Data-Driven District Energy Management with surrogate models and deep reinforcement learning

Giuseppe Pinto [a], Davide Deltetto [a], Alfonso Capozzoli [a*]

[a] *Politecnico di Torino, Department of Energy, BAEDA Lab, Corso Duca degli Abruzzi 24, 10129 Torino, Italy*

* Corresponding author: Tel: +39-011-090-4413, fax: +39-011-090-4499, e-mail: alfonso.capozzoli@polito.it

Abstract

Demand side management at district scale plays a crucial role in the energy transition process, being an ideal candidate to balance the needs of both users and grid, by managing the volatility of renewable sources and increasing energy flexibility. The presented study aims to explore the benefits of a coordinated approach for the energy management of a cluster of buildings to optimise the electrical demand profiles and provide services to the grid without penalising indoor comfort conditions. The proposed methodology makes use of a fully data-driven control scheme which exploits Long Short-Term Memory (LSTM) Neural Networks, and Deep Reinforcement Learning (DRL). A simulation environment is introduced to train a DRL controller to manage the operation of heat pumps and chilled and domestic hot water storage for a cluster of four buildings. LSTM models are trained with synthetic data set created in EnergyPlus and are integrated into simulation environment to evaluate the indoor temperature dynamics in each building. The developed DRL controller was tested against a manually optimised Rule Based Controller (RBC). Results show that the DRL algorithm is able to reduce the overall cluster electricity costs, while decreasing the peak energy demand by 23% and the Peak to Average Ratio (PAR) by 20%, without penalizing indoor temperature control.

Keywords: Coordinated Energy management, Deep Reinforcement Learning, Long Short-Term Memory Neural Network, Data-driven modelling, Building Energy Flexibility

1. Introduction

Building sector accounts for 40% of global energy consumption, thus playing a key role in the energy transition process [1]. The increasing population and rapid urbanization are causes of the growing energy demand, which can be sustainably satisfied by exploiting in a great extent renewable energy source (RES). However, the volatility of renewable energy production can lead to potential grid instability [2]. In that scenario, demand side management (DSM) has become relevant, considering the high operating and maintaining cost of flexibility sources on supply side [3]. In addition, proper DSM strategies can represent an additional source to increase supply efficiency and reducing investment cost related to facilities for the centralised generation, transmission and distribution [4].

DSM strategies can contribute to exploit building energy flexibility, defined as the ability of adapting energy consumption without compromising technical and comfort constraints [5]. This could be achieved especially by means of thermal and electric storage, that allow to decouple energy demand and local production, shifting the energy consumption from period of high electricity price to period of low electricity price. However, the adoption of price-based programs could lead to new undesirable peaks of demand (peak shifting) during periods with low electricity prices [6]. Moreover, the energy flexibility of a single building is typically too small to be bid into a flexibility market, highlighting the necessity to analyse the aggregated flexibility provided by a cluster of buildings [7]. To overcome these limitations, a novel approach to energy management is represented by the coordinated energy management for cluster of buildings, which aims to exploit the benefits of DSM and demand response (DR) programs while avoiding peak rebounds on the grid and enhancing energy flexibility. Coordinated management of buildings can be addressed with a centralised or decentralised control [8]: in the first case a single controller is assumed to have the information on the current states of the entire building cluster, while in the second case decentralised controllers can act at single building level [9]. By means of a coordinated energy management a number of buildings can cooperate or compete to achieve global objectives (e.g., cost minimisation, peak shaving) [10]. The following

subsection provides an overview on coordinated energy management related works, together with identified literature gaps and paper contributions.

1.1 Related works

The coordinated approach in building energy management has recently received a lot of interest, with particular attention to the participation in demand response programs [11,12] mainly through the exploitation of electric vehicles charging strategies [13], schedulable appliances [14,15], peer-to-peer transactions [16] and incentive based programs [17].

Despite many studies have demonstrated the advantages of optimal management of HVAC systems, by means of adaptive [18] and predictive [19,20] control strategies, few efforts have been devoted to the coordination of their operation for a cluster of buildings. In fact, most of recent works reported in literature exploited co-simulation environment based on white-box modelling such as Modelica [21] and EnergyPlus [22] to perform energy management strategies at single building scale. However, when the interest shifts from a single building to a cluster of buildings, the computational cost associated to the simulation of energy performance of building and HVAC system is not negligible, making the forward approach unsuitable for the effective implementation of online control strategies. Early studies tried to face the computational burden of district energy management by decoupling building energy demand and local production, focusing the attention on the formulation of control strategy for supply systems coupled with thermal [23,24] and electrical storage [25]. In those cases, the control strategies act on HVAC system or storage operation to meet ideal building energy demand that are pre-calculated by considering fixed schedule of indoor set-point conditions. By adopting this modeling approach, storage control strategies have showed to be effective in providing grid services at both single building [26] and multiple buildings scale [27,28] or in improving energy management [29]. However, this approach poses limitations to the exploitation of building thermal mass and indoor temperature control as additional flexibility source.

The authors in [30] discussed the importance of modeling building dynamics to control HVAC systems operation for the effective implementation of DR programs. Moreover, in [31] and [32] the

trade-off between thermal demand reduction and acceptable indoor temperature was analyzed; while in [33] the same concept was extended to a cluster of buildings, highlighting the role of indoor set point temperature as a key flexibility source.

Some recent works [34,35] assessed the advantages to implement model predictive control (MPC) for regulating HVAC systems and controlling indoor temperature in a cluster of buildings.

However, the main barriers behind the implementation of district energy management are represented by *i)* the computational cost necessary to properly model local supply systems and energy demand considering indoor temperature control in each building of the cluster *ii)* the complexity associated to the optimization of a district of buildings, characterized by different energy systems and energy demand patterns.

To overcome computational cost necessary to model multiple buildings considering the indoor temperature evolution, a recent approach takes advantage from the implementation of data driven models (e.g., artificial neural networks (ANN). This opportunity has gained popularity in recent years, due to the increasing availability of building-related data and to the necessity of computationally lightweight the predictive models of indoor environmental conditions. Ruano et al. [36] proposed a radial basis function neural network to predict the indoor air temperature of a public building, while in [37] and [38] nonlinear autoregressive models were exploited for the same purpose. In [39] a long-short term memory (LSTM) neural-network was employed to predict the indoor air temperature in a multi zone building..

Due to their versatility, neural networks have been exploited to predict the indoor microclimatic conditions, also coupling them with advanced control strategies. Huang et al. [40] implemented a predictive controller coupled with a neural network predicting the indoor air temperature of a multi-zone building, to optimise the start and stop of an HVAC system. In [41] the application of an autoregressive neural network for the indoor temperature prediction integrated in a fuzzy logic controller was implemented to regulate the volumetric flow rate supplied by the HVAC system..

However, while this approach seems to be well-established at single building scale, it has not been fully explored at district level, due to the computational complexity associated to model based control. Moreover, MPC showed good performances when applied at single building scale but at the expense of defining detailed models, whose development requires a great effort.

Recent research has tried to develop more efficient model-free control, such as Reinforcement Learning (RL). Reinforcement learning is less expensive to be implemented because it does not require a model of the system and could learn through the interaction with both the environment and historical data. Moreover, a peculiarity of the RL lies in its adaptability [43], making it able to automatically adapt to the environment's changes, as well as to human preferences, that can be directly integrated into the control logic.

RL controllers have proven to be effective to control the operation of several energy systems in residential or commercial buildings, including gas boiler [44], electric water heater [45], domestic hot water (DHW) [46] or heat pumps [47]. In addition, [48] deeply reviewed the application of RL for demand response, emphasizing the opportunity provided by such control approach. Recently, few studies have started to put emphasis on the application of reinforcement learning for the cooperative and competitive coordination mechanisms [49] to account for demand peak shifting in cluster of buildings [24].

In Figure 1 a Venn diagram is reported, with the aim to underline the different main contributions provided by some relevant papers presented in literature in the field of the building energy management. The diagram shows that most of the previous works focused on the energy management strategies with specific objectives at single building scale, namely on demand response and grid-interaction, demand side management and indoor temperature control, or demand independent supply side management. Few papers have been focused on multiple building coordination, albeit analysing only specific aspects of the energy management problem. This paper intends to provide a contribution that accounts the multi-objectives nature of energy management at a district scale.

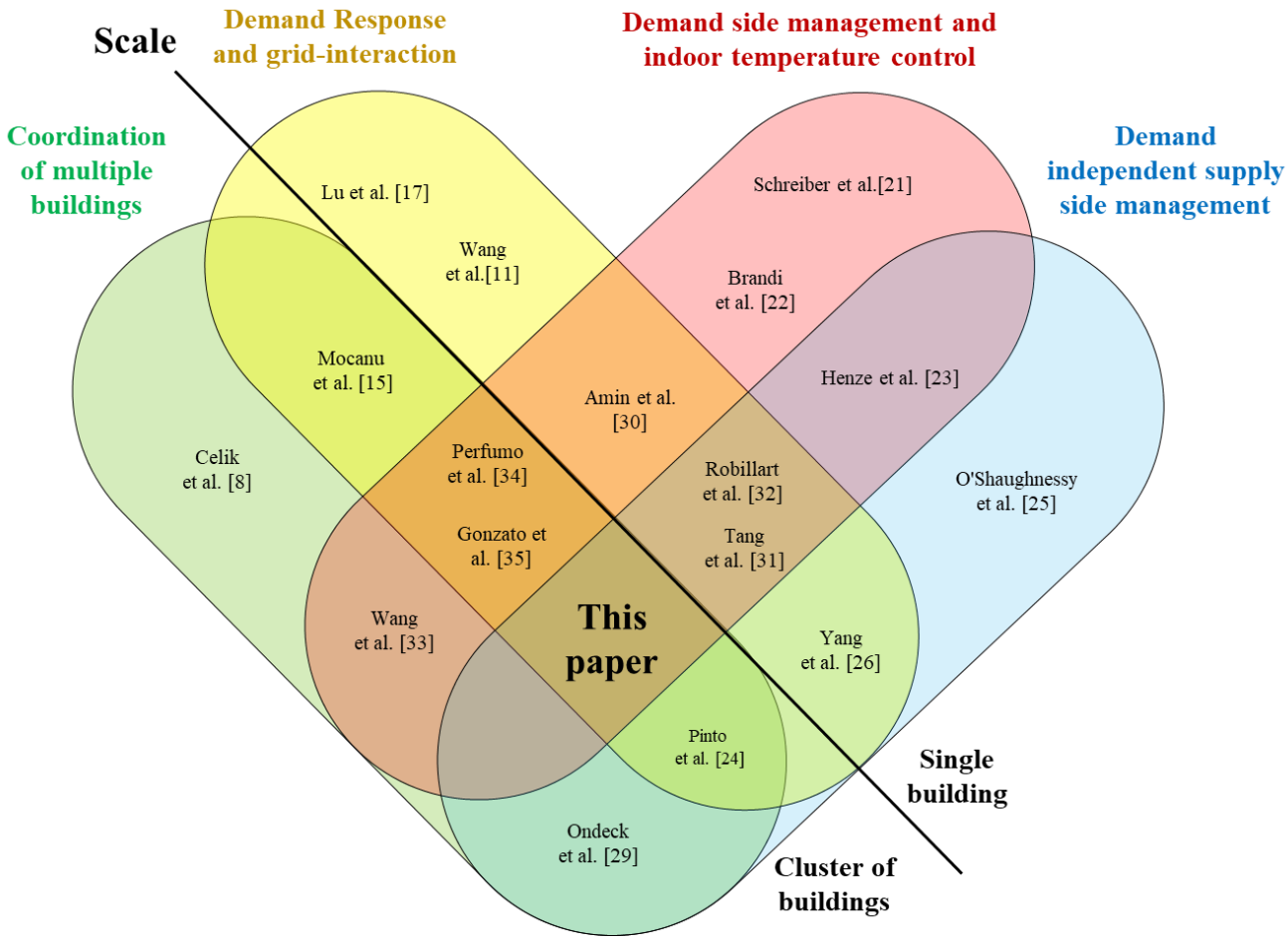

Figure 1: Venn diagram displaying the four pillars of advanced control for district energy management: buildings coordination, grid-interaction, indoor comfort and storage technologies

In summary, the following gaps have been identified from the existing literature, which require further investigations:

1. Current energy management strategies for multiple buildings mainly focused on the coordination of schedulable appliances, neglecting the potentialities of controlling HVAC systems.
2. Coordinated district energy management was often implemented only on local production side, considering pre-computed ideal building energy demand. This approach disregards to assess user comfort and to exploit the indoor temperature control as an additional flexibility source.

3. The control optimization of multiple energy systems is challenging with MPC, which requires huge effort for model development and lacks adaptability. In this context, RL seems to provide a viable alternative that needs to be still analysed for large scale environment.

To overcome current limitations of district energy management, this paper proposes a fully data-driven framework to coordinate multiple energy systems (heat pumps and thermal storage) for a group of four buildings modelling the building thermal dynamics and the indoor temperature evolution by means of deep neural networks (DNN).

To this purpose a new simulation environment, CityLearn [50], was used and specifically built to enable training and evaluation of reinforcement learning models in a cluster of buildings. The centralised DRL controller was designed to coordinate the energy demand of four buildings, by controlling the cooling power supplied by the heat pump and the operation of cold and DHW thermal storage for optimising both operational costs and peak demand without jeopardizing indoor temperature control.

The primary contributions of the present paper can be summarized as follows:

1. A number of LSTM neural networks were developed to predict the indoor temperature evolution of different buildings with the aim of reducing computational cost needed to take into account of the building dynamics at district level.
2. The forecasting models were integrated into a data-driven simulation environment (CityLearn), with the possibility to coordinate the control of heat pumps and thermal storage considering the indoor temperature evolution during the optimization process.
3. A Soft Actor Critic (SAC) reinforcement learning agent was implemented to coordinate the energy demand, indoor comfort, and grid-interaction for a cluster of four buildings, analysing the effect of the coordinated management on multiple levels.

The paper is organised as follows: Section 2 introduces the methods adopted, including LSTM neural network architecture and RL algorithm. Then, Section 3 describes the case study and the control

problem. Section 4 introduces the proposed methodological framework, while Section 5 describes the implementation of the methodology, with particular attention to the training process and controller design. Section 6 presents the results of the training and deployment phase, while discussion of results is given in Section 7. Lastly, conclusion and future works are reported in Section 8.

2. Methods

2.1. Long Short-Term Memory Neural Networks

Long Short-Term Memory networks, usually just called "LSTMs", are a special kind of recurrent neural networks, capable of learning long-term dependencies [51]. This property of LSTMs is due to a particular gating mechanism and to the presence of two states:

- Hidden state: responsible of maintaining the short-term memory.
- Cell state: responsible of maintaining the long-term memory and capturing long term dependencies.

The scheme of the LSTM cell is shown in Figure 2.:

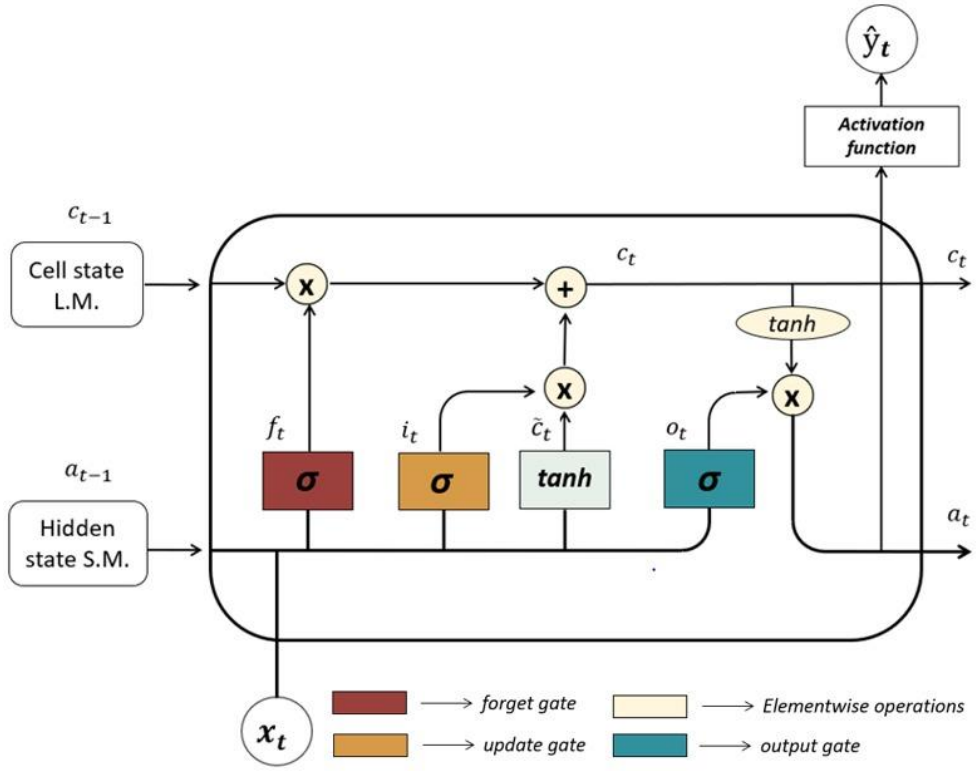

Figure 2: LSTM architecture

The main feature of LSTMs is the cell state, which is responsible for maintaining long term dependencies: information is removed or added to the cell state by means of three gates.. The forget gate decides what information has to be deleted from the cell state, the update gate decides which information is going to be stored in the cell state and the output gate is used to compute the output of the LSTM.

2.2.Reinforcement Learning

Reinforcement learning is a branch of machine learning specialized in solving control problems. It combines the advantages of dynamic programming, with a trial-and-error approach. RL uses an agent-based control, in which the agent learns through the interaction with the controlled environment.

Reinforcement learning can be formalized as a Markov decision process (MDP), a discrete-time stochastic control process [52]. MDP is useful when the decision maker deals with partly random or unknown environment.

Markov Decision Process are represented using a 4-tuple (S, A, P, R) made up of:

1. *S*: State space

   The states represent a mathematical description of the environment.

2. *A*: Action space

   The action is the decision made by the agent on how to control the environment.

3. *P*: Transition probability

   The transition probability $P(s_{t+1} = s' | s_t = s, a_t = a) = P: S \times A \times S'$ is the probability that, starting in s and performing action a at the time t, the next state will be s'.

4. *R*: Reward function

   The reward function is used to map the immediate reward r with the tuple $S \times A \times S'$.

The ultimate goal of the agent is to find the optimal control Policy (π). A control policy is a mapping between states and actions $\pi: S \rightarrow A$, and it has the aim to maximize the cumulative reward over a time horizon, called return $G = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}$. Return is defined as a discounted cumulative where

$\gamma \in [0,1]$ is the discount factor for future rewards. An agent employing γ equal to 1 considers future rewards as important as current ones, while an agent with γ equal to 0 assign higher values to states that lead to high immediate rewards. For sake of clarity an example with an energy system is provided in Figure 3.

In that case the controller (*agent*) is connected to a heat pump. The controller has the role of minimising electricity cost while guaranteeing comfort conditions (*reward function*). The amount of cooling power delivered to the building (*actions*) influences both terms of the reward charging and discharging (*actions*) the storage to satisfy the demand. The exploitation of information such as heat pump efficiency or low electricity price period (*states*) allows the controller to find the optimal policy and maximize the reward.

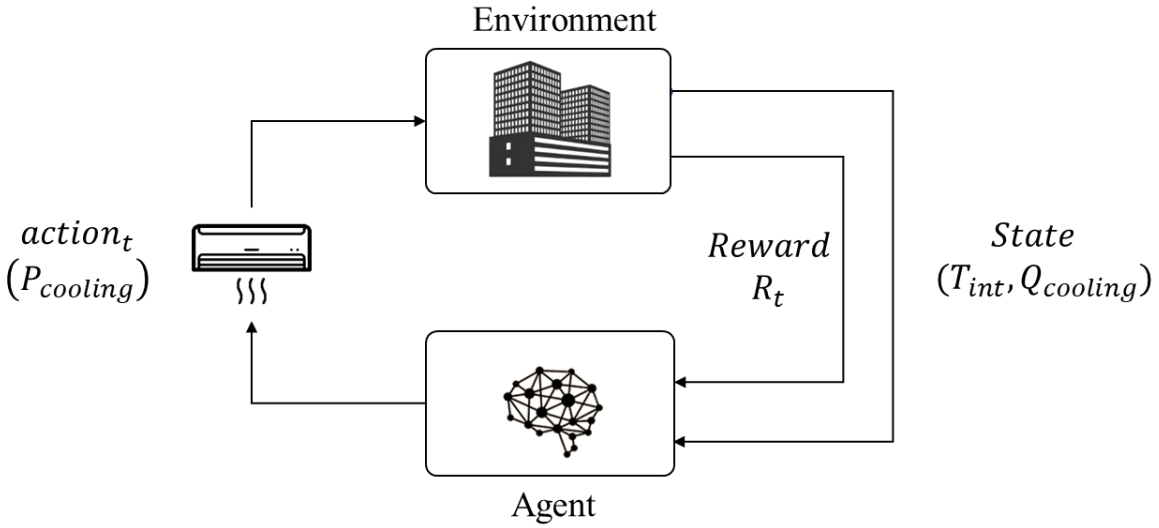

Figure 3: Reinforcement learning control framework

Among RL algorithms, the most used one, due to its simplicity, is Q-learning [53]. In Q-learning transitions can be represented with a tabular approach that stores the state-action values (Q-values) that are updated as follows:

$$Q(s,a) \leftarrow Q(s,a) + \mu[Q(s,a) + \gamma \max_{a}(Q(s',a') - Q(s,a)] \quad (1)$$

Where s' is the next state and μ ∈ [0,1] is the learning rate, which determines to what degree new knowledge overrides old knowledge. When μ is equal to 1 new knowledge completely substitutes old knowledge, while for μ set equal 0 no learning happens. Despite the advantages, a tabular representation of real-world problem may be unfeasible, due to large state and action spaces that needs to be stored.

2.2.1. Soft actor critic

The combination of RL and high-capacity function approximators such as deep neural networks renewed the interest for the RL topic and promoted its extension to complex problems [54]. Among Deep Reinforcement Learning (DRL) algorithms, an actor-critic method was selected for its ability to combine advantages of both value-based and policy-based methods.

The key components of soft actor-critic [55] are:

- An actor-critic architecture used to map policy and value function with different networks.
- The off-policy formulation that allows reusing previously collected data, stored in a replay buffer (D) to increase data efficiency.
- The entropy maximization formulation, that helps stabilize the algorithm and the exploration.

Differently from the standard RL algorithm, maximum entropy reinforcement learning optimises policies to maximize both the expected return and the expected entropy of the policy as follows:

$$\pi^* = \arg\max_{\pi_\phi} \sum_{t=0}^{T} \mathbb{E}_{(s_t,a_t)\sim\rho_\pi} \left[r(s_t, a_t) + \alpha H\left(\pi_\phi(\cdot\,|s_t)\right) \right] \quad (2)$$

Where $(s_t, a_t)_{\sim\rho_\pi}$is a state-action pair sampled from the agent's policy, and $r(s_t, a_t)$ is the reward for a given state-action pair. Due to the entropy term, H, the agent attempts to maximize the returns while behaving as randomly as possible.

SAC is influenced by the temperature parameter α, that determines the relative importance of the entropy term against the reward, and thus controls the stochasticity of the optimal policy. A high value of the temperature parameters may lead to a uniform behaviour, while a low value of the temperature parameter will only maximize the reward. To reduce the effort of tuning this hyper-

parameter, this paper exploits a recent version of the SAC that employs alpha automatic optimization [56]. To ease the comprehension, the algorithm is summarized in Table 1.

Table 1: soft actor-critic

Input: θ_1, θ_2, ϕ	Initial parameters
$\bar{\theta}_1 \leftarrow \theta_1, \bar{\theta}_2 \leftarrow \theta_2$	Initialize target network weights
$\mathcal{D} \leftarrow 0$	Initialize an empty replay buffer
for each iteration **do**	
for each environment step **do**	
$a_t \sim \pi_\phi(a_t \mid s_t)$	Sample action from the policy
$s_{t+1} \sim p(s_{t+1} \mid s_t, a_t)$	Sample transition from the environment
$\mathcal{D} \leftarrow \mathcal{D} \cup \{(s_t, a_t, r(s_t, a_t), s_{t+1})\}$	Store the transition in the replay buffer
end for	
for each gradient step **do**	
$\theta_i \leftarrow \theta_i - \lambda_Q \nabla_{\theta_i} J_Q(\theta_i)$ for $i \in \{1,2\}$	Update of the Q-function parameters
$\phi_i \leftarrow \phi_i - \lambda_\pi \nabla_\phi J_\pi(\phi)$	Update policy parameters
$\alpha \leftarrow \alpha - \lambda \nabla_\alpha J(\alpha)$	Adjust temperature
$\bar{\theta}_i \leftarrow \tau\bar{\theta}_i + (1-\tau)\bar{\theta}_i$ for $i \in \{1,2\}$	Soft update of the target network weight
end for	
end for	
Output: θ_1, θ_2, ϕ	Optimised parameters

2.3.The CityLearn simulation environment

To train and deploy the developed RL controller CityLearn simulation environment [57] was adopted. The aim of the environment is to ease and standardize the implementation of RL agents in smart cities. In particular, it allows to control multiple energy storage devices within each building of a cluster, including domestic hot water and chilled water tanks for the storage of cooling and heating energy. The cooling energy is supplied by air-to-water heat pumps, and the heating energy is supplied by electric heaters, taking into account the presence of photovoltaic plants. The environment is highly flexible, allowing the implementation of both centralised and decentralised control agents, together

with the possibility to easily add other technologies and state variables. A detailed description of the CityLearn environment and its architecture is presented in [50], while the code is available at [58].

3. Case study and control problem

The proposed methodology, described in detail in next section, is applied to a cluster of 4 commercial buildings, including a small office, a retail, a restaurant and a medium office. The four buildings analyzed belong to commercial reference buildings developed by U.S. Department of Energy (DOE). The energy demand of the buildings was evaluated from June to October considering the Albuquerque (New Mexico, 4B climate zone) climatic conditions.

Each building is equipped with a heat pump, hot and cold thermal storage and electric heater to meet heating, cooling and domestic hot water energy demand respectively. Figure 4 shows a schematic of the control architecture with a detail on energy systems for a representative building. Heat pump serves for both space heating and cooling, with the possibility to charge the cold storage and/or to directly supply cooling energy in order to control indoor temperature, while electric heater and hot storage meet DHW demand. Moreover, the heat pump operation at part load conditions was modelled according to UNI EN 14825 [59]. The energy systems are managed by a centralised controller, which conceived to optimise operational costs and to flatten the aggregated load profile of the entire cluster reducing peaks on the grid.

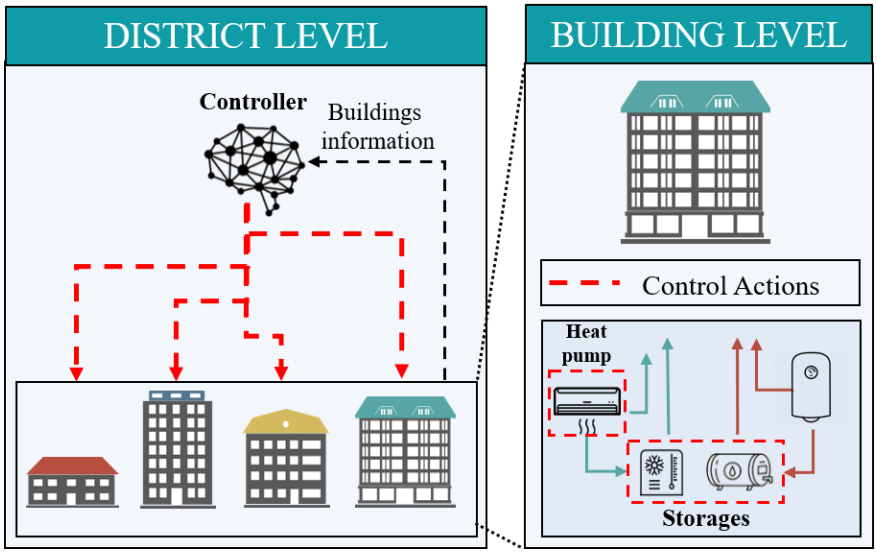

Figure 4: Schematics of the district energy management systems analysed

To simulate a realistic scenario, a variable electricity price was considered, with a tariff varying from 0.03025 $/kWh during off-peak night-time period (8 p.m. - 7 a.m.) to 0.06605 $/kWh during on-peak day-time period (7 a.m. - 8 p.m.). Table 2 reports the geometrical features and the nominal capacity of the different systems for each building analyzed, including the PV capacity installed only in Building 4.

Table 2: Building and energy systems properties

	Type	Surface [m^2]	Volume [m^3]	Heat Pump Capacity [kW]	Cold Storage Capacity [kWh]	Hot Storage Capacity [kWh]	PV Capacity [kW]
Building 1	Small Office	511	2280	31	53	0	0
Building 2	Retail	2294	13993	130	225	6	0
Building 3	Restaurant	511	2415	95	162	50	0
Building 4	Medium Office	4981	19777	295	505	13	120

Figure 5 summarizes the electric load profile for three days of simulation for each building calculated with EnergyPlus, together with aggregated load profile of the entire cluster of buildings. In detail, the bottom part of the figure shows the aggregated load profile, highlighting the contribution of the

photovoltaic generation on the right. The breakdown of the total electrical load is reported on the left, considering appliances (non-shiftable), DHW and cooling demand. This representation is useful to underline cooling and DHW contribution, on which controller can act to enhance the flexibility of the cluster of buildings. Due to the high cooling demand needed to maintain indoor comfort condition, the analysis was focused only on the summer period (1st June to 31st August).

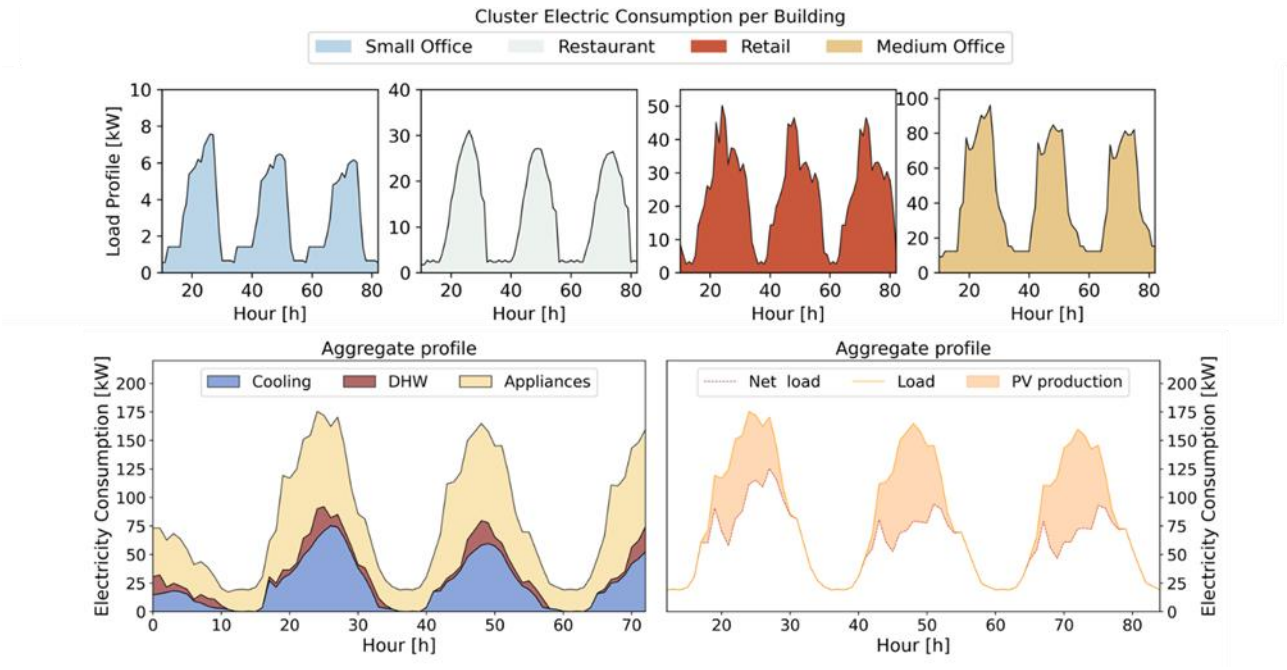

Figure 5: Load profile for each building (up) and cluster electricity profile and PV production (down)

4. Methodology

The methodology takes advantage of two machine learning techniques to fully exploit the energy flexibility associated to a cluster of buildings using a coordinated energy management approach. As shown in Figure 6 the methodological framework exploits LSTM neural networks to predict indoor temperature evolution for each building. The neural networks were trained with synthetic datasets obtained through the modelling of each building with EnergyPlus. LSTM models were then coupled with CityLearn simulation environment to enable also the possibility to act on heat pump to control the indoor temperature, overcoming a limitation of the CityLearn environment, which allowed to work only with a pre-computed building energy demand.

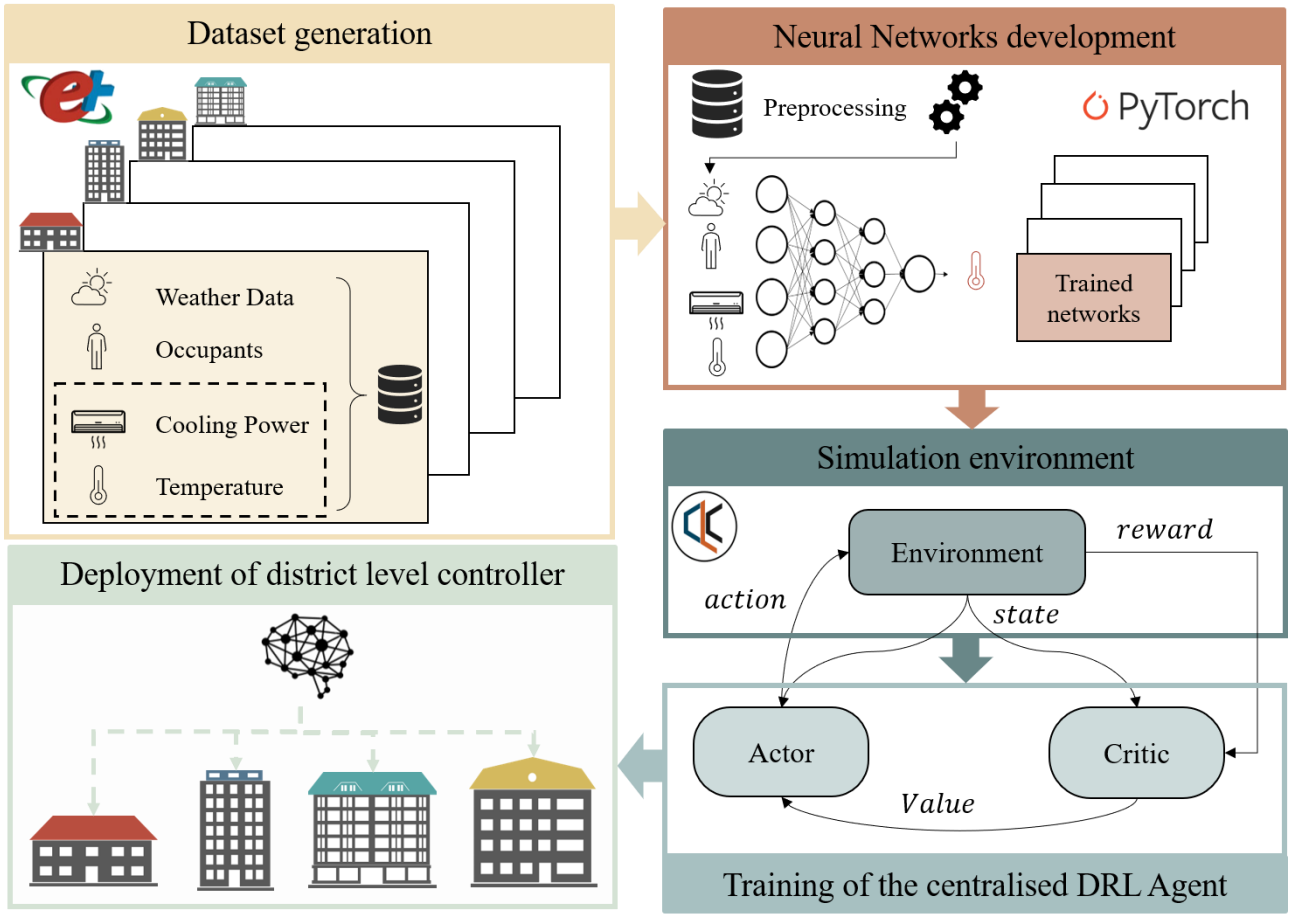

Figure 6: Proposed framework for the district energy management

Then, a centralised DRL controller based on SAC algorithm was trained and deployed to perform a coordinated control of the energy systems of the cluster of buildings. Eventually, after a trial-and-error interaction with the environment, the agent learnt how control indoor temperature in the different buildings, coordinating heat pump and storage operation to reduce costs and peak demand.

4.1.Development of artificial neural networks

In order to generate a labelled dataset for training and testing LSTM models, the four buildings of the cluster were preliminary modelled and simulated through EnergyPlus. For each building, several simulations were performed to analyse the effect of supply cooling load on indoor temperature. In particular, the set of simulations designed to create the synthetic data set include the variation of the percentage of cooling load supplied with respect to EnergyPlus ideal load.

The synthetic dataset resulted of 11520 rows with an hourly granularity corresponding to 4 months of hourly simulations obtained by randomly varying the supply cooling load. The variables reported

in Figure 7 were used as input variables of the DNNs to predict indoor temperature for each building. In particular, to assess the feasibility towards a real-world implementation, were selected time variables, weather variables, together with the cooling load and the internal temperature related to previous time steps. Temporal variable was encoded using sine and cosine transformation and data was normalized using a min-max normalization. Then, a series-to-supervised procedure was performed using a sliding window. Since the aim of the problem is to forecast the internal temperature, the latter has a lag of one hour with respect to the other variables.

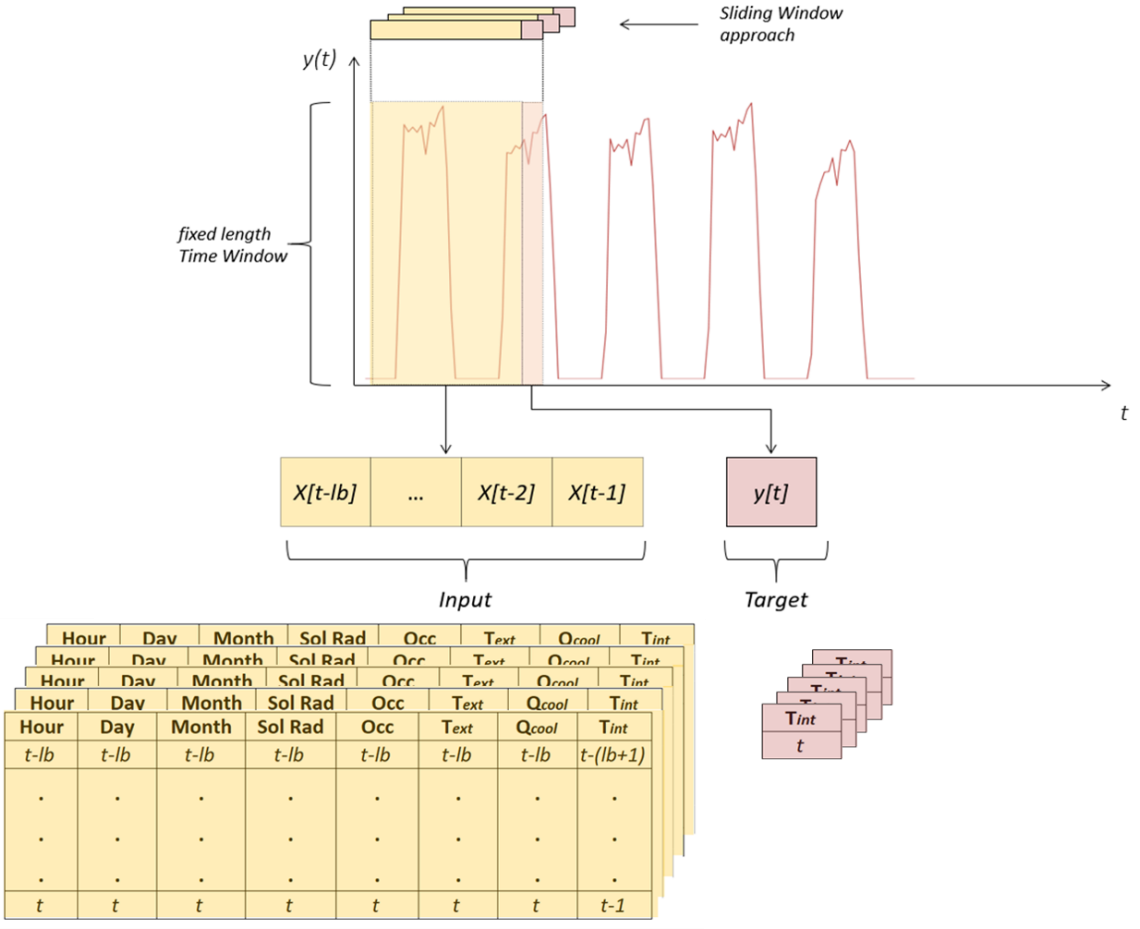

Figure 7: Sliding window approach and DNN inputs

To select the best architecture for each LSTM model, different hyperparameters were analysed. A sensitivity analysis was performed changing batch size, number of hidden neurons and layers, lookback and learning rate iteratively and finally selecting the set of parameters, after a training period

of 100 epochs, leading to the highest accuracy. The accuracy was evaluated computing the following metrics:

- $RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)}{n}}$
- $MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right|$

The selected parameters resulted from the sensitivity analysis for each neural network are reported in Table 3:

Table 3: Hyperparameters for each building model

	Small Office	Retail	Restaurant	Medium Office
Batch size	100	100	100	100
n° hidden	8	8	8	50
Lookback	12	12	12	12
Learning rate	0.008	0.005	0.008	0.005
n° layers	2	2	2	1

4.2.Deployment strategy of the neural network

The trained neural networks were then tested both in one step ahead and recursive configuration. This latter is a strategy to perform multi step ahead predictions in simulation mode as shown in Figure 8. More in detail, a single model is trained to perform one-step ahead forecast given the input sequence. Then, during the operational phase, the forecasted output is recursively fed back and used as input for the next predictions. The recursive neural networks were integrated into CityLearn environment, adding the possibility to simulate the evolution of the indoor temperature in each building of the district.

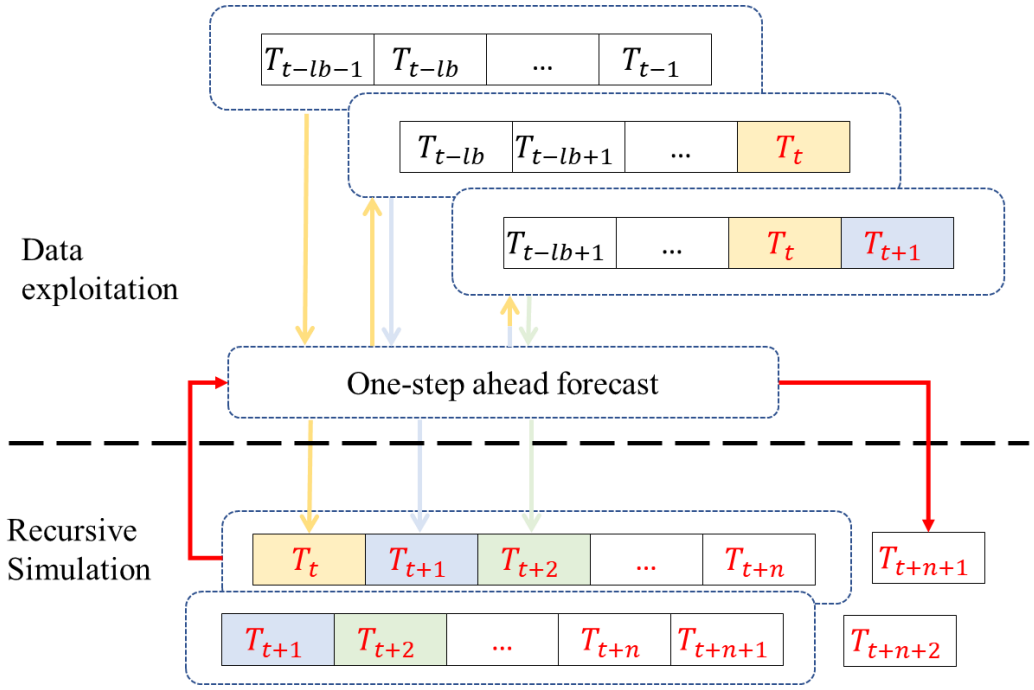

Figure 8:Recursive strategy used to perform indoor temperature prediction

The coupling of the trained neural networks with CityLearn, provided twofold advantages: first, in addition to controlling storage operation, the possibility to control the cooling energy supplied by heat pumps; furthermore, the interactions of the neural networks with the controller allowed to evaluate the indoor temperature evolution in each building, with the opportunity to find the trade-off between comfort, energy consumption and grid requirements.

4.3. Training of the centralised DRL

After defined the environment, the control problem was formulated. Firstly, the *action-space* was set, which represents the set of possible control actions performed by the agent. Then the *state-space*, a set of variables related to the controlled environment, was defined and fed to the agent to learn the optimal control policy. Lastly, the *reward function* was formulated to optimise the operation of the agent according to the control objectives. The agent was trained in an offline fashion using a training episode multiple times to constantly refine the control policy [47].

4.4.Deployment of the centralised DRL

The agent was statically deployed in the same climatic conditions used during the training phase, to assess the effect of the control policy on the objectives. The performances of the DRL controller were

benchmarked against a RBC controller by evaluating several key performances indicators (KPI) including: system costs, maximum peak, average daily peak, peak-to-average ratio (PAR), daily peak-to-average ratio, and flexibility factor [60]. The latter KPI is defined as the ratio between off-peak energy consumption and total energy consumption.

The KPIs have been selected to highlight the advantages of DRL control strategies at single building scale (electricity cost), district scale (maximum peak and peak-to-average ratio) and to evaluate the effect of the coordinated energy management on the grid (average daily peak, daily peak-to-average ratio and flexibility factor).

5. Implementation

The section describes the design of baseline control strategy used as benchmark, followed by a detailed description of the DRL controller design.

5.1.Baseline control

A manually designed rule-based controller was used as a baseline in order to evaluate the performance of the SAC algorithm.

This control strategy was designed to control for each building the heat pump operation to satisfy building cooling demand, and the operation of hot and cold storage. In particular, the heat pump control strategy was designed to satisfy the ideal load of the building, defined as the load necessary to always ensure 26°C when the building is occupied, evaluated through EnergyPlus. This strategy was considered as benchmark to evaluate the effect of a control strategy to meet the ideal cooling load of the building cluster.

In the RBC strategy the actions to control the operation of the storage were optimised to reduce energy costs, taking advantage from the electricity price tariffs. In particular, to limit peak demand, hot and cold storage units are uniformly charged during the night period, exploiting the lower electricity tariff, and discharged during the day homogeneously to flatten the load profile of the entire cluster of buildings.

5.2.Design of the DRL controller

SAC control strategy was conceived to manage energy demand of each building, while satisfying indoor comfort conditions and flattening the aggregated load profile at district level. In the next sub-sections, action space design is presented, along with the description of the state-space and the reward function.

5.2.1. Action-space design

The case study deals with multiple buildings, each one equipped with a heat pump and thermal storage, whose operation can be controlled. The size of the action space is equal to 11 since all buildings except the small office have 3 controlled variables: the heat pump cooling power supply, the chilled water storage charge/discharge and the DHW storage charge/discharge. The actions related to the heat pump cooling power can vary from 0 to 1; the selected action is then multiplied by the available nominal thermal power of the heat pump in the corresponding time step. Moreover, the cooling power delivered to the building is set to 0 during non-occupancy period. The control actions on the storage can vary between -1 and 1. However, considering that a full charge/discharge in a single timestep is not feasible, in this work, the action space was constrained into the interval [-0.33,0.33], imposing therefore a complete charge or discharge time of 3 hours according to [61].

5.2.2. State-space design

The agent learns the optimal control policy observing the effects of its actions on the environment states. The definition of the state space, together with the reward function, is crucial to help the learning process of the controller. In particular, a robust space of states should include variables easy to measure and meaningful. The variables selected are reported in Table 4 and further described below.

Table 4: State-space variables

Variable	**Unit**

Weather	
Temperature	[°C]
Temperature Forecast (6,12,24h)	[°C]
Direct Solar Radiation	[W/m^2]
Direct Solar radiation Forecast (6,12,24h)	[W/m^2]
Diffuse Solar Radiation	[W/m^2]
District	
Electricity Price	[€/kWh]
Electricity Price forecast (1,2,3h)	[€/kWh]
Hour of day	[h]
Day of the week	[-]
Month	[-]
Building	
Non-shiftable load	[kW]
Heat pump efficiency	[-]
PV generation	[kW]
Chilled water Storage SOC	[-]
DHW storage SOC	[-]
Heat pump supply cooling power @t-1	[kW]
Temperature Setpoint	[°C]
ΔT Setpoint - LSTM indoor temperature @t-1	[°C]
Occupancy	[-]

The variables are classified as weather, district and building related variables. Weather information such as the *outdoor air temperature* and *solar radiation* were included into the state space considering the strong influence they have on the cooling load and heat pump efficiency. Moreover, weather forecasts have been introduced to exploit the predictive nature of the controller.

District states include variables common to all buildings, such as *hour of day*, *day of the week, month*, *electricity price* and *electricity price forecast*.

Building states include variables related to the electricity production (*PV generation*) and consumption of the buildings (*non-shiftable load*). Additionally, *heat pump efficiency, cooling and domestic hot water state of charge of storage* were included. Lastly, to characterize building indoor environment, *heat pump supply cooling power* chosen by the agent and temperature difference between the indoor setpoint and that predicted trough the LSTM model during the previous hour (*ΔT Setpoint - LSTM indoor temperature @t-1)* were introduced, together with *occupancy* information. Figure 9 shows the variables included in the state-space and the actions of the DRL controller. The centralised controller receives high-level information (district and weather variables), and low-level information (building variables), to optimise building and district electric electrical load profile.

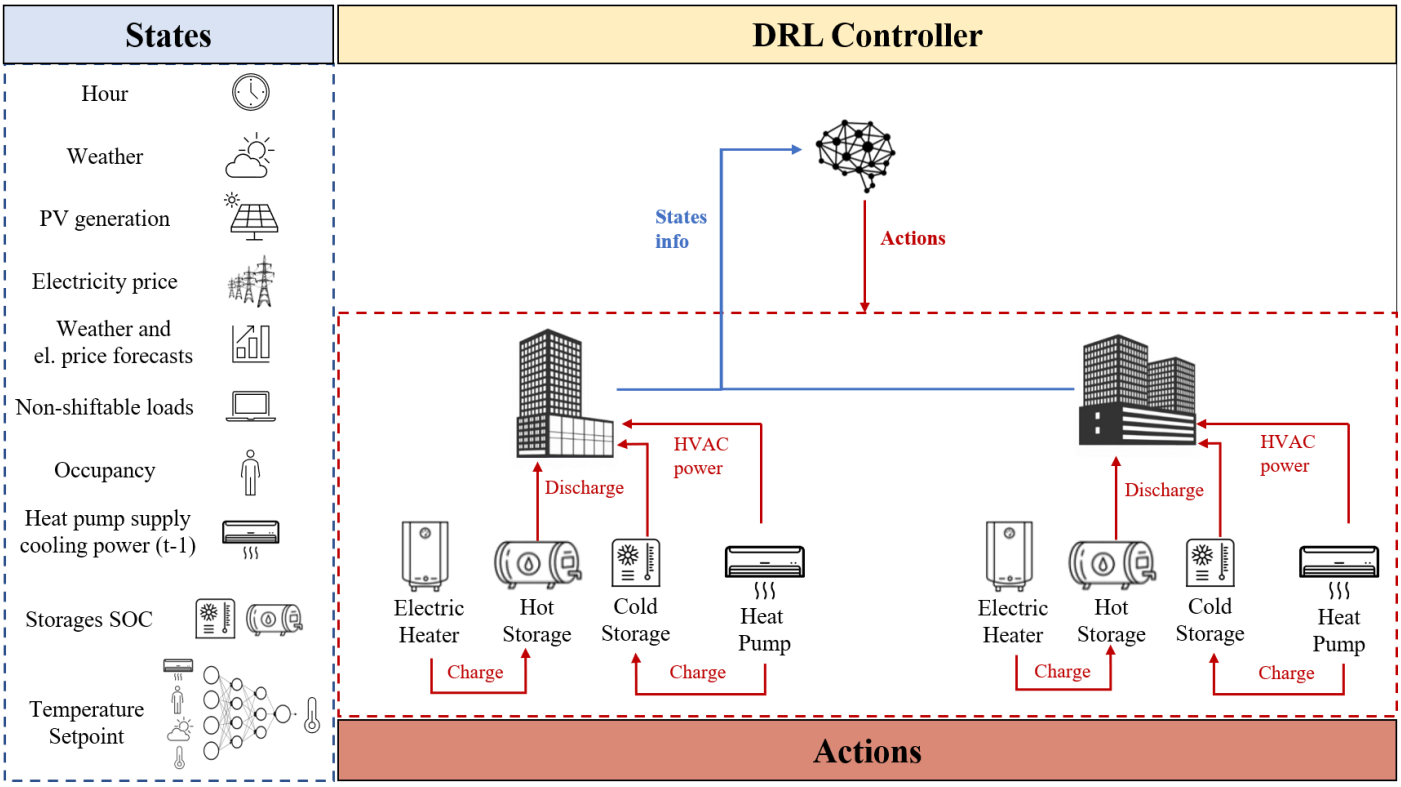

Figure 9: State and action spaces of the control strategy

5.2.3. Reward function

The reward function describes how the agent should behave; it has to be representative of the control problem under attention. In this case study, definition of the reward function was particularly challenging to properly take into account the cluster electrical load profile without jeopardizing

indoor thermal comfort in each building of the cluster. As a result, the defined reward is a combination of different contributions formulated as:

$$R = \sum_{i=1}^{n} Comf_p + \sum_{i=1}^{n} Storage_p + Peak_p \quad (3)$$

where n is the number of buildings.

The *comfort related term* ($Comf_p$) was introduced to minimize the temperature violations, with the aim to maintain the indoor air temperature within a comfort band ranging from 25°C to 27°C.

The comfort term is structured as follows:

$$Comf_p = \begin{cases} -m(SP - T_{in})^3, & T_{in} < T_{low} \\ -m(SP - T_{in}), & T_{low} \leq T_{in} < SP \\ 0, & SP \leq T_{in} < T_{up} \\ -m(T_{in} - SP)^2, & T_{in} \geq T_{up} \end{cases} \quad (4)$$

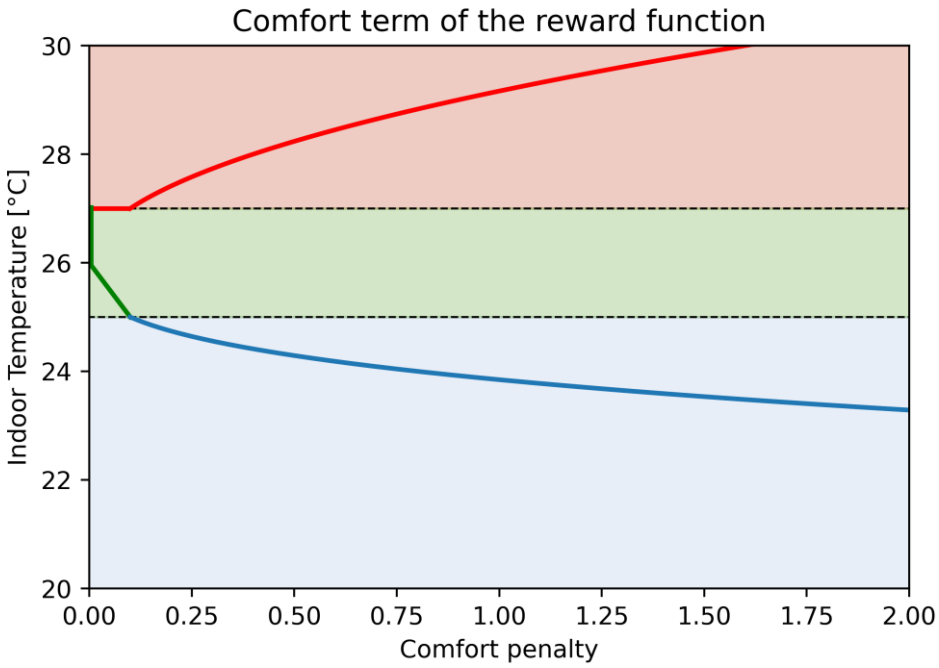

Figure 10: Comfort term of the reward function

The comfort term of the reward, shown in Figure 10, was conceived to encourage the controller to stay as much as possible close to 26 °C, with slight preference towards the 27°C, to consume less energy. When the indoor temperature falls out of the lower or the upper bound of indoor temperature acceptability range, the penalty becomes exponential; for lower temperatures, the exponent is cubic

instead of quadratic since it would generate both thermal discomfort and additional energy consumption.

The *storage price* is the only positive term, and it is computed only during off-peak periods, encouraging charge during the night periods. This term is based on the storage state of charge (SOC) and it is calculated as follows:

$$Storage_p = \max(0, \Delta SOC_{DHW}) * K_1 + \max(0, \Delta SOC_{chilled}) * K_2 \quad (5)$$

Lastly, the peak term is computed starting from the overall district energy consumption. Depending on the electricity price, it assumes different values according to the following equation:

$$Peak_p = \begin{cases} c_{el} = \max(c_{el}), & -\max(0, e - th_1) * K_p \\ c_{el} < \max(c_{el}), & -\left[\max(0, e - th_2) * K_p + \max(0, th_3 - e) * K_p\right] \end{cases} \quad (6)$$

Threshold th_1 was set equal to 120 kW to limit peak demand during peak hours. Moreover, th_2 and th_3, equal to 65 and 35 kW were used to flatten the load curve during off-peak hours. The values of the thresholds were chosen according to the RBC load duration curve, deeply described in 6.2.2: th_1 represents the 99th percentile of the load duration curve, th_2 is the median value and th_3 is the 10th percentile.

The design of the reward function highly influences reinforcement learning performances, and the coefficients $m, K_1, K_2 \; and \; K_p$ in equation (6) weight the relative importance of temperature violations and peak shaving actions. Moreover, since the reward magnitude influences the behaviour of SAC algorithm [55], these coefficients were used to tune exploration-exploitation trade-off of the agent. Their values are shown in Table 5:

Table 5: Reward function coefficients

Coefficient	Value
m	0.12
K_1	3

K_2	2
K_p	0.6

5.2.4. Hyperparameters setting of deep reinforcement learning

Reinforcement learning is characterised by several hyperparameters, which highly influence agent behaviour. To allow the reproducibility of the results, the hyperparameter settings is reported in Table 6. As explained in section 2.2, α highly influences the outcome of the policy, therefore a version of SAC algorithm that optimises the temperature parameter was adopted. For temperature α and entropy coefficient H both starting value and optimised values are reported below.

Table 6: Hyperparameter settings

	Variable	**Value**
1	DNN architecture	2 Layers
2	Neurons per hidden layer	256
3	DNN Optimiser	Adam
4	Batch size	512
5	Learning rate λ	0.001
6	Discount rate γ	0.9
7	Decay rate τ	0.005
8	Temperature* α	Starting = 1, Final = 0.1
9	Entropy coefficient* H	Starting = 8, Final = 5
10	Target model update	1
11	Episode Length	2196 Control Steps (92 days)
12	Training Episodes	30

6. Results

This section describes the results of the implemented framework. Firstly, the results related to the development and training of LSTM models are discussed. Then, the DRL agent performances are

reported at district level and single building level to show outcomes related to comfort and energy system operation.

6.1. Artificial neural network testing results

To check the quality of the developed models, mean absolute percentage error (MAPE) and root mean square error (RMSE) have been computed using a recursive deployment on a testing dataset. The results are summarized in the following table:

Table 7: Evaluation metrics

	MAPE [%]	**RMSE [°C]**
Small Office	0.80	0.28
Retail	0.45	0.15
Restaurant	0.78	0.26
Medium Office	0.81	0.28

As shown in Table 7 a MAPE smaller than 1% was obtained for all models, highlighting the ability of neural networks to capture building thermal dynamics, with a RMSE always smaller than 0.3 °C. Figure 11 shows on the left side the comparison between indoor temperature predicted with LSTM and EnergyPlus for the small office, while on the right is reported the temperature error distributions for each building of the cluster, highlighting the ability of the neural networks to provide an accurate forecasting.

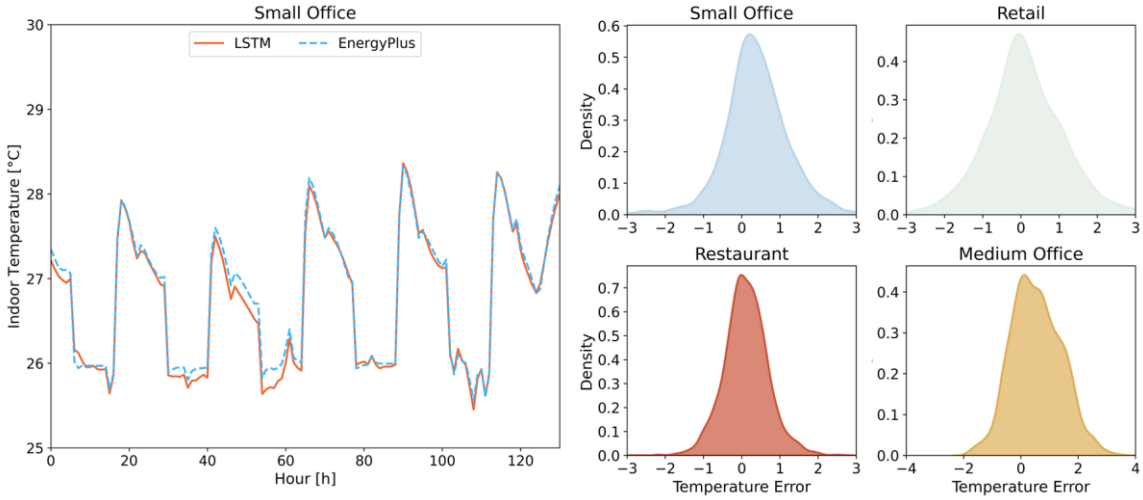

Figure 11: Comparison between indoor temperature predicted with LSTM model and simulated with EnergyPlus (left) and relative error distribution of indoor temperature predicted with LSTM models (right)

6.2.Deployment of the deep reinforcement learning controller

The section presents the results of the developed controller, with particular attention to the benefits provided at district scale, together with a detail on the results of the control strategy on the building indoor temperature control and energy systems operation. Finally, the section includes the results obtained at grid level.

6.2.1. Comparison at district level

The carpet plots in Figure 12 shows a comparison between the aggregate energy consumption at cluster level with the RBC and the DRL controller.

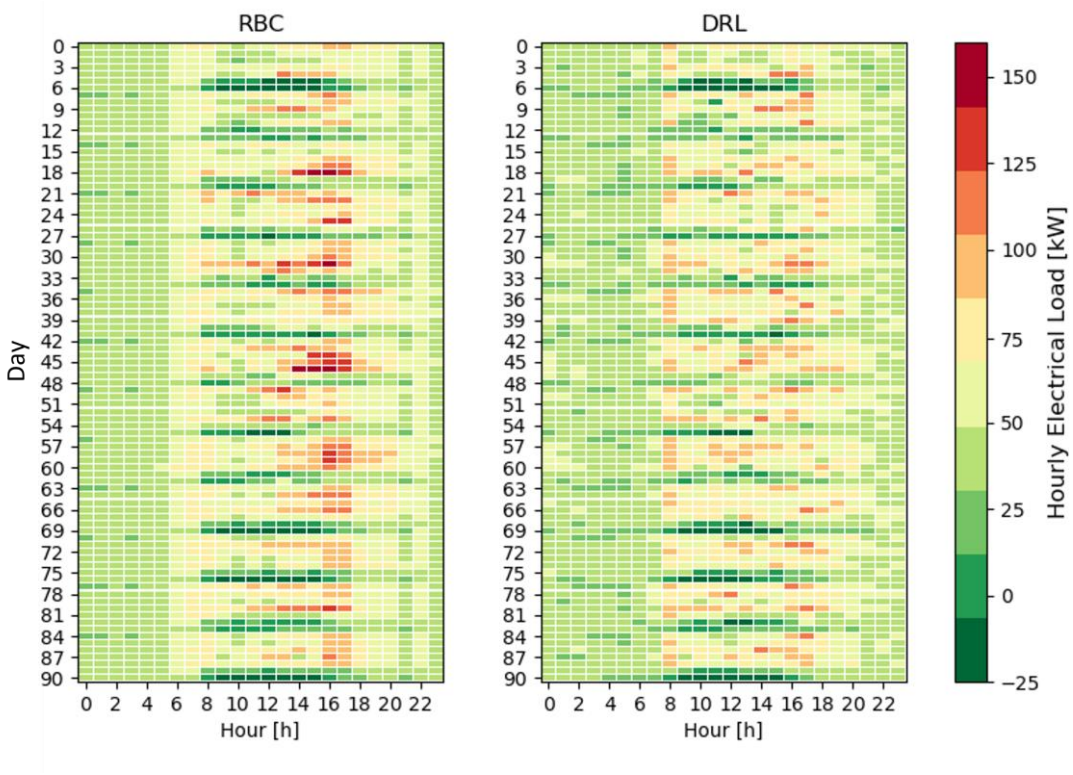

Figure 12: Carpet plot of RBC and coordinated energy consumption of the cluster of buildings

The DRL controller is able to flatten the cluster load profile in comparison to RBC due to the optimal management of the charge and discharge process of the storage installed in each building. On the

other hand, the carpet plot of the electrical load with RBC is characterized in average by higher electrical loads during the time period 14-18.

Furthermore, the charging process with the DRL control strategy is more uniform: storage units are charged in the earlier hours of the night to reduce morning load peaks, when the heat pumps are turned on. To understand how these results have been achieved, Figure 13 shows charge and discharge process (SOC) of the four chilled water storage installed in each building of the cluster. The agent adopts a control policy that spreads both charging and discharging over the day to prevent new undesirable peaks, while still exploiting the low electricity price during the night. The control policy exploits storage SOC information to optimise their operations, spreading the charge over the night period and reducing the peak loads. On the other hand, the discharge is optimised to increase energy efficiency during operation of the heat pumps.

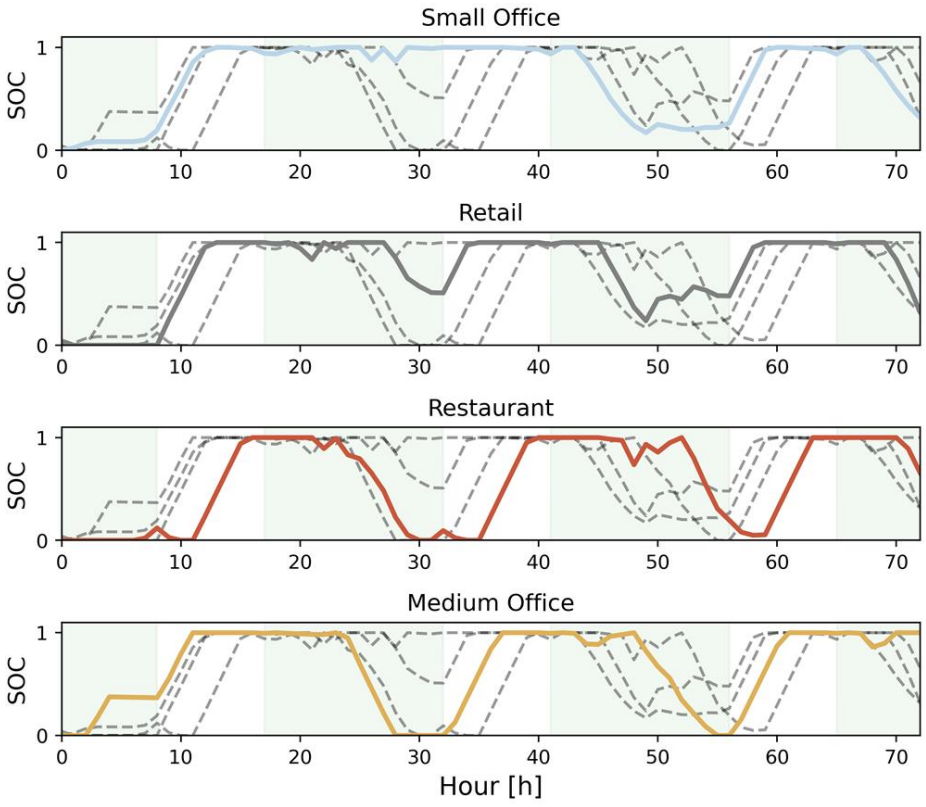

Figure 13: Chilled water storage control strategy

Figure 14 shows the distribution of the indoor temperature for the four controlled buildings during occupancy period. As can be seen in Figure 14 both office and restaurant buildings show very limited discomfort periods, while retail is characterized by a higher discomfort rate. In particular, retail has a large number of lower violations, influenced by the external temperature during the early morning hours, when it is open.

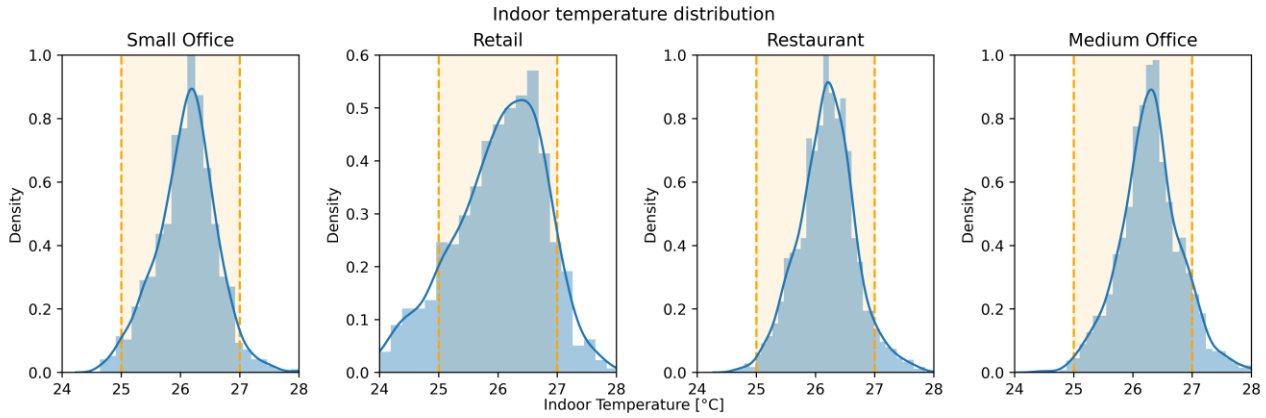

Figure 14: Indoor temperature distribution of the four buildings

Moreover, to fully characterize the effects of the DRL control policy on the indoor temperature control the cumulative values of degrees associated to comfort violations, the number of hours of discomfort and the average temperature difference between the indoor temperature and the upper and lower threshold are reported in Table 8.

Table 8: Metrics related to indoor temperature control

	Cumulative T<25 [°C]	**Hours of Discomfort T<25 °C**	**Average lower T discomfort [°C]**	**Cumulative T>27 [°C]**	**Hours of Discomfort T>27 °C**	**Average upper T discomfort [°C]**
Small Office	2.1	13	0.15	6.2	21	0.29
Retail	7.7	41	0.18	29.1	107	0.28
Restaurant	1.8	10	0.18	27.7	94	0.29
Medium Office	1.4	8	0.18	33.4	106	0.31

The table shows that, considering the 3 months of simulation, discomfort conditions are highly unusual, and that the distribution of violations reflects the reward function behaviour, which penalizes high temperature violations. In particular, the control policy lead to higher cumulative values of indoor temperature exceeding the upper threshold, where violations are less penalized, as a result of a trade-off between thermal comfort and energy consumption.

Figure 15 reports internal temperature evolution and storage operation for the small office for both control strategies, where the RBC uses an ideal load, considering a constant temperature at 26°C during occupancy periods. In detail, the Figure 15a) shows that, on average, the controller is able to maintain the indoor temperature close to the upper limit of the admitted range, leading to a reduction of energy consumption. Figure 15b) shows how the DRL agent tries to meet the cooling load either fully discharging the chilled water storage or running the heat pump ensuring its more efficient operation. Figure 15 c) focuses on the RBC strategy, whose control leads to the simultaneous operation of both heat pump and thermal storage to meet the cooling load. As a result, the heat pump often works at partial load operation with lower efficiency.

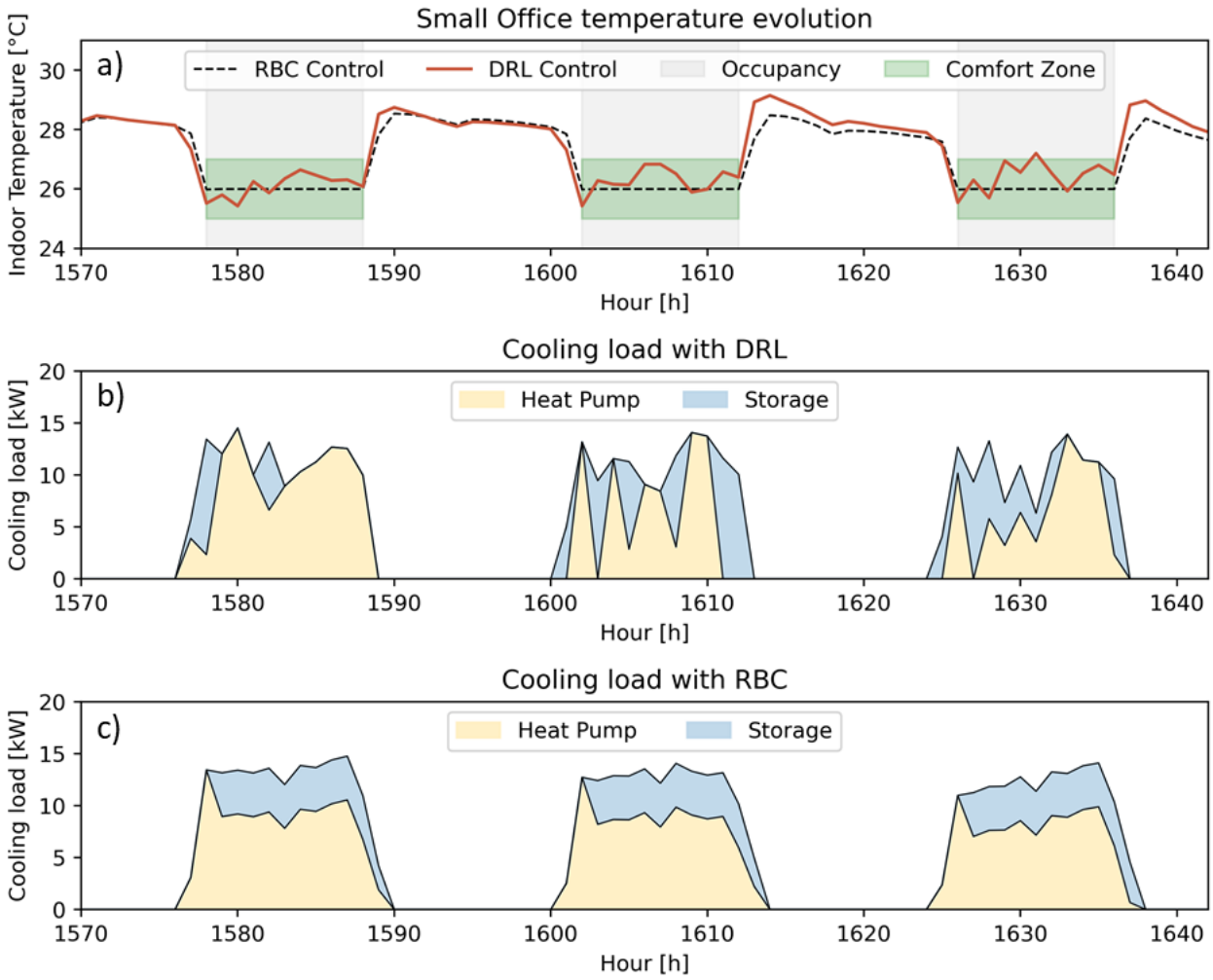

Figure 15: Temperature profile and cooling load

6.2.2. Analysis at grid level

The analysis was then shifted towards the benefits provided by the coordinated control strategy on the grid. In particular, Figure 16 shows the load duration curve for different control strategies considering as a benchmark the electrical load curve of the cluster resulting from no-storage installation., This benchmark makes it possible to highlight the impact on peak reduction of thermal storing in combination with control strategies. The values of the cluster load peaks for the different cases (i.e., no storage, RBC, DRL) related to the entire period of simulation are reported with horizontal dashed lines. In addition, in the bottom right of the figure can be noticed the increase of baseload as a result of storage installation, leading to a more uniform use of energy.

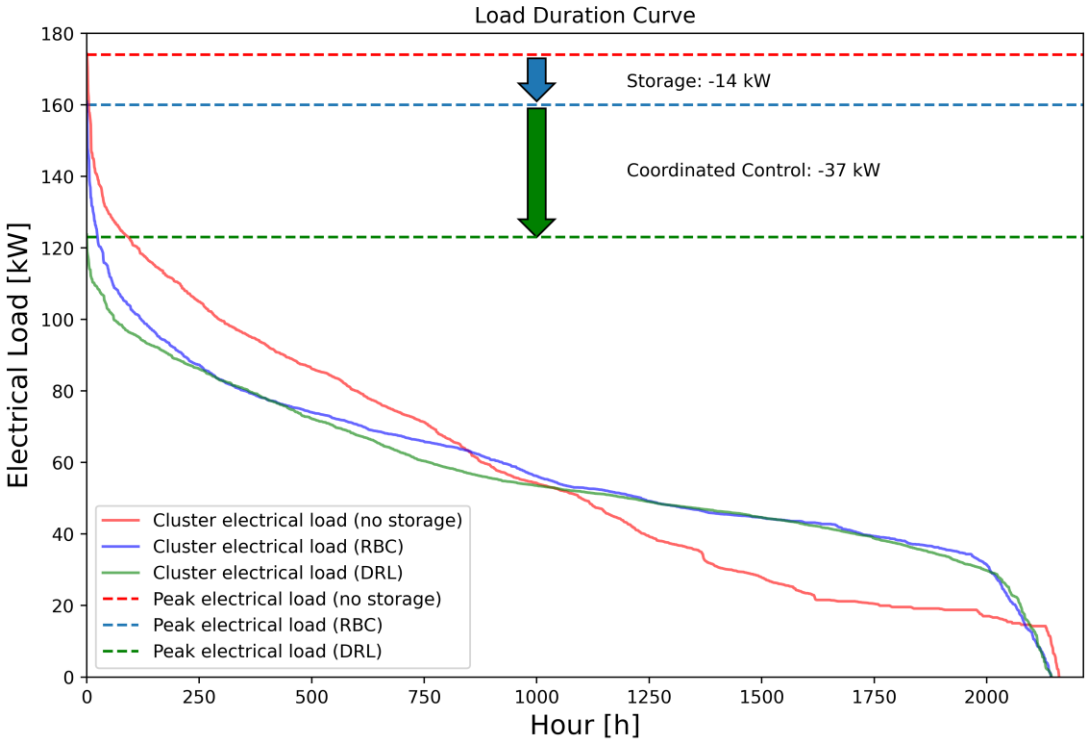

Figure 16: Load duration curve for the different control strategies

Table 9 summarizes the performance of the two control strategies with respect to the main KPIs selected. To allow an easier comparison, the values are normalised on those resulted from the implementation of the RBC strategy.

Table 9: Comparison between performances of the two control strategies

	Electricity Cost	Maximum Peak	Peak-to-average ratio (PAR)	Average daily peak	Average daily PAR	Flexibility Factor
Manually Optimised RBC	1	1	1	1	1	1
DRL	0.97	0.77	0.80	0.88	0.92	1.04

DRL controller exploits the possibility to modulate the heat pump cooling power to avoid peak load and takes advantage from storage charge and discharge process to increase heat pump efficiency, while slightly reduces electricity costs. As pointed out by Table 9 and Figure 16, the coordinated approach shows very good results at district level, reducing maximum peak by 23% and average daily peak by 12%. Moreover, the PAR and average daily PAR reduction of 20 and 8% respectively highlights the benefits of building coordination that can be translated in a more uniform baseload. Finally, the controller also shows the ability to better exploit energy flexibility of the multiple energy systems highlighted by a 4% increase of flexibility factor.

7. Discussion

The presented paper aims to exploit DNN and model-free DRL to enhance district energy management. LSTM models have been exploited to develop lightweight building models, to predict indoor environment evolution with a low computational effort. Once trained and tested, the DNNs building models have been integrated into CityLearn, an openAI gym environment used to train the DRL controller.

The centralised DRL controller was designed to coordinate electric load profile of the cluster of buildings, by regulating the heat pump supply cooling power and the operation of the thermal storage to optimise both economic costs and peak demand without jeopardizing indoor temperature control in each building. The main novelty is related to the introduction of DNN models coupled with DRL controller that enabled the opportunity of controlling indoor temperature through the modulation of heat pump operation, adding a flexibility sources to the control problem.

The optimal control policy of the agent is obtained through a *trial-and-error* interaction with the environment; in particular LSTM models receives as input the supply cooling power and return the corresponding indoor temperature in order to optimise heat pump operation, while electricity price information is used to optimise storage operation.

To analyse the effectiveness of the proposed approach, a manually optimised RBC controller was introduced. The proposed RBC ensures an internal temperature of 26°C during occupied periods, while taking advantage of the low-price tariff to charge the storage. On the other hand, DRL was designed to maintain indoor comfort conditions, exploiting the comfort band to minimize energy consumption and thermal mass during start-up and shut-down periods. Moreover, the agent found a better control strategy for the thermal storage, consuming energy more efficiently and flattening the electric load profile.

The paper analyses the role of the state-space and the reward function on the optimal control strategy. The reward function was designed with the aim of searching a trade-off between indoor temperature control, energy costs and grid requirements. Moreover, forecast information regarding weather conditions, occupancy information and electricity price resulted to be effective to learn the optimal control policy, highlighting the crucial role of the state-space design in the DRL problem.

As a result, DRL outperformed RBC, proving to be simultaneously able to find a compromise between indoor temperature control and energy consumption, with the additional capability to coordinate the operation of multiple buildings to reduce peak demand and flatten the load profile.

Lastly, the strength of the proposed approach lies in the lightness of the data-driven methodology, which helps the scalability of district energy management. In order to assess the computational cost advantages, a comparison between the proposed fully data-driven approach with a forward simulation environment coupling EnergyPlus and the DRL agent through BCVTB [22] was performed. The simulations were run on a single building using a workstation with i9-10900X CPU @ 3.7 GHz and 128 GB RAM. The training period of the DRL agent for 30 episodes using the proposed approach took 1920 seconds, while the forward simulation run for 2300 seconds. During the deployment of the trained DRL agent, episode was run within 60 seconds by the proposed approach and 87 seconds with the alternative forward approach. In summary a computational advantage of 20% during training and around 50% during deployment was found. Moreover, it should be highlighted that as the number of buildings increases, the simulation environment coupling Energyplus with DRL through BCVTB

needs to collect and share multiple idf files while the proposed fully data-driven approach shares data more efficiently exploiting the same environment for the entire district.

The analysis highlights how building-related data could be exploited to develop data-driven models used to coordinate a district of buildings. Moreover, the adaptive nature of DRL is very effective in large evolving environments, such as districts, where consumption patterns can be modified by retrofitting operations, PV installation, EV charging or demand response programs.

8. Conclusion & future perspectives

The present work proved the feasibility and advantages of a data-driven and adaptive control scheme for district energy management. In the first part, the study focused on the development of LSTM models, one for each building, to describe thermal dynamics. DNNs were used to perform multi-step temperature forecasts with a recursive strategy. Successively, the models were integrated into CityLearn environment, where a centralised DRL controller was designed. The main contribution of the study regards the introduction of indoor temperature regulation into the control problem, which is the result of the interaction between the heat pump supply cooling power and the LSTM models.

The developed DRL controller was able to maintain indoor comfort conditions for each building, while reducing costs of around 3%. In addition, the DRL controller allowed to reduce the peak by 23% and PAR by 20%. Lastly, the DRL controller was able to exploit the interaction between different flexibility sources, increasing flexibility factor by 4%.

In conclusion, the work has shown that a data-driven coordinated energy management is effective at district scale, being able to find an optimal trade-off between indoor temperature control, energy consumption and district electric load shape.

Future works will be focused on:

- The implementation and comparison between the proposed centralised controller with different management architectures, such as decentralised DRL approach, in which the controllers can cooperate or compete, or hierarchical multi-agent architecture, in which a

high-level agent controls low-level agent. The analysis will introduce a comparison among the different architectures, highlighting pros and cons of the approaches in the district energy management.

- The implementation of dynamic electricity price tariffs and demand response programs, to study building-to-grid interaction. The use of a dynamic electricity price tariff, together with the exploitation of indoor comfort conditions, can pave the way towards tailored demand response programs, in which each building can find the optimal compromise among costs and comfort.
- The study of the effectiveness of transfer learning for the indoor environment representation, easing the extension of the proposed control architecture in different buildings and allowing the scalability of the methodology. Moreover, the analysis will study the feasibility of transferring the DRL control policy, to ease real-world implementation.

Nomenclature

Symbols

A = Action space

a = Action

c_{el} = Electricity price

$\mathcal{D}$ = Replay Buffer

e = Energy consumption

G = Return

K_1 = Reward hot storage weight

K_2 = Reward cold storage weight

K_p = Reward peak weight

m = Reward temperature weight

P = Transition Probabilities

q = Action-value

r = Reward

S = State space

SP = Set Point

s = State

T = Temperature

th = Power thresholds

α = Temperature parameter

γ = Discount factor

θ = Soft-Q network parameters

λ = Learning rate

ϕ = Policy network parameters

τ = Decay rate

H = Shannon Entropy of the policy

π = Policy

π^* = Optimal Policy

Abbreviations

ANN = Artificial Neural Network

COP = Coefficient of Performance

DHW = Domestic Hot Water

DNN = Deep Neural Network

DR = Demand Response

DRL = Deep Reinforcement Learning

DSM = Demand Side Management

HVAC = Heating, Ventilation and Air Conditioning

KPI = Key Performance Indicator

LSTM = Long-short Term Memory

MAPE = Mean Absolute Percentage Error

MDP = Markov Decision Process

MPC = Model Predictive Control

PAR = Peak-to-average ratio

PV = Photovoltaic

RBC = Rule Base Control

RES = Renewable Energy Sources

RL = Reinforcement Learning

RMSE = Root Mean Square Error

SAC = Soft Actor-Critic

SOC = State-of-Charge

References

[1] IEA. World Energy Outlook 2019. World Energy Outlook 2019 2019:1.

[2] Lund PD, Lindgren J, Mikkola J, Salpakari J. Review of energy system flexibility measures to enable high levels of variable renewable electricity. Renew Sustain Energy Rev 2015;45:785–807. https://doi.org/10.1016/j.rser.2015.01.057.

[3] Auer H, Haas R. On integrating large shares of variable renewables into the electricity system. Energy 2016;115:1592–601. https://doi.org/10.1016/j.energy.2016.05.067.

[4] Jabir HJ, Teh J, Ishak D, Abunima H. Impacts of demand-side management on electrical power systems: A review. Energies 2018;11:1–19. https://doi.org/10.3390/en11051050.

[5] Haider HT, See OH, Elmenreich W. A review of residential demand response of smart grid. Renew Sustain Energy Rev 2016;59:166–78. https://doi.org/10.1016/j.rser.2016.01.016.

[6] Hui H, Ding Y, Liu W, Lin Y, Song Y. Operating reserve evaluation of aggregated air conditioners. Appl Energy 2017;196:218–28. https://doi.org/10.1016/j.apenergy.2016.12.004.

[7] Jensen SØ, Marszal-Pomianowska A, Lollini R, Pasut W, Knotzer A, Engelmann P, et al. IEA EBC Annex 67 Energy Flexible Buildings. Energy Build 2017;155:25–34. https://doi.org/https://doi.org/10.1016/j.enbuild.2017.08.044.

[8] Celik B, Roche R, Suryanarayanan S, Bouquain D, Miraoui A. Electric energy management in residential areas through coordination of multiple smart homes. Renew Sustain Energy Rev 2017;80:260–75. https://doi.org/10.1016/j.rser.2017.05.118.

[9] Fiorini L, Aiello M. Energy management for user's thermal and power needs: A survey. Energy Reports 2019;5:1048–76. https://doi.org/10.1016/j.egyr.2019.08.003.

[10] Guerrero J, Gebbran D, Mhanna S, Chapman AC, Verbič G. Towards a transactive energy system for integration of distributed energy resources: Home energy management, distributed optimal power flow, and peer-to-peer energy trading. Renew Sustain Energy Rev 2020;132. https://doi.org/10.1016/j.rser.2020.110000.

[11] Wang S, Xue X, Yan C. Building power demand response methods toward smart grid. HVAC R Res 2014;20:665–87. https://doi.org/10.1080/10789669.2014.929887.

[12] Deltetto D, Coraci D, Pinto G, Piscitelli MS, Capozzoli A. Exploring the potentialities of deep reinforcement learning for incentive-based demand response in a cluster of small commercial buildings. Energies 2021;14. https://doi.org/10.3390/en14102933.

[13] Verschae R, Kawashima H, Kato T, Matsuyama T. Coordinated energy management for inter-community imbalance minimization. Renew Energy 2016;87:922–35. https://doi.org/10.1016/j.renene.2015.07.039.

[14] Chang TH, Alizadeh M, Scaglione A. Real-time power balancing via decentralized coordinated home energy scheduling. IEEE Trans Smart Grid 2013;4:1490–504. https://doi.org/10.1109/TSG.2013.2250532.

[15] Mocanu E, Mocanu DC, Nguyen PH, Liotta A, Webber ME, Gibescu M, et al. On-Line

Building Energy Optimization Using Deep Reinforcement Learning. IEEE Trans Smart Grid 2019;10:3698–708. https://doi.org/10.1109/TSG.2018.2834219.

[16] Wang X, Liu Y, Zhao J, Liu C, Liu J, Yan J. Surrogate model enabled deep reinforcement learning for hybrid energy community operation. Appl Energy 2021;289. https://doi.org/10.1016/j.apenergy.2021.116722.

[17] Lu R, Hong SH. Incentive-based demand response for smart grid with reinforcement learning and deep neural network. Appl Energy 2019;236:937–49. https://doi.org/10.1016/j.apenergy.2018.12.061.

[18] 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.

[19] Serale G, Fiorentini M, Capozzoli A, Bernardini D, Bemporad A. Model Predictive Control (MPC) for enhancing building and HVAC system energy efficiency: Problem formulation, applications and opportunities. Energies 2018;11. https://doi.org/10.3390/en11030631.

[20] Afram A, Janabi-Sharifi F. Theory and applications of HVAC control systems – A review of model predictive control (MPC). Build Environ 2014;72:343–55. https://doi.org/10.1016/J.BUILDENV.2013.11.016.

[21] Schreiber T, Eschweiler S, Baranski M, Müller D. Application of two promising Reinforcement Learning algorithms for load shifting in a cooling supply system. Energy Build 2020;229:110490. https://doi.org/10.1016/j.enbuild.2020.110490.

[22] Brandi S, Piscitelli MS, Martellacci M, Capozzoli A. Deep reinforcement learning to optimise indoor temperature control and heating energy consumption in buildings. Energy Build 2020;224:110225. https://doi.org/10.1016/j.enbuild.2020.110225.

[23] Henze GP. Predictive Optimal Control of Active and Passive Building Thermal Storage Inventory. Univ Nebraska - Lincoln Archit Eng -- Fac Publ 2003;110 PART 1.

[24] Pinto G, Piscitelli MS, Vázquez-Canteli JR, Nagy Z, Capozzoli A. Coordinated energy management for a cluster of buildings through deep reinforcement learning. Energy

2021;229:120725. https://doi.org/10.1016/j.energy.2021.120725.

[25] O'Shaughnessy E, Cutler D, Ardani K, Margolis R. Solar plus: Optimization of distributed solar PV through battery storage and dispatchable load in residential buildings. Appl Energy 2018;213:11–21. https://doi.org/10.1016/j.apenergy.2017.12.118.

[26] Yang L, Nagy Z, Goffin P, Schlueter A. Reinforcement learning for optimal control of low exergy buildings. Appl Energy 2015;156:577–86. https://doi.org/10.1016/j.apenergy.2015.07.050.

[27] Vazquez-Canteli JR, Henze G, Nagy Z. MARLISA : Multi-Agent Reinforcement Learning with Iterative Sequential Action Selection for Load Shaping of Grid-Interactive Connected Buildings. In: ISBN, editor. BuildSys '20, Yokohama, Japan: Association for Computing Machinery; 2020. https://doi.org/10.1145/3408308.3427604.

[28] Huang P, Fan C, Zhang X, Wang J. A hierarchical coordinated demand response control for buildings with improved performances at building group. Appl Energy 2019;242:684–94. https://doi.org/10.1016/j.apenergy.2019.03.148.

[29] Ondeck AD, Edgar TF, Baldea M. Impact of rooftop photovoltaics and centralized energy storage on the design and operation of a residential CHP system. Appl Energy 2018;222:280–99. https://doi.org/10.1016/j.apenergy.2018.03.131.

[30] Amin U, Hossain MJ, Fernandez E. Optimal price based control of HVAC systems in multizone office buildings for demand response. J Clean Prod 2020;270:122059. https://doi.org/10.1016/j.jclepro.2020.122059.

[31] Tang R, Wang S. Model predictive control for thermal energy storage and thermal comfort optimization of building demand response in smart grids. Appl Energy 2019;242:873–82. https://doi.org/10.1016/j.apenergy.2019.03.038.

[32] Robillart M, Schalbart P, Chaplais F, Peuportier B. Model reduction and model predictive control of energy-efficient buildings for electrical heating load shifting. J Process Control 2019;74:23–34. https://doi.org/10.1016/j.jprocont.2018.03.007.

[33] Wang A, Li R, You S. Development of a data driven approach to explore the energy flexibility potential of building clusters. Appl Energy 2018;232:89–100. https://doi.org/10.1016/j.apenergy.2018.09.187.

[34] Perfumo C, Kofman E, Braslavsky JH, Ward JK. Load management: Model-based control of aggregate power for populations of thermostatically controlled loads. Energy Convers Manag 2012;55:36–48. https://doi.org/10.1016/j.enconman.2011.10.019.

[35] Gonzato S, Chimento J, O'Dwyer E, Bustos-Turu G, Acha S, Shah N. Hierarchical price coordination of heat pumps in a building network controlled using model predictive control. Energy Build 2019;202:109421. https://doi.org/10.1016/j.enbuild.2019.109421.

[36] Ruano AE, Crispim EM, Conceição EZE, Lúcio MMJR. Prediction of building's temperature using neural networks models. Energy Build 2006;38:682–94. https://doi.org/10.1016/j.enbuild.2005.09.007.

[37] Mustafaraj G, Lowry G, Chen J. Prediction of room temperature and relative humidity by autoregressive linear and nonlinear neural network models for an open office. Energy Build 2011;43:1452–60. https://doi.org/10.1016/j.enbuild.2011.02.007.

[38] Afroz Z, Urmee T, Shafiullah GM, Higgins G. Real-time prediction model for indoor temperature in a commercial building. Appl Energy 2018;231:29–53. https://doi.org/10.1016/j.apenergy.2018.09.052.

[39] Mtibaa F, Nguyen KK, Azam M, Papachristou A, Venne JS, Cheriet M. LSTM-based indoor air temperature prediction framework for HVAC systems in smart buildings. Neural Comput Appl 2020;32:17569–85. https://doi.org/10.1007/s00521-020-04926-3.

[40] Huang H, Chen L, Hu E. A neural network-based multi-zone modelling approach for predictive control system design in commercial buildings. Energy Build 2015;97:86–97. https://doi.org/10.1016/j.enbuild.2015.03.045.

[41] Marvuglia A, Messineo A, Nicolosi G. Coupling a neural network temperature predictor and a fuzzy logic controller to perform thermal comfort regulation in an office building. Build

Environ 2014;72:287–99. https://doi.org/10.1016/j.buildenv.2013.10.020.

[42] Ellis MJ, Chinde V. An encoder–decoder LSTM-based EMPC framework applied to a building HVAC system. Chem Eng Res Des 2020;160:508–20. https://doi.org/10.1016/j.cherd.2020.06.008.

[43] Mason K, Grijalva S. A review of reinforcement learning for autonomous building energy management. Comput Electr Eng 2019;78:300–12. https://doi.org/10.1016/j.compeleceng.2019.07.019.

[44] Coraci D, Brandi S, Piscitelli MS, Capozzoli A. Online Implementation of a Soft Actor-Critic Agent to Enhance Indoor Temperature Control and Energy Efficiency in Buildings. Energies 2021;14:997. https://doi.org/10.3390/en14040997.

[45] Ruelens F, Claessens BJ, Quaiyum S, Schutter B De, Babuška R, Belmans R. Reinforcement Learning Applied to an Electric Water Heater : From Theory to Practice 2018;9:3792–800. https://doi.org/10.1109/TSG.2016.2640184.

[46] Kazmi H, D'Oca S, Delmastro C, Lodeweyckx S, Corgnati SP. Generalizable occupant-driven optimization model for domestic hot water production in NZEB. Appl Energy 2016;175:1–15. https://doi.org/10.1016/j.apenergy.2016.04.108.

[47] Vázquez-Canteli J, Kämpf J, Nagy Z. Balancing comfort and energy consumption of a heat pump using batch reinforcement learning with fitted Q-iteration. Energy Procedia 2017;122:415–20. https://doi.org/10.1016/j.egypro.2017.07.429.

[48] 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.

[49] Kofinas P, Dounis AI, Vouros GA. Fuzzy Q-Learning for multi-agent decentralized energy management in microgrids. Appl Energy 2018;219:53–67. https://doi.org/10.1016/j.apenergy.2018.03.017.

[50] Vázquez-Canteli JR, Nagy Z, Dey S, Henze G. CityLearn: Standardizing Research in Multi-

Agent Reinforcement Learning for Demand Response and Urban Energy Management n.d.

[51] Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Comput 1997;9:1735–80. https://doi.org/10.1162/neco.1997.9.8.1735.

[52] Sutton RS, Barto AG. Reinforcement Learning: An Introduction. MIT Press Cambridge 1998. https://doi.org/10.1016/S0140-6736(51)92942-X.

[53] Watkins CJCH, Dayan P. Technical Note: Q-Learning. Mach Learn 1992;8:279–92. https://doi.org/10.1023/A:1022676722315.

[54] Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, et al. Playing Atari with Deep Reinforcement Learning 2013:1–9.

[55] Haarnoja T, Zhou A, Abbeel P, Levine S. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. 35th Int Conf Mach Learn ICML 2018 2018;5:2976–89.

[56] Haarnoja T, Zhou A, Hartikainen K, Tucker G, Ha S, Tan J, et al. Soft Actor-Critic Algorithms and Applications 2018.

[57] Vázquez-Canteli JR, Kämpf J, Henze G, Nagy Z. CityLearn v1.0: An OpenAI gym environment for demand response with deep reinforcement learning. BuildSys 2019 - Proc 6th ACM Int Conf Syst Energy-Efficient Build Cities, Transp 2019:356–7. https://doi.org/10.1145/3360322.3360998.

[58] Vázquez-Canteli JR, Kämpf J, Henze GP NZ. CityLearn Github repository 2019 n.d. ttps://github.com/intelligent-environments-lab/CityLearn.git.

[59] UNI EN 14825:2019 “Condizionatori d’aria, refrigeratori di liquido e pompe di calore, con compressore elettrico, per il riscaldamento e il raffrescamento degli ambienti - Metodi di prova e valutazione a carico parziale e calcolo del rendimento stagionale.” Italy: 2019.

[60] Clauß J, Finck C, Vogler-finck P, Beagon P. Control strategies for building energy systems to unlock demand side flexibility – A review Norwegian University of Science and Technology , Trondheim , Norway Eindhoven University of Technology , Eindhoven , Netherlands

Neogrid Technologies ApS / Aalborg. 15th Int Conf Int Build Perform 2017:611–20.

[61] Henze GP, Schoenmann J. Evaluation of reinforcement learning control for thermal energy storage systems. HVAC R Res 2003;9:259–75. https://doi.org/10.1080/10789669.2003.10391069.