Energy & Buildings 351 (2026) 116655

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enb

Experimental insights on CIRLEM: Enhancing energy efficiency and flexibility in buildings☆

Mohammad Hosseini [a], Panayiotis Papadopoulos [b], Muhammad Salman Shahid [c], Kavan Javanroodi [d], Silvia Erba [e], Etta Grover-silva [f], Peter Riederer [f], Amin Moazami [a,g], Frederic Wurtz [c], Benoit Delinchant [c], Salvatore Carlucci [b,h], Mohammadreza Aghaei [a,*], Vahid Nik [d,i]

[a] Department of Ocean Operations and Civil Engineering, Norwegian University of Science and Technology, 6008 Ålesund, Norway
[b] Energy, Environment, and Water Research Centre, The Cyprus Institute, Nicosia, Cyprus
[c] Univ. Grenoble Alpes, CNRS, Grenoble INP, G2Elab, F-38000 Grenoble, France
[d] Division of Building Physics, Department of Building and Environmental Technology, Lund University, SE-223 63 Lund, Sweden
[e] Department of Architecture and Urban Studies, Politecnico di Milano, Milano 20133, Italy
[f] Centre Scientifique et Technique du Bâtiment, 06 904 Sophia Antipolis, France
[g] Department of Architectural Engineering, SINTEF Community, SINTEF AS, Oslo, Norway
[h] Department of Theoretical and Applied Sciences, University of Insubria, Varese, Italy
[i] CIRCLE - Centre for Innovation Research, Lund University, Box 118, 221 00 Lund, Sweden

ARTICLE INFO

Keywords:
Energy flexibility
Decentralized control
Reinforcement learning
Human-centric design
Experimental evaluation

ABSTRACT

The growing complexity of urban energy systems, climate uncertainties, and geopolitical disruptions highlight the need for energy flexibility and smart management. Recent developments in smart buildings enable real-time adaptability and collective energy behavior through the deployment of Reinforcement Learning (RL), which optimizes energy use, integrates distributed resources, and enhances demand response. However, challenges in communication, system diversity, and user intervention must be addressed for scalable and secure multi-agent RL-based management. This study evaluates the application of CIRLEM, a previously developed and introduced Energy Management system that integrates Collective Intelligence (CI) with an online, value-based, model-free RL algorithm. The experiment is carried out in Building Energy Living Lab in France, as one of the pilots of COLLECTiEF, an European funded Horizon 2020 project, equipped with an advanced building management system for one year. The control algorithm interacts with the building management system every 15 min, optimizing setpoints based on real-time monitoring of energy use and indoor environmental conditions. The results indicate an 18% reduction in overall energy use compared to the reference baseline, with heating and cooling demands decreasing by 5% and 32%, respectively. Additionally, peak power demand is curtailed up to 15% for heating and 50% for cooling. The performance of the control algorithm is in an excellent level for more than 50% of the time in 1-month analyses through achieving load reduction and shifting. This experimental study demonstrates that CIRLEM effectively enhances energy flexibility while maintaining thermal comfort, demonstrating its potential for broader implementation, paving the way decentralized energy management solutions in smart buildings and urban energy networks.

☆ This article is part of a special issue entitled: 'Decarbonising Built Env' published in Energy & Buildings.

* Corresponding author.
E-mail addresses: mohammad.hosseini@ntnu.no (M. Hosseini), p.papadopoulos@cyi.ac.cy (P. Papadopoulos), Muhammad-Salman.Shahid@g2elab.grenoble-inp.fr (M.S. Shahid), kavan.javanroodi@byggtek.lth.se (K. Javanroodi), silvia.erba@polimi.it (S. Erba), Etta.grover-silva@cstb.fr (E. Grover-silva), peter.riederer@cstb.fr (P. Riederer), amin.moazami@ntnu.no (A. Moazami), frederic.wurtz@g2elab.grenoble-inp.fr (F. Wurtz), benoit.delinchant@g2elab.grenoble-inp.fr (B. Delinchant), s.carlucci@cyi.ac.cy (S. Carlucci), mohammadreza.aghaei@ntnu.no (M. Aghaei), vahid.nik@byggtek.lth.se (V. Nik).

https://doi.org/10.1016/j.enbuild.2025.116655
Received 19 May 2025; Received in revised form 28 September 2025; Accepted 28 October 2025
Available online 31 October 2025

Nomenclature

AHU	Air Handling Unit
ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers
API	Application Programming Interface
ANN	Artificial Neural Network
ALD	ASHRAE Likelihood of Dissatisfaction
BELL	Building Energy Living Lab
BMS	Building Management System
R^2	Coefficient of Determination
CI	Collective Intelligence
DRL	Deep Reinforcement Learning
DFF	Demand Flexibility Factor
EM	Energy Management
G2ELab	Grenoble Electrical Engineering Lab
HVAC	Heating, Ventilation, and Air Conditioning
KDE	Kernel Density Estimate
LPD	Long-term Percentage of Dissatisfaction
MAE	Mean Absolute Error
MSE	Mean Squared Error
MLP	Multi-Layer Perceptron
NMBE	Normalized Mean Bias Error
POE	Post Occupancy Evaluation
RL	Reinforcement learning
SEM	Standard Error Mean
ReLU	Rectified Linear Unit

1. Introduction

1.1. Background

The increasing complexity of urban energy systems, coupled with the growing penetration of renewable energy sources (RES) creates new opportunities challenges in building energy management. Furthermore, climate change intensifies the frequency and magnitude of extreme weather events, such as heatwaves, cold snaps, and storms, placing unprecedented stress on energy infrastructure and leading to potential technical failures and surge of energy poverty [1]. Global economic and political disruptions, including inflationary pressures and geopolitical conflicts, further exacerbate energy insecurity [2,3]. In response, energy flexibility emerges as a crucial strategy enabling dynamic demand-side management through Demand Response (DR) and decentralized decision-making [4,5]. The integration of smart solutions in buildings enables collective energy system behavior, allowing for a harmonized operation that adapts to real-time conditions and enhances overall performance [6].

Building energy management faces challenges due to the dynamic interactions of multiple entities with diverse characteristics within the urban ecosystem and energy networks [7]. Buildings function as autonomous agents within multi-agent systems, each with distinct energy demands influenced by user behavior, occupancy, weather, and functionality [8]. The heterogeneity of buildings, along with the infrastructure needed for effective and customized energy control, adds further complexity [9]. Addressing these challenges requires robust control strategies capable of real-time decision-making under uncertain and dynamic conditions [10]. Moreover, interactions between entities in energy systems involve security and privacy concerns, where distributed decision-making and localized data storage appear as potential solutions [11]. Among various methods, Reinforcement Learning (RL) demonstrates effective performance in distributed decision-making within a multi-agent environment [12].

RL has emerged as a promising solution for optimizing building energy management [9]. Among other uses, RL algorithms enable buildings to autonomously adapt energy use to environmental and market dynamics [13,14]. Their applications include variety of applications such as optimizing Heating, Ventilation, and Air Conditioning (HVAC) systems, managing distributed energy resources, and participating in demand response programs [15]. Moreover, multi-agent RL systems enhance energy management coordination across urban areas, improving grid stability and overall efficiency [16]. Nonetheless, RL algorithms face challenges such as the curse of dimensionality, when applied to large-scale urban networks with diverse energy agents [17], privacy concerns, computation power, constraints of systems' controllers. To overcome these challenges, in the previous work of the authors [18], an innovative EM framework, called CIRLEM, is introduced and combines Collective intelligence (CI) with RL. CIRLEM utilizes an online, value-based, model-free RL approach and is designed to operate by receiving flexibility signals. The application of CIRLEM was evaluated, initially, in virtual environment through simulation platforms to examine its performance, fine-tuning, and adjustments [6,19].

1.2. Related studies

Review studies show that the deployment of reinforcement learning (RL) in real buildings is promising, although still limited [15,20,21]. Indeed, the transition from simulation to real-world implementation is particularly complex, facing challenges related to the training process, the control security and robustness and the need of greater generalizability of control policies [15]. Moreover, the successful implementation of RL in real-world buildings demands significant domain expertise from control system engineers and field installers, including knowledge in computer science, building physics, HVAC systems, control system architectures, and engineering documentation [22].

Often, RL models are trained offline in a simulated environment and then deployed in the real-world built environment. For example, Kurte et al. [23] evaluated the adaptability of the reinforcement learning (pre-trained) based HVAC control for residential houses. They observed around 30 % of cost reduction by pre-trained model over baseline when validated in a simulation environment and achieved up to 21 % cost reduction when deployed in the real building. Schmidt et al. [24] tested their RL controller in a real school and were able to reduce the energy consumption by almost 33 % while maintaining comfort level similar to the reference period.

Nweye et al. [25] proposed a framework for the training, evaluation, deployment and transfer of control policies for distributed energy resources in grid-interactive communities for different levels of data availability and demonstrated that transfer learning cuts training costs and maintains performance. Utilizing historical data, Zhang et al. [26] introduce a data-driven pre-trained RL algorithm, implemented in real heat pumps in an office building. The results show improvements in the performance of the RL-based control, yet the complexity of access to historical data remains a challenge.

Silvestri et al. [27] applied Deep Reinforcement Learning (DRL) for controlling the thermal dynamics of an office building used as a living lab equipped with thermally activated building system. The DRL controller performances were compared with those of the best-performing rule-based controllers, enhancing indoor temperature control by 68 % without increasing energy consumption. Another study by Silvestri et al. [28] shows the better performance of RL by using mimic learning including a rule-based layer. Mokhtari et al. [29] evaluates the performance of a DRL algorithm in real environment over a winter period, trained initially by simulations and using price signals. Zhang et al. [30] proposed a practical control framework based on DRL which

was implemented and assessed in an existing office building as a case study. They achieved 16.7 % heating demand reduction with more than 95 % probability compared to the old rule-based control.

In a recent review, Kannari et al. [31] highlight that, although extensive research through simulation studies demonstrates the great potential of applying RL-based methods to the control and optimization of building systems, real-world implementations remain limited. The review also emphasizes concerns regarding the security and scalability of such solutions in practice. While existing implementations have demonstrated the potential of RL control in real-life environments, they have also highlighted challenges and research questions [22] that necessitate further experimentation to address these gaps.

1.3. Contributions and aims

Real-world validation is essential to assess the integration of a RL-based control algorithm with existing systems, evaluate the maturity of the algorithm and its practicality under dynamic operating conditions [27,32]. The distinctive contribution of this work lies in its real-world implementation. While existing state-of-the-art RL implementations are focused on short-term periods, this study demonstrates a long-term deployment of the RL-based control algorithm within a multi-agent system (the CI component of CIRLEM), explicitly incorporating human interaction and user feedback. This experiment aims to address key challenges leading to the following contributions:

1. Unlike virtual environments, real-world energy systems face major deployment limitations. DR-enabled energy flexibility depends on energy system controls and their interaction with the environment. This research demonstrates a long-term implementation of an RL-based control, focusing on energy use, thermal comfort, and flexibility; without relying on predefined building or HVAC models, weather forecasts, or energy consumption predictions.
2. Real-time learning and decision-making in energy management require stable and reliable communication between the environment, BMS, and controllers. However, physical constraints, delays, and varying response times due to diverse controllers and communication speeds pose challenges. This study addresses these challenges while exploring mitigation strategies for reliable RL-based energy management.
3. In a real-life setting, occupants actively interact with their environment and can modify setpoints as part of regular building operations. These manual adjustments override the values set by the control algorithm, introducing additional variability that may affect system performance. This research evaluates the impact of occupant interventions on the real-time performance of the control system.

Building on the previous development of CIRLEM, this study further develops the core algorithm through the following updates:

- Two different reward functions based on Fanger and Adaptive comfort models are developed and integrated into the algorithm.
- An adapter is created in the script to couple the control algorithm with the building energy management, overcoming the differences in naming, protocols, and communication.

The impact assessments in this work also include two extensions compared to the earlier study:

- Energy flexibility is evaluated based on demand flexibility factors and detailed responses to weather variations.
- Thermal comfort is assessed based on objective indicators, complemented by subjective indicators including user feedback.

The Building Energy Living Lab (BELL) is a part of the Grenoble Electrical Engineering Lab (G2ELab) in France. This experiment is conducted as part of the demonstration phase of the EU H2020 COLLECTiEF project. This article is structured as follows: Section 2 details the experiment design, covering the living lab structure, communication, implementation, and the experiment framework. Section 3 outlines the methods used, including the energy management approach, baseline prediction, occupant feedback, and key performance indicators. Section 0 presents the results, including metered data and occupant feedback, followed by analysis and discussion.

2. Design of the experiment

This experiment is designed as part of the demonstration phase of the COLLECTiEF project. COLLECTiEF (Collective Intelligence for Energy Flexibility) is an EU-funded Horizon 2020 project running from June 2021 to May 2025 [33]. COLLECTiEF aims to enhance, implement, test, and evaluate an interoperable and saleable energy management system based on collective intelligence that allows easy and seamless integration of legacy equipment into a collaborative network within and between existing buildings and urban energy systems with reduced installation cost, data transfer, and computational power while increasing user comfort, energy flexibility, climate resilience, and data security [34].

2.1. The living lab

The BELL, as a part G2ELab is situated in the GreEn-ER building in the Presqu'île district of Grenoble. The ground floor of GreEn-ER building, designed for public use, houses a cafeteria and event spaces. Academic functions occupy the first three floors, home to the engineering school of the University of Grenoble-Alpes. BELL has its facilities on the top two floors. The purpose of this living lab is to realize interdisciplinary research while considering human activity, where user engagement and feedback serve to perform research to support the energy transition [35,36].

BELL consists of 8 rooms including offices, a laboratory, and a classroom. Room 4A013 is dedicated to the prototyping and small size fabrication of electronic devices, generally occupied by a single person. The room 4A014 generally hosts the interns during the spring semester. The number of interns varies between one and four. The rooms 4A015 to 4A019 are office spaces for professors and researchers. Finally, room 4A020 is a classroom for the students and its occupancy varies as a function of the classes scheduled in this room. Fig. 1 shows the layout of the living lab and an aerial view of the GreEn-ER building whereas Table 1 presents the rooms' features.

Indoor air quality at the living lab is maintained through mainly an airborne system which is supplied from a rooftop Air Handling Unit (AHU). The air supplied to the rooms is a mixture of return air and pre-heated fresh air. Air flow at the room level is controlled by variable flow dampers, which are triggered by the CO_2 level at the room. Additional thermal load demand is provided by radiant ceilings in the offices and the laboratory, and an active chilled beam is mounted in the classroom. All rooms are equipped with several pre-installed sensors including temperature and CO_2 concentration meters to ensure adequate indoor air quality. Fig. 2 depicts the air conditioning system diagram in the living lab.

The indoor air quality in the rooms can be modulated by modifying the setpoints in the BMS, including temperature setpoint (heating in winter and cooling in summer) and CO_2 concentration setpoint. These setpoints can either be modified by the occupants through the LCD screen installed in each room (see Fig. 3), or they can be changed by using the web interface of the BMS as shown in Fig. 4.

The HVAC system utilizes built-in controls provided by the supplier, which communicate with the available sensors to regulate system performance. To ensure the integrity and longevity of the equipment, internal controls of the HVAC system are set up with highly restricted access for external interventions, which causes a serious challenge for developers to interact with the system. For example, compressors and valves do not act immediately by receiving a command, or pumps

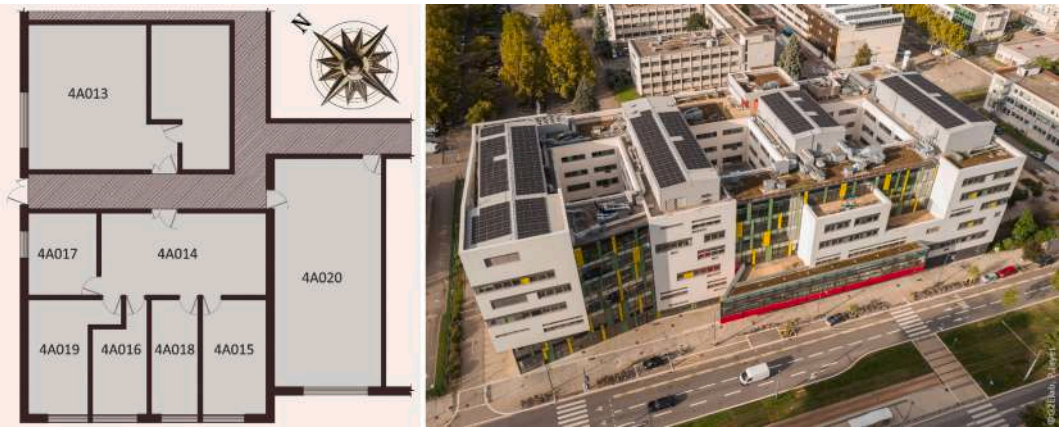

Fig. 1. Layout of the Building Energy Living Lab (BELL)in G2ELab (left) situated in the GreEn-ER building (right) [37,38]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Rooms in the living lab: features in addition to the utilized sensors and controllers.

Room ID	Use type	Comfort category	Surface area (m^2)	Sensors		Controllers	
				Room Temp.	Energy meter	Temp. setpoint	CO_2 Setpoint
4A013	Laboratory	II	63.5	✓	✓	✓	×
4A014	Office	II	42.0	✓	✓	✓	✓
4A015	Office	II	25.0	✓	✓	✓	✓
4A016	Office	II	18.5	✓	✓	✓	✓
4A017	Office	II	19.0	✓	✓	✓	✓
4A018	Office	II	18.5	✓	✓	✓	✓
4A019	Office	II	27.5	✓	✓	✓	✓
4A020	Classroom	II	78.5	✓	✓	✓	✓

change the flow slightly instead of a sharp increase or decrease to protect the electromotor from damage. Such internal controls may lead to a lag in modifications of room conditions.

The sensors in the BMS are connected to the HVAC system control and also record metered data on a server. The data stored on the server could be retrieved through an API and used as historical input for generating a reference baseline (see 3.2). The collected dataset includes room-level measurements for the hydronic system (radiant ceilings) and aggregated living lab–level measurements for the airborne system. The measurements of energy consumption by energy meters are incremental, i.e., for every rise in 1 unit of energy consumption (in this case kWh), the data is recorded in the database, stamped at time instance T. Therefore, the precision of these energy meters is 1kWh. The sensors measuring different temperature variables have a precision of tenth decimal place in Celsius. These sensors measure the values instantaneously. The measurement of water flow rate is also instantaneous with a precision of hundredth decimal place in the unit of m^3/h.

The experiment also benefits from specific sensors, commercially known as “Sphensor”, that are installed in the living lab for the small-scale test of the COLLECTiEF project. Sphensors have a data resolution of 1 min and are placed in each room as close as possible to the work location of the occupants to be capable of characterizing the physical environment around the person. The data is transferred to the dedicated server of the experiment and collected through an Application Programming Interface (API). Table 1 presents the BMS sensors and Sphensors in addition to the control setpoints in each room that are deployed in this experiment whereas Fig. 5 illustrates the placement of Sphensors in three sample rooms.

2.2. The experiment framework

The experiment was designed to evaluate the performance of CIRLEM as a model-free RL-based control algorithm aiming at reducing energy use, maintaining adequate thermal comfort conditions, and enhancing energy flexibility. To include various weather conditions; the experiment was designed to run for an entire year, a period long enough to capture seasonal variations and diverse climatic conditions.

Conducting a one-year field experiment is inherently demanding, and the occurrence of all rare extreme events cannot be guaranteed. However, capturing seasonal variations and possible extreme events could provide a broad representation of performance of CIRLEM under diverse conditions. It is recognized that a single experimental period cannot fully encompass all possible extremes. Nevertheless, the proposed method was designed to function robustly under both typical and extreme conditions, and the insights derived from this long-term implementation can therefore be considered transferable to a wider range of scenarios.

The experiment includes all eight rooms of the living lab (as described in Table 1) for one year from April 1st, 2023, to March 31st, 2024. As illustrated in Fig. 1, the rooms have various sizes, orientations, and occupancy patterns; highlighting the necessity for CIRLEM to adapt to heterogeneous energy demands and thermal conditions. Fig. 6 presents the schematic diagram of the experiment including the time frame, reward functions, occupants’ engagement and feedback, data sharing, control, and monitoring via SG-InterOp, decision-making, and algorithm development workflow.

Due to data privacy and cyber-security concerns, it is not possible to provide direct access to the BELL’s BMS from outside the building. Therefore, an intermediate platform is used to share the collected data and acquire new setpoints for each controller, known as SG-InterOp (acronym of Smart-Grid Inter-Operability). Moreover, the HVAC system has its own internal controls and setup, which are not accessible and completely known to the developers. Hence, a coupling component is designed to connect the control algorithm to the HVAC controller and set the boundaries for setpoint values. MHI-server, as a local device in the living lab, hosts the running code of the EM (i.e., CIRLEM).

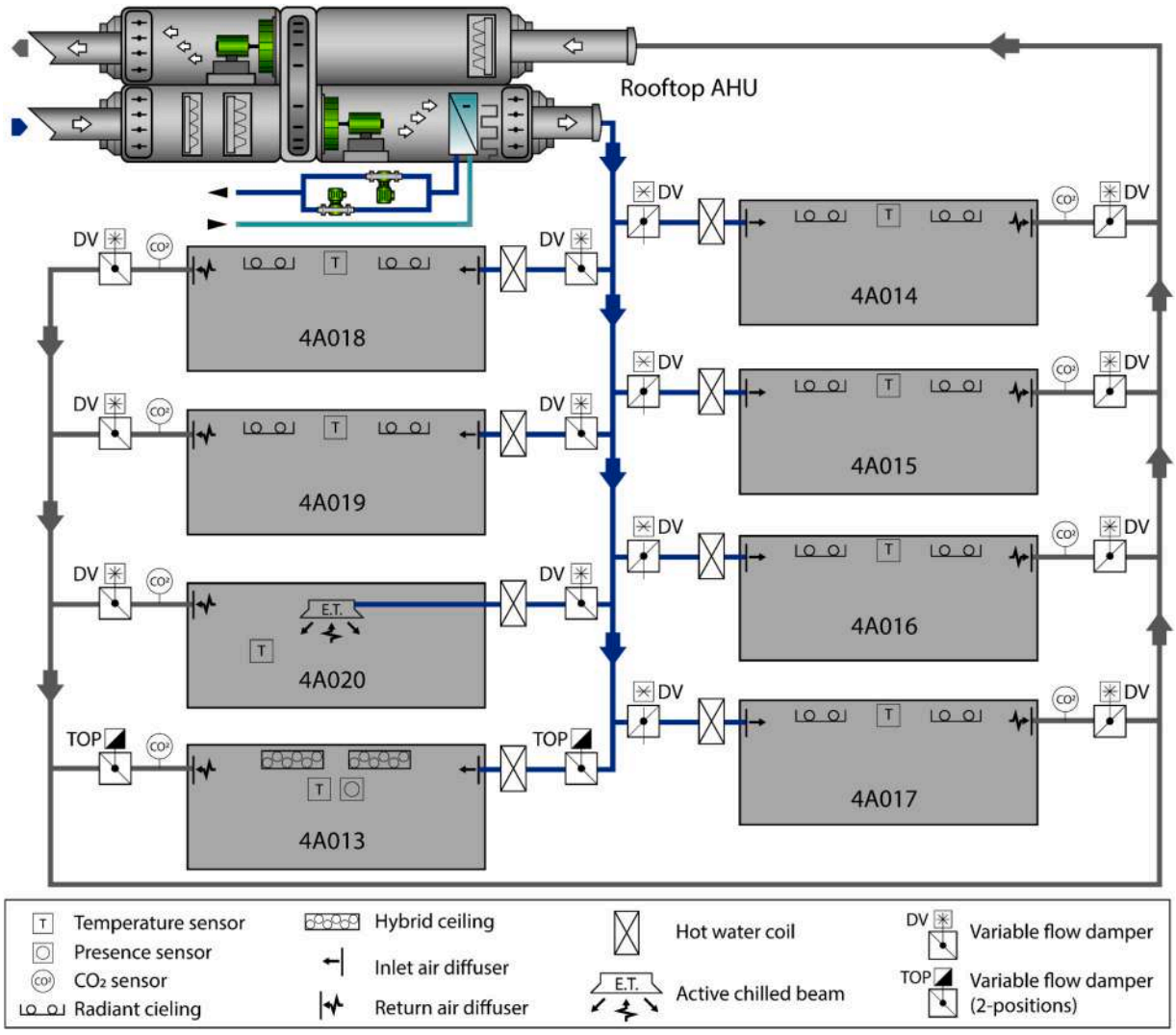

Fig. 2. Simplified layout of the HVAC system and room-level equipment of the living lab. The rooftop AHU supplies the airborne system providing ventilation, fresh air, and partly thermal load.

Developers can update and modify the scripts and receive the measured data and setpoint values through a git repository which includes the scripts to run CIRLEM, actuators and signal values generated at each timestep, and the metered data from the living lab. The developers are able to push the updates to the running codes and pull the generated and metered data. On the other side, MHI-server pulls the updated codes and pushes the generated and recorded data. Indeed, the git repository is the bridge between developers and the MHI-server.

The control algorithm runs every 15 min to collect monitored data and initiate the decision-making process on the MHI server in the living lab. Once the setpoint values are calculated, they are transmitted to the BMS for application to the controllers. A 2-minute lag is incorporated between data collection and transmission to account for communication delays and computation time. The occupancy pattern and user behavior are not modified during the experimental period. There is no intervention or interaction with users, and their existing behavior is treated as a fixed condition in the living lab. Users have access to the control screens in the rooms and can change the HVAC settings (see Section 2.1 and Fig. 3); however, CIRLEM overrides setpoints at 15-minute intervals to prevent inefficient energy use or unnecessary thermal drift.

3. Methodology

3.1. Implemented energy Management: CIRLEM

The implemented energy management algorithm in the living lab is based on the combination of CI and RL so called CIRLEM, as introduced elsewhere [18]. CIRLEM merges key components of CI and RL, and functions by receiving a flexibility signal by agents within the energy system, prompting their adaptive actions. The flexibility signal captures the collective behavior of all agents in the network, while the agents themselves use a value-based, model-free RL engine to learn and make decisions. Using a model-free approach eliminates the need for detailed building models, while RL follows a 'learning by doing' paradigm that reduces dependence on historical data. As a result, the computational requirements and data storage needs of the algorithm are minimal compared to conventional machine learning methods, making it well-suited for edge computation. This is a key feature of CIRLEM, enabling its deployment on single-board computers or other smart devices for energy management systems.

To understand better the CIRLEM approach, we provide some basics for CI and RL based models and their functionalities. CI is a universally distributed, self-improving intelligence that operates in real time with minimal centralized control, enabling efficient skill mobilization across

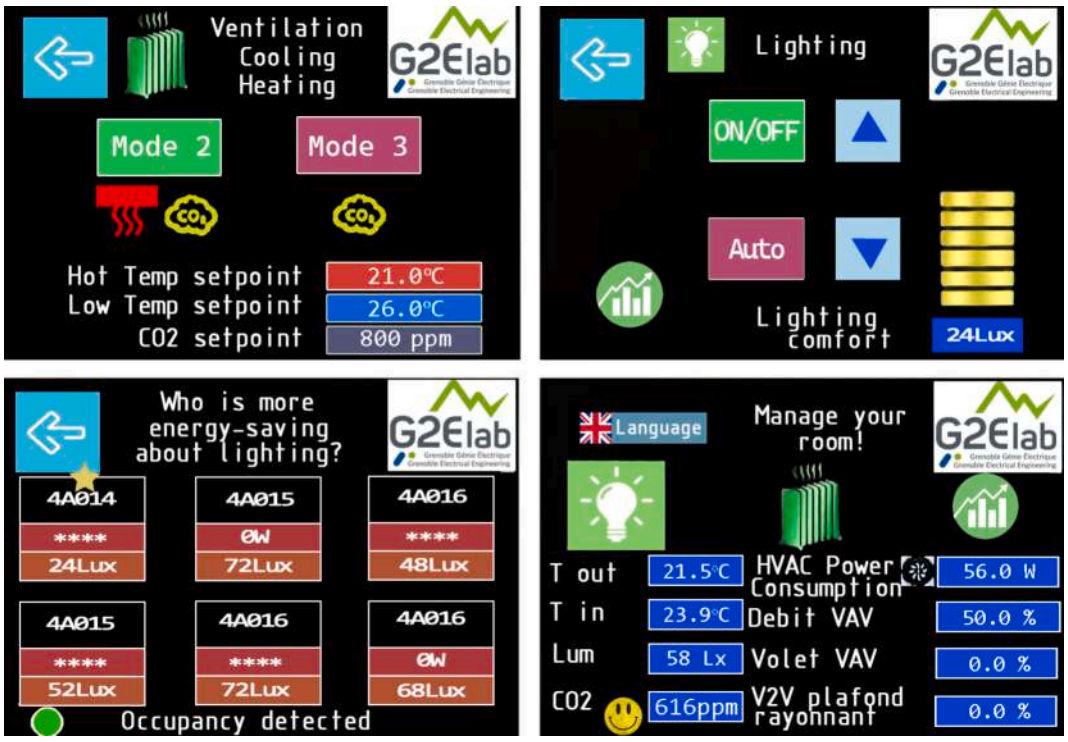

Fig. 3. LCD screen to control (top) and visualize (bottom) each room environment quality.

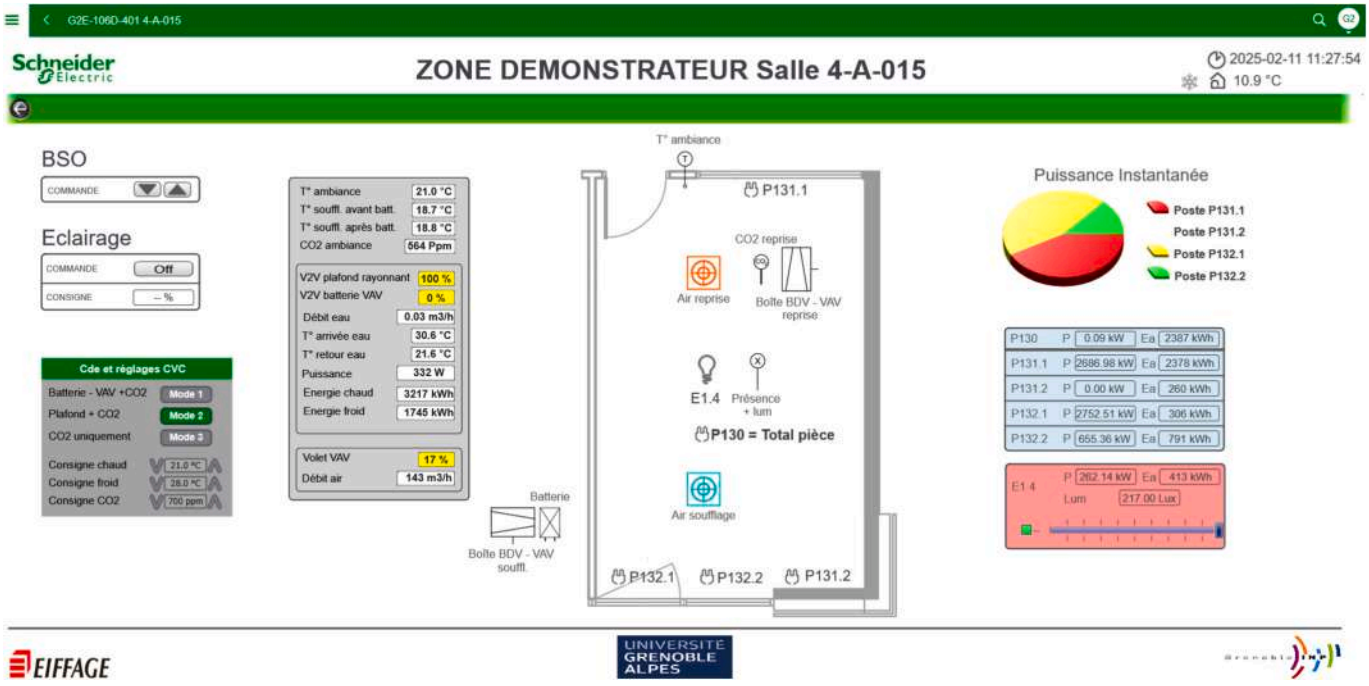

Fig. 4. Control and monitoring through the web interface of the building management system of GreEn-ER. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

large, interactive agent groups. In a CI system, agents use Reinforcement Learning (RL) to enhance both individual and overall system performance [6]. RL is a machine learning paradigm where an agent learns to make sequential decisions by interacting with a dynamic environment [15,39]. The core principle involves training agents to adopt desirable behaviors by associating them with rewards [40]. In an RL system, the agent's interaction with its environment is structured around (i) the policy, which determines the agent's actions, (ii) and the reward function, which provides feedback to inform learning and guide decisions. The environment is often modeled as a tuple (S, A, P, R). Here, S represents the set of possible states, A defines the set of actions, P(s'|s, a) specifies the transition probability from the state s to s', and R(s, a) denotes the reward function [15].

RL algorithms are generally categorized into model-free and model-based approaches based on whether the agent relies on an explicit model of the environment [39]. Model-free methods do not require a

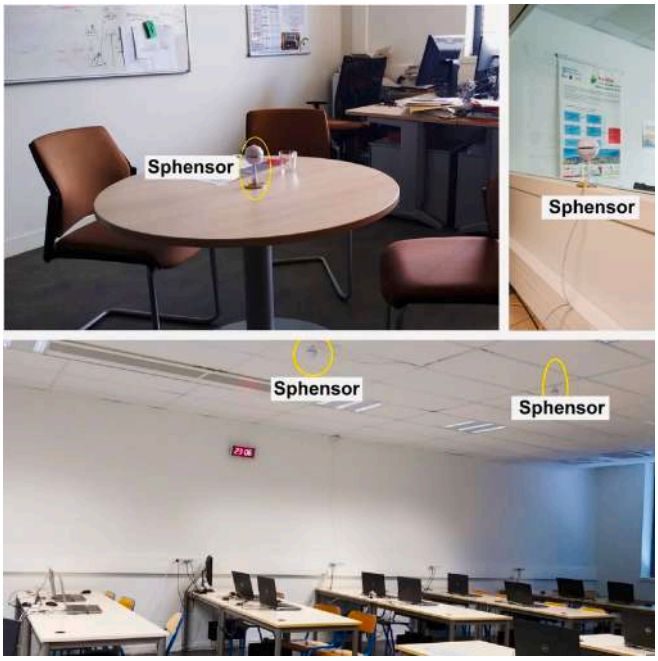

Fig. 5. Placement of Sphensors in different rooms.

predefined model; instead, they learn optimal policies through direct interaction with the environment. By employing trial-and-error strategies, these methods enable the agent to iteratively improve its actions to maximize cumulative rewards [64]. Consequently, model-free RL is favored for its ability to learn directly from the environment without the need for extensive data collection or model development [22].

. As shown in Fig. 7, CIRLEM observes the environment by monitoring the state using sensors. The state includes the conditions of the rooms such as temperature, occupancy, and energy use, as well as outdoor conditions, mainly weather. It is important to mention that monitoring does not occur for every parameter but relies on the impacts on a few measurable parameters. For example, the presence of people has an impact on CO2 and temperature. The RL-engine makes the decision based on the real-time monitored data and arranged policy, thereafter, sending the action to the actuators to modify the state. In the next timestep, the monitored values are used to calculate rewards. The flexibility signal stimulates this cycle, transfers the need for action, and indicates the intensity of the severity of the state's condition.

3.1.1. *Sensors and actuators*

The agent monitors its conditions (i.e., state) based on the gathered data from sensors. The state of each room is represented by different metrics including air temperature and energy use. The impacts of other variables in the room, such as internal sensible and latent heat gains from lighting, equipment, and occupants, as well as operable windows or solar gain, are indirectly captured through air temperature. The list of utilized sensors is provided in Table 1.

Agents modify their conditions by changing the settings of their energy systems through various system controllers and actuators, depending on their availability. In this experiment, the deployed actuators, among all available controllers, are the air temperature setpoint (for space heating and cooling) and the CO_2 concentration setpoint. The temperature setpoints regulate the thermal load from the HVAC system while the CO_2 concentration setpoint controls the ventilation rate at the room level using the adjustable airflow dampers. The list of available actuators (controllers) for each room is detailed in Table 1.

3.1.2. *Flexibility signal*

The flexibility signal is a crucial component in a CI system that activates flexibility by enabling the collective behavior of the entities in the system. In this experiment, the flexibility signal, hereafter mentioned as 'signal', is generated based on the energy demand and takes an integer value between 0 and 5; where 0 means no need for taking action and 5 indicates the maximum intensity. The signal value is based on the ratio of the current energy use against the historical seasonal energy use data. In this experiment, historical data includes the measured energy use data from January 2018 to December 2022 divided into two seasons: summer (cooling period) and winter (heating period). The signal gets a value of 0 if the current energy use falls below the 40th percentile of the historical seasonal data, while a value of 5 is assigned if

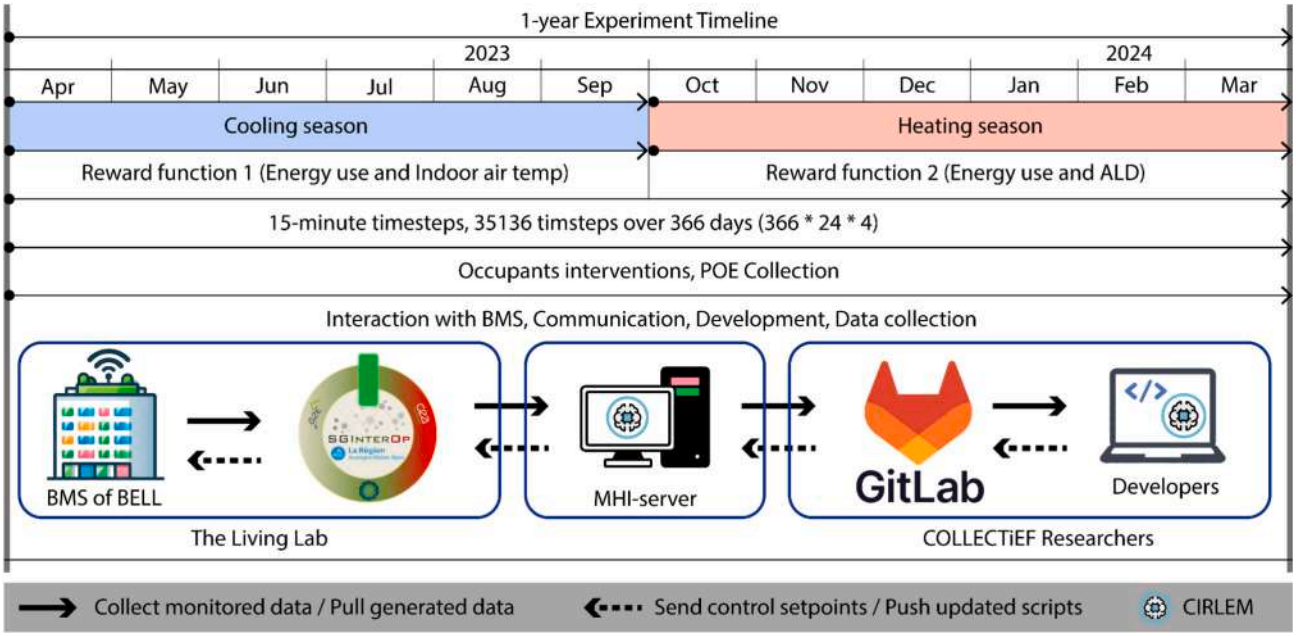

Fig. 6. Schematic diagram of the experiment demonstrating the time frame, reward functions, user engagement, data sharing and reception of new setpoints using SG-InterOp, and communication with developers via MHI-server.

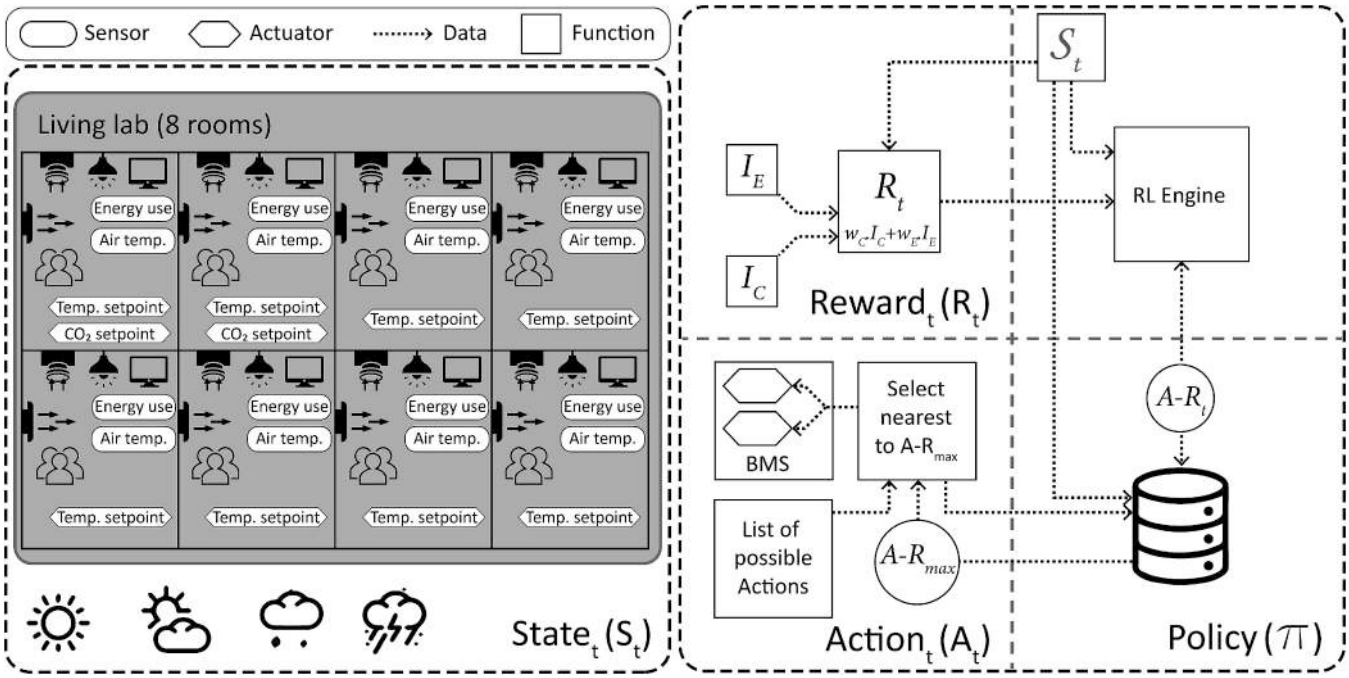

Fig. 7. Flowchart of the RL algorithm showing the relation between State, Action, Signal, and Reward at time t. The schematic layout of the living lab, including eight rooms with sensors, actuators and their equipment, is illustrated in a given state.

it exceeds the 90th percentile. The intermediate signals (1 to 4), are distributed proportionally between the 40th and 90th percentiles, as illustrated in Fig. 8.

3.1.3. Reward function

A reward is feedback from the environment to lead the system to improve its performance. In this experiment, reward is a function of energy use and thermal comfort which obtains a value between 0 and 1. Two reward functions are deployed during the experiment. The first one is a function of energy use and indoor temperature as the indicator of thermal comfort. For the second reward function, the ALD is used as the indicator of thermal comfort instead of the indoor air temperature or other thermal comfort indices.

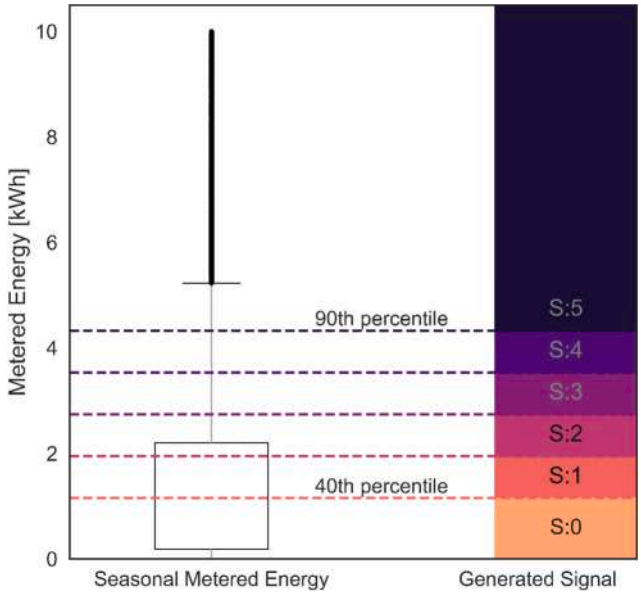

Fig. 8. Signal generation based on the historical energy use; the boxplot is ana example that represents the historical metered energy use in winter.

The first reward function is a weighted summation of the comfort indicator (I_C) and energy indicator (I_E) as denoted in Eq. (1). The weighting factors adjust the relative importance of comfort and energy under varying conditions, thereby defining the system's priorities through the reward function. A higher weighting factor assigned to either indicator directs the system to prioritize that parameter (i.e., comfort or energy) as the main objective. In this experiment, two predefined ratios, previously tested in a simulation environment [41], were applied for evaluation. When the Signal value is 5, indicating that the system is under stress, the comfort weight (w_C) is set to 15 %, while the energy weight (w_E) is 85 %. This means that under extreme conditions, decision-making prioritizes energy efficiency over thermal comfort. In all other conditions, comfort and energy are given equal weight (50 % each), resulting in a balanced impact on the reward calculation.

$$R = w_C \times I_C + w_E \times I_E; w_C + w_E = 1 \quad (1)$$

I_C is a function of indoor air temperature and its relation against thermal comfort thresholds as presented in Eq. (2) where t_l and t_u are the lower and upper limits of the comfort zone, respectively, based on EN16798-1:2019 [42] in relation to the correlated thermal comfort category. i is an indicator for thermal comfort category where it gets the values of 1, 2, and 3 for the zone's thermal comfort category I, II, and III, respectively. I_E is the relative change of the energy use compared to the previous timestep as presented in Eq. (3).

$$I_C = \begin{cases} 1 & t_l < T < t_u \\ i/4 & otherwise \end{cases} \quad (2)$$

$$I_E = \frac{E_t - E_{t-1}}{E_t} \quad (3)$$

The second reward function deploys a different I_C (thermal comfort indicator) in Eq. (1) which corresponds to ALD. ALD is a right-here and right-now thermal comfort index that enables the assessment of the Likelihood of dissatisfaction according to the ASHRAE adaptive comfort models as presented in the ASHRAE standard 55 [43]. It is worth noticing that, according to ASHRAE 55, the adaptive comfort model is

used for naturally ventilated buildings and other approaches should be used in mechanically heated and cooled buildings. However, following the intentions by McCartney and Nicol [44] who developed an Adaptive Comfort Algorithm (ACA) based on the European adaptive comfort model, we used the ASHRAE adaptive comfort model for the development of one of the reward functions of the CIRLEM control algorithm. The reasons behind this choice are that (i) occupants have direct control of several adaptive options available in the rooms, including opening of doors and windows, setpoint adjustment, clothing change, position change, and (ii) adjusting the temperature setpoints of heating and cooling systems according to the adaptive comfort models provides room for energy saving with respect to referring to the Fanger comfort model [45]. Thus, the expression used as the second reward function is given by Eq. (4) (i.e., in the second reward function Eq. (4) is replaced with Eq. (2) to calculate I_C).

$$I_C = ALD(\Delta T_{op}) = \frac{e^{-3.057+0.419\Delta T_{op}+0.007\Delta T_{op}^2}}{1+e^{-3.057+0.419\Delta T_{op}+0.007\Delta T_{op}^2}} \quad 4$$

where $\Delta T_{op} = \left| T_{op,in} - T_{\text{ASHRAE, comf}} \right|$ is the offset of the indoor operative temperature $T_{op,in}$ from the ASHRAE optimal operative temperature $T_{\text{ASHRAE, comf}}$, derived from the prevailing mean outdoor air temperature $T_{pma,out}$. The theoretical comfort temperature according to the ASHRAE adaptive thermal comfort model is defined as Eq. (5).

$$T_{\text{ASHRAE,comf}} = 0.31 \cdot T_{pma,out} + 17.8 \quad 5$$

where $T_{pma(out)}$ is the is the prevailing mean outdoor air temperature. The prevailing mean outdoor air temperature is defined as "the arithmetic average of the mean daily outdoor temperatures calculated over some period of days that have to be no fewer than seven and no more than 30 sequential days prior to the day in question" [46]. Besides the arithmetic average of the mean daily outdoor temperatures, since the 2020 version of the ANSI/ASHRAE 55 standard, the use of the running mean of external temperature is also permitted when the adaptive comfort model is used. This is an exponentially weighted running mean of the daily external temperature $t_{pma,out}$ as denoted in Eq. (6).

$$\overline{T_{mpa,out}} = (1-\alpha)\cdot\left[T_{e(d-1)} + \alpha \cdot T_{e(d-2)} + \alpha^2 \cdot T_{e(d-3)} + \alpha^3 \cdot T_{e(d-4)} + \alpha^4 \cdot T_{e(d-5)} + \cdots\right] \quad 6$$

α is a constant ranging between 0 and 1, and $T_{e(d-1)}$ is the daily mean external air temperature for a time d of a series of equal intervals (day). In the last two versions, ANSI/ASHRAE 55 suggests a value of 0.9 for those climates where the day-to-day temperature variation is relatively minor, such as humid tropics, and a lower α-value of 0.6 for the mid-latitude climates where the day-to-day temperatures variation is more pronounced. An approximated and more practical of the running mean of external temperature is presented in the EN 16798–1 standard that uses only the daily temperatures of the previous seven days.

3.2. Reference energy profile

To evaluate the performance of the control algorithm, it is crucial to define the energy performance of the building without applying the control algorithm as the baseline reference. A high-quality historical dataset for energy use at the living lab level was available, detailing heating and cooling consumption in the lab separately from the rest of the building. This dataset corresponds to the exact same sensors that were deployed during the experiment. Therefore, a data-driven approach using the Multi-Layer Perceptron (MLP) method is chosen for generating reference baseline. MLP is a type of Artificial Neural Network (ANN) that can effectively capture nonlinear relationships for classification and regression tasks by mapping input parameters to output predictions [47]. The MLP, as a modern feedforward ANN, consists of fully connected neurons with nonlinear activation functions, making it suitable for estimating complex targets, such as heating and cooling loads in buildings [48].

In this work, the input parameters are historical energy use (heating and cooling), outdoor air temperature, solar radiation, wind speed, and time features including the numerical values of hour (0–23), day (1–31), and month (1–12). The output parameters are the heating and cooling load of the living lab. The MLP consists of three hidden layers with 256, 128, and 64 neurons, uses ReLU (Rectified Linear Unit) as the activation function, and optimizes the weights using stochastic gradient descent (sdg) with back propagation (initial learning rate, and maximum iteration equal to 0.001 and 2000). These hyperparameters are adjusted manually to reach the acceptable reference profile. Scikit-learn library in Python is deployed to generate the reference profile [49].

To validate the trained model, two widely used performance indicators are utilized, namely, Mean Squared Error (MSE) and Mean Absolute Error (MAE), according to Eq. (7) and Eq. (8), respectively [50]. Models with values for MSE and MAE closer to zero are considered better-calibrated [51]. Here, y_i and $\hat{y}_i$ the metered and reference values, respectively, while n represents the number of samples (i.e., timesteps).

$$\text{MSE}\ (y,\hat{y}) = \frac{1}{n}\sum_{i=0}^{n-1}(y_i - \hat{y}_i)^2 \quad 7$$

$$\text{MAE}\ (y,\hat{y}) = \frac{1}{n}\sum_{i=0}^{n-1}|y_i - \hat{y}_i| \quad 8$$

Additionally, the suggested indicators in ASHRAE Guideline 14 [52] including Normalized Mean Bias Error (NMBE) and Coefficient of Determination (R-squared) are incorporated into the evaluation as denoted in equations Eq. (9) and Eq. (10). NMBE is normalized by the mean value of the dataset; therefore, the result is in percentage and comparable with the values in the dataset. R-squared indicates how well the reference dataset aligns with the regression line of the metered data.

$$NMBE(y,\hat{y}) = 100 \times \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{n\cdot\bar{y}} \quad 9$$

$$R^2(y,\hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad 10$$

3.3. Post occupancy evaluation

POE is used to assess the interactions between objective data (i.e. spot-measured, and monitored) and subjective data (i.e. questionnaires, interviews, and surveys). These data types are used to construct a quantitative representation of the actual environmental conditions inside the building and the subjective judgments of the occupants, which are complex and intertwined information. The analysis accounts for the physical building attributes, features of the mechanical systems, electronic systems, and physical equipment installed, while simultaneously considering occupant behavior and the actual usage patterns. POE relies on designed questionnaires to gather feedback from building occupants in conjunction with environmental monitoring data. Specifically, occupants' subjective judgements on thermal discomfort, preferences, acceptability, perception and evaluation are used for (i) fine-tuning and (ii) performance assessment of the CIRLEM control algorithm. Table 2 presents the section of POE related to the thermal comfort assessment of the indoor environment. Data from the POE questionnaire are collected and treated statistically as ordinal data to avoid any erroneous alteration of the results as recommended by Favero et al. (2021). Occupants were asked to complete an online POE survey twice a week if possible or upon an event of discomfort to capture the extremes.

Occupants' responses on thermal preference, acceptability, judgment and perception will be considered to evaluate the selection of objective methods and assess the performance of the control algorithm. The utilization, process and interpretation of subjective data is presented in the following section.

Table 2
POE survey on thermal comfort.

Questions	Options
At this precise moment, would you prefer the room temperature to be …?	Lower, Without change, Higher
At this precise moment, how do you judge the room temperature on a personal level?	Clearly unacceptable, Just unacceptable, Just acceptable, Clearly acceptable
At this precise moment, how do you find the room temperature?	Extremely uncomfortable, Very uncomfortable, Uncomfortable, Slightly uncomfortable, Comfortable
At this precise moment, how do you perceive the room temperature?	Cold, Cool, Slightly cool, Neutral, Slightly warm, Warm, Hot
At this precise moment, would you prefer the humidity in the air of the room to be …?	Lower, Without change, Higher
How do you judge, on a personal level, the humidity in the air of the room?	Clearly unacceptable, Just unacceptable, Just acceptable, Clearly acceptable
At this precise moment, how do you find the humidity in the air of the room?	Extremely uncomfortable, Very uncomfortable, Uncomfortable, Slightly uncomfortable, Comfortable

3.4. Key performance indicators

This section describes the Key Performance Indicators (KPIs) utilized in this research to evaluate the performance of the living lab before and after the implementation of the developed algorithm. Indicators have been identified to characterize the target domains of the research, i.e., energy use, indoor thermal comfort, and energy flexibility.

3.4.1. Energy use

To assess the energy performance of the building under the operation of CIRLEM, at the first stage, the energy use in monthly, daily, and hourly scales is investigated. Moreover, two KPIs are defined, namely, (1) the peak power reduction and (2) cumulated energy savings which allow for a global view of the change in energy usage during a target analysis period. For the energy use analysis, the aggregated data for the entire living lab is used since the target of the algorithm is to minimize the overall energy use in not only one room but also rooms inside the living lab that have considerable heat exchange through walls and openings so that their thermal load is mixed.

The peak power reduction (P_f) calculates the average power reduction achieved during a target period as defined in Eq. (11). The cumulated energy savings (E_f) is calculated through Eq. (12). $E_{r,t}$ is the reference energy use, $E_{m,t}$ is the metered energy use, and P_f is the average difference of reference profile and metered energy use (kW), and n represents the number of timesteps.

$$P_f = \frac{1}{n}\sum_{t=0}^{n}\left(E_{r,t} - E_{m,t}\right) \quad 11$$

$$E_f = \sum_{t=0}^{n}\left(E_{r,t} - E_{m,t}\right) \quad 12$$

3.4.2. Thermal comfort

This section presents the indicators chosen to evaluate the performance of the proposed control algorithm in providing thermal comfort in indoor spaces. The selected KPIs consider both objective and subjective methods. Further, the evaluation methods consider different temporal and spatial granularities, i.e., a) time-series (right-here-right-now) thermal comfort analysis for each zone/room, b) monthly long-term thermal discomfort analysis for each zone/room, and c) monthly thermal unacceptability through occupants' direct feedback in comparison to the likelihood of dissatisfaction at the building level. The exact methodology applied in each evaluation method is described below.

a) *Time-series (right-here-right-now) thermal comfort analysis for each zone/room*: This analysis is performed at the zone level utilizing the ALD indicator introduced by Carlucci et al. (2021). Time-series analysis is made for each building zone's predefined time window of 8 days, given that going beyond 8 days back it time does not affect significantly due to the exponential decay of terms. The indoor air temperature (T_{sensor}) used for this analysis is measured by the Sphensor devices, the prevailing mean average outdoor air temperature calculated based on a 7-day time window ($T_{mpa(out)}$) and the optimal operative indoor air temperature (T_{ASHRAE}) as defined by the ASHRAE 55 standard. Python's Meteostat open-source library is used to collect historical weather and climate data and following the approach of the nearest/closest available weather station to the case study building [53]. The time-series analysis is performed on the highest available time resolution from the Sphensor devices, which is close to 1 min.

b) *Monthly long-term thermal discomfort analysis for each zone/room*: This analysis is performed by calculating the Long-term Percentage of Dissatisfaction (LPD) using ALD with a monthly time resolution. Its analytical expression accounts for the hourly-predicted ALD calculated for each zone, which is weighted for the number of people inside the zone and is normalized over the total number of people inside the building and overall time corresponding to the seasonal calculation period. Sensor time-aggregated data retrieved through API on the time resolution of 1 h. Data is collated every month and the calculation of the LPD is made for each month.

c) *Monthly thermal unacceptability through occupants' direct feedback in comparison to the likelihood of dissatisfaction at the building level*: This analysis is performed using occupants' direct feedback via POE digital questionnaires. Table 3 shows the classification of responses for estimating the percentage of thermal unacceptability of occupants. A backtracking method is applied to link subjective feedback to environmental conditions recorded at the time of response and it's used to compare the estimated thermal dissatisfaction with the ALD model and the average feedback directly provided by the occupants.

3.4.3. Energy flexibility

To assess the response of the control algorithm to the variations in the environment, the Demand Flexibility Factor (DFF) is used. While there is no standard definition for the flexibility factor, *DFF* shows the deviation of the metered energy operated by a flexibility-based control algorithm from the baseline [54]. The *DFF* value ranges from −1 to 1. A maximum value of 1 occurs when adaptation measures reduce energy demand to zero. The *DFF* value of 0 indicates that the measures provide no improvement, leaving demand unchanged from the reference profile. However, if the algorithm performs worse than the reference profile ($E_{r,t} < E_{m,t}$), *DFF* can take negative values meaning a range of $(-\infty, 1]$. In this experiment, *DFF* is calculated according to Eq. (13) and Eq. (14) where *dff* represents the demand flexibility factor for each time unit, $E_{r,t}$ and $E_{m,t}$ refer to predicted and metered energy at the time t. n is the number of timesteps in the assessment.

$$dff = \frac{E_{r,t} - E_{m,t}}{E_{r,t}} \quad 13$$

Table 3
Classification of occupants' responses to thermal unacceptability.

	Unacceptable	Acceptable
Thermal perception	Cold – Cool – Warm – Hot	Slightly cool – Neutral – Slightly warm
Thermal judgement	Clearly unacceptable – Just unacceptable	Just acceptable – Clearly acceptable
Thermal preference	Lower – Higher	Without change
Thermal comfort	Extremely uncomfortable – Very uncomfortable – Uncomfortable	Comfortable

$$DFF = \frac{1}{n}\sum_{t=0}^{n} dff \quad (14)$$

Besides the numeric values of *dff*, to achieve a better understanding of the performance, the values of *dff* are divided into five ranges of performance rate, according to Table 4, to provide a qualitative assessment of the control algorithm. The ranges are not divided linearly based on numeric values, as neither the equation nor the performance follows a linear pattern. Values around 1 indicate that the algorithm's adaptation results in minimal or nearly zero energy use compared to the reference value, demonstrating excellent performance. Values of *dff* near zero suggest that the algorithm's actions have little impact on energy use, with metered energy remaining close to the baseline. Conversely, values lower than −0.8 indicate significantly higher metered energy use compared to the reference profile, reflecting poor and inefficient performance.

Since one of the objectives of the control system is to respond to climate variations and provide flexibility against them, its performance is evaluated based on the correlation between energy use and indoor temperature with outdoor air temperature. If the correlation is high, it indicates that the building's energy use remains strongly dependent on outdoor conditions. Conversely, a low correlation suggests that the control system is successfully decoupling energy performance from external climate fluctuations. This analysis provides insight into how the implemented RL-based EM can reduce the dependence of energy use on climate variations while maintaining a stable indoor temperature and ensuring occupant comfort.

4. Results and Discussions

The recorded and analyzed results of the experiment are presented herewith divided into energy use, thermal comfort, and energy flexibility. We remind that the experiment period runs for one continuous year, from 1st of April 2023 to 31st of March 2024. The timestep of the running algorithm is 15 min, therefore the entire experiment includes 366 days and 35,136 timesteps. The progress was continuously monitored to ensure a smooth and effective implementation of the algorithm during the experiment. The algorithm successfully interacted with the living lab for 20,600 timesteps (59 % of the total). However, in the remaining timesteps, The CIRLEM algorithm and the BELL technical systems did not interact due to ongoing adjustments, developments, system failures, and communication errors.

Fig. 9 presents a heatmap of the generated signal, where each data point refers to the signal value at a 15-minute timestep. The purple gradient corresponds to signal values ranging from 0 to 5, whereas the gray points represent NaN (Not a Number) values. NaN values appear when the code is not running; therefore, no signal is generated. There are three major gray zones in the graph representing malfunction in algorithm operation. The first one, in April and May 2023, in the early stage of the implementation, occurred mainly because of bugs in the script, mainly caused by different syntaxes in the control algorithm and HVAC system. Also, capturing NaN values from the API resulted in errors in calculations of reward in several cases. These errors helped to improve the coupling between CIRLEM, the intermediate system, and the HVAC controller.

The second major issue happened in August 2023, when, during the summer holidays, the HVAC system was turned off, resulting in calculation errors, which were solved in the coding. Also, not having direct access to the MHI-server during holidays caused delays in solving the issues. Accordingly, the process of updating the scripts through the git repository and updating the running script on the server was modified. The third and main gap in the operation occurred from November 2023 to January 2024 due to major modifications in the script, changing the reward function, updating the signal generator function, and reshaping the policies database in the RL engine.

Table 4
Performance categories based on dff values.

dff Values	Performance rate	Interpretation
$dff \geq 0.7$	Excellent	Near-optimal efficiency
$0.3 \leq dff < 0.7$	Good	Improvement in performance
$-0.2 \leq dff < 0.3$	Baseline	Neutral performance
$-0.8 \leq dff <$ -0.2	Poor	Energy waste is increasing
$dff <$ -0.8	Very Poor	Highly inefficient

The first part of the results shows the statistics of the signal. Fig. 10 illustrates the frequency of each signal value. In 37 % of the timesteps, the signal value 0 is generated most frequently, indicating no decision-making happens from CIRLEM for the HVAC system due to low energy use. Signal 5 is generated in 17 % of the time, which signifies the urgency of action in response to high energy use in the living lab.

MLP models are trained by minimizing an error function on the training set, making their performance highly dependent on data quality. During training, the model aims to fit data points as closely as possible, which can lead to improper mapping if outliers or erroneous patterns are present [55]. As a result, MLP tends to neglect outliers in the training dataset, and when a new input falls outside the trained distribution, the network struggles to recognize it, leading to low confidence in its predictions of those values [56]. The metered historical data (see Fig. 13 − right) includes outliers and significantly large values due to fluctuation in the system or possible wrong measurements, yet the quality of data is high. The predicting MLP model neglects some of the outliers and they are not considered in the prediction; nonetheless, these values have quite low frequency in the dataset. These outliers are visible in Fig. 13 on the right where the metered data exceeds the maximum value of the reference data.

The results for validating the predicting model for the baseline are presented in Table 5. It is important to mention that the value for R^2 for cooling season is below the recommended range in ASHRAE Guideline 14. However, as the guideline suggests, R^2 alone is not always a representative indicator for highly variable data with outliers and thus it should be used together with other indicators. Therefore, considering the R^2 value for heating season close to the recommended value of 75 % in ASHRAE and the acceptable values for MSE, MAE, and NMBE, the model performance is sufficiently robust for the purposes of this study.

Among other parameters, the weather conditions are one of the most influencing parameters in building energy systems. Fig. 11 shows the distribution of the measured hourly air temperature in Grenoble for a period from 2016 to 2024. The graph on the left shows the annual data, the ones in the middle and right represent the data for winter and summer, respectively. The dashed green line shows the average temperature over nine years while the orange lines on each boxplot show the average temperature for the corresponding year and season. The dotted gray lines show the 5th and 95th percentile to demonstrate the coldest and warmest hours (i.e., extremes in this period), respectively.

The annual data for 2023 shows an average temperature approximately 0.8 °C higher than the long-term period. The year 2023 also recorded the highest temperature since 2016, reaching 38.7 °C. During winter 2023–24, the average temperature was 6.1 °C while the average winter temperature in the long-term was equal to 4.0 °C. In this period, the city experienced less than 1 % of hours below the 5th percentile of the long-term data, which is −3 °C. In summer 2023, the average temperature was 22.5 °C against 21.9 °C for the eight-year average. Moreover, around 6.5 % of hours (150 h) are above the 95th percentile of long-term data. All in all, the comparison between historical data and the experiment period shows higher temperatures during the experiment period, particularly in summer, which affects the performance of the algorithm by introducing challenges related to solar radiation and air temperature.

The monthly distribution of outdoor air temperature during the experiment period is presented in Fig. 12 along with the historical data in gray boxes. July 2023 showed the highest average temperature at

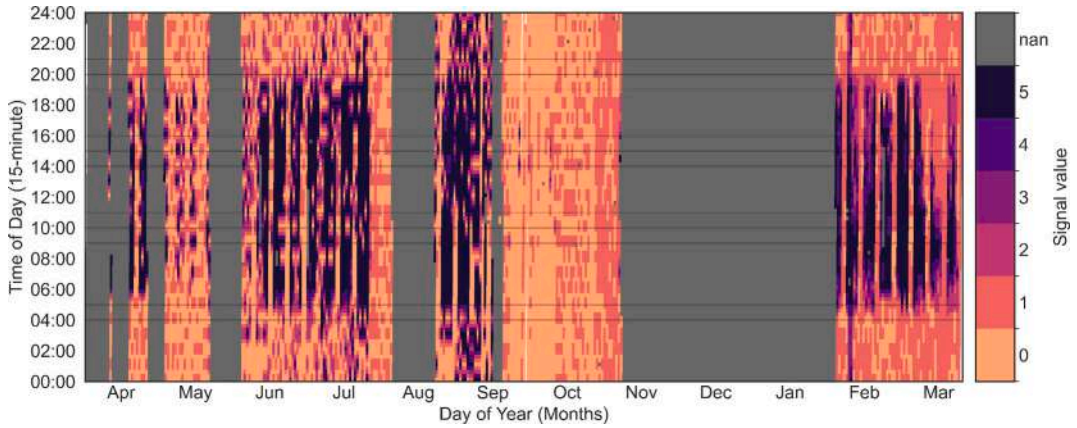

Fig. 9. Signal values in 15-minute timesteps over a year of the experiment. The NaN values are marked with gray.

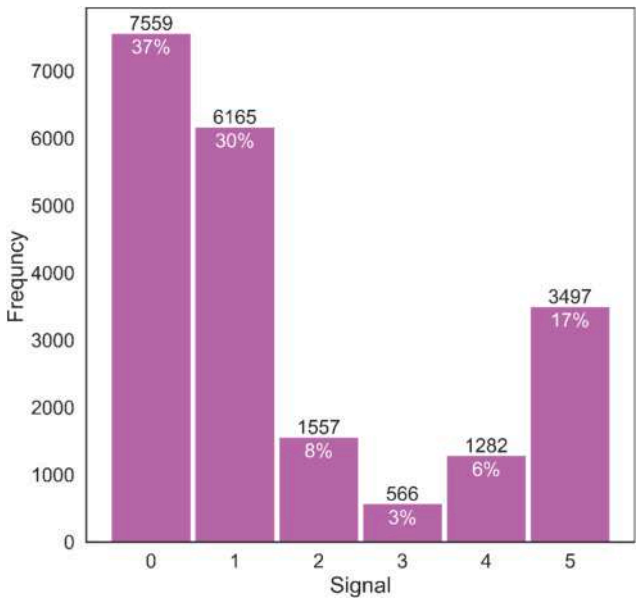

Fig. 10. The frequency of signal generation.

Table 5
Reference profile validation indicators.

	MSE	MAE	NMBE	R^2
Cooling season (Apr-Sep)	1.39	0.45	0 %	65 %
Heating season (Oct-Mar)	2.43	0.91	−5%	74 %

23.5 °C, approximately 1 °C higher than historical data. During winter, December had the lowest average temperature, and January had the coldest hours. Nonetheless, historical data shows lower average and minimum for both December and January indicating that the experiment period had a warmer winter compared to the previous years.

4.1. Energy use

Accumulated metered (solid lines) and reference (dashed lines) energy use is illustrated in Fig. 13, divided into space cooling (blue) and space heating (red). The total metered energy uses for space cooling and heating are 4.3 and 6.7 MWh, respectively. During the cooling period, the metered energy use is 68 % of the reference energy, while metered heating energy reaches 95 % of the reference energy.

The boxplot on the right shows the distribution of the data and the peak power. Metered energy shows higher peak power with 64 hourly values (0.003 % of total timesteps) above the maximum value of the reference energy. The high peak power values with low frequency are not captured by the MLP algorithm because of low frequency and not having a certain pattern, even though the training dataset may include them. It can be noticed that the difference between the reference energy and the metered data is larger in the cooling period (summer). This can be explained by the higher temperature in summer 2023 during the experiment compared to the previous years (see Fig. 12). The winter period during the experiment is warmer than the previous years, which likely results in lower energy use; therefore, the impact of the control system is lower because of lower signal values which activates the control algorithm in less timesteps, and also because of more potential for reducing energy use under extreme conditions as compared to typical conditions.

The monthly energy use including cooling and heating from 2020 until the experiment period in 2023–24 are presented in Fig. 14 and Fig. 15. Historical data shows the highest cooling energy use in July 2022, with a total of 3,170 kWh, which is very close to the reference value of 3,070 kWh for July 2023. Under the experiment, the highest cooling energy use is observed in July 2023, with a metered value of 1,400 kWh, less than half of the reference value. On the other hand, metered heating energy shows the highest total value in January 2024, around 2,300 kWh, followed by 1,600 kWh in December 2023. The reference values for January and December are 2,490 kWh and 2,160 kWh, showing 8 % and 34 % higher values than the metered energy in the same period. The hourly energy use in July and December is analyzed in a higher resolution later in this article.

The hourly values for cooling and heating energy use based on the days of the week for the entire experiment period are presented in Fig. 16. The scatter plot shows the hourly values for hours of the day, the solid lines show the hourly average of the metered energy, and the dashed lines show the average reference for each hour over the experiment period. The visible daily and weekly pattern can be explained by the use type of rooms (see Table 1) which are occupied mainly during working hours. The differences between the dashed and solid lines show the demand reduction by the control algorithm, which happens mostly during working hours.

The cooling energy shows the highest reference values in the afternoon of each working day whereas heating energy shows the highest reference values in the morning. Subsequently, the maximum hourly load reduction happens in the afternoon and morning for the metered cooling and heating energy, respectively. The cooling energy, however, shows a larger difference between the reference and metered values. This is discussed in more detail in the following paragraph.

Following the weekly energy use pattern, Fig. 17 presents the average value for each hour of the day across all weekdays during the experiment period for cooling (left) and heating (right) energy. Solid

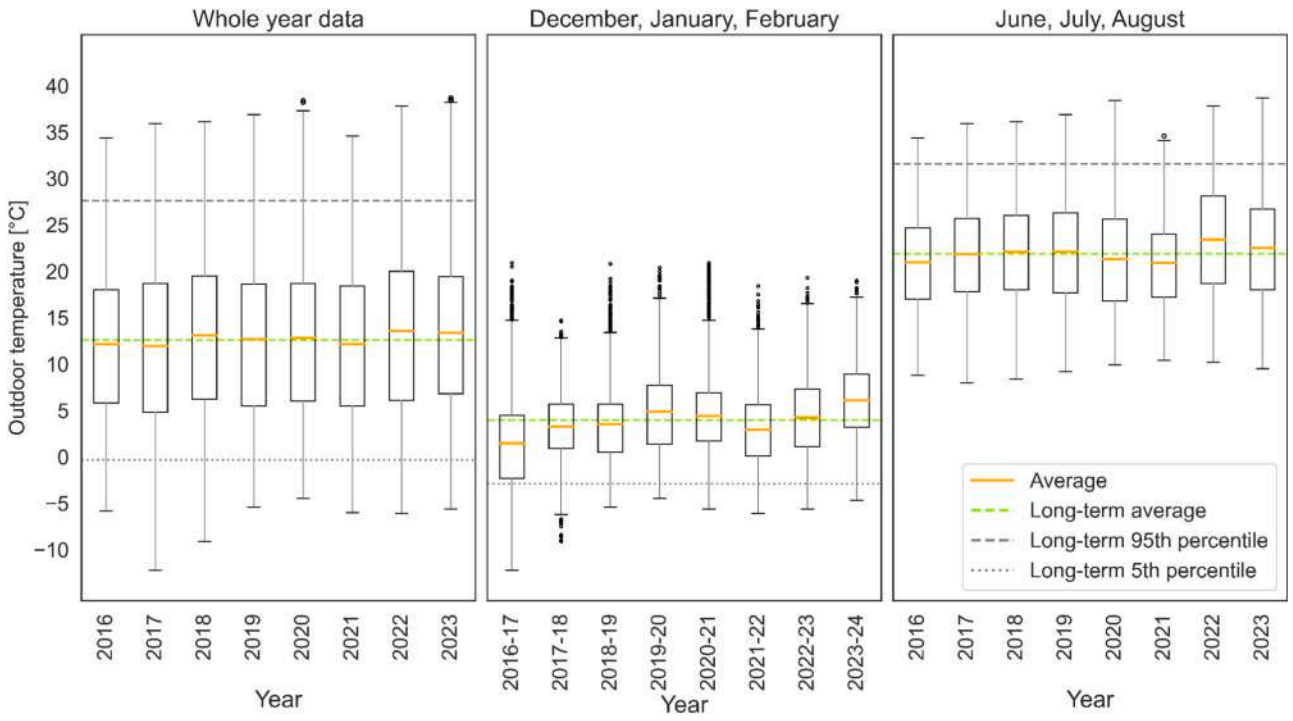

Fig. 11. Distribution of measured outdoor air temperature in Grenoble from 2016 to 2024 on an hourly scale for the whole year (left), winter (middle), and summer (right).

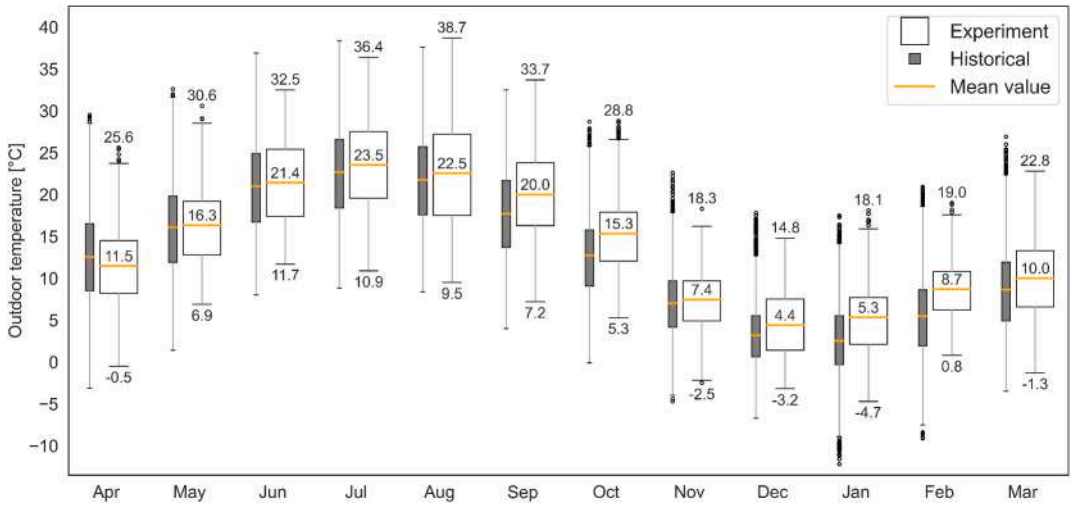

Fig. 12. Outdoor air temperature during the experiment period (white) and historical data (gray) for each month.

lines show the metered values, and dashed lines show the reference values while the Standard Error Mean (SEM) is illustrated as a shade around each line. The average value of air temperature in each hour of the day for the corresponding periods is demonstrated with the gray line on the right axis.

In the cooling season, the load reduction happens from the beginning of the day, but in the afternoon the difference is more significant. The largest difference occurs from 12:00 to 16:00 by around a 40 % reduction with the maximum reduction at 14:00 when the metered cooling energy is almost 50 % of the reference one. These peak energy use, as indicated by the temperature graph (gray line), corresponds to the warmest hours of the day. This high energy use is further influenced by accumulated heat gains from occupants, equipment, and solar radiation since the start of the working hours. Therefore, accumulating energy due to ambient temperature and internal gains, into the building structures, the control algorithm can curtail the demand and peak power significantly.

Heating energy, on the other hand, shows different behavior. From the graph, it is possible to notice that for a few hours during the day (11:00 to 15:00), metered energy is slightly higher than the reference energy. This can be attributed to the optimization strategy of the control algorithm, which prioritizes maximizing long-term rewards rather than achieving the maximum reward at each timestep. As a result, energy use for certain hours may increase to achieve overall energy savings over a longer period or for providing thermal comfort. Also, a large difference between metered and reference heating energy use occurs in the early morning. For example, the average energy use at around 7:00 in the morning, which corresponds to the peak heating demand, shows a 15 %

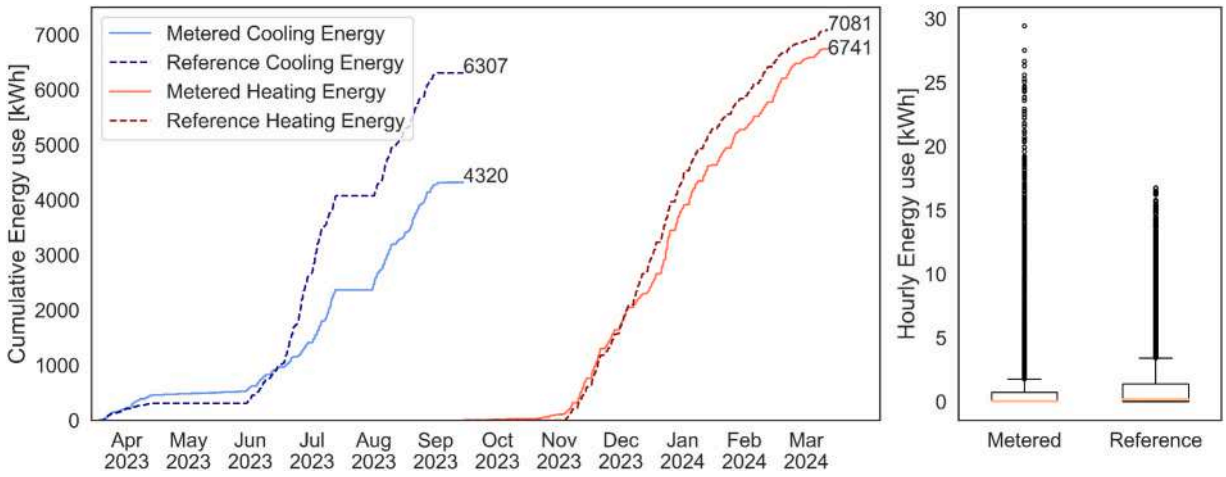

Fig. 13. Reference and metered energy use, cumulative summation (left), and hourly values (right).

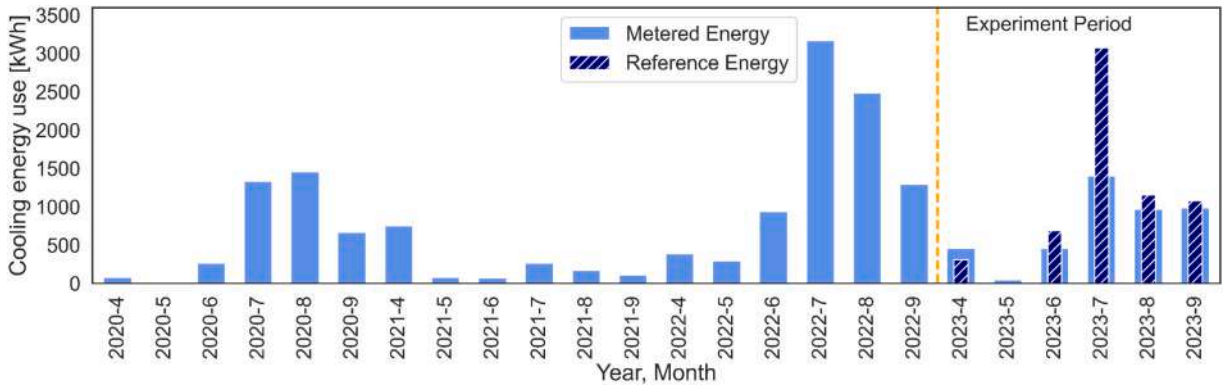

Fig. 14. Total monthly cooling energy use, including metered data, along with reference values during the experiment.

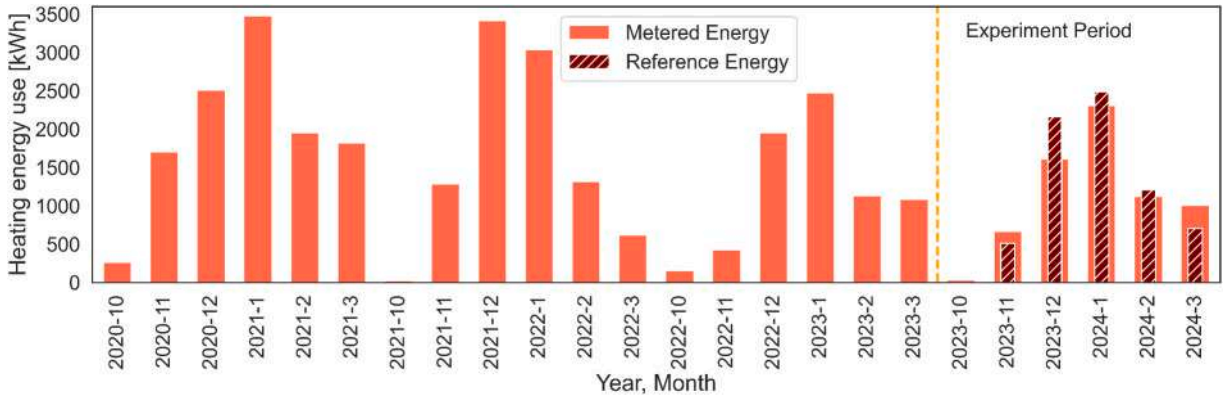

Fig. 15. Total monthly heating energy use, including metered data, along with reference values during the experiment.

reduction in the metered compared to the reference energy. This peak in heating demand is caused by low outdoor temperatures (6:00 and 7:00 in the morning show the lowest hourly average temperatures around 4.8 °C) and unoccupied hours during nighttime.

An investigation is carried out on the warmest and coldest month in the experiment period. Fig. 18 and Fig. 19 show the hourly values for July and December, respectively. Each graph contains data for four weeks starting from Monday and showing the weekends on the right side. The left axis shows the hourly metered and reference energy use (dots) whereas the right axis shows the cumulative sum during each day (dashed lines) for the same data. The signal values are illustrated by purple gradient bars along the upper x-axis with a 15-minute timestep. The length and color intensity of the bars represent the magnitude of the signal: shorter and lighter bars indicate lower values, while longer and darker bars correspond to higher values. These graphs demonstrate the hourly energy use, and the variations compared to the reference energy. These high-resolution timesteps also provide the opportunity to visualize and correlate signal values in 15-minute timesteps with the algorithm performance.

The cooling energy use in four consecutive weeks in July is presented

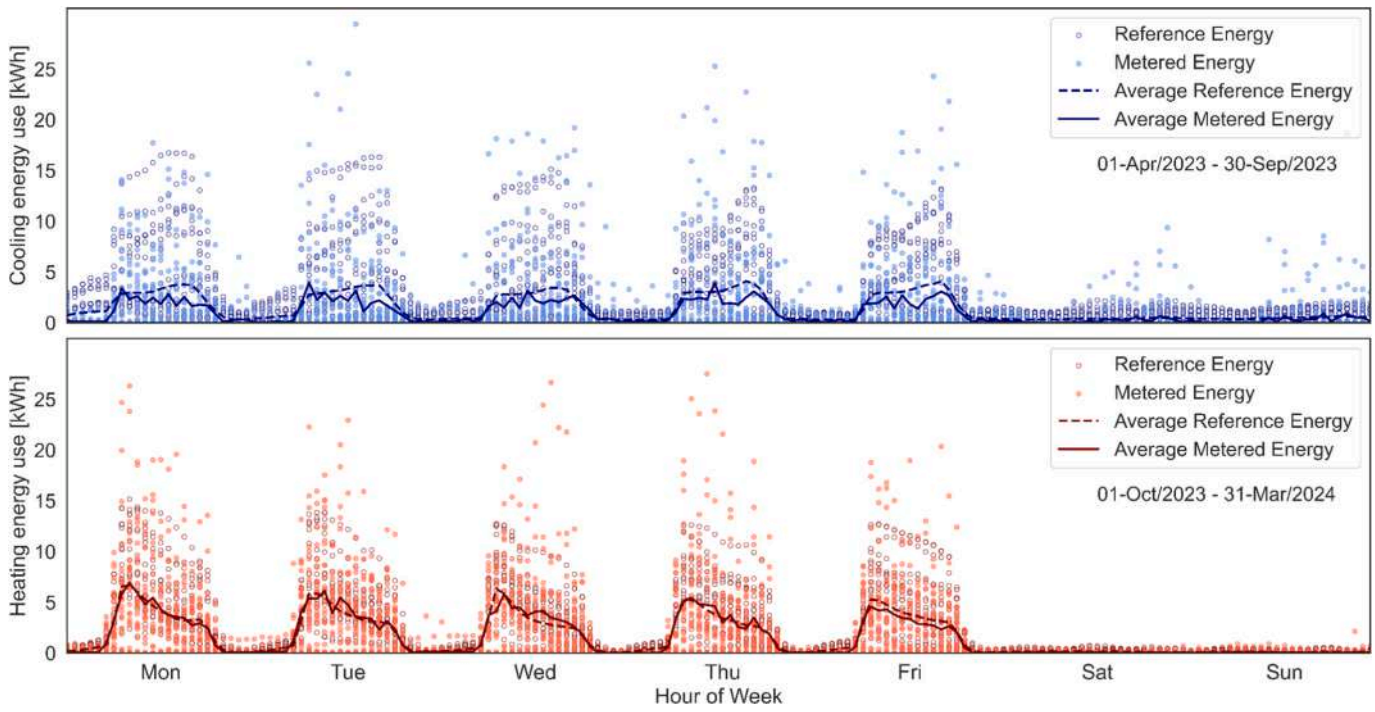

Fig. 16. Energy use for cooling (top) and heating (bottom) based on days of the week for metered and reference data.

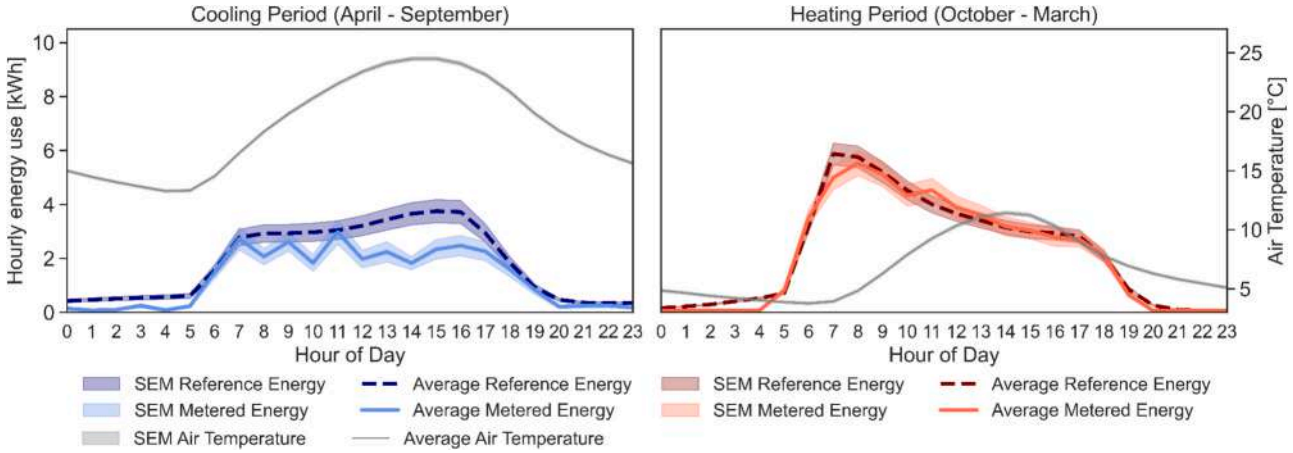

Fig. 17. Average energy use for each hour of the day on weekdays for cooling (left) and heating (right) for metered (solid) and reference (dashed) data, hourly average metered outdoor temperature on the secondary y-axis.

in Fig. 18. It can be noticed that the cumulative metered energy use during almost all days is lower than the reference one, except on July 25 and 26 (the 4th week). On these two days, the signal values are relatively low, especially on the 26th, indicating that the control algorithm is not attempting to reduce energy use considerably. Similarly, the hourly energy use also shows lower values in most of the hours. Nevertheless, there are some hours with high metered energy use, for example, July 6th in the morning, and July 21st around noon. This can be explained by (i) based on the prediction methodology, it is known that the extreme values are not captured by the MLP model, potentially being neglected, and (ii) the operation of the control algorithm causes unusual patterns in the energy use which is different from the historical pattern and making them less correlated with the outdoor conditions which are used for the prediction. Therefore, this is not an error in the control algorithm, and the daily demand is reduced through load shaving or shifting. It can be noticed that for almost all instances of hourly energy use higher than the baseline, the signal values gradually decrease after several timesteps. This suggests that the control algorithm is likely adjusting and returning closer to the normal operation of the HVAC system.

The heating energy use in four consecutive weeks in December is presented in Fig. 19. During this period, eight days recorded higher metered energy use (solid lines) compared to the reference values (dashed lines). As previously mentioned, the winter period during the experiment showed higher average temperatures compared to historical weather data (2016–2022), leading to a lower energy demand. However, when energy use is relatively low, meaning low value of signal, CIRLEM has limited potential to take further actions.

Table 6 presents the KPI values for P_f, E_f, and E_m/E_r for heating and cooling periods (see Eq. (12) and Eq. (13). According to the equations, positive values for P_f and E_f indicate reduction and savings, respectively. Hence, in both seasons, the algorithm achieves its objective, with a greater effect during the cooling season. The total annual metered

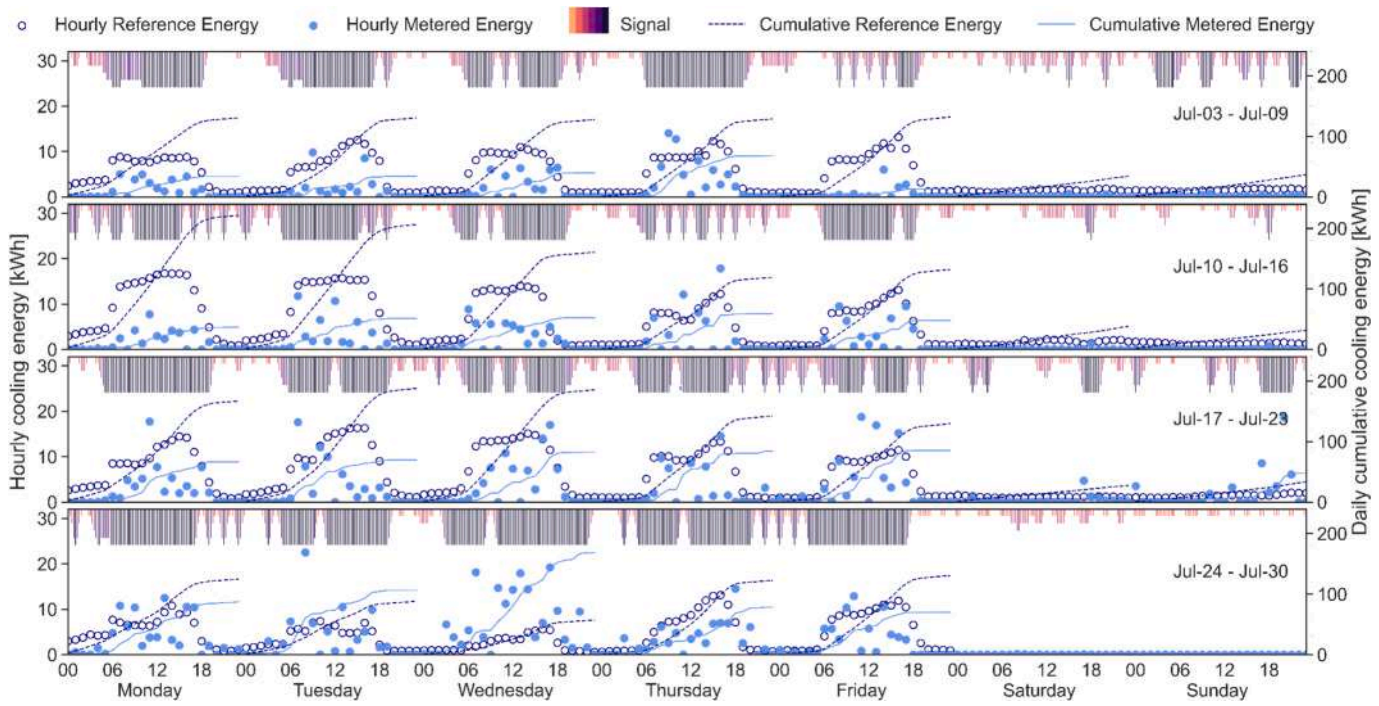

Fig. 18. Hourly (left y-axis) and cumulative (right y-axis) cooling energy use for four weeks from Jul-03 to Jul-30 on an hourly scale. The signal values (purples) are demonstrated on the upper axis of each graph in 15-minute timesteps, the purple gradient and the length of the bars correspond to the signal values from 0 to 5 (lightest to darkest). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

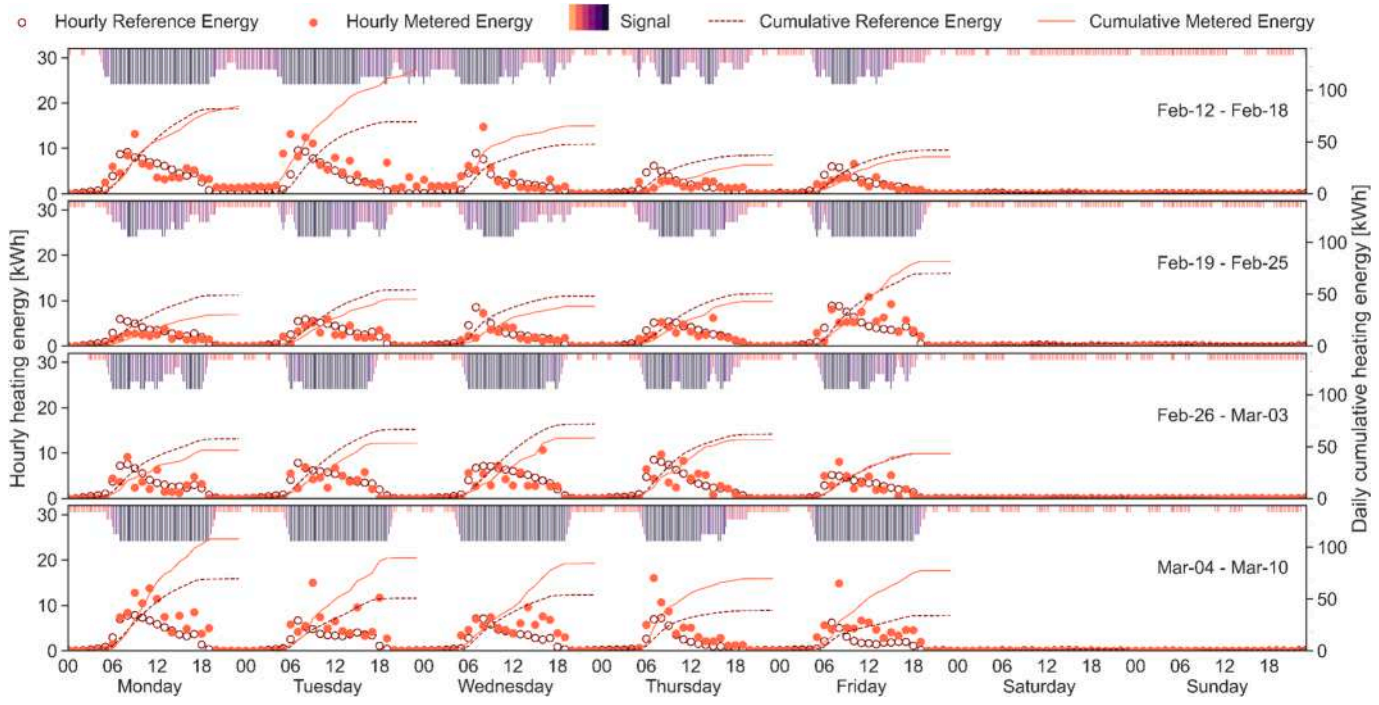

Fig. 19. Hourly (left y-axis) and cumulative (right y-axis) heating energy use for four weeks from Jul-03 to Jul-30 on an hourly scale. The signal values (purples) are demonstrated on the upper axis of each graph in 15-minute timesteps, the purple gradient and the length of the bars correspond to the signal values from 0 to 5 (lightest to darkest). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

energy use is 18 % lower than the reference value, resulting in savings of 2,326 kWh.

Another comparison uses weather-normalized energy use to neglect the impact of weather conditions over different years. This is also a common method among practitioners in building and property management [57,58] as introduced by Energy Star Portfolio Manager Technical Reference [59]. Fig. 20 shows the weather normalized for heating (red) and cooling (blue) energy for historical (circles) and experiment (diamond) periods. As explained before, the impact of the control algorithm is more visible on cooling energy with 7 % reduction

Table 6
Energy KPIs for cooling and heating periods.

KPI	Heating energy (Winter)	Cooling energy (Summer)	Total energy (Entire year)
Average peak power reduction (P_f)	0.08 kWh	0.45 kWh	0.26 kWh
Cumulated energy savings (E_f)	339 kWh	1,987 kWh	2,326 kWh
Metered to Reference energy ratio (E_m/E_r)	95 %	68 %	82 %

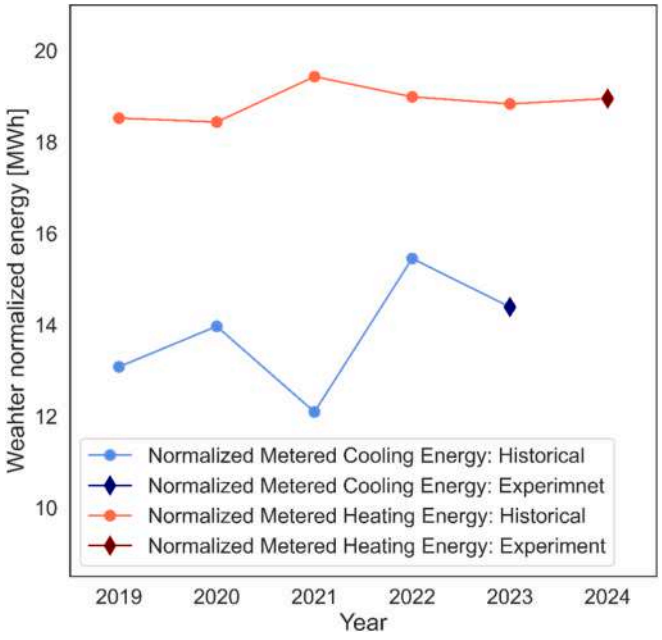

Fig. 20. Weather normalized energy use for historical and experiment period.

as compared to summer 2022. The value for weather-normalized heating is almost constant.

4.2. Thermal comfort

This section presents the thermal comfort analysis for the experimental period, using both subjective and objective KPIs as presented in Section 3.4.2. Fig. 21 presents LPD based on the ALD calculated monthly for the rooms 4A015-4A019. The period between the dashed red line indicates the experiment period (i.e., April 2023 to March 2024). It can be observed that the application of the proposed control approach, which primarily aims to provide energy flexibility, due to the consideration of thermal comfort in the reward function (as described in Section 3.1.3) can give the same and in some cases improved performance concerning thermal comfort at room level. For example, by comparing October 2022 (before the experiment), October 2023 (during the experiment) and October 2024 (after the experiment) for Office 4A015, it can be observed that the LPD during the experiment is lower. The majority of the time, LPD follows the same pattern with and without the application's proposed control algorithm stating that energy flexibility can be obtained without compromising the thermal comfort in the indoor spaces. Poor thermal comfort manifests mainly during unoccupied periods (August due to summer break, room 4A016 was not occupied for long durations) since the HVAC system was locally or centrally off.

Moreover, to analyze the performance taking into consideration both the subjective and objective methodologies, Fig. 22 presents the thermal unacceptability via occupants' direct feedback in comparison to the ALD indicator for all zones of the G2ELab building for each month. The dashed red line indicates the time the experiment started. Since it is not feasible to collect occupants' responses homogeneously over time, the analysis is presented only for periods with recorded responses. Overall, Fig. 22 shows similar patterns between ALD and the direct responses of occupants on thermal unacceptability. That indicates that ALD, which is primarily an objective method based on sensing data, can represent occupant's feedback on thermal comfort adequately. Therefore, even if the adaptive comfort model is currently limited in scope to the naturally ventilated buildings, one of the outcomes of this study shows that it can be used to assess people's thermal acceptability in buildings with mechanically controlled conditions, in line with what originally McCartney and Nicol [44] proposed with the so-called ACA model standing for Adaptive Comfort Algorithm. Considering both objective (i.e., ALD) and subjective (i.e., occupants' feedback) data, an improvement of thermal comfort can be observed during the experiment period compared to the period before the experiment where room air temperature setpoints are not given by the control algorithm (i.e., controlled locally by occupants and partially through predefined schedules). For example, comparing the months 9–2022 and 9–2023, we can observe similar scores for ALD (top plot) i.e., a median of around 10 %. For the same months, as presented in Fig. 14, metered energy use for cooling in the month 9–2023 (experimental) is less than that of 9–2022, giving an advantage in energy savings.

Following the time windows chosen for energy use analysis (see Section 4.1), the right-here-right-now assessment of thermal comfort analysis performed in each room of the experiment are presented in Fig. 23 and Fig. 24 for the summer and winter seasons, respectively. Both figures present the dry-bulb air temperature (purple solid line) measured by the Sphensors, the running mean outdoor air temperature (blue solid line), and calculated ASHRAE operative temperature (orange solid line) in comparison with the ALD indicator (scatters with inferno colormap (black-red)) for each room taking part in the experiment. Fig. 23 shows the results between July 3rd to July 8th, 2023, and Fig. 24 the results for the period March 4th to March 10th, 2024.

In Fig. 23, corresponding to cooling season, poor indoor thermal comfort is observed during the weekends when offices are not occupied when the HVAC system is centrally turned off. Further, during weekdays the overall conditions for thermal comfort are close to optimal, except for a few short-term and low-magnitude discomfort events, which are observed mainly outside working hours. A higher indoor air temperature is observed in rooms 4A016, 4A017 and 4A019 during weekdays since their orientation leads to higher solar gains (i.e., rooms are in the south part of the building as can be observed in Fig. 1).

Fig. 24, corresponds to the heating season, during the weekend, even though the HVAC system is centrally turned off and there is no energy use (see Fig. 19), due to the high insulation of the building, air temperature in all zones does not drop below 20 or 18 °C, maintaining a comfortable indoor environment.

4.3. Energy flexibility

The values for *dff*, calculated based on Eq. (14), are presented in Fig. 25 based on hourly and daily energy use (i.e., daily *dff* uses daily timesteps and total daily energy use). The analysis period for this calculation is the same as those used in Fig. 18 and Fig. 19, from February 12 to March 10 for the winter season and July 3 to 30 for the summer period. During this period, the weather conditions and signal generation make the algorithm interact more with the building. There are 80 and 19 values in winter and summer below −1.3 which appear in the dataset because of a large difference between reference and metered energy use. These very negative values are placed in the lowest category all together (*dff* < -0.8) and, for the sake of clarity, they are trimmed in the graphs.

The hourly values of *dff* (without clipping the extreme negatives) are

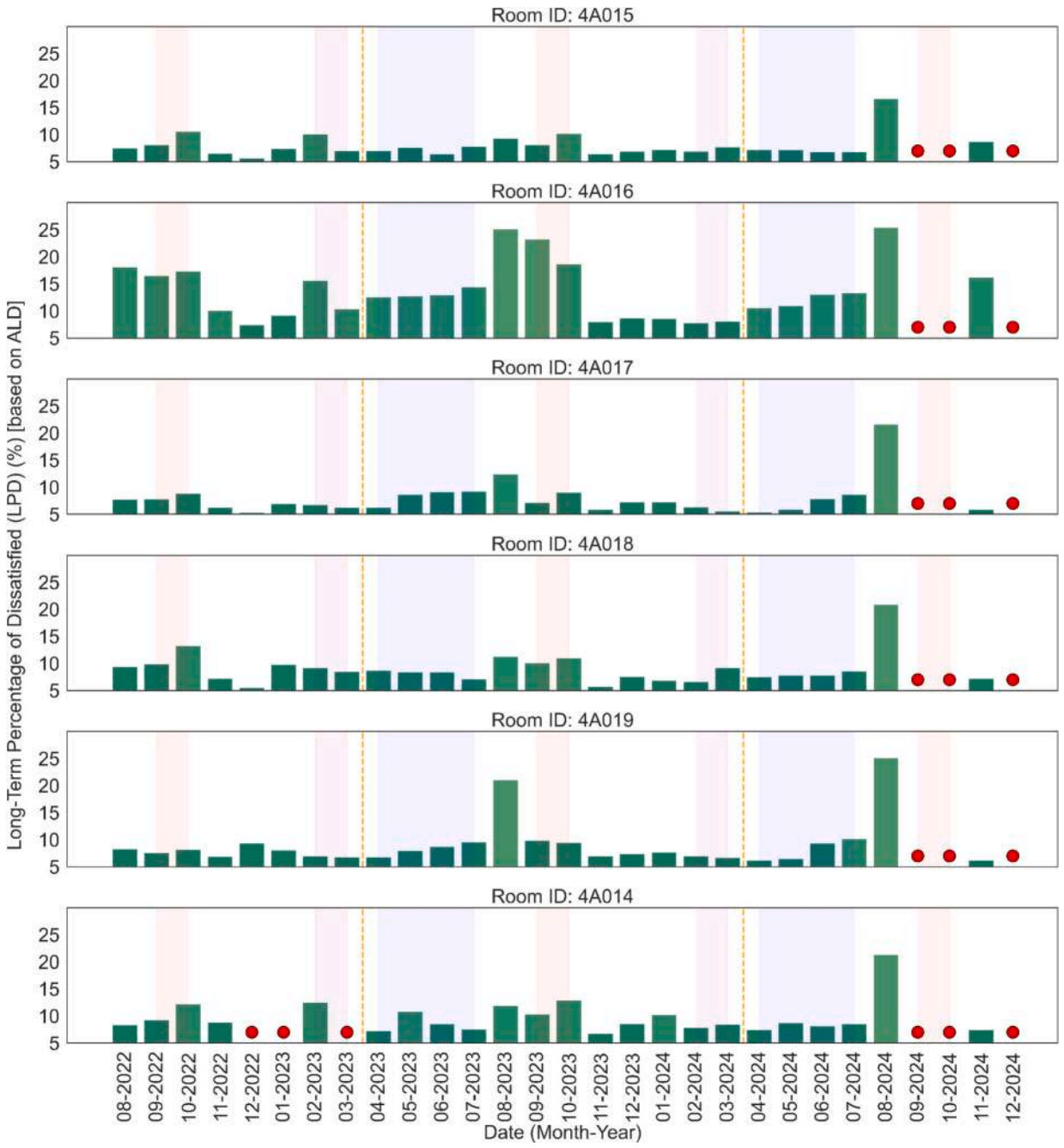

Fig. 21. Monthly LPD based on the ALD model at room level, before, during and after the experiment period. Dashed orange lines indicate the beginning and end of the experiment period and red dots represent the absence of measured data. Highlighted periods within the experimental period indicate the period when the control algorithm was active, and colors indicate the same period outside the experimental period for comparison purposes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

presented in Fig. 26, on the left, for winter (top) and summer (bottom) period divided into four weeks for each season. These four weeks correspond to the analysis weeks in Fig. 18 and Fig. 19. *DFF* values for each week are demonstrated with dashed lines, along with the notation of the values. The histograms on the right show the frequency of *dff* values in each of the qualitative categories mentioned in Table 4. The number of occurrences and percentage of each category mentioned on the graphs.

DFF value for winter period is equal to −0.07, meaning that the control algorithm, on average, does not make a change in energy performance, placing in category Baseline with overall improvements in the second and third weeks and inefficient performance in weeks one and four (according to the values of *DFF*). From the graphs, it can be noticed that the values of *dff* are distributed around zero, showing improvements in some timesteps and inefficiencies in other timesteps. The performance of the algorithm is excellent in 35 % of timesteps, followed by 20 % good performance. The decisions of the algorithm made poor and very poor impacts on energy use in 9 % and 17 % of the actions, respectively.

For the summer period in July, *DFF* is equal to 0.65, showing good performance over one month with weekly *DFF* values all above zero. The performance of the algorithm is in Excellent category for more than 70 % of the timesteps, followed by 13 % in Good. The high performance of the algorithm during summer (cooling season) is discussed also in

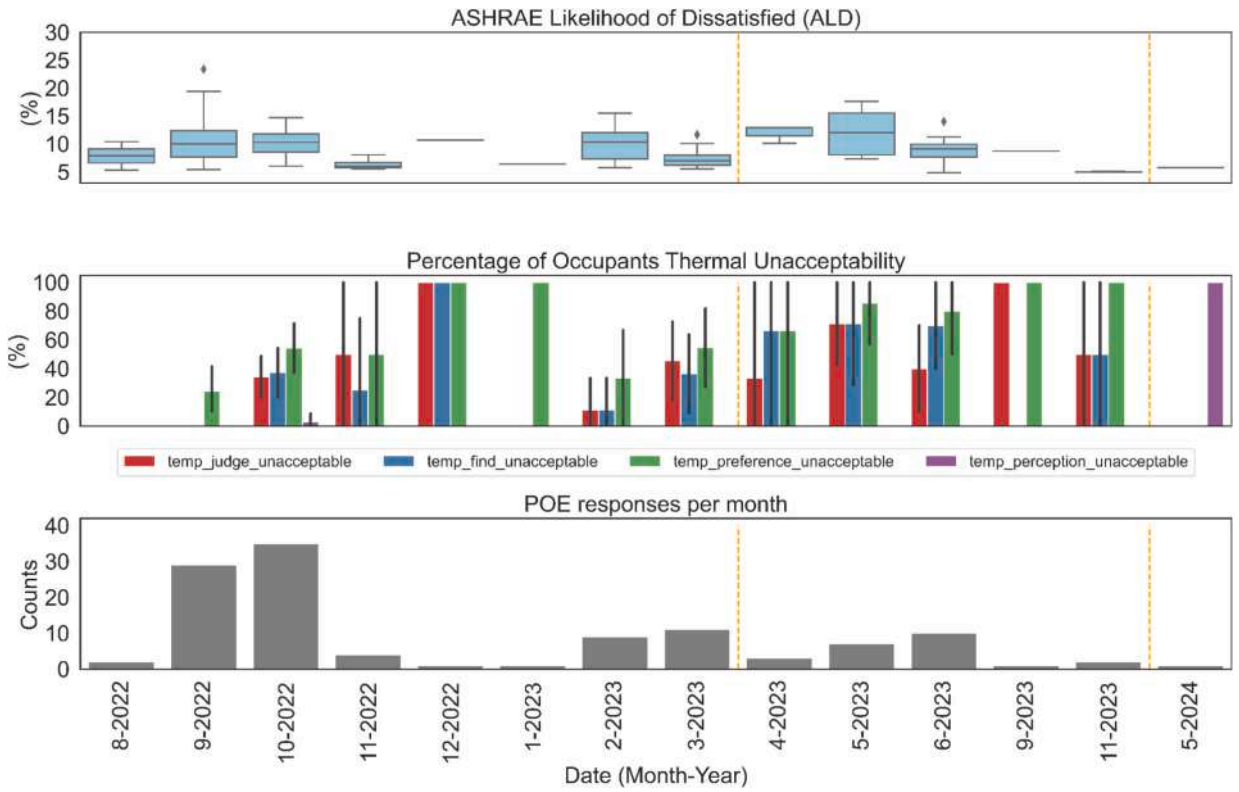

Fig. 22. Comparison of ALD distribution (top graph) with the rate of occupants' thermal unacceptability (middle graph) based on POE questionnaires for each month and all rooms of the living lab. The bottom graph shows the counts of POE responses.

section 4.1, where the weather data shows higher temperature during the experiment period compared to the historical data.

4.3.1. *Responses to weather variations*

Outdoor ambient temperature is an influential parameter on building energy performance, and a major goal of energy flexibility is to respond to the variations in outdoor conditions. Hence, in this section, the energy use is compared with outdoor air temperature. Fig. 27 presents the daily average outdoor air temperature against heating (top) and cooling (bottom) energy divided into daytime (left) and nighttime (right). The polynomial regression curves, along with their 95 % confidence intervals, are illustrated in green. Using daily averages helps reduce the influence of large values in metered energy, as the control algorithm does not necessarily aim to reduce energy use at every timestep but rather optimizes overall system performance. First and foremost, the graphs demonstrate that in both day- and nighttime, for both heating and cooling energy, daily average metered energy has lower variations in relation to daily average outdoor air temperature. Moreover, the lower energy use and variations during nighttime are a result of the building's schedule, as it is unoccupied. As explained earlier, the graph also indicates that the discrepancy between metered and reference energy is greater in summer than in winter.

The variation of indoor temperature concerning outdoor temperature is presented in Fig. 28, with historical data (left) and the experiment period data (right). The top and bottom rows show the data related to the heating (winter) and cooling (summer) periods, respectively. Due to the high density of data points, contour lines are added to show the density of the points using a Kernel Density Estimate (KDE). Unlike histograms, KDE applies a Gaussian kernel to each observation, generating a continuous density estimate [60].

To study the correlation between indoor and outdoor temperatures, the slope of the regression line (regression coefficient) is used. A comparison of historical and experiment period data for both heating and cooling seasons shows a higher slope during the experiment period, indicating a stronger correlation between indoor and outdoor temperatures. This happens due to the fact that the control algorithm reacts to the variations in the outdoor temperature by adjusting the HVAC system setpoints. As a result, indoor temperatures show a greater variation in alignment with outdoor temperature changes. However, this does not imply a deviation from acceptable thermal comfort ranges (see *4.2*, and Fig. 24). This reaction leads to energy flexibility aiming at reducing the impacts of outdoor temperature on energy use (see Fig. 21) while maintaining thermal comfort conditions. The next section provides a more detailed discussion of thermal comfort conditions.

4.4. *Limitations and mitigation strategies*

Overcoming real-life limitations is a key objective of this research, paving the way for the practical implementation of RL-based control strategies. In the course of this experiment, several challenges related to communication, data access, computation, and system integration were identified and addressed. These challenges primarily arose due to the complexity of real-time energy management in a multi-system environment, where different components from different suppliers must interact seamlessly to ensure reliable operation. Specifically, this research utilizes two different sources of monitored data including BMS sensors and Sphensors, collected on different servers with various timesteps, units, and accuracy.

4.4.1. *Communication Issues: Delays, Security, and developer access*

Reliable communication between monitoring systems, BMS, and control algorithms is critical for real-time energy management. However, communication lags, data transmission delays, and security constraints posed significant challenges during the experiment. To mitigate the impact of communication and computation delays, a 2-minute buffer was introduced between the monitoring and control loops. This ensured that any minor delays in data transmission and processing did not compromise the performance of the decision-making algorithm.

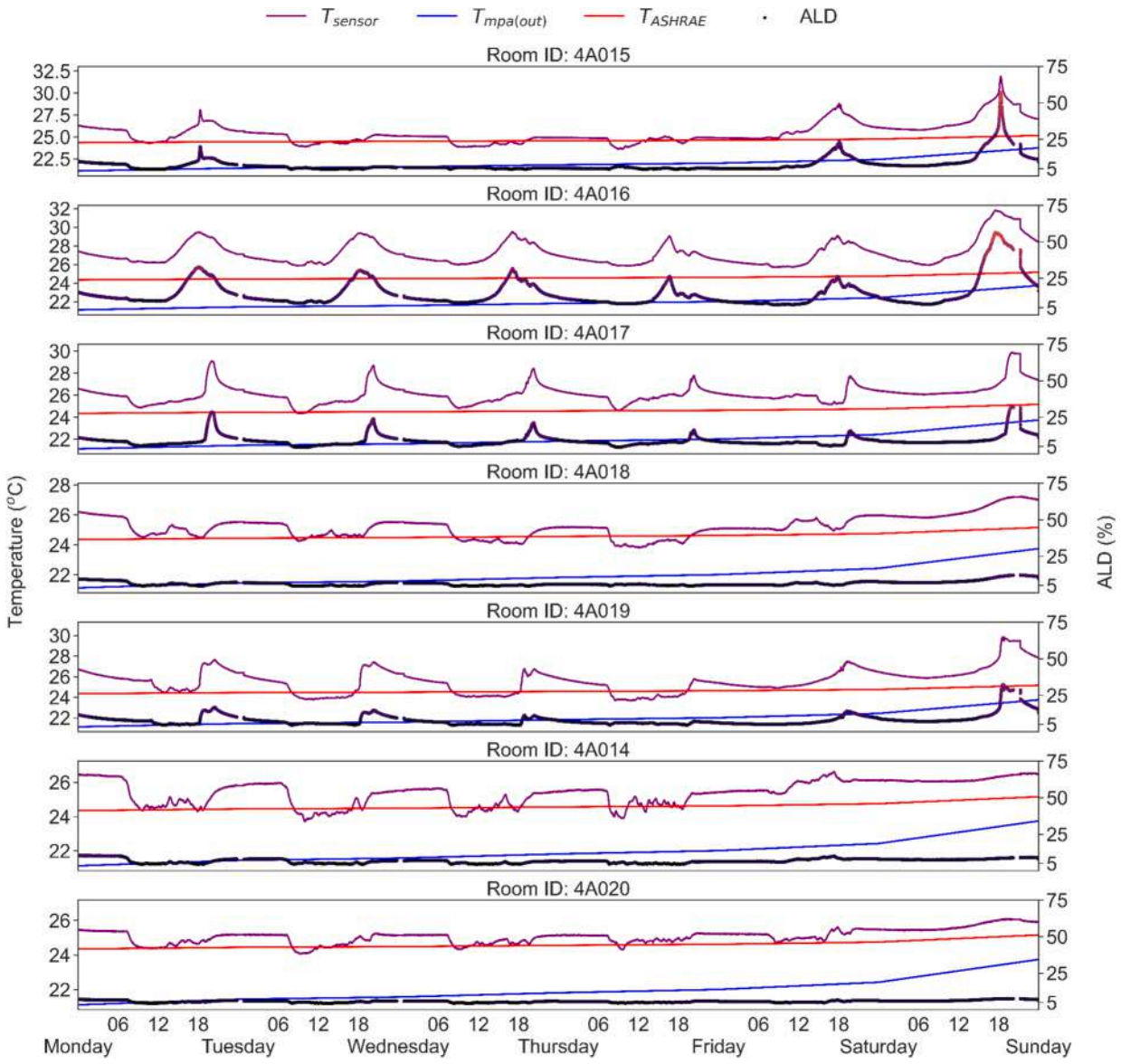

Fig. 23. Right-here right-now thermal comfort assessment for July 3rd to July 8th, 2023 (cooling season). Dry bulb air temperature (purple line), running mean outdoor air temperature (blue line), and ASHRAE operative temperature (red line). The scatter plot with inferno colormap (black-red) shows the magnitude of discomfort assessed by the ALD indicator. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Additionally, to address cyber security concerns and facilitate safe developer access, a Git repository was implemented. This provided a secure version-controlled environment where developers could retrieve monitored data, update scripts, and deploy new configurations without directly interfering with the operational infrastructure.

4.4.2. System Compatibility and data Standardization

The experiment involved interaction with multiple monitoring and control systems, each with its own syntax, naming conventions, data types, and units of measurement. These discrepancies led to integration challenges, making it difficult to interpret and process the collected data effectively. To overcome this, a coupling component was developed within the script, acting as an intermediary between CIRLEM and the various monitoring and control systems. This component standardized data formats, translated naming conventions, and ensured consistency in measurement units before feeding the data into the algorithm. By automating this process, CIRLEM was able to operate independently of the specific system configurations, improving scalability and reducing manual developers' intervention.

4.4.3. Occupants intervention challenges

One of the key challenges encountered in this research was the impact of occupant interventions on the control system's decision-making process. During the experiment, occupants of the living lab were allowed to adjust setpoints through control screens, replicating real-life conditions. As a result, setpoints could be modified at any time within a control timestep, creating discrepancies between the previously set value and the newly recorded one at the next monitoring point. This uncertainty complicated the learning process of the RL-based control algorithm, which relies on consistent state-action relationships. To mitigate this issue, the control algorithm first detects differences between the setpoints in the preceding timestep and the current timestep. Then, the algorithm uses the setpoint from the preceding timestep (set by the algorithm) to calculate the reward function, while the adjusted setpoint by the occupants at the current timestep serves as the reference for determining the proceeding timestep's setpoint, ensuring closer alignment with user preferences. Additionally, an updated reward function is under development to incorporate the newly adjusted setpoint as an indicator of user satisfaction. This approach allows occupant

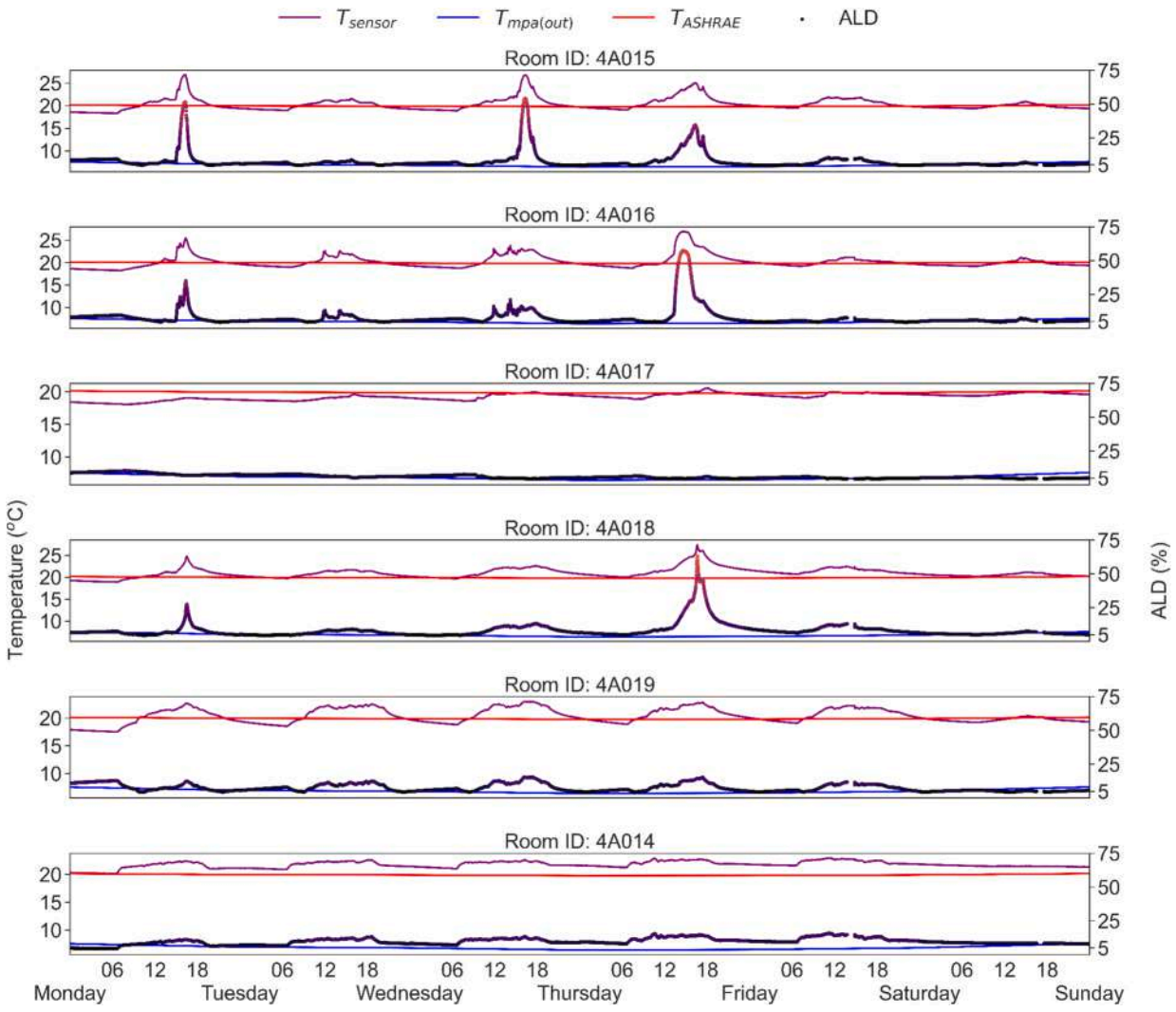

Fig. 24. Right-here right-now thermal comfort assessment for March 4th to March 10th, 2024. Dry bulb air temperature (purple line), running mean outdoor air temperature (blue line), and ASHRAE operative temperature (orange line). The scatter plot with inferno colormap (black-red) shows the magnitude of discomfort assessed by the ALD indicator. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

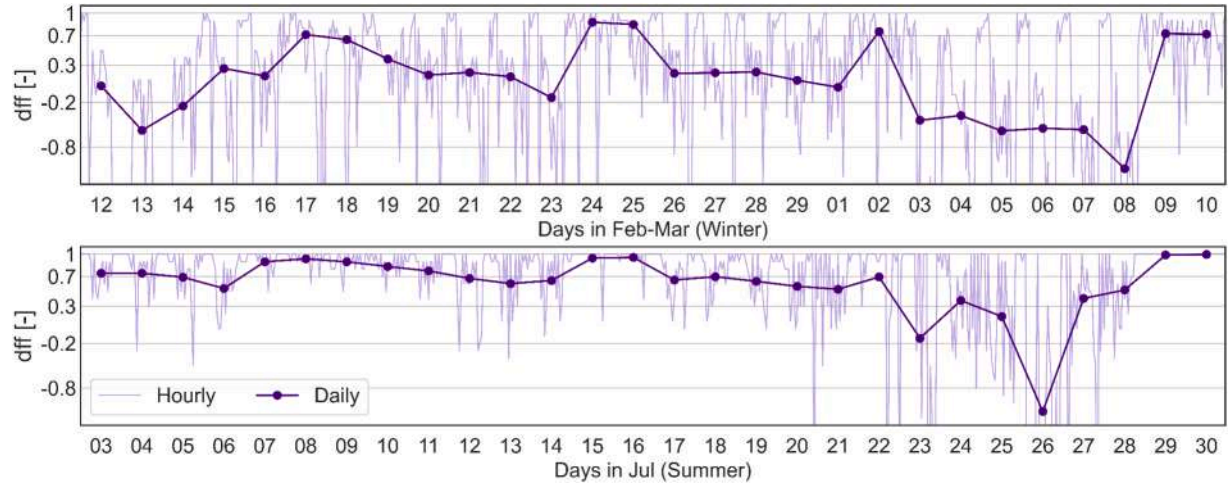

Fig. 25. Results for dff in hourly and daily timesteps for the data related to heating energy in winter (top) and cooling energy in summer (bottom). The performance categories are delineated with gray horizontal lines. The graphs are trimmed for dff < -1.3 to enhance the clarity of the graph.

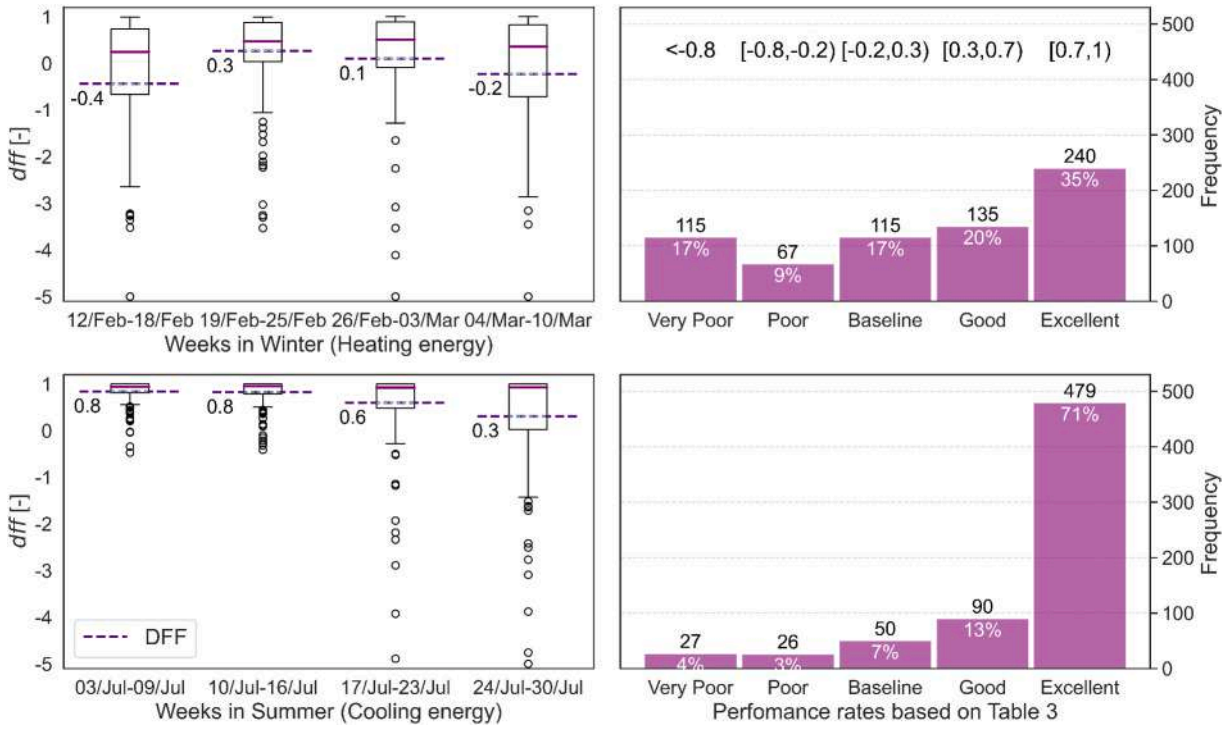

Fig. 26. The distribution of dff (boxplots) and values for DFF (dashed lines) in weekly periods during winter (top) and summer (bottom). The histogram presents the frequency of the values in five performance ranges.

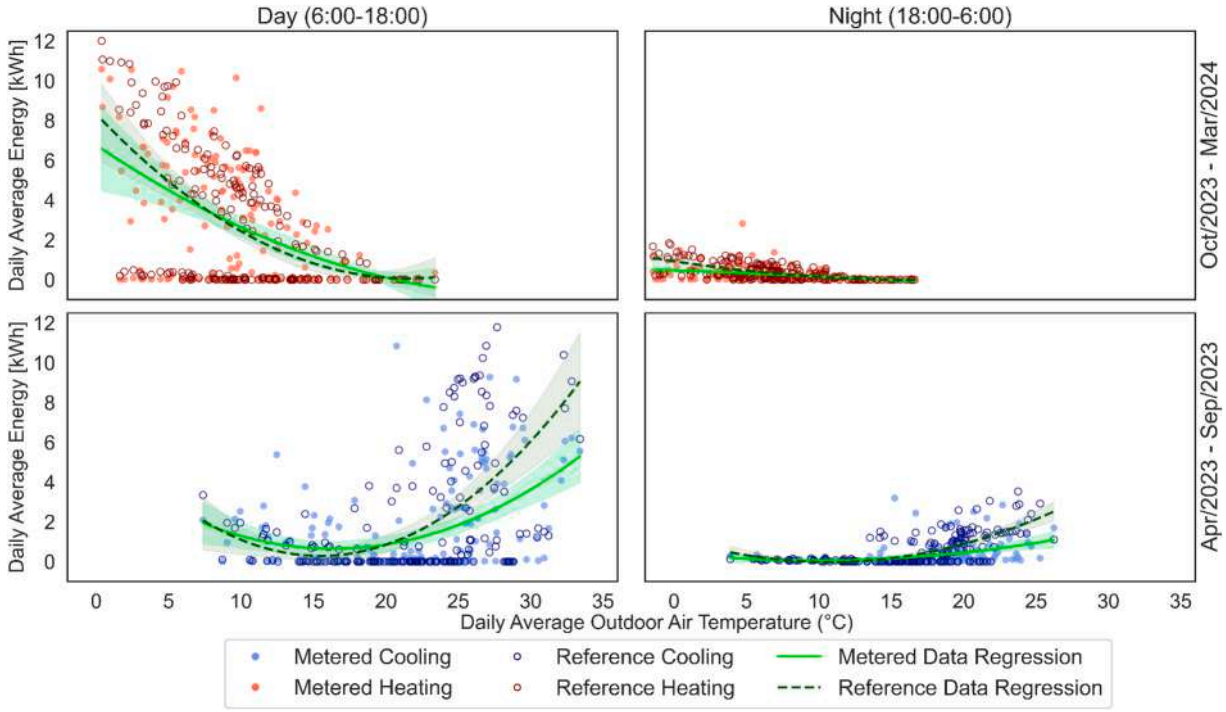

Fig. 27. Daily average heating (top) and cooling (bottom) energy use against daily average outdoor temperature divided into day (left) and night (right), the reference values are presented with dark circles.

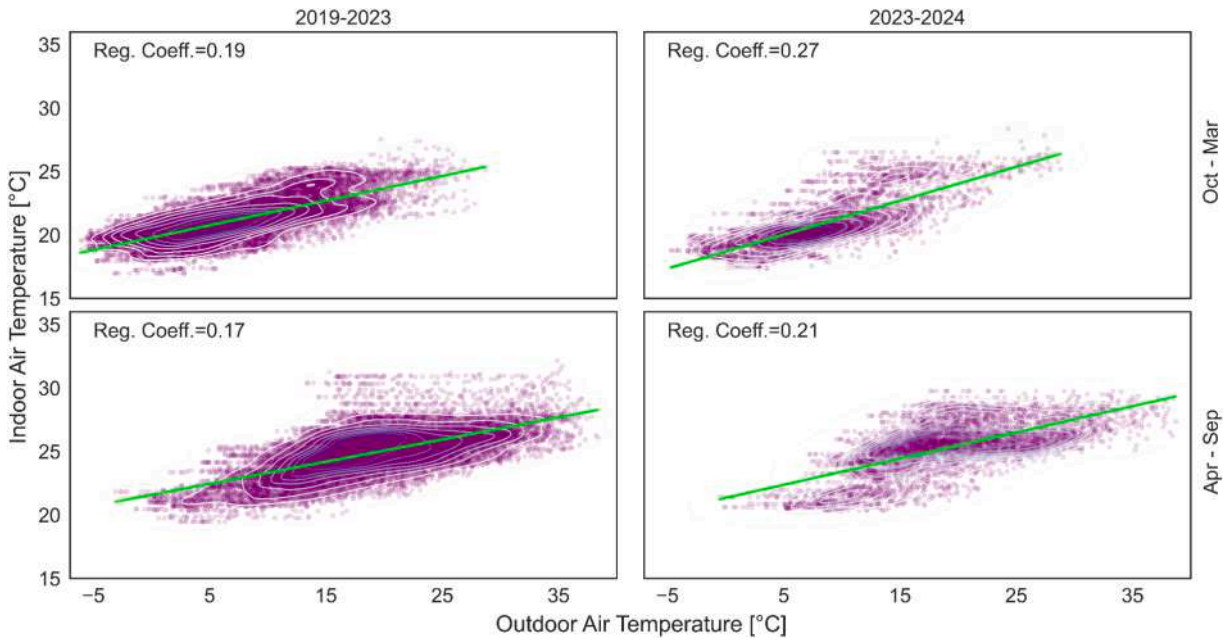

Fig. 28. Indoor against outdoor air temperature for heating (top) and cooling (bottom) seasons. The contour lines show the density of the data points. The data is related to the working hours from 8:00 to 16:00.

preferences to be integrated into the decision-making process in real time.

5. Conclusion

In the face of increasing energy complexity, climate uncertainties, and geopolitical disruptions, energy flexibility has become essential for effective urban energy management. Smart building solutions, facilitate real-time adaptability and collective energy behavior. Reinforcement Learning (RL) has emerged as a promising approach, enabling buildings to optimize energy use, integrate distributed energy resources, and participate in demand response programs. Despite its potential, challenges remain in communication, system diversity, and privacy when implementing multi-agent RL-based control. Addressing these barriers through decentralized decision-making and secure data handling is crucial for realizing scalable, intelligent energy management solutions that enhance grid stability, reduce energy costs, and improve overall system performance.

The CIRLEM framework, introduced in previous works, integrates Reinforcement Learning (RL) for energy management utilizing an online, value-based, model-free RL approach. To validate its real-world functionality, challenges with technical constraints, and impacts of occupants' intervention, the RL-based control is implemented in a living lab equipped with an advanced building management system at Grenoble Electrical Engineering Lab (G2ELab). The experiment is designed for one continuous year from April 2023.

During the experiment, the control algorithm runs every 15 min, retrieving data, making decisions, and sending optimized setpoints. Sensors and actuators allow the control algorithm to monitor conditions and implement necessary adjustments in the rooms. The living lab is equipped with air temperature and sensors and energy meters, while temperature and CO_2 level setpoints regulate energy use at the room level according to CIRLEM's generated values. A reference energy use profile is generated using the Multi-Layer Perceptron (MLP) method, incorporating historical energy use and weather conditions to provide a baseline for the algorithm's performance evaluation.

The metered energy use shows an 18 % reduction in the annual period as compared to the reference profile whereas heating and cooling energy show 5 % and 32 % reduction in each period, respectively. Peak power is curtailed by up 15 % and 50 % for heating and cooling energy use. Hourly scale analysis shows better performance of CIRLEM during warmer or colder conditions rather than moderate temperatures. The metered energy use in July 2023 shows significant reduction to a half compared to the reference profile.

Demand Flexibility Factor (DFF) is used to evaluate the performance of the algorithm regarding energy flexibility. Five categories are defined for the performance of the algorithm based on the values of DFF. Accordingly, the algorithm's performance is classified as Excellent level in 1-month analysis for 35 % of timesteps in winter and 71 % in summer.

Given the thermal comfort-oriented control design, experimental results obtained from both objective and subjective methods show that thermal comfort is maintained around the optimal level, examined both at building and room level, keeping occupants in a thermally satisfying indoor environment, while reducing building's energy use and providing energy flexibility to the grid.

As the experiment progressed, communication issues including delays, security, and access for the developers are solved. The possible communication and computation lags are considered with a 2-minute delay between monitoring and controlling loops. By using a git repository, developers could access the monitored data and update the scripts in a safe manner. Interacting with different control and monitoring systems caused issues related to the syntax of scripts which is solved by developing a coupling component in the script to interpret various naming, data type, and units to the one used in CIRLEM.

All in all, this experiment demonstrated the application of CIRLEM in building energy management to achieve energy efficiency and thermal comfort. The generated signal based on historical energy use enables CIRLEM to perform in a stand-alone building demonstrating the scalability of the control logic. Further studies are ongoing to enhance the decision-making process through the implication of user feedback and intervention of the setpoints in the reward function. The developer's platform, including the Git repository and the intermediate computer for secure access, is utilized for further developments of smart control algorithms.

CRediT authorship contribution statement

Mohammad Hosseini: Writing – original draft, Visualization,

Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Panayiotis Papadopoulos:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Muhammad Salman Shahid:** Writing – original draft, Visualization, Software, Resources, Methodology. **Kavan Javanroodi:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Silvia Erba:** Writing – review & editing, Validation. **Etta Grover-silva:** Writing – review & editing. **Peter Riederer:** Writing – review & editing, Project administration. **Amin Moazami:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **Frederic Wurtz:** Writing – review & editing, Resources. **Benoit Delinchant:** Writing – review & editing, Resources. **Salvatore Carlucci:** Writing – review & editing. **Mohammadreza Aghaei:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Vahid Nik:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

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.

Acknowledgements

This work is supported by the European Union's Horizon 2020 research and innovation programme under grant agreement for the COLLECTiEF - Collective Intelligence for Energy Flexibility project (grant agreement ID: 101033683).

Data availability

The data that has been used is confidential.

References

[1] A.T.D. Perera, T. Hong, Vulnerability and resilience of urban energy ecosystems to extreme climate events: a systematic review and perspectives, Renew. Sustain. Energy Rev. 173 (2023) 113038, https://doi.org/10.1016/j.rser.2022.113038.
[2] H. Banna, A. Alam, X.H. Chen, A.W. Alam, Energy security and economic stability: the role of inflation and war, Energy Econ. 126 (2023) 106949, https://doi.org/10.1016/j.eneco.2023.106949.
[3] M. Baranowski, Welfare over Warfare? Russia's War on Ukraine through the Prism of Europe's Energy Security, Int. J. Energy Econ. Policy 12 (2022) 226–231, https://doi.org/10.32479/ijeep.13415.
[4] A.H. Nebey, Recent advancement in demand side energy management system for optimal energy utilization, Energy Rep. 11 (2024) 5422–5435, https://doi.org/10.1016/j.egyr.2024.05.028.
[5] L. Toderean, T. Cioara, I. Anghel, E. Sarmas, V. Michalakopoulos, V. Marinakis, Demand response optimization for smart grid integrated buildings: review of technology enablers landscape and innovation challenges, Energ. Buildings 326 (2025) 115067, https://doi.org/10.1016/j.enbuild.2024.115067.
[6] M. Hosseini, S. Erba, P. Hajialigol, M. Aghaei, A. Moazami, V.M. Nik, Enhancing climate resilience in buildings using Collective Intelligence: a pilot study on a norwegian elderly care center, Energ. Buildings 308 (2024) 114030, https://doi.org/10.1016/j.enbuild.2024.114030.
[7] Basu S, S. E. Bale C, Wehnert T, Topp K. A complexity approach to defining urban energy systems. Cities 2019;95:102358. https://doi.org/10.1016/j.cities.2019.05.027.
[8] V. Vahidinasab, C. Ardalan, B. Mohammadi-Ivatloo, D. Giaouris, S.L. Walker, Active building as an energy system: concept, challenges, and outlook, IEEE Access 9 (2021) 58009–58024, https://doi.org/10.1109/ACCESS.2021.3073087.
[9] J.R. Vázquez-Canteli, Z. Nagy, Reinforcement learning for demand response: a review of algorithms and modeling techniques, Appl. Energy 235 (2019) 1072–1089, https://doi.org/10.1016/j.apenergy.2018.11.002.
[10] T. Wei, Y. Wang, Q. Zhu, in: Deep Reinforcement Learning for Building HVAC Control, Association for Computing Machinery, New York, NY, USA, 2017, pp. 1–6, https://doi.org/10.1145/3061639.3062224.
[11] Y. He, Z. Zhou, Y. Pan, F. Chong, B. Wu, K. Xiao, et al., Review of data security within energy blockchain: a comprehensive analysis of storage, management, and utilization, High-Confid. Comput. 4 (2024) 100233, https://doi.org/10.1016/j.hcc.2024.100233.
[12] K. Hu, M. Li, Z. Song, K. Xu, Q. Xia, N. Sun, et al., A review of research on reinforcement learning algorithms for multi-agents, Neurocomputing 599 (2024) 128068, https://doi.org/10.1016/j.neucom.2024.128068.
[13] S. Hosseini, P. Hajialigol, M. Aghaei, S. Erba, V. Nik, A. Moazami, Improving climate resilience and thermal comfort in a complex building through enhanced flexibility of the energy system, Int. Conf. Smart Energy Syst. Technol. (SEST) 2022 (2022) 1–6, https://doi.org/10.1109/SEST53650.2022.9898453.
[14] T. Wei, Y. Wang, Q. Zhu, in: Deep Reinforcement Learning for Building HVAC Control, Association for Computing Machinery, New York, NY, USA, 2017, pp. 1–6, https://doi.org/10.1145/3061639.3062224.
[15] Z. Wang, T. Hong, Reinforcement learning for building controls: the opportunities and challenges, Appl. Energy 269 (2020) 115036, https://doi.org/10.1016/j.apenergy.2020.115036.
[16] D. Zhu, B. Yang, Y. Liu, Z. Wang, K. Ma, X. Guan, Energy management based on multi-agent deep reinforcement learning for a multi-energy industrial park, Appl. Energy 311 (2022) 118636, https://doi.org/10.1016/j.apenergy.2022.118636.
[17] X. Li, C.-G. Lee, ΔV-learning: an adaptive reinforcement learning algorithm for the optimal stopping problem, Expert Syst. Appl. 231 (2023) 120702, https://doi.org/10.1016/j.eswa.2023.120702.
[18] V.M. Nik, M. Hosseini, CIRLEM: a synergic integration of Collective Intelligence and Reinforcement learning in Energy Management for enhanced climate resilience and lightweight computation, Appl. Energy 350 (2023) 121785, https://doi.org/10.1016/j.apenergy.2023.121785.
[19] M. Hosseini, A. Mazaheri, V.M. Nik, Enhancing the smart readiness of buildings: combining collective intelligence and reinforcement learning in building energy management, E3S Web of Conf 562 (2024) 10004, https://doi.org/10.1051/e3sconf/202456210004.
[20] K. Al Sayed, A. Boodi, R. Sadeghian Broujeny, K. Beddiar, Reinforcement learning for HVAC control in intelligent buildings: a technical and conceptual review, J. Build. Eng. 95 (2024) 110085, https://doi.org/10.1016/j.jobe.2024.110085.
[21] A. Shaqour, A. Hagishima, Systematic review on deep reinforcement learning-based energy management for different building types, Energies 15 (2022) 8663, https://doi.org/10.3390/en15228663.
[22] Z. Nagy, G. Henze, S. Dey, J. Arroyo, L. Helsen, X. Zhang, et al., Ten questions concerning reinforcement learning for building energy management, Build. Environ. 241 (2023) 110435, https://doi.org/10.1016/j.buildenv.2023.110435.
[23] K. Kurte, J. Munk, O. Kotevska, K. Amasyali, R. Smith, E. McKee, et al., Evaluating the adaptability of reinforcement learning based HVAC control for residential houses, Sustainability 12 (2020) 7727, https://doi.org/10.3390/su12187727.
[24] M. Schmidt, M.V. Moreno, A. Schülke, K. Macek, K. Mařík, A.G. Pastor, Optimizing legacy building operation: the evolution into data-driven predictive cyber-physical systems, Energ. Buildings 148 (2017) 257–279, https://doi.org/10.1016/j.enbuild.2017.05.002.
[25] K. Nweye, S. Sankaranarayanan, Z. Nagy, MERLIN: multi-agent offline and transfer learning for occupant-centric operation of grid-interactive communities, Appl. Energy 346 (2023) 121323, https://doi.org/10.1016/j.apenergy.2023.121323.
[26] W. Zhang, Y. Yu, Z. Yuan, P. Tang, B. Gao, Data-driven pre-training framework for reinforcement learning of air-source heat pump (ASHP) systems based on historical data in office buildings: field validation, Energ. Buildings 332 (2025) 115436, https://doi.org/10.1016/j.enbuild.2025.115436.
[27] A. Silvestri, D. Coraci, S. Brandi, A. Capozzoli, E. Borkowski, J. Köhler, et al., Real building implementation of a deep reinforcement learning controller to enhance energy efficiency and indoor temperature control, Appl. Energy 368 (2024) 123447, https://doi.org/10.1016/j.apenergy.2024.123447.
[28] A. Silvestri, D. Coraci, S. Brandi, A. Capozzoli, A. Schlueter, Practical deployment of reinforcement learning for building controls using an imitation learning approach, Energ. Buildings 335 (2025) 115511, https://doi.org/10.1016/j.enbuild.2025.115511.
[29] R. Mokhtari, M. Montazeri, H. Cai, P. Heer, R. Li, Price-responsive control using deep reinforcement learning for heating systems: simulation and living lab experiment, Energy (2025) 138517, https://doi.org/10.1016/j.energy.2025.138517.
[30] Z. Zhang, A. Chong, Y. Pan, C. Zhang, K.P. Lam, Whole building energy model for HVAC optimal control: a practical framework based on deep reinforcement learning, Energ. Buildings 199 (2019) 472–490, https://doi.org/10.1016/j.enbuild.2019.07.029.
[31] L. Kannari, N. Wessberg, S. Hirvonen, J. Kantorovitch, S. Paiho, Reinforcement learning for control and optimization of real buildings: identifying and addressing implementation hurdles, J. Build. Eng. 104 (2025) 112283, https://doi.org/10.1016/j.jobe.2025.112283.
[32] A. Lavin, C.M. Gilligan-Lee, A. Visnjic, S. Ganju, D. Newman, S. Ganguly, et al., Technology readiness levels for machine learning systems, Nat. Commun. 13 (2022) 6039, https://doi.org/10.1038/s41467-022-33128-9.
[33] Collective Intelligence for Energy Flexibility | COLLECTiEF Project | Fact Sheet | H2020. CORDIS | European Commission 2024. https://cordis.europa.eu/project/id/101033683 (accessed January 27, 2025).
[34] M. Aghaei, A. Moazami, S. Erba, M. Hosseini, I.A.C. Avendano, M.-S. Shahid, et al., Collective intelligence for energy flexibility – collectief: an EU H2020 project for enhancing energy efficiency and flexibility in existing buildings, Int. Conf. Fut. Energy Solut. (FES) 2023 (2023) 1–6, https://doi.org/10.1109/FES57669.2023.10182779.
[35] N.K. Twum-Duah, M. Amayri, S. Ploix, F. Wurtz, A comparison of direct and indirect flexibilities on the self-consumption of an office building: the case of predis-MHI, a smart office building, Front. Energy Res. 10 (2022), https://doi.org/10.3389/fenrg.2022.874041.

[36] F. Wurtz, B. Delinchant, “Smart buildings” integrated in “smart grids”: a key challenge for the energy transition by using physical models and optimization with a “human-in-the-loop” approach, C. R. Phys. 18 (2017), https://doi.org/10.1016/j.crhy.2017.09.007.
[37] Boustie N. Inauguration de la centrale photovoltaïque de GreEn-ER : 1re expérience d’auto-consommation collective dans une université. Université Grenoble Alpes 2023. https://www.univ-grenoble-alpes.fr/inauguration-de-la-centrale-photovoltaique-de-green-er-1re-experience-d-auto-consommation-collective-dans-une-universite-1226307.kjsp?RH=1680617282213 (accessed November 16, 2024).
[38] Delinchant B, Wurtz F, Ploix S, Schanen J-L, Marechal Y. GreEn-ER Living Lab - A Green Building with Energy Aware Occupants, 2024, p. 316–23.
[39] R. Sutton, A. Barto, Reinforcement Learning: an introduction, The MIT Press, 2018.
[40] A.T.D. Perera, P. Kamalaruban, Applications of reinforcement learning in energy systems, Renew. Sustain. Energy Rev. 137 (2021) 110618, https://doi.org/10.1016/j.rser.2020.110618.
[41] M. Hosseini, S. Erba, A. Mazaheri, A. Moazami, V.M. Nik, From flexible building to resilient energy communities: a scalable decentralized energy management scheme based on collaborative agents, Energ. Buildings 337 (2025) 115651, https://doi.org/10.1016/j.enbuild.2025.115651.
[42] Standards European. BS EN 16798-1:2019 Energy performance of buildings. Ventilation for buildings Indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality, thermal environment, lighting and acoustics. Module M1-6. Https://WwwEn-StandardEu 2019. https://www.en-standard.eu/bs-en-16798-1-2019-energy-performance-of-buildings-ventilation-for-buildings-indoor-environmental-input-parameters-for-design-and-assessment-of-energy-performance-of-buildings-addressing-indoor-air-quality-thermal-environment-lighting-and-acoustics-module/ (accessed November 2, 2023).
[43] ANSI, ASHRAE. Standard 55 – Thermal Environmental Conditions for Human Occupancy 2023.
[44] K.J. McCartney, N.J. Fergus, Developing an adaptive control algorithm for Europe, Energ. Buildings 34 (2002) 623–635, https://doi.org/10.1016/S0378-7788(02)00013-0.
[45] S. Attia, S. Carlucci, Impact of different thermal comfort models on zero energy residential buildings in hot climate, Energ. Buildings 102 (2015) 117–128, https://doi.org/10.1016/j.enbuild.2015.05.017.
[46] S. Carlucci, S. Erba, L. Pagliano, R. de Dear, ASHRAE Likelihood of Dissatisfaction: a new right-here and right-now thermal comfort index for assessing the Likelihood of dissatisfaction according to the ASHRAE adaptive comfort model, Energ. Buildings 250 (2021) 111286, https://doi.org/10.1016/J.ENBUILD.2021.111286.
[47] Y. Xu, F. Li, A. Asgari, Prediction and optimization of heating and cooling loads in a residential building based on multi-layer perceptron neural network and different optimization algorithms, Energy 240 (2022) 122692, https://doi.org/10.1016/j.energy.2021.122692.
[48] O.A. Alawi, H.M. Kamar, Z.M. Yaseen, Optimizing building energy performance predictions: a comparative study of artificial intelligence models, J. Build. Eng. 88 (2024) 109247, https://doi.org/10.1016/j.jobe.2024.109247.
[49] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., Scikit-learn: machine Learning in Python, J. Mach. Learn. Res. 12 (2012).
[50] Q. Yin, C. Han, A. Li, X. Liu, Y. Liu, A Review of research on building energy consumption prediction models based on artificial neural networks, Sustainability 16 (2024) 7805, https://doi.org/10.3390/su16177805.
[51] R. Olu-Ajayi, H. Alaka, I. Sulaimon, F. Sunmola, S. Ajayi, Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques, J. Build. Eng. 45 (2022) 103406, https://doi.org/10.1016/j.jobe.2021.103406.
[52] M. Hosseini, K. Javanroodi, V.M. Nik, High-resolution impact assessment of climate change on building energy performance considering extreme weather events and microclimate – investigating variations in indoor thermal comfort and degree-days, Sustain. Cities Soc. 78 (2022) 103634, https://doi.org/10.1016/j.scs.2021.103634.
[53] Lamprecht CS. Meteostat Python n.d.
[54] V.M. Nik, A. Moazami, Using collective intelligence to enhance demand flexibility and climate resilience in urban areas, Appl. Energy 281 (2021), https://doi.org/10.1016/j.apenergy.2020.116106.
[55] M. Kordos, A. Rusiecki, Reducing noise impact on MLP training, Soft. Comput. 20 (2016) 49–65, https://doi.org/10.1007/s00500-015-1690-9.
[56] V. Hodge, J. Austin, A Survey of Outlier Detection Methodologies, Artif. Intell. Rev. 22 (2004) 85–126, https://doi.org/10.1023/B:AIRE.0000045502.10941.a9.
[57] S.J. Won, X.H. Wang, H.E. Warren, Climate normals and weather normalization for utility regulation, Energy Econ. 54 (2016) 405–416, https://doi.org/10.1016/j.eneco.2015.12.016.
[58] S. Beheshti, A. Sahebalam, E. Nidoy, Structure dependent weather normalization, Energy Sci. Eng. 7 (2019), https://doi.org/10.1002/ese3.272.
[59] Portfolio Manager Technical Reference: Climate and Weather | ENERGY STAR 2024.
[60] S. Weglarczyk, Kernel density estimation and its application, ITM Web Conf. 23 (2018) 00037, https://doi.org/10.1051/itmconf/20182300037.