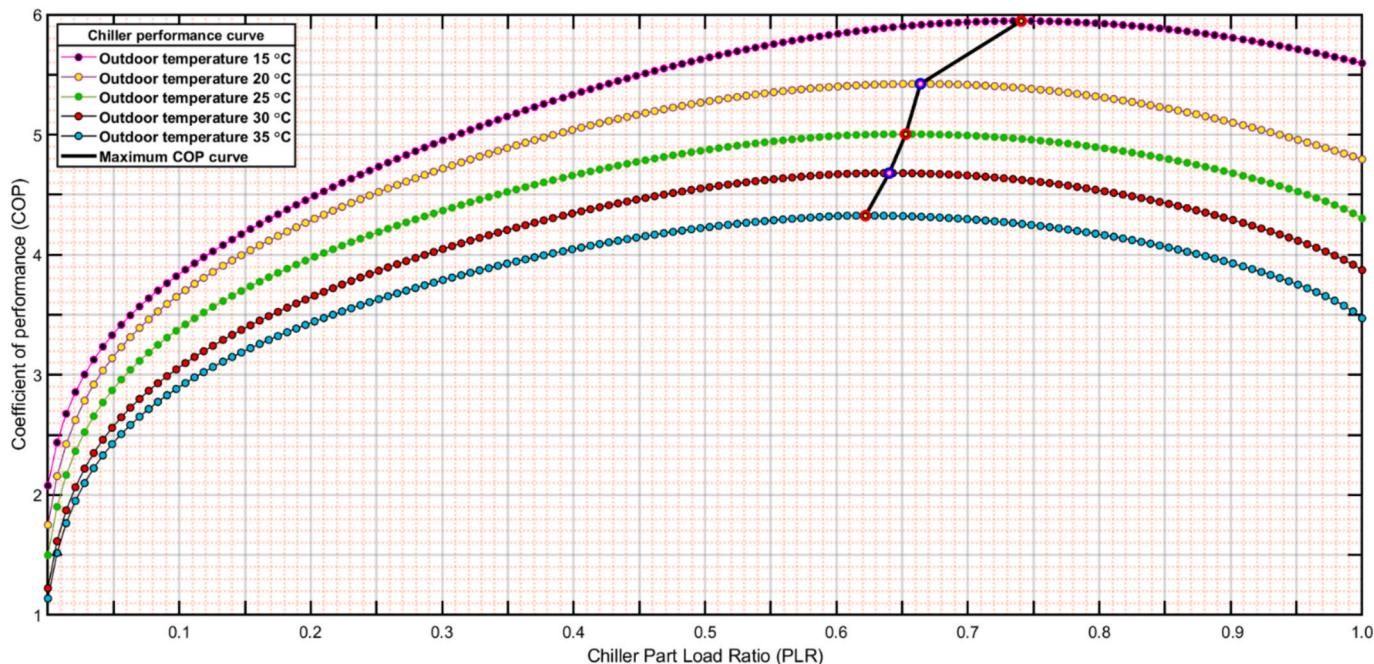


Fig. 1. Effects of varying partial load ratio (PLR) values on the coefficients of performance (COP) [10].

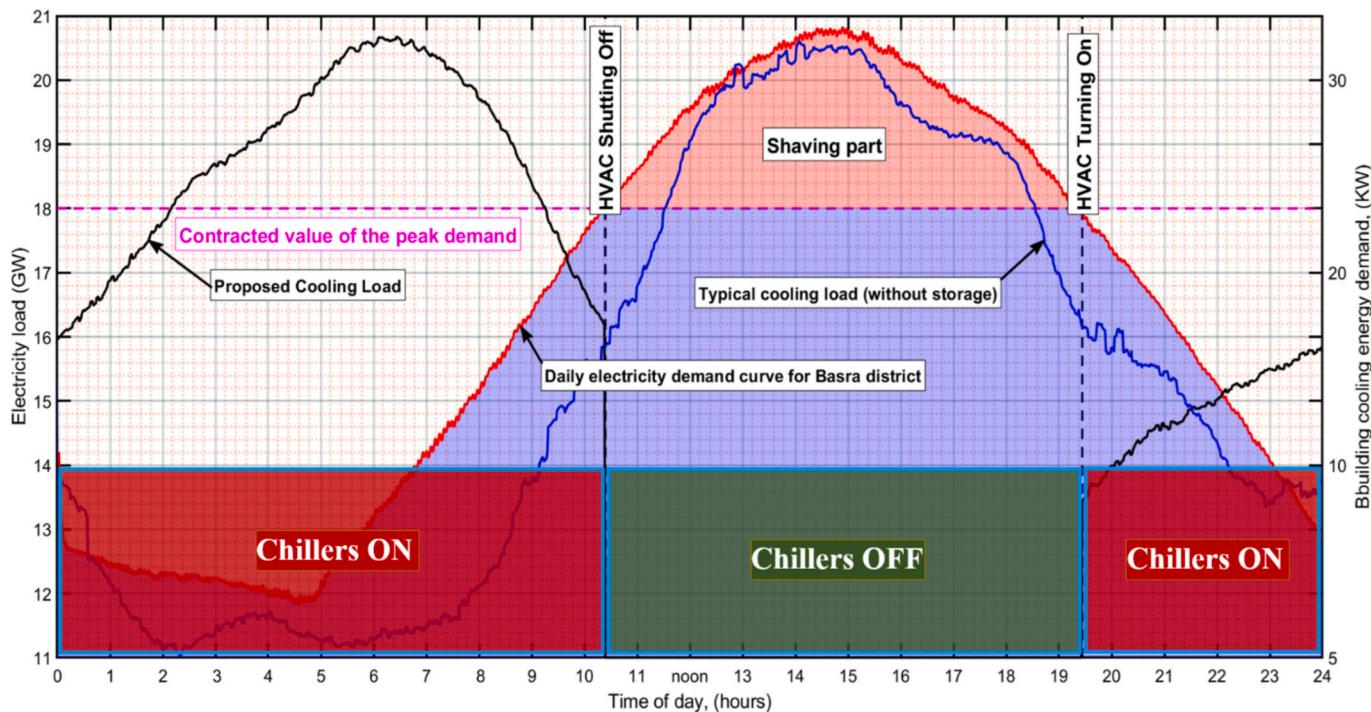


Fig. 2. The optimal timing for shifting peak load using CTES, based on daily cooling load demand [12].

low latent heat [17]. Consequently, extensive research has examined several types of phase change materials (PCMs) and shown diverse behaviors in energy storage for each material type [16].

Phase change material (PCM)-based thermal energy storage (TES) has emerged as a versatile technology with applications across various sectors, including renewable energy integration, energy conservation in buildings, solar power generation, heating and cooling systems, battery thermal management, and electronic devices [18]. TES systems are classified into three main types: (a) sensible heat storage, which involves changes in material temperature without a phase transition; (b) latent heat storage, which employs PCMs to absorb or release latent heat during phase transitions; and (c) thermochemical storage, which relies on reversible chemical reactions to store and release heat [19]. Among these, latent heat thermal energy storage (LHTES) using PCMs is considered the most promising due to its high energy storage density and ability to maintain and release thermal energy at nearly constant temperatures during phase changes. Recent advancements in passive and active strategies have been developed to optimize PCM-based TES systems, addressing challenges such as the low thermal conductivity of PCMs and improving heat transfer efficiency [20].

PCMs are classified into three primary categories: inorganic (e.g., metals, molten salts, hydrated salts), organic (e.g., paraffin, fatty acids, polyethylene glycol), and eutectic mixtures of inorganic and organic substances [21]. Each type has distinct advantages, disadvantages, and unique properties based on factors such as stability, corrosion resistance, thermal properties (e.g., phase change points and heat of fusion), and transport properties (e.g., thermal stability and liquid-state volume expansion) [22]. Inorganic PCMs, while offering recyclability, melting without component separation, and being safe, non-corrosive, and non-toxic, also present challenges such as incongruent melting, where they melt into a solid phase (typically a lower hydrate of the same salt) and a saturated aqueous phase with weak nucleation properties. Additionally, inorganic PCMs can be corrosive to materials such as copper, aluminum, and metal alloys. Organic PCMs, however, are non-corrosive, non-toxic, and do not exhibit supercooling, offering an alternative with more favorable handling properties [23].

Eutectic phase change materials (EPCMs) are homogeneous mixtures

of many PCM components, each with unique thermal characteristics and a melting point that is lower than the aggregate of its parts. The three types of eutectic mixtures (combinations of organic and inorganic phase change materials) that constitute EPCMs are often binary or ternary. These pairings are also categorized as organic-organic, organic-inorganic, and inorganic-inorganic [24]. Organic phase change materials (OPCMs) are the most often used PCMs due to their properties, including resistance to thermal cycling. One benefit of EPCMs over individual PCMs is their capacity for facile adjustment of the phase change point and latent heat by synthesizing EPCMs from various OPCMs [25].

Yang et al. [26] reported that the binary mixture of paraffin waxes as PCMs achieved excellent thermal performance and effective cool storage. Tetradecane ($CH_3 - (CH_2)_{12} - CH_3$) and hexadecane ($CH_3 - (CH_2)_{14} - CH_3$) made up the blend, which was purposefully designed to optimize the heat of fusion. The optimal melting point is achieved with a composition of 3.8 % hexadecane and 96.2 % tetradecane [27], which is suitable for integration with air handling unit operations. Weng et al. [28] studied a new battery cooling strategy that combines phase-change material with dynamic liquid cooling. This approach improves cooling performance and reduces energy consumption, especially in high-temperature environments.

This research presents an innovative combination ratio that offers substantial cost reductions compared to traditional cooling methods while significantly minimizing spatial requirements, hence creating several application possibilities. An extensive analysis of spatial and financial considerations indicates a remarkable $1681\% \pm 20\%$ reduction in volume relative to conventional chilled water tanks, highlighting the efficacy of this method. The reduced operational expenses, realized via compact chillers, cooling storage units, and supplementary components, counterbalance the original investments in storage apparatus. The research introduces a novel approach employing deep reinforcement clustering for adaptive decision policy (DRCADP), which optimizes tank sequence control (OTSC) and enhances the chiller COP, thereby improving system performance and energy efficiency by utilizing off-peak cooling hours. The integration of deep reinforcement learning (RL) clustering with multi-stage TES systems, employing PCMs, effectively connects traditional energy management with modern

advancements, providing a comprehensive solution for enhancing chiller COP, minimizing peak demand, optimizing costs, and maintaining grid stability. This unique method signifies a substantial progress in sustainable energy management for intelligent buildings, integrating sophisticated machine learning algorithms with cutting-edge TES and PCM technologies to optimize chiller operations, augment storage capacity, and boost overall system efficiency. This study aims to enhance the coefficient of performance (COP) of chillers by mitigating energy peak demand and rising costs through the integration of artificial intelligence (AI) in energy management systems utilizing phase change materials (PCMs). In summary, the primary contributions of this study are highlighted in the following aspects:

1. Shifting running a full load of chillers into the off-peak period (at night) to achieve maximum COP by the policy of CMARLDC.
2. Utilized two different types of melting points of PCM to get optimal energy efficiency by achieving a higher temperature of returned chilled water.
3. Multi-agent reinforcement learning (MARL) policies are created to interact in their environment and then represented by Lagrangian trajectory to deal with the PCM tanks based on OTSC.

2. Problem identification and modelling

This section will provide a comprehensive and brief explanation of the strategies for carrying out the proposed technique. It will be organized into subsections that outline the necessary conditions for executing this idea. By integrating PCMs into the Heating, Ventilation, and Air Conditioning (HVAC) system instead of traditional water tanks, a significant decrease in system dimensions is attained. This modification is also evident in a notable change in the management of peak load, shifting from the afternoon, around 14:00, to the overnight. The deployment of PCMs efficiently replaces the usually heavy daytime peak load on the power plant, thereby reducing its efficiency.

At night, when the chiller is operating with a low load and approaches its optimal condition, the Coefficient of Performance (COP) of the HVAC system decreases. As a result, the electricity usage is adjusted to match the reduced demand, effectively utilizing the building's thermal load at night. The strategic arrangement of the schedule leads to a higher COP for the system. This is due to the lower surrounding temperature and the efficient condensing processes, which are especially noticeable in air-based cooling systems. This work presents a dual-stage methodology in which chilled water flows consecutively through two stages that contain PCM. A single PCM demonstrates a low critical temperature, which is beneficial in certain situations, while it has a limitation in terms of heat capacity. In contrast, the second PCM has a greater melting point and improved heat capacity. In order to optimize the COP and improve the system's thermal conductivity, fins have been integrated. This comprehensive strategy is in line with the overarching objective of maximizing HVAC performance and energy savings.

2.1. Ideal eutectic PCM

Paraffin waxes stand out among the variety of PCMs discussed in the previous section due to their descending melting point, affordability, safety, dependability, predictability, non-corrosiveness, and chemical inertness [26]. These materials are notable for their minimum volume change during phase transitions, uniform melting without component segregation, and progressive melting point increase with increasing carbon atom count. Due to its inherent flexibility, the mix of material components inside the eutectic can be adjusted to obtain a nuanced alignment between the melting point range of PCMs and the system's operating temperature. Eutectic PCM mixtures, such as paraffin waxes, are utilized in cold storage systems due to their exceptionally high energy storage density, enhancing the efficiency and compactness of the storage system. Due to the organic-organic eutectic exhibits properties

such as high latent heat, cost-effectiveness, chemical stability, and non-toxicity. On the other hand, one of the major disadvantages of organic PCMs is their flammability, making them unsuitable for high-temperature applications. These mixtures are utilized in various applications, from industrial processes to everyday products. Particularly, the mixture of hexadecane ($CH_3 - (CH_2)_{12} - CH_3$) and tetradecane ($CH_3 - (CH_2)_{14} - CH_3$) appears as a promising candidate after a thorough investigation of organic-organic eutectic types in the literature (Fig. 3).

This configuration aligns with the design temperature of the air handling unit (AHU), as illustrated in Fig. 3 (sourced from Reference [27]). The behavior of different mixture ratios is depicted in Fig. 3, highlighting the green zone, which represents the optimal operating temperature, beginning at 75 vol% of tetradecane. To ensure efficient thermal energy storage (TES), the two types of PCM heat storage systems must be characterized by distinct melting temperatures based on chilled water operation and have to achieve the highest possible heat of fusion. Accordingly, the optimal compositions for PCM1 and PCM2 are determined to be 99 vol% tetradecane and 1 vol% hexadecane (indicated by the pink dashed line), and 92 vol% tetradecane and 8 vol% hexadecane (indicated by the yellow dashed line), respectively. These concentrations have been demonstrated to be ideal, producing the lowest temperature and maximum cooling storage capacity.

The storage configuration incorporating two distinct melting points, as proposed in this study, achieved significant thermal storage capacity, as illustrated in Fig. 4. The schematic not only depicts the essential features of the two PCM storage units, but it also emphasizes a carefully adjusted heat transfer configuration that is necessary to guarantee ideal thermal conductivity when the PCM tank storage is being charged and discharged. Furthermore, our previously developed fin tree-type network [29] is integrated to improve thermal conductivity, facilitating efficient heat transfer without necessitating the physical relocation of the PCMs. This structure is well visualized in Fig. 5's top view, which also highlights the interaction between the fin network and PCM tanks. By greatly enlarging the heat exchange area, this structural modification enables chilled water to flow through the piping coil inside the tanks and efficiently reach every surface of the fins network [30]. The incorporation of this structural element emphasizes how crucial it is to maximize the thermal energy storage system's overall efficacy and efficiency.

2.2. Building, PCM tanks, and HVAC systems modelling

When a pre-cooling coil is installed in an air handling unit (AHU), several alterations take place. The pre-cooling loop is primarily designed to lower the air temperature prior to it entering the principal cooling coil. This system helps the incoming fresh air become less humid by allowing vapour to condense into droplets.

Thermophysical properties of construction materials cannot be 100 % accurate when simulating interior conditions. By combining physical and empirical methods, hybrid modelling has successfully characterized true internal thermal perceptions. This research leverages a previously validated model established in the author's prior work [31], which was implemented by integrating HVAC and building fixtures. The focus here shifts towards the model's novel application in the smart building energy optimization domain. The established validation in the above work serves as a robust foundation for the model's reliability and accuracy in this new context. Therefore, while the current study may not explicitly present a specific validation for the model within the smart building application, the prior validation lends significant credibility and strengthens the model's applicability in this novel scenario. In order to set up the model's internal conditions, physical properties and the empirical RLF were used. This was achieved by applying the energy and mass balance Eqs. (1) and (2) [32].

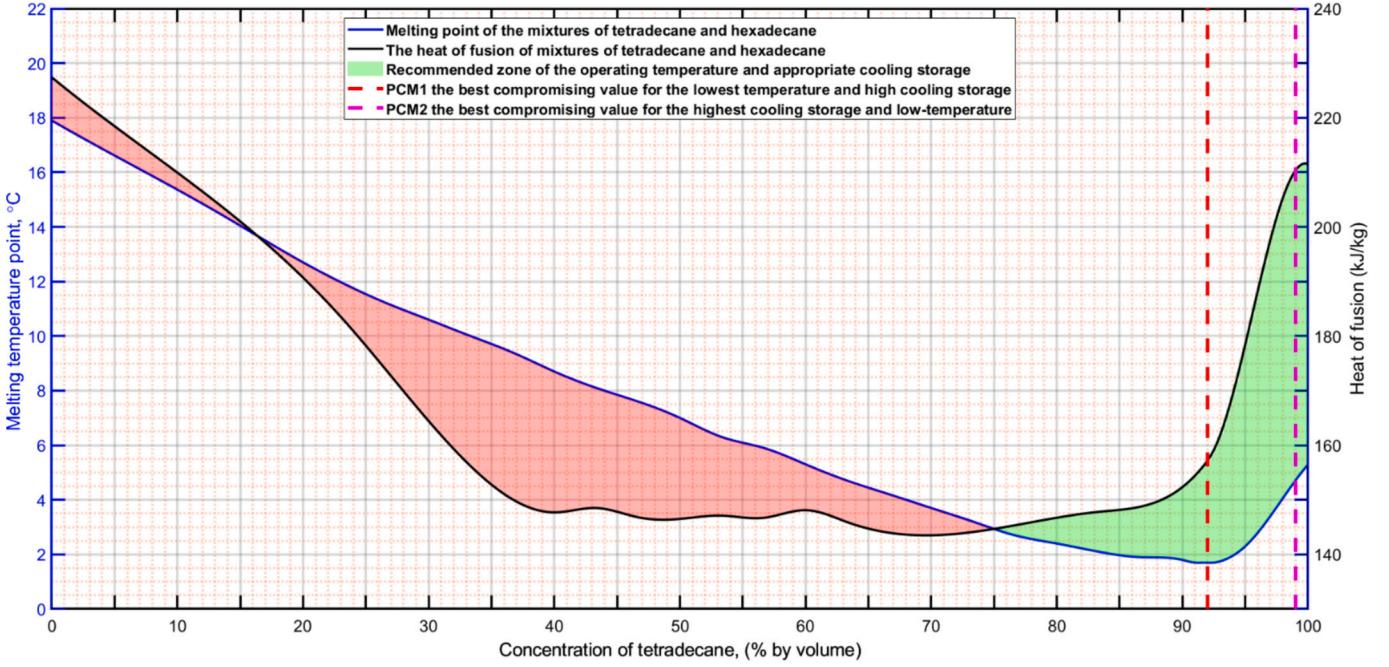


Fig. 3. The optimal concentration values for each of the double stages of PCM, based on the heat of fusion and melting points of hexadecane and tetradecane mixtures, as measured by a thermo-sensor: this data was adapted from Ref. [27].

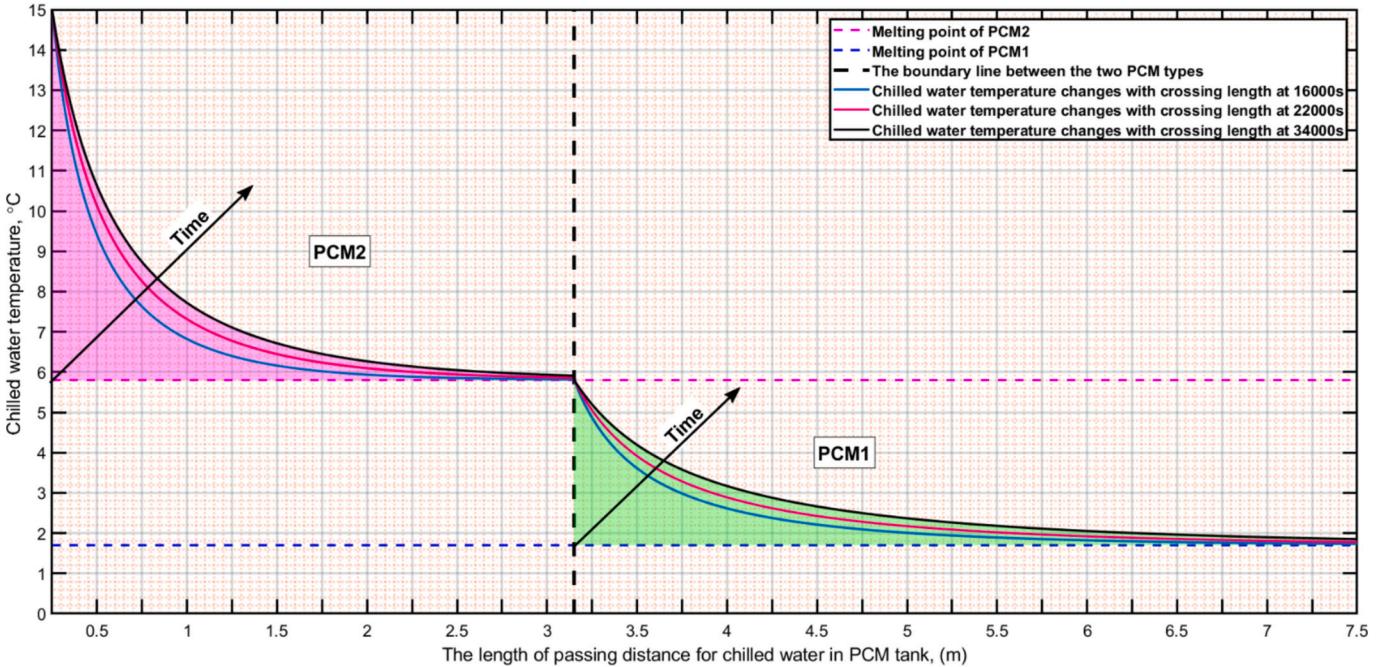


Fig. 4. The behavior of chilled water in relation to the melting points of PCM1 and PCM2 over time and at various tank depths.

$$\overbrace{\dot{Q}_s}^{\text{Cooling load}} = \overbrace{\dot{Q}_{air} + \dot{Q}_{fur}}^{\text{Accumulation or storage energy}} + \overbrace{\dot{Q}_{opq} + \dot{Q}_{fen} + \dot{Q}_{slab} + \dot{Q}_{inf} + \dot{Q}_{ig,s}}^{\text{Difference between input and output of energy}} \quad (1)$$

where $\dot{Q}_{s,t}$ is the cooling load exerted by AHU, \dot{Q}_{air} is the storage energy at air mass, \dot{Q}_{fur} is storage energy at furniture mass, \dot{Q}_{opq} is the convection heat gain from opaque surfaces, \dot{Q}_{fen} is the conduction and solar radiation heat gain, \dot{Q}_{slab} is convection heat gain from slab floors, \dot{Q}_{inf} , is the heat gain due to infiltration and $\dot{Q}_{ig,s}$ is the sensible cooling load from internal gains [33].

$$\dot{Q}_{s,t} = \overbrace{\dot{m}_m c p_a (T_{r,t} - T_{s,t})}^{\text{Cooling load exerted by AHU}},$$

$$\dot{Q}_{air} = \overbrace{M_{air} c p_a \frac{dT_{air}}{dt}}^{\text{storage energy at air mass}},$$

$$\dot{Q}_{fur} = \overbrace{\sum_j M_{fur,j} c p_{fur,j} \frac{dT_{fur}}{dt}}^{\text{storage energy at furniture mass}},$$

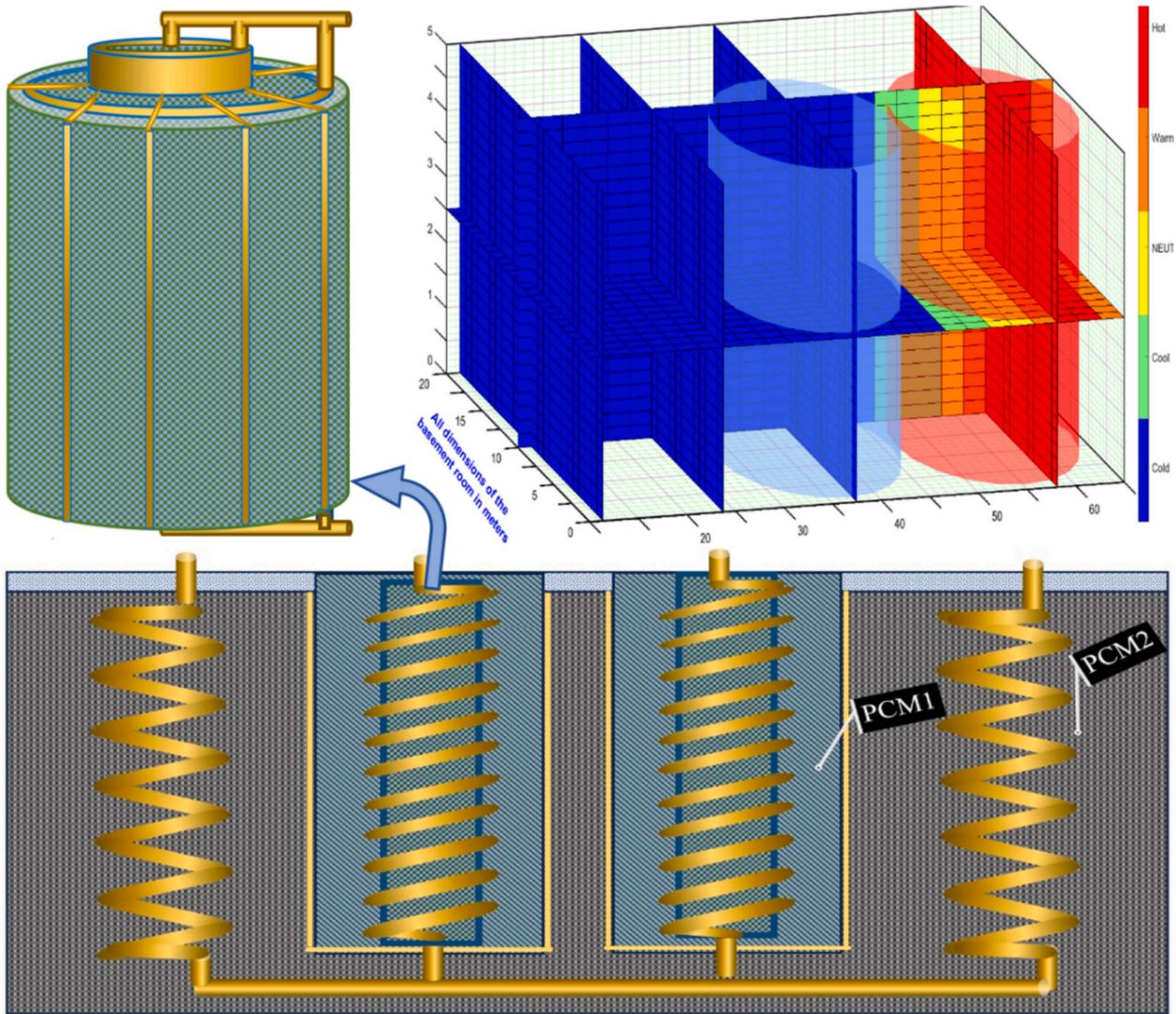


Fig. 5. The design of the proposed dual PCM heat storage system and the optimized configuration of the heat transfer method, ensuring efficient thermal conductivity during the charging and discharging processes of the PCM tank.

$$\begin{aligned} \dot{Q}_{opq} &= \underbrace{\sum_j A_{wj} h_{ij} (T_{Wl_{in}} - T_r)}_{\text{convection heat gain from opaque surfaces}}, \\ \dot{Q}_{fen} &= \underbrace{\frac{(T_{g_{in}} - T_r)}{R_g}}_{\text{conduction heat gain}} + \underbrace{\sum_j A_{fenj} PXI_j \times SHGC_j \times IAC_j \times FF_{sj}}_{\text{solar radiation heat gain}}, \\ \dot{Q}_{slab} &= \underbrace{\sum_j A_{slbj} h_{ij} (T_{slb_{in}} - T_r)}_{\text{convection heat gain from slab floors}}, \\ \dot{Q}_{inf} &= \underbrace{C_s \times A_L \times IDF(T_{o,t} - T_{r,t})}_{\text{heat gain due to infiltration}} \text{ and} \\ \dot{Q}_{ig,s} &= \underbrace{136 + 2.2A_{cf} + 22N_{oc}}_{\text{sensible cooling load from internal gains}}. \end{aligned}$$

Substituting these quantities into Eq. (1) yields:

$$\begin{aligned} \dot{m}_m c p_a (T_{r,t} - T_{s,t}) &= M_r c p_a \frac{dT_{r,t}}{dt} + \sum_j M_{furj} c p_{furj} \frac{dT_{fur,t}}{dt} \\ &+ \sum_j A_{wj} h_{ij} (T_{Wl_{in},t} - T_{r,t}) + \frac{(T_{g_{in},t} - T_{r,t})}{R_g} \\ &+ \sum_j A_{fenj} PXI_j \times SHGC_j \times IAC_j \times FF_{sj} \\ &+ \sum_j A_{slbj} h_{ij} (T_{slb_{in}} - T_r) + C_s \times A_L \times IDF(T_{o,t} - T_{r,t}) \\ &+ 136 + 2.2A_{cf} + 22N_{oc} \end{aligned} \quad (2)$$

The building latent heat gain is related to moisture transfer, which can be evaluated by applying the conservation of time dependent mass law on the building control volume as shown in Eq. (3) [33]

$$\begin{aligned} \overbrace{\dot{m}_s (\omega_{r,t} - \omega_{s,t})}^{\text{rate of moisture withdrawal by AHU}} &= \overbrace{\frac{d M_r \omega_{r,t}}{dt}}^{\text{rate of moisture change}} + \overbrace{\frac{\dot{Q}_{ig,l}}{h_g}}^{\text{rate of moisture generation}} \\ &+ \overbrace{\dot{m}_o \dot{t}_{rr,t,inf}}^{\text{rate of moisture transfer}} \end{aligned} \quad (3)$$

Eq. (4) represents the integration of the thermofluid formulas for the

building and air handling unit (AHU) in HVAC units to determine the interior thermal dynamics [34].

temperature and humidity ratio, while the $G_{1,1}(s)$, $G_{1,2}(s)$, ..., $G_{1,12}(s)$, $G_{2,1}(s)$, $G_{2,2}(s)$, ..., $G_{2,12}(s)$ denote

$$\begin{bmatrix} T_r(s) \\ \omega_r(s) \end{bmatrix} = \begin{bmatrix} G_{1,1}(s) & G_{1,2}(s) & G_{1,3}(s) & G_{1,4}(s) & G_{1,5}(s) & G_{1,6}(s) & G_{1,7}(s) & G_{1,8}(s) & G_{1,9}(s) & G_{1,10}(s) & G_{1,11}(s) & G_{1,12}(s) \\ G_{2,1}(s) & G_{2,2}(s) & G_{2,3}(s) & G_{2,4}(s) & G_{2,5}(s) & G_{2,6}(s) & G_{2,7}(s) & G_{2,8}(s) & G_{2,9}(s) & G_{2,10}(s) & G_{2,11}(s) & G_{2,12}(s) \end{bmatrix}$$

$$\begin{bmatrix} \dot{m}_w(s) \\ \dot{m}_{mw}(s) \\ \dot{m}_{os}(s) \\ \dot{m}_r(s) \\ T_o(s) \\ \omega_o(s) \\ f_4 \\ \dot{Q}_{ig,l} \\ A_{slab} \\ f_{DR} \\ k_2 \\ T_r(s) \end{bmatrix} = \begin{bmatrix} \dot{m}_w(s) \\ \dot{m}_{mw}(s) \\ \dot{m}_{os}(s) \\ \dot{m}_r(s) \\ T_o(s) \\ \omega_o(s) \\ f_4 \\ \dot{Q}_{ig,l} \\ A_{slab} \\ f_{DR} \\ k_2 \\ T_r(s) \end{bmatrix} \quad (4)$$

Eq. (4) represents a concise formulation that encapsulates the key elements of the model under investigation, providing a succinct and precise description of the relevant variables and their relationships. Where the variables on the left-hand side of Eq. (4) represent the indoor

the factors of the twelve input transfer functions. The aim is to show the parameters of the twelve input functions.

As illustrated in Fig. 6, the 12 parameters controlled in the full structure provide the independent factors of the sub-model transfer function.

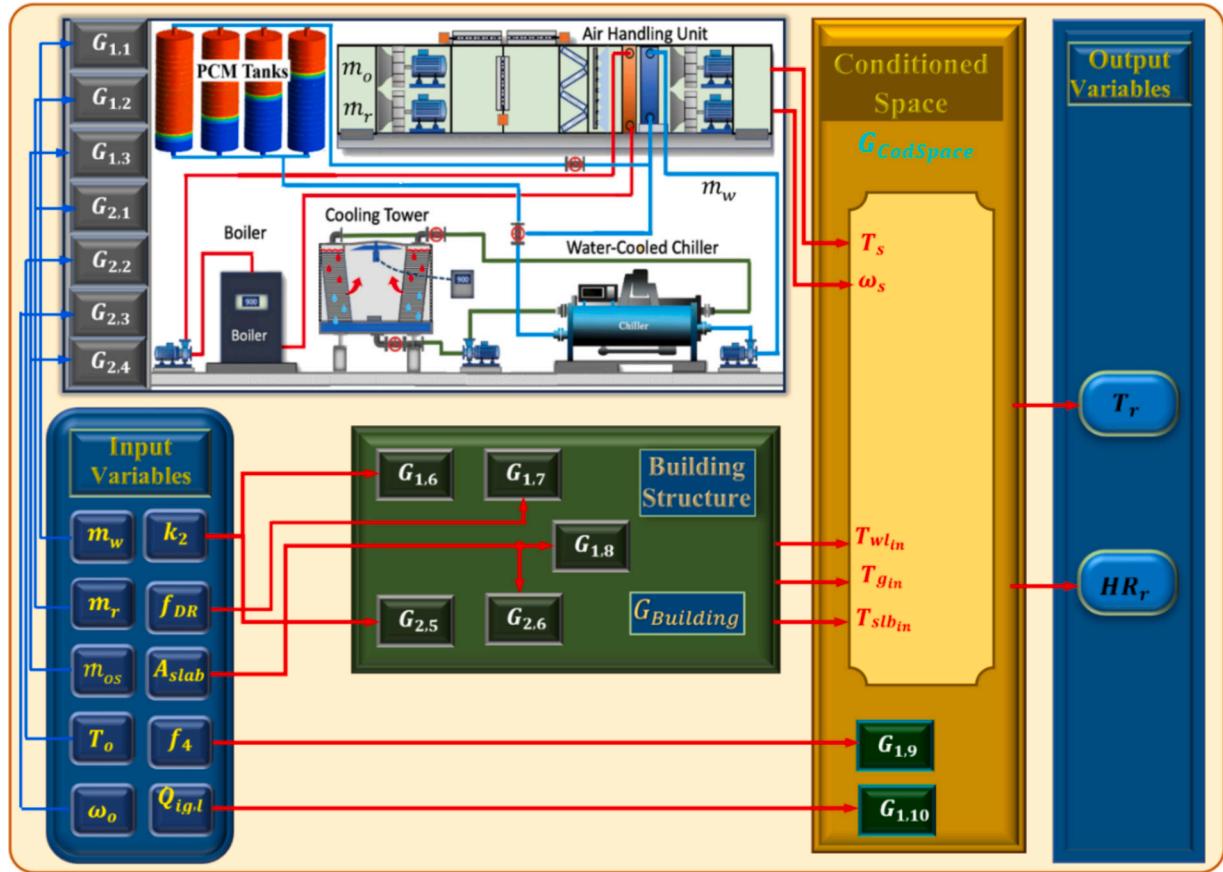


Fig. 6. The schematic diagram of the building submodules sequencing chain, including the submodules for PCM storage tanks and HVAC systems.

2.3. Formulate the problem of PCM tank control as a DRCADP

Energy conservation and ideal interior conditions through environmental interactions are the overriding goals of the structural multi-agent system, which uses a sequential decision-making approach to understand varied activities. A dynamic organizational method guides intelligent agent selection, especially during the investigative phase. Room temperature, humidity, and hot water outlet temperature are all aspects of the environment that might affect the structure's condition during this inspection, and all of this information is reported. These readings were probably taken during an experimental batch stage when the Markov Decision Process (MDP) was being trained using tuples. It is from these measurements that subsequent states are derived. In the context of Markov Decision Processes (MDPs), the stationary policy states that the probability of going to the next state (s') depends only on the present state (s) and ignores the previous sequence of event changes. In Eq. (4), the transition probability matrix (P) is defined, which is a discrete calculation that describes the statistics of the whole structure under different cases. When this matrix is taken into account, it shows a structured organization in which the model creates a systematic relationship between state-action pairings and their future states ($S_{t+1} = P(\mu_{st,at}, \theta_{st,at})$) where $\theta_{st,at}$ is the mean function, and $\mu_{st,at}$ is the transition function's factors [35].

$$p_{ss'} = \begin{bmatrix} S_{1,t} & S_{2,t} & \dots & S_{n,t} \\ S_{1,t+1} & S_{2,t+1} & \dots & S_{n,t+1} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n,t} & S_{n+1,t} & \dots & S_{n+n,t} \end{bmatrix} \quad (5)$$

S is represented by S_t and S' is designated by S_{t+1} , where S is the current state and S' is the following step, respectively. The probability matrix of transitioning from one condition to another nation is $p_{ss'}$.

A Markov Decision Process (MDP) is the best way to represent the unsupervised operation of the air handling units (AHUs), chiller, and entire structure while thinking about the effective total advantage technique. The energy-storing needs of the building and the optimal configuration form the basis of this approach. This problem and its solution can be addressed by utilizing algorithms that are based on machine learning. Thus, Markov Decision Processes (MDPs) are the

dynamic RL approach for sequential decision-making, allowing agents in fields encompassing the entire building to evaluate action laws. The finite Markov Decision Process (MDP) can be defined using a 5-tuple (S, A, P, R, γ) . S denotes the criteria for the outgoing and incoming air conditions, while A denotes a limited set of choices accessible under all conditions at any one time (t). The procedure's transition probability from state (s) to state (s') is denoted by $P(s, s'|s, a)$. The agent finds the reward it got from a specific state at time step t , denoted as $r(s_t, a_t) \in R$ after it has observed the current states and made its actions. In terms of discounting, the reduction factor is denoted by $\gamma \in [0, 1]$. A simplified graphic representation of the primary RL cycle is shown in Fig. 7.

The agent's mastery of the best strategy is paramount inside the domain of a Markov Decision Process (MDP). An ideal policy model is constructed by extracting insights from environmental data using reinforcement learning (RL), a basic machine learning technique. Finding the optimal course of action that maximizes future rewards while taking the discount value into account is the ultimate objective of the agents. The value function, $V(s)$, that is unique to each state determines how cumulative return is computed. As we iteratively solve the estimated $V(s)$ to get the ideal value, Bellman's formula becomes the crucial and pivotal instrument. A particular state can be valued by building an implicit and an explicit function for mathematical sequences, which requires two essential components. The first part is an encapsulation of the overall benefit from that state's actions, while the second part is a connection between the value of the next state-action pair. Two other ways can these relationships be expressed: one using an algebraic formula (Eq. (6)), and the other using a square matrix formula (Eq. (7)) [36].

$$V(s) = R_s + \gamma \sum_{s' \in S} p_{ss'} V(s') \quad (6)$$

R_s is a scalar recompense that, in the given context, indicates the actual income an agent receives upon moving from state (s) to state (s'). The value characteristic of the subsequent state is denoted by $V(s')$, whereas the lowered future profits are represented by γ [37].

$$\begin{bmatrix} V(1) \\ V(2) \\ \vdots \\ V(n) \end{bmatrix} = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_n \end{bmatrix} + \gamma \begin{bmatrix} P_{1,1} & P_{1,2} & \dots & P_{1,n} \\ P_{2,1} & P_{2,2} & \dots & P_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n,1} & P_{n,2} & \dots & P_{n,n} \end{bmatrix} \begin{bmatrix} V(1) \\ V(2) \\ \vdots \\ V(n) \end{bmatrix} \quad (7)$$

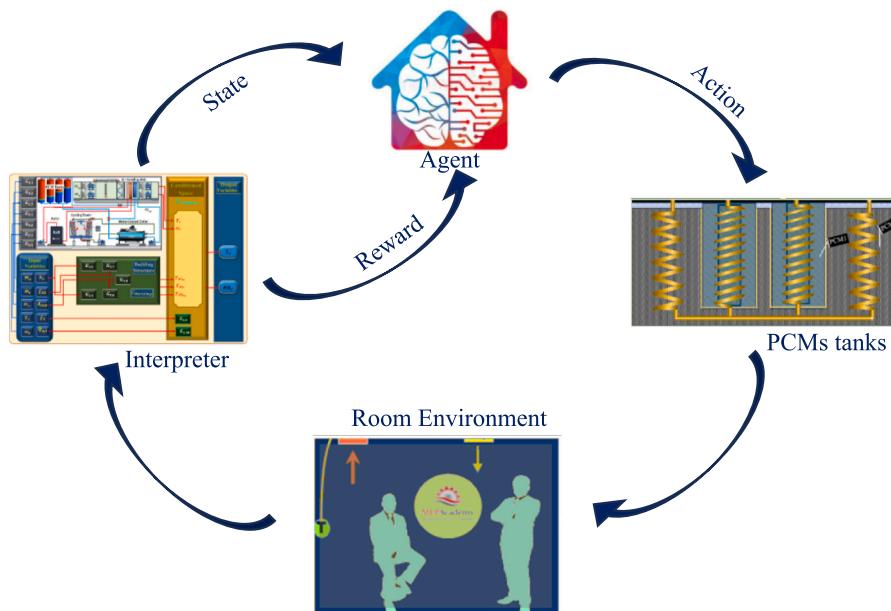


Fig. 7. How the RL agent learns through continuous actions based on the dynamic algorithm of the MDP to achieve its goal.

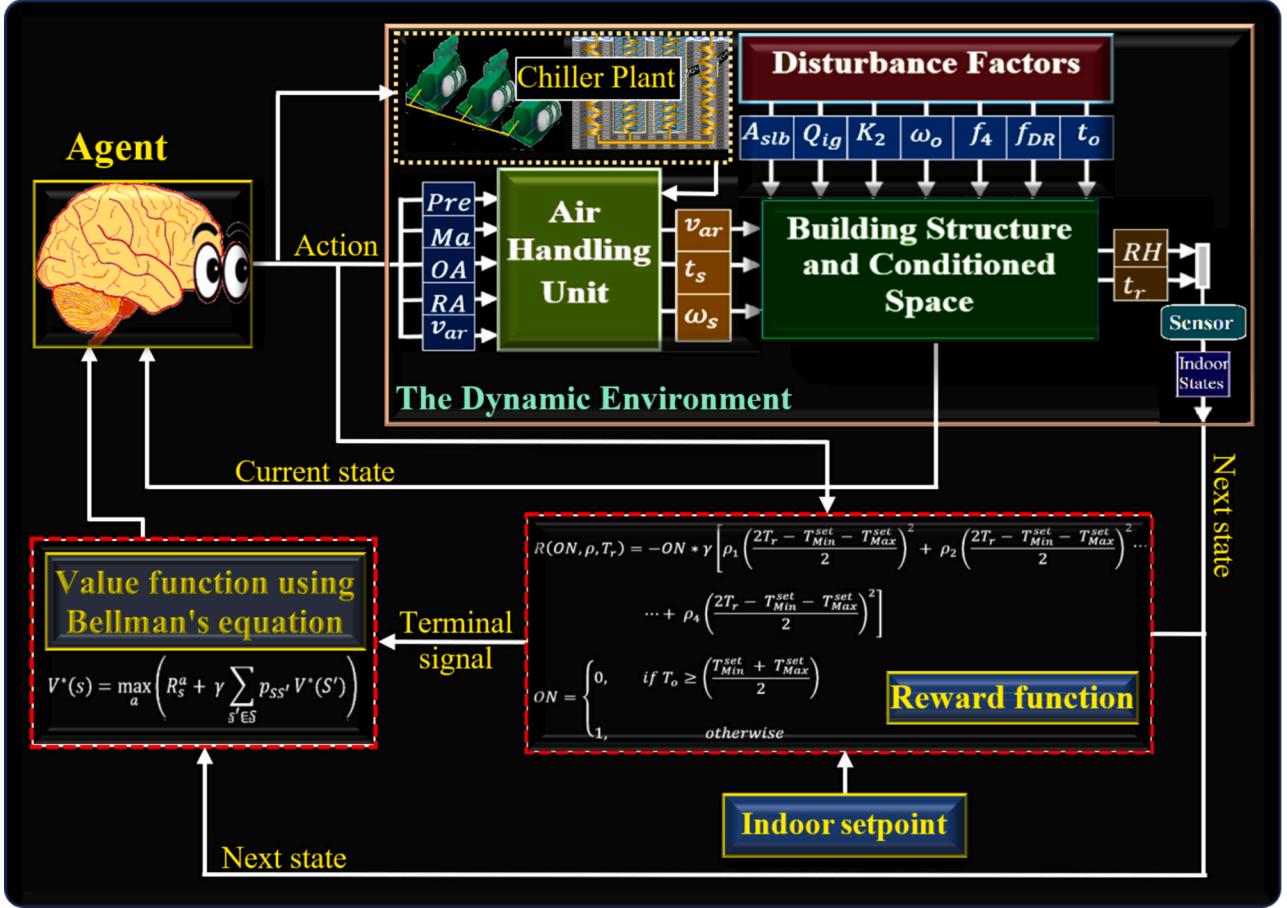


Fig. 8. The general schematic of how RL solves the MDP to enable the CMAS to tune the AHU inputs.

A collaborative multi-agent system (CMAS) adjusts their behavior in response to one another, synchronizing it and transmitting agent selection criteria based on the circumstances at hand. Using these manipulative action approaches, the structure's lighting, OTSC (optimal tank sequencing control), ventilation, interior temperature, and other features can all be altered. The primary variables that constantly alter the parameters are the operation of the windows and doors, the OTSC, the way the fuel valve status of the chiller relay is altered, how lights brighten or dim, and if they are turned on or off. The learning process for an agent arises from these discrete-time actions. The agents send out the following activities each period: $\{U = u_{closWin}^1, u_{openWin}^2, u_{onLigh}^3, \dots, u_{openVal}^m\}$. The agents act in sets $A = \{a^1, a^2, \dots, a^p\}$ of the OTSC in physical space by using the multi-objective optimization strategy, $p = m^{no.Equ}$, taking into account all changing parameters to accomplish energy savings by implementing MAS learning procedures. The p polynomials function of CMAS based on the number of state variables may be seen using the AHU's input lines in Fig. 8. This suggests that the growth of the stats dataset causes the mathematical space of action sets A to grow exponentially, which significantly reduces agent action efficiency. But even though reinforcement learning (RL) is among the best approaches to deal with the maximum degree of plausibility (MDP) and produce best-case scenarios based on setting input, multi-agent actions will not work well when the nonlinear relationship between the number of states and the space of contends in a system increases. Furthermore, by using a Bellman equation in the optimal $V^*(s)$ provided by Eq. (8), the Bellman optimality equation offers an iterative way to find the ideal value of the operation in an MDP [38].

$$V^*(s) = \max_a \left(R_s^a + \gamma \sum_{s' \in S} p_{ss'} V^*(s') \right) \quad (8)$$

The multi-policy MDP scenario is evaluated using scalar incentive (or cost) variables that specify the reinforcement output produced by the environment in each condition. A tuple of five components (S, A, P, R, γ) dependent on the benefit of MDP defines the computational framework for modelling making choices. These components can be thought of as link weights that have been changed with respect to time. At time t , the advantages of the current solution are at their highest value. The MDP tuple uses a MORL technique to fine-tune its components by learning the multi-agent to follow the reward function, which enforces a priority level based on the order of the agents. This allows for sufficient pathfinding reliability to be achieved. The Bellman equation carries out the registration procedure to provide the optimal (max) incentive, as seen in Fig. 9, after the relationship between the reward function and energy use is explained, and a trade-off analysis between energy usage and thermal comfort is conducted. The feedback incentive of the agent of the PCM tanks arranging was provided by Eqs. (9) and (10) [38].

$$ON = \begin{cases} 0, & \text{if } T_o \geq \left(\frac{T_{Min}^{set} + T_{Max}^{set}}{2} \right) \\ 1, & \text{otherwise} \end{cases} \quad (9)$$

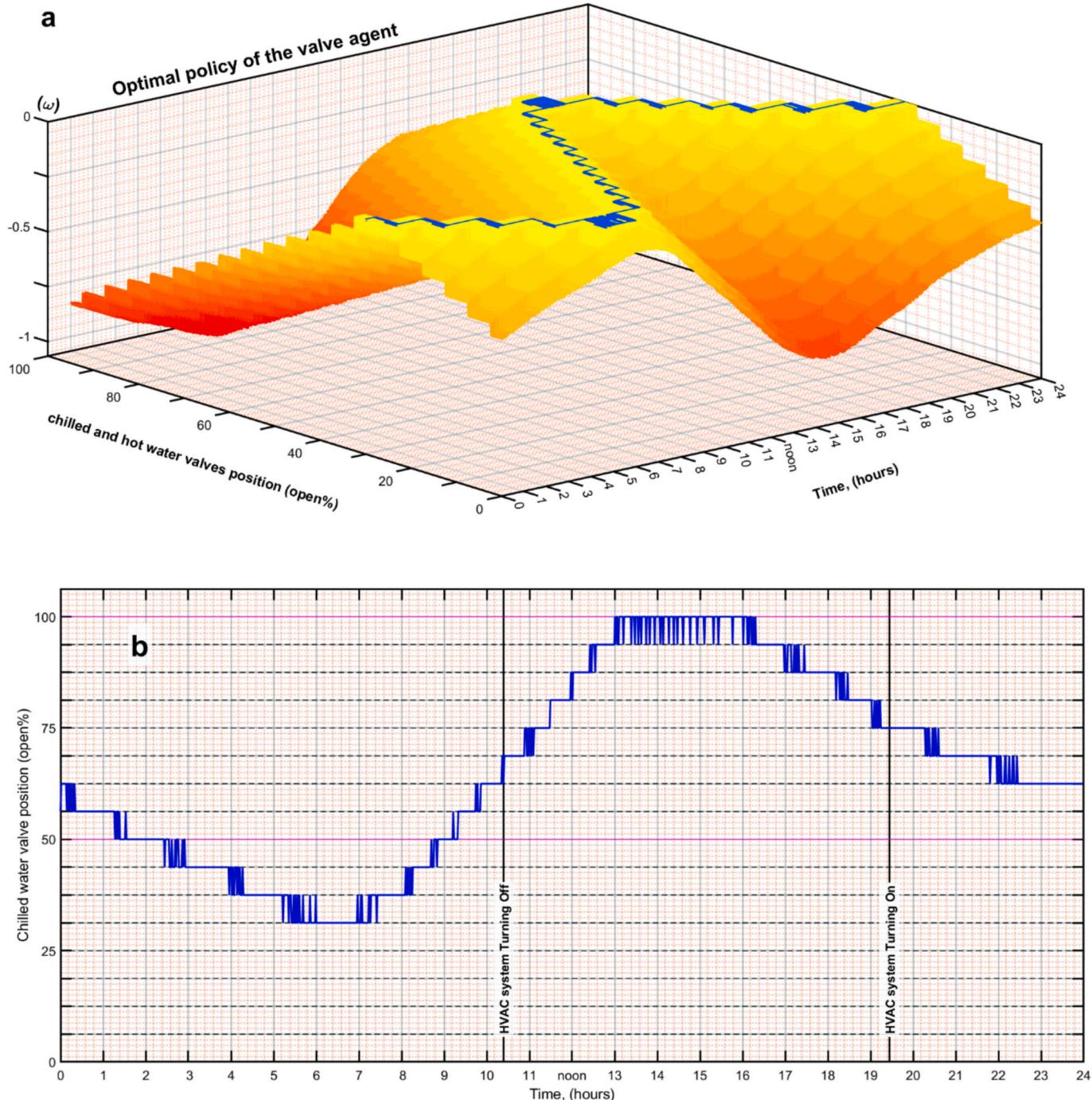


Fig. 9. Both 3D (a) and 2D (b) views to derive the optimal agent policy using the Bellman equation, based on the reward weights illustrated in (a).

$$R(ON, \rho, T_r) = -ON^* \gamma \left[\rho_1 \left(\frac{2T_r - T_{Min}^{set} - T_{Max}^{set}}{2} \right)^2 + \rho_2 \left(\frac{2T_r - T_{Min}^{set} - T_{Max}^{set}}{2} \right)^2 \dots + \rho_4 \left(\frac{2T_r - T_{Min}^{set} - T_{Max}^{set}}{2} \right)^2 \right] \quad (10)$$

where γ is the value of the reward's periodic declining element ($\gamma \in [0, 1]$), γ was 0.98 in the appear examination, and ρ is the threshold value used to initiate the sequential transformation decision for tanks, which will be stated in paragraph (4.2), starts the ensuing finding for dual PCMs tanks. The environment incentivizes the agents $R(ON, \rho, T_r)$ based on the condition states of indoor and outdoor temperatures (T_r, T_o). All

chillers are turned off as the cooling load exceeds the peak load deadband line ($ON = 0$). The default values for the top and bottom deadband state temperatures and moisture levels are, respectively, $T_{Min}^{set} = 20^\circ\text{C}$, $T_{Max}^{set} = 24^\circ\text{C}$, $RH_{Min}^{set} = 45\%$, and $RH_{Max}^{set} = 55\%$.

When the preceding method is applied, continuous learning maximizes value for states; both V and π are indicated by an asterisk (*); the agents' objective is to reap the yield (benefits) across an infinite horizon. Using a best-case scenario of $a^* = \pi^*$, the change approach creates an ideal activity that indicates whether iterative RL algorithms will eventually come to an ideal value function $V^*(s) = V \pi^*$. Alternatively, Eq. (11) shows that the RL method of determining the best MDP strategy maximizes the v value over time [39].

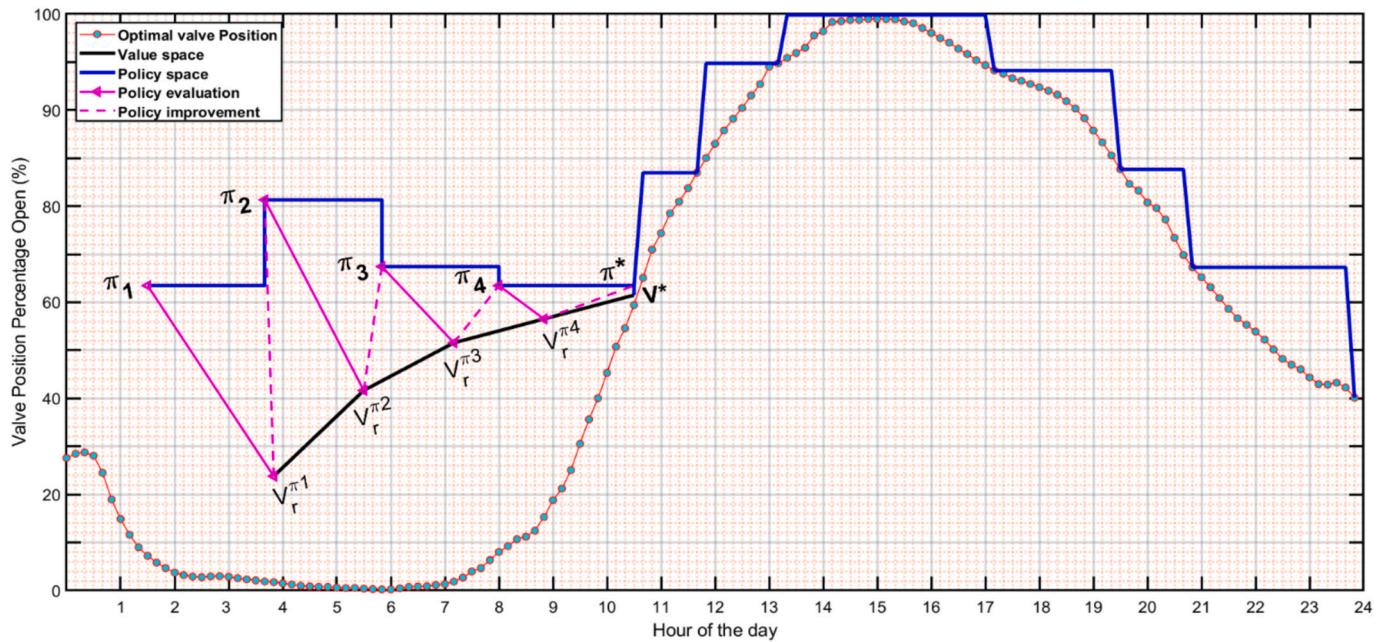


Fig. 10. The recursive tracking of the state by the value function, which iteratively refines the policy until it converges to the optimal policy, π^* .

$$\pi^*(a|s) = \operatorname{argmax}_{a \in A} V^*(S') \quad (11)$$

The updated path in Fig. 10 illustrates how a broad range of Eqs. (8) and (11) will be used to carry out the series of regulatory upgrades after two crucial steps (policy development and policy assessment) have been repeatedly generalized.

3. Description of design DRCADP

The fact that the multi-agent rules include as much material as a set

of tuples has made processing that enormous quantity of data more challenging. It would be better to create a novel system with large-scale areas and information storage as its main features. Unlike DRL, the multi-agent can handle continuous large-scale multi-action areas because of the recommended form DRCADP of multi-objective reinforcement learning (MORL), which is powerful enough to store and analyze data quickly. Several cooperative multi-agent laws are designed to maximize energy efficiency while preserving indoor thermal comfort; these objectives manifest over very long-time horizons when applied to the best possible arrangements of chillers and the relocation of the fuel

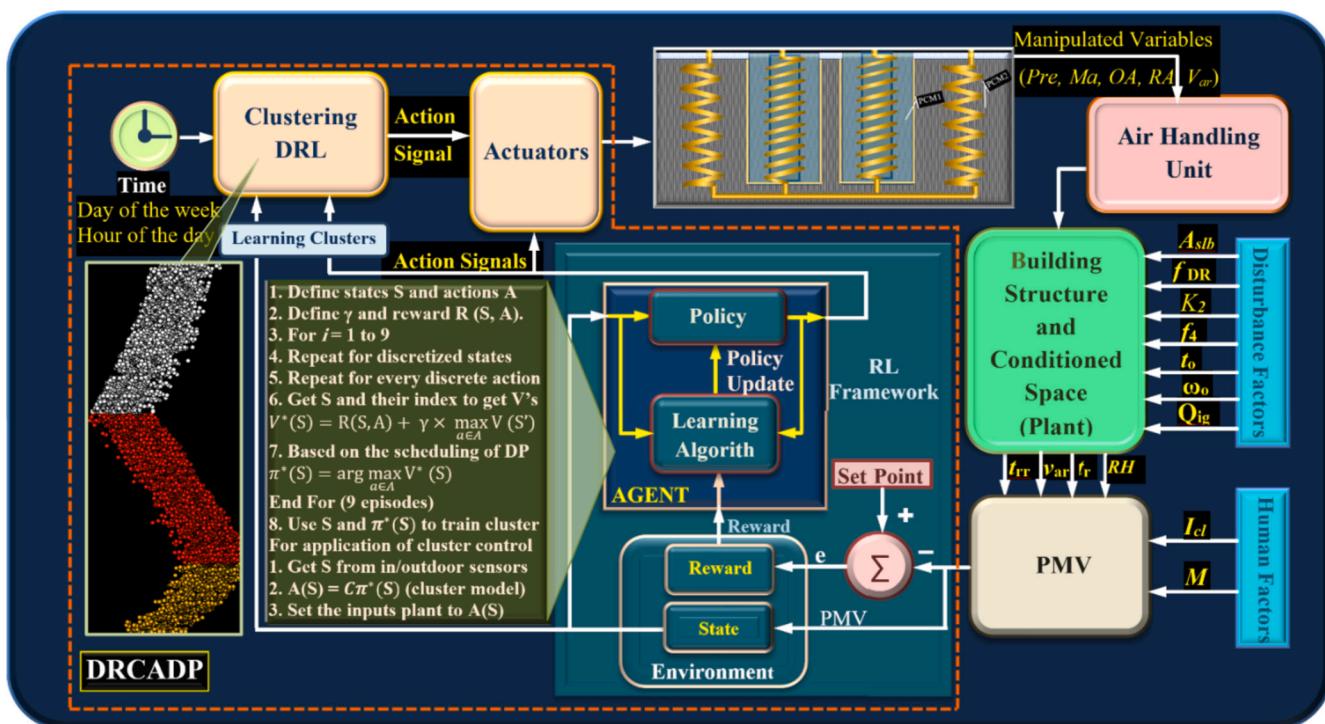


Fig. 11. The configuration of DRCADP and its implementation of a double-stage PCM, utilizing multi-agent interaction to synchronize activities and mitigate peak load.

valve for the chiller solenoid. To ensure a comprehensive analysis of DRCADP certainty methodologies, this part defines the variables of units for each agent's procedure rules using a cluster of multi-agent selections generated using the bang-bang method. Consequently, the location of the motor-driven fresh and return air dampers, the OTSC system, the rate at which chilled water is delivered, and the fan speed that is adjusted to adjust air flow rates are all of considerably greater quality. As it goes around the building's surroundings, the agent modifies the signals of these motors and actuators until it reaches the interior PMV threshold value. $[T_r(t) \quad RH_r(t)]_{DES}^T$ at the appointed time and in an appropriate condition. By putting into practice, the incentive system that follows its model-based policy, the PMV vector value is achieved. Consequently, $e = [T_r(t) \quad RH_r(t)]^T - [T_r(t) \quad RH_r(t)]_{DES}^T$ is the state error between the expected and actual values. Several elements influence indoor thermal convenience. The unit is impacted by predicting the reward function.

The initial design of DRCADP needs to be produced effectively for various hybrid layers and backed by the ability to handle issues involving numerous agents with sizable status and activity spaces. Updated calculations for the set point and advantage level coverage rates were made for the hybrid layer's key factors (physical and neural network layer load). Each agent's cluster centres of action routes (the RL policy displayed) regress in accordance with the updated or altered variables, as illustrated in Fig. 11. Considering the DRCADP's focus on observation and multi-agent interaction, it is possible to build this framework in two phases: one offline for cluster establishment and the other online for updates. Nonetheless, the well-structured hybrid layers could be helpful in modelling situations. Before starting the two tuning processes, three successive creation procedures must be finished. The clustering, Lagrangian determination, and basic mapping stage of the RL approach are the logical steps.

3.1. DRCADP policy generation based on CMAS

The suggested bang-bang regulator is appropriate for cooperative multi-agent systems (CMAS) and multi-objective optimization. The main goal of this chapter is to present the optimal rules framework for RL CMAS, which results in cost-effective HVAC unit management. By lowering the quantity of air circulation, the variable air volume (VAV) AHU assists HVAC systems in achieving energy-efficient thermal

comfort on cold days [40,41]. The programmable or distinct variables in the suggested framework are defined by a finite set of states and a finite action space, and the agents of the RL algorithm function as the link between its inputs and outputs. The primary/pre-cooling coil valve spots, the VAV unit tune dampers, the on/off mechanical ventilation, the open/close windows, the on/off light, and other functions are all managed by a multi-agent unit or CMAS.

The state uses Bellman Eq. (8) to maximize the V^* value while accounting for energy savings and interior thermal comfort level, which stays within the setpoint. After every action, the instantaneous reward value is given back because of the state transition probability (p) to a certain site (s). Relative humidity (RH) and working temperature were included as independent factors for the value related to states in the reinforcement learning process because of their notable impact on the perception of interior thermal comfort. The relative humidity (RH) and internal operating temperature (IOAT) are also used as critical inputs to construct the projected mean value (PMV), which has been verified in compliance with ISO 7730 and ASHRAE standards [42–45]. By reducing the amount of cooling needed at reasonable outside temperatures, setting the state's setpoint value in accordance with the ASHRAE standard's lower and upper bounds makes it easier to use such a range. HVAC systems are occasionally turned off in the morning to save electricity. The current condition of its result is transited into a specific location following a probability that occurs for each action of $a \in d$ at A in the input. It seems clear that each CMAS agent has access to an iterative periodic time step (sample time slot t) in order to ascertain the best course to follow (policy) for every input parameter. As shown in Fig. 12, indoor suggested states that are continually changed by disturbance variables are used to develop the appropriate CMAS policies utilizing RL extended slots. The pseudocode procedures for agent-environment interactions along the state change potential are shown in Table 1 in order to acquire the optimum rule's tuple values.

3.2. Structure of deep clustering

As directed by the RL approach, after the rule is formed, the weights of the CMAS actions are updated automatically. Deep clustering is an approach to clustering where the clustering approach is expressed using multi-hidden layer neural networks. Network loss, deep neural network (DNN), and clustering loss are the three primary types of deep approaches for clustering. There is no distinct cluster organization in the

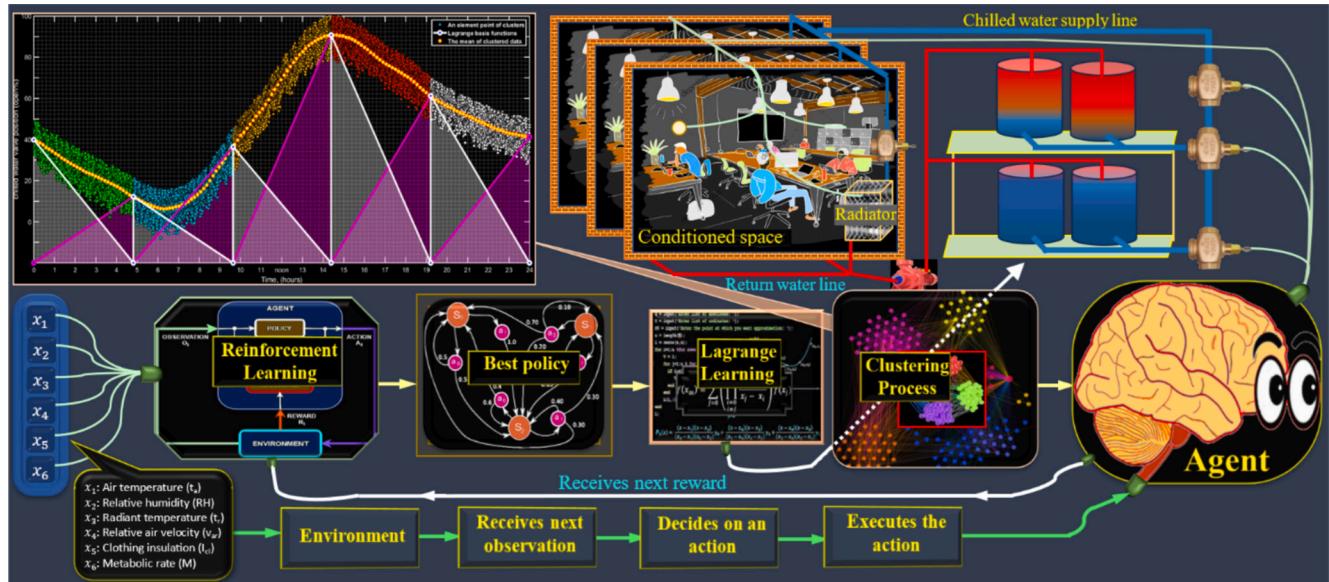


Fig. 12. The framework of the expanding view of the clustering approach in the deep reinforcement clustering for adaptive decision policy (DRCADP), based on the double-stage PCM.

Table 1

A pseudocode of learning algorithm of agent policy for chiller plant and PCM tanks.

The pseudocode steps of the policy-based RL algorithm
Model Inputs: The range of chilled water flow rate for pr and main cooling coil, the content of fresh, return and exhaust air dampers, open/close windows etc. The indoor set-point temperature is varied between 20°C to 24°C and RH between 45% to 55%.
Define model (s, a) for all $s \in S$ and $a \in A(s)$ and initialize functions V and π
For each iteration, Do
Iterate every environment step for disturbance factors
For environment setup, Do
$S \leftarrow$ current states (interior and exterior RH and temperatures)
$A \leftarrow$ Execute step (t) of action A for each agent in CMAS
Observes and receive the immediate reward (t) and execute the next state's $s(t+1)$ or S'
Model of the environment $(S, A) \leftarrow R, S'$ (tabular values)
Loop for iterate n times:
$S \leftarrow$ Store transitions of the detected state
$A \leftarrow$ Store earlier executed action in S
$R, S' \leftarrow$ Model of the environment (S, A)
End
End
Update and store the best agents' old actions for CMAS: $\pi_{old}^*(a s) = \pi^*(a s)$
End

dataset, according to the current study. As a result, hybrid DNN clustering makes the deep shape more effective. Even in the absence of supervision or prior knowledge of the rule, each action an agent does is categorized into a set of related data points. The c-means algorithm is used to locate the centres of the complex clusters that the agent has formed. Assigning each data point $\{Y_i^c\}$ into the characteristic space of a specific cluster is the basic c-means technique. The approach produces c cluster centres, every one of which corresponds to a cluster, by calculating the mean coordinate values from the locations allocated for every cluster. This study focuses on the clustering process that separates the data space $\{U\}$, $U = \{u_1, u_2, \dots, u_K\}$ into a group of clusters $\{C_i\}$, where $i = 1, 2, \dots, c$ and $[2 < c \leq K]$. The following set of theoretic equations

describes the clustering operation:

$$Y_{i=1}^c C_i = U \quad (12)$$

The inherent nonlinearity of each agent's policy, influenced by the dynamic environment, prevents representation by the Lagrangian formula. Consequently, this high nonlinearity necessitates segmenting the target policy into clusters. The weights of neural networks for each cluster derived from the Lagrangian formulation are structured into matrices according to dynamic memory cells, with each matrix (layer) representing a cluster load. The memory cell load layers are constructed and then arranged in an orderly manner using a nonlinear Lagrangian framework. The above layers are changed so that an agent action can

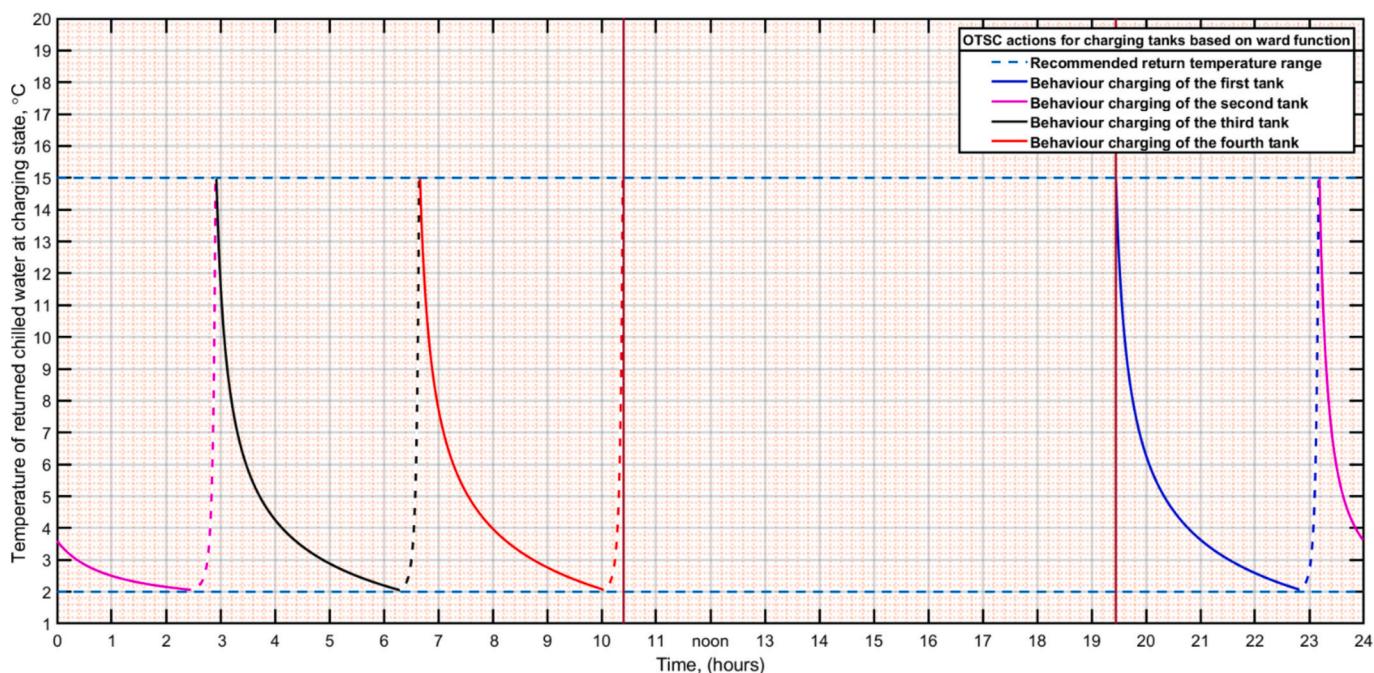


Fig. 13. The threshold point in the RL reward function of the agent action of OTSC, based on the double-stage PCM.

modify the input settings [46]. CMAS divides each agent action into five or more clusters based on how it is carried out. The offline code that sets the specifications in dynamic memory zones is based on base reactions to the trained RL methods of the CMAS.

The Quasi-Newton (QN) optimization method is employed to adjust the weights of the ANN derived from the Lagrangian formulation, ensuring optimal alignment with the agent's policy. Additionally, the QN algorithm is utilized to make appropriate decisions for CMAS, considering the multi-objective nature of the problem. An online reactive dynamic load adjustment is necessary to modify the action loads and achieve the desired condition. The online adaptive QN enhanced its response time to CMAS commands by employing forward tuning. It is possible to increase the structure of the weights layers to include additional weights, hence simplifying the analysis of huge, complex data sets. The ability to expand the three-dimensional structure of loads allows for high-correlation matrix-based independent variable addition. In addition to the three dimensions, time is a fourth dimension. Fig. 11's time-dependent heating and cooling loads suggest a link between the period and each input factor.

To elucidate the agent action of OTSC within CMAS, after segmenting its target policy into clusters and optimizing the weight of each cluster to fit with the action generated by RL. The action behavior is determined by the returned temperature of chilled water from PCM tanks, as depicted in Fig. 13, which serves as the threshold point in the RL reward function.

Each activity pattern of CMAS is divided into clusters to define the agent target rules. The Lagrangian equation is utilized in this method to direct each section in handling problems resulting from variations in cooling/heating loads within a designated degree of comfort range while guaranteeing that inside factors stay within the advised range. By extracting the weight matrix via the Lagrangian formula and aggregating the values using the multilayer perceptron (MLP) network, the agent seeks to reflect the parameters of the rule's characteristic. This makes it possible to express the relationships from a single tabulated load matrix of MLP of agent activities that define the agent operation. The Lagrangian weight box model can also be obtained by increasing the number of clusters [47]. The offline learning values were used to initialize the Lagrangian weight box model, which significantly

decreased the amount of time needed for online tuning. By starting the learning process from the parameter value rather than from zero, this was accomplished.

The air handling unit's input parameters, such as the location of the air absorbers (for both back and fresh air), the velocity of the air fan, and the state of the main and pre-cooling coil valves, were all adjusted by the DRCADP. Other building amenities were also adjusted, including the lighting systems' on-and-off cycles and the opening and closing of windows. These adjustments were made in order to improve indoor conditions in accordance with the suggested goals of attaining energy conservation and thermal comfort. As a result, the CMAS category includes the output classified by DRCADP. Division of all data points into ellipsoidal clusters employing the independent distance technique is the task assigned to each agent in CMAS. Finding the values of the rule vectors and normalizing them by scaling to the interval $[0, 1]$ are prerequisites for starting the clustering process. The preceding stages demonstrate how this validation process improves clustering efficiency.

$$Norm(x_i) = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (13)$$

where

x_{max} : is the maximum value throughout all data variables for each agent policy,

x_{min} : is the minimum value across all data values for each agent policy and

x_i : is a current data value in the policy.

3.3. Lagrangian formulation

In order to maximize long-term advantages and attain optimum criteria, the rules need a CMAS in the classroom. Each agent's action rule needs to be represented by the Lagrangian form once the optimal policies have been determined, as shown in Fig. 12. Conversely, there exists a highly nonlinear interaction among the guidelines. Their dynamic nature necessitates the application of the Lagrangian model due to their sensitivity to both internal and external conditions. The nonlinear Lagrangian estimate procedure can be carried out inside each cluster

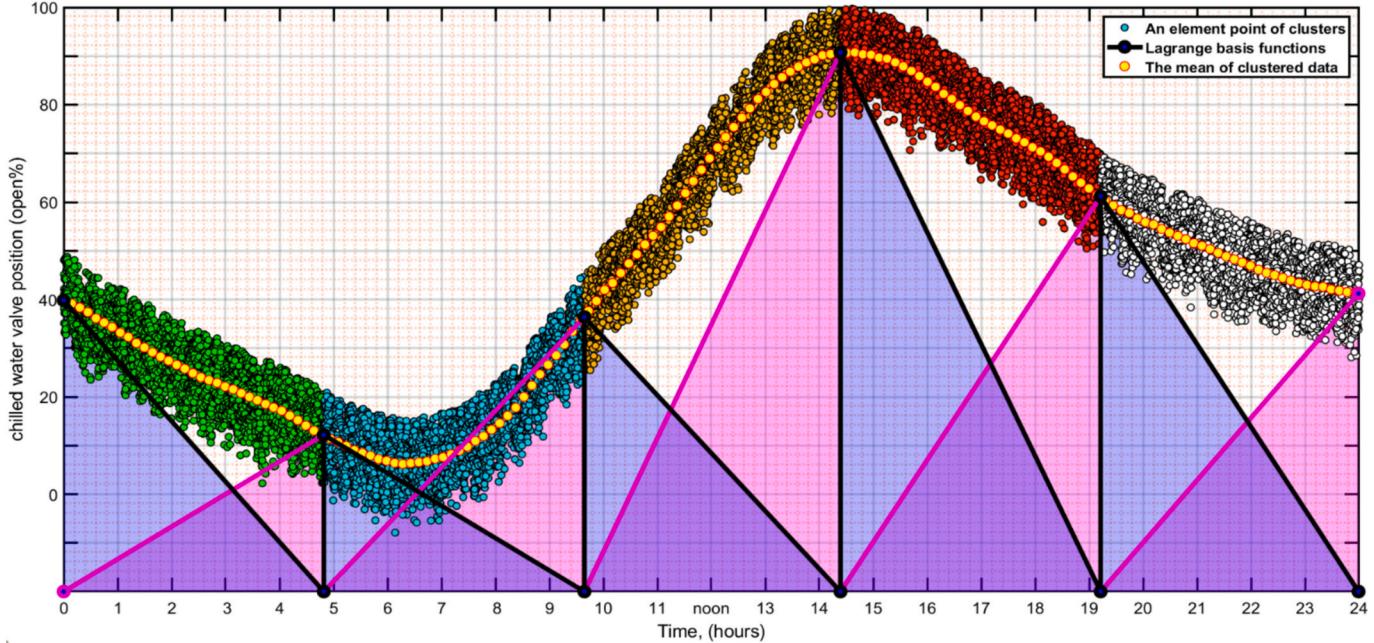


Fig. 14. How clustering the action policy of the agent controlling the chilled water flow rate valve position within the double-stage PCM establishes the local value of the basis Lagrangian function.

Table 2

The arrangement of the weight matrix is based on the number of rows (represented by samples of training) and the number of columns (represented by simulation elements).

Training weight	Simulated weight for predicted vector values			
	No. of iterations	$m = 1$	$m = \text{No. of points}$
	$i = 0$	$\omega_{0,1}$	\ddots	$\omega_{0,m}$
	\vdots	\vdots	\ddots	\vdots
	\vdots	\vdots		\vdots
	$i = k$ (No. of training sets)	$\omega_{k,1}$		$\omega_{k,m}$

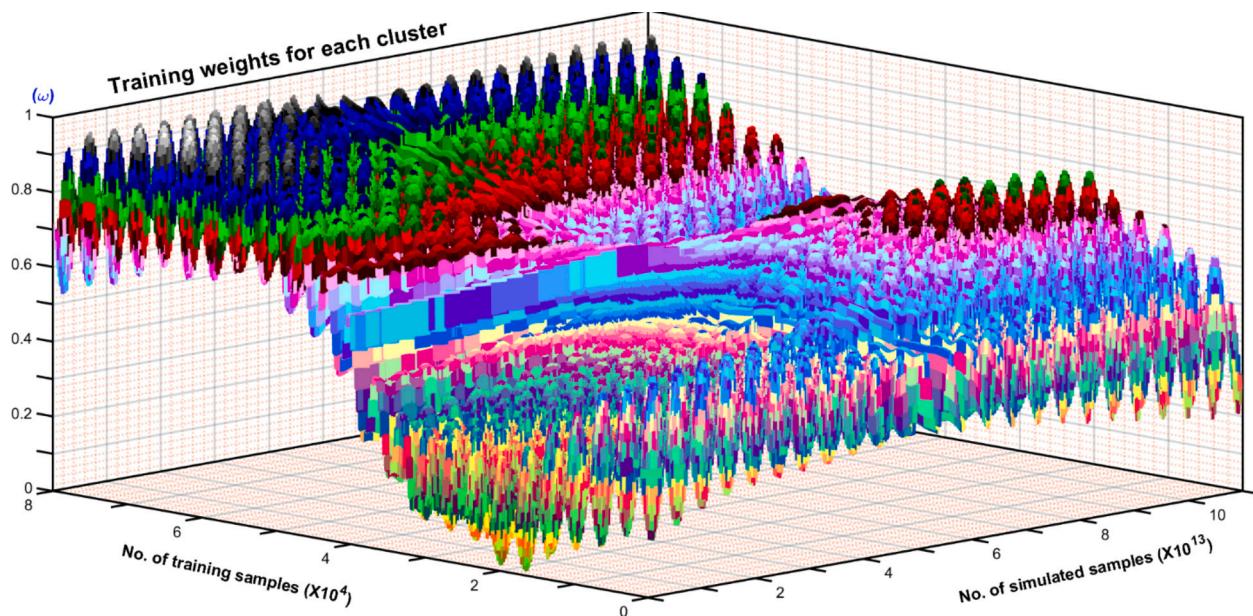


Fig. 15. Two dependent variable schemes (training and simulation) that elucidate the tabulated weights of each cluster within the Lagrangian matrix.

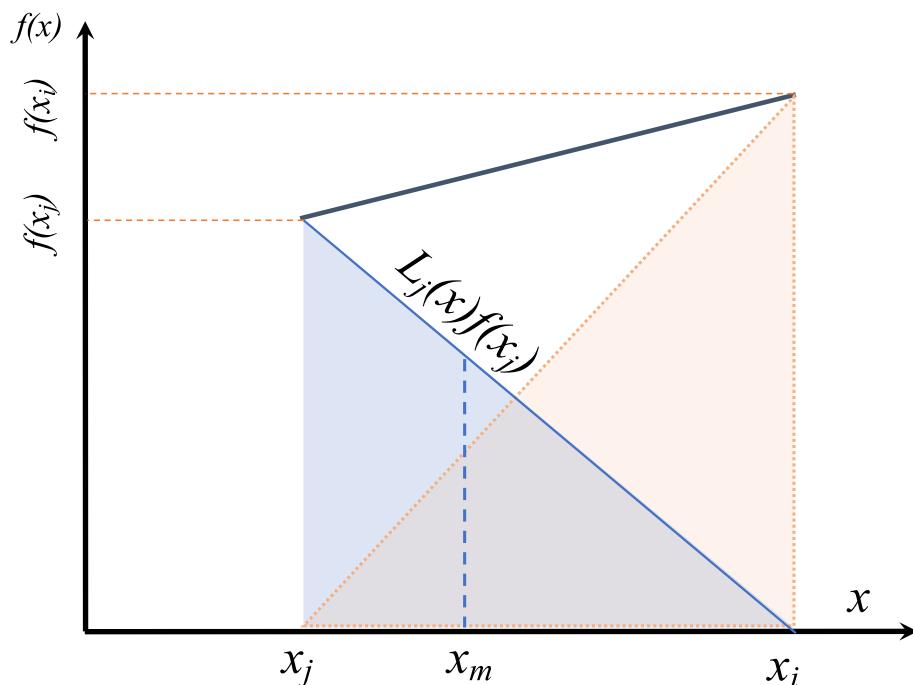


Fig. 16. The interpolation of the δ_{ji} value within the range of 0 to 1, representing the weight in the Lagrange polynomial equation.

using the following tools:

Each CMAS policy is achieved at individual quantities of an independent determined of $k + 1$ points $(x_0, y_0), (x_1, y_1), (x_2, y_2), \dots (x_k, y_k)$, then let us imagine these quantities on the scale of $L(x_j)$ such that $L_j(x) = \text{actions } (A_j), j = 0, 1, 2, \dots k$.

$$A[x_0, x_1, x_2, \dots x_k, x] = 0 \quad (14)$$

That is

$$\frac{f_0}{(x_0 - x_1) \dots (x_0 - x_k)(x_0 - x)} + \frac{f_k}{(x_k - x_0) \dots (x_k - x_{k-1})(x_k - x)} + \dots + \frac{f_x}{(x - x_0) \dots (x - x_k)} = 0 \quad (15)$$

$$f(x) = \frac{(x - x_0) \dots (x - x_k)}{(x_0 - x_1) \dots (x_0 - x_k)(x_0 - x)} f_0 + \dots + \frac{(x - x_0) \dots (x - x_{k-1})}{(x_k - x_0) \dots (x_k - x_{k-1})(x_k - x)} f_k \quad (16)$$

While a circuit of $k + 1$ points explores the operator $f(x)$, the nonlinear algebraic polynomial expression for the k^{th} degree Lagrangian, or the technique, is provided by formula (16).

$$f(x_m) = \sum_{j=0}^k L_j(x_i) f(x_j) \quad (17)$$

Formula (18) illustrates that the load part of formula (17) shows all normalized variables of the $k - 1$ and neglects the models at $j = i$.

$$L_j(x_i) = \prod_{i=0, i \neq j}^k \frac{x_m - x_i}{x_j - x_i} \quad (18)$$

Two elements of the Lagrangian concept cross one site and are zero at the other, as shown in Fig. 14. The network that the DRCADP algorithm produced was then weighted by a matrix and organized in compliance with Table 2's recommendations. The tabular representation demonstrates the systematic arrangement of weights attributed to each cluster. One of the two factors that determines the loads assigned to each cluster in the table is the practice cases. Fig. 15 provides more information on the relationship between the variables in question and the load displayed in the table. As illustrated in the graphic, the load for every

cluster is computed using both real-world and numerical situations. Using this strategy, the DRCADP algorithm may effectively aggregate and arrange relevant data to improve the matrix-weighted unit's overall utility.

When sum forms are used, the result can be expanded as the sum of the elements in formula (17), which is also referred to as Kronecker's delta notation. The Lagrangian hypothesis could be expressed as follows: when $m = I$, the result factors are $\frac{x_m - x_i}{x_j - x_i} = 0$, and when $m = j$, the product factors are $\frac{x_m - x_i}{x_j - x_i} = 1$.

$$L_j(x_i) = \delta_{ji} = \begin{cases} 1, & \text{if } m = j \\ 0, & \text{if } m = i \end{cases} \quad (19)$$

The quantity for $L_j(x_i)$ can be calculated using the size of the triangle in Fig. 16 if m is not equivalent to either i or j . The value of $L_j(x_i)$ is shown by the following Eq. (20).

$$L_j(x_i) = \delta_{ji} = \frac{x_m - x_i}{x_j - x_i} \quad (20)$$

The following illustrates the meaning of Kronecker's delta in the context of computing for the Lagrange evaluating polynomial:

$$L(x_i) = \sum_{j=0}^k y_j L_j(x_i) = \sum_{j=0}^k y_j \delta_{ji} = y_i \quad (21)$$

The following formula can be used to display the Lagrangian formula in its basic format:

$$f(x_m) = \sum_{j=0}^k \left(\prod_{i=0, i \neq j}^k \frac{x_m - x_i}{x_j - x_i} \right) f(x_j) = \omega_{n-1} \frac{f(x_{n-1})}{f(x_n)} + \omega_{n+1} \frac{f(x_{n+1})}{f(x_n)} \quad (22)$$

Eq. (22) states that the instantaneous weights (ω) are reflected in the DRCADP technique outputs through iterative learning. The Lagrange basis functions, which are presented in matrix form and based on the number of training samples for each cluster, are depicted in Fig. 14 as the weights signal.

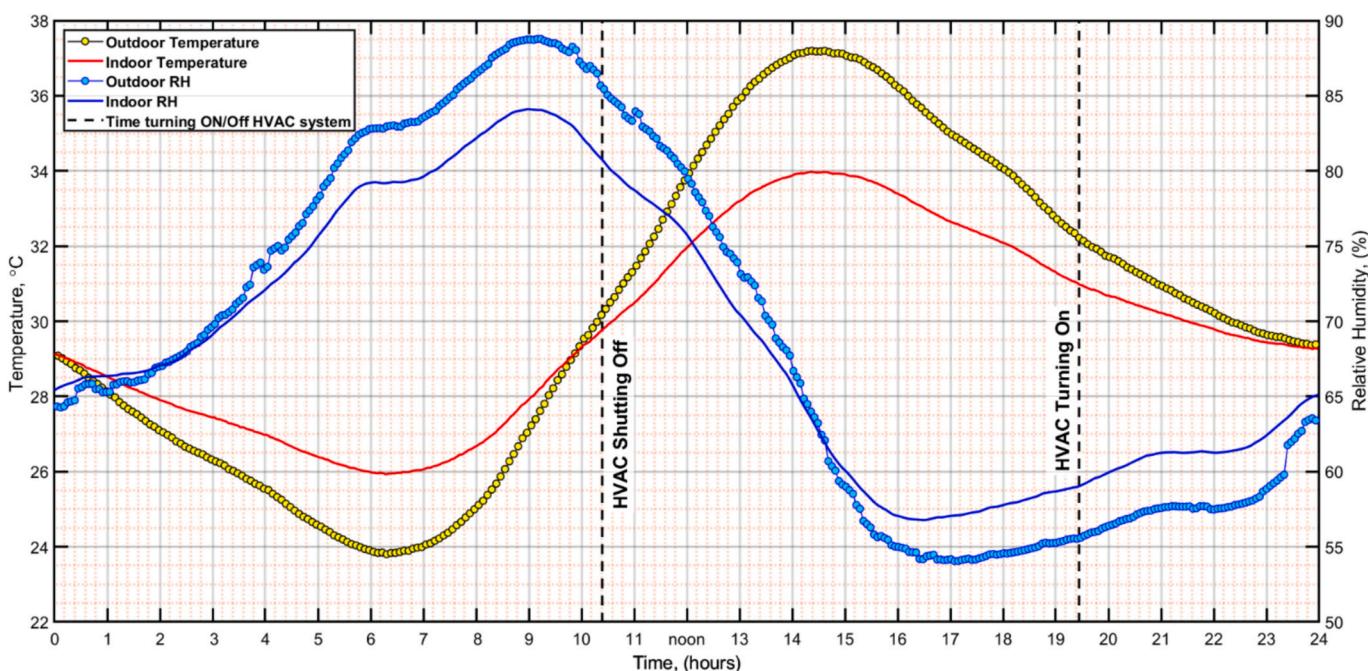


Fig. 17. The indoor and outdoor variation conditions over 24 h, utilized as a learning environment with the HVAC systems turned off.

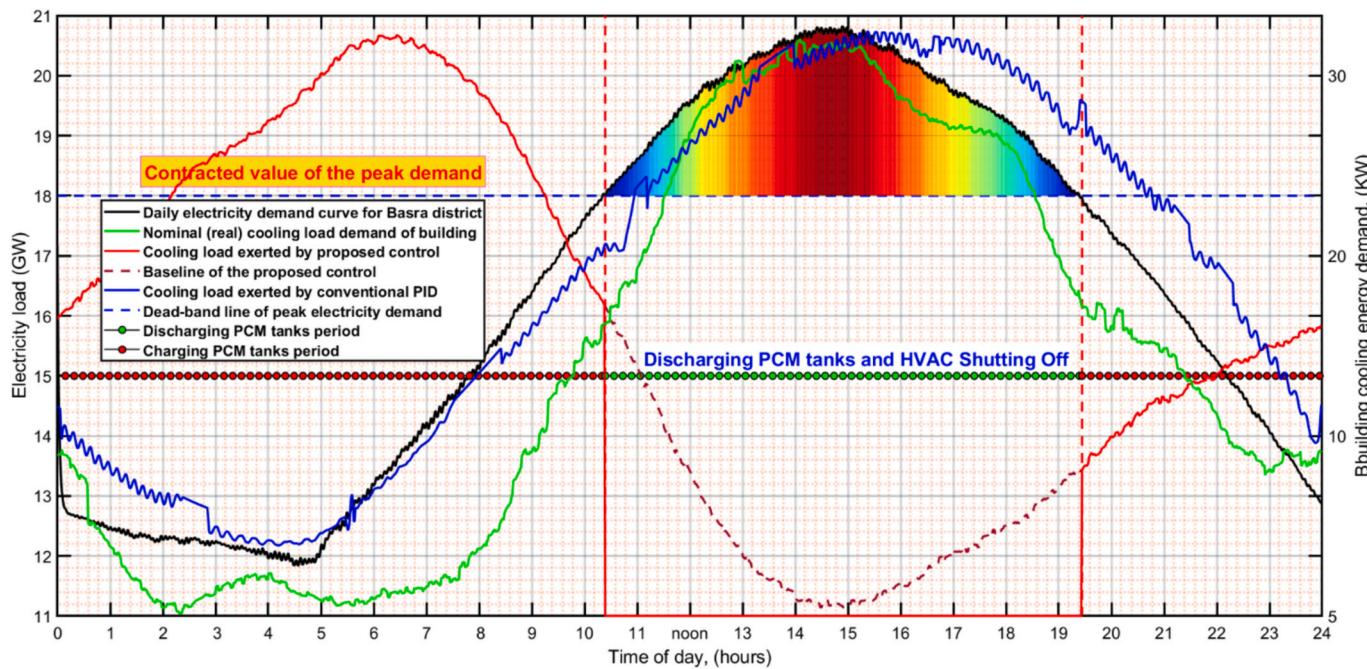


Fig. 18. The performance response of the multi-agent policy in manipulating controllable variables within the PCM tanks to track the cooling load's operating band.

4. Physical description for the DRCADP and evaluation criteria

4.1. The environmental conditions of test and learning

The building model is applied to a typical one-story residential house with a simple structure. The building has dimensions of 22.6 m in length, 11 m in width, and 4.5 m in height. The net area, excluding the garage, is 195.3 square meters. The gross area of windows and exposed walls is 126.2 square meters, while the net exterior wall area is 108.5 square meters. The overall house volume excluding garage volume is 468.7 m³. Based on observational parameters collected during an autumn day in

Basra City, MATLAB software was used to assess the effectiveness of the proposed controller. Fig. 17 delineates the behavioural occurrences transpiring both within and outside the architectural confines over the course of the day. The congruence in temperature and relative humidity patterns, observed both internally and externally when the chiller system is inactive, underscores the impact of building materials' thermal mass and wall insulation, colloquially referred to as the "thermal flywheel." This dynamic results in a delayed elevation of internal temperature relative to external conditions.

Between the hours of 5:00 and 8:00 in the morning, the nadir of temperature is reached for both indoor and outdoor locations,

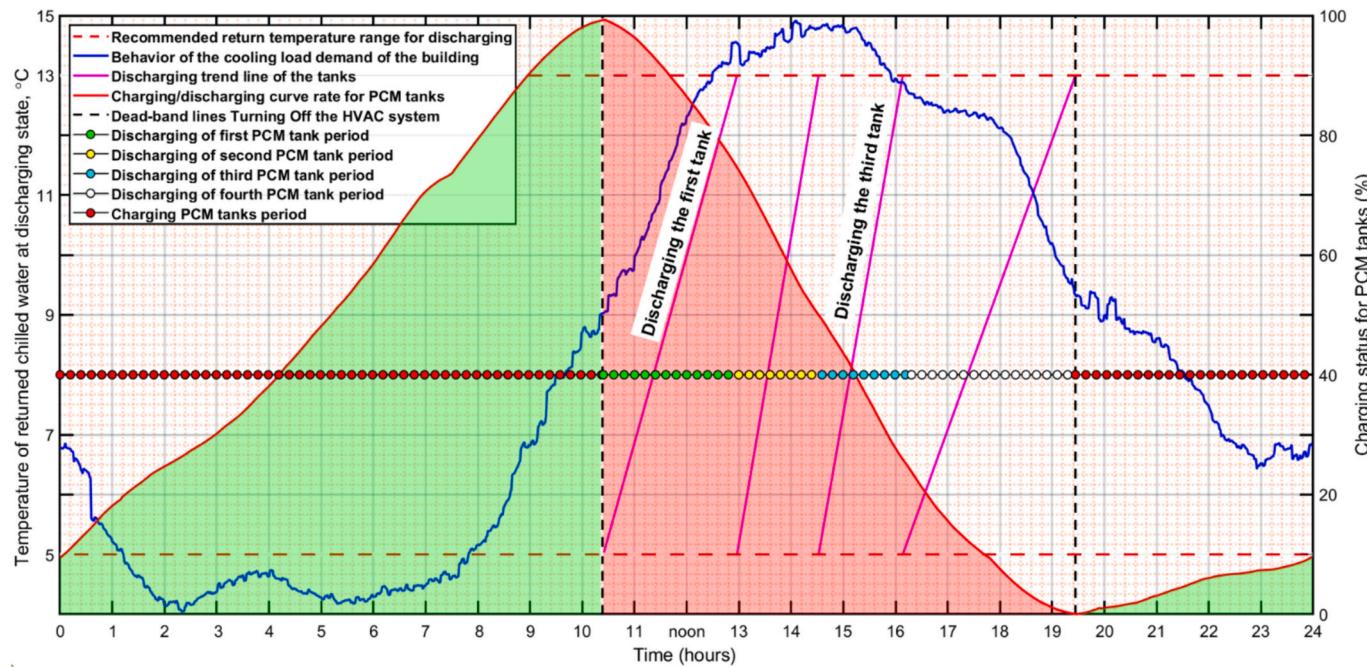


Fig. 19. The variation in the inclination angles of discharge lines between two recommended ranges of returned water temperatures from PCM tanks, as influenced by changes in the cooling load.

registering minimum values of 23.5 °C and 26 °C, respectively. Subsequent to this period, a gradual ascent in temperature prevails until it culminates at 14:30, manifesting as 37 °C externally and 34 °C internally. The pinnacle of humidity occurs around 9:00 in the morning, subsequently undergoing a decremental trend commencing at 16:00 to 18:00 until reaching its minimum.

The chiller units adhere to a programmed regimen, ceasing operations at 10:20 and recommencing at 19:20. This decision is contingent upon the presumption that, within the stipulated duration, the cooling load persists above the recommended deadband threshold. During intervals when the chillers are inactive, electrical consumption is nonexistent, and the provision of chilled water by both PCM tanks serves to sustain a thermally comfortable interior.

Fig. 18 demonstrates that the electrical demand in the Basra district follows typical semi-tropical patterns, peaking between 10:20 and 19:20. The figure also visually elucidates the performance of the multi-agent policy in adapting to variable controllable parameters within PCM tanks, which is crucial for monitoring the cooling load's operational range. Each point on the solid red line represents a unique chiller tank and PCM layout, highlighting the tailored configurations for specific scenarios. The challenge of achieving optimal interior thermal comfort within prescribed limits arises due to the restricted range of feasible shifts in peak cooling loads. This limitation, inherent due to its heat capacity characteristics, necessitates the assignment of these PCM tanks to the specific building. As shown in **Fig. 19**, the discharge line angles are restricted for this building. Furthermore, the PID controller is observed to extend the nominal peak load period, prompting the incorporation of recommended PCM tanks, as illustrated in **Fig. 18**, to mitigate the peak load and alleviate electric peak demand effectively. This integrated analysis underscores the intricate interplay of variables and strategies in managing the electrical demand dynamics of the Basra district.

The adaptability to adjust the return chilled water temperature to accommodate small-scale PCM tanks is apparent from the unique slope discharge lines displayed by each tank. This characteristic is determined by the tanks' capacities and immediate cooling requirements, as illustrated by the blue line in **Fig. 19**. An extensive analysis of the upper and lower limits for melting and solidification (10:20 to 19:20) is conducted to ensure sufficient coverage during periods of high demand. The diagram in **Fig. 19** demonstrates the precise alignment of PCM tanks with the deadband lines, confirming their effectiveness. These graphics provide significant insights into the subtle differences in cooling load requirements across a wide range of operational circumstances. **Fig. 19** presents a detailed representation of data by examining the water temperatures in each tank at various time intervals. This allows for a comprehensive understanding of the factors influencing the highest needs for cooling load. This analytical methodology reveals trends and patterns that are essential for assessing the effectiveness of differently ordered PCM tanks in satisfying specific cooling needs. The water temperatures that are returned serve as measurable indicators for the sequencing of PCM tanks. They provide a quantitative measure of cooling load requirements when there is a high demand for electricity. This systematic examination of fluctuations adds to the improvement of the PCM tank system, guaranteeing efficient energy usage and maintaining a thermally pleasant interior atmosphere.

4.2. Reward setting as bang–bang (on/off)

For producing the output advantage in the context of reinforcement learning (RL), the current work employs a binary (on and off) strategy. "Chiller sequencing control" is the term used to describe the current approach of chiller sequencing control. The main goal of using a specific level value trigger is to ascertain the best mode. When the indoor scenario reaches a threshold value, the working unit of the Bang-Bang agent switches between different PCMs tank (each tank has dual PCMs) modes. The agent runs in mode 1 and uses the first PCM tank originally. The agent switches to mode 2, which uses the second PCM tank to give cold

water, when the threshold value is exceeded. The agent will transition to mode 3 (using a third tank) and finally mode 4 (using a fourth tank) if the threshold value rises. The dead band's mathematical equation describes the operational part of the bang-bang agent when its sensitivity is zero. After passing through PCM1, which has a low critical temperature, the chilled water from the chiller travels through PCM2, which has a considerably higher critical point.

The bang-bang agent is a reinforcement learning agent that learns how to control a unit by choosing between two different actions: activation and deactivation. To control the interior temperature and keep it within a preset range, the bang-bang agent may learn to alternate between tanks one, two, three, and four in the PCM tank sequence management arrangement. An incentive system that pushes the agent to switch to the proper mode when the interior temperature rises over a certain threshold is what trains the bang-bang agent. To maximize the reward function, the bang-bang agent makes mistakes and learns from them. As the agent learns, it becomes more adept at switching to the optimal mode as necessary, which reduces energy usage and boosts enjoyment. The operating rate of the dead band is important because it can affect how energy-efficient the PCM tank sequencing process is. It is anticipated that energy consumption will drop with an increase in the active fraction of the dead band. However, it can also result in reduced comfort indoors. The dead band working rate is an estimate of the frequency at which the bang-bang agent switches between states.

$$\rho_1 = \frac{t_{md1}}{t_{md1} + t_{md2} + t_{md3} + t_{md4}}, \rho_2 = \frac{t_{md2}}{t_{md1} + t_{md2} + t_{md3} + t_{md4}}, \rho_3 = \frac{t_{md3}}{t_{md1} + t_{md2} + t_{md3} + t_{md4}} \text{ and } \rho_4 = \frac{t_{md4}}{t_{md1} + t_{md2} + t_{md3} + t_{md4}} \quad (23)$$

where ρ is the threshold value used to initiate the sequential action for tanks, $t_{md1} = \frac{1}{r} \ln \frac{T_{inset} - \frac{1}{2} T_{hys} - T_{md1}}{T_{inset} + \frac{1}{2} T_{hys} - T_{md1}}$, $t_{md2} = \frac{1}{r} \ln \frac{T_{inset} - \frac{1}{2} T_{hys} - T_{md2}}{T_{inset} + \frac{1}{2} T_{hys} - T_{md2}}$, $t_{md3} = \frac{1}{r} \ln \frac{T_{inset} - \frac{1}{2} T_{hys} - T_{md3}}{T_{inset} + \frac{1}{2} T_{hys} - T_{md3}}$ and $t_{md4} = \frac{1}{r} \ln \frac{T_{inset} - \frac{1}{2} T_{hys} - T_{md4}}{T_{inset} + \frac{1}{2} T_{hys} - T_{md4}}$

The pace at which the temperature of a given room will change if the actions that control its temperature are stopped is indicated by the time constant (r) for the decay rate of the interior temperature. The optimal temperature for the shared area is represented by the temperature set point (T_{set}). The temperature range that needs to be exceeded when turning on cooling or heating equipment is known as the dead band or resonance of the bang-bang. When the heating or cooling unit is operating in mode 1, the room's equilibrium temperature is represented by the steady-state temperature during mode 1 activation (T_{md1}). The space will get closer to thermal equilibrium (T_{md2}) when the heating or cooling mechanism runs in mode 2 at a consistent temperature throughout the mode 2 operation. The proportion or rate at which the heating or cooling system is operated is known as the duty cycle.

The duty cycle is important since it could affect how energy-efficient the heating or cooling system is. Energy usage and duty cycle increases are positively connected. As a result, it is essential to choose a duty cycle that is high enough to maintain the desired temperature and uses the least amount of energy.

5. Results and discussion

This study utilized a strategic sequencing methodology to maximize the COP values of the Multiple TES system, hence ensuring the optimal operation of the chillers. The excess cooled water, which was not immediately needed, displaced the warmer water in the PCMs tanks. In contrast, if the cooling load of the chiller surpassed the peak load deadband line, chilled water was extracted from the PCM tanks in order to satisfy the necessary flow rate for the air handling unit (AHU). Furthermore, a new methodology was utilized during the charging process of the PCM tanks to optimize the COP of the chillers. This entailed employing the suggested technique, deep reinforcement clustering for adaptive decision policy (DRCADP), deliberately leveraging

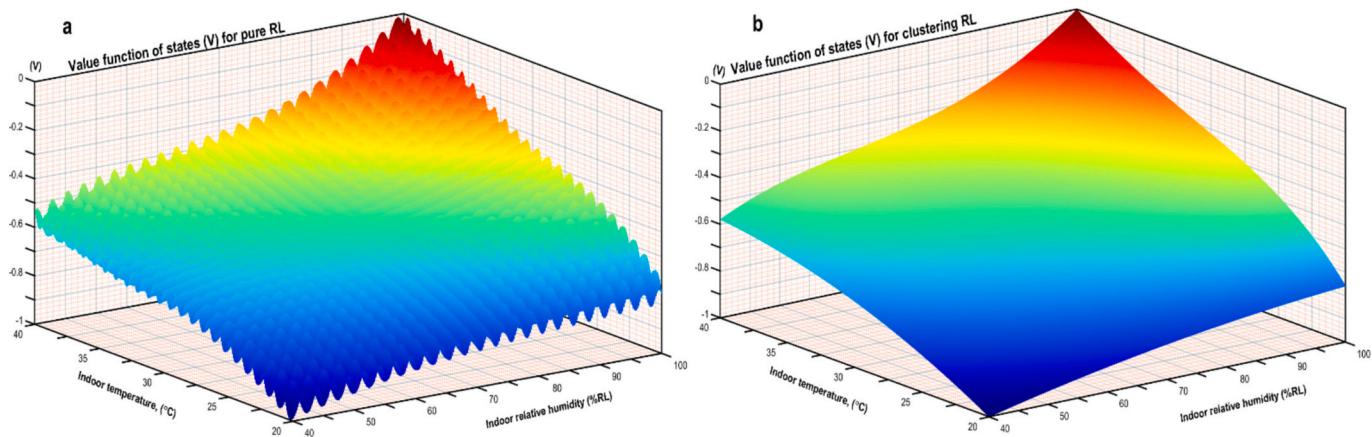


Fig. 20. The impact of episode count (discretization) on the optimal state's value function. (a) The surface of the optimal value function for ($N_t = NRH = 26$, $N_v = 10$), (b) The surface of the optimal value function for ($N_t = NRH = 58$, $N_v = 10$) episodes.

the cooler temperatures during the evening. This novel approach guarantees the smooth integration of system elements, hence improving overall effectiveness and productivity. Each storage tank has two PCMs, which improve the efficiency of both the charging and discharging processes. This enhancement is mostly justified by the second PCM's greater heat capacity. The use of two unique PCMs within each tank strategically capitalizes on the second PCM's superior thermal properties, contributing to increased efficacy in the overall thermal energy storage system. This strategy is consistent with the goal of enhancing system performance by exploiting the complementing characteristics of the selected PCMs to enhance the efficiency of the energy transfer processes during both the charging and discharging phases.

5.1. Performance analysis of the DRCADP

Two different discretisation levels were considered to analyze the impact of discretization of continuous states and actions on reinforcement learning. First, the outputs of the system (indoor temperature and

relative humidity) and input of the system (valve position) are discretized into 26 and 10 steps each ($N_t = 26$, $N_{RH} = 26$, $N_v = 10$). This gives 676 possible states, and 10 possible actions for the MDP, as illustrated in Fig. 9. In the second case, the output levels are discretized into 58 steps and the input is discretized into 10 steps ($N_t = 58$, $N_{RH} = 58$, $N_v = 10$), which provides 3364 possible states and 10 possible actions for the MDP. Then, in the second case, the input (agent policy) was represented by deep clustering based on the nonlinear Lagrangian framework. To assess the impact of these two distinct scenarios, the optimal value function's surface will serve as the criterion for evaluating the proposed CMARLDC structure and obtaining actionable feedback based on surface smoothness quality.

Fig. 20 shows the value function plays a crucial role in calculating the value (V) of a state and guiding the implementation of bang-bang actions in the context of reinforcement learning (RL).

The comparison between episodes ($N_t = N_{RH} = 26$) and ($N_t = N_{RH} = 58$) is primarily based on the optimal value error. During software operation, the error significantly increases until ($N_t = N_{RH} = 26$), after

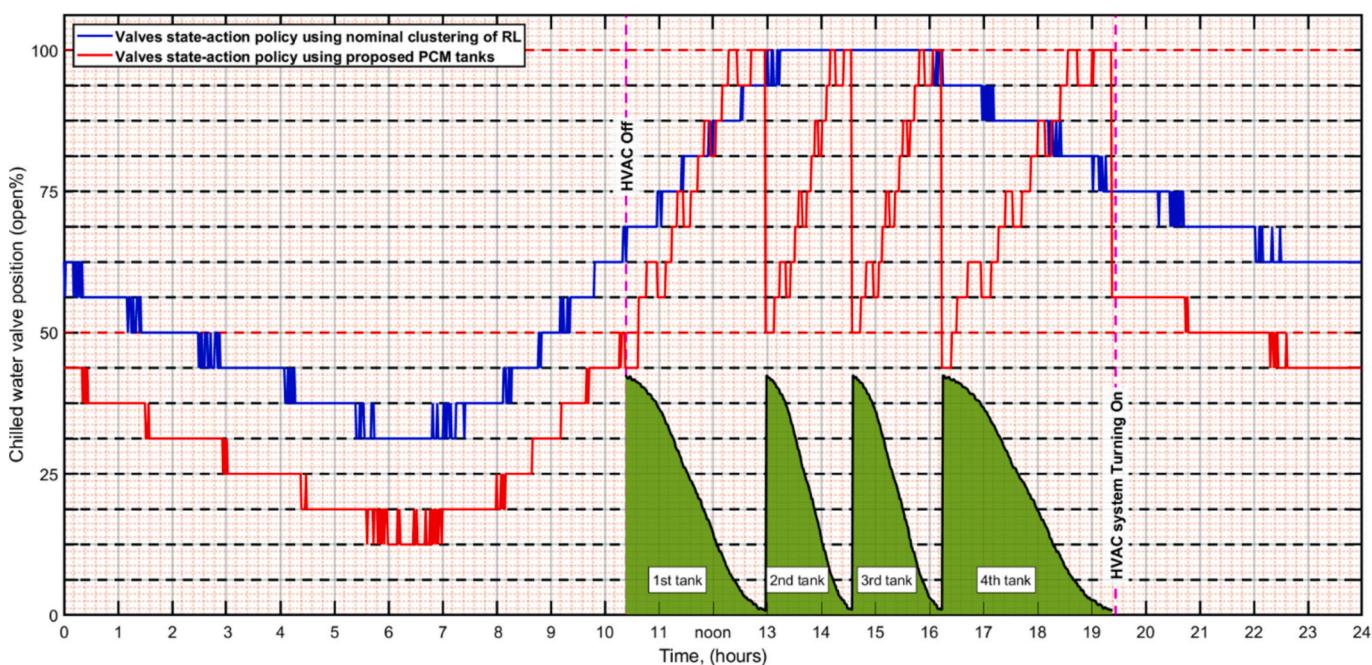


Fig. 21. The optimal actions (policies) of two cooperative agents managing optimal tank sequencing control (OTSC) and regulating the flow rate of chilled water into PCM tanks via valve position. The action response profiles of both agents correspond to the cooling load demand.

which it gradually decreases, though it remains considerable until ($N_t = N_{RH} = 58$). Beyond episode 53, the error becomes negligible, making further episodes unnecessary for error reduction. Balancing accuracy and computational efficiency, a higher number of episodes would unduly extend software operation times. Therefore, the optimal number of episodes is determined to be 58.

Fig. 20a, produced for the first case, displays a prominently uneven surface caused by the restriction because of the high nonlinear dataset, particularly noticeable in the z-axis, indicating the value function of state and the x and y axes showing interior temperature and humidity, respectively. Whereas **Fig. 20b**, which corresponds to the second case, displays a more even surface, making it easier to compare the impact of different episode numbers on the value function based on indoor temperature, and relative humidity. A comparison of **Figs. 20a** and **19b** shows that the finer details of the optimal value function are lost in the proposed CMARLDC value approximation approach, thus leading to the proposed surface of the optimal value function being much smoother than pure RL. The RL system's susceptibility to discretisation modifications is emphasized, as alterations in discretisation levels (N) result in a range of value function profiles that affect the value of criteria. This highlights the need to carefully choose the number of episodes in reinforcement learning applications to maximize system efficiency and performance.

This indicated that the value function is sensitive to changes in discretization (N); by increasing iteration or the episode (N) gets more missing values and improves the overall performance of the RL system. However, increasing (N) leads to increasing time in learning so that, the optimal N was fined around 58.

Fig. 21 illustrates the strategic measures used by two cooperative agents involved in optimal tank sequencing control (OTSC) and the regulation of chilled water flow rates into PCMs tanks through valve modifications. The occurrence of these measures is dependent on the cooling load demand and takes place between 10:30 and 19:30. The initial agent coordinates the placement of PCM tanks, while the second agent, working together, modifies valve locations to control the rates of chilled water flow. The figure presented offers essential information regarding the step-by-step functioning of tanks and emphasizes the significant influence of building load and environmental factors on

discharge rates. Utilizing two PCMs in each tank increases the efficiency of the charging/discharging processes since the second PCM has a high heat capacity. The decision to commence the initial PCM tank discharge at 10:20, coordinated with the shutdown of HVAC systems, is supported by the rising external temperature that requires cooling interventions. The discharge patterns that follow are carefully scheduled to match the changing cooling requirements. This is demonstrated by the second tank's faster discharge around 13:00, which coincides with the highest occupancy of the building and increased cooling demand. The discharge processes of the third and fourth tanks, which began at 14:40 and 16:15, respectively, illustrate the adaptive utilization to address varying cooling requirements. This discovery confirms the efficient utilization of tanks for maintaining steady cooling, even during periods of high demand. The figure outlines a detailed operational plan that guarantees a continuous supply of chilled water, hence improving the overall efficiency of the cooling system. The coordination of the HVAC system shutdown with tank operations from 10:20 to 19:20 provides evidence for the justification of synchronizing these procedures, ensuring the building's cooling needs are met. In addition, strategically adjusting the placements of chilled water valves during tank operations enhances energy efficiency by utilizing colder water at the bottom of the tank and maximizing the Coefficient of Performance (COP). This detailed operational plan enhances the total energy efficiency of the system and maximizes COP, hence improving the sustainability of the cooling process.

5.2. Comparison of indoor conditions with nominal RL and conventional PID

The previous comparison of **Fig. 21** shows that the control actions taken by the proposed controller are more active than the nominal control performance, due to its agent acting in a wide range with the clustering RL control strategy. This is reflected in the proposed surface of the optimal value function which is much smoother than the nominal control pure RL. **Fig. 22**, provides a comprehensive evaluation of indoor temperature regulation by comparing three different types of control systems. Key elements include the proposed dead-band interval, ranging from 10:20 to 19:20, along with corresponding indoor and outdoor

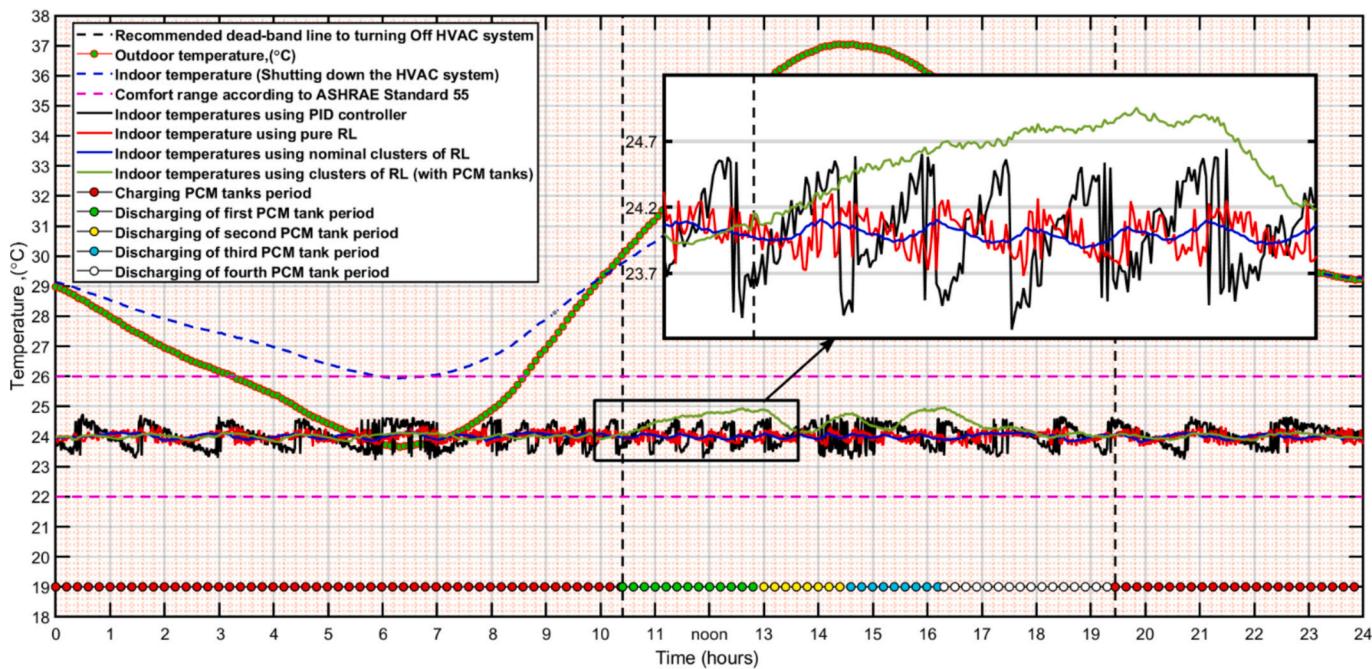


Fig. 22. The time sequencing control of the charge/discharge tanks and compares indoor profile conditions for three different control types based on temperature evaluation.

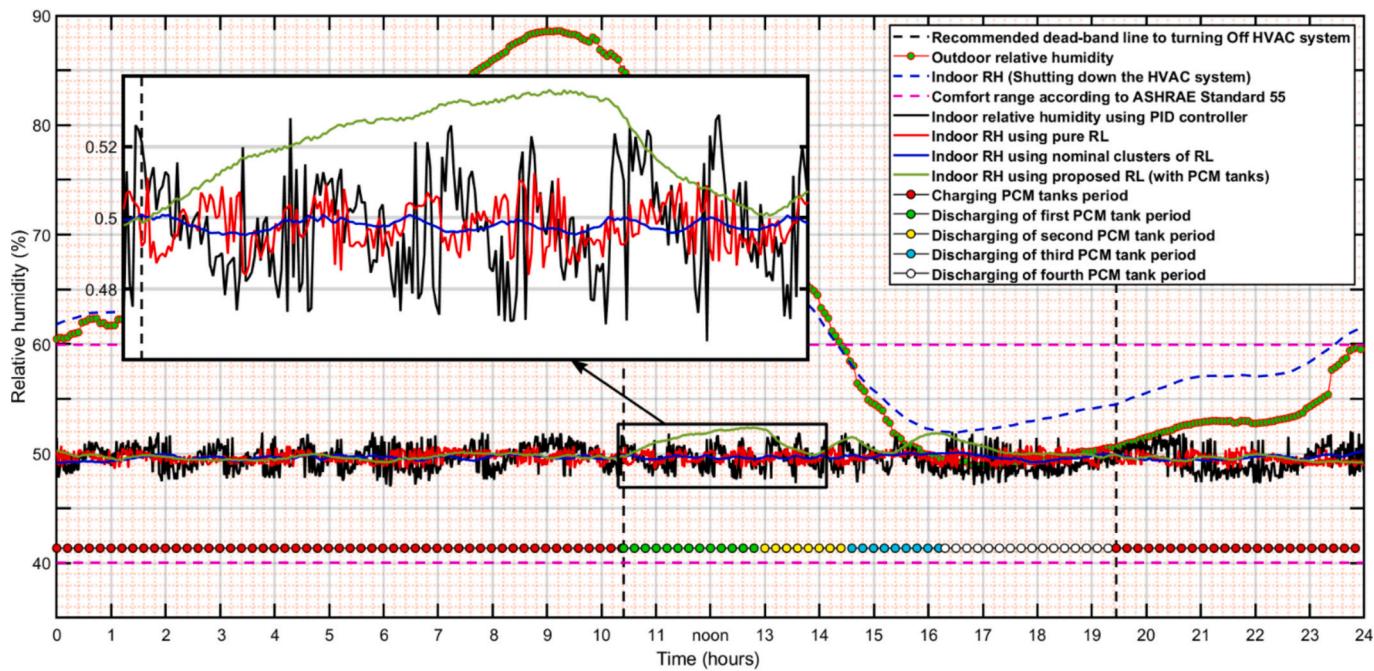


Fig. 23. The time sequencing control of the charge/discharge tanks and compares indoor profile conditions for three different control types based on relative humidity (RH) evaluation.

temperature data. At approximately 10:20, when the dead-band line is initially established, both indoor and outdoor temperatures are recorded at 30 °C. Subsequently, temperatures gradually rise, peaking around 14:30 at 37 °C indoors and 33 °C outdoors. The diagram demonstrates the effects of three control strategies based on the same setpoints: the PID controller, the proposed reinforcement learning (RL) strategy with PCM tanks, and the RL agent under normal conditions. This analysis demonstrates the impact of various strategies on building temperature profiles during HVAC operation and setpoints. According to ASHRAE Standard 55, the comfort range is between 22 °C and 26 °C. The RL

approach with PCM tanks is identified as the most efficient control scenario, exhibiting minimal temperature fluctuation within a narrow range of recommended indoor temperatures. Also, Fig. 22 clearly illustrates the superior efficacy of the proposed approach in ensuring reliable and efficient interior temperature control, particularly when incorporating PCM tanks. The adaptability and comprehensive evaluation of multiple elements in the proposed technique contribute to its optimization in aligning indoor temperature with the desired setpoint. It is crucial to note the system's charging cycle, which begins at 19:20 and concludes at 10:20 the following day, followed by a predefined sequence

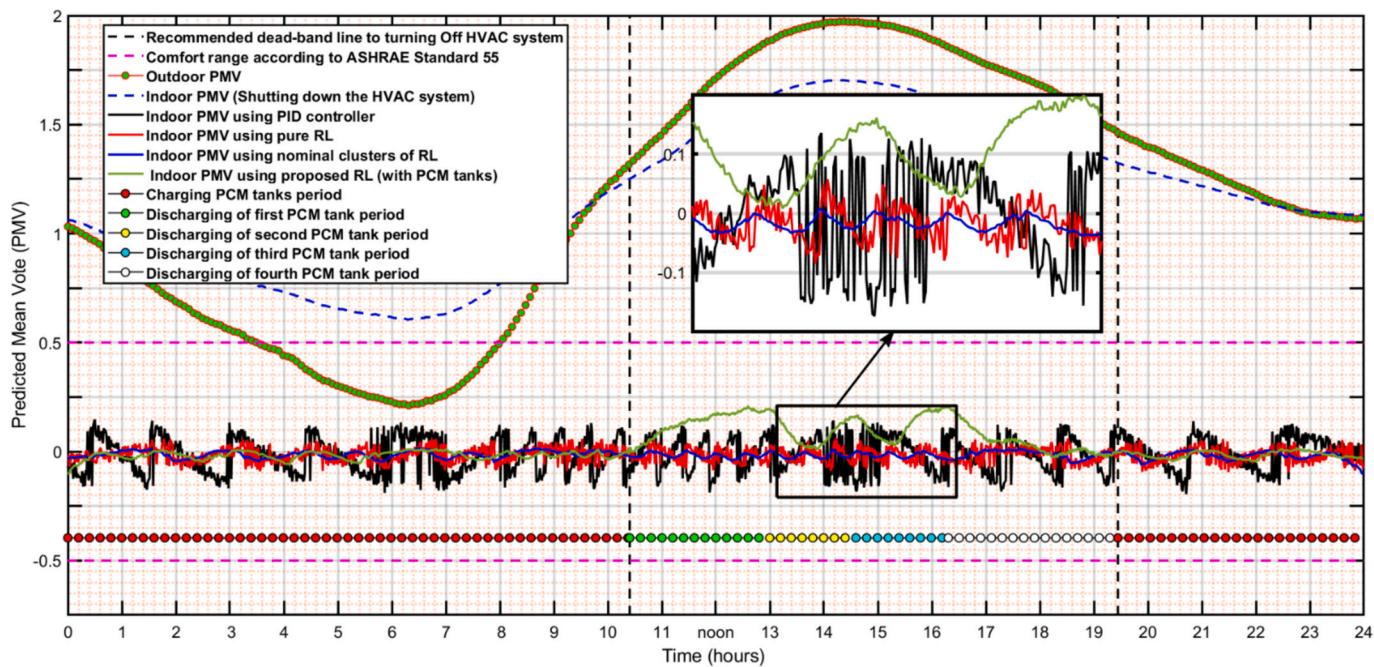


Fig. 24. The time sequencing control of the charge/discharge tanks and compares indoor profile conditions for three different control types based on Predicted Mean Vote (PMV) evaluation.

for tank discharge. The use of two PCMs per tank enhances the efficiency of both charging and discharging due to the substantial heat capacity of the PCM2.

When the HVAC systems and PCM tanks shut down, the indoor temperature peaks at 14:30, coinciding with the maximum cooling load. Consequently, the angle of the PCM tanks' discharging line reaches its maximum value at this time, as illustrated in Figs. 17, 18 and 19, to maintain indoor conditions within the recommended parameters. The proposed structure (the RL approach with PCM tanks) enables the action response of both agents to correspond to the cooling load demand (maintains the indoor conditions within the ASHRAE comfort range). Consequently, the main challenge encountered by the proposed CMARLDC arises at this pivotal moment, where the CMARLDC demonstrates high performance, as demonstrated in Fig. 22. In summary, the results show the proposed strategy maintains greater stability within the recommended range when compared to other types, which exhibit fluctuations within this range.

Fig. 23 clearly illustrates the chronological control of charge/discharge tanks and compares indoor relative humidity (RH) profiles across three control methods. It emphasizes the effectiveness of the reinforcement learning (RL) approach incorporating PCM tanks. The graphic shows humidity levels inside and outside, along with target values between 40 % and 60 %, consistent with ASHRAE Standard 55 comfort guidelines. Integrating four PCM tanks into the RL algorithm is demonstrated as the most efficient method for forecasting and controlling relative humidity.

Fig. 23 clearly illustrates the superior efficacy of the analyzed strategy in accurately predicting and maintaining the desired relative humidity within the prescribed range, as compared to other control methods. The RL technique efficiently manages the charging and discharging cycles, strategically deploying PCM tanks to attain a relative humidity level of 41 %. The deliberate utilization of tanks highlights a methodical approach to preserving ideal humidity levels in the regulated setting. The image also provides crucial data on variations in ambient humidity levels during the assessment period. Significantly, the highest level of external humidity, reaching 89 %, is observed around 9:00; however, it decreases to 50 % during the period of inactivity of the HVAC system known as the dead-band line. These observations

emphasize the ever-changing nature of external humidity, requiring strong management measures to effectively maintain internal relative humidity within specific limits. The comprehensive analysis of the three control methods depicted in Fig. 23 confirms that the RL strategy utilizing PCM tanks surpasses other approaches in effectively predicting and managing relative humidity. Achieving the desired indoor relative humidity levels necessitates the charging and discharging of tanks at a humidity level of 41 %. The figures supplied offer useful insights into the changes in humidity, highlighting the importance of efficient management systems in maintaining optimal conditions indoors.

When strategically developing and managing outdoor spaces, especially in chronically cold climates, it is crucial to consider outdoor thermal comfort. The Predicted Mean Vote (PMV) index is used to objectively quantify human thermal comfort, which is a crucial characteristic in this context. The PMV index considers multiple climate parameters, such as air temperature, relative humidity, wind speed, and sun radiation. The PMV scale measures the perceived thermal comfort in a space. A score of -3 suggests a very cold environment, while a score of +3 indicates a very hot environment. A score of 0 represents a neutral or comfortable environment. The PMV index is commonly used in indoor settings to define a comfortable thermal experience, with a range of -0.5 to +0.5 as the desired set points. Fig. 24 illustrates the PMV values in a detailed manner for both indoor and outdoor settings, taking into account various management techniques implemented throughout the day. Significantly, at approximately 14:30, the highest PMV value of 2 in the outdoor setting is situated within the range of thermal comfort. An extensive evaluation of three management techniques (PID control, RL agent action, and the recommended RL approach with PCM tanks) demonstrates the effectiveness of each strategy in impacting the PMV value.

When it comes to maintaining a comfortable temperature and following certain temperature guidelines, the recommended way is using RL with PCM tanks, which outperforms the other two methods (RL agent action and PID control). Because the second PCM has a high heat capacity, using two PCMs in each tank boosts the efficiency of the charging/discharging processes. This result highlights the benefits of the RL method, especially when integrating PCM tanks to control and enhance outdoor thermal comfort. Fig. 24 offers information about the

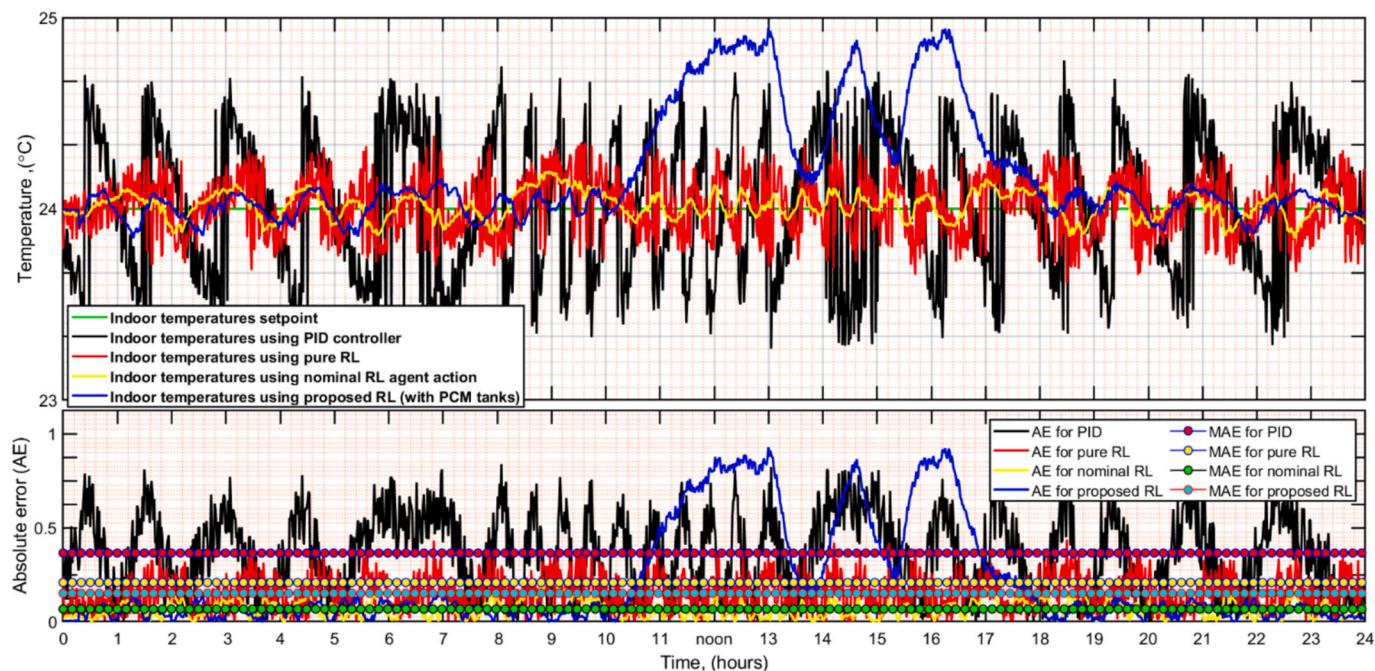


Fig. 25. The evaluation of the three controllers' performance using two primary statistical criteria: absolute error (AE) and mean absolute error (MAE), both related to the indoor temperature set point value.

Table 3

Comparison of the statistical indices of three different controllers according to indoor temperature response.

Indices	MxAE	MAE	RMSE	MAPE	RAE	r^2
PID	3.3978	0.2993	0.4893	0.0120	0.3062	0.9452
Nominal RL	3.6987	0.2204	0.5738	0.0114	0.2280	0.9571
DRCADP	2.4691	0.1297	0.4189	0.0068	0.1284	0.9831

outdoor PMV values between 1:30 and 8:00 in the morning. It shows that the outdoor PMV stays within the tolerable range (-0.5 to 0.5) currently, mainly due to the relatively low temperature. This empirical analysis strengthens the need to use the recommended RL technique with PCM tanks to provide efficient and optimum outdoor thermal comfort. It demonstrates that this technique is superior to other management options.

5.3. Statistical tools to compare different indices

An important discovery in this study is the notable clustering detected in Lagrangian trajectory curves, which visually represent the cooperative behavior displayed by agents. Fig. 25 presents two important statistical measures, specifically absolute error (AE) and mean absolute error (MAE), which are used to evaluate the effectiveness of three controllers in relation to indoor temperature setpoint values. The data presented in Fig. 25 indicates that the highest temperature (25 °C) occurs between 13:00 and 16:20. This coincides with the implementation of the reinforcement learning (RL) technique using PCM tanks. It is worth noting that this period corresponds to the peak outdoor temperature. This observation highlights the heating system's ability to maintain inside temperatures within the required range, even when the outdoor temperatures are higher than the internal temperatures. The replicated behavioural patterns of the RL model (with cooling tanks), PID control, and RL agent activity, as depicted in the figure, unexpectedly demonstrate significant resemblances. This demonstrates the effectiveness of the deep clustering approach in determining the optimal strategy for controlling the PCM tank system.

Notably, the suggested RL approach exhibits a maximum absolute error of 1, while the PID technique reaches a peak of 0.8 at different

instances during system activation, especially during the startup phase. The extended duration of the chilling process in the chiller, which is longer than expected, leads to a cooling load that exceeds the chiller's capability. This disparity is ascribed to the distinctive attributes of the chiller. The proposed RL technique with PCM tanks achieves a maximum mean absolute error of 0.1 during the day, which is regarded as acceptable. However, for situations including pure RL agent activity and PID control, the mean absolute error stands at 0.3 and 0.4, respectively. These inconsistencies underscore the difficulties in precisely forecasting and controlling the system's actions when the chiller is activated, underscoring the necessity for additional improvements in the control strategies to optimize performance and minimize mistakes.

Table 3 presents a comprehensive quantitative analysis of three controllers - PID, Nominal RL, and DRCADP - using important statistical measures to assess their effectiveness in regulating indoor temperature. DRCADP demonstrates superior performance over both PID and Nominal RL in terms of Maximum Absolute Error (MxAE), with a value of 2.4691. This represents a 27.3 % improvement compared to PID (3.3978) and a 33.4 % improvement compared to Nominal RL (3.6987). DRCADP obtains the lowest Mean Absolute Error (MAE) value at 0.1297, which represents a significant reduction of 56.7 % compared to PID (0.2993) and a 41.1 % reduction compared to Nominal RL (0.2204). Regarding the Root Mean Square Error (RMSE), DRCADP outperforms once again, showing a 14.3 % enhancement compared to PID and a 27.1 % enhancement compared to Nominal RL. When considering the Mean Absolute Percentage Error (MAPE), DRCADP shows a significant decrease of 43.3 % compared to PID and a 40.0 % decrease compared to Nominal RL, with values of 0.0068. DRCADP demonstrates a 58.0 % enhancement compared to PID and a 43.7 % improvement compared to Nominal RL, as measured by the Relative Absolute Error (RAE). Regarding r^2 , DRCADP achieves the greatest value of 0.9831, which corresponds to a 3.7 % enhancement compared to PID and a 2.7 % improvement compared to Nominal RL. The results provide quantitative evidence that DRCADP routinely surpasses PID and Nominal RL in all indices, showcasing its greater accuracy, precision, and overall efficacy in predicting interior temperature. The lower values of MxAE, MAE, RMSE, MAPE, RAE, and the better r^2 for DRCADP demonstrate its superior effectiveness in reducing prediction errors and delivering accurate interior temperature forecasts.

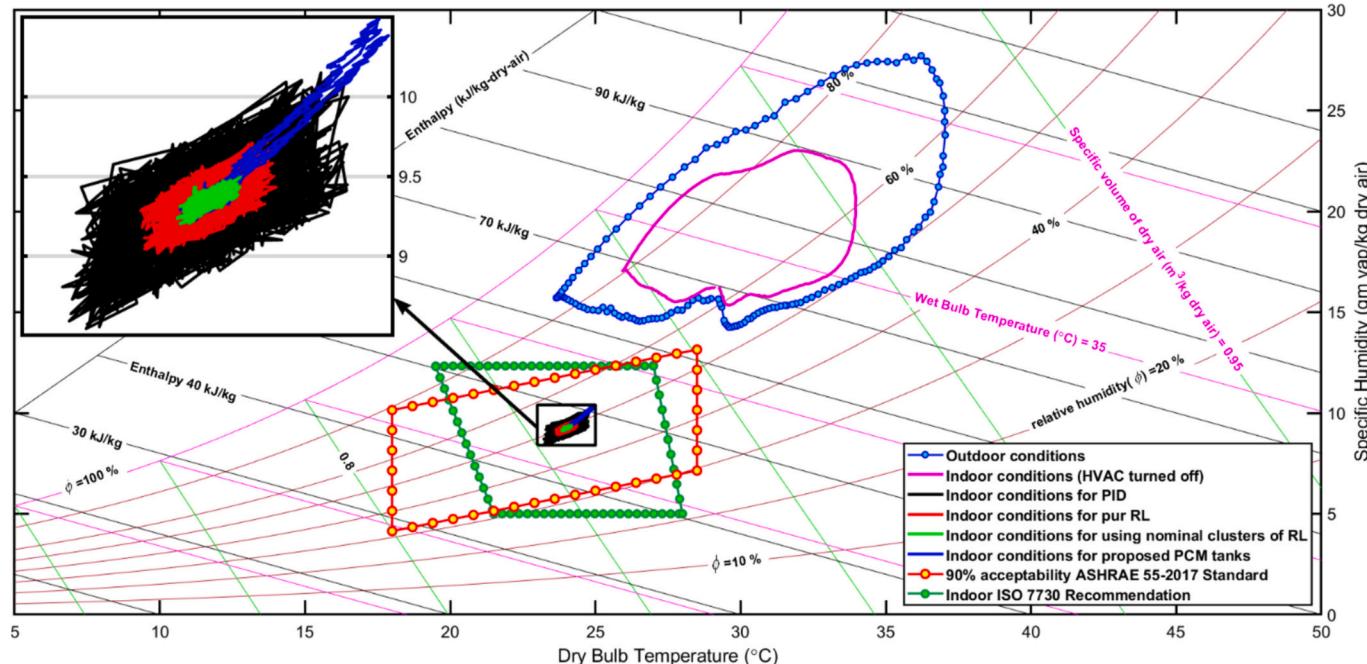


Fig. 26. The performance of the three controllers was assessed using the Psychrometric profile response, by benchmark zone recommendations.

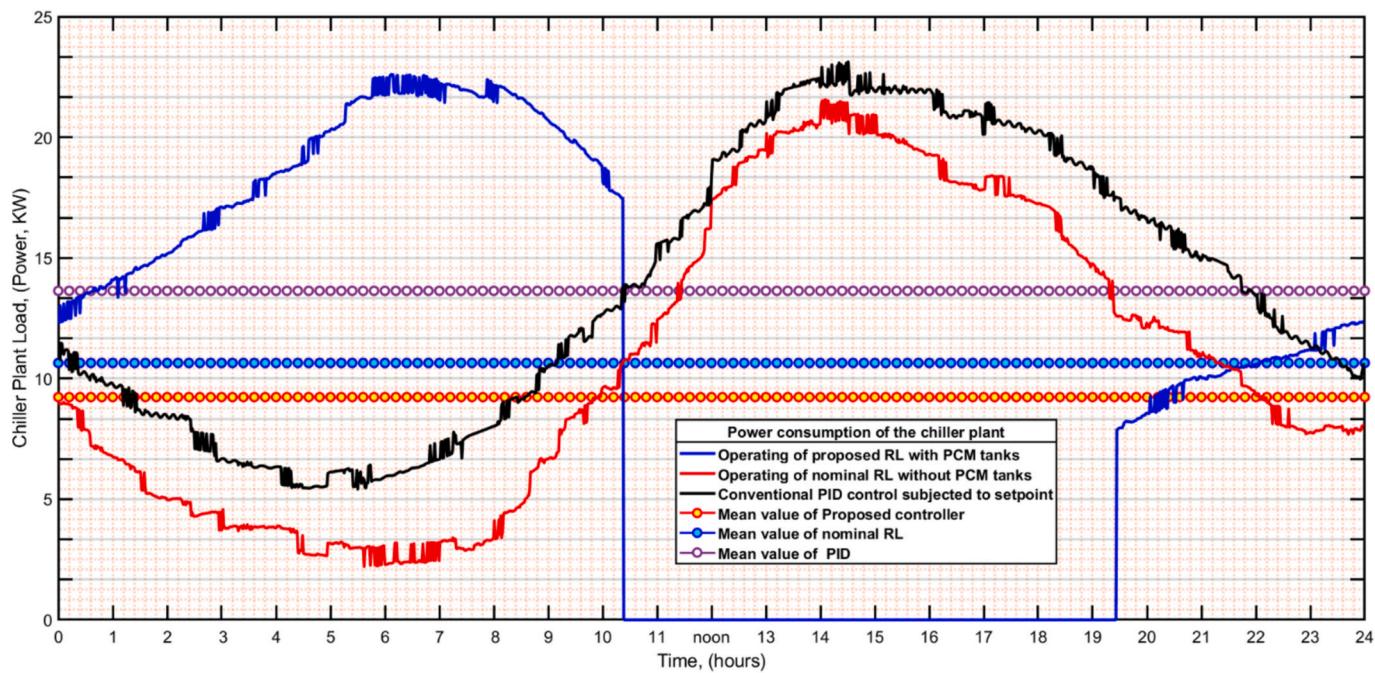


Fig. 27. The three control scenarios exhibit distinct power consumption profiles, which correlate with their performance with the cooling plant power.

5.4. Thermal comfort analysis by psychrometric chart

A psychrometric chart serves as a valuable tool for visually representing the intricate thermodynamic characteristics of air, aiding in practical applications such as cooling process calculations. By providing a clear depiction of saturation points, this chart becomes instrumental in evaluating the dynamic hygrothermal comfort of occupants. Fig. 26, showcasing the ASHRAE comfort zone, serves as a benchmark for establishing regulatory standards related to the assessment of thermal comfort levels. To conduct a comprehensive analysis of the outputs generated by three controllers, a 24-h psychrometric study was conducted using profile data from both indoor and outdoor settings, as

illustrated in Fig. 22. The results obtained from the psychrometric procedures applied to cooling and dehumidification highlight the controllers' effectiveness in maintaining optimal indoor climates, ensuring adherence to the fundamental range of thermal comfort standards.

Conversely, some controllers exceed their recommended capacity, leading to an expansion of the comfort zone coverage area. Notably, all regulated systems consistently remain within the specified comfort range outlined by ISO 7730 and ASHRAE Standard 55. In terms of performance, the DRCADP model distinguishes itself by demonstrating reduced fluctuation in the average temperature, particularly around 24 °C. The proposed control approach consistently operates within the comfort zone, as visually represented in Fig. 26, while occasional

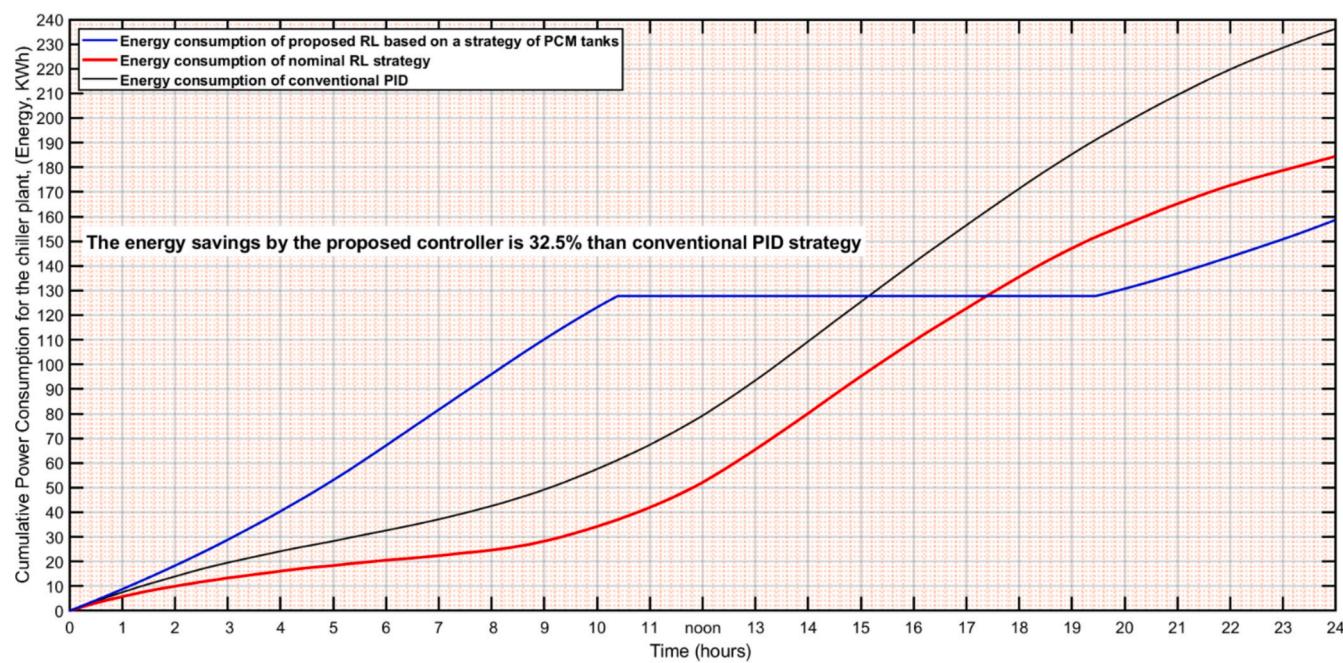


Fig. 28. The three control scenarios exhibit distinct energy consumption profiles, which correlate with their performance with the cooling plant energy.

deviations beyond the comfort zone are observed with the PID method. This nuanced understanding of the psychrometric dynamics contributes to a comprehensive assessment of the controllers' capabilities in ensuring optimal thermal conditions.

5.5. Comparative analysis of energy efficiency

Fig. 27 unveils three distinctive control scenarios, each delineating discernible power consumption profiles intricately linked to their performances within the cooling plant. The mean output power levels were meticulously ascertained for the proposed controller with PCM tanks, notional reinforcement learning (RL) without PCM tanks, and the conventional proportional-integral-derivative (PID) settings, yielding values of 9.1 kW, 10.8 kW, and 14 kW, respectively. Modifications were implemented to enhance the efficiency and adaptability of each approach. The RL with PCM tanks strategy reveals a notable peak cooling load of 23.5 kW at 6:30 am, strategically utilizing PCM for cooling during the temporary chiller disablement from 10:20 to 19:20. In contrast, the chiller load in the other scenarios experiences a nadir between 3:00 and 7:00 am, attributed to enhanced chiller efficiency in lower ambient temperatures. However, on days with elevated external temperatures, the chiller operates at peak capacity from 13:30 to 16:00. Over a 24-h cycle, the conventional PID controller exhibits limitations in maintaining consistent cooling requirements, a common challenge in traditional control systems. In contrast, the integration of RL with PCM tanks emerges as the most effective strategy for handling cooling loads, adeptly navigating external temperature fluctuations, and showcasing superior adaptability and performance in regulating the desired cooling load for specific building conditions.

Fig. 28 displays a thorough examination of power usage data for the building's cooling needs over the day. This analysis utilizes three different models: PID, nominal RL agent activation, and the recommended RL with dual PCMs. The power consumption of all three models starts at a baseline of 0 kWh before the cooling load is activated, indicating a condition of no consumption. Significantly, the proposed RL technique exhibits a greater beginning rate of consumption in contrast to the other two scenarios. Additionally, its consumption trajectory experiences a subsequent increase once the slope hits zero between 10:20 and 19:20, which corresponds to the discharge phase of the PCM tanks. Every model demonstrates predictable patterns of power consumption throughout time. From the start until 10:00 am, the utilization of the chiller for the other two methods shows a slight increase, which can be attributed to the rising temperatures outdoors.

Consequently, these models demonstrate decreased energy usage in order to sustain the appropriate indoor temperature. The RL model with PCM tanks consistently exhibits lower energy consumption compared to both the active RL agent and PID models, as seen by the results presented in **Fig. 28**. More precisely, the suggested RL model consistently maintains lower power consumption levels throughout the day, which is beneficial for saving energy. The proposed Reinforcement Learning (RL) model exhibits a 32.5 % reduction in energy consumption compared to the Proportional-Integral-Derivative (PID) model, as evidenced by a measured value of 131 kW-hours (kWh). Significantly, the PID model demonstrates the most power consumption by the conclusion of the day, reaching 236 kWh, suggesting possible deficiencies in its capacity to regulate heating demand and optimize energy usage. So, the proposed dual-stage eutectic PCM system achieves a 32.5 % reduction in energy consumption, compared to the 25.47 % reduction achieved by the previous study using chilled water tanks [39]. Energy consumption is closely linked to electricity costs and is further influenced by price volatilities during peak load periods. The findings indicate that the proposed control scenarios achieve significant electricity cost savings, surpassing those of conventional PID control scenarios. This suggests that AI-based control is more effective than traditional methods in optimizing electricity costs for residential HVAC systems.

The data presented in **Fig. 28** highlights a significant improvement in

energy efficiency when comparing the RL and PID models to the recommended RL with the PCM tanks model. The recommended RL model has a constant energy usage pattern, resulting in significant daily savings in energy consumption. On the other hand, the PID model's increased power usage towards the end of the day indicates that there is a requirement for improvement in its energy optimization methods. The use of two PCMs in the suggested reinforcement learning (RL) method is identified as a critical element that enhances its energy efficiency and ability to meet cooling requirements. The results confirm that using RL with PCM tanks is a more effective method, providing both energy efficiency and enhanced regulation of heating requirements compared to alternative models.

6. Conclusions

The amalgamation of automation with dual-phase change materials (PCMs) offers a novel approach to enhancing energy efficiency in building cooling systems. This research effectively illustrates that a multi-stage thermal energy storage (TES) system, using two separate PCM composites, may improve energy storage capacity, decrease costs per kilowatt-hour, and mitigate strain on power networks. Major findings include a 32.5 % decrease in energy use, a reduced storage footprint, and enhanced temperature management. Optimizing the sequencing of PCM tanks with the Deep Reinforcement Clustering for Adaptive Decision Policy (DRCADP) approach considerably improved the chiller's performance, notably its Coefficient of Performance (COP), especially during low chilling demand times. The DRCADP methodology surpassed conventional control techniques such as nominal reinforcement learning (RL) and PID control for accuracy, energy economy, and error reduction.

The research reveals that the coordinated functioning of PCM tanks and the HVAC system guarantees ideal interior temperatures, especially during high demand periods. Psychrometric assessments verified that the system sustained pleasant indoor conditions within designated areas. Moreover, the use of an optimal reinforcement learning approach yielded significant energy savings and enhanced system efficiency. This research's principal contribution is the effective application of an improved sequencing mechanism integrated with DRCADP, enhancing thermal comfort, energy efficiency, and the overall performance of PCM-based cooling systems in smart buildings.

Future study need to investigate enhanced optimization, encompassing the incorporation of renewable energy sources, innovative control algorithms, and smart grid systems, in addition to multi-objective optimization and durability assessments. The suggested system has promise for extensive deployment in smart buildings, with future prospects for managing supplementary equipment such as lighting, ventilation, and security systems; nevertheless, the extended learning time for new agents poses a barrier.

The proposed CMARLDC aims to be implemented in large smart buildings by increasing the number of agents and their policies to control all building equipment, including lighting, ventilation, window operations, and security services. In addition to energy saving by the CMARLDC in buildings, its centralization reduces installation and maintenance costs. However, a potential challenge is the increased learning time required as the number of agents grows.

CRediT authorship contribution statement

Raad Z. Homod: Project administration, Conceptualization. **Hayder I. Mohammed:** Writing – original draft, Validation. **Abdellatif M. Sadeq:** Investigation, Data curation. **Bilal Naji Alhasnawi:** Validation, Software. **Ali Wadi Al-Fatlawi:** Resources, Formal analysis. **Ahmed Al-Manea:** Visualization, Validation. **Omer A. Alawi:** Methodology, Formal analysis. **Ali Alahmer:** Resources, Formal analysis. **Jasim M. Mahdi:** Software, Resources, Data curation. **Wael Al-Kouz:** Software, Investigation. **Zaher Mundher Yaseen:** Validation, Software, Data

curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

- [1] C. Song, T. Wang, X. Chen, Q. Shao, X. Zhang, Ensemble framework for daily carbon dioxide emissions forecasting based on the signal decomposition-reconstruction model, *Appl. Energy* 345 (2023) 121330, <https://doi.org/10.1016/j.apenergy.2023.121330>.
- [2] E.K. Avenyo, F. Tregenna, Greening manufacturing: technology intensity and carbon dioxide emissions in developing countries, *Appl. Energy* 324 (2022) 119726, <https://doi.org/10.1016/j.apenergy.2022.119726>.
- [3] H.A. Al-Salam, N.S. Dhaidan, H.H. Abbas, F.N. Al-Mousawi, R.Z. Homod, Review of PCM charging in latent heat thermal energy storage systems with fins, *Thermal Science and Engineering Progress* 51 (2024) 102640, <https://doi.org/10.1016/j.tsep.2024.102640>.
- [4] A. Nandy, Y. Hou, W. Zhao, N.A. D'Souza, Thermal heat transfer and energy modeling through incorporation of phase change materials (PCMs) into polyurethane foam, *Renew. Sust. Energ. Rev.* 182 (2023) 113410, <https://doi.org/10.1016/j.rser.2023.113410>.
- [5] S. Mousavi, B. Rismanchi, S. Brey, L. Aye, PCM embedded radiant chilled ceiling: a state-of-the-art review, *Renew. Sust. Energ. Rev.* 151 (2021) 111601, <https://doi.org/10.1016/j.rser.2021.111601>.
- [6] F. Guo, A. Li, B. Yue, Z. Xiao, F. Xiao, R. Yan, A. Li, Y. Lv, B. Su, Improving the out-of-sample generalization ability of data-driven chiller performance models using physics-guided neural network, *Appl. Energy* 354 (2024) 122190, <https://doi.org/10.1016/j.apenergy.2023.122190>.
- [7] A. Waqas, Z. Ud Din, Phase change material (PCM) storage for free cooling of buildings—a review, *Renew. Sust. Energ. Rev.* 18 (2013) 607–625, <https://doi.org/10.1016/j.rser.2012.10.034>.
- [8] G. Wang, S. Ghoddousi, L. Ding, L. Song, Investigation of different cooling tower fan control strategies using COP of actual chillers and calibrated models of actual cooling towers and fans, *Energ. Buildings* 277 (2022) 112585, <https://doi.org/10.1016/j.enbuild.2022.112585>.
- [9] H. Pieper, I. Krupenski, W. Brix Markussen, T. Ommen, A. Siirde, A. Volkova, Method of linear approximation of COP for heat pumps and chillers based on thermodynamic modelling and off-design operation, *Energy* 230 (2021) 120743, <https://doi.org/10.1016/j.energy.2021.120743>.
- [10] S.K. Anka, K. Mensah, S. Boahen, T.I. Ohm, Y. Cho, J.W. Choi, S.H. Choo, H. Kim, J. M. Choi, Performance optimization of an air source HVAC system for an internet data center building using the integrated COP method, *Journal of Building Engineering* 61 (2022) 105308, <https://doi.org/10.1016/j.jobe.2022.105308>.
- [11] W. Ho, F. Yu, Predicting chiller system performance using ARIMA-regression models, *Journal of Building Engineering* 33 (2020) 101871, <https://doi.org/10.1016/j.jobe.2020.101871>.
- [12] Q. Fu, X. Chen, S. Ma, N. Fang, B. Xing, J. Chen, Optimal control method of HVAC based on multi-agent deep reinforcement learning, *Energ. Buildings* 270 (2022) 112284, <https://doi.org/10.1016/j.enbuild.2022.112284>.
- [13] B. Rismanchi, R. Saidur, G. BoroumandJazi, S. Ahmed, Energy, exergy and environmental analysis of cold thermal energy storage (CTES) systems, *Renew. Sust. Energ. Rev.* 16 (8) (2012) 5741–5746, <https://doi.org/10.1016/j.rser.2012.06.002>.
- [14] G. Campos, Y. Liu, D. Schmidt, J. Yonkoski, D. Colvin, D.M. Trombly, N.H. El-Farra, A. Palazoglu, Optimal real-time dispatching of chillers and thermal storage tank in a university campus central plant, *Appl. Energy* 300 (2021) 117389, <https://doi.org/10.1016/j.apenergy.2021.117389>.
- [15] A.H. Hassan, L. O'Donoghue, V. Sánchez-Canales, J.M. Corberán, J. Payá, H. Jockenhöfer, Thermodynamic analysis of high-temperature pumped thermal energy storage systems: refrigerant selection, performance and limitations, *Energy Rep.* 6 (2020) 147–159, <https://doi.org/10.1016/j.egyrep.2020.05.010>.
- [16] D. Gowthami, R. Sharma, Influence of hydrophilic and hydrophobic modification of the porous matrix on the thermal performance of form stable phase change materials: a review, *Renew. Sust. Energ. Rev.* 185 (2023) 113642, <https://doi.org/10.1016/j.rser.2023.113642>.
- [17] H. Shakibi, A. Shokri, B. Sobhani, M. Yari, Numerical analysis and optimization of a novel photovoltaic thermal solar unit improved by Nano-PCM as an energy storage media and finned collector, *Renew. Sust. Energ. Rev.* 179 (2023) 113230, <https://doi.org/10.1016/j.rser.2023.113230>.
- [18] O. Younis, H. Laiedoudi, A. Abderrahmane, A. Belazreg, N.A. Qasem, R.Z. Homod, M. Rawa, A.M. Hassan, Thermal pattern of nano-encapsulated PCM in a lid-driven cavity with presence of a heated body, magnetic field and limited permeability, *Case Studies in Thermal Engineering* (2023) 103469, <https://doi.org/10.1016/j.csite.2023.103469>.
- [19] J. Jeon, J., Lee, J.H., Seo, J. et al. Application of PCM thermal energy storage system to reduce building energy consumption. *J. Therm. Anal. Calorim.* 111, 279–288 (2013). doi:<https://doi.org/10.1007/s10973-012-2291-9>.
- [20] X. Wang, W. Li, Z. Luo, K. Wang, S.P. Shah, A critical review on phase change materials (PCM) for sustainable and energy efficient building: design, characteristic, performance and application, *Energ. Buildings* 260 (2022) 111923, <https://doi.org/10.1016/j.enbuild.2022.111923>.
- [21] H. Togun, H.S. Sultan, H.I. Mohammed, A.M. Sadeq, N. Biswas, H.A. Hasan, R. Z. Homod, A.H. Abdulkadhim, Z.M. Yaseen, P. Talebizadehsardari, A critical review on phase change materials (PCM) based heat exchanger: different hybrid techniques for the enhancement, *Journal of Energy Storage* 79 (2024) 109840, <https://doi.org/10.1016/j.est.2023.109840>.
- [22] A. Palacios, M. Navarro-Rivero, B. Zou, Z. Jiang, M. Harrison, Y. Ding, A perspective on Phase Change Material encapsulation: guidance for encapsulation design methodology from low to high-temperature thermal energy storage applications, *Journal of Energy Storage* 72 (2023) 108597, <https://doi.org/10.1016/j.est.2023.108597>.
- [23] S. Al Arni, J.M. Mahdi, A.M. Abed, K.A. Hammoodi, H.A. Hasan, R.Z. Homod, N. B. Khedher, Novel multi-layer nano-modified PCM configuration for efficient thermal management of photovoltaic-thermal systems, *Journal of Energy Storage* 103 (2024) 114352, <https://doi.org/10.1016/j.est.2024.114352>.
- [24] L. Abdolmaleki, S. Sadrameli, A. Pirvaram, Application of environmental friendly and eutectic phase change materials for the efficiency enhancement of household freezers, *Renew. Energy* 145 (2019) 233–241, <https://doi.org/10.1016/j.renene.2019.06.035>.
- [25] S. Dinesh, R. Saminathan, M.M. Patil, P. Ramchandra Baviskar, H. Hadidi, S. Vignesh, P. Manoj Kumar, Investigating the single pass baffled solar air heater (SAH) with an organic PCM (OPCM), *Materials Today: Proceedings* 62 (2021) 5245–5249, <https://doi.org/10.1016/j.mtpr.2022.03.216>.
- [26] L. Yang, U. Villalobos, B. Akhmetov, K.J. Onn, A. Gil, W.L. Tan, A. Romagnoli, Active TES with PCM for refrigeration applications, in: *Encyclopedia of Energy Storage*, 2021, pp. 479–497, <https://doi.org/10.1016/B978-0-12-819723-3.00029-9>.
- [27] H. Bo, E. Gustafsson, F. Setterwall, Tetradecane and hexadecane binary mixtures as phase change materials (PCMs) for cool storage in district cooling systems, *Energy* 24 (12) (1998) 1015–1028, [https://doi.org/10.1016/S0360-5442\(99\)00055-9](https://doi.org/10.1016/S0360-5442(99)00055-9).
- [28] Jingwen Wang, Changren Xiao, Xiaqing Yang, Dongxu Ouyang, Mingyi Chen, Guoqing Zhang, Eric Lee Waiming, Richard Kwok Kit Yuen, Jian Wang, An energy-saving battery thermal management strategy coupling tubular phase-change-material with dynamic liquid cooling under different ambient temperatures, *Renew. Energy* 195 (2022) 918–930, <https://doi.org/10.1016/j.renene.2022.06.025>.
- [29] N.A. Qasem, A. Abderrahmane, A. Belazreg, O. Younis, R.Z. Homod, M. Oreijah, K. Guedri, Influence of tree-shaped fins to enhance thermal storage units, *International Communications in Heat and Mass Transfer* 151 (2024) 107220, <https://doi.org/10.1016/j.icheatmasstransfer.2023.107220>.
- [30] J. Zamani, A. Keshavarz, Genetic algorithm optimization for double pipe heat exchanger PCM storage system during charging and discharging processes, *International Communications in Heat and Mass Transfer* 146 (2023) 106904, <https://doi.org/10.1016/j.icheatmasstransfer.2023.106904>.
- [31] R.Z. Homod, K.S.M. Sahari, H.A. Almurib, F.H. Nagi, Double cooling coil model for nonlinear HVAC system using RLF method, *Energ. Buildings* 43 (9) (2011) 2043–2054, <https://doi.org/10.1016/j.enbuild.2011.03.023>.
- [32] R.Z. Homod, K.S.M. Sahari, H.A. Almurib, Energy saving by integrated control of natural ventilation and HVAC systems using model guide for comparison, *Renew. Energy* 71 (2014) 639–650, <https://doi.org/10.1016/j.renene.2014.06.015>.
- [33] R.Z. Homod, Analysis and optimization of HVAC control systems based on energy and performance considerations for smart buildings, *Renew. Energy* 126 (2018) 49–64, <https://doi.org/10.1016/j.renene.2018.03.022>.
- [34] R.Z. Homod, K.S. Gaeid, S.M. Dawood, A. Hatami, K.S. Sahari, Evaluation of energy-saving potential for optimal time response of HVAC control system in smart buildings, *Appl. Energy* 271 (2020) 115255, <https://doi.org/10.1016/j.apenergy.2020.115255>.
- [35] R.Z. Homod, Z.M. Yaseen, A.K. Hussein, A. Almusaed, O.A. Alawi, M.W. Falah, A. H. Abdelrazek, W. Ahmed, M. Eltaweel, Deep clustering of cooperative multi-agent reinforcement learning to optimize multi chiller HVAC systems for smart buildings energy management, *Journal of Building Engineering* 65 (2023) 105689, <https://doi.org/10.1016/j.jobe.2022.105689>.
- [36] S.M. Dawood, A. Hatami, R.Z. Homod, Trade-off decisions in a novel deep reinforcement learning for energy savings in HVAC systems, *J. Build. Perform. Simul.* 15 (6) (2022) 809–831, <https://doi.org/10.1080/19401493.2022.2099465>.
- [37] R.Z. Homod, H.I. Mohammed, A. Abderrahmane, O.A. Alawi, O.I. Khalaf, J. M. Mahdi, K. Guedri, N.S. Dhaidan, A. Albahri, A.M. Sadeq, Z.M. Yaseen, Deep clustering of Lagrangian trajectory for multi-task learning to energy saving in intelligent buildings using cooperative multi-agent, *Appl. Energy* 351 (2023) 121843, <https://doi.org/10.1016/j.apenergy.2023.121843>.
- [38] R.Z. Homod, B.S. Munahi, H.I. Mohammed, M.A.A. Albadr, A. Abderrahmane, J. M. Mahdi, M.B. Ben Hamida, B.N. Alhasnawi, A. Albahri, H. Togun, U.F. Alqsair, Z. M. Yaseen, Deep clustering of reinforcement learning based on the bang-bang principle to optimize the energy in multi-boiler for intelligent buildings, *Appl. Energy* 356 (2024) 122357, <https://doi.org/10.1016/j.apenergy.2023.122357>.
- [39] R.Z. Homod, H.I. Mohammed, M.B. Ben Hamida, A. Albahri, B.N. Alhasnawi, O. Albahri, A. Alamoodi, J.M. Mahdi, M.A.A. Albadr, Z.M. Yaseen, Optimal shifting of peak load in smart buildings using multiagent deep clustering reinforcement learning in multi-tank chilled water systems, *Journal of Energy Storage* 92 (2024) 112140, <https://doi.org/10.1016/j.est.2024.112140>.

- [40] N. Torabi, H.B. Gunay, W. O'Brien, T. Barton, Common human errors in design, installation, and operation of VAV AHU control systems – a review and a practitioner interview, *Build. Environ.* 221 (2022) 109333, <https://doi.org/10.1016/j.buildenv.2022.109333>.
- [41] H. Kim, Y. Cho, Optimization of supply air flow and temperature for VAV terminal unit by artificial neural network, *Case Studies in Thermal Engineering* 40 (2022) 102511, <https://doi.org/10.1016/j.csite.2022.102511>.
- [42] M. Jang, C. Koh, I. Moon, Review of thermal comfort design based on PMV/PPD in cabins of Korean maritime patrol vessels, *Build. Environ.* 42 (1) (2007) 55–61, <https://doi.org/10.1016/j.buildenv.2005.07.025>.
- [43] S. Zhang, W. He, D. Chen, J. Chu, H. Fan, A dynamic human reliability assessment approach for manned submersibles using PMV-CREAM, *International Journal of Naval Architecture and Ocean Engineering* 11 (2) (2019) 782–795, <https://doi.org/10.1016/j.ijnaoe.2019.03.002>.
- [44] H. Liu, Z. Lian, Z. Gong, Y. Wang, G. Yu, Thermal comfort, vibration, and noise in Chinese ship cabin environment in winter time, *Build. Environ.* 135 (2018) 104–111, <https://doi.org/10.1016/j.buildenv.2018.02.041>.
- [45] Z. Tabaie, A. Omidvar, J. Kim, Non-uniform distribution of clothing insulation as a behavioral adaptation strategy and its effect on predicted thermal sensation in hot and humid environments, *Energ. Buildings* 271 (2022) 112310, <https://doi.org/10.1016/j.enbuild.2022.112310>.
- [46] M. Lehna, J. Viebahn, A. Marot, S. Tomforde, C. Scholz, Managing power grids through topology actions: a comparative study between advanced rule-based and reinforcement learning agents, *Energy and AI* 14 (2023) 100276, <https://doi.org/10.1016/j.egyai.2023.100276>.
- [47] R.Z. Homod, A. Albahri, B.S. Munahi, A. Alamoodi, A.K. Hussein, O. Albahri, B. N. Alhasnawi, W.J. Al-Mudhafar, J.M. Mahdi, Z.M. Yaseen, Hybrid weights structure model based on Lagrangian principle to handle big data challenges for identification of oil well production: a case study on the North Basra oilfield, Iraq, *Engineering Applications of Artificial Intelligence* 138 (2024) 109465, <https://doi.org/10.1016/j.engappai.2024.109465>.