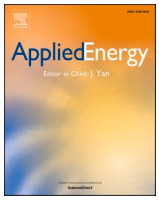
Applied Energy 350 (2023) 121785

Contents lists available at ScienceDirect

Applied Energy

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



CIRLEM: a synergic integration of Collective Intelligence and Reinforcement learning in Energy Management for enhanced climate resilience and lightweight computation

Vahid M. Nik [a,b,*], Mohammad Hosseini [c]

[a] Division of Building Physics, Department of Building and Environmental Technology, Lund University, SE-223 63 Lund, Sweden
[b] CIRCLE – Centre for Innovation Research, Lund University, Box 118, 221 00 Lund, Sweden
[c] Department of Ocean Operations and Civil Engineering, Faculty of Engineering, NTNU Norwegian University of Science and Technology, Ålesund, Norway

HIGHLIGHTS

- Integrating Collective intelligence (CI) and Reinforcement Learning (RL) to form a novel energy management (EM) system called CIRLEM.
- Two levels of control and decision making are tested at the edge node and cluster levels.
- Novel approaches are developed for defining flexibility signals and the agent's policy.
- CIRLEM becomes a lightweight algorithm that converges quickly.
- CIRLEM improves the autonomy in absorbing shocks effectively through distributed intelligence.

ARTICLE INFO

Keywords:
Energy management
Reinforcement Learning
Collective intelligence
Extreme climate
Energy flexibility
Climate Resilience

ABSTRACT

A novel energy management (EM) approach is introduced, integrating core elements of collective intelligence (CI) and reinforcement learning (RL) and called CIRLEM. It operates by distributing a flexibility signal from the energy supplier to agents within the grid, prompting their responsive actions. The flexibility signal reflects upon the collective behaviour of the agents in the grid and agents learn and decide using a value-based model-free RL engine. Two ways of running CIRLEM are defined, based on doing all the decision making only at the edge node (Edge Node Control or ENC) or together with the cluster (Edge node and Cluster Control or ECC). CIRLEM's performance is thoroughly investigated in an elderly building situated in Ålesund, Norway, specifically during extreme warm and cold seasons in the future climate. The building is divided into 20 thermal zones, each acting as an agent with three control strategies. CIRLEM undergoes comprehensive testing, evaluating policies with 24 and 48 sets of actions (referred to as L24 and L48) and six different randomness levels. The results demonstrate that CIRLEM swiftly converges to an optimal solution (the optimum set of policies), offering both enhanced indoor comfort and significant energy savings. Among the CIRLEM algorithms, ENC-L24, the fastest and simplest one, showcased outstanding performance. Overall, CIRLEM offers a remarkable improvement in energy flexibility and climate resilience for a group of grid-connected agents, ensuring energy savings without compromising indoor comfort.

1. Introduction

Climate change intensifies climate variations which can result in stronger and more frequent extreme events [1]. The Scandinavian cities have been recently experiencing more extreme climate events, e.g. frequent warm summers with excessive heatwaves since 2018 [2],

Abbreviations: BMS, Building Management System; CI, Collective Intelligence; CI-DSM, Collective Intelligence-based Demand-Side Management; CMIP5, Coupled Model Intercomparison Project 5; DSM, Demand-Side Management; DSO, Distribution System Operators; ECC, Edge node and Cluster Control; ECY, Extreme Cold Year; ENC, Edge Node Control; EWY, Extreme Warm Year; HVAC, Heating, Ventilation, Air Conditioning; ICT, Information and Communication Technology; MDP, Markov Decision Process; ML, Machine Learning; MPC, Model Predictive Control; RL, Reinforcement Learning; TDY, Typical Downscaled Year.

* Corresponding author at: Division of Building Physics, Department of Building and Environmental Technology, Lund University, SE-223 63 Lund, Sweden.
E-mail address: vahid.nik@byggtek.lth.se (V.M. Nik).

https://doi.org/10.1016/j.apenergy.2023.121785
Received 31 December 2022; Received in revised form 30 July 2023; Accepted 16 August 2023
Available online 23 August 2023

increased morbidity and mortality rates [3], and cold snaps such as 2021 with a record breaking cold December during the last fifty years, leading to a dramatic rise in energy prices and transport system disruption [4]. Experiencing other extreme events such as COVID-19 and Ukraine crisis, explains how fragile the situation can become and how rapid the impacts can propagate in different sectors and at different scales.

Given the significance of human comfort and health, the impacts of climate change at the building scale are likely to gain increasing importance in the future. This is particularly true for countries where buildings are not adequately designed to withstand extreme weather events. For example, there is a considerable relationship between extreme hot and cold weather events in Sweden, associated with cardiovascular and respiratory hospitalizations [5]. Heat waves can significantly increase all-cause mortality and mortality caused by coronary heart disease, by approximately 10% and 15%, respectively [6]. In Norway, data were collected on hospital visits and general practitioners by elderly (<70 years) in Oslo, indicating a significant correlation between increased patient admissions during summer months with extreme warm conditions [7]. Different studies point to the increased consequences and death rates among vulnerable groups such as elderly and people with pre-existing medical conditions and lack of mobility [8–10]. Climate change also impacts the release of biological airborne allergens, such fungal spores and plant pollen which will likely have a significant impact on the prevalence of allergic respiratory diseases, such as asthma [11]. The significance of indoor comfort has been underscored in the wake of the COVID-19 crisis, sparking discussions about the necessity for improved standards and ventilation strategies [12,13]. This becomes even more critical in conjunction with climate change, as extreme weather events can exacerbate indoor comfort issues [14], heighten the risk of mould growth [15], and lead to increased hospitalizations [5]. It is expected that the role of energy and HVAC systems in providing indoor comfort and good air quality becomes more important in the future since extreme events force people to stay longer inside buildings with controlled indoor environments [16]. Consequently there will be a bigger need for active energy-consuming solutions in the future, leading to higher energy demand.

Proper climate change adaptation of urban areas is crucial, and the role of buildings and energy systems are very important in this regard. In combination with increased urbanization, extreme climate events put larger loads on urban energy systems, meanwhile increasing the failure risk of energy systems and critical infrastructures [17,18], which can also result in cascading failures [19]. Conditions can get worse when energy supply is largely dependent on renewable generation, due to its intermittent characteristics [20]. Resilient energy solutions are needed, not only to cope with extreme climate events, but also to support the transition towards carbon neutrality [18]. They should account for multi-sectoral impacts and cascading failures, considering users, their preferences and comfort [21].

Although the scientific community have developed several methods and models to design and control energy systems, there exist major gaps to evaluate and enhance the climate resilience and flexibility of energy systems, especially in connection to complex urban environments and uncertainties. As discussed in a review work [18], the topic of climate resilience of energy solutions is quite immature and suffers from its loose connection to climate change modelling and not accounting for urban complexities. In our previous works, we have contributed to the field by introducing new approaches to account for thermal comfort [22], microclimate [23], energy price and CO2 mitigation [24], interconnected infrastructures [17], and resilient energy system design [25] in connection to climate change. When moving towards a finer temporal and spatial resolution in the analysis, e.g. assessing hourly energy demand and indoor thermal comfort in single buildings, the role of building control strategies become stronger. In this regard, topics of energy management (EM), demand response (DR) and Demand Side Management (DSM) are very relevant when the behaviour of buildings and urban energy systems together are discussed. In a previous work, we presented a DSM approach based on Collective Intelligence (CI), calling in CI-DSM, showing that it can increase the energy flexibility and climate resilience in urban areas [26]. Nevertheless, a significant challenge lies in identifying optimal building controls for a sizable cluster of interconnected buildings, where each structure has an impact on the others. The reality is complex and more enhanced approaches are needed when the number of buildings and control strategies increase, especially when preparing for future extreme and unprecedented events [27]. As discussed in [26], the implementation of DSM has been lagging behind due to increased complexity of the system operation [28] and the need for expensive ICT solutions and implying privacy and security risks [29]. There is a need for simpler holistic solutions [30], therefore simpler approaches and lighter algorithms are needed for the successful control of multi-variant extended systems such as networks of buildings and energy grids. To this end, implementation of Reinforcement Learning (RL) has shown promising results [31,32], especially the model-free RL approaches (e.g. Q-learning) [24].

A major limitation in the current state of the art in implementing RL (or ML) for energy calculations is the oversimplification of energy problems [33,34] and there is a big need to develop RL-based methods for real building applications to accelerate training and enhance control robustness, especially with bigger participation of multiple agents with different priorities [32]. Robust and reliable RL policies are needed to address environmental shifts and mismatched configurations [34]. RL algorithms with reduced variance are needed to control multi-agent systems in non-stationary environments [31,35]. Addressing this gap is crucial to effectively adapt to the escalating frequency and intensity of climate variations caused by climate change. To enhance energy flexibility and climate resilience within urban energy systems, it is imperative to develop methods that empower energy management in this context. This work significantly focuses on this pivotal aspect, presenting a novel approach to increase the energy flexibility and climate resilience in the energy network.

This work is designed to integrate RL-based decision making into the previously suggested CI-DSM [26] to enhance the performance of the energy management system (focusing on demand response) in coping with uncertainties and variations of the outdoor climate, focusing on extreme climate conditions. The novel EM based on combining CI and RL is called CIRLEM and this work explains the theory and results for implementing that at the building scale. The performance of CIRLEM is investigated for an elderly building in Ålesund, Norway during one extreme warm summer and one extreme cold winter for future climate. The building is divided into 20 separate zones with active controls for heating and cooling set-points, ventilation, and appliances. Multiple approaches for controlling the buildings are investigated, which differ in the level of control and data sharing, strategies for setting policies and randomness. In the following, Section 2 explains the background and methodology of CIRLEM, Section 3 presents the results and their assessment, and conclusions are discussed in Section 4.

2. Background and methodology

As mentioned, the present work is shaped around integrating RL-based decision making into a previously presented DSM approach based on collective intelligence [26]. This section is divided into four subsections, explaining how we reached from CI-DSM to CIRLEM and introducing the case study.

2.1. The need for moving from CI-DSM to CIRLEM

DSM refers to the set of means to change the pattern and/or magnitude of energy use, which usually appear as a set of actions and strategies to reduce, increase, or reschedule the demand [29]. DSM plays an important role in increasing the flexibility of energy systems [36]. A higher flexibility helps to increase the share of distributed generation, resulting in a higher reliability and carbon reduction of energy solutions

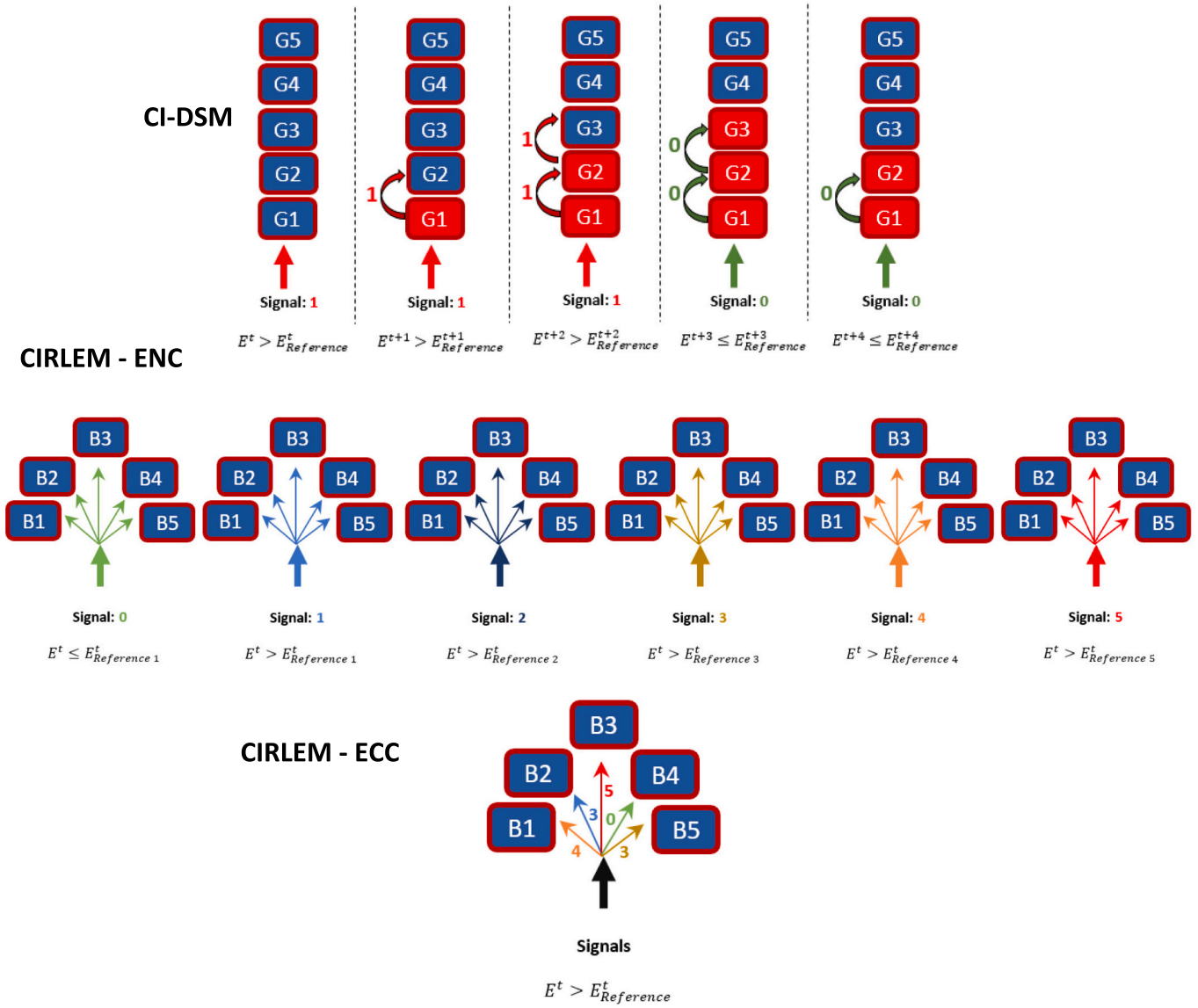


Fig. 1. Schematic presentation of the communication strategies for CI-DSM and CIRLEM.

without heavy network investments [28]. DR plays an important role in DSM by increasing the demand flexibility [31], helping the energy users to go for more economic and greener solutions [37]. We introduced a DSM approach based on Collective intelligence, calling it CI-DSM [26]. CI is a form of universally distributed intelligence, working based on collaborative problem solving and decision making [38]. The CI-based systems are identified by their robustness, flexibility, and scalability. They can organize themselves autonomously and adapt to unknown environments, showing emergent behaviours that enhance their ability to cope with uncertain conditions [39]. Since CI-DSM is based on distributed intelligence, the data/privacy security considerably increases compared to traditional DSMs.

By analysing the performance of CI-DSM for extreme climate conditions in Stockholm, we showed that it can enhance the flexibility of an energy system and consequently make it more resilient against environmental variations or external shocks [26]. Running the energy simulations and applying the adaptation measure (or control strategy) was a time consuming process in the Stockholm case study, although the only adaptation measure was changing the set-point temperature for the whole building. This is while buildings usually have multiple control strategies, so reaching an optimum control can become computationally demanding. Moreover, there exist several other influencing factors and indicators, such as user behaviour/comfort and appliances, which all can affect the energy performance of buildings. Reaching an optimum control strategy in the existence of multiple influencing factors can become very challenging, both when simulating and controlling buildings in connection to the energy grid. We need lighter algorithms to be implemented in energy management systems, enabling a big number of agents (with multiple control strategies) to communicate and collaborate in a complex environment. In this regard, RL-based methods have shown substantial potential in resolving increasing complexities within the energy domain, considering both supply and demand [34,40]. In CIRLEM, RL is implemented to use the global CI-based knowledge for optimum decision making at two levels, as explained in the following.

2.2. *Flexibility signal and collective behaviour in the grid*

The need for flexibility is usually transferred from the energy provider to the consumer through the energy network, which is called flexibility signal hereafter. In many cases, the price signal is also used as the flexibility signal, affected by the market and bidding strategies that govern the energy grid/market. In CI-DSM, the flexibility signal was 0 or

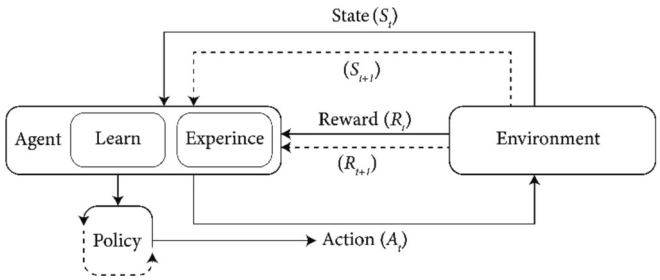


Fig. 2. Overall interaction in an RL-based system between the agent and the environment.

1, which 1 indicates the need for flexibility, asking buildings (or agents) to lower their energy demand by adopting their adaptation measures (or changing their control settings), which was only changing the indoor set-point temperature in [26]. Signal 0 allowed lifting the adaptation measures or continuing the control settings as is. Signal 1 was generated whenever the energy demand during operation time (or during extreme weather conditions) was higher than the reference values (which was considered as the energy demand at that time during typical weather conditions). The flexibility signal was transferring from the first group of buildings to the second and so on per time step, engaging one group of buildings per time step. Lifting the adaptation strategies was starting from the last group; the last to engage would be the first to release. In Fig. 1, the figure on top explains the CI-DSM communication strategy (for more details check [26]). Thanks to the simple communication approach (only 0 or 1 signals), the size and need of data transfer/storage decreases enormously (cheap ICT solutions).

In CIRLEM, the flexibility signal is a signal between 0 and 5, which 0 means there is no need for flexibility and 5 asks for the maximum possible flexibility in the grid. The approach of sending the signal can differ depending on the intelligence and control level at the energy supply or cluster side (which can be interpreted as Distribution System Operators or DSO). Higher control access often corresponds to increased data transfer while compromising data protection and user privacy. In this work, we are assessing two control strategies: 1) Edge Node Control (ENC), and 2) Edge node and Cluster Control (ECC). In ENC, the need for data transfer between buildings and the energy provider is at a minimum level (or keeping the user privacy at a maximum level). Meanwhile, there is no need for a high computational power at the cluster level since only one flexibility signal is transferred to the whole grid per time step (e.g. every 15 min) without any decision making at this level, as is visualized in Fig. 1. Although all buildings (or agents in the grid) receive the same signal per time step in ENC, they can respond differently depending on their capability and preferences. In ECC, the privacy measures at the building level are less strict than ENC and the computational power is stronger on the supply side. This creates the opportunity to optimize the distribution of signals for the buildings connected to

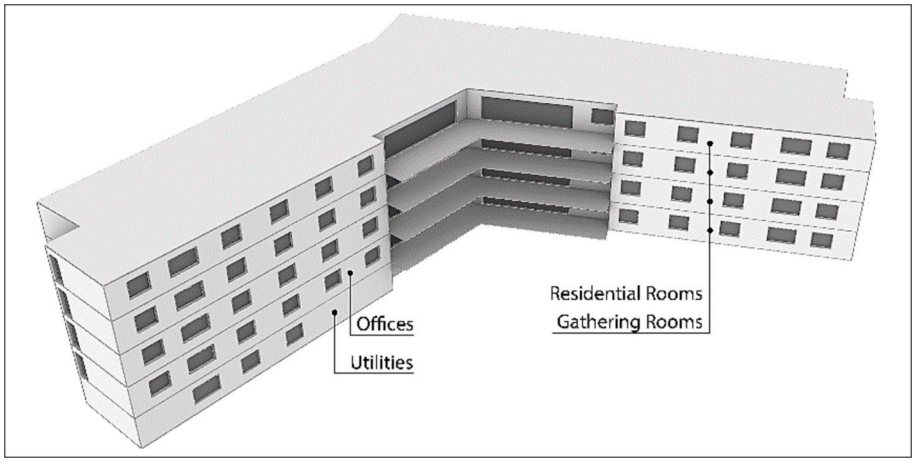


Fig. 3. The 3D model of the Eidet building.

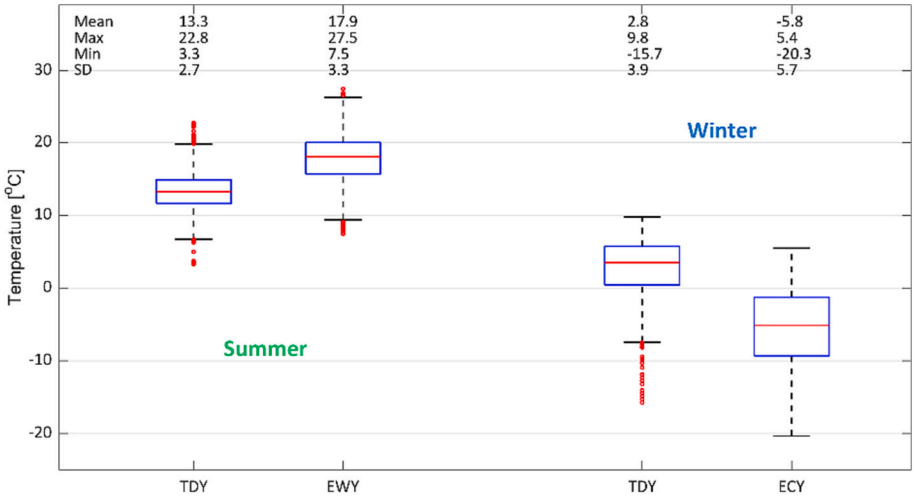


Fig. 4. The 3D model of the Eidet building.

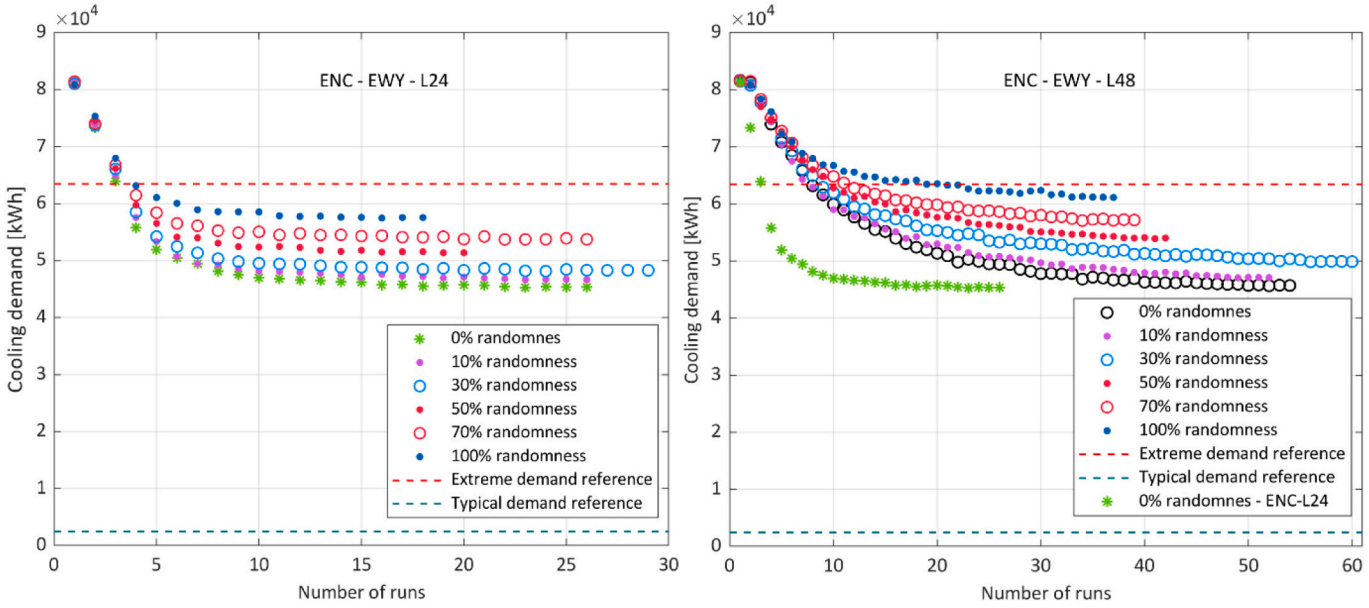


Fig. 5. Comparing the convergence speed and energy saving of algorithms with different randomness levels (0% to 100%), action library size (left: 24 and right: 48) for cooling demand during EWY summer.

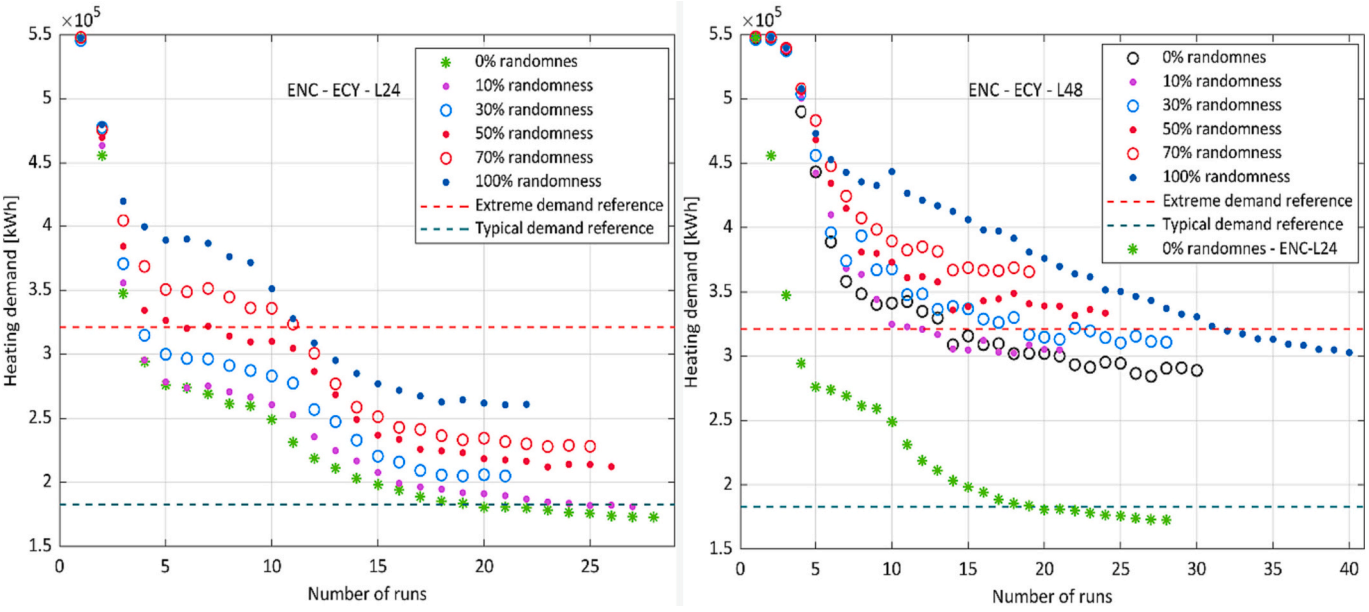


Fig. 6. Comparing the convergence speed and energy saving of algorithms with different randomness levels (0% to 100%), action library size (left: 24 and right: 48) for cooling demand during ECY winter.

the grid and transfer specific signals to different buildings as shown in Fig. 1. In both ENC and ECC, the energy supplier/distributor does not know about the specific control actions taking place within buildings. In ECC the distributor understands the impact of single buildings/agents in the network, while such knowledge is not needed in ENC.

Having a set of signals (0–5) and adaptation measures in buildings implies a big difference between CI-DSM and CIRLEM, making the decision making procedure more advanced (selecting the optimum adaptation measures per buildings depending on the signal), especially considering the connection of many buildings with different demand profiles and adaptation strategies to the grid [41,42]. This is where RL plays an important role and helps to reach an optimum solution. In other words, RL enables the buildings/agents to reach a collective behaviour in response to the flexibility signal, helping to increase climate resilience and safely pass extreme events.

2.3. Integrating RL into decision-making

In a CI-based system, agents encounter repeated tasks and situations where learning from the past behaviours and consequences could improve the coming collective behaviours and achievements [43]. Moreover, agents can deploy RL to improve their individual and the whole system performance [44]. RL is a machine learning training method based on rewarding desired behaviours of (an) agent(s) [32] which can be also interpreted as the process of learning to act optimally in an environment through experience [34]. RL is a game between an agent and an environment, where the environment is represented by a Markov Decision Process (MDP) [34], formed by a quadruple (S, A, P,

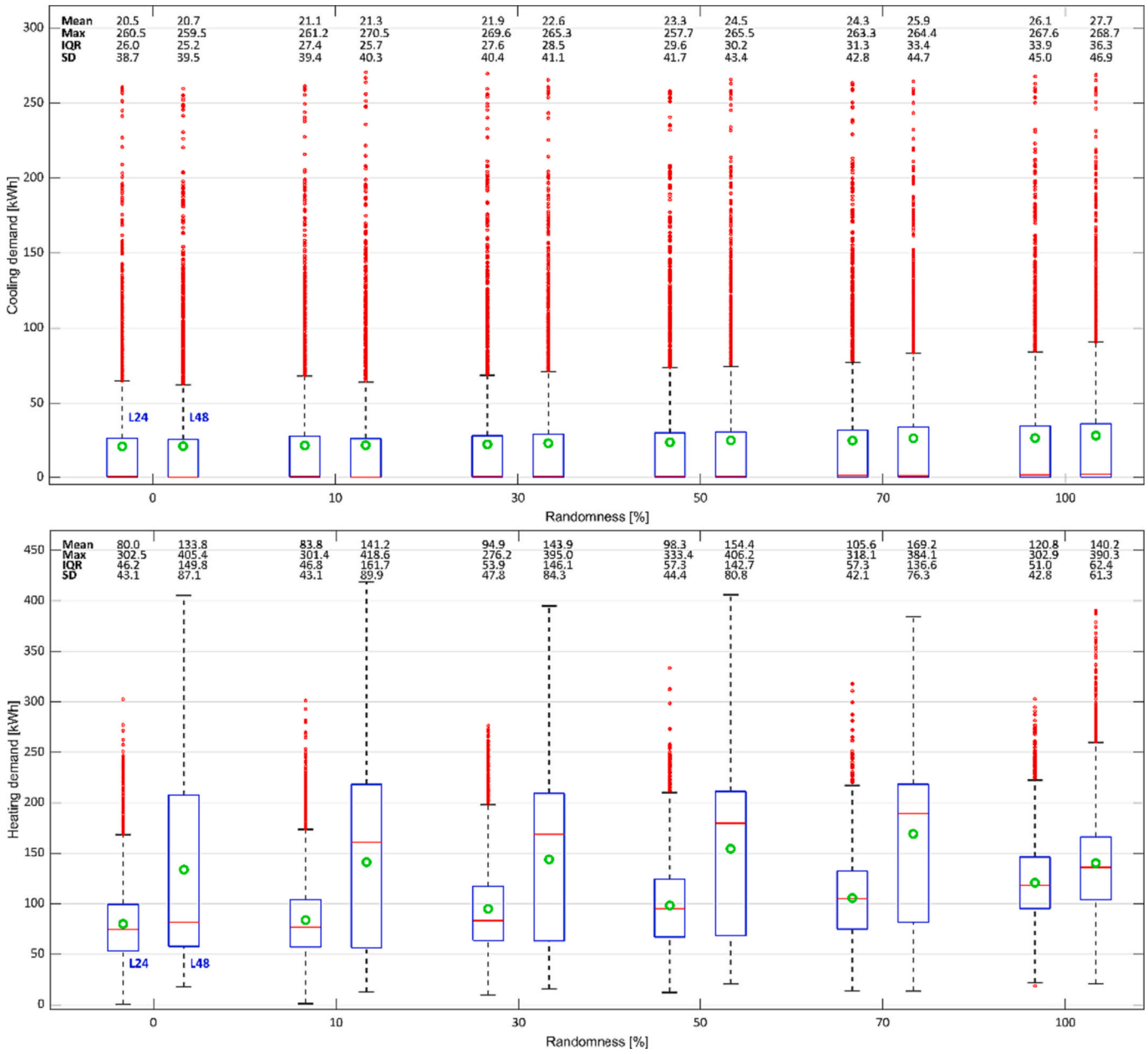


Fig. 7. Comparing the distribution of the hourly (top) cooling demand in EWY summer and (bottom) heating demand in ECY winter.

R), representing State, Action, Environment and Reward [32]. In an MDP environment, the current state characterizes the whole process completely [45] saying that each state depends on the preceding states. Additionally, all the parameters in the quadruple values of (S, A, P, R) necessarily have finite numbers of elements, and actions and states interact in discrete time. Fig. 2 depicts the overall interactions between entities in an RL-based system where the agent perceives the state from the environment, takes an action and modifies its conditions in the environment, and earns a reward accordingly from the environment. The policy is to update after each new experience and reward for improving coming actions.

Eq. (1) formulates an MDP environment where s' is the successor state of s and p(s', r|s, a) is the probability of transition to the state s' with reward r, from the state s and the action a. Agents aim to maximize the expected return $\mathbb{E}[G_t]$ from the actions. G_t is calculated based on Eq. (2) where γ is the discount factor between and equal to 0 and 1, R_t is the reward at time t, and $\gamma^k R$ is the value of receiving reward R after k + 1 time-steps [46].

$$p(s', r|s, a) \doteq Pr\{S_t = s', R_t = r|S_{t-1} = s, A_{t-1} = a\} \tag{1}$$

$$G_t = \sum_{k=0}^{T} \gamma^k R_{t+k+1} \tag{2}$$

RL algorithms tend to estimate the value function which gives a quantitative assessment of the agent for being in the given state. This value refers to the expected return ($\mathbb{E}[G_t]$) or future/expected reward which depends on what action the agent takes. Accordingly, the value function outcome is the "ways of acting", so-called policies (π) [46]. In each interaction with the environment, the agent receives a reward from the environment. The agent then updates the current policy according to the earned reward leading to an improvement in the actions.

Based on whether the model of the environment is accessible by the agent, RL algorithms can act model-free and model-based. A model-free RL method primarily relies on learning while model-based method mainly relies on planning [47]. In the contexts of buildings and energy

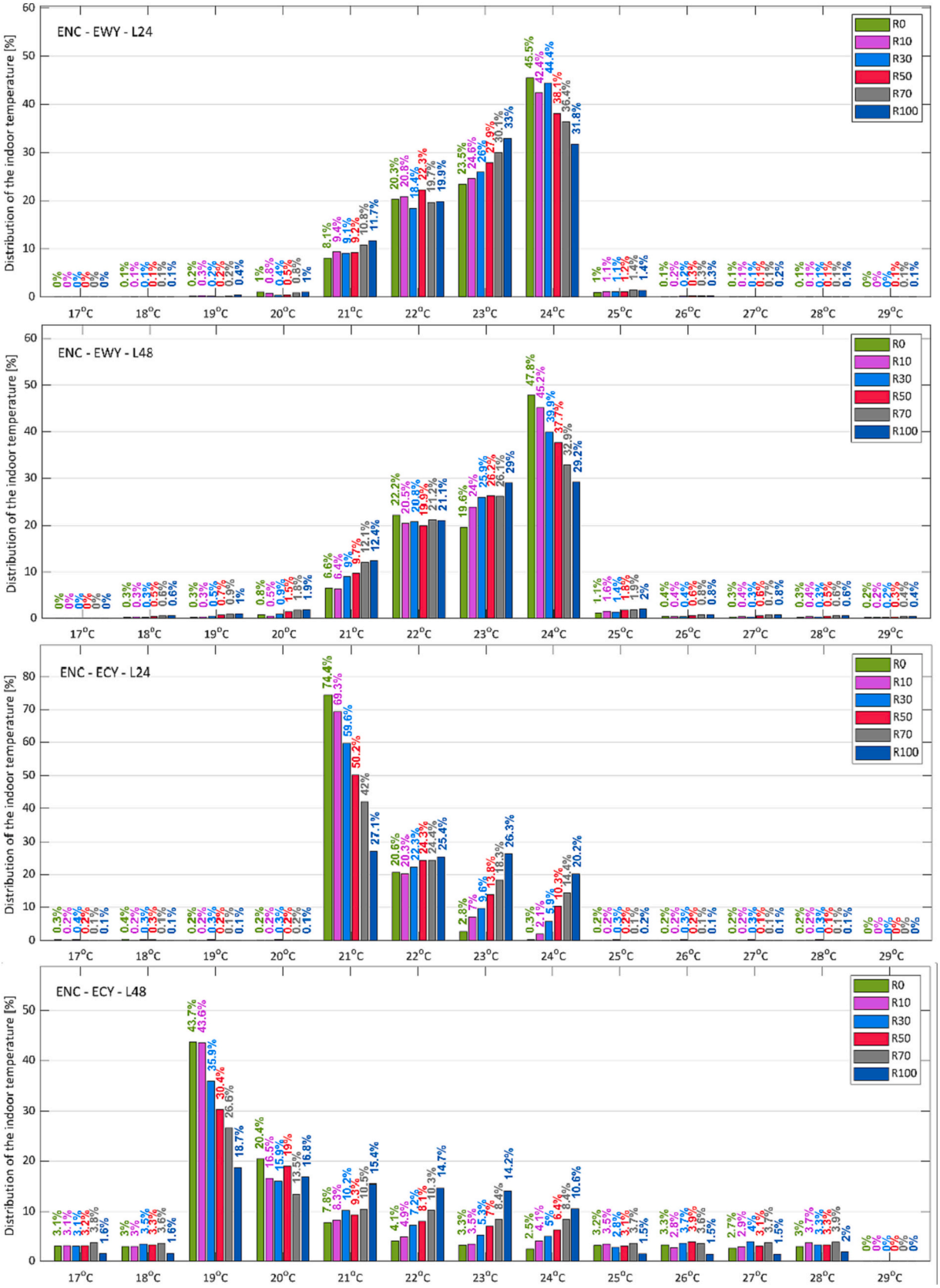


Fig. 8. Percentage of indoor temperature during EWY summer and ECY winter for L24 and L48 runs with different randomness levels. For ENC-ECY-L48, the indoor comfort range has been extended to 19-24 °C instead of 21-24 °C.

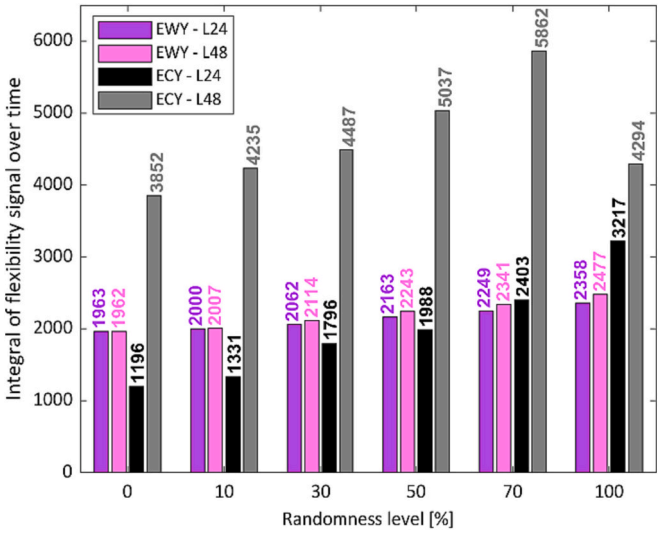


Fig. 9. Integral of the flexibility signal over time for ENC with L24 and L48 over EWY summer and ECY winter for different randomness levels.

systems, providing the model requires enormous amounts of data and yet the training efficiency is low [48] due to the large amount of uncertain and unknown data. In contrast, model-free RL can perform with the least amount of data by real-time learning and learning directly from the actions while it does not require to store the full description of the model [49]. The model-free approach avoids complexities in building energy modelling, energy systems identification, user behaviour, and weather conditions [50]. During the interaction with the environment, agents learn how to maximize their reward over time by taking proper actions in response to their environment based on the real-time data. Model-free approaches are divided into three categories, namely, (1) policy based (e.g., REINFORCE), (2) value based (e.g., Q-learning), and (3) hybrid actor-critic style [51,52]. Model-based methods, on the other hand, are divided into model-based RL with a known model (e.g., AlphaZero) or a learned model (e.g., Dyna-style) [53].

Nagy et al. [51] and Gao and Wang [53] carry out comprehensive studies about model-free and model-based RL methods while value and policy based methods are investigated by Nachum et al. [52] and Nagy et al. [51]. Zhuang et al. [54] proposes an RL-MPC (Model Predictive Control) coupled method for HVAC control using time series forecasting. Biemann et al. [55] applies a model-free actor-critic control algorithm for HVAC in an experiment showing the robustness and data efficiency emphasizing the implementation complexities. Suman et al. [56] deploys RL to learn occupants behaviour using MDP to overcome the uncertainties related to the occupants. Fu et al. [57] shows the outperformance of the model-free RL compared to MPC in a simulation-based study for load shifting while Wang et al. [58] compare the performance of the model-free RL against MPC in HVAC control optimization. Zhou et al. [59] introduce a combination of RL with a rule-based control and decision-tree to enhance the building energy flexibility. Coraci et al. [60] propose an online transfer learning approach to increase the scalability of RL in building control.

The RL engine in CIRLEM is a value-based model-free engine which uses the flexibility signal to learn about its environment and state. In this work, the flexibility signal is defined as a number between 0 and 5, comparing the energy demand at time t to the reference energy demand at time t. The reference energy demand is the demand during typical weather conditions or TDY. Signal 0 means that the energy demand is less than or equal to the TDY demand while values 1–5 indicate a higher demand; the larger the signal, the higher the demand in comparison to the TDY demand. Since the signal is generated every 15 min in this work, the reference values are calculated by considering the average and maximum 15-min energy demand values for TDY seasonal periods. For example, the maximum 15-min cooling demand in TDY summer and the average 15-min value for the whole TDY summer. Afterwards, the range of calculated values is divided into five equal sections (can be non-equal depending on the need and capacity of the energy provider). Knowing these values, the corresponding flexibility signals are generated per time step (15-min) using a simple function.

The self-knowledge of the agent is generated by calculating rewards per time step or values. In this work, energy demand and indoor comfort are used to define the value functions. If both the energy demand and indoor discomfort at time t are smaller than the corresponding value for the extreme reference case (the default setting of agents/buildings during extreme weather conditions, ECY or EWY), the value function (or reward at that time t for the adopted actions) is equal to 1, if energy demand is smaller but discomfort is larger, it is equal to 0.5, and if both are larger, the value function at time t is zero. Actions with the value of 1 are selected as a set of actions that form the policy (unless the algorithm does not converge to a solution, then actions with the value function of 0.5 are also considered). The selected actions are added in the library of

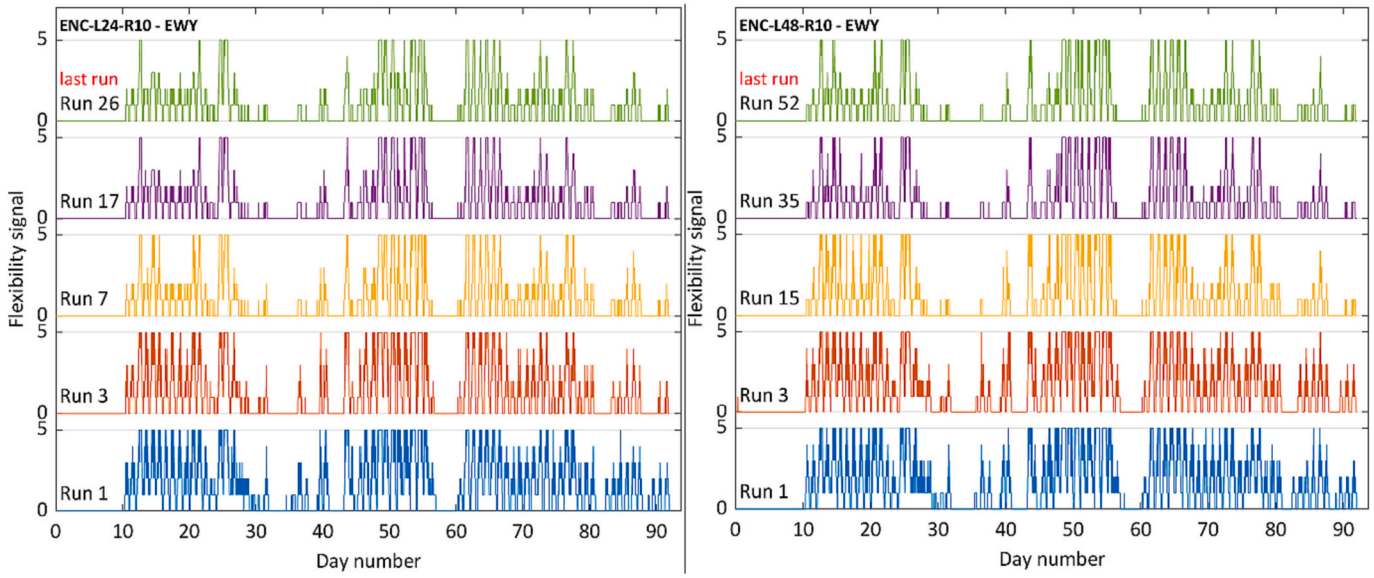


Fig. 10. Flexibility signal profiles for five runs of ENC with R10 for L24 (left) and L48 (right) during EWY summer.

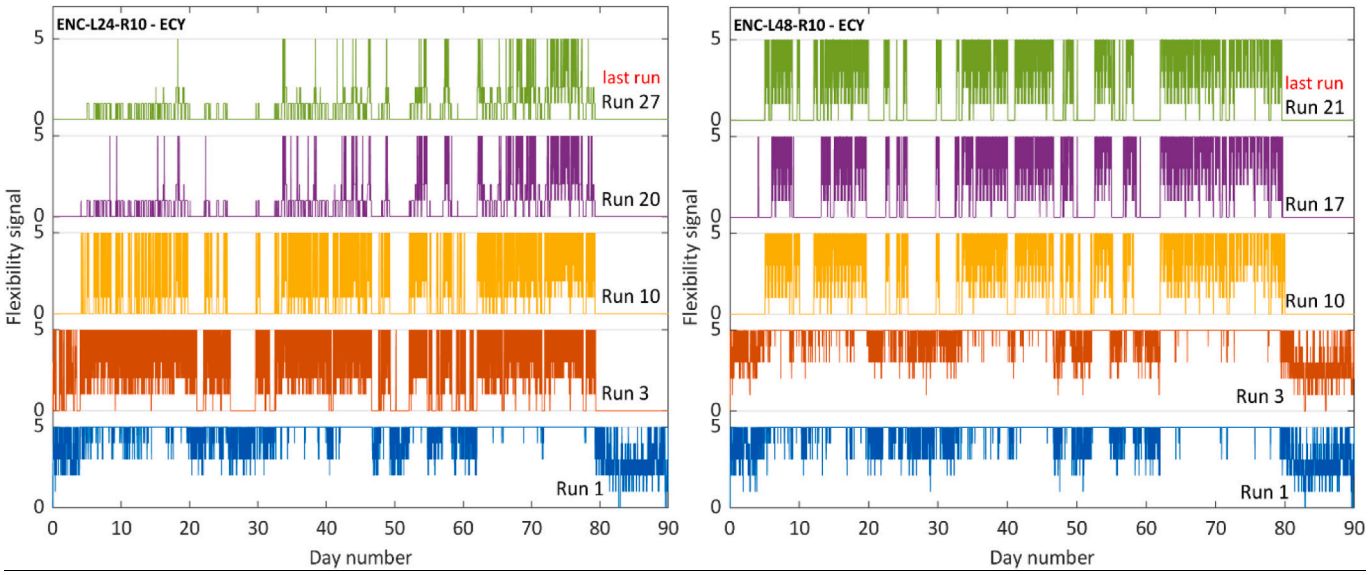


Fig. 11. Flexibility signal profiles for five runs of ENC with R10 for L24 (left) and L48 (right) during ECY winter.

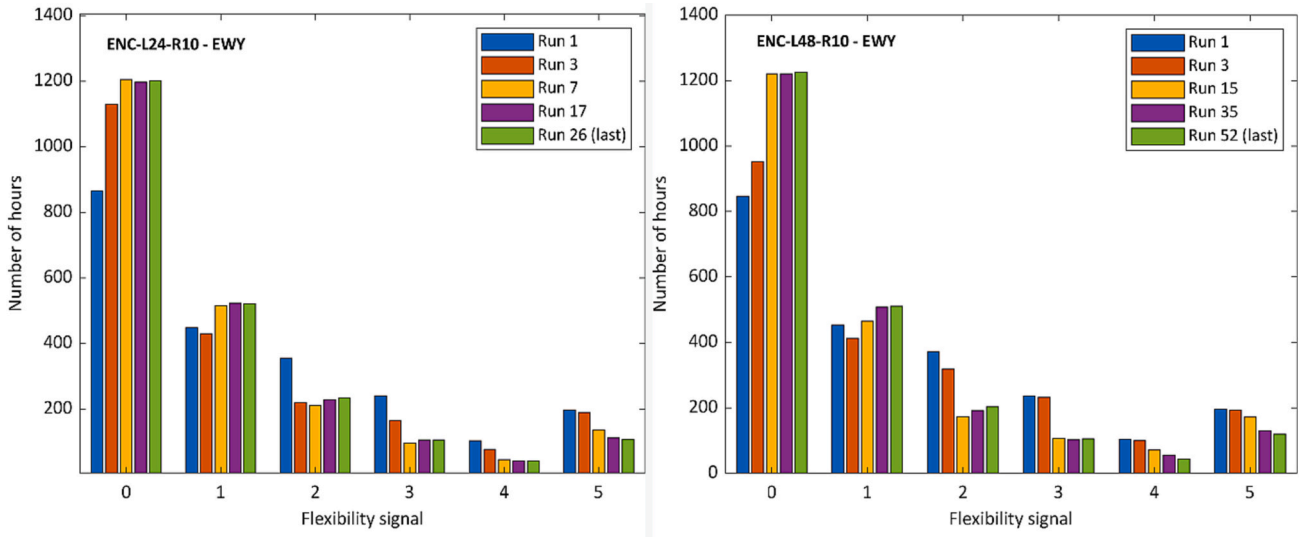


Fig. 12. Flexibility signal distribution for five runs of ENC with R10 for L24 (left) and L48 (right) during EWY summer.

actions, forming the final and optimized library or the policy of the agent. In other words, the agent's policy is a set of optimal actions which have been selected through an iterative process. We set a limit for the length of the library in this work, which is 24 and 48 actions and respectively called L24 and L48 hereafter. In this work, we put an equal weight to energy saving and comfort when selecting the actions. This can be changed based on the need, for example putting a higher weight on energy saving (e.g. 70%) and lower on comfort (e.g. 30%).

The control strategies inside agents are interpreted as the actions of the agents, which are divided into three groups of 1) changing the cooling set-point between 18 and 29 °C in summer and heating set-point between 17 and 28 °C in winter, both with the intervals of 1 °C (12 possible actions), 2) changing the ventilation rate per area between 0 and 1.5 [l/m^2/s] with 0.3 [l/m^2/s] intervals (5 possible actions), and 3) changing the internal loads (including equipment and plug loads) between 0 and 9 [W/m^2] with the intervals 3 [W/m^2] (4 possible actions). This results in 240 possible actions per agent. These actions are also called adaptation actions/measures, since they are the actions of the agent to adapt to the new environmental conditions. The actions are selected randomly at the beginning and through the rewarding mechanism, the agents pick the suitable actions till creating the policy. To not stick all the time to the same policy and try different actions, we defined a randomness factor in the algorithm, which allows the algorithm to pick a random action even after fixing the policy. This opens doors to update the policy if (by any chance) a better action is being experienced by the agent. In this work, we have defined six randomness levels of 0%, 10%, 30%, 50%, 70% and 100%, which 0% means there is no random choice (sticking all the time to the learnt policy) and 100% means to go all the time with random actions.

2.4. The case study

The performance of CIRLEM is assessed for an elderly care centre in Ålesund, Norway, which is called Eidet and owned by the Ålesund municipality (Fig. 3). The Eidet building is operated by the municipality property manager who operates around 400 public buildings in the town. The majority of the buildings are equipped with a building management system (BMS) that provides online access and control to

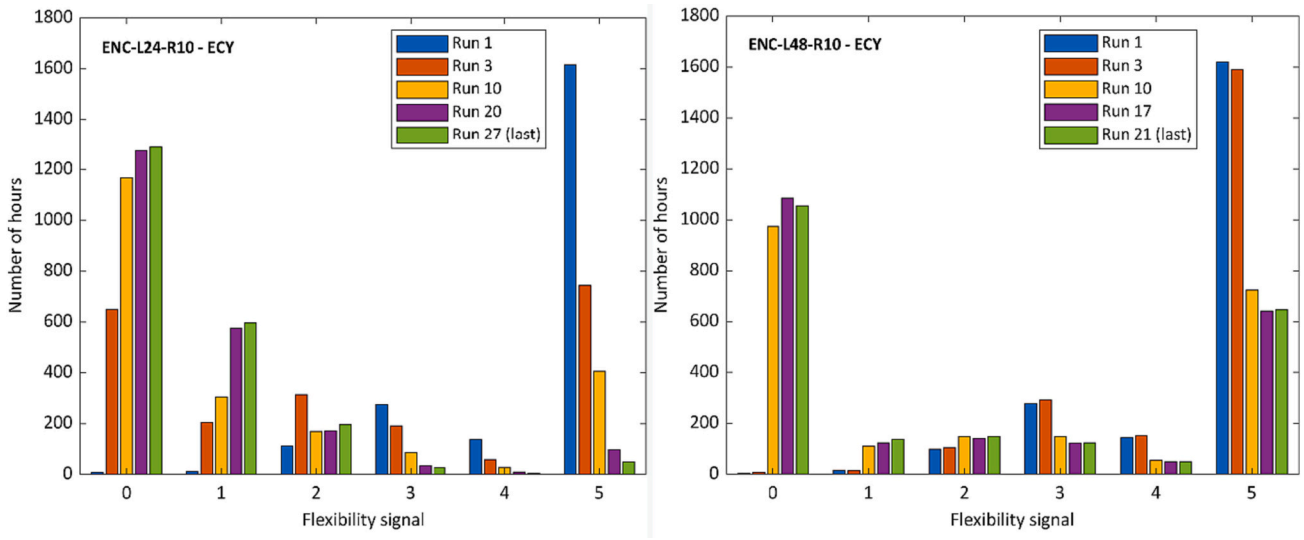


Fig. 13. Flexibility signal distribution for five runs of ENC with R10 for L24 (left) and L48 (right) during ECY winter.

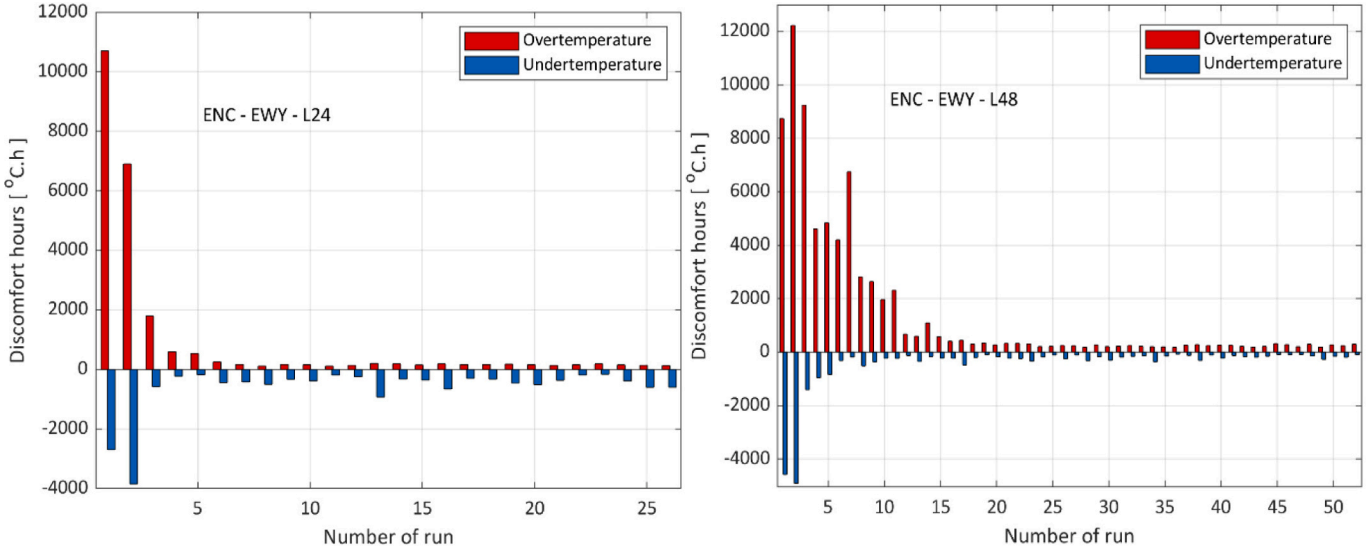


Fig. 14. Evolution of the discomfort hours over runs of the algorithm for ENC-L24-R10 during EWY (left) and ECY (right).

the buildings. The building was built in 2017 in five floors and a conditioned area of 7000 m2. The first to fourth floors accommodate private residential units and gathering rooms with a number of 32 rooms per floor. The basement includes facility rooms, kitchen, fridges, and storages. Like other municipal buildings, Eidet elderly care centre is also equipped by BMS, smart meters, and renewable sources of energy consists of photovoltaic cells, solar collector, boreholes, and thermal storage which are fully controlled by the BMS [61]. The residential rooms are the majority of the units, each around 32 m2 including a bathroom. Rooms are private with one occupant and an independent set-point which can be adjusted by the user $\pm 3^{\circ}C$ relative to the BMS setpoint. There are also some gathering rooms and offices in the building. The Eidet building presents an ideal case study for evaluating the performance of CIRLEM, owing to its controlled and monitored environment encompassing separate zones, users, and controllers. Additionally, the building's seamless connection to the grid and the availability of detailed information and data further contribute to its suitability for in-depth analysis.

For the purpose of this work, a high spatiotemporal resolution building performance simulation model was developed in EnergyPlus, including multiple thermal zones on each floor. EnergyPlus was selected to develop the building energy model (BEM) because of the proper integration with Python and the possibility of interaction with the dynamic simulation at each timestep via EnergyPlus API to mimic the flexible energy management. EnergyPlus API provides the ability to fetch the results at each timestep, modify the settings of the BEM and continue the dynamic simulation accordingly. Two groups of variables are defined including sensors and actuators to send and receive values from the API. Sensors represent the simulation results such as energy demand and air temperature which fetch the data from the simulation after each run and send to CIRLEM script. Actuators carry the required values to run the energy simulation such as cooling and heating setpoint, ventilation rate (per area and per people), lighting load, equipment load, occupancy, etc. to send them to the simulation according to the decisions made by CIRLEM. In total, 20 different thermal zones are generated which are considered as agents in this study. The model is verified against measured electricity, using historic weather data (check [61] for more details).

Typical and extreme weather data sets are generated using 13 future climate scenarios from the "Coupled Model Intercomparison Project 5"

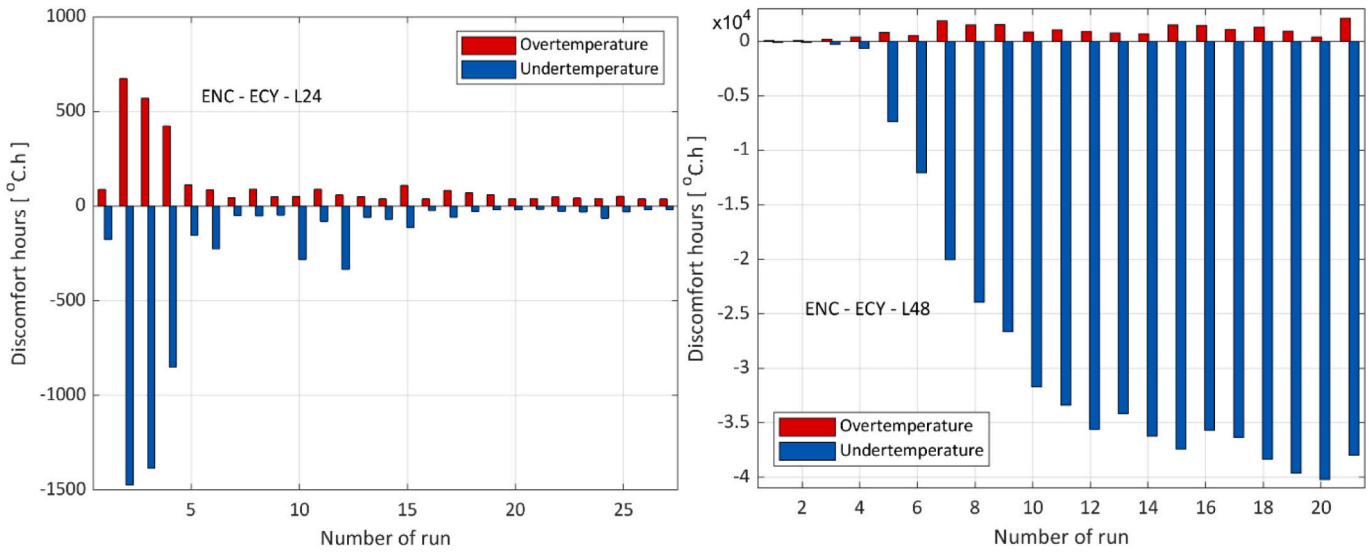


Fig. 15. Evolution of the discomfort hours over runs of the algorithm for ENC-L48-R10 during EWY (left) and ECY (right).

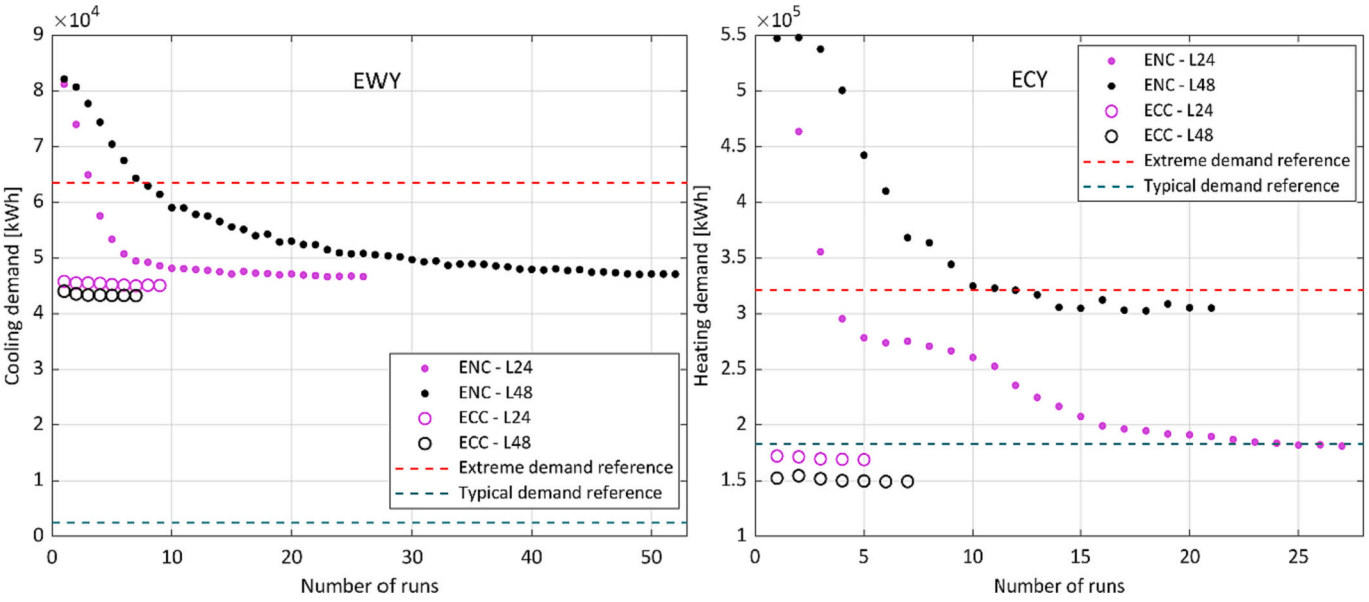


Fig. 16. Comparing the convergence and energy saving of ENC and ECC algorithms for cooling in extreme warm summer and heating in extreme cold winter.

(CMIP5) for the period of 2040–2069. Three sets of data are generated for building simulation using Nik's approach [62], namely Typical Downscaled Year (TDY), Extreme Warm Year (EWY) and Extreme Cold Year (ECY). CIRLEM is used to control the performance of the 20 zones in the Eidet building during EWY summer and ECY winter in comparison with the typical (TDY) summer and winter. The distributions of the outdoor temperature together with some statistics are shown Fig. 4 for the typical and extreme weather data sets.

3. Results

Results are discussed in two sections and the first one is focused on ENC where all the decision making is happening at the edge node (i.e. inside agents or buildings). The performance of CIRLEM for different randomness levels is also investigated in this case. In the second section, the performance of CIRLEM with 10% randomness is investigated for ENC and ECC, investigating how transferring some of the decision making to the cluster level affects the performance of CIRLEM.

3.1. ENC and different randomness levels

In ENC all the decision making is happening at the edge node (inside the agent) and only one flexibility signal is transferred from the cluster to all the agents. The assessment is conducted by considering policies with two distinct numbers of actions: 24 and 48, referred to as libraries or sets of actions (designated as L24 and L48). Additionally, six different randomness levels are considered: 0%, 10%, 30%, 50%, 70%, and 100%.

In this work, the convergence criteria were based on checking the last five runs and comparing the total energy demand of the building. If the standard deviation of the last five runs is 0.002 times smaller than their average, then the final solution (or the optimum policy) converges. This is very strict convergence criteria and to streamline the process without compromising the outcome, we made the necessary adjustment by relaxing the convergence criteria for L48 to 0.02. This decision was taken to prevent prolonged runs when there was no certainty of a significantly improved solution being achieved. Moreover, for L48 during ECY winter, we had to extend the acceptable range of indoor

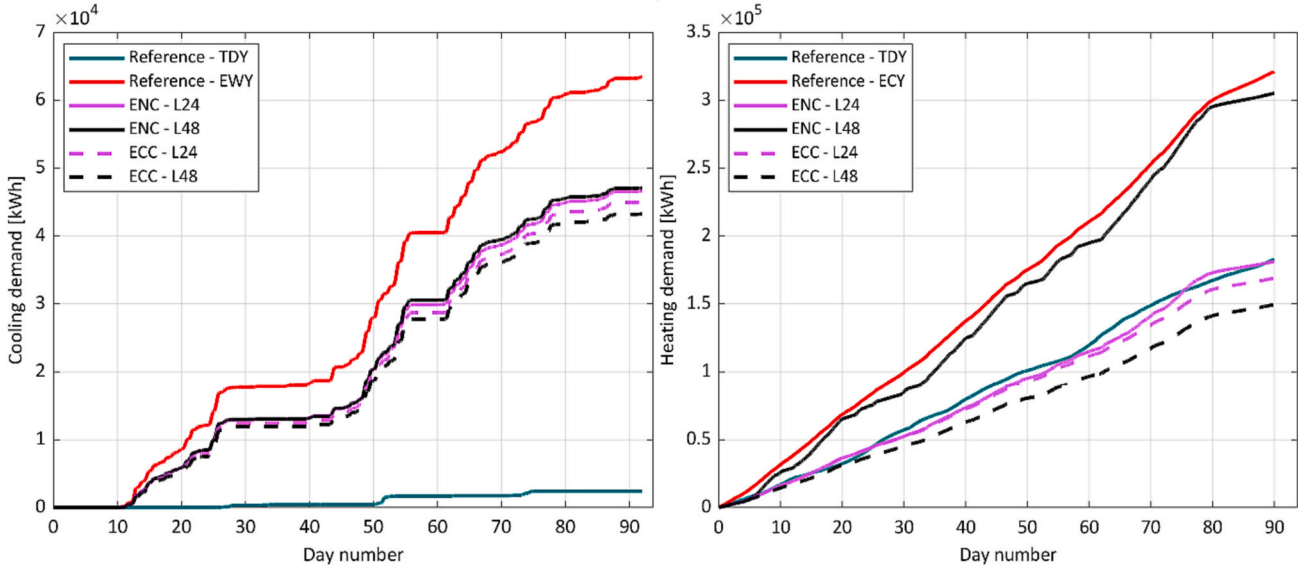


Fig. 17. Cumulative distribution of the hourly cooling (left) and heating (right) demand for the reference case during typical and extreme conditions and ENC and ECC with L24 and L48 during extreme weather conditions.

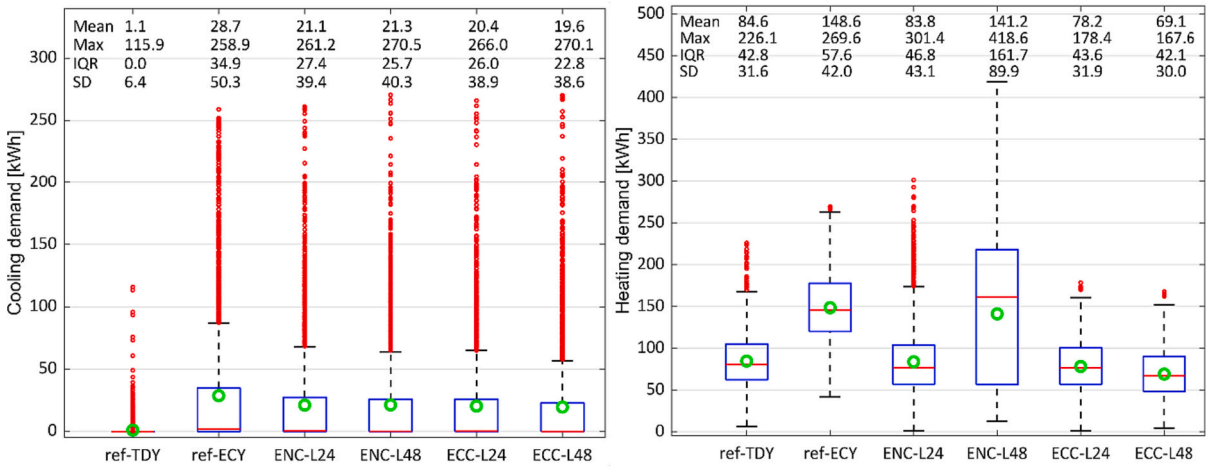


Fig. 18. Distribution of the hourly cooling (left) and heating (right) demand for the reference case during typical and extreme conditions and ENC and ECC with L24 and L48 during extreme weather conditions.

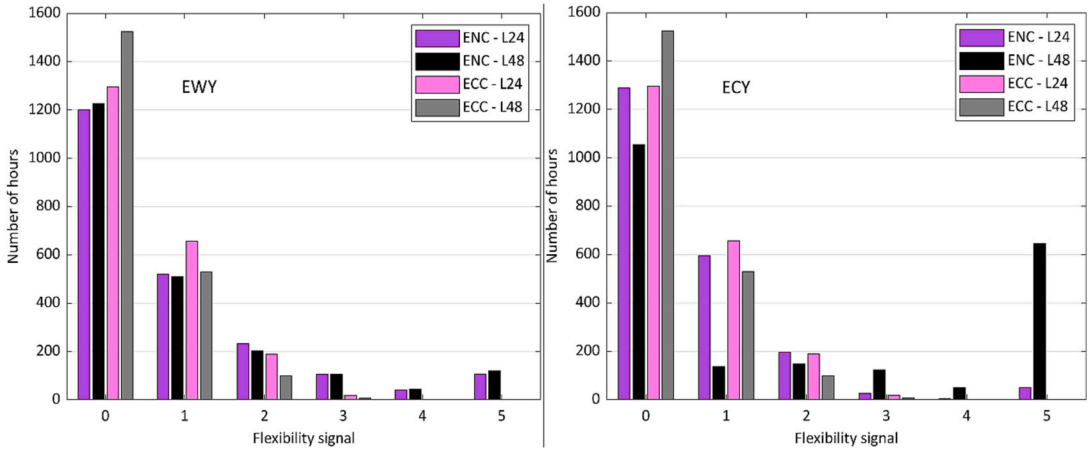


Fig. 19. Number of hours for each flexibility signal for ENC and ECC with L24 and L48 during EWY (left) and ECY (right).

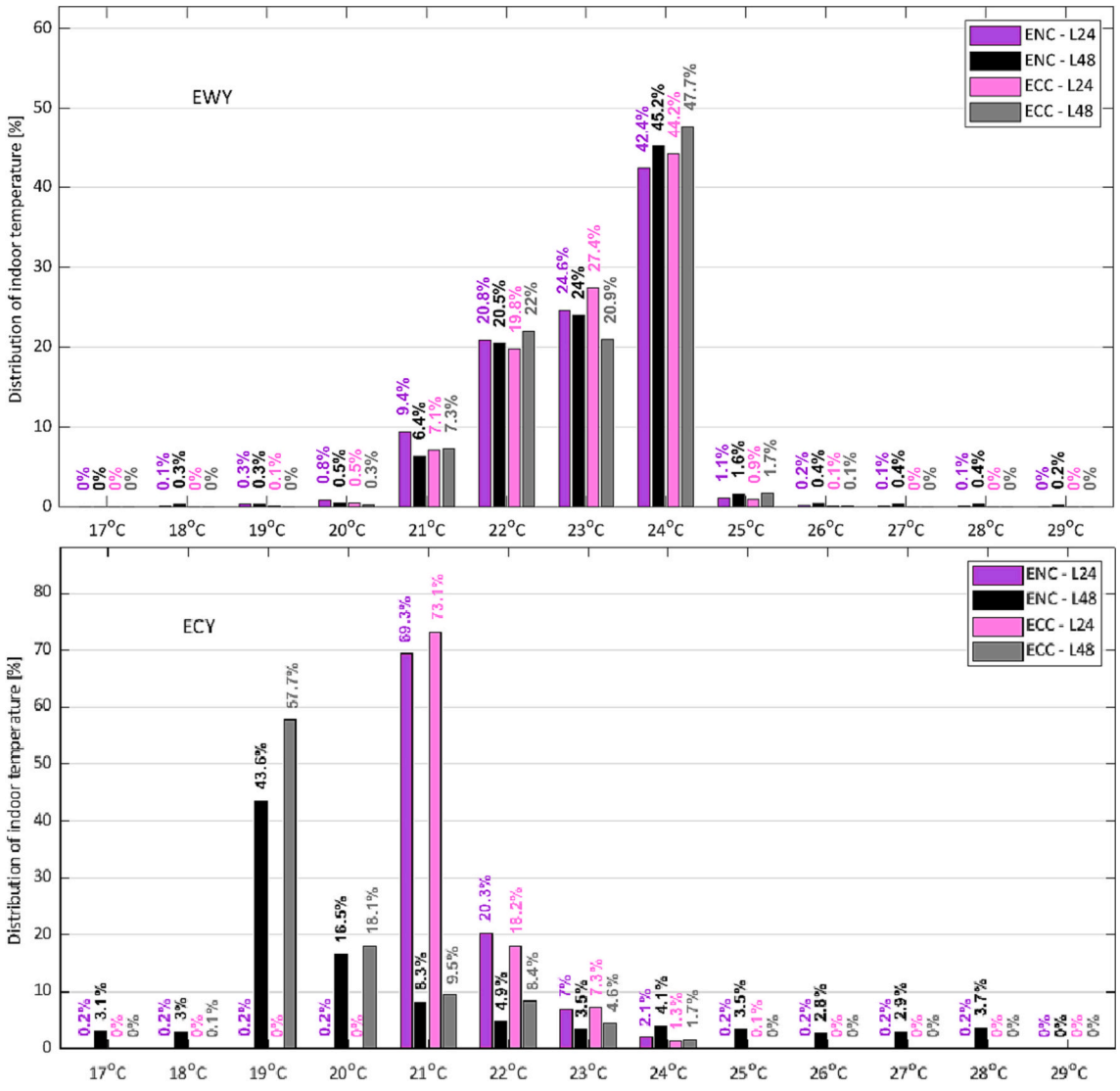


Fig. 20. Distribution of indoor temperature during EWY summer and ECY winter for ENC and ECC with L24 and L48 policies.

Table 1
Percentage of indoor temperature being maintained within the comfort range across various CIRLEM algorithms.

CIRLEM	Comfort during EWY [%]	Comfort during ECY [%]	Comfort during ECY - extended comfort [%]
ENC-L24	97.2	98.7	N/A
ENC-L48	96.1	20.8	80.9
ECC-L24	98.5	99.9	N/A
ECC-L48	97.9	24.2	100

temperature and set the minimum to 19 °C (instead of 21 °C). Fig. 5 compares the convergence of the algorithms to reach an optimum cooling policy in the building for EWY summer, while Fig. 6 makes a similar comparison for heating policy in ECY winter. The reference energy values are shown as dashed blue (for typical weather conditions) and red (for extreme weather conditions) lines. The reference control strategy for both cases is having ventilation rate per area of 0.3 [l/m^2/s] and internal load of 6 [W/m^2], while the cooling set-point in summer is 23 °C and heating set-point in winter is 24 °C. For both the cases in Fig. 5 and Fig. 6, a lower randomness level results in lower energy demand. We reach an optimum policy faster with a smaller library size, or L24, however the convergence speed does not show any correlation with the randomness level. For cooling in summer in Fig. 5, L24 drops the energy demand very quickly after the fourth run, in comparison to L48 (compare the green asterisk with the other graphs in Fig. 5-right). L24 works better for both cooling and heating the building, decreasing the energy demand to lower values and converging faster. The difference between L24 and L48 is considerable for heating demand as is visible in Fig. 6.

For all the cases, a lower randomness results in a lower final energy demand (for the fixed policy). The distributions of the hourly energy demand values are plotted in Fig. 7 together with some statistics above each boxplot. For both the cooling and heating demands, a higher randomness increases the average energy demand but not necessarily the other statistics. For example, the maximum hourly energy demand (or peak values) can slightly decrease by increasing the randomness. In ENC, increasing the library size for the policy does not provide any advantage during the cooling season and even worsens the performance of CIRLEM during the heating season.

Knowing that indoor comfort has the same weight as energy saving in this study, distribution of the indoor temperature for different randomness levels and run cases are compared in Fig. 8. For all the cases,

CIRLEM keeps the indoor temperature in the defined comfort range most of the time. By reducing the randomness level, there is an observed increase in the percentage of maintaining temperatures at the upper limit (24 °C) during the warm months (EWY) and at the lower limits (21 °C or 19 °C) during the cold months (ECY). This can result in a lower energy demand.

Impacts of randomness level on the flexibility signal is shown in Fig. 9 by calculating the integral of the flexibility signal over time and comparing it between different cases. As visible, for each case (e.g. EWY-L24) a higher randomness level results in a higher integral of flexibility signals. The only exception is ECY-L48 for the randomness of 100%, which is not a case to make conclusions about the performance of CIRLEM itself, since L48 is not an optimum policy length for heating in cold season, as it obvious based on the results (a bigger number of actions can increase the risk of extreme conditions for this heating dominated case). In general, we can see that the R10 algorithm can be considered as a safe choice and R0 a conservative one. For both the cooling and heating cases in Fig. 9, L24 keeps the total value of the signal integral lower than L48.

Picking 10% as the optimum randomness level in this work, the evolution of CIRLEM over different runs, as it approaches convergence, is studied by comparing the flexibility signals between different runs. As it was discussed in the theory section, the flexibility signal represents the state of the environment and reflects upon the collective behaviour of the agents. Therefore, studying its variations between different cases can provide more information about their performances. As time progresses, the algorithm is expected to learn and effectively manage energy demand, aiming to optimize the performance of the energy system by minimizing peak demands and achieving a smoother energy consumption profile. The profiles of flexibility signals for some runs (including the first and last run) are shown in Fig. 10 and Fig. 11, respectively for EWY summer and ECY winter. As is visible, the intensity of signals over 3 decreases by reaching the final run. This is better illustrated in Fig. 12 and Fig. 13 by plotting bar charts for exactly the same cases. As visible, the quantity of 2, 3, 4 and 5 signals decrease as the solution converges while the number of 0 and 1 signals increase (the only minor exception is ENC-L48-R10-ECY for signal 2 in Fig. 13). This indicates that CIRLEM learns by time to keep the energy demand at lower levels and decrease the number of hours with intense energy demand.

It is interesting to see how the indoor comfort conditions change per run until converging to a final solution. This is visualized in Fig. 14 and Fig. 15, respectively showing the discomfort hours for EWY summer and ECY winter. Each bar chart shows the over (plus values) and under-temperature (minus values) hours separately for two policies with 24 and 48 sets of actions. The data presented clearly indicates a decrease in the number of discomfort hours across all cases, with some fluctuations observed during the convergence process. However, it is worth noting that the ENC-ECY-L48 case exhibits a larger number of under-temperature hours. This discrepancy is due to our consideration of any temperature below 21 °C as discomfort hours, while the ENC-ECY-L48 case includes an extended discomfort range of 19–24 °C, resulting in a significant number of hours falling within the 19–21 °C range.

3.2. ENC and ECC with 10% randomness

This section presents a comparison between two distinct approaches for running CIRLEM. As explained in the theory section, the ENC approach involves all control and optimization taking place at the edge side, where agents receive a single signal per time step from the energy provider. In ECC, each agent has its own policy (which is developed through running a process similar to ENC), however there is a higher control at the cluster level, enabling optimized distribution of the flexibility signal among agents. In essence, ECC involves a greater level of control that optimizes the allocation of signals across agents. All the results in this section are based on having a randomness level of 10% when running CIRLEM.

The convergence of ENC and ECC approaches are compared in Fig. 16. Naturally, all the ECC cases converge much faster and start with lower values for energy demand than ENC since ECC starts from the point ENC ends, using the optimum policies that are converged in ENC. So, we cannot neglect the time required for agents to reach their optimum policies. Having the ENC policies, ECC can help to better distribute them among agents and reach lower energy demand values, as is visible in Fig. 16.

The cumulative energy demand profiles are plotted in Fig. 17 for all the cases in comparison to the reference cases for typical and extreme weather conditions. According to the results for cooling demand, ENC-L24 and ENC-L48 algorithms perform quite similarly (ENC-L24 performs slightly better, saving energy for 1% more). Adding a higher control at the cluster level can decrease the cooling demand a bit more, which is relatively bigger for ECC-L48 (~3% for L24 and ~ 8% for L48; compare ENC-L48 with ECC-L48 in Fig. 17-left). For heating demand, ENC-L24 performs very well and decreases the heating demand to values very close to the TDY reference. ECC-L24 slightly enhances the performance of CIRLEM, saving around 7% more energy in comparison to ENC-L24. Unlike the L24 case, the difference between ENC and ECC is huge for the L48 case. The heating demand for ENC-L48 is very high and close to the extreme ECY case. By adding a higher control at the cluster level for ECC-L48, it is possible to decrease the heating demand for 51% to values lower than the TDY reference case. Based on the results, having a bigger library size for ENC does not necessarily help with saving more energy and can even increase the energy demand and complicate the optimization process, especially for the case of heating during ECY winter. A bigger library can be advantageous in ECC where the higher control at the cluster level can become helpful in reaching optimum distributions of the policies.

By checking the distribution of the hourly energy values in Fig. 18, we see that differences in reducing the peak cooling demand are not considerable between different algorithms, unlike heating demand which ENC-L48 shows the worst performance in this case as well. ECC helps to decrease the outliers for heating demand, which might be considered as unprecedented future peaks, depending on how to interpret them [25]. Considering that the case study is already designed for cold climate conditions and the energy system can cope with extreme cold events, ENC-L24 can be a safe and cheap solution.

The distribution of flexibility signals for different algorithms are compared in Fig. 19 for cooling (left) and heating (right). Interestingly, ECC results in diminishing 4 and 5 signals in both cases, meaning that it successfully manages to avoid putting high pressure on the energy supplier during extreme weather events. While ENC-L48 is excluded for heating during ECY, it is worth noting that ENC still demonstrates a remarkable performance in reducing the necessity for high flexibility.

Distribution of the indoor temperature among the cases are compared in Fig. 20 for EWY summer and ECY winter. For all the cases, temperature is mostly distributed in the comfort range (assuming that the comfort range is 19-24 °C for L48-ECY and 21-24 °C for the rest). The performance of ENC-L24 is interestingly well for both EWY and ECY. For example, ENC-L24 has the highest percentage of 21 °C during EWY summer, while it is at the lower end of the comfort limit (which may require spending more energy in summer as visible in previous figures). Table 1 summarises the percentage of indoor temperature being maintained within the comfort range across all the CIRLEM algorithms assessed in this work. Given the remarkable performance of all algorithms in meeting the comfort criteria with consistently high rankings, this analysis shows that having a larger policy library and higher control at the cluster level does not necessarily and/or considerably enhances the indoor comfort in comparison to the fastest and simplest CIRLEM algorithm, namely ENC-L24.

4. Conclusions

In this work, CIRLEM was introduced which is a novel energy

management (EM) approach developed based on the synergic integration of the fundamental concepts of collective intelligence (CI) and reinforcement learning (RL). The RL engine in CIRLEM is a value-based model-free engine which uses the flexibility signal to learn about its environment and state. The flexibility signal reflects upon the collective behaviour of the agents in the grid. Two ways of running CIRLEM were investigated: 1) based on doing all the decision making and optimization at the edge node (or agents), called Edge Node Control or ENC, and 2) based on adding a higher control at the cluster level and controlling building from outside, called Edge node and Cluster Control or ECC. In ENC, there is no need to know about the control options and behaviour of single buildings, ensuring maximum user privacy and minimum data transfer, while in ECC there is a need to understand the impact of the policies adopted by buildings, however still there is no need to know about the actions in the policy. The performance of CIRLEM was investigated for an elderly building in Ålesund, Norway during two extreme periods; summer in an extreme warm year (EWY) and winter in an extreme cold year (ECY). The building was divided into 20 separate zones, each helping the whole group of zones and the energy supplier to pass the extreme conditions safely. Four reference cases were considered to perform a comparative analysis where two of them are the default running mode of the building during summers in typical (TDY) and extreme warm (EWY) years, and the other two are the default mode during typical and extreme cold (ECY) winters. The performance of CIRLEM was assessed for policies with 24 and 48 sets of actions (called L24 and L48), considering six randomness levels of 0, 10, 30, 50, 70 and 100% (R0–100) for ENC, while e.g. R10 means that the agent selects a random action for 10% of time instead of picking the best action from its policy.

Based on the results, CIRLEM converges quickly to an optimum solution (optimum set of policies), providing an enhanced indoor comfort and energy saving, with variations between cases based on running ENC or ECC, size of the policy library and randomness level. There is no obvious correlation between randomness and convergence, however a higher randomness level increases the average energy demand but not necessarily the other statistics such as maximum energy demand. According to the results, having a bigger library size for ENC does not necessarily help with saving more energy and can even increase the energy demand and make the optimization process lengthier, especially for the case of heating during ECY winter. A bigger library can be advantageous in ECC where the higher control at the cluster level can become helpful in reaching optimum distributions of the policies. Moreover, considering the fact that all algorithms performed remarkably well in meeting the comfort criteria, we can conclude that having a larger policy library and higher control at the cluster level does not necessarily and/or considerably enhance the indoor comfort. Overall, CIRLEM could enhance the energy flexibility and climate resilience of the building, saving energy without compromising indoor comfort, while the fastest and simplest CIRLEM algorithm, namely ENC-L24, demonstrated an excellent performance for both cooling in EWY summer and heating in ECY winter.

This work is based on some ongoing projects and further research is under development to enhance CIRLEM. The future work is focused on integrating other selection criteria such as energy price into decision making. Moreover, by integrating other factors into the energy price, such as sustainability of the energy source and flexibility, the price signal itself can be used as a comprehensive flexibility signal. The performance of CIRLEM is going to be assessed for other climate regions over Europe considering different building designs and urban areas. Moreover, the developed algorithms will be implemented in Raspberry Pi control units and tested in controlled environments.

CRediT authorship contribution statement

Vahid M. Nik: Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft. **Mohammad Hosseini:** Resources, Software, Validation, Writing – original draft.

Declaration of Competing Interest

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

Data availability

The authors do not have permission to share data.

Acknowledgment

This work was 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), the Joint ERA-Net Call 2020 (MICall20) on digital transformation for green energy transition under DigiCiti project (project ID: 108807), the Crafoord Foundation under CARUACI project, and the Swedish Research council for Sustainable Development, Formas (project ID: 2022-01120).

References

[1] IPCC. Climate change 2022: Impacts, Adaptation and Vulnerability. 2022.
[2] Sommaren. Extremt varm och solig | SMHI 2018. https://www.smhi.se/klimat/klimatet-da-och-nu/arets-vader/sommaren-2018-extremt-varm-och-solig-1.138134; 2018 (accessed November 18, 2018).
[3] Ruuhela R, Votsis A, Kukkonen J, Jylhä K, Kankaanpää S, Perrels A. Temperature-related mortality in Helsinki compared to its surrounding region over two decades, with special emphasis on intensive heatwaves. Atmosphere 2021:12. https://doi.org/10.3390/atmos12010046.
[4] Brutally cold weather in Scandinavia. Mkweather; 2021. https://mkweather.com/brutally-cold-weather-in-scandinavia-legendary-frosts-in-sweden-375c-norway-369c-and-finland-340c-and-below-40c-is-forecasted-the-coldest-seasonal-temperatures-in-50/ [accessed January 20, 2022].
[5] Fonseca-Rodríguez O, Sheridan SC, Lundevaller EH, Schumann B. Effect of extreme hot and cold weather on cause-specific hospitalizations in Sweden: a time series analysis. Environ Res 2021;193:110535. https://doi.org/10.1016/j.envres.2020.110535.
[6] Oudin Åström D, Åström C, Forsberg B, Vicedo-Cabrera AM, Gasparrini A, Oudin A, et al. Heat wave-related mortality in Sweden: a case-crossover study investigating effect modification by neighbourhood deprivation. Scand J Public Health 2020;48:428–35. https://doi.org/10.1177/1403494818801615.
[7] Venter ZS, Krog NH, Barton DN. Linking green infrastructure to urban heat and human health risk mitigation in Oslo. Norway Sci Total Environ 2020;709:136193. https://doi.org/10.1016/j.scitotenv.2019.136193.
[8] Rumpca J. Impacts of climate change on indoor thermal comfort in typical Swedish residential buildings - assessing risks for human health. Lund University; 2021.
[9] Vandentorren S, Bretin P, Zeghnoun A, Mandereau-Bruno L, Croisier A, Cochet C, et al. August 2003 heat wave in France: risk factors for death of elderly people living at home. Eur J Pub Health 2006;16:583–91. https://doi.org/10.1093/eurpub/ckl063.
[10] Bouchama A, Dehbi M, Mohamed G, Matthies F, Shoukri M, Menne B. Prognostic factors in heat wave related deaths: a meta-analysis. Arch Intern Med 2007;167:2170–6. https://doi.org/10.1001/archinte.167.20.ira70009.
[11] Smith KR. Human health: impacts, adaptation, and co-benefits. In: Climate Change 2014: Impacts, adaptation, and vulnerability. Part A: Global and Sectoral Aspects Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change; 2014.
[12] Calls for post-Covid "revolution" in building air quality. BBC News; 2021.
[13] Morawska L, Allen J, Bahnfleth W, Bluyssen PM, Boerstra A, Buonanno G, et al. A paradigm shift to combat indoor respiratory infection. Science 2021;372:689–91. https://doi.org/10.1126/science.abg2025.
[14] Hosseini M, Javanroodi K, Nik VM. 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 2022;78:103634. https://doi.org/10.1016/j.scs.2021.103634.
[15] Nik VM, Sasic Kalagasidis A, Kjellström E. Assessment of hygrothermal performance and mould growth risk in ventilated attics in respect to possible climate changes in Sweden. Build Environ 2012;55:96–109. https://doi.org/10.1016/j.buildenv.2012.01.024.
[16] Todeschi V, Javanroodi K, Castello R, Mohajer N, Mutani G, Scartezzini J-L. Impact of the COVID-19 pandemic on the energy performance of residential

neighborhoods and their occupancy behavior. Sustain Cities Soc 2022:103896. https://doi.org/10.1016/j.scs.2022.103896.
[17] Perera ATD, Javanroodi K, Nik VM. Climate resilient interconnected infrastructure: co-optimization of energy systems and urban morphology. Appl Energy 2021;285:116430. https://doi.org/10.1016/j.apenergy.2020.116430.
[18] Nik VM, Perera ATD, Chen D. Towards climate resilient urban energy systems: a review. Natl Sci Rev 2021:8. https://doi.org/10.1093/nsr/nwaa134.
[19] Buldyrev SV, Parshani R, Paul G, Stanley HE, Havlin S. Catastrophic cascade of failures in interdependent networks. Nature 2010;464:1025–8. https://doi.org/10.1038/nature08932.
[20] Wei C, Bai X, Kim T. Advanced control and optimization for complex energy systems. Complexity 2020;2020:e5908102. https://doi.org/10.1155/2020/5908102.
[21] Magnan AK, Schipper ELF, Burkett M, Bharwani S, Burton I, Eriksen S, et al. Addressing the risk of maladaptation to climate change. WIREs Clim Change 2016;7:646–65. https://doi.org/10.1002/wcc.409.
[22] Kramer T, Garcia-Hansen V, Omrani S, Nik VM, Chen D. A machine Learning approach to enhance indoor thermal comfort in a changing climate. Switzerland: Lausanne; 2021.
[23] Javanroodi K, Nik VM. Interactions between extreme climate and urban morphology: investigating the evolution of extreme wind speeds from mesoscale to microscale. Urban Clim 2020;31:100544. https://doi.org/10.1016/j.uclim.2019.100544.
[24] Perera ATD, Wickramasinghe PU, Nik VM, Scartezzini J-L. Introducing reinforcement learning to the energy system design process. Appl Energy 2020;262:114580. https://doi.org/10.1016/j.apenergy.2020.114580.
[25] Perera ATD, Nik VM, Chen D, Scartezzini J-L, Hong T. Quantifying the impacts of climate change and extreme climate events on energy systems. Nat Energy 2020;5:150–9. https://doi.org/10.1038/s41560-020-0558-0.
[26] Nik VM, Moazami A. Using collective intelligence to enhance demand flexibility and climate resilience in urban areas. Appl Energy 2021;281:116106. https://doi.org/10.1016/j.apenergy.2020.116106.
[27] Zhou Y. Advances of machine learning in multi-energy district communities–mechanisms, applications and perspectives. Energy AI 2022:10. https://doi.org/10.1016/j.egyai.2022.100187.
[28] Strbac G. Demand side management: benefits and challenges. Energy Policy 2008;36:4419–26. https://doi.org/10.1016/j.enpol.2008.09.030.
[29] Lund PD, Lindgren J, Mikkola J, Salpakari J. Review of energy system flexibility measures to enable high levels of variable renewable electricity. Renew Sust Energ Rev 2015;45:785–807. https://doi.org/10.1016/j.rser.2015.01.057.
[30] McIlvennie C, Sanguinetti A, Pritoni M. Of impacts, agents, and functions: an interdisciplinary meta-review of smart home energy management systems research. Energy Res Soc Sci 2020;68:101555. https://doi.org/10.1016/j.erss.2020.101555.
[31] Vázquez-Canteli JR, Nagy Z. Reinforcement learning for demand response: a review of algorithms and modeling techniques. Appl Energy 2019;235:1072–89. https://doi.org/10.1016/j.apenergy.2018.11.002.
[32] Wang Z, Hong T. Reinforcement learning for building controls: the opportunities and challenges. Appl Energy 2020;269:115036. https://doi.org/10.1016/j.apenergy.2020.115036.
[33] Lu R, Hong SH, Zhang X. A dynamic pricing demand response algorithm for smart grid: reinforcement learning approach. Appl Energy 2018;220:220–30. https://doi.org/10.1016/j.apenergy.2018.03.072.
[34] Perera ATD, Kamalaruban P. Applications of reinforcement learning in energy systems. Renew Sust Energ Rev 2021;137:110618. https://doi.org/10.1016/j.rser.2020.110618.
[35] Papini M, Binaghi D, Canonaco G, Pirotta M, Restelli M. Stochastic variance-reduced policy gradient. In: Proc. 35th Int. Conf. Mach. Learn. PMLR; 2018. p. 4026–35.
[36] Sheikhi A, Rayati M, Ranjbar AM. Demand side management for a residential customer in multi-energy systems. Sustain Cities Soc 2016;22:63–77. https://doi.org/10.1016/j.scs.2016.01.010.
[37] Gelazanskas L, Gamage KAA. Demand side management in smart grid: a review and proposals for future direction. Sustain Cities Soc 2014;11:22–30. https://doi.org/10.1016/j.scs.2013.11.001.
[38] Suran S, Pattanaik V, Draheim D. Frameworks for collective intelligence: a systematic literature review. ACM Comput Surv 2020;53(14):1–14. https://doi.org/10.1145/3368986.
[39] Schut MC. On model design for simulation of collective intelligence. Inf Sci 2010;180:132–55. https://doi.org/10.1016/j.ins.2009.08.006.
[40] Nweye K, Liu B, Stone P, Nagy Z. Real-world challenges for multi-agent reinforcement learning in grid-interactive buildings. In: ArXiv211206127 Cs Eess; 2022.
[41] Zhang L, Gao Y, Zhu H, Tao L. A distributed real-time pricing strategy based on reinforcement learning approach for smart grid. Expert Syst Appl 2022:191. https://doi.org/10.1016/j.eswa.2021.116285.
[42] Charbonnier F, Morstyn T, McCulloch MD. Scalable multi-agent reinforcement learning for distributed control of residential energy flexibility. Appl Energy 2022:314. https://doi.org/10.1016/j.apenergy.2022.118825.
[43] Sasaki T, Biro D. Cumulative culture can emerge from collective intelligence in animal groups. Nat Commun 2017;8:15049. https://doi.org/10.1038/ncomms15049.
[44] Qin X, Li X, Liu Y, Zhou R, Xie J. Multi-agent cooperative target search based on reinforcement Learning. J Phys Conf Ser 2020:1549. https://doi.org/10.1088/1742-6596/1549/2/022104.
[45] Levin E, Pieraccini R, Eckert W. Using Markov decision process for learning dialogue strategies. In: Proc. 1998 IEEE Int. Conf. Acoust. Speech Signal Process. ICASSP 98 Cat No98CH36181. vol. 1; 1998. p. 201–4. vol.1, https://doi.org/10.1109/ICASSP.1998.674402.
[46] Dey S, Marzullo T, Zhang X, Henze G. Reinforcement learning building control approach harnessing imitation learning. Energy AI 2023:14. https://doi.org/10.1016/j.egyai.2023.100255.
[47] Sutton R, Barto A. Reinforcemnet learning. In: An Introduction. Second. Cambridge, Massachusetts: The MIT Press; 2023.
[48] Li Y, Wang Z, Xu W, Gao W, Xu Y, Xiao F. Modeling and energy dynamic control for a ZEH via hybrid model-based deep reinforcement learning. Energy 2023:277. https://doi.org/10.1016/j.energy.2023.127627.
[49] Li G, Shi L, Chen Y, Chi Y. Breaking the sample complexity barrier to regret-optimal model-free reinforcement learning. Inf Inference J IMA 2023;12:969–1043. https://doi.org/10.1093/imaiai/iaac034.
[50] Yang F, Gao F, Liu B, Ci S. An adaptive control framework for dynamically reconfigurable battery systems based on deep reinforcement Learning. IEEE Trans Ind Electron 2022;69:12980–7. https://doi.org/10.1109/TIE.2022.3142406.
[51] Nagy Z, Henze G, Dey S, Arroyo J, Helsen L, Zhang X, et al. Ten questions concerning reinforcement learning for building energy management. Build Environ 2023;241:110435. https://doi.org/10.1016/j.buildenv.2023.110435.
[52] Nachum O, Norouzi M, Xu K, Schuurmans D. Bridging the gap between value and policy based reinforcement. Learning 2017. https://doi.org/10.48550/arXiv.1702.08892.
[53] Gao C, Wang D. Comparative study of model-based and model-free reinforcement learning control performance in HVAC systems. J Build Eng 2023;74:106852. https://doi.org/10.1016/j.jobe.2023.106852.
[54] Zhuang D, Gan VJL, Duygu Tekler Z, Chong A, Tian S, Shi X. Data-driven predictive control for smart HVAC system in IoT-integrated buildings with time-series forecasting and reinforcement learning. Appl Energy 2023;338:120936. https://doi.org/10.1016/j.apenergy.2023.120936.
[55] Biemann M, Scheller F, Liu X, Huang L. Experimental evaluation of model-free reinforcement learning algorithms for continuous HVAC control. Appl Energy 2021;298:117164. https://doi.org/10.1016/j.apenergy.2021.117164.
[56] Suman S, Rivest F, Etemad A. Toward personalization of user preferences in partially observable smart home environments. IEEE Trans Artif Intell 2023;4:549–61. https://doi.org/10.1109/TAI.2022.3178065.
[57] Fu Y, Xu S, Zhu Q, O'Neill Z, Adetola V. How good are learning-based control v.s. model-based control for load shifting? Investigations on a single zone building energy system. Energy 2023;273:127073. https://doi.org/10.1016/j.energy.2023.127073.
[58] Wang D, Zheng W, Wang Z, Wang Y, Pang X, Wang W. Comparison of reinforcement learning and model predictive control for building energy system optimization. Appl Therm Eng 2023;228:120430. https://doi.org/10.1016/j.applthermaleng.2023.120430.
[59] Zhou X, Du H, Sun Y, Ren H, Cui P, Ma Z. A new framework integrating reinforcement learning, a rule-based expert system, and decision tree analysis to improve building energy flexibility. J Build Eng 2023;71:106536. https://doi.org/10.1016/j.jobe.2023.106536.
[60] Coraci D, Brandi S, Hong T, Capozzoli A. Online transfer learning strategy for enhancing the scalability and deployment of deep reinforcement learning control in smart buildings. Appl Energy 2023;333:120598. https://doi.org/10.1016/j.apenergy.2022.120598.
[61] Hosseini S, Hajialigol P, Aghaei M, Erba S, Nik V, Moazami A. Improving climate resilience and thermal comfort in a complex building through enhanced flexibility of the energy system. In: 2022 Int. Conf. Smart Energy Syst. Technol. SEST; 2022. p. 1–6. https://doi.org/10.1109/SEST53650.2022.9898453.
[62] Nik VM. Making energy simulation easier for future climate – synthesizing typical and extreme weather data sets out of regional climate models (RCMs). Appl Energy 2016;177:204–26. https://doi.org/10.1016/j.apenergy.2016.05.107.