



Politecnico
di Torino

ScuDo

Scuola di Dottorato - Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation

Doctoral Program in Energetics (35th cycle)

Scaling energy management in buildings with artificial intelligence

By

Giuseppe Pinto

Supervisor(s):

Prof. A. Capozzoli, Supervisor

Doctoral Examination Committee:

Prof. Enrico Fabrizio, Politecnico di Torino

Prof. Cheng Fan, Shenzhen University

Prof. Rongling Li, Technical University of Denmark - DTU

Prof. Adolfo Palombo, Università degli studi di Napoli Federico II

Dr. Marco Pritoni, Lawrence Berkeley National Laboratory - LBNL

Politecnico di Torino

2023

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Giuseppe Pinto
2023

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

Abstract

The growing adoption of automation and control systems, and internet of things sensors in smart buildings has contributed to the unprecedented availability of monitoring data of the built environment, that could enable the deployment of Energy Management and Information Systems (EMIS) at scale. This dissertation aims at analyzing the potentialities provided by the exploitation of Artificial Intelligence (AI) techniques to scale EMIS, identifying promising directions and potential barriers for its real-world application. In this context, Grid-Interactive Efficient Buildings (GEB) are ideal candidates for the application of advanced energy management strategies. GEBs are energy-efficient buildings that uses smart technologies to provide demand flexibility while co-optimizing for energy cost, grid services, and occupant needs. However, when energy management is faced with shifting from a single building to multiple buildings, uncoordinated strategies for exploiting energy flexibility may have negative effects on the grid reliability, causing undesirable new peaks. The recent development of AI supported the creation of advanced data-driven control strategies, such as Deep Reinforcement Learning (DRL), however implementation focused on single buildings, neglecting the potentialities of applying this control strategy in multiple buildings. In this dissertation, three different applications that leveraged DRL at scale are conceived and tested. DRL is a control method based on the paradigm of learning from interaction, encoding the environment using deep neural networks. The developed applications used CityLearn, a simulation environment for the implementation of DRL in multiple buildings, focusing on 4 buildings equipped with thermal energy storage and renewables, benchmarking the DRL controller against a rule-based controller. In the first application, a centralized DRL controller was implemented to optimize the electrical demand profiles, reducing costs and peaks, understanding the effects of advanced control strategies at different scales (single building, district, grid). This application showed the potential of applying DRL in multiple buildings, achieving a 4% cost and 12% peak reduction. Then, a

second application analyzed the role of different reinforcement learning architectures, comparing a centralised (coordinated) controller and a decentralised (cooperative) controller to also consider different renewable energy systems. The two controllers reduced the costs by 3% and 7% respectively, and 10% and 14% respectively for peak demand. The study showed that the multi-agent cooperative approach may be more suitable for districts with heterogeneous objectives within the individual buildings. In a third application, the role of HVAC flexibility was investigated, exploiting deep neural networks to simulate the building thermal dynamics of the buildings, adding to the previously introduced framework the possibility to control the HVAC. In this case, the DRL controller was conceived to optimise the electrical demand profiles and provide services to the grid without penalising indoor comfort conditions. The developed DRL controller reduced the overall district electricity costs, while decreasing the peak energy demand by 23% and the Peak to Average Ratio by 20%, without penalizing indoor temperature control. The third application showed how deep neural networks are effective as a lightweight data-driven model to predict building thermal responses, highlighting their reliance on a large amount of data, that clashes with the potential limited data availability in most existing buildings. Therefore, the last application focused on data-driven models, identifying in transfer learning a way to overcome data reliance, describing its role in supporting building energy management. The application conducted a suite of experiments that leveraged 250 data-driven models based on a synthetic dataset of a building to study the influence of several features, isolating their contribution. The performance of the transfer learning process was compared against a classical machine learning approach, identifying guidelines for its application in buildings. Lastly, findings and outcomes of the present research study were discussed, providing a robust reasoning on the application of DRL controllers at large scale and how data-driven models can boost their adoption. Eventually, a wide overview on the lessons learned is proposed, outlining the future opportunities and barriers of scaling energy management in buildings using artificial intelligence.

Contents

| | |
|--|-------------|
| List of Figures | xi |
| List of Tables | xv |
| Nomenclature | xvii |
| 1 Introduction | 1 |
| 1.1 Motivations of the research | 4 |
| 1.2 Research outline | 6 |
| 1.3 Research questions | 10 |
| 1.4 Objective of the thesis and novelty | 12 |
| 1.5 Organization of the thesis | 13 |
| 2 Theoretical background on data-driven controllers | 15 |
| 2.1 Literature review | 16 |
| 2.1.1 Scale of analysis | 17 |
| 2.1.2 Control strategies and architectures | 21 |
| 2.1.2.1 Control strategies | 21 |
| 2.1.2.2 Architectures | 22 |
| 2.1.2.3 Optimal control | 27 |
| 2.1.3 Energy systems and objective functions | 29 |

| | | |
|----------|--|-----------|
| 2.1.4 | Level of detail | 35 |
| 2.1.5 | Discussion of the literature review | 37 |
| 2.2 | Reinforcement learning | 40 |
| 2.2.1 | Multi-agent reinforcement learning | 44 |
| 2.2.1.1 | Markov game | 45 |
| 2.2.1.2 | Decentralized partially observable markov decision process | 45 |
| 2.2.1.3 | Partially observable markov game | 46 |
| 2.2.1.4 | Challenges | 47 |
| 2.2.1.5 | Approaches | 48 |
| 2.2.2 | From reinforcement learning to deep reinforcement learning | 49 |
| 2.2.3 | Soft-actor critic | 51 |
| 3 | Scale-up energy management in buildings with data-driven controllers | 56 |
| 3.1 | CityLearn environment | 58 |
| 3.1.1 | Building | 61 |
| 3.1.2 | Heat pump | 62 |
| 3.1.3 | Electric heater | 62 |
| 3.1.4 | Thermal storage | 63 |
| 3.1.5 | Solar photovoltaic panels | 65 |
| 3.1.6 | Key performance indicators | 65 |
| 3.2 | Enhancing energy management in grid-interactive buildings with deep reinforcement learning | 66 |
| 3.2.1 | Motivations and novelty of the proposed approach | 67 |
| 3.2.2 | Methodological framework | 68 |
| 3.2.3 | Case study | 69 |
| 3.2.3.1 | Description of the cluster of buildings | 70 |

| | | |
|---------|---|-----|
| 3.2.3.2 | Energy systems and control objectives | 71 |
| 3.2.3.3 | Baseline rule-based control | 71 |
| 3.2.3.4 | Design of the deep reinforcement learning controller | 72 |
| 3.2.3.5 | Training and deployment | 75 |
| 3.2.4 | Results | 77 |
| 3.2.4.1 | Training results | 77 |
| 3.2.4.2 | Deployment of deep reinforcement learning controller in different climatic conditions | 82 |
| 3.2.5 | Discussion | 82 |
| 3.3 | A comparison among coordinated and cooperative deep reinforcement learning architectures in buildings | 84 |
| 3.3.1 | Motivations and novelty of the proposed approach | 85 |
| 3.3.2 | Methodology | 87 |
| 3.3.3 | Case study district & control problem | 88 |
| 3.3.3.1 | District | 88 |
| 3.3.3.2 | Energy systems at building level | 89 |
| 3.3.3.3 | Definition of the control problem | 92 |
| 3.3.3.4 | Key performance indicator design | 93 |
| 3.3.4 | Design of multi-agent reinforcement learning control strategies | 93 |
| 3.3.4.1 | Design of action-space | 94 |
| 3.3.4.2 | Design of state-space | 95 |
| 3.3.4.3 | Design of reward functions | 97 |
| 3.3.5 | Results | 99 |
| 3.3.5.1 | Comparison with baseline RBC | 100 |
| 3.3.5.2 | Deployment of RL controllers for different climates | 106 |
| 3.3.6 | Discussion | 107 |
| 3.3.6.1 | Limitations | 109 |

| | | |
|----------|--|------------|
| 4 | 3DEM: A methodology to combine data-driven models and controllers | 111 |
| 4.1 | Motivations and novelty of the proposed approach | 112 |
| 4.1.1 | Building load prediction models | 112 |
| 4.1.2 | Building thermal dynamic models | 113 |
| 4.2 | Case study and control problem | 116 |
| 4.3 | Methodology | 117 |
| 4.3.1 | Development of artificial neural networks | 118 |
| 4.3.2 | Deployment strategy of the neural network | 121 |
| 4.3.3 | Training of the centralised DRL | 122 |
| 4.3.4 | Deployment of the centralised DRL | 122 |
| 4.4 | Implementation | 123 |
| 4.4.1 | Baseline control | 123 |
| 4.4.2 | Design of the DRL controller | 124 |
| 4.4.2.1 | Action-space design | 124 |
| 4.4.2.2 | State-space design | 124 |
| 4.4.2.3 | Reward function | 126 |
| 4.4.2.4 | Hyperparameters setting of deep reinforcement learning | 128 |
| 4.5 | Results | 129 |
| 4.5.1 | Artificial neural network testing results | 129 |
| 4.5.2 | Deployment of the deep reinforcement learning controller | 130 |
| 4.5.2.1 | Comparison at district level | 130 |
| 4.5.2.2 | Analysis at grid level | 133 |
| 4.6 | Discussion | 135 |
| 5 | Scale-out energy management in buildings with data-driven models | 138 |
| 5.1 | Theoretical background on transfer learning | 139 |

| | | |
|----------|---|------------|
| 5.1.1 | Transfer learning | 139 |
| 5.1.2 | Literature review on transfer learning applications | 143 |
| 5.1.2.1 | Energy systems control | 143 |
| 5.1.2.2 | Building thermal dynamic models | 145 |
| 5.2 | Motivations and novelty of the proposed approach | 146 |
| 5.3 | Case study | 149 |
| 5.4 | Methodology | 151 |
| 5.4.1 | Source building selection | 152 |
| 5.4.2 | Machine learning model optimization | 152 |
| 5.4.3 | Design of ML and TL configurations | 154 |
| 5.4.4 | Design of experiments | 154 |
| 5.4.5 | Assessment of TL performance | 155 |
| 5.4.6 | Comparison in an online fashion | 156 |
| 5.5 | Results | 156 |
| 5.5.1 | Machine learning and transfer learning performance | 157 |
| 5.5.2 | Negative transfer learning | 162 |
| 5.5.3 | Jumpstart performance | 164 |
| 5.5.4 | Online deployment | 164 |
| 5.6 | Discussion | 166 |
| 6 | Conclusions | 168 |
| | References | 175 |
| | Appendix A | 204 |
| A.1 | CityLearn Documentation | 204 |
| A.1.1 | Input Attributes | 204 |
| A.1.2 | Internal Attributes | 205 |

| | | |
|-------|---|-----|
| A.1.3 | CityLearn Methods | 206 |
| A.1.4 | Methods inherited from OpenAI Gym | 206 |
| A.1.5 | States | 206 |
| A.1.6 | Actions | 208 |
| A.1.7 | Rewards | 208 |
| A.1.8 | Evaluation metrics | 209 |
| A.2 | Deep reinforcement learning hyperparameters | 210 |
| A.3 | List of articles | 212 |

List of Figures

| | | |
|------|---|----|
| 1.1 | Model-based and model-free flowchart comparison | 5 |
| 1.2 | Thesis contributions according to model and control scale | 7 |
| 1.3 | Description of applications involving data-driven models and controllers | 10 |
| 1.4 | Conceptual organisation of the thesis | 13 |
| 2.1 | Organization of the literature review | 17 |
| 2.2 | Demand response programs classification | 18 |
| 2.3 | Representation of different sizes of building clusters | 19 |
| 2.4 | Representation of different architectures for multiple building energy management | 23 |
| 2.5 | Non-exhaustive taxonomy of optimization techniques for building energy management | 27 |
| 2.6 | Summary of the three modeling paradigms features | 37 |
| 2.7 | Venn diagram displaying the four pillars of advanced control for district energy management: coordination of multiple buildings, grid-interaction, indoor comfort and management of supply technologies | 40 |
| 2.8 | Schematic representation of the RL framework | 42 |
| 2.9 | Non-exhaustive taxonomy of RL | 44 |
| 2.10 | Multi-agent RL problem classification | 45 |
| 2.11 | Multi-agent RL problem representation | 46 |

| | | |
|------|--|-----|
| 2.12 | Multi-agent RL control architectures | 49 |
| 2.13 | Soft Actor-Critic architecture | 54 |
| 3.1 | Contribution of the dissertation on data-driven controllers for the energy management of multiple buildings | 58 |
| 3.2 | Flowchart of the CityLearn environment | 59 |
| 3.3 | CityLearn code architecture | 60 |
| 3.4 | Energy systems in the CityLearn environment with corresponding energy flows | 63 |
| 3.5 | Framework of the application of DRL control | 68 |
| 3.6 | Load profile for each building (left) and cluster profile electricity and PV production (right) | 70 |
| 3.7 | State-action space representation of the DRL controller | 74 |
| 3.8 | Temperature distribution of the different deployment climates | 76 |
| 3.9 | State of charge of storage and forcing variables scaled between 0 and 1 | 78 |
| 3.10 | Comparison between uncoordinated and coordinated energy management | 79 |
| 3.11 | State of charge of storage averaged over a day | 80 |
| 3.12 | Load duration curve for the base case without energy storage in buildings and the two control strategies | 81 |
| 3.13 | KPI comparison for the four-deployment case | 83 |
| 3.14 | Methodological framework overview | 87 |
| 3.15 | Building energy management control scheme | 90 |
| 3.16 | Electrical load profile for each building in the district for Climate 2A | 91 |
| 3.17 | Coordinated and cooperative control architectures | 95 |
| 3.18 | Cost related to the energy term for each building (left) and total district cost, sum of energy and peak terms (right), for the different control strategies over the entire simulation period | 100 |

| | | |
|------|---|-----|
| 3.19 | District electrical load profile for each control strategy during a three-days period | 101 |
| 3.20 | Comparison of control strategies for Building 1 | 102 |
| 3.21 | Daily average hourly scale profiles of SOC with relative standard deviations for the three control strategies in Building 1 | 103 |
| 3.22 | Comparison of district cumulative exported electricity between control strategies over the entire simulation period (3 months) | 104 |
| 3.23 | District energy disaggregation comparison over the entire simulation period (3 months) | 105 |
| | | |
| 4.1 | Schematic of the district energy management and controlled energy systems | 117 |
| 4.2 | Electrical load profile for each building (up) and electrical load profile and PV production for the cluster of buildings (down) | 118 |
| 4.3 | Proposed framework for the district energy management | 119 |
| 4.4 | Proposed framework for the district energy management | 120 |
| 4.5 | Proposed framework for the district energy management | 122 |
| 4.6 | State and action spaces of the DRL control strategy | 126 |
| 4.7 | Comfort term of the reward function | 127 |
| 4.8 | 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) | 130 |
| 4.9 | Carpet plot of electrical load at cluster of buildings level with RBC and DRL strategy | 131 |
| 4.10 | State of charge profile of thermal storage for each building of the cluster | 132 |
| 4.11 | Indoor temperature distribution for each building of the cluster | 133 |
| 4.12 | Profiles of indoor temperature and cooling load for the small office building | 134 |
| 4.13 | Load duration curve for the different control strategies | 135 |

| | | |
|------|--|-----|
| 5.1 | Schematic representation of machine learning and transfer learning problem in buildings | 141 |
| 5.2 | Feature-extraction (top) and weight-initialization (bottom) transfer learning schematization | 143 |
| 5.3 | A schematic representation of medium office geometry and thermal zones for a single floo | 149 |
| 5.4 | Distribution of the outdoor air temperature for each month and climate considered during the analysis (left) and occupancy profile distribution (right) | 150 |
| 5.5 | Methodological framework | 151 |
| 5.6 | Input of the neural networks and sliding window approach | 153 |
| 5.7 | Transfer learning metrics used to quantify the performances of the new model | 155 |
| 5.8 | Performance of the different techniques over the control horizon | 158 |
| 5.9 | Performance of the different techniques over different zones | 159 |
| 5.10 | MAE distribution over different periods and techniques | 160 |
| 5.11 | Categorical plot of the error distribution for each technique over all the influencing factors | 160 |
| 5.12 | Performance comparison with isolated effects of features | 161 |
| 5.13 | Error distribution for each technique over different climate and data availability (top) Asymptotic performance for each technique over different climate and data availability (bottom) | 163 |
| 5.14 | Categorization of transfer learning effectiveness and negative transfer analysis | 164 |
| 5.15 | Prediction evolution for the first time-step with different techniques for effective and negative TL | 165 |
| 5.16 | Jumpstart comparison over different training time | 166 |
| 5.17 | Performance comparison between online ML and online TL | 166 |
| A.1 | Evolution of the reward function with episodes | 211 |

List of Tables

| | | |
|------|--|-----|
| 3.1 | KPIs used in CityLearn | 66 |
| 3.2 | Building and energy systems properties | 70 |
| 3.3 | State-space for the case study | 73 |
| 3.4 | Hyperparameter settings | 75 |
| 3.5 | Reward and KPI evolution over training period | 77 |
| 3.6 | Comparison between performances of the two control strategies | 81 |
| 3.7 | Climate zones (per ASHRAE definitions) considered in this study | 89 |
| 3.8 | Summary of building geometrical features and energy systems in district | 90 |
| 3.9 | Electricity tariff including energy terms and peak terms | 92 |
| 3.10 | KPIs Used in MARL controller comparisons | 94 |
| 3.11 | State-space description for coordinated and cooperative DRL agents | 96 |
| 3.12 | Reward function hyperparameter values | 98 |
| 3.13 | Results of the MARL controllers deployed on Climate 2A (performance improvement in brackets) | 106 |
| 3.14 | Results of the MARL controllers deployed on Climate 3A (performance improvement in brackets) | 107 |
| 3.15 | Results of the MARL controllers deployed on Climate 5A (performance improvement in brackets) | 107 |
| 4.1 | Building and energy systems properties | 118 |

| | | |
|-----|---|-----|
| 4.2 | DNN hyperparameters for each building model | 121 |
| 4.3 | State-space variables | 125 |
| 4.4 | Reward function coefficients | 128 |
| 4.5 | Hyperparameter settings | 129 |
| 4.6 | Evaluation metrics | 130 |
| 4.7 | Metrics related to indoor temperature control | 133 |
| 4.8 | Comparison between performances of the two control strategies . . . | 135 |
| 5.1 | Parameters and modified features used for the design of experiment | 150 |
| 5.2 | Neural network hyperparameter optimization process | 153 |
| A.1 | Settings of the DRL hyperparameters for coordinated and cooperative architectures | 210 |
| A.2 | Settings of the control problem hyperparameters for coordinated and cooperative architectures | 210 |

Nomenclature

Acronyms / Abbreviations

A/S Ancillary Service

AC Air Conditioning

AI Artificial Intelligence

ANN Artificial Neural Network

ARMAX Autoregressive-Moving-Average with Exogenous Inputs

ASO Automated System Optimization

BAS Building Automation System

BEMS Building Energy Management System

BESS Battery Energy Storage System

BNN Bayesian Neural Network

CHP Combined Heat and Power

CNN Convolutional Neural Network

COP Coefficient Of Performance

CPP Critical Peak Pricing

CV Computer Vision

CV-RMSE Coefficient of Variation of Root Mean Squared Error

| | |
|------|--|
| DC | Declared Capacity |
| DER | Distributed Energy Resources |
| DH | District Heating |
| DHW | Domestic Hot Water |
| DHW | Domestic Hot Water |
| DLC | Direct Load Control |
| DNN | Deep Neural Network |
| DOE | Department of Energy |
| DOE | Department of Energy |
| DP | Dynamic Programming |
| DQN | Deep-Q Network |
| DR | Demand Response |
| DRL | Deep Reinforcement Learning |
| DSM | Demand Side Management |
| EIS | Energy Information System |
| EMIS | Energy Management and Information System |
| EUI | Energy Use Intensity |
| EV | Electric Vehicle |
| FC | Fully Connected |
| FDD | Fault Detection and Diagnosis |
| FF | Flexibility Factor |
| GA | Genetic Algorithm |
| GEB | Grid-interactive Efficient Buildings |

| | |
|-------|---|
| GRU | Gated Recurrent Unit |
| HEMS | Home Energy Management System |
| HVAC | Heating, Ventilation and Air Conditioning |
| IAQ | Indoor Air Quality |
| ICT | Information and Communication Technology |
| IES | Integrated Energy System |
| IoT | Internet of Things |
| KPI | Key Performance Indicators |
| LP | Linear Programming |
| LSTM | Long Short Term Memory |
| MADRL | Multi-Agent Deep Reinforcement Learning |
| MAE | Mean Absolute Error |
| MAE | Mean Absolute Error |
| MAPE | Mean Absolute Percentage Error |
| MAPE | Mean Absolute Percentage Error |
| MAS | Multi-Agent System |
| MDP | Markov Decision Process |
| MEL | Miscellaneous Electric Load |
| MILP | Mixed Integer Linear Programming |
| ML | Machine Learning |
| MLP | Multi-Layer Perceptron |
| MPC | Model Predictive Control |
| MSE | Mean Squared Error |

| | |
|-------|--|
| MSE | Mean Squared Error |
| NILM | Non-Intrusive Load Monitoring |
| NLP | Non-Linear Programming |
| NN | Neural Network |
| NZEB | Nearly-Zero Energy Building |
| PAR | Peak-to-average Ratio |
| PCC | Pearson Correlation Coefficient |
| PCM | Phase Change Material |
| PID | Proportional-Integrative-Derivative |
| PINN | Physics-Informed Neural Network |
| PLR | Partial Load Ratio |
| POMDP | Partially Observable Markov Decision Process |
| POMG | Partially Observable Markov Game |
| QP | Quadratic Programming |
| RBC | Rule-Based Control |
| RELU | Rectified Linear Unit |
| RES | Renewable Energy Sources |
| RL | Reinforcement Learning |
| RMSE | Root Mean Squared Error |
| RNN | Recurrent Neural Network |
| SAC | Soft Actor-Critic |
| SF | Self-Sufficiency |
| SHW | Sanitary Hot Water |

SOC State of Charge

SVM Support Vector Machine

TCL Thermostatically Controlled Loads

TES Thermal Energy Storage

TESS Thermal Energy Storage System

TL Transfer Learning

TOU Time of use

V2G Vehicle to Grid

VAV Variable Air Volume

XGBoost EXtreme Gradient Boosting

Chapter 1

Introduction

The growing adoption of automation and control systems, information, and communication technologies (ICT), and internet of things (IoT) sensors in smart buildings has contributed to the unprecedented availability of long-term monitoring data related to the energy performance and indoor quality of the built environment. As a result, complex building-related databases are more available than in the past, and their exploration provides the opportunity to effectively characterise the actual building energy behaviour to optimise the performance of its energy systems during operation. The size, complexity, and heterogeneity of building-related databases make it increasingly necessary for the introduction of frameworks based on an effective coupling of machine learning and energy domain knowledge to extract ready-to-implement strategies for building energy management and information systems (EMIS) [1]. Machine learning (ML) methods proved to be effective tools to valorize the knowledge that can be extracted from data and have been applied in various applications across the building life cycle to improve building performance [2, 3] and occupant comfort and health [4]. The most promising applications for building energy management are the prediction of energy demand required for the efficient operation of a building [5] and the optimization of building operation [6–8], the detection and commissioning of operational failures of building equipment [9, 10], the energy benchmarking analysis [11, 12], the characterisation of energy demand profiles [13–15], and the assessment of the impact of user behaviour [16]. Currently, the building industry is exploiting ML with the progressive introduction of energy management and information systems, which enhance and integrate the functionalities of traditional building

automation system (BAS) to analyse and control building energy use and system performance.

The EMIS includes the energy information systems (EIS) and fault detection and diagnostic (FDD) systems, which are aimed to support the decisions using informative solutions (one-way communication with the BAS), and automated system optimization (ASO) tools, which optimize the control settings (two-way communication paradigm with the BAS) [17, 18]. EIS includes both predictive and descriptive analytics for performing tasks such as load prediction, anomaly detection, advanced benchmarking, load profiling, and schedule optimisation of building energy systems. FDD systems help to detect abnormal system states whose identification and diagnosis can lead to significant energy savings. The 2016–2020 Smart Energy Analytics Campaign [17] assessed the costs and benefits of EMIS installations for several different building types and sizes, including 104 commercial organizations across the United States and more than 6,500 buildings. By the second year of installation, a median annual energy savings of three percent with EIS, and nine percent with FDD tools, were evaluated, supporting the use of such technologies in buildings. Energy information systems are helpful tools to provide data-driven insights in buildings, however, their inability to directly act can strongly affect potential savings, if not coupled with actions. On the other hand, ASO tools actively operate on energy systems and they can include predictive and adaptive control solutions (e.g., model predictive control or reinforcement learning-based control), able to optimise the settings of building energy systems considering the trade-off between multiple and contrasting objectives, achieving an annual cost reduction that ranges from 11% to 16% [19].

Furthermore, the complexity of the built environment has been increasing, due to the introduction of distributed energy resources (DER) and demand side management (DSM), which paved the way for the definition of Grid-interactive Efficient Buildings (GEB) [20], with a key role in the energy transition [21]. The deployment of smart meters and grid automation technologies offers enormous potential to improve the efficiency, flexibility, and resilience of GEB energy systems. Grid-interactive efficient buildings are the ideal candidate for the application of machine learning techniques, since can use analytics supported by sensors and controls to optimize energy use for occupant patterns, while considering their preferences and modifying their consumptions according to utility price signals, weather forecasts, on site energy generation and storage. In this context, ASO tools can exploit the energy flexibility

provided by GEB to further decrease costs, participating to Demand Response (DR) programs. Energy flexibility is defined as the ability to adapt energy consumption and storage operation without compromising technical and comfort constraints, to increase on-site renewable energy consumption, reduce costs and provide services to the grid (i.e. load shifting, peak shaving) [6]. However, when energy management is faced with shifting from a single building to multiple buildings, uncoordinated strategies for exploiting energy flexibility may have negative effects on the grid reliability, causing undesirable new peaks. 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 district of buildings. To overcome the problem, coordinated optimization and collaborative management of various smart grid actors have been proposed [22], paving the way for power systems to fully enter the digital era, leveraging new technologies such as the Internet of Things (IoT), real-time monitoring and control, peer-to-peer energy, and smart contracts [23] to ensure more efficient, reliable, and sustainable electricity dispatch. In this context, the recent development of artificial intelligence (AI) supported the creation of advanced control strategies, able to exploit forecasting and online analytics to enhance energy management [24]. However, their implementation is still limited due to a lack of guidelines and case studies, able to showcase the effectiveness of advanced control strategies. Furthermore, despite some applications existing at the single building level, there is a lack of proof-of-concept at the district level able to exploit new technologies and scale the previous advantages.

This dissertation aims at analyzing the potentialities provided by the exploitation of AI techniques to scale energy management in buildings. The main purpose of the thesis is to identify promising directions and potential barriers for the real-world application of AI in buildings at scale. The thesis will firstly study the application of AI-based advanced control strategies to optimize the energy management of multiple grid-interactive buildings. Then, it will analyze how data-driven models can be used for building operation and control, automating decision-making and easing the deployment of energy management systems at scale.

1.1 Motivations of the research

The increasing complexity of the built environment offers great potential for improving energy management, leveraging ASO and EIS tools to enhance grid-integration and optimize the performance of energy systems at scale. However, building management systems (BMS) are based on classical approaches such as rule-based control (RBC) and proportional-integrative-derivative (PID) controllers. Despite their simplicity, these controllers are characterized by a reactive approach, thus being unable to be optimized in a changing environment, and to handle the multi-objective nature of the building energy management problem, that involves multiple stakeholders [25, 26]. To overcome these limitations, new control paradigms, henceforth referred to as “advanced control strategies”, use a predictive and adaptive approach to perform optimal or near-optimal energy management. The last few years have been the breeding ground for many publications of such techniques in the built environment [27, 28]. Different advanced controllers are categorized based on model reliance. Indeed, there are two main approaches used to represent a system, a model-based approach and a model-free approach, briefly described below:

- **Model-based:** the model-based controller leverages a model of the control environment to obtain information that will be used by an optimizer to maximize a specific objective function [8]. In addition, the optimizer can exploit predictions of the environment to increase the quality of the resulting control policy.
- **Model-free:** model-free controllers do not require a model of the environment, learning a near-optimal control policy, interacting with the environment, using a trial-and-error approach, and a reward mechanism to increase control policy performance as new experience is acquired [29].

Despite being non-exhaustive, the proposed classification highlights the main paradigm of the two approaches, schematized in Figure 1.1.

Model-based approaches can leverage physical laws and varying levels of complexity to support the decision-making problem. As it can be expected, an increasing model complexity is associated with a more refined control, however, its complexity directly limits its scalability. Indeed, the creation of a detailed model is

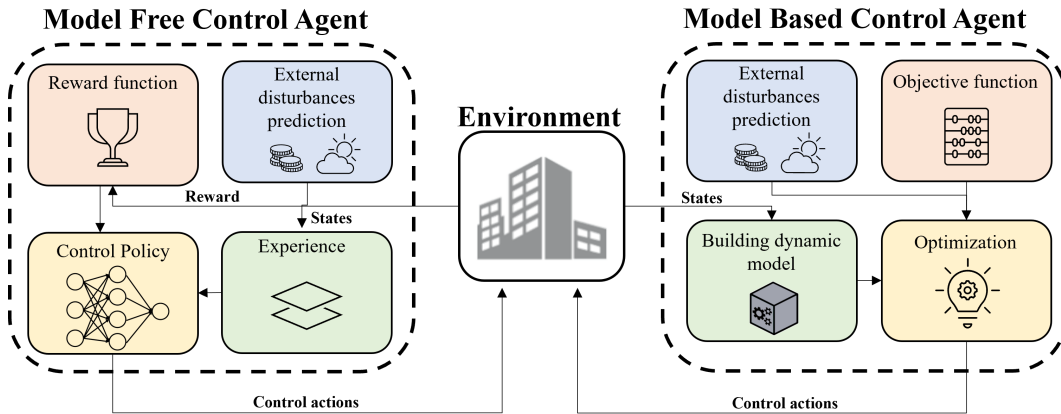


Fig. 1.1 Model-based and model-free flowchart comparison

a labor-intensive process that requires a human expert. Among model-based controllers, MPC stands out for its ability to consider complex systems characterized by non-linear and time-varying dynamics, performing an optimization process over a receding time horizon [30]. However, due to the necessity of a complex and tailored model, there is still not a broad implementation for these kinds of controllers in the building industry [31]. On the other hand, model-free controllers were born to overcome the reliance on a detailed model of the control environment, thus being more scalable. Among model-free methods, reinforcement learning is rapidly emerging in the built environment. Reinforcement Learning (RL) is a branch of machine learning conceived to solve control problems and sequential decision-making processes [32]. RL uses an agent-based control, where the agent learns through interaction with the controlled environment. Among the main limitations of classic RL, there is the data efficiency when dealing with complex problems, such as the ones faced within building energy systems. The evolution of artificial intelligence supported the development of new techniques able to handle such complexity, leading to the creation of Deep Reinforcement Learning (DRL). DRL employs Deep Neural Networks (DNN) as function approximates of the control policy, achieving nearly-human performances on a variety of tasks [33]. The advantage of DRL lies in its ability to adapt to new conditions while requiring minimal human intervention. This is particularly effective for building energy systems, exposed to degradation, retrofit, and stochastic use. However, the exploration phase needed by the controller to reach near-optimal performance may influence user comfort, compromising its real-world availability. As a result, this control strategy obtained high research interest at the single building

level, but it is still in its infancy for large-scale application in the built environment, needing further exploration. The application of DRL at the district level introduces additional challenges, including the higher computational power needed to simulate such scale and the longer exploration phase required by the agents to understand the interaction between multiple buildings, in addition to the one among the building and the environment. Indeed, to overcome these challenges a series of assumptions and simplifications are often made, such as fixed energy demand profiles. For example, the previous assumption neglects the importance to consider humans in the control loop and the possibility to exploit building thermal mass.

In conclusion, due to the opportunities provided by the integration of artificial intelligence (AI) in the building sector, the thesis will investigate how AI can help to overcome the introduced challenges, helping to scale energy management applications. The first part of the thesis will help to scale the application of controllers based on DRL from single buildings to districts, while the second one will study how to increase the level of detail of building simulation at scale, investigating the role of surrogate models based on data-driven techniques.

1.2 Research outline

To assess the effectiveness of different AI techniques to scale energy management in buildings, several applications and case studies were investigated. Figure 1.2 shows the scale of analysis considered for both control and thermal load modeling purposes. The division into quadrants derives from the clashing nature of models and controllers. A more accurate model produces more efficient control, but it also comes at the price of becoming more computationally complex. As a result, very often when the scale of control increases, there is a compromise in the model's accuracy. The thesis aims to understand how to leverage artificial intelligence to increase the scale of analysis while not compromising model's accuracy. The starting point of the thesis is the analysis of the common ground of existing literature, that focused its attention on the control of single buildings, modeled using detailed physics-based models. On the other hand, the thesis will consider applications that are firstly aimed to assess the potentialities of data-driven control at cluster of building scale, using pre-computed fixed demand, and then shifting the attention towards the possibility to employ data-driven surrogate models to simulate building

thermal loads, paving the way for an advanced and detailed control at cluster of building scale. Lastly, an application that tries to speed up the creation of multiple surrogate models is proposed. In particular, control applications studied the adoption of deep reinforcement learning in multiple buildings, leveraging a novel simulation environment created to ease building simulation at scale and the implementation of reinforcement learning based controllers, described in Section 3.1. Different levels of complexity were included in the investigation of data-driven controllers.

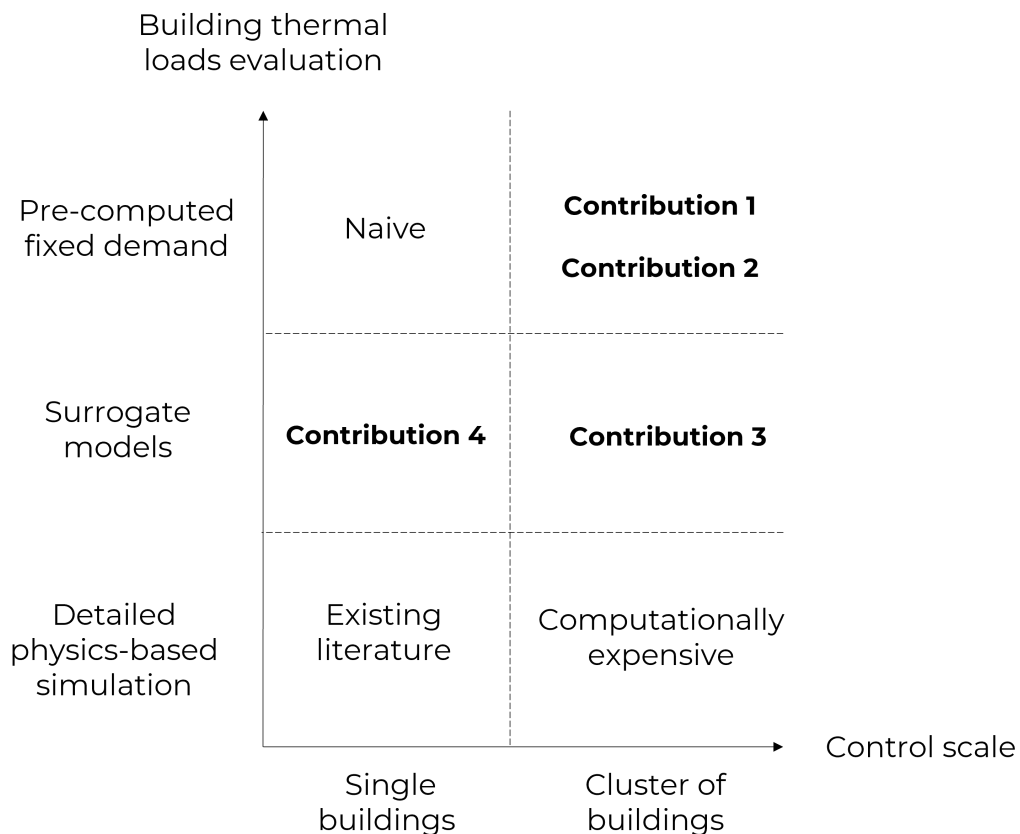


Fig. 1.2 Thesis contributions according to model and control scale

1. Enhancing energy management in grid-interactive buildings

The first case study considered four buildings equipped with thermal storage and PV. The application aimed at studying the feasibility of a reinforcement learning based controller for energy management in multiple buildings. The methodology used a centralized approach that tried to pursue different objective functions, studying the effect of an advanced control strategy at three different

levels: single buildings, district, and grid. Furthermore, the adaptability of the controller was studied, testing its performance in different climates. A detailed description of the application is discussed in Section 3.2

2. Comparing multi-agent architectures in grid-interactive buildings

The second application further pushes the capabilities of data-driven controllers, comparing different multi-agent architectures in a case study similar to the first application, introducing another objective in the control problem, represented by the presence of PV electricity surplus. The application compared a centralized coordinated controller with decentralized cooperative controllers, studying their pros and cons for the energy management of multiple buildings. The application provides guidelines and limitations for a real-world application of these solutions. The comparison and a detailed description are provided in Section 3.3

Then, to assess the potentialities of combining data-driven models and controllers a specific framework, described below, was conceived.

3. A methodology to combine data-driven models and data-driven controllers

This application aimed to couple data-driven models able to represent building thermal dynamics with a DRL controller based on the one developed in the previous applications. The main benefit of employing a model that described the thermal dynamic was the possibility to leverage the thermal mass as an additional source of flexibility, allowing to manage the HVAC and introducing comfort in the control problem. To achieve so, a new simulation environment was conceived and built. To test the proposed methodology, a case study of four commercial buildings was analyzed. In addition to economic savings and better energy management at the grid level, the case study showed the potentialities of a fully data-driven scheme for district energy management, that leveraged the concepts previously introduced within the thesis. A detailed description of the methodology and its application is provided in Section 4.

Lastly, the thesis focuses on how to scale-out data-driven models, assessing the limitations of their effective application in the building field. As a result, the

thesis identified and reviewed the application of transfer learning (TL) in buildings. Furthermore, it also proposes a methodology to assess its strengths and limitations.

4. **Sharing building dynamic models to support energy management**

This application tries to study real-world potentialities and limitations of transfer learning to ease the simulation of building thermal dynamics at scale. Transfer learning aims to exploit knowledge in a similar building to increase the performance on another building. However, the building similarity is hard to define and it is still not clear how to quantify it. Leveraging a synthetic dataset made up of hundreds of simulations, the application tries to quantify the influence of data availability, energy efficiency level, occupancy, and climate on model performances, building several neural networks to represent thermal dynamics. The methodology identified case study applications and limitations of the technique, together with suggestions for its online implementation. A detailed discussion of the findings is provided in Section 5.6.

All the developed tools leverage machine learning frameworks to scale energy management in buildings. To this purpose, the developed data-driven frameworks exploited deep reinforcement learning to scale-up controllers, from individual buildings to multiple grid-interactive buildings. Furthermore, artificial neural networks such as LSTM were used to represent building thermal dynamics, speeding up simulations, and allowing an efficient coupling with data-driven controllers previously cited. Lastly, transfer learning was adopted to scale-out data-driven building dynamics models. This can lead to the development of a digital twin that will further adopt advanced controllers.

Figure 1.3 shows thesis contributions for data-driven models and controllers, highlighting the scale, the main goal of the applications, and the methods used, further described in the next sections. The thesis refers to multiple independent buildings to highlight that the analysis involves various buildings. Still, they do not have any interaction, while the terms numerous buildings and cluster of buildings refer to a group of buildings that can interact with each other. Lastly, it also points out the main novelties introduced by each contribution, which will be deeply described in each application.

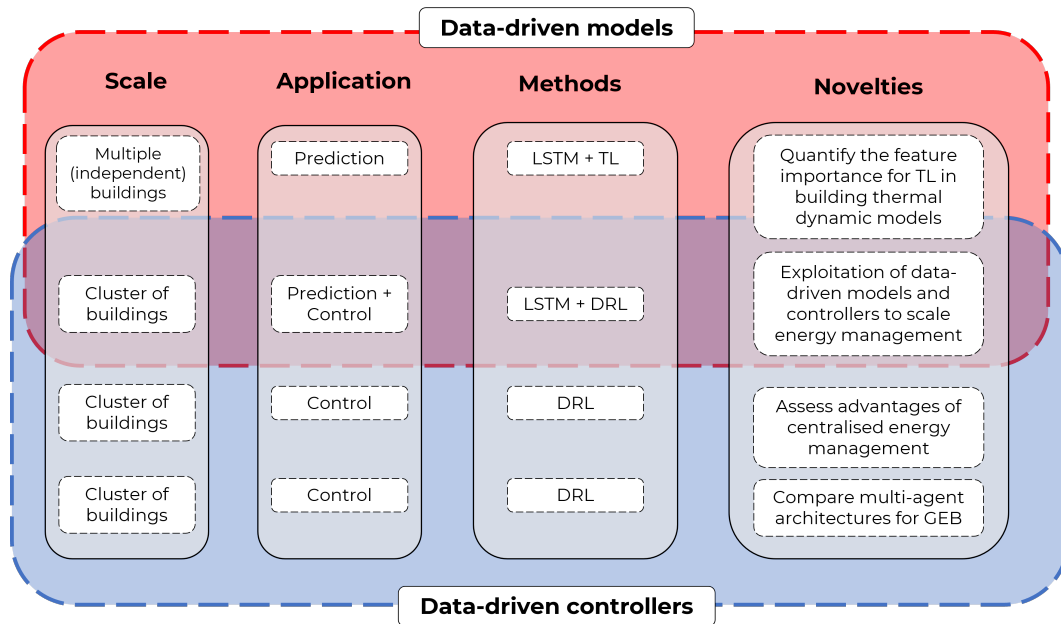


Fig. 1.3 Description of applications involving data-driven models and controllers

1.3 Research questions

The previous sections detailed the opportunities of using machine learning for improving energy management, also discussing the importance of machine learning as a tool to speed-up simulations or forecasting models for renewable energy and electricity prices.

The application of advanced control strategies in buildings that leverage time series analytics is not unusual. Extracted knowledge from building-related data can be exploited to understand the relations between building energy needs and control problem constraints (e.g., weather, occupancy, electricity price). Elaborating on this point, machine learning proved to be effective in providing forecasting for such variables, to optimize the control problem over a certain time-horizon, enabling the use of advanced controllers. To further explore this aspect, environments that simulate the internal dynamics of the building can be used to ensure users' comfort while optimizing the energy use of the buildings.

However, despite their proven effectiveness, data-driven dynamics simulation frameworks are not widely adopted, due to their reliance on historical data. This is even more true when the scale of the analysis is shifted from single buildings to cluster of buildings. Indeed, regardless of the recent interest in multi-agent energy

management in smart cities, the research field still needs significant contributions to provide a generalisable and robust framework that can untap the potential of data-driven applications in energy management at scale.

For this purpose, the dissertation aims to address the following question:

How artificial intelligence can be leveraged to scale the applications of data-driven energy management in buildings?

In particular, the thesis defines the energy management scaling process as follows:

- **Scale-up:** the term refers to the action of increasing in size or number. Declined to the energy management systems, the scale-up process considers the management of multiple buildings or energy systems, to provide services to the grid thanks to the application of data-driven controllers.
- **Scale-out:** the term refers to the action of adding more components in parallel to spread out the load. In the context of building energy management, multiple data-driven models can be used to substitute standard simulation environments, speeding up the simulation process.

As a consequence, the main research question is further articulated in more specific topics as follows:

- *What are the best control algorithms to support the scale-up of energy management?*
- *What are the most effective control architectures for grid-interactive buildings?*
- *How data-driven controllers can exploit the energy flexibility of the building sector?*
- *How to scale-out energy management with data-driven models?*
- *How to combine scale-out and scale-up to achieve a data-driven energy management framework?*

The present dissertation aims at discussing and proposing solutions to the aforementioned questions concerning the scaling process of data-driven energy management in buildings with robust and generalisable frameworks based on artificial intelligence

1.4 Objective of the thesis and novelty

Machine learning in buildings is a fast-growing discipline that has already shown its potential in single building studies. However, its application at a large scale has not been fully explored, posing limits on the advantages that such frameworks could provide in a grid-interactive environment. The present study aims at conceiving and testing several methodologies to scale machine learning applications in multiple buildings, providing insights into the effectiveness and scalability of each proposed approach. From this perspective the main research objectives can be summarized as follows:

- Assess the effectiveness of data-driven control strategies for energy management at the district level. Despite the recent interest in advanced control strategies (e.g., reinforcement learning, model predictive control) their application was mainly limited to single buildings. Data-driven controllers represent a viable alternative to tackle the complexity of the control problem.
- Address the advantages and disadvantages of different multi-agent architectures in heterogeneous environments. Advanced control strategies at district level should optimize grid-interaction without penalizing specific users, finding compromises between optimization at multiple levels.
- Demonstrate the adaptability of data-driven models in multiple buildings. Machine learning models can be powerful tools to support building energy management, but their application is limited to the building in which they are used.
- Summarize the advantages provided by transferable data-driven models and controllers and their future perspectives. Support the scaling process of energy management models and controllers, ensuring their ability to be generalisable, allowing for sharing and fair benchmarking.
- Create a fully data-driven framework for district energy management. Many data-driven models at district scale make strong assumptions on a fixed demand to avoid complex thermal dynamic simulation. In this perspective, data-driven models can help speed-up simulations to obtain more accurate results at large scale.

The main objective of the research is to demonstrate how machine learning can efficiently support energy management in buildings, untapping its potential at a district scale. The main novelty of the research is related to the combination of different machine learning techniques to support the scaling process in the two directions previously identified (scale-up and scale-out), intending to create data-driven frameworks that can be generalised among multiple buildings in different conditions (e.g., climate, occupancy, efficiency level).

1.5 Organization of the thesis

The thesis consists of 6 chapters, that can be divided into two main areas. An overview of the outline and its relation to the thesis aim is shown in Figure 1.4.

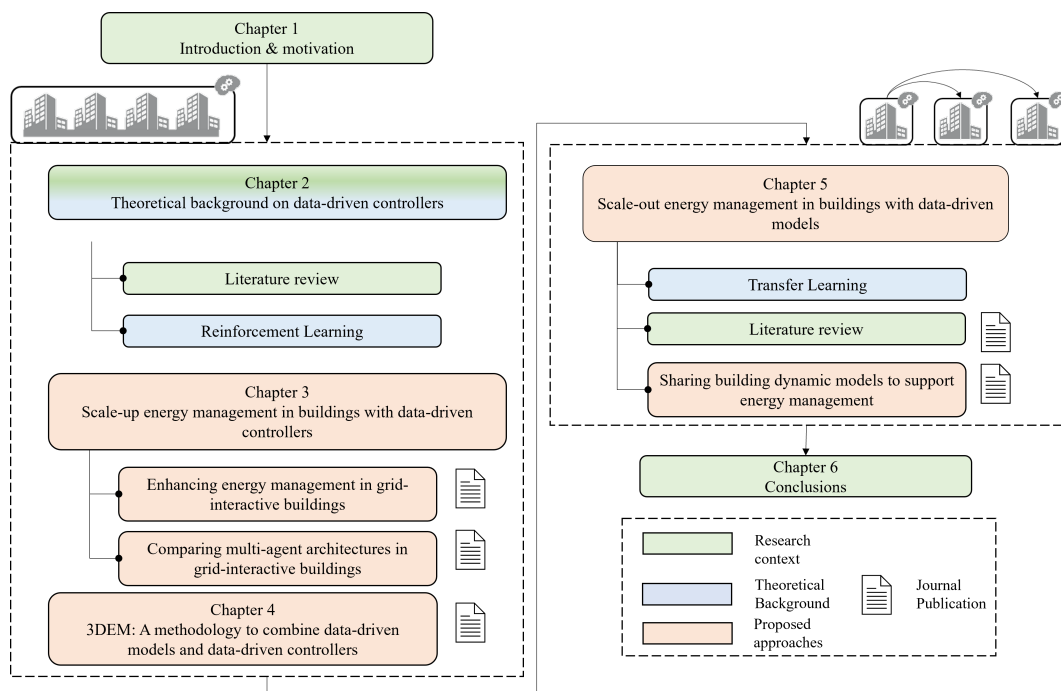


Fig. 1.4 Conceptual organisation of the thesis

Chapter 1 presents the motivation of the research, the objectives, and the organization of the thesis.

Chapter 2 discusses the role of data-driven controllers in scale-up energy management in buildings, while section 2.1 provides insights into the technical concepts treated in the thesis, performing an overview of the most common control strategies

in buildings. Section 2.2 describes the theoretical background that supports the analysis.

Chapter 3 presents the simulation environment used for the deployment of the contributions. Section 3.2 present the first application used to test the proposed methodological frameworks with a centralised RL agent to control the energy storage of four buildings, analysing the effectiveness of the proposed approach for buildings, district, and the grid. Lastly, Section 3.3 further analyses the role of multi-agent architectures in grid-interactive buildings, comparing two advanced control strategies to provide insights into the main advantages and limitations of data-driven controllers in terms of robustness and scalability.

Chapter 4 combines models and controllers to test the effectiveness of data-driven district energy management. The chapter briefly describes the role of machine learning in the creation of building thermal dynamics models and their role in energy management. Then, it presents a methodology that uses LSTM to simulate the indoor temperature dynamics in four buildings, to exploit thermal mass for demand side management. The models are integrated into a simulation environment coupled with a centralised DRL controller, creating a methodology that has the aim to quantify the robustness of a fully data-driven energy management framework at scale, contributing to studying the strengths and limitations of this approach.

Chapter 5 reviews the role of data-driven models in scale-out energy management in buildings, with particular attention to the rule of transferability. Then, it shows an application of transfer learning in smart buildings that aims to analyse the most influencing features for the transferability of building thermal dynamic models, as well as presenting methodologies for their online implementation.

Lastly, Chapter 6 summarizes the work presented in the thesis, giving an overview of the application of data-driven techniques to scale building energy management, identifying opportunities, challenges, and future directions of machine learning in buildings.

Chapter 2

Theoretical background on data-driven controllers

The scope of the present chapter is to analyse current scientific literature to investigate the best practices used to achieve advanced control in multiple buildings, starting from the analysis of the different scale considered, focusing on architectures, control strategies, energy systems, control objectives, and simulation paradigms. Then, the chapter provides a theoretical background of reinforcement learning, which has been selected as control strategy for the thesis. Portions of the present Chapter were already published in the following scientific papers:

- Giuseppe Pinto, Silvio Brandi, Josè Ramòn Vazquez-Canteli, Zoltán Nagy, and Alfonso Capozzoli. Towards Coordinated Energy Management in Buildings via Deep Reinforcement Learning.pdf. pages 1–14, 2020 [34]
- Giuseppe Pinto, Marco Savino Piscitelli, José Ramón Vázquez-Canteli, Zoltán Nagy, and Alfonso Capozzoli. Coordinated energy management for a cluster of buildings through deep reinforcement learning. *Energy*, 229:120725, 2021 [35]
- Giuseppe Pinto, Davide Deltetto, and Alfonso Capozzoli. Data-driven district energy management with surrogate models and deep reinforcement learning. *Applied Energy*, 304:117642, 2021 [36]
- Giuseppe Pinto, Anjukan Kathirgamanathan, Eleni Mangina, Donal P. Finn, and Alfonso Capozzoli. Enhancing energy management in grid-interactive

buildings: A comparison among cooperative and coordinated architectures. *Applied Energy*, 310:118497, 2022 [37]

- Davide Deltetto, Davide Coraci, Giuseppe Pinto, Marco Savino Piscitelli, and Alfonso Capozzoli. Exploring the Potentialities of Deep Reinforcement Learning for Incentive-Based Demand Response in a Cluster of Small Commercial Buildings. *Energies*, 14(10), 2021 [38]

2.1 Literature review

This section provides the theoretical background of the chapter, that aims to describe the different opportunities introduced by the application of advanced controllers in buildings. Figure 2.1 shows a conceptual scheme of four different criteria used to classify energy management in buildings, discussing previous works and opportunities.

To highlight how data-driven controllers can help to scale energy management, the starting point is the definition of the scale of analysis.

Section 2.1.1 defines the concept of a cluster of buildings, how the scale of analysis affects the complexity of the control problem, and the introduced limitations. Then, Section 2.1.2 shows an overview of the control architectures and strategies adopted in this scale of analysis, introducing the most common advanced control techniques. After the analysis of control strategies, Section 2.1.3 presents a description of the energy systems used in buildings, with detail on renewable energy sources (RES), storage equipment, and HVAC systems. Furthermore, the analysis focuses on the different ways in which these systems can be integrated and how they can be fully exploited to maximize several objective functions, closely linked to available energy systems. Section 2.1.4 describes the different environment configurations and levels of detail associated with the building energy simulation. Lastly, Section 2.1.5 discusses a brief literature review and identifies the proposed scale of analysis, level of detail, and control techniques chosen for the thesis.

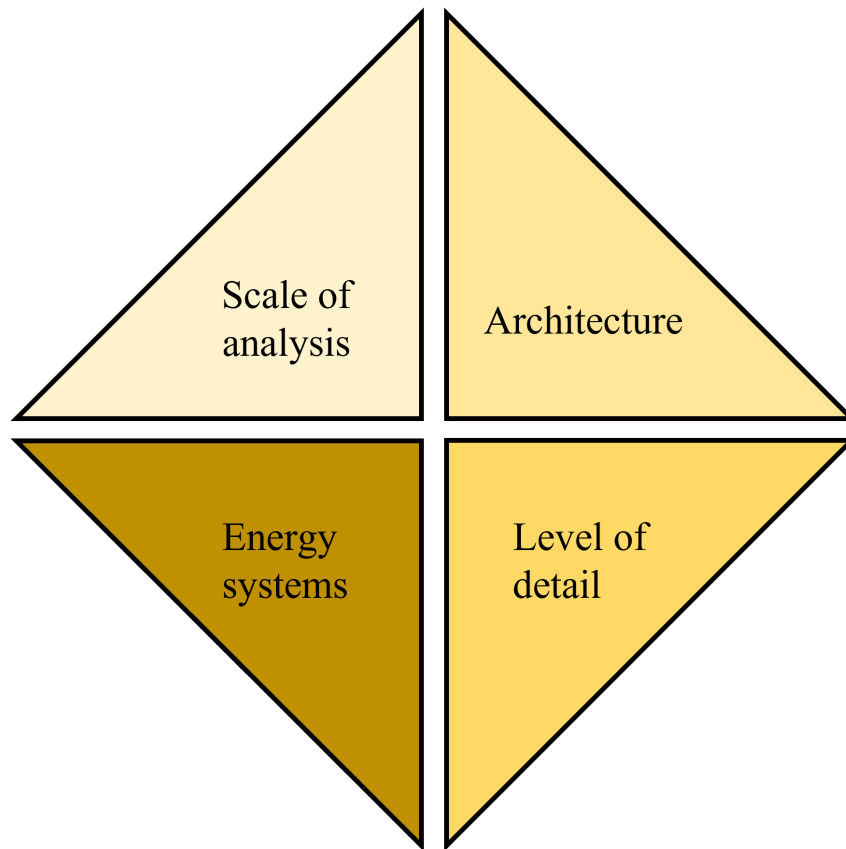


Fig. 2.1 Organization of the literature review

2.1.1 Scale of analysis

The vast majority of literature in the past years focused its interest on the optimization of energy systems in single buildings, both residential and commercial [39, 40]. The increasing complexity of the built environment and its ability to interact with the grid have renewed the interest in the coupling of advanced control and flexible energy systems. In this context, energy flexibility, defined as "*the ability of adapting energy consumption and storage operations without compromising technical and comfort constraints*", is exploited in energy management for different purposes, including increasing on-site renewable energy consumption, reducing costs, and providing services to the grid (e.g., load shifting, peak shaving) [41]. This can be achieved thanks to Demand Side Management (DSM) [42] or Demand Response (DR) programs [43]. DR programs incentives users to curtail or shift their building load according to grid requirements, rewarding the users. This kind of programs

can be classified into two main categories, displayed in Figure 2.2: time-based and incentive-based programs [44].

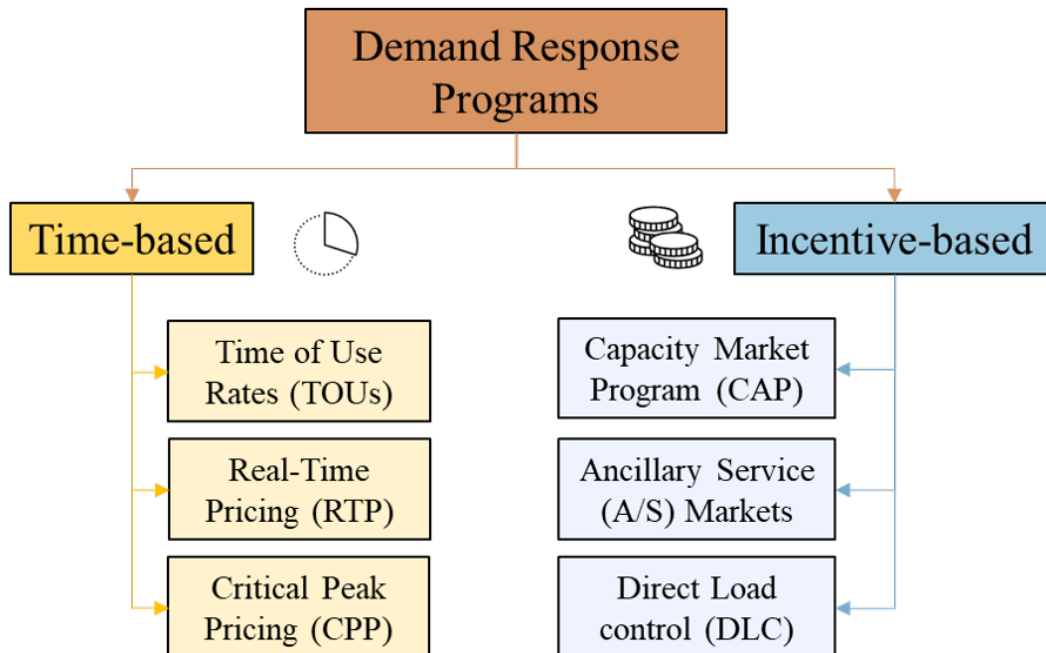


Fig. 2.2 Demand response programs classification [38]

Time-based programs aims to indirectly change consumption patterns through a time-varying price signal established in advance, that incentives energy users to shift their consumption. Different tariffs have different objectives; for example, critical peak pricing (CPP) is a technique used to reduce the demand for a few hours, after an electricity price increase of around 3 times the peak price. This approach is usually used a few times per year and it is useful to protect the grid from stress during peak hours [45]. Real-time pricing is used for customers with high flexibility, which can reduce their consumption according to the price of the electrical market. The program is designed for commercial buildings, that with a high amount of energy can help improve the performance of the power system, as well as benefit from hourly changes in electricity prices and consequently lower their electricity bills. [46]. Lastly, time-of-use (TOU) rates have been introduced to shift demand from peak to off-peak hours. However, the adoption of price-based programs in some circumstances is a double-edged sword, causing new peaks of demand during times of low electricity prices, due to a massive shift of aggregated demand from multiple residential buildings [47].

On the other hand, incentive-based DR leverage a remuneration of the participants that manually or automatically reduce the electrical load after a request from the service provider, while being under specific constraints with penalties for non-conformance of these [48]. Incentive-based DR usually acts on a large scale, using aggregators for capacity market programs, or directly allowing the service provider to curtail some appliances, as in the case of direct load control (DLC). The operation of building equipment under DLC may lead to sub-optimal economic performance, that needs to be balanced by a higher economic incentive [49]. Incentive-based programs are also used to alleviate problems associated with grid congestion at different scales using Ancillary Service (A/S) Markets.

The common idea behind DR programs is the reduction of a certain amount of demand over a certain period. As a result, the scale of analysis plays a key role in the success of such programs. Indeed, the energy flexibility of a single building is typically too small to be bid into a flexibility market, requiring figures such as aggregators. Investigating the effect of energy flexibility at scale helps service providers to address environmental and financial benefits before a real implementation. To study how the scale of analysis plays a role in the energy management problem, the thesis use the word "building cluster" to refer to two or more buildings or units (apartments in a multi-unit residential building), up to neighborhoods, districts, and communities controlled at a substation level in a micro-grid. The cluster of buildings offers the opportunity to coordinate multiple heterogeneous energy systems, including electric vehicles (EVs), battery energy storage systems (BESS), thermal energy storage (TES), renewable energy sources (RES), unlocking flexibility potential at scale. Figure 2.3 shows different dimensions of building clusters, that can be classified as follows:

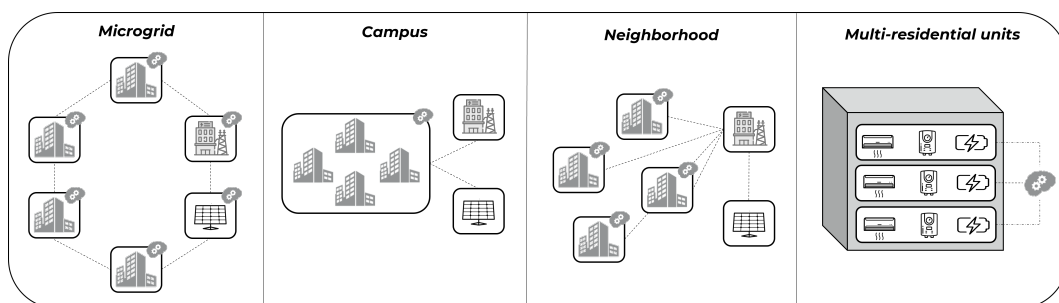


Fig. 2.3 Representation of different sizes of building clusters

- **Multi-unit residence:** a multi-unit residence refers to multiple apartments physically located in the same building, which operation is associated with different owners. These units may or not have common energy systems, such as centralised temperature control systems [50].
- **Neighborhood:** a neighborhood is characterised by the presence of multiple independent houses with no common energy systems that aim to minimize their electricity bill. In this case, if all homeowners are subject to the same dynamic profile, peak rebound may shift the demand during low price periods, limiting the benefits of DR for the grid [51].
- **Campus:** a campus includes multiple buildings with different energy systems managed by a single user. In this case, the use of commercial tariffs based on peak-power highly incentive coordinated management and facilitate access to DR programs [52].
- **Microgrid:** a microgrid is a group of interconnected loads and distributed energy resources within clearly defined electrical boundaries that acts as a single controllable entity for the grid. A microgrid can disconnect from the grid to enable it to operate in island mode, with the main difference to balance voltage in addition to energy flows [53].

Hu et al. [52] and Kaspar et al. [54] provided an overview on the application of DSM at building cluster level, while recent analysis started to investigate the role of aggregated flexibility in multiple buildings [55, 56]. Kazmi et al. [57] discussed how even retrofit design can benefit from shifting the scale of analysis to the neighborhood, while Taniguchi et al. [58] studied the effect of energy management of around 500 homes on grid peaks. Similarly, Perfumo et al. [59] showed the potentiality of the aggregate flexibility of around 10000 residential air conditioners for DSM purposes and Hu and Xiao [60] proved that the aggregated demand flexibility is less sensitive to user stochasticity, representing a reliable source for the grid. However, these studies only focused on flexibility quantification, without real exploitation at large scale. Several studies aimed to make steps in this direction, proposing energy management strategies in multi-residential units. Van Pruissen et al. [61] controlled heating and Domestic Hot Water systems of 79 apartments over two floors, while Comodi et al. [62] presented the result of real-life implementation of six apartments controlled to act like a microgrid, reducing costs thanks to better storage and PV management.

Moving at the neighborhood level, recent studies tried to focus the attention on the ability of energy management at scale to reduce peak consumption. In particular, Huang et al. [63] propose a hierarchical controller for DR in multiple buildings, achieving peak and cost reduction using a planning horizon of 24 hours ahead, while Angizeh et al. [64] achieved a 20% cost reduction by optimizing the operational schedule of community operated assets in a real building cluster. Furthermore, current literature analysed how everything above multi-unit residents can act as a single entity in a microgrid context, as shown in [65], in which an office building was managed as a microgrid to balance the electric power exchange using virtual energy storage systems for short-term and vehicle-to-building exchanges for ultra-short term power balance. This approach has been facilitated by the introduction of EVs and BESS, which provide greater flexibility to buildings, and the vast deployment of Distributed Energy Resources (DER), which expanded the interest in microgrid communities [66].

Despite the advantages achievable in such configurations, these kinds of control require specific sensors, protocols, and architectures. The next subsection will describe the most common techniques and architecture used for building energy management at scale.

2.1.2 Control strategies and architectures

2.1.2.1 Control strategies

The building energy management problem at scale involves multiple interconnected entities that aim to pursue their objective functions. The interaction among these entities can be described using concepts related to social behaviors within society. Common social behaviors in buildings are coordination, cooperation, and negotiation, as described below.

- **Coordination** is an arrangement of group efforts to harmonize individual efforts in pursuit of common goals. The limitations of this control strategy are the following: i) the exponential growth of the state and action spaces with the number of reactive agents may limit real-world implementation; ii) the coordination control may result in sub-optimal solutions for specific

buildings; and iii) private information collection (and their possible sharing) may discourage user participation in a real-world setting.

- **Cooperation** is a voluntary effort of individuals to work together to help each other. In cooperative settings, each building is represented by an agent that learns the optimal policy according to the specific objective function. The limitations of this approach are the following: i) the interaction between multiple control strategies can lead to a non-stationary environment thus challenging the learning process; ii) while the number of agents grows, a large number of models need to be tuned and trained, requiring considerable effort for the definition of reward functions.
- **Negotiation** is a more sophisticated social behavior to solve conflicts among multiple entities in a non-cooperative environment. Multiple buildings need to negotiate to solve their conflicting goals, i.e., maximize the payoffs of both sides. Negotiation is often used in peer-to-peer systems, that try to maximize advantages for the grid and multiple users [67]. In the domain of residential microgrids, game-theory based techniques are normally used to solve conflict situations.

2.1.2.2 Architectures

The energy management classification in multiple buildings is also diversified by the type of entity that determines the action and the sharing of information within the buildings, that can exploit coordination, cooperation or negotiation. If the actions are made by aggregators or utilities between multiple homes, the architecture is centralized and is often associated with a coordinated environment. On the other hand, if each house determines its action, the architecture can either be decentralized or distributed. The difference between these two architectures is related to the information shared among buildings; if there is no information sharing it results in a decentralized architecture, otherwise in a distributed one. Furthermore, the distributed architecture can be classified into hierarchical distributed or non-hierarchical distributed, depending on how the information flows between multiple buildings. Figure 2.4 shows the different architecture types, while a detailed description of the proposed architectures is provided below.

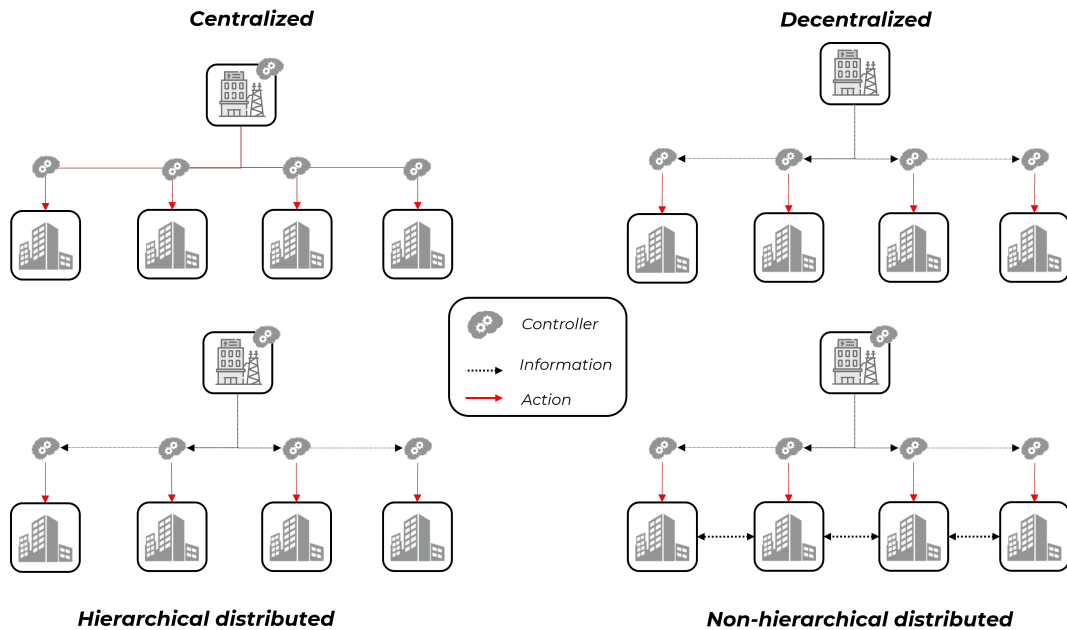


Fig. 2.4 Representation of different architectures for multiple building energy management

- Centralized:** this setting exploits a centralized architecture called cognitive-reactive, in which a cognitive agent uses as inputs the observations of all the buildings (reactive agents), that do not have decision-making capabilities, but respond as actuators to the decision taken by the cognitive agent. Often coordination and centralization are used in the same settings and coordinated energy management is referred to as centralized training with centralized execution.
- Decentralized:** this setting uses a decentralized architecture, in which multiple agents try to find the optimal control strategy using a certain amount of information about the environment but do not have any information or interaction with other agents. Decentralized architecture is often used for cooperative energy management and is referred to as decentralized training with decentralized execution.
- Hierarchical distributed:** the hierarchical distributed architecture takes decisions using a decentralized architecture, but also shares some information among buildings in a centralized way, usually employing a specific agent. The concept behind this idea is that information sharing may increase the

effectiveness of the control policy, allowing multiple agents to account for specific changes of the environment.

- **Non-hierarchical distributed:** in the non-hierarchical distributed architecture, the actions are taken in a decentralized architecture, while the information is shared among peers, differently from the hierarchical distributed structure. Recently this architecture has seen a growing interest due to the opportunity provided by peer-to-peer energy systems in buildings.

To fully exploit the flexibility associated with buildings, the scale of analysis should be between single buildings and aggregated demand, in the so-called neighborhood, communities, districts or integrated micro-grid. The following provides an overview of the architectures used in literature to enhance energy management in multiple buildings. Looking at centralized energy management, Nguyen and Le [68] developed an algorithm to optimize the schedule and usage of HVAC systems and EVs in a residential community. The authors compared the performance of a centralized coordinated agent with one of multiple individual optimizations for each building. The paper showed how a centralized approach can achieve significant savings in electricity cost and allows more flexibility, with the opportunity to reduce peak loads. Tushar et al. [69] also analysed how to exploit EVs charging and discharging, as well as home appliances and distributed generation (PV panels and wind turbines) to reduce the energy consumption of 200 homes with 1400 appliances. The authors compared 3 control architectures, a naïve scheduling framework, a decentralized approach that employed game theory and a centralized approach that used mixed integer linear programming (MILP), observing the superiority of the centralized method. Ouammi [70] used a centralized MPC-based controller to manage the power consumption of multiple smart residential buildings, characterized by a high share of distributed generation that included PV panels, wind turbines, EVs, BESS, cogeneration heat plant (CHP) and home controllable appliances. The proposed centralized control approach enabled the interconnected residential buildings to deal with the uncertainties in the loads and RESs, and to maximize the use of local renewable energy generations in a cooperative manner. Logenthiran et al. [71] used a heuristic evolutionary algorithm to optimize the load shifting of a smart grid of over 2500 appliances in a centralized fashion. The approach led to a 5% electricity cost reduction and a peak power load reduction of 18%. The centralized architecture is often used to fulfill coordination purposes, since the cognitive agent has information

about the environment as a whole, being able to provide the global optimal solution at the system level. However, as previously stated, a centralized architecture may result in sub-optimal solution for some subsystems, as it is not fault-tolerant, relies on a single agent, and the computational burden of the centralized controller limits its application at large scale [72].

On the other hand, decentralized architecture has been used in cases in which the subsystems needed to prioritize the optimization of their performances. Molitor et al. [73] used a two-step decentralized coordination method to reduce the power fluctuations controlling the heating systems of 66 homes included in a residential district. The first step aimed at finding a set of near-optimal schedules for each heating system, followed by the selection of a single schedule per building, chosen to minimize a global objective function at a high level. Cole et al. [74] compared a centralized and decentralized approach controlling the air conditioning systems of around 900 homes to minimize the peak power. The results showed that the coordinated approach reduced the peak of 8.8%, while the decentralized control by 5.7%. As a result, they associate the information sharing with an additional 3.1% of peak power reduction. However, by tuning the penalty-based term the authors suggest that a similar result can be achieved with a reduced computational cost. Decentralized architecture is usually associated with cooperation, even if using a two-step methodology also a coordinated approach can be used. The decentralized architecture decomposes the main task into various sub-tasks solved using multiple local controllers. A drawback of the decentralized architecture lies in the difficulty to achieve optimal cooperation, which usually leads to a selfish optimization of some sub-systems.

To solve the problem of information sharing in decentralized architecture, hierarchical distributed have been explored in recent years. This architecture proved its effectiveness at different levels. Safdarian et al. [75] used a hierarchical distributed architecture to perform a coordinated DR in 50 homes to address peak rebound issues. The hierarchical distributed approach was developed in two stages: during the first one several home energy management systems scheduled the loads to reduce electricity consumption. Then, in the second stage, the load service provider used an iterative approach to update the global profile (and consequent costs) and the HEMSs adjusted their consumption accordingly until an optimum was found. Chavali et al. [76] proposed a distributed framework for DR to control a community of around 100 users. Each user in the systems used an approximate greedy method to optimize their

consumption, which depends in turn on the global profile. As a result, a penalty term was introduced to find a solution that coordinated the buildings using penalties to shave the peak load, providing cost reduction for users and peak reduction for utility companies. Roche et al. [77] exploited a MAS distributed architecture to reduce the peak load of over 5000 homes in a residential community equipped with ACs, water heaters and EVs. A coordinator collected the flexibility bids from the end-users and selected some of them to participate to the DR program, remunerating the participants without explicit effects on comfort conditions. In summary, usually the hierarchical distributed architecture exploits coordination for decision making. This kind of coordination can help overcome sub-optimal solutions for the decentralized architecture.

Lastly, non-hierarchical distributed energy management has shown its effectiveness in several use cases. Very often, game theory is coupled with this kind of control to achieve optimal solutions. Mohsenian-Rad et al. [78] firstly conceived a methodology to combine non-hierarchical distributed architecture in demand side management using game-theory and then used the methodology to an appliance scheduling problem involving multiple homes. The work used dynamic electricity pricing to obtain an optimal aggregate profile using the Nash equilibrium of the resulting game, encouraging users to minimize their electricity bills. As a result, the users can maintain privacy and do not need to reveal the details on their energy consumption schedules to other users, while still reducing the peak-to-average ratio community, the total energy costs, as well as each user's individual daily electricity charges. Basir Khan et al. [79] combined non-hierarchical distributed architecture with non-cooperative game theory to optimize a microgrid constituted of multiple distributed generators, outperforming a conventional centralized control system. On the other hand, Chang et al. [80] developed a non-hierarchical distributed coordinated home energy management (CoHEM) architecture to orchestrate the energy scheduling of multiple households. Compared with selfish HEMS, the proposed CoHEM exchanges information with the neighboring HEMSs, providing the optimal appliance schedules for each household. The work demonstrated the feasibility of a non-hierarchical distributed approach, that can improve real-time power balancing. The main advantage of the non-hierarchical distributed architecture is the ability to be used in negotiation problems, unavailable for the above mentioned architectures.

2.1.2.3 Optimal control

After the description of control type and architecture used to classify the energy management problem in buildings, the optimization technique is the third pillar of effective energy management. Optimization algorithms can be classified according to several criteria: problem linearity, the presence of hard or soft constraints, the objective function considered (e.g., single-objective, multi-objective), the problem nature (e.g., deterministic, stochastic). This work summarizes the optimization algorithms in three main categories, broadly described below and represented in Figure 2.5.

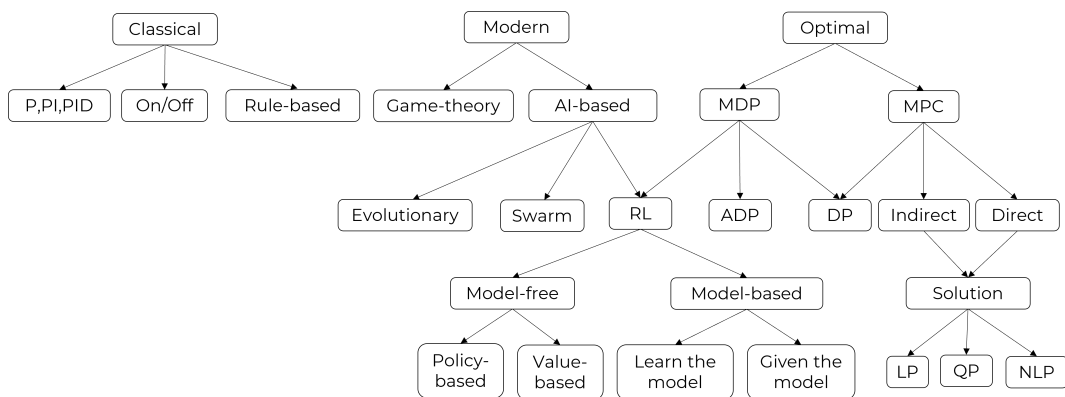


Fig. 2.5 Non-exhaustive taxonomy of optimization techniques for building energy management (based on [81])

- Classical:** this approach includes techniques commonly used in building energy management. On/off techniques is the simplest, this technique is as simple as inefficient and represented one of the first approaches in the industry, overcome by rule-base control (RBC) [82] and Proportional-Integrative-Derivative (PID) [26], applied at several levels of BEMS . These controllers are defined as reactive, since they can only leverage previous information about the environment and the controlled variables, optimizing the control signal to track a certain set-point. The main drawbacks lie in the reactive approach, which lead to sub-optimal solutions. Furthermore, the parameters of these controllers or the adopted rules are often the results of experience and rule of thumbs, rather than an optimization process; in addition, these systems cannot handle multi-objective problems, strongly limiting their application in the modern built environment. Lastly, rules and parameters are often static, meaning that

they do not automatically adapt to the changing environment, degrading their performances over time. Indeed, the manual tuning of PID and RBC is based on domain expertise, leading to a cost-intensive procedure to update them.

- **Modern:** these algorithms have been recently introduced to overcome the limitations of classical controllers. Modern controllers are based on game-theory and AI [3]. While game theory is traditionally associated with social sciences, control strategies have strong ties with it. In particular, control problems can be classified as zero-sum or min-max games. Building energy management at scale includes a large number of decision makers with different objectives and a recent area of interest includes distributed control systems [83]. Smart grids are a clear example of distributed/networked control systems, in which multiple prosumers try to optimize the production, consumption or storage of energy, according to the evolution of the environment [84]. The other common approach used involves AI, that in its broad definition includes nature-inspired evolutionary algorithms and reinforcement learning. Nature-inspired algorithms have been used for planning and control purposes and are often meta-heuristic, motivated by evolution, biological swarms, or physical processes. Meta-heuristic refers to the class of stochastic algorithms that employs a local search to discover near-optimal solutions finding the best trade-off between exploration and exploitation of a control strategy [85]. Nature-inspired algorithms are often used in buildings to schedule loads in HEMS, or to bid energy flexibility in electricity markets. The main advantage is related to the low computational cost required to obtain a near-optimal solution, making them a perfect fit for control tasks that require a real-time operation. Similarly, evolutionary algorithms are heuristic-based approaches that mimic some of the core principles of evolution, as reproduction, intuition, recombination and selection. The strength of evolutionary algorithms lies in their ability to optimize the system without gradient information, which allow their parallelization. However, there is no guarantee of finding the optimal solution. The most common evolutionary algorithm is genetic algorithm (GA) [86], which encoded Charles Darwin's theory of natural selection. As for the meta-heuristic algorithms, also the heuristic-algorithms are often used for appliances scheduling in aggregators price scheme optimization. Lastly, Reinforcement Learning (RL) is a control method based on the paradigm of learning from interaction. RL frames the problem considering an agent with a

predetermined goal that interact with the environment. This technique tries to find trade-off between exploration and exploitation using a trial and error type of search and a delayed reward approach [32]. The RL framework has seen an increasing interest in building energy management, due to its model-free nature [87]. Its application has been studied for appliance scheduling with occupant interaction, HVAC and storage management for DR [88].

- **Optimal:** optimal control embeds advanced control strategies that uses mathematical optimization techniques. The first applications of optimal control appeared in the 1950's with the introduction of Dynamic Programming (DP), a recursive method for multi-stage decision processes. There are mainly two approaches able to handle the complexity and the large scale faced by building energy management problems: Markov Decision Process (MDP) and Model Predictive Control (MPC) [89]. The first one aims at describing the stochastic processes and can be solved with several approaches, including the above mentioned DP, RL and approximate DP. In particular, RL represents the method selected for this work and will be fully described in the next section. On the other hand, MPC is a framework commonly adopted in control theory, that employs a model to simulate a dynamic system and optimize it over a receding time horizon, using common mathematical tools such as Linear Programming, Quadratic Programming (QP) and Non-Linear Programming, depending on the nature of the problem. Moreover, despite its effectiveness, MPC is based on a detailed model, that requires a lot of effort, especially for buildings, that are intrinsically unique [8]. As a consequence, despite providing optimal solutions, MPC is still not widely adopted in the building industry, especially at large scale.

2.1.3 Energy systems and objective functions

Another criterion used to characterize the problem of energy management in buildings is the controlled energy system and its relative objective function. This section describes the most common energy systems used to produce and transform energy in buildings, together with the main objectives and challenges associated with their control. The energy systems are described starting from the lowest level (single building appliances) up to an energy system able to satisfy multiple building loads.

- **Electrical appliances:** such as refrigerator, televisions, and microwaves are defined as non-schedulable appliance, since their control directly affect occupants' comfort and are usually associated with an on/off control. On the other hand, schedulable electrical appliances are loads that can be reduced or shifted without directly affecting user perception. This category is particularly useful for DR events and usually includes dimmable lights, air conditioning (AC), EV, washing machines, and dishwashers.
- **HVAC Loads:** HVAC systems account for 40-50% of overall consumption in commercial buildings (such as schools, university campuses, shopping malls, and offices). Thermostatically controlled loads such as heat pumps, electric water heaters, and air conditioning represent a major contribution to energy consumption and peak power and due to the building thermal inertia, they are prime candidates for DR since they can shift their consumption without affecting users' comfort. Concerning space cooling and heating, the two most common strategies for HVAC equipment subject to a dynamic electricity price are pre-cooling/pre-heating and zone temperature reset. Yoon et al. [90] adjusted the internal temperature set-point according to electricity price to reduce HVAC consumption. The results showed a 10% cost reduction and a 25% peak power reduction. Mtibaa et al. [91] proposed a novel online algorithm based on MPC and GA that optimized the operation of a multi-zone HVAC, reducing energy costs, peak demand, and discomfort. Yoon and Moon [92] used an RL controller combined with a data-driven model that predicted personal comfort to optimize an HVAC.

At a large scale, HVAC and variable-speed AC can also be coordinated to provide ancillary services to the grid. Chassin et al. [93] designed a residential thermostat able to provide ancillary services in a reliable and aggregated way. The thermostat was able to provide a large amount of load elasticity (10-25%) during on-peak times, favoring DER integration. Hu et al. [94] developed a frequency-based MPC controller for variable-speed ACs able to interact with the grid in response to real-time prices with a 5 minute resolution. Compared to a standard PID controller, the proposed method was able to reduce average power consumption during on-peak hours by up to 38% and reduce costs by up to 22%. Lastly, Chen et al. [95] proposed a methodology used for the coordination of thermostatically controlled loads (TCLs) that can modulate

energy demand, decrease operating costs, and increase grid resiliency over a large number of buildings.

- **Electric Vehicles (EV):** vehicle electrification is a trend in line with the energy transition of several countries, helping reduce greenhouse gas emissions and saving energy. According to the U.S. Transportation Department, approximately 70% of EVs are charged at home [96], resulting in higher peak demand in residential buildings. Techniques such as vehicle-to-grid (V2G) have been proposed to exploit the large capacities of EVs' batteries, using them as distributed energy resources for grid stability and DR purposes, rewarding the users with financial benefits [97]. Electric vehicles are seen both as appliances (when charged) and as storage (if discharged), and recently a large amount of literature has investigated the control of single [98, 99] and aggregated EVs [100, 101]. [102] designed a DRL controller to manage a residential building equipped with a heat pump and an electric vehicle, resulting in around 40% of cost savings compared to a naive baseline in a real case-study.
- **Battery Energy Storage Systems (BESS):** are among the most used energy storage systems today, and usually employ electrochemical batteries such as lead-acid, nickel-cadmium, and lithium-ion to temporarily store electricity. Many studies investigated the flexibility provided by the installation of such energy systems in single buildings and microgrids. [103] reported a 15% peak power reduction with the installation of a battery system of 1 kWh capacity in a residential building, while [104] explored different BESS sizes in Canadian houses for load shaving purposes, according to the electricity intensity of the different homes. [105] studied the effect of coupling a battery with a PV, using a HEMS for demand response. At a larger scale, [106] exploited MPC as an aggregator for multiple energy systems including BESS, tailoring charge and discharge based on real time prices.
- **Thermal energy storage system (TESS) :** thermal storage showed a great ability to shift peak power loads to low-price times, especially for cooling loads. A TESS can store thermal energy exploiting a cooling/heating process without material change (sensitive heat storage), or a phase changing material that can either solidify, melt, vaporize or condense (latent heat storage). Due to the increase in complexity of systems, commercial buildings and residential buildings use different types of storage. Commercial buildings often employ

ice/chilled water tanks to pre-cool or reduce peak power consumption. On the other hand, residential buildings commonly use hot storage for domestic hot water or space heating [107]. Comodi et al. [62] studied the effect of installing water TESS in a real residential microgrid made up of six apartments, highlighting the effect of storage size on self-sufficiency and demand side management. Alimohammadisagvand et al.[108] integrated TESS to shift the space heating load in residential buildings equipped with heat pumps. Furthermore, Fiorentini et al. [109] analyzed the effect of the introduction of Phase Change Material (PCM) storage coupled with a photovoltaic-thermal system in a nearly-zero energy building (NZEB) , showing higher efficiency of the whole system. Additionally, a recent trend has emerged for the exploitation of building thermal mass, seen as passive thermal energy storage, allowing for pre-cooling [110] or pre-heating buildings [111]. Turner et al. [110] investigated how pre-cooling strategies can be used for load shifting, achieving a shift of around 50% of load peaks from a period from 4pm to 8pm. Similarly, Reynders et al.[112] exploited the energy flexibility provided by the structural thermal mass of a residential building, lowering electricity consumption during peak periods. Lastly, at large scale Dominikovic et al. [113] evaluated the potential of building thermal mass in district heating systems, quantified between 5 and 8% of the total district heating demand, reaching about 6 hours of flexibility for some buildings.

- **Renewable energy sources (RES):** includes a series of energy systems that exploit renewable energy, such as PV panels, solar thermal panels, wind power, biomass plants and geothermal plants. Despite an increasing trend in their use, the main limitation of these technologies lies in their stochasticity, which can jeopardize the grid. Nevertheless, their coupling with BESS and TES is further increasing their use even in the building sector. Anvari-Moghaddam et al. [114] proposes an ontology-driven multi-agent based energy management system of an integrated microgrid system with various renewable energy resources and controllable loads, controlling battery operation to reduce imported electricity. Raman et al. [115] compared the effectiveness of multiple controllers, including RBC, MPC and RL, to manage an integrated energy system (IES) consisting of PV and battery. Tascikaraoglu et al. [116] studied the effect of the forecasting and demand-side management strategies in a house equipped with wind turbines and solar panels achieving a 4.2% cost

reduction. Bilardo et al. [117] explored the ability of a solar cooling system to meet the summer energy demand of a multi-family building. The resulting optimal design reduced the non-renewable primary energy demand by 48%, increasing the renewable energy ratio up to 83%. In conclusion, previous studies highlighted how RESs and other energy systems are interconnected and mutually influenced. Building energy flexibility benefits the penetration of RESs, while EMS can achieve higher savings with storage and RESs.

- **Combined heating and power (CHP):** these systems have been widely utilized as distributed energy resources in recent years due to their high energy efficiency, cost effectiveness, low greenhouse gas emissions and high reliability [118]. CHP combined-cycle power plants can deliver concurrent production of electricity and useful thermal energy from a common fuel. The captured thermal energy (steam or hot water) can be used for processes like heating and cooling, and to generate power. Zhang et al. [119] proposed two-stage coordinated energy management strategy for supply side, consisting of CHP and other RESs, and supply side, including electric and thermal loads, taking into account RESs uncertainties. Results showed higher efficiency and economic benefits using a coordination between the two sides. Additionally, CHP can be combined with district heating (DH), an underground infrastructure asset where thermal energy is provided to multiple buildings from a central energy plant or plants. Steam or hot water produced at the plant is transmitted 24/7 through highly insulated underground thermal piping networks. Guelpa et al. [120] analysed the opportunities for peak load shaving in district heating systems managing the thermal request profile of multiple buildings using local storage systems. Pinto et al. [121] presented a methodology to support decision making about carbon-neutral technologies for district heating, using a multi-criteria approach that encoded different objective functions, including economic, environmental and technical objectives.

As buildings are becoming more complex, the flexibility provided by different energy systems allows considering more complex objective functions. In particular, depending on the energy systems considered and the scale of analysis, the following major goals have been identified in the literature:

- **Energy conservation:** aims at minimizing energy consumption (thermal and electrical) of the controlled systems. This goal can be achieved through a

retrofit intervention that aims to increase equipment efficiency or implementing advanced control strategies able to enhance how energy is used [122].

- **Cost reduction:** aims at minimizing the operating cost of the controlled energy systems. Despite being strongly related to energy consumption, costs are mainly influenced by energy price schedules and by the presence of flexible sources, RES production and the participation in a flexibility market, which allows a more efficient shift of the demand to low-price periods [123].
- **Peak reduction:** aims at reducing the magnitude of peak absorption from the electrical grid. This objective is particularly important as the scale of analysis increases, being useful to both single homes and grid operators. Furthermore, many commercial buildings also employ tariffs dependent on both energy consumption and maximum peak absorption. Lastly, this objective can also be relevant in Demand Response (DR) scenarios [124] that often aim to avoid undesirable peaks of demand [125].
- **Peak-to-average ratio (PAR):** as the name describes, peak-to-average ratio is the ratio between the maximum peak absorption and the average power absorption. A high peak-to-average ratio leads to the installation of new energy systems that often do not operate, resulting in inefficiencies [126].
- **Grid stability:** such as voltage control or frequency regulation have assumed a crucial role in RESs integration, especially with the large adoption of DERs, which can jeopardize grid stability [127]. Despite the large scale of the problem involved, multiple energy systems as HVAC, EVs and storage, if aggregated, can help with grid stability.
- **Comfort maximization:** this is a primary goal for HVAC systems and appliance. Especially in residential buildings, the satisfaction of occupant comfort along with appliances use is responsible for 80% of energy consumption in buildings [128]. Indeed, maintaining comfort is a key-aspect to ensure morale, working efficiency and productivity of the occupants [129] highlighted by the fact that electricity is a resource whose value for consumers is much higher than its price [130]. In particular, HVAC systems are responsible for thermal comfort and Indoor Air Quality (IAQ). Thermal comfort is challenging to be evaluated, especially in physical implementations and most application relies

on indoor air temperature measurements to evaluate thermal comfort and CO₂ measurements to evaluate IAQ.

2.1.4 Level of detail

As explained above, the scale of analysis has a direct influence on the level of detail used in the analysis. Especially for multiple buildings, most of the works are simulated in several different ways. For this purpose, researchers developed different simulation environments that employ surrogate models to represent building thermal dynamics [131], to fully characterize the energy management problem in buildings. A complete description of the different modeling paradigms can be found in [129]. The surrogate models can be broadly classified into three main approaches:

- **White-box models:** also known as physics-based models or engineering models exploit physical knowledge to describe systems dynamics [132]. In the framework of buildings, they use equations based on the principles of heat transfer, energy and mass conservation. One of the strengths of this method is its physical significance, which use parameters that can be retrieved from technical documentation of real systems, standards, or guidelines. If properly tuned, they are capable of correctly emulating the physical properties and dynamics of the building system, providing great advantages for the application of advanced control strategies. However, despite being physics-based, these models suffer from inaccuracies to the large number of parameters required for their definition, as well as needing a lot of time and effort to define the building features, which currently represent the major barrier for the application at scale of these kinds of models [133]. A compromise between model accuracy and complexity is represented by simplified solutions such as first or second order models.
- **Grey-box models:** The gray-box category encompasses a wide range of models that include simplified physical relationships but also necessitate parameter estimation using measurable data. In most gray-box models, the physics is reduced through state space dimensionality reduction or linearization. A typical concept in gray-box modeling is the RC analogy that defines any model by its affinity with a resistor-capacitor electrical circuit [7]. Theoretically, gray-box

models can overcome the limitations of both physics-based and purely data-driven approaches. Since part of the knowledge regarding the physics of the system is already present in the model structure, gray-box are more likely to perform correctly outside the calibration range [134]. Moreover, they require less information than white-box models to be developed. In practice, the main drawbacks are related to the necessity of a robust parameter identification method.

- **Black-box models:** Black-box models or data-driven models learn the building dynamics directly from the measured data, without making any prior assumptions regarding any physical relationships [131]. The main advantages of the black-box models are the lower development cost and the flexibility of any measured signal as an input or output, due to the absence of physics involved. On the other hand, these approaches require extensive datasets with enough information to capture the building dynamics. The training datasets need to be large and rich, so it has to cover all possible operational conditions as well as weather conditions [135], since these kinds of models are not reliable outside the training range.

Modeling is one of the main barriers to the implementation of advanced controllers in multiple buildings. The introduced techniques (white-box, gray-box and black-box modeling) are three different paradigms used in this field. The choice of which paradigm use is mainly influenced by the data availability, the scale of analysis, and the kind of error that can be accepted in such simulations. Indeed, Figure 2.6 shows the main properties of these paradigms, respectively: accuracy, transferability, data independency, smoothness (required by some optimization solvers), and reliability (generalization capabilities). As the image shows, black-box techniques are dependent on the amount of data and their transferability still needs to be proven. These two characteristics will be studied in the following chapters with specific analysis aimed at increasing black-box modeling generalizability and transferability. Very often, if technical documentation and physics-based modeling expertise are available it is preferable to use a white-box approach, due to its reliability and interpretability, however, its application at large scale is time consuming. On the other hand, if a robust dataset is available, a black-box approach is a valid alternative to obtain accurate results in a short amount of time. Lastly, if both technical information

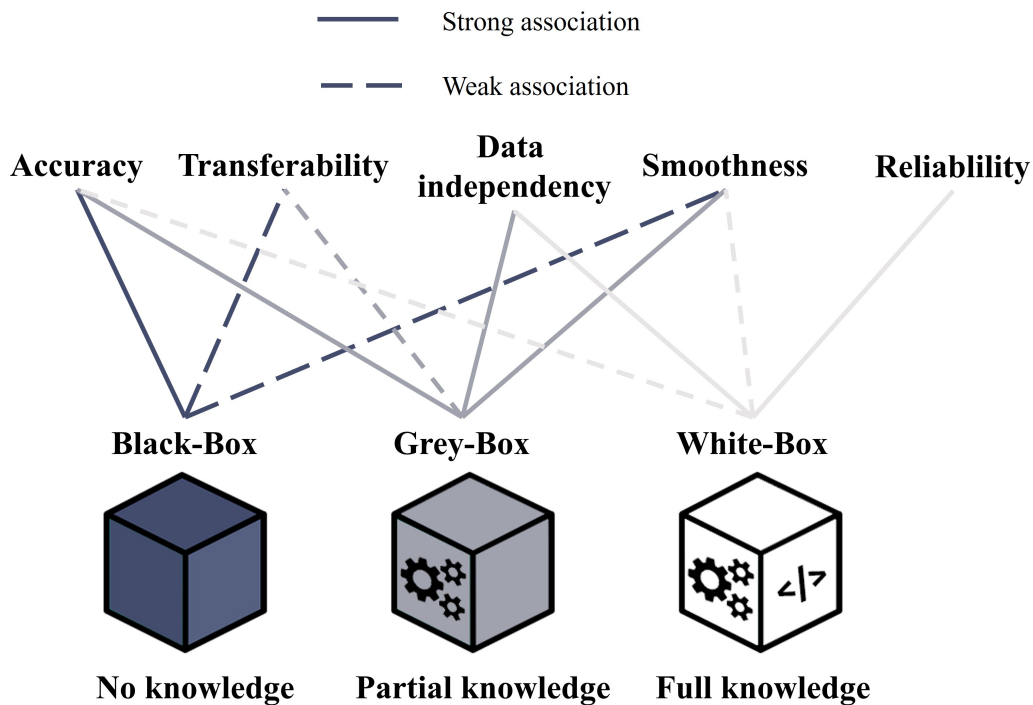


Fig. 2.6 Summary of the three modeling paradigms features (based on [7])

and data are available, a gray-box approach can represent a valid alternative, despite its application at large scale is still time consuming.

2.1.5 Discussion of the literature review

Advanced control strategies represent a powerful opportunity to decarbonize the building sector. Furthermore, the applications introduced and discussed in the previous sections highlighted even greater benefits when the scale of analysis is shifted towards multiple buildings and optimal multi-agent architecture is employed. For these reasons, energy management in multiple buildings has recently received a lot of interest. Among the most used applications there are the participation in demand response programs [136] using time-of-use tariffs to efficiently charge electric vehicles [137] or schedule appliances [80, 138], or incentive-based programs [139] and the possibility to exchange energy in multiple buildings with peer-to-peer transactions [140]. However, despite many studies have demonstrated the effectiveness of adaptive and predictive control strategies for HVAC systems, few efforts have been devoted to the simulation of their operation for a cluster of buildings.

Indeed, most of the recent works reported in literature made use of co-simulation environments based on white-box modeling such as Modelica [141] and EnergyPlus [142], limiting their application to single buildings.

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 [143] and electrical storage [144]. In those cases, the control strategies act on HVAC system or storage operations to meet ideal building energy demands that are pre-calculated by considering a fixed schedule of indoor set-point conditions. By adopting this modeling approach, storage control strategies have shown to be effective in providing grid services at both single buildings [145] and multiple buildings scale [146, 147] or in improving energy management [148].

However, this approach strongly limits the amount of flexibility that can be leveraged by control strategies, excluding building thermal mass and indoor temperature set-points. Tang and Wang [149], Robillart et al. [150] analyzed the trade-off between thermal demand reduction and acceptable indoor temperature, while Wang et al. [55] extended this concept to multiple buildings, demonstrating the effectiveness of indoor set point temperature as a source of flexibility. Recently, Perfumo et al. [59], Gonzato et al. [151] assessed the advantages of implementing MPC for regulating HVAC systems and controlling the indoor temperature in small groups of buildings. 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.

As explained in Section 2.1.4, a recent approach takes advantage of the implementation of data-driven models, due to the increasing availability of building-related data and the necessity of computationally lightweight models of indoor environmental conditions. Additionally, recent studies tried to develop more efficient model-free controllers using reinforcement learning. RL is less expensive to be implemented because it does not require a model of the system and could learn through interaction with both the environment and historical data. Moreover, a peculiarity of the RL lies in its adaptability [152] 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 boilers [153], electric water heaters [154], domestic hot water (DHW) [155] or heat pumps [156]. Vazquez-Canteli et al. [157] deeply reviewed the application of RL for demand response, emphasizing the opportunity provided by such a control approach. Wang and Hong [29] provided a detailed breakdown of the existing RL studies, identifying the main barriers of RL controller applications in actual buildings, namely the time consuming training process, the robustness and the generalization capabilities.

Recently, just a few studies have started to emphasize the application of reinforcement learning in multi-agent systems using cooperative and competitive coordination mechanisms [158] to account for demand peak shifting in a cluster of buildings. Qiu et al. [159] formulated the peer-to-peer trading problem between cluster of buildings as a multi-agent coordination problem and propose a novel multi-agent deep reinforcement learning (MADRL) method to address it. On the other hand, Charbonnier et al. [160] proposed a novel scalable type of multi-agent reinforcement learning-based coordination for distributed residential energy. Cooperating agents learn to control the flexibility offered by electric vehicles, space heating, and flexible loads in a partially observable stochastic environment.

Despite the great interest aroused by techniques based on RL, their use in the energy and buildings field is still in its infancy, especially at scale, with limited adoption in physical case studies. The dissertation seeks to overcome current literature limitations through the development of three innovative DRL applications considering different HVAC systems, control objectives, architectures, and levels of detail. Figure 2.7 shows a Venn diagram that aims to underline the different contributions provided by some relevant papers presented in the literature in the field of building energy management and highlight the contributions provided by the thesis for energy management at district scale. The diagram shows that most of the previous works focused on 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.

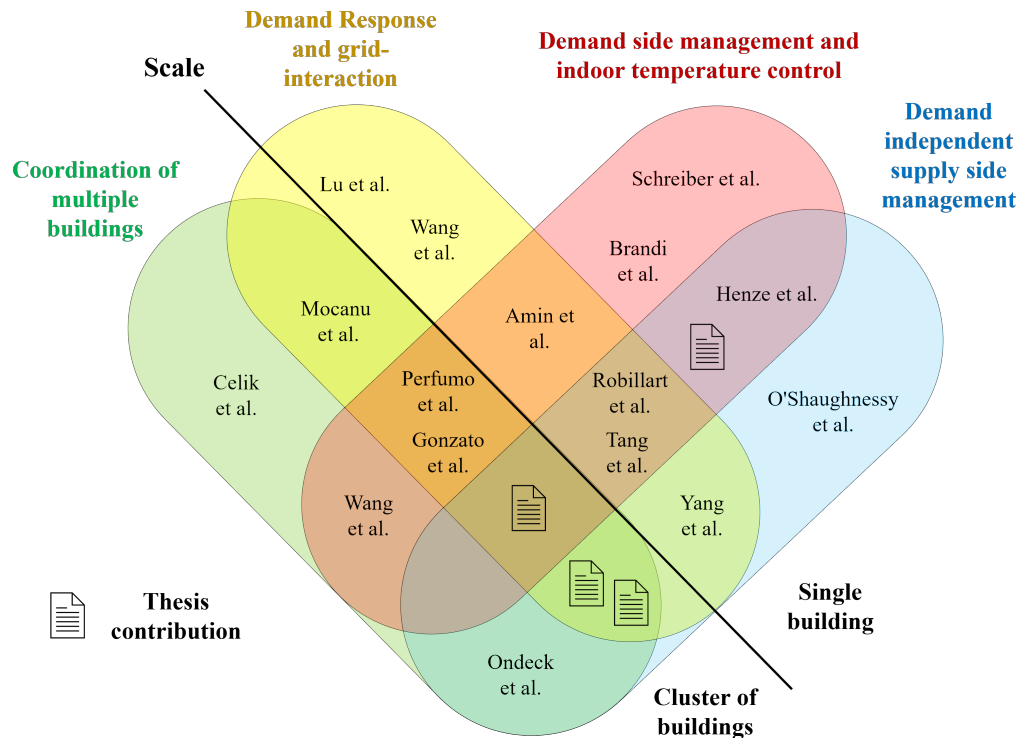


Fig. 2.7 Venn diagram displaying the four pillars of advanced control for district energy management: coordination of multiple buildings, grid-interaction, indoor comfort and management of supply technologies (based on [36])

2.2 Reinforcement learning

Reinforcement Learning (RL) is a branch of machine learning conceived to solve control problems and sequential decision-making processes [32]. RL uses an agent-based control, where the agent learns through the interaction with the controlled environment. RL can be formalized as Markov Decision Process (MDP), a discrete-time stochastic control process. MDP provides a mathematical framework for modeling decision making in situations where outcomes are partly random and partly under the control of an agent. Markov Decision Process is represented using a 4-tuple (S, A, P, R) made up of:

- **State (S):** The state describes the environment completely. The state term is used to represent, while the information seen by the agent, that is a mathematical description of the environment, relevant and informative to the decision to be made is called observations. Often, the agent can see only a part of

the state, dealing with the so-called Partially Observable Markov Decision Process (POMDP). In the context of energy management in buildings, typical examples of state variables are the outdoor air temperature and the electricity price.

- **Action (A):** The action is the decision performed by the control agent. In the context of energy management in buildings, the action could be represented by the charging/discharging of storage devices.
- **Transition Probabilities (P):** 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 time t , the next state will be s' . MDP satisfies the Markov Property, which states the memoryless of the stochastic process, represented as $P(s_{t+1} = s' | s_t) = P(s_{t+1} = s' | s_1, s_1, \dots, s_t)$. In the context of energy management in buildings the transition probabilities are generally unknown since this process will require the development of a detailed model of the controlled environment.
- **Reward (R):** The reward is the feedback received by the control agent for taking a specific action a_t in certain state s_t , mapping the tuple $S \times A \times S'$. The reward is evaluated through a function that depends on the control objectives of the specific control problem. In the context of energy management in buildings, the reward could be represented by a combination between energy consumption and costs.

Figure 2.8 shows a schematic representation of the RL framework. The four elements of the MDP are depicted in the figure as interactions between the control agent and the controlled environment. The control agent observes the current state of the environment (S) and selects a control action at each control timestep (A). The control action causes a change in the controlled environment, leading to a new state configuration (S') according to the transition probability (P). The reward (R), with the role of quantifying the goodness of the changes in the environment, is successively forwarded to the control agent, along with information about the new state of the environment.

In the reinforcement learning framework, the control agent directly learns the optimal control policy (π) by interacting with the controlled environment through a trial-and-error approach. The policy represents a mapping between states and actions

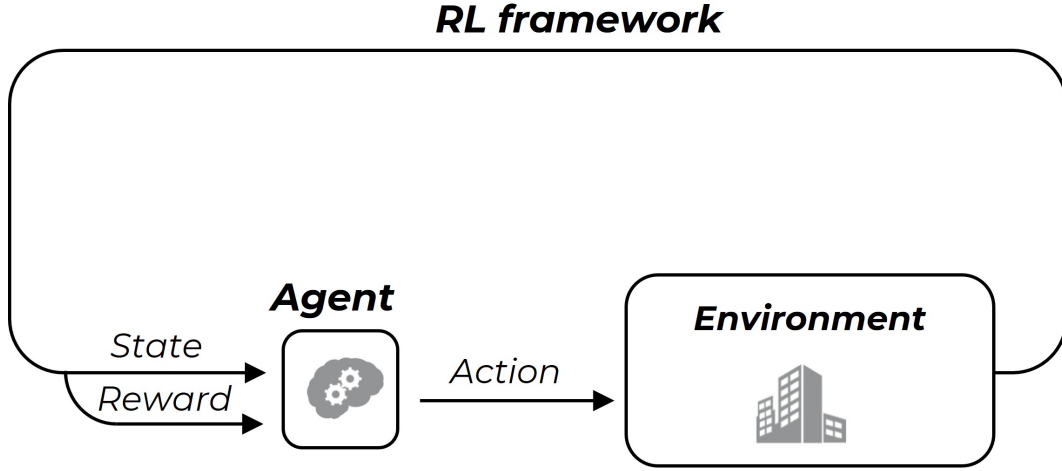


Fig. 2.8 Schematic representation of the RL framework

$\pi : S \rightarrow A$ and is the core of the reinforcement learning agent. The goal of the agent is to find an optimal control policy π^* , a policy that aim to maximize the cumulative reward over a time horizon. This concept is summarised by introducing the expected return G , which represent the cumulative sum of the reward $G = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}$, 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 assigns higher values to states that lead to high immediate rewards. The optimal control policy can be found employing two closely related value functions, namely the state-value function $v_{\pi}(s)$ and state-action value function $q_{\pi}(s, a)$, used to show the expected return of a control policy π at a state or a *state, action* tuple, as follows:

$$\begin{aligned} v_{\pi}(s) &= \sum_a \pi(a|s) \sum_{s', r} p(s', r|s, a) [r + \gamma v_{\pi}(s')] \\ &= \mathbf{E}[r_{t+1} + \gamma v_{\pi}(s') | S_t = s, S_{t+1} = s'] \end{aligned} \quad (2.1)$$

$$q_{\pi}(s, a) = \mathbf{E}[r_{t+1} + \gamma q_{\pi}(s', a') | S_t = s, A_t = a] \quad (2.2)$$

These functions represent, respectively, the goodness of being in a certain state S_t with respect to the control objectives [161] and the goodness of taking a certain action A_t in a certain state S_t following a specific control policy π using the concept

of expected return. If the transition probabilities p and the rewards r are known, the solution of the state-value function, equation (2.1) can be found through direct approaches, retrieving the optimal policy using dynamic programming algorithms or with direct approaches such as policy or value iteration [162].

However, often the transition probabilities and the rewards are not known, and to estimate the optimal policy the agent needs to interact with the environment, observing its responses. Based on the information availability (rewards and transition probability), reinforcement learning can be categorized as model-based and model-free RL. In this context, a model-based algorithm uses the transition probability function to estimate the optimal policy. The main difference with dynamic programming is that the agent might have access only to an approximation of the transition function, which can be learned by the agent while it interacts with the environment. In general, in a model-based algorithm, the agent can potentially predict the dynamics of the environment using estimates of the transition probability, however these transition probabilities may be approximations, leading to sub-optimal solutions. A model-free algorithm estimates the optimal policy without using or estimating the dynamics of the environment. In practice, a model-free algorithm either estimates a value function (equation 2.1) or the policy directly from the interaction with the environment.

Model-free approaches, which represent a large majority in the energy management field, can be further divided into value-based and policy-based methods. Value-based methods aim at learning the value function, which estimates the goodness of taking a specific action a starting, from state s . On the other hand, policy-based methods do not employ the value function as a proxy and directly try to learn the optimal control policy π [163]. Each of the two methods has its advantages; value-based methods are more sample efficient, while policy-based methods have better convergence properties and are capable to handle continuous problems characterized by high stochasticity. Additionally, RL algorithms can be also characterized by the approach used to update the optimal policy, on-policy and off-policy methods. On-policy RL algorithms directly try to improve the policy that is used by the agent to generate decisions, updating the policy based on estimates of the optimal policy. Off-policy methods evaluate a policy that is different from the one used to select actions, allowing them to learn from historical data and previous experience. Looking at the energy management problem, on-policy training is particularly challenging, due to the necessity of the agent during the training phase to explore sub-optimal solutions

that may compromise user comfort conditions. On the other hand, the application of on-policy learning ensures better state-action space exploration, converging to the optimal solutions faster than off-policy algorithms.

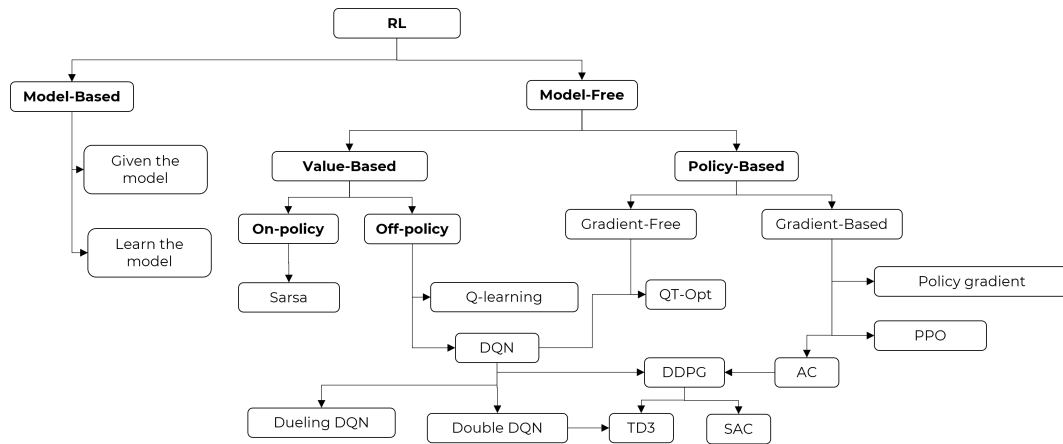


Fig. 2.9 Non-exhaustive taxonomy of RL

Figure 2.9 displays a non-exhaustive taxonomy of RL, highlighting in bold the various classification previously cited, together with the most famous algorithms, some of which will be further described in the next subsections. As can be seen, some algorithms can be both value-based and policy-based; this kind of algorithms, called actor-critic try to combines the benefits of the two techniques.

2.2.1 Multi-agent reinforcement learning

Despite MDPs having proven to be effective when dealing with optimal decision-making in single-agent stochastic environments, multiagent environments require a different representation. In multiagent RL, a set of autonomous agents interact within the environment to learn how to achieve their objectives. However, the state dynamics and the expected rewards change according to the effect of joint actions, violating the Markov property. In a multi-agent setting, the problem representation depends on both the agent interaction (cooperative or competitive) and whether the agents take actions sequentially or simultaneously.

Figure 2.10 shows an overview of the theoretical frameworks used in MARL problems. Depending on the interaction between the agent, the problem can be viewed as Decentralized partially observable Markov decision process, or as a Partially observable Markov game, in which the agents collaborate to maximize

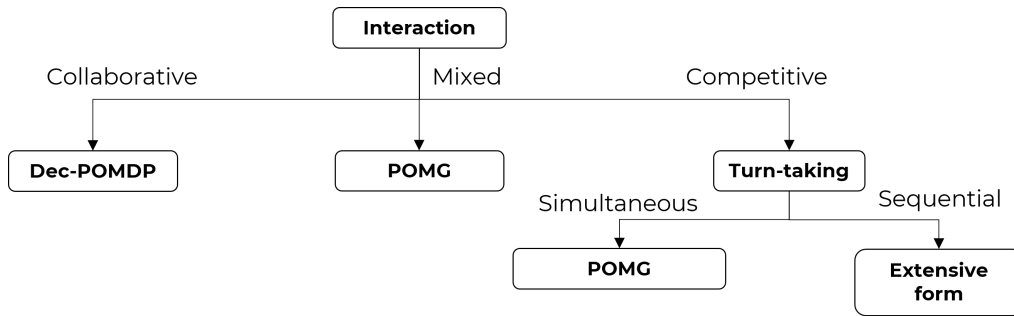


Fig. 2.10 Multi-agent RL problem classification

a common reward. Additionally, if the agents take turns sequentially rather than simultaneously the problem can be viewed as an extensive form of Markov game.

2.2.1.1 Markov game

Markov games [164] provide a theoretical framework to study multiple interacting agents in a fully observable environment and can be applied to cooperative, collaborative and mixed settings. A Markov game is a collection of normal-form games (or matrix games) that the agents play repeatedly. Each state of the game can be viewed as a matrix representation with the payoffs for each joint action determined by the matrices. In its general form, a Markov game is a tuple (I, S, A, R, T) where I is the joint set of N agents, S is a finite state space, $A = A_1 \times A_2 \times \dots \times A_N$ is the joint action space of N agents, $R = (r_1, r_2, \dots, r_N)$ where $R_i : S \times A \rightarrow \mathbb{R}$ is each agent's reward function, $T : S \times A \times S \rightarrow [0, 1]$ is the transition function. In a team Markov game, agents work together to achieve a goal and share the rewards function $r_1 = r_2 = \dots = R_N$. In a Markov game all agents take their actions simultaneously.

2.2.1.2 Decentralized partially observable markov decision process

In a decentralized partially observable Markov decision process (Dec-POMDP) all agents attempt to maximize the joint reward function, while having different individual objectives. A Dec-POMDP consists of a state space S , the transition probabilities $P(s'|s, a_1, \dots, a_N)$ and expected rewards $R(s; a_1, \dots, a_N)$. Σ_i is a finite set of observations for agent i , and $O(o_1, \dots, o_N | a_1, \dots, a_N, s')$ are observed by agents $1, \dots, N$, respectively, given that each action tuple (a_1, \dots, a_N) was taken and let to state s' . At every time step, each agent takes an action, receives a local observation

that is correlated with the state and a joint immediate reward. A local policy is a mapping from local histories of observations to actions, and a joint policy is a tuple of local policies. Dec-POMDPs are very hard to solve and searching directly for an optimal solution in the policy space is intractable [165]. One approach is to transform the Dec-POMDP into a simpler model and solve it with planning algorithms, for instance, using a centralised controller that receives all agents' private information and converts the model into a POMDP.

2.2.1.3 Partially observable markov game

The partially observable Markov game (POMG) [166], is the counterpart of the Dec-POMDP, in which agents optimise their individual reward functions instead of a joint reward function, within a partially observable environment. The POMG implicitly models a distribution over other agents' belief states. Formally, a POMG is a tuple (I, S, A, O, P, R) where I is the set of N agents, S is the set of states, A_i is the action set of agent i and $A = A_1 \times A_2 \times \dots \times A_N$ is the joint action set, O_i is a set of observations for agent i and $O = O_1 \times O_2 \times \dots \times O_N$ is the joint observation set. P is a set of state transitions and observation probabilities, where $P(s', o|s, a)$ is the probability of moving into state s' and joint observation o when taking joint action a in state s . $R_i : S \times A \rightarrow \mathbb{R}$ is the reward function for agent i where S refers to the joint state (s_1, \dots, s_N) and A refers to the joint actions (a_1, \dots, a_N) . The model can be reduced to a POMDP when $I = 1$. In this framework, dynamic programming algorithms are not suitable for high-dimensional problems, further complicated by the adoption of competing goals, nonstationarities and incomplete information.

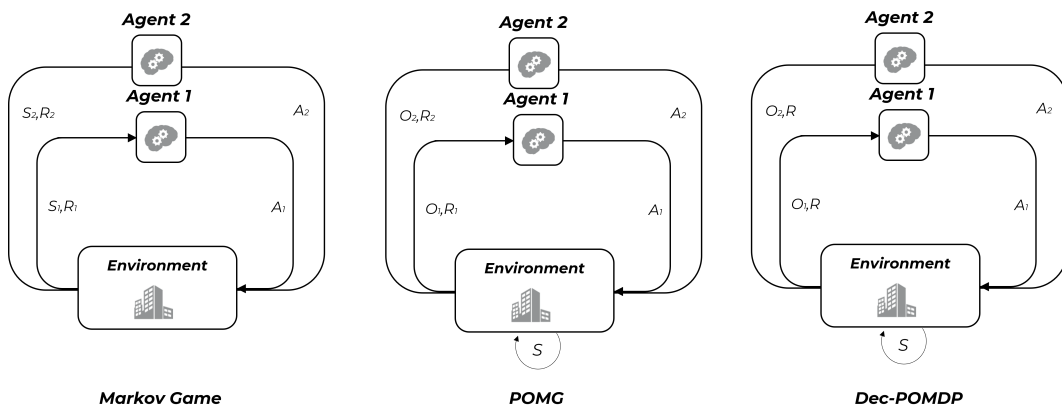


Fig. 2.11 Multi-agent RL problem representation

Figure 2.11 displays the three ways in which a multi-agent problem can be represented, as a function of how the reward is shared and the information that the agent can use to update its policy, either full knowledge of the environment (state) or partial observations.

2.2.1.4 Challenges

The application of multi-agent reinforcement learning faces several challenges. In particular, the 4 main challenges have been identified and will be further discussed:

1. **Computational Complexity:** Training a deep RL model for a single agent already requires substantial computational resources and when the problem is shifted to multi-agent, the state-action space increases exponentially with every agent, suffering from the curse of dimensionality.
2. **Non-stationarity:** In a multiagent environment, all agents learn, interact and change the environment concurrently. Consequently, the state transitions and rewards are no longer stationary, violating the Markov property and leading to agents that keep adapting to a changing environment, undermining the convergence guarantee of the algorithms. Recent works have addressed non-stationarity differently, focusing on various variables: such as the setting (cooperative, competitive or mixed), the training process of agents, the availability of other agents' information, and whether the execution of actions is centralised or decentralized, and among the way to address non-stationarity exchange information between different agents seems a promising alternative.
3. **Partial Observability:** In a partially observable environment, agents do not have access to the global state and have to make decisions based on local observations, which can lead to suboptimal solutions. Furthermore, the effect of other agents on the environment is difficult to attribute to specific actions, since an agent only know its action. Partial observability has been mainly studied in the setting where a group of agents maximise a team reward via a joint policy (e.g. in the Dec-POMDP setting). The two main approaches are the centralised training and decentralised execution paradigm and using communication to exchange information about the environment.

4. **Credit Assignment:** There are two main problems related to credit assignment (reward function). The first one is that, as previously cited, for an agent is hard to determine its contribution to the joint reward, due to the effect of other agents. Moreover, the second problem is related to the formulation of the reward function to promote collaborative behavior, finding optimal policies for each agent.

2.2.1.5 Approaches

To solve the challenges associated with MARL, several methods have been proposed. The three most common training schemes are following described and shown in Figure 2.12.

1. **Centralised controller:** the most simple multi-agent training scheme is to train multiple agents reducing the problem to single agent with a centralized controller. The agents send their observations to a central controller that decides which action to take for each agent, with a single policy. The method mitigates partial observability problem but suffer from curse of dimensionality.
2. **Decentralised controller:** the counterpart of the centralized controller includes multiple independent controllers that aim to optimize their objective function not coordinating explicitly or sharing information, ignoring the non-stationarity of the environment, thus with the possibility of converging to a suboptimal solution.
3. **Distributed controller:** A third approach tries to combine the best of both worlds. A distributed controller shares only a certain amount of information between the controllers, allowing the evaluation of multiple policies with a larger observation space. However, such information may be sensitive and this does not fully solve the non-stationarity problem, since the problem still violates the Markov property. There is also the possibility to combine both centralised and decentralised processing in a distributed setting. Information could be shared during the training phase to stabilize the learning environments and executed locally during the deployment phase.

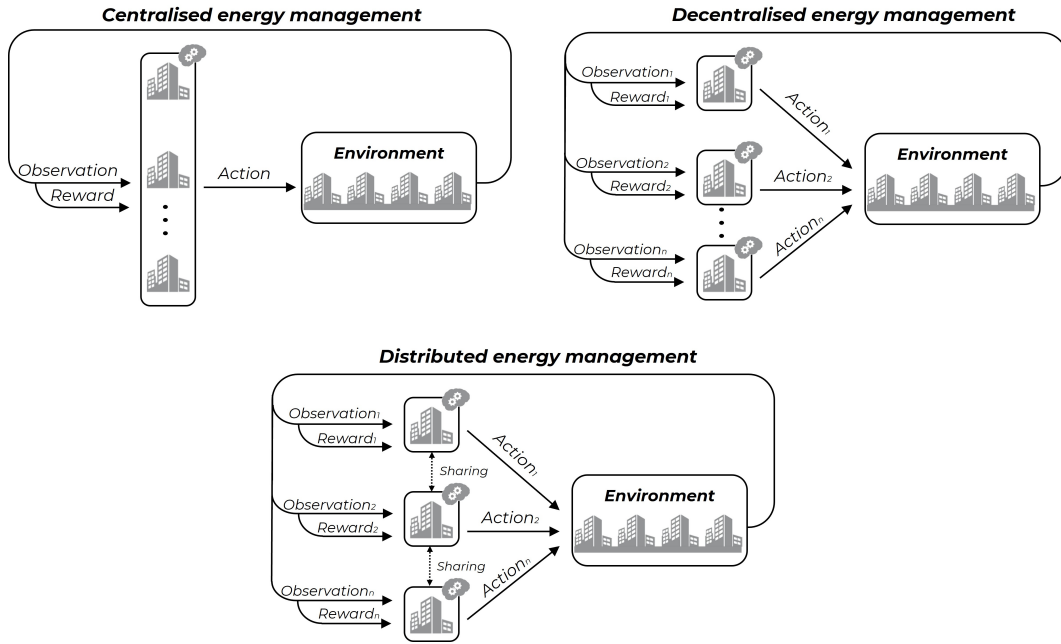


Fig. 2.12 Multi-agent RL control architectures

2.2.2 From reinforcement learning to deep reinforcement learning

After having discussed the theoretical background of RL and its multi-agent version, this subsection describes the most popular algorithms and the necessity to shift from reinforcement learning to deep reinforcement learning, with a detailed description of the algorithm used within the thesis. Among the most popular algorithms, Q-learning [167] arise for its simplicity and quickly became one of the widest algorithms adopted. According to the classification adopted in the previous section, Q-learning is a value-based off-policy method, that estimates state-action values (also called Q-values) to maximize the expected return. Q-learning stores the Q-values into a tabular structure, this approach is called tabular Q-learning. The Q-values are updated during the training according to the following equation:

$$Q_{s,a} \leftarrow Q_{s,a} + \mu [r(s,a) + \gamma \max_a Q(s',a) - Q(s,a)] \quad (2.3)$$

Where $\mu \in [0, 1]$ 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. The

agent observes the current state s of the environment at each interaction and chooses an action a based on the Q-values stored in the table relative to the same state s . This action is forwarded to the environment, which changes states to s' and sends this information along with the reward signal r to the agent, updating the relative Q-values according to equation 2.3.

A key-aspect to consider is the strategy employed to select the action. In this context, the maximization of the Q-values relies on the identification of the optimal trade-off between exploration and exploitation. In particular, the agent should select actions that proved to be associated with high rewards (exploitation) but should also select new actions to explore the Q-values of other state-action tuples. The simplest method to balance exploration and exploitation is the ϵ -greedy method. In the ϵ -greedy the agent acts greedy most of the time, using the actions with the highest Q-values. Then, the agent explores random actions with a probability of ϵ , which value is usually small [168]. The Q-learning algorithm, in its most basic form, uses lookup tables to store and retrieve state-action values, with each entry representing a state-action tuple (s,a) . In practical problems with large state and action spaces, however, using a tabular representation may be impossible. A solution to this problem is to use a function approximator to represent Q-values. This allows state-action values to be represented using only a fixed amount of memory that is determined solely by the function used to approximate the problem. The combination of RL and high-capacity function approximators such as Deep Neural Networks (DNN) demonstrated to overcome computational problems renewing the interest for the RL topic and promoting its extension to complex problems. [169] created the first work that combined Q-learning and DNNs. The Q-value function in Deep-Q-Networks (DQN) is parametrized by θ which represent the weights of the network. The input layer has a number of neurons equal to the number of states, while the output layer has many neurons as the number of actions that the agent takes at each control timestep. The neural network is used to learn the relation between states and the Q-values for each action. The true Q-value for each state-action pair is not known a priori in the RL paradigm, but it is learned through repeated interactions with the controlled environment, updated according to equation 2.3. To improve the efficiency of the DQN, a replay memory was introduced, storing previous experience obtained by the agent. In the optimization process of the network weights a random mini batch is extracted from the replay memory and used to fit DNN-regression using as targets Q-values.

2.2.3 Soft-actor critic

After DQN, a lot of DRL algorithms were proposed, each one with its advantages and disadvantages. The algorithms can deal with continuous or discrete actions and their effectiveness is influenced by the control problem. In the last years, among different DRL algorithms, actor-critic methods arise for their ability to combine advantages of both value-based and policy-based methods, introduced in the previous chapter. The main idea behind actor-critic is to split the problem using two deep neural networks. The actor maps the current state to the action that it estimates to be optimal (policy-based), while the critic evaluates the actions by computing the value function (value-based). One of the main advantages of actor-critic methods is that they can learn stochastic policies through a direct approach which represents an important advantage for stochastic processes [170]. Amidst actor-critic methods, [171] proposed the Soft Actor-Critic (SAC) algorithm an off-policy maximum entropy actor-critic algorithm, which showed excellent performance in solving several continuous control tasks. The soft actor-critic is based on three key pillars:

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

SAC learns three different functions: (i) the actor (mapped through the policy function with parameters φ), (ii) the critic (mapped with the soft Q-function with parameters θ) and (iii) the value function V , defined as:

$$\begin{aligned}
 V(s_t) &= \mathbf{E}_{a_t \sim \pi} [Q(s_t, a_t) - \alpha \log \pi(a_t | s_t)] \\
 &= \mathbf{E}_{a_t \sim \pi} [Q(s_t, a_t)] + \alpha \mathbf{E}_{a_t \sim \pi} [\log \pi(a_t | s_t)] \\
 &= \mathbf{E}_{a_t \sim \pi} [Q(s_t, a_t)] + \alpha \mathcal{H}
 \end{aligned} \tag{2.4}$$

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

$$\pi^* = \arg \max_{\pi_\varphi} \sum_{t=0}^T \mathbf{E}_{(s_t, a_t) \sim \rho_\pi} [r(s_t, a_t) + \alpha \mathcal{H}(\pi_\varphi(\cdot | s_t))] \quad (2.5)$$

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, \mathcal{H} , the agent attempts to maximize the returns while behaving as randomly as possible. The final policy used in the evaluation of the algorithm can be made deterministic by selecting the expected value of the policy as the final action. The parameters of the critic networks are updated by minimizing the expected error J_Q , which is given by:

$$J_Q(\theta) = \mathbf{E}_{(s_t, a_t) \sim D} \left[\frac{1}{2} (Q_\theta(s_t, a_t) - (r(s_t, a_t) + \gamma \mathbf{E}_{s_{t+1} \sim \rho} [V_\theta(s_{t+1})]))^2 \right] \quad (2.6)$$

where the value function is implicitly parameterized through the soft Q-function parameters in Equation 2.6. Furthermore α , called temperature parameter, determines the relative importance of the entropy term against the reward, thus controlling the stochasticity of the optimal policy. A high value of the temperature parameters may lead to uniform behavior, while a low value of the temperature parameter will only maximize the reward. For sake of clarity, the main algorithm logic is reported and the framework of the algorithm is shown in Figure 2.13

The effectiveness of any DRL algorithm is highly influenced by the hyper-parameters tuning, which represents a crucial task for the deployment of DRL controllers.

Hyperparameters can be organized according to the following classification:

- **General RL hyper-parameters:** these hyper-parameters are common to all RL frameworks. Among the general RL hyper-parameters, the discount factor (γ) is one of the most important, weighting the importance of future

Algorithm 1: SAC algorithm adapted from [171]

Input: Policy (actor) and soft-Q (critic) DNNs

Initialise target network weights

Initialise experience replay buffer with random policy samples

for *each episode* **do** **for** *each step* **do**

sample actions from policy

sample transition from the environment

store the transition in the replay buffer

end **for** *each gradient update step* **do**

update the soft-Q DNN weights

update the policy DNN weights

update the target DNN weights

end**end****Output:** Optimised actor and critic DNNs

rewards. The discount factor assumes a value included between 0 and 1, where values close to 1 gives greater importance to rewards obtained far in the future compared to the moment in which the control action is taken, while values close to 0 favor immediate rewards. This is a mathematical object introduced to prevent the cumulative sum of future rewards from going infinite, ensuring the convergence of the algorithm. Three other important hyper-parameters characterizing off-policy DRL frameworks (like SAC) are the Replay Memory Size, the Batch Size and the Number of Gradient Steps. Replay memory stores the results of previous interactions of the agents with the controlled environment. The size of this memory determines the amount of previous knowledge that can be leveraged by the algorithm to refine the control policy. Batch size regulates the number of elements drawn from Replay Memory during the learning phase. Small values of the batch size can guarantee faster convergence properties with the risk of being stuck in near-optimal solutions. Higher values of the batch size may result in slower convergence properties with the benefit of mitigating the risk of learning sub-optimal policies [172]. The number of gradient steps is a hyper-parameter that regulates the number of batches randomly drawn from memory buffer on which gradient update is performed at each control time-step. Typically, this hyper-parameter is

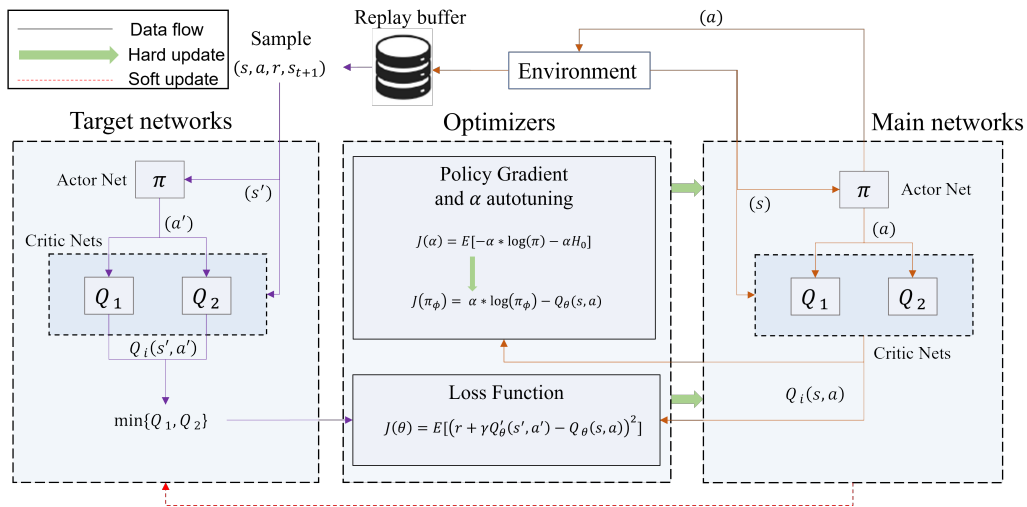


Fig. 2.13 Soft Actor-Critic architecture

set equal to 1, but sometimes this value can be increased to encourage faster learning.

- **Specific algorithm hyper-parameters:** these hyper-parameters are specific of an algorithm, characterizing their behavior and converging properties. A description of the most common hyper-parameters used for SAC is provided below:
 - **Target Model Update Frequency:** determines the frequency at which the parameters of the online network are copied into the target network.
 - **Entropy Coefficient (α):** is the temperature parameter that determines the relative importance of the entropy term versus the reward, and thus controls the stochasticity of the optimal policy [173].
- **DNN hyper-parameters:** these hyper-parameters characterize the architecture of DNN employed by DRL algorithms as function approximators. DNNs are the most widely used function approximators, thanks to their abilities to successfully mapping nonlinear relationships. However, they are also characterized by several hyper-parameters adding a further degree of complexity to the development of RL controllers:
 - **Neural Network Structure:** The most widely applied neural network architecture is the Multi-Layer Perceptron (MLP). The number of hid-

den Layers and neurons for each hidden layer are the hyper-parameters determining this architecture.

- **Activation Function:** The choice of the activation function may influence convergence of DRL algorithms. The most widely applied functions are the REctified Linear Unit (RELU) [174] and Hyperbolic Tangent (tanh).
 - **Optimizer:** The choice of the optimizer may influence convergence and performance of DRL algorithms. The most widely applied optimizers is Adam [175].
 - **Optimizer Learning Rate:** The learning rate of the optimizer implemented in DNNs is a hyper-parameter that controls the degree of change of the network in response to the estimated error each time the weights are updated. Increasing the value of the learning rate may be useful in some circumstances to speed-up the learning processes.
- **Environment hyper-parameters:** these hyper-parameters characterize the controlled environment and can strongly influence stability and convergence of implemented control agents and include:
 - **Episode Length:** The length of the episode depends on the specific control problem being studied. Usually, for energy management problems it can range from a few weeks up to a year.
 - **Number of training episodes:** The number of training episodes must be tuned to provide the agent with a sufficient amount of experience to identify the optimal control policy
 - **Reward coefficients:** As previously introduced the reward function combines in a mathematical expression the different objectives that an agent seeks to maximize (or minimize). The relative importance of these different objectives is commonly managed through the introduction of weight factors and their tuning plays a key role in the definition of a robust reward function. Furthermore, in algorithms like SAC, also the magnitude of the reward function can influence exploration and exploitation, posing even more importance in the definition of these weight factors.

Chapter 3

Scale-up energy management in buildings with data-driven controllers

This chapter discusses in detail the development of two deep reinforcement learning applications for the energy management of storage in multiple buildings. Portions of the present Chapter were already published in the following scientific papers:

- Giuseppe Pinto, Silvio Brandi, Josè Ramòn Vazquez-Canteli, Zoltán Nagy, and Alfonso Capozzoli. Towards Coordinated Energy Management in Buildings via Deep Reinforcement Learning.pdf. pages 1–14, 2020 [34]
- Giuseppe Pinto, Marco Savino Piscitelli, José Ramón Vázquez-Canteli, Zoltán Nagy, and Alfonso Capozzoli. Coordinated energy management for a cluster of buildings through deep reinforcement learning. *Energy*, 229:120725, 2021 [35]
- Giuseppe Pinto, Davide Deltetto, and Alfonso Capozzoli. Data-driven district energy management with surrogate models and deep reinforcement learning. *Applied Energy*, 304:117642, 2021 [36]
- Giuseppe Pinto, Anjukan Kathirgamanathan, Eleni Mangina, Donal P. Finn, and Alfonso Capozzoli. Enhancing energy management in grid-interactive buildings: A comparison among cooperative and coordinated architectures. *Applied Energy*, 310:118497, 2022 [37]

-
- Davide Deltetto, Davide Coraci, Giuseppe Pinto, Marco Savino Piscitelli, and Alfonso Capozzoli. Exploring the Potentialities of Deep Reinforcement Learning for Incentive-Based Demand Response in a Cluster of Small Commercial Buildings. *Energies*, 14(10), 2021 [38]

Advanced control strategies are enabling the participation of buildings in the flexibility market, allowing building owners to reduce their electrical bills, as well as providing grid services. These advantages can be even greater when aggregating flexibility from multiple buildings, managing neighborhoods, or small communities, as described in Section 2.1.1. Advanced controllers demonstrated their advantages in facing complex energy systems, forecast information, and variable electricity price or DR participation. In particular, looking at commercial buildings, storage represents the energy systems with the highest flexibility, due to their ability to decouple production and demand. As stated in Sections 2.1.2 and 2.1.3 the effectiveness of these control strategies in such a complex environment is influenced by the architecture chosen and the control strategy. Section 2.1.4 highlighted the benefits introduced by DRL, such as its ability to adapt to complex control problems characterized by high dimensionality and contrasting objectives, and its potential model-free nature. Despite the advantages provided by the application of DRL, many studies focused their attention on single buildings HVAC systems or EVs, needing further exploration for multiple buildings. To demonstrate the potential of DRL at scale, two applications related to the implementation of DRL controllers in multiple buildings with heterogeneous energy systems are investigated in this dissertation and shown in Figure 3.1.

The first application aims at demonstrating how DRL can be beneficial at multiple levels, for single buildings, the district, and the grid. Once the potentialities of this approach have been unveiled, the second application compares two different architectures to understand their advantages, scalability, and robustness. The novel framework proposed in this analysis leverages methodological procedures that exploit CityLearn environment, an open source OpenAI gym environment for the implementation of Multi-Agent Reinforcement Learning for building energy coordination and demand response in cities, described below. The next sections present the main research challenges analyzed in the thesis, introducing the motivations and novelties of the proposed approaches.

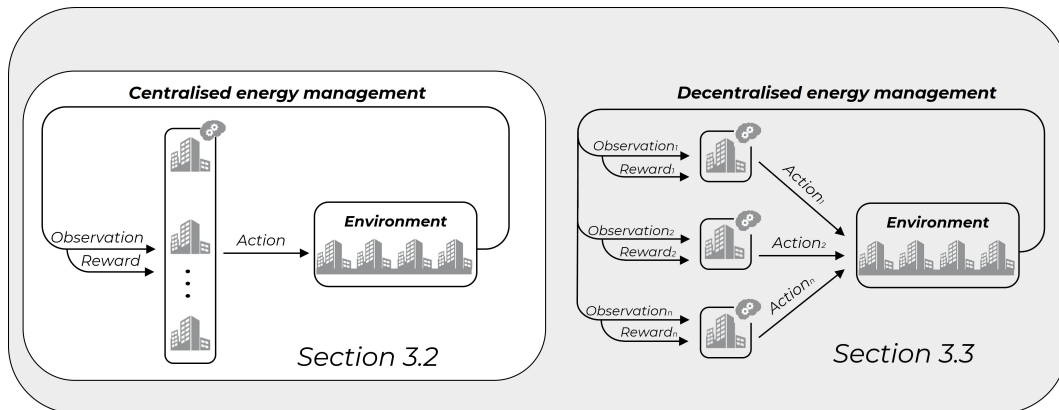


Fig. 3.1 Contribution of the dissertation on data-driven controllers for the energy management of multiple buildings

3.1 CityLearn environment

CityLearn is a framework for the implementation of multi-agent or single-agent reinforcement learning algorithms for urban energy management, load-shaping, and demand response using the OpenAI Gym standard [176], developed in Python [177]. CityLearn is a self-contained environment, since it does not require co-simulation with other buildings energy simulation programs (EnergyPlus, Modelica, TRNSYS). CityLearn overcomes the dependence on simulation environments using building hourly data from pre-simulated models and assumes that the building indoor temperature does not vary with the controlled actions. This can be achieved through the control of thermal storage, which decouples the demand and the production. Indeed, CityLearn does not control passive storage such as building thermal mass or HVACs systems. Despite the inability to leverage these additional flexibility sources, this approach guarantees that the heating and cooling energy demand of buildings are satisfied at any time, regardless of the control actions. Furthermore, not evaluating the building thermal response allows an increase of the scale of analysis, with the advantage for the controllers to focus on shaping building loads without compromising the comfort of the occupants.

Figure 3.2 shows the CityLearn framework, with the pre-simulated inputs on top, that includes building energy models, solar production, weather conditions, electricity price and forecasts. Then it also takes as inputs other files with several parameters defined by the user to characterize specific energy systems or buildings. These parameters include the building attributes and all the energy subsystems that

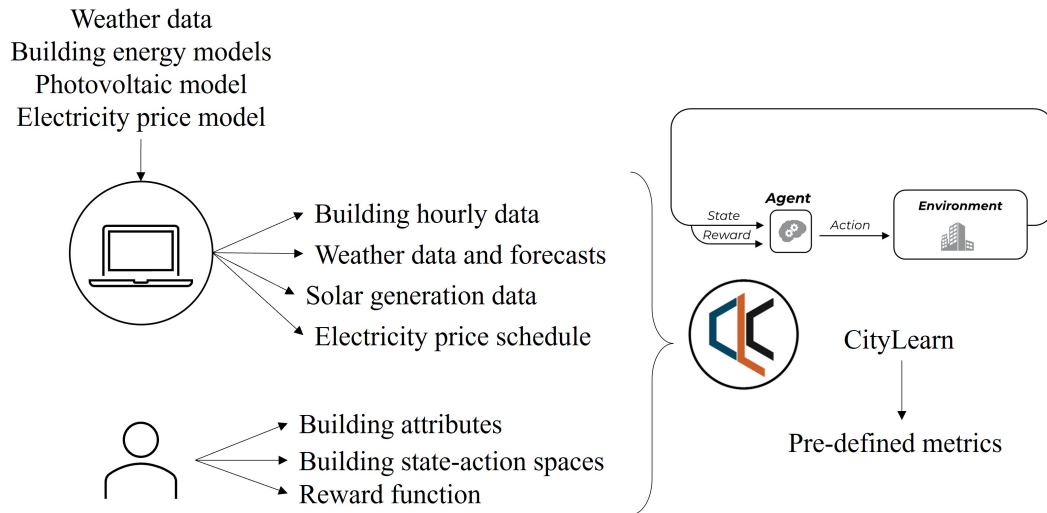


Fig. 3.2 Flowchart of the CityLearn environment

characterize the building: heat pumps, thermal energy storage tanks and electric heaters. Then, among the information in input, it includes the building state-action spaces and the reward. These parameters are specific to an RL controller and include a list of information that each building returns at each simulation time-step (state), the actions that each building can perform, depending on the energy systems in that specific building, and the reward function used by the CityLearn environment to evaluate the performance of the controllers. As previously said, CityLearn can work in either centralized or decentralized mode, with the latter being the default mode. Using the decentralized mode, CityLearn returns hierarchical lists to represent the rewards, the states and the actions. If a centralized mode is selected, then the reward will be a single value each time-step, and the states and actions can be encoded in a single list. To avoid computational complexity, common states such as outdoor temperature, solar radiation and electricity price are appended just once to the list of returned states. Therefore, the dimensions of the state-action space increase linearly with the number of buildings.

The following framework was chosen for the development of the thesis thanks to its characteristics:

- It is conceived for both single-agent and multi-agent RL implementations.
- The reward function is fully customizable and tailored according to the environment mode (centralized, decentralized).

- The environment is modular and open source. Energy systems are encoded in classes, therefore different building architectures can be implemented easily.
- User-friendly: CityLearn requires few dependencies and no co-simulation environment.

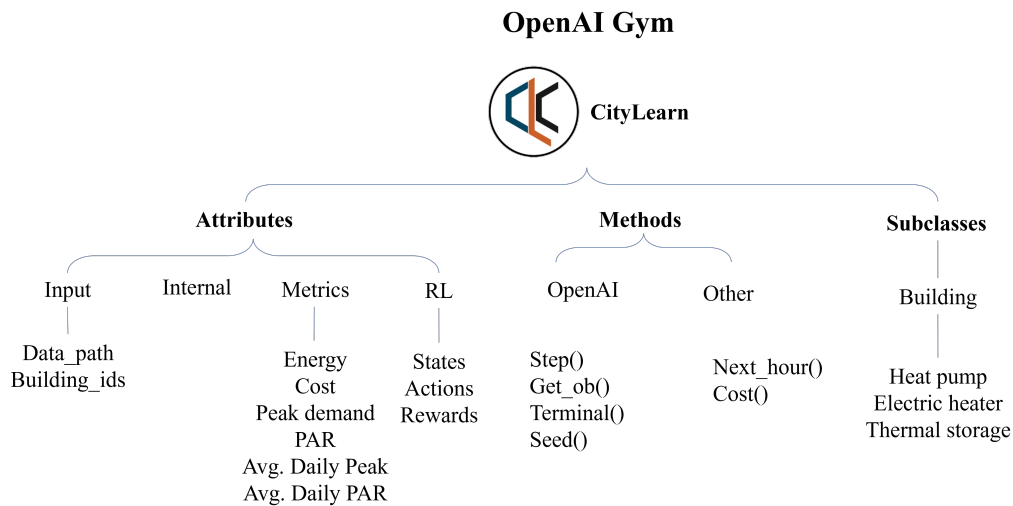


Fig. 3.3 CityLearn code architecture

Figure 3.3 shows how CityLearn inherits methods and attributes from the OpenAI Gym super class, and also contains further attributes, methods, and subclasses (energy systems). The following subsections will discuss the subclasses and metrics, while a detailed description of Attributes and Methods is provided in Appendix A.

Despite the many advantages provided by CityLearn, that made it a user-friendly environment, this also poses some limitations:

- Limited use cases: creating a dataset that has all the information required by CityLearn needs expertise with energy simulators.
- Pre-determined fixed demand: The energy loads and the schedules used in the environment are pre-computed, requiring less computational effort to run the simulations, but introducing an approximation for the definition of the models, limiting their real-world application.

These two limitations will be better explored in the next sections when the thesis will try to introduce additional test cases and surrogate models to overcome the pre-determined fixed demand.

3.1.1 Building

As previously mentioned, each building has pre-simulated thermal and electrical loads on an hourly timescale. Each building is characterized by electrical loads, cooling and heating loads. To ensure to use of diverse load profiles, preventing the buildings from behaving very similar to each other, several techniques have been used. For non-shiftable electrical loads, Pecan Street data was used [178] using the inputs of multiple households in Austin, TX, and training a probabilistic regression model to generate schedules used in EnergyPlus. A similar approach was used for the domestic hot water (DHW) profiles, when data from Solar Row project [179] was used to deal with the stochastic use of DHW. Lastly, for more realistic energy consumption profiles, different setpoints for different thermal zones of multi-family buildings were generated according to the data from the ResStock Project [180]. Each building has also specific information that allows a non-RL control loop to simulate the control loop of the HVAC system of the building. Each building uses as an input a .csv file with the following information:

- month of the year (1-12)
- Hour of day (1-24)
- Day type using the EnergyPlus day type variable (1= Sunday, 8 = Holiday)
- Daylight saving status (0,1)
- Indoor average temperature: the average indoor building temperature across all thermal zones weighted by their floor area (°C)
- average unmet cooling setpoint difference: the unmet cooling setpoint difference, defined as the difference between the thermal zone temperature and its setpoint, weighted over all the thermal zones by their floor area.
- Indoor relative humidity: Average indoor relative humidity over all thermal zones weighted by their floor area (0-100%)
- Equipment electrical power: is the electricity consumed by all electrical appliance excluding HVAC equipment (kWh)
- Heating energy for DHW: is the thermal energy consumed for the production of DHW (kWh)

- Total cooling load: is the total thermal energy demand for cooling (kWh).

As previously said, the total cooling load of the building is satisfied at each time step, either by the thermal storage tank or the energy supply unit (heat pump). To satisfy the assumption that the building temperature set-points are always met, the energy supply devices are sized using a specific safety factor. This is done using the function *building_loader()*, which reads the input csv and uses the maximum hourly thermal energy consumption and the safety factor to size the system.

3.1.2 Heat pump

Concerning the classic implementation of the heat pump implemented in CityLearn (air-to-water heat pump), its efficiency was modified to take into account partial load ratio (PLR) and the effect of external temperature not only on the coefficient of performance (COP), but also on the design capacity. A description of the relation among COP, declared capacity (DC) and the external temperature was defined according to real data sheet of heat pumps. Moreover, the heat pump operation at part load conditions was modeled according to UNI EN 14825 [181]. Eventually, COP was evaluated according to Equation 3.1.

$$COP(T, PLR) = COP_T(T_i) * \left(\frac{\frac{Q_{cooling}(action)}{DC(T_i)}}{0.9 * \frac{Q_{cooling}(action)}{DC(T_i)} + 0.1}} \right) \quad (3.1)$$

The external temperature rise has a twofold effect, firstly it reduces COP (increasing electricity consumption) and secondly it increases the maximum cooling power deliverable by the heat pump. Moreover, heat pump efficiency is influenced not only by external variables, but also by controller actions affecting the cooling load. Finally, the fraction in Equation 3.1 accounts for part load ratio and intermitting operation of the heat pump.

3.1.3 Electric heater

The electric heater provides heating energy, Q_{heater} , (i.e. for DHW) consuming electricity from the grid, E_{heater} , and following the equation: Where η_{heater} is

the user-defined heater efficiency and is usually greater than 0.9, while Q_{heater} is the algebraic sum of the heating needed from the DHW and the consequent charging/discharging of the storage.

3.1.4 Thermal storage

The current version of CityLearn considers two types of thermal energy storage: a DHW or sanitary hot water (SHW) tank and a chilled water tank. The DHW is charged using the electric heater, while the chilled water tank is charged using the heat pump. A scheme related to the energy flows from each energy system is illustrated in Figure 3.4.

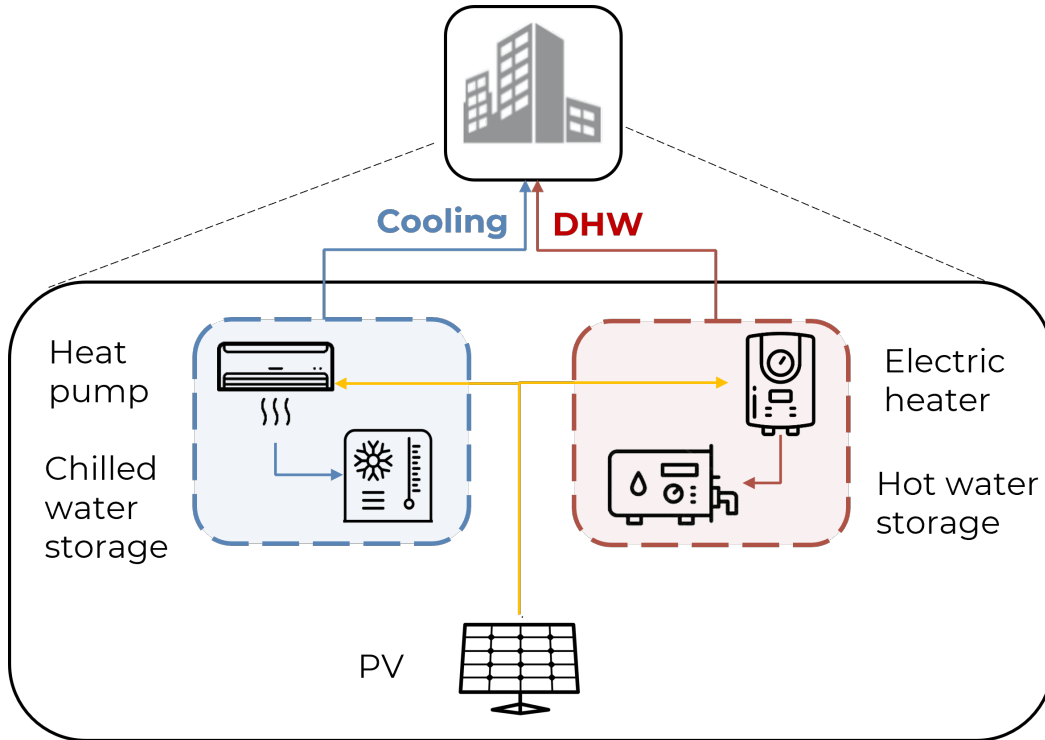


Fig. 3.4 Energy systems in the CityLearn environment with corresponding energy flows

Each storage is characterized by a state of charge, evaluated as follows:

$$SoC_{t+1} = SoC_t * (1 - e_{loss}) + Q_{in}^{sto} - Q_{out}^{sto} \quad (3.2)$$

A thermal loss coefficient e_{loss} , which represent the fraction of energy storage that is lost in each time-step, and a roundtrip efficiency, represents the efficiency of the

complete process of charging and discharging the storage. In particular, the charging and discharging processes are regulated according to the following equations:

$$Q_{out}^{sup} = \frac{Q_{in}^{sto}}{\sqrt{\eta_{eff}}} \quad (3.3)$$

$$Q_{out}^{sup} = -\frac{Q_{dem}}{\sqrt{\eta_{eff}}} \quad (3.4)$$

Where Q_{dem} is the demand associated with that specific timestep. To ensure that the energy systems can provide the right amount of energy to the building, several physical constraints and a backup controller have been added. This controller overrides the actions taken by the DRL controller if necessary: The DRL controller sends a control signal indicating how much energy it wants to charge or discharge from the thermal energy storage devices (chilled water and DHW) and the backup controller checks the energy balances in both cases. If the storage device is going to be charged, the energy supply device will prioritize the meeting of the building energy demand. If the storage device is going to be discharged, it can only do so by an amount of energy that is no greater than the energy demand of the building, to avoid thermal discomfort to the occupants.

The state of charge (SOC) is calculated every time-step in CityLearn for both cooling energy storage and DHW energy storage as follows:

$$SoC_{t+1} = SoC_t + \max_{\min} a_t * C, Q_{t_{max}} - Q_b, -Q^{dem} \quad (3.5)$$

Where:

- $C \geq 0$: represent the maximum energy storage capacity of the storage device
- $-\frac{1}{C} \leq a_t \leq \frac{1}{C}$: represents the action encoded by the controller
- $Q^{dem} \geq 0$: is the building energy demand for cooling or DHW
- $Q_t^{max} \geq Q^{dem}$: is the maximum thermal power of the heating device

In this case Q_t^{max} can change over time, since both coefficient of performance (COP) and Declared Capacity (DC) change as a function of outdoor air temperature, influencing the maximum possible thermal power output. Furthermore, the action of the controller is limited between $-\frac{1}{C}$ and $\frac{1}{C}$, since the storage capacity C is defined

as a multiple of the maximum thermal energy consumption by the building for the worst case scenario. If the controller decides to use actions beyond the limits, the action will be overridden by the backup controller.

3.1.5 Solar photovoltaic panels

CityLearn uses pre-simulated data of photovoltaic generation per kW of installed solar PV power capacity. By doing so, the user can define the solar capacity installed in each building, that will be used by the environment to obtain the final value of solar production. Solar panels are particularly useful to offset part of the electricity consumption of the buildings, since they have equipment and appliance consumption and are equipped with an air-to-water heat pump and electric heaters.

3.1.6 Key performance indicators

The environment defines several key performance indicators (KPI) to quantify the goodness of a controller. In particular, the modified version of the environment uses seven metrics, described below and quantified in Table 3.1.

- Energy consumption: energy consumed within the episode.
- Electricity cost: cost expenditure for the amount of energy imported by the grid.
- Peak: maximum amount of electricity withdrawn from the grid during the episode.
- Average daily Peak: average maximum amount of electricity withdrawn from the grid per day.
- Peak-to-average ratio (PAR) : the ratio between the maximum and the average demand during the entire episode.
- Average daily PAR: average PAR evaluated over a period of 1 day.
- Flexibility Factor: the amount of energy used during on-peak period over the total amount of energy.

Table 3.1 KPIs used in CityLearn

| KPI | Formula | Units |
|-----------------------------|---|-------|
| Cost | $\sum_i^n e_i * c_i$ | [\$] |
| Peak | $\max \sum_i^n \frac{e_i}{\Delta t}$ | [kW] |
| Daily-Peak | $\frac{\sum_i^{n_{day}} Peak_{day}}{n_{day}}$ | [kW] |
| Peak-to-average ratio (PAR) | $\frac{Peak}{\sum_i^n e_i / n_{day}}$ | [-] |
| Daily-PAR | $\frac{\sum_i^{n_{day}} PAR_{day}}{n_{day}}$ | [-] |
| Self-sufficiency (SF) | $\frac{\sum_i^n \sum_{j=1}^T \min(PV_{i,j}, e_{i,j})}{\sum_i^n e_i}$ | [%] |
| Flexibility factor (FF) | $\frac{\sum_i^n e_{i,off-peak}}{\sum_i^n (e_{i,off-peak} + e_{i,on-peak})}$ | [-] |

3.2 Enhancing energy management in grid-interactive buildings with deep reinforcement learning

DRL has recently gained popularity among energy management controllers, due to its ability to adapt to very complex control problems, thanks to the exploitation of deep neural networks as function approximators, able to capture non-linear relationships of the controlled environment. However, their application at scale is still in its infancy and requires further analysis.

The next section presents the main research challenges analyzed and introduces the motivations and novelty of the proposed methodological approach.

3.2.1 Motivations and novelty of the proposed approach

As highlighted in the literature review, most studies neglected the opportunity provided by a coordinated control on multiple buildings to flatten the peaks on the grid rather than shifting them and more efforts are necessary to study the application of these techniques in multiple buildings. It is not surprising that in the past years the need for multi-agent coordination in DR applications was not fully adopted, as the lack of it does not always lead to shifts in the peak demand or “rebound” effects in the daily load profiles. Indeed, in urban settings where the amount of energy storage capacity is not very high, building agents can enable DR without coordination and still be successful in reducing the peaks of electric demand. However, due to the trend towards wide adoption of electric vehicles and other storage devices such as batteries, this is subject to change shortly [182]. As energy storage devices become more abundant and electrical demand more volatile due to the presence of more renewable energy resources and EV charging stations, properly adaptively coordinating all these energy systems can be critical without a centralised control or multi-agent cooperation.

This work explores the opportunity of enhancing demand flexibility for a cluster of buildings by implementing coordinated energy management, using Deep Reinforcement Learning (DRL) to manage the thermal storage of four buildings equipped with different energy systems. The controller was designed to flatten the total load profile of the cluster while optimizing the energy consumption of each building. For benchmarking purposes, the coordinated energy management is then tested and compared against a manually optimised rule-based controller. Based on the literature review the main novelty of the work can be summarised as follows:

- The work exploits a multi-agent RL centralised controller with a strategy explicitly designed to consider the benefits at multiple levels (i.e., single building, cluster, and grid level), against a most common rule-based control strategy that optimises single buildings.
- The work makes use of a novel simulation environment, CityLearn, an OpenAI Gym environment specifically designed to allow RL implementations for the built environment, enhanced to consider a variable electricity price.
- The DRL controller used in this work exploits a state-of-the-art continuous control algorithm i.e., soft actor critic (SAC). The control performances of the

agent were deeply analysed to highlights the benefits provided by coordinated energy management.

The work is organised as follows: Section 3.2.2 describes the methodological framework at the basis of the analysis. Section 3.2.3 introduces the case study, explaining in detail the energy modelling of the system and the controllers design and training. Section 3.2.4 presents the results of the training and deployment phase, while a discussion of the results is given in Section 3.2.5.

3.2.2 Methodological framework

The section reports the methodological framework adopted in the present work, to describe each stage of the process, including the development, training and deployment of DRL control agent. The framework unfolds over four different stages, as shown in Figure 3.5

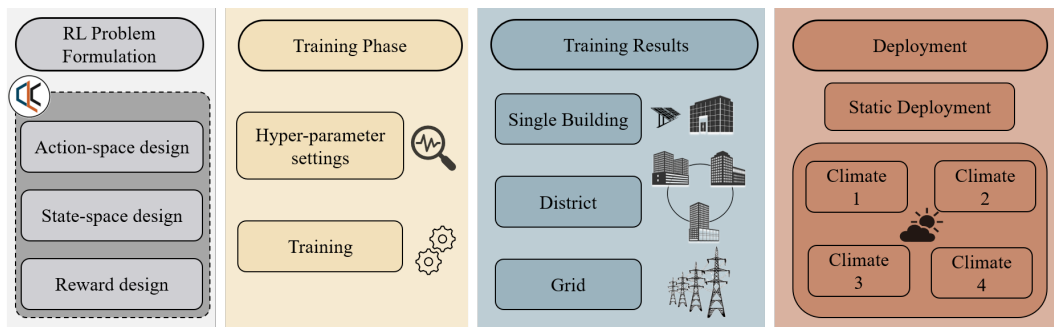


Fig. 3.5 Framework of the application of DRL control [35]

RL Problem formulation: the first stage of the framework was aimed at defining the main components of the reinforcement learning control problem. The action-space includes all the possible control actions that can be taken by the control agent. Considering that the work aims is to coordinate multiple buildings, the action space includes multiple actions, 2 for each building. The state-space is a set of variables related to the controlled environment which are fed to the agent to learn the optimal control policy which maximizes the reward function. Eventually, a reward function was formulated to describe the performance of the control agent concerning the control objectives. **Training phase:** in the second stage of the process the DRL agent was trained. As previously introduced in Section 2.2, DRL agent is characterised by many

hyperparameters which require appropriate tuning. To enhance the reproducibility of the work, a description about the setting of hyperparameters was provided. The training process was implemented in an off-line fashion using the same training episode (i.e., a time period representative of the specific control problem) multiple times to refine agent's control policy. Training Results: the agent was firstly tested with the same climate used for training, to specifically analyse the effect of the learned policy on multiple levels, including single buildings, cluster and then on the grid. The performances of the DRL controller were analysed against an RBC controller, by evaluating various key performances indicators, specifically tailored for each scale of analysis (i.e., single building level or cluster of buildings level). Deployment: to evaluate the robustness of the trained agent, the algorithm was deployed in four different climates, which also lead to different building thermal-related loads. The agent was tested through a static deployment in one episode and compared with the RBC also during this stage.

3.2.3 Case study

The SAC algorithm described in Section 2.2.3 was used to control a complex environment that consists of a cluster of 4 commercial buildings, whose load profiles have been assessed through dynamic simulations in EnergyPlus. Each building is equipped with a heat pump, to satisfy cooling demand, an electric heater to meet DHW demand and both cooling and DHW storage. For each building, the heat pump is sized to always match cooling demand, considering a safety factor to account for reduced capacity in case of low external temperature. On the other hand, storage capacity is three times the maximum hourly demand for both cooling and DHW loads [183]. Moreover, two out of four buildings are equipped with photovoltaic systems.

Table 3.2 reports for each building their geometrical features and the main design details of the energy systems. The energy systems are managed by a multi-agent centralised DRL controller, which aims to reduce costs and to flatten the aggregated load profile of the cluster reducing peaks.

Table 3.2 Building and energy systems properties [35]

| | Type | Surface [m^2] | Volume [m^3] | Cold Storage Capacity [kWh] | Electric Heater Capacity [kW] | Hot Storage Capacity [kWh] | PV Capacity [kW] |
|------------|------------|----------------------|---------------------|--------------------------------------|--|-------------------------------------|------------------------|
| Building 1 | Office | 5000 | 13700 | 235 | 17 | 50 | 0 |
| Building 2 | Restaurant | 230 | 710 | 150 | 25 | 75 | 25 |
| Building 3 | Retail | 2300 | 14000 | 200 | 23 | 70 | 20 |
| Building 4 | Retail | 2100 | 10800 | 185 | 35 | 105 | 0 |

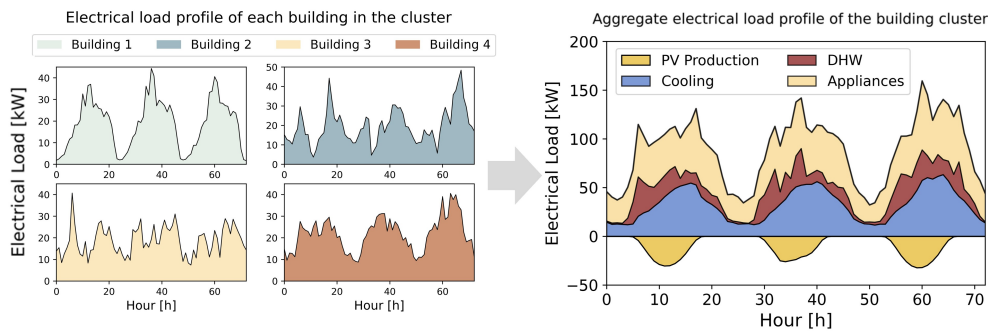


Fig. 3.6 Load profile for each building (left) and cluster profile electricity and PV production (right) [35]

3.2.3.1 Description of the cluster of buildings

The aggregated load pattern of the cluster can result from heterogeneous single building profiles, characterised by both very different intensities and shape. Figure 3.6 shows the electrical consumption patterns for the first three days analysed. On the left part, it is displayed the load profile for each of the 4 buildings included in the cluster analysed, while on the right part it is showed the total profile. In particular, Building 1 and 4 are characterised by homogeneous daily load profiles, with a peak in the morning, while Building 2 has a sudden peak during the evening and Building 3 may have more than a peak per day. As a result, considering that the load profile of the cluster is highly influenced by the single building energy behaviour, to achieve optimal control at cluster level coordination at both low and high levels is needed. Moreover, the right part of Figure 3.6 shows the breakdown of the cluster electrical demand for cooling, DHW and appliance and the PV production in Buildings 2 and 3. This representation is useful to underline the electrical demand for cooling and DHW, on which the RL controller can act to enhance cluster flexibility. Since

electrical cooling demand represents a large part of the cluster load, the analysis focused only to the summer period (1st June to 31st August).

3.2.3.2 Energy systems and control objectives

The control problem consists in the optimal management (i.e., charging and discharging process) of the 8 storage to satisfy cooling and DHW demand of the four buildings. The goal of the control policy is to minimize costs and to avoid peaks at cluster level. The most influencing factor to take into account are the energy cost and the heat pump efficiency. In particular, the energy cost considered in the work is based on the Austin (Texas) electricity tariffs [184]. In detail, were assumed an off-peak electricity tariff during night-time period 20:00-7:00 (0.03025 \$/kWh) and an on-peak electricity tariff during daytime period 7:00-20:00 (0.06605 \$/kWh). On the other hand, efficiency of the heat pump was modified from CityLearn original implementation to consider partial load ratio and the effect of external temperature not only on the coefficient of performance, but also on the design capacity. A description of the relation among COP, declared capacity and external temperature was defined according to real data sheet of heat pumps. Moreover, the heat pump operation at part load condition was modelled according to UNI EN 14825. Eventually, COP was evaluated according to Equation 3.1. The external temperature rise has a twofold effect, firstly it reduces COP (increasing electricity consumption) and secondly it increases the maximum cooling power deliverable by the heat pump. Moreover, heat pump efficiency is influenced not only by external variables, but also from controller actions affecting the cooling load. Finally, the fraction in Equation 3.1 accounts for part load ratio and intermitting operation of the heat pump.

3.2.3.3 Baseline rule-based control

The effectiveness of the DRL controller was assessed through a comparison with a manually optimised rule-based controller. In the baseline strategy, both cooling and DHW storage is charged during the night period, when the electricity price is lower and heat pumps can take advantage, in terms of efficiency, of lower temperature (i.e., higher values of COP). To limit peak demand, the charging process was spread over the whole night period, while the discharging process is homogeneous throughout the day.

3.2.3.4 Design of the deep reinforcement learning controller

The DRL control algorithm described in section 2.2.3 was trained and tested in the CityLearn environment, including constraints related to the maximum charging and discharging rate of the storage and ensuring that cooling and DHW demands are always met. In the next sub-sections, action space design is presented, together with the description of the reward function design and of the state-space, to properly characterize the DRL control problem.

Action-space design The analysed case study deals with multiple buildings, each one with two storage that could be controlled. Therefore, the two actions have different targets: the first one is related to the operation of the cold storage, which can be charged by the heat pump to store energy or discharged to meet building cooling load; the second action is related to the operation of the hot storage, which can be charged by an electric heater or discharged to meet DHW demand. Since each building has different storage and heat pump capacities, the action space makes use of normalized values. In particular, the controller uses actions between -1 and 1, where -1 represents the full storage discharge in the control timestep and 1 represents the full storage charge. 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 [184]. In conclusion, at each control time step the agent selects 8 values (one for each storage) to charge or discharge the energy storage devices. This information is used to select the best actions that maximize the reward function.

State-space design The states represent the environment as it is observed by the control agent. At each control timestep, the agent chooses among the available actions according to the values assumed by the states. In particular, states should be easy to measure in real-world implementation, and they should be selected according to the meaningfulness of the information they provide for predicting the reward function. The variables used for representing the state space are reported in Table 3.3 and in the following further described. The variables used for representing the state-space can be categorised as weather, district and building states.

Weather states, such as outdoor temperature and direct solar radiation, were selected because of their strong influence on the magnitude of building loads for

Table 3.3 State-space for the case study [35]

| Variable group | Variable | Unit |
|----------------|--------------------------------------|------------------|
| Weather | Temperature | °C |
| | Temperature Forecast (6h) | °C |
| | Direct Solar Radiation | W/m ² |
| | Direct Solar radiation Forecast (6h) | W/m ² |
| District | Total Load | kW |
| | Electricity Price | €/kWh |
| | Electricity Price Forecast (1,2,3 h) | €/kWh |
| | Hour of day | h |
| Building | Non-shiftable load | kW |
| | Heat Pump Efficiency | [-] |
| | Solar generation | W/m ² |
| | Cooling Storage SOC | [-] |
| | DHW SOC | [-] |

space cooling. Additionally, their 6 hours-ahead values were used to provide useful information about temperature and solar radiation changes and enhance the predictive capabilities of the controller. District states include variables that assume the same value for all the buildings over time, such as hour of day, electricity price, forecast of the electricity price and weather conditions. Building states include variables related to the electricity production (photovoltaic system) and consumption of the buildings (non-shiftable load). These states are specific to the single building, which can have different energy systems (PV) or trends of consumption. Additionally, heat pump efficiency, cooling and domestic hot water state of charge were included. Figure 3.7 summarizes the states and action space interaction selected in this work. The central controller receives as states high level information such as weather conditions and electrical demand of the whole cluster of buildings. Moreover, it also receives low-level information from each building such as appliance loads and energy systems information.

Reward design The reward function plays a key role in defining how the agent assesses the quality of the control policy during the learning phase. It was conceived

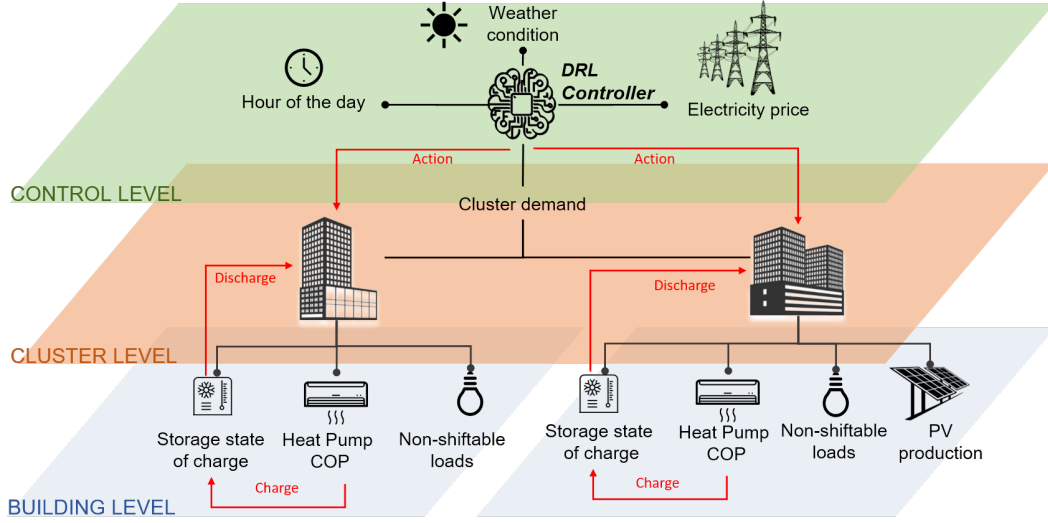


Fig. 3.7 State-action space representation of the DRL controller [35]

to allow the agent learning a control policy during training period which minimize the energy cost at cluster level and reduce the demand peaks. In particular, the reward was formulated as follows:

$$R = \sum_{i=1}^n e_i^2 * c_{el_i} \quad (3.6)$$

where e_i^2 is the squared energy consumption of the i -th building, while c_{el} is the electricity tariff in that time step. To obtain a more uniform load profile at cluster level, the controller tries to minimize the sum of the squared consumption of each building for each time step. This formulation was chosen since the squared minimization approach tries to flatten the profile rather than shifting the consumption to low electricity tariff, avoiding simultaneous charge (and discharge) of storage. To consider the economic aspect of the problem, the electricity tariff in the specific timestep was included. Moreover, due to the relation between consumed energy and costs, the controller tries to minimize energy consumption, increasing system efficiency. The design of the reward function highly influences DRL performances, searching compromises between energy savings and grid stability.

3.2.3.5 Training and deployment

The subsection describes the setting of hyperparameters during the training phase. Then, a description of the different climatic conditions analysed for the deployment phase is presented.

Training phase The DRL framework is characterised by several hyperparameters that strongly affect the behaviour of the control agent. This subsection aims to illustrate the hyperparameters set during the formulation of the control problem. For the sake of reproducibility, Table 3.4 reports the value of the main hyperparameters. In particular, the two hyperparameters that mostly influence the results are the number of training episodes and the temperature α .

Table 3.4 Hyperparameter settings [35]

| | Variable | Value |
|----|--------------------------|------------------------------|
| 1 | DNN architecture | 2 Layers |
| 2 | Neurons per hidden layer | 256 |
| 3 | DNN Optimizer | Adam |
| 4 | Batch size | 512 |
| 5 | Learning rate λ | 0.003 |
| 6 | Decay rate τ | 0.005 |
| 7 | Temperature* α | Starting =1, Final = 0.05 |
| 8 | Entropy coefficient* H | Starting = 8, Final = 5 |
| 9 | Target model update | 1 |
| 10 | Episode Length | 2208 Control Steps (92 days) |
| 11 | Training Episodes | 5 |

Differently from many other control fields, the number of training episodes is relatively low. This is justified by the problem nature, in which actions are constrained by energy balance, finding the optimal policy quickly. Furthermore, as explained in Section 2.2.3 a highly influences the outcome of the policy. While in certain applications a could be set a-priori as a constant, in this study a version of SAC algorithm that optimizes the temperature parameter was adopted. As a reference, both starting and final value of temperature and entropy coefficient are provided below. As previously stated in Section 4.1, a training episode includes 3 months, from 1st of June to 31st of August (2208 control steps). The weather file

used in this work for the training phase is referred to the climatic zone of the USA named 2A, Hot-Humid.

Deployment phase In the last phase of the process the agent was deployed for the same cluster of buildings but considering four different climates to assess the adaptability capabilities of the learned control policy to different configurations related to the controlled environment. Each agent was deployed for one episode including the period between 1st June and 31st August. The first climate is 2A Hot-humid: this climate is the same on which the agent was trained on. This scenario is compared to the baseline RBC to assess the effectiveness of the trained agent. Then, the adaptability is tested with the deployment of the agent in warm-humid climate (3A), mixed-humid climate (4A) and cold humid climate (5A). The thermal related load patterns changed according to climatic conditions. Figure 3.8 shows the patterns of outdoor air temperature in the four climates selected, highlighting how the external temperature is strongly different in amplitude and distribution. In particular, the Climate 2A, the one on which the agent is trained on, has a distribution with a narrow amplitude, with a mean temperature of about 27.5 °C. On the other hand, climates 3A and 4A have a different temperature distribution, but the same mean value of about 25.5 °C. Lastly, Climate 5A is the coldest climate considered, with a mean temperature of 22.5 °C and a more uniform distribution with respect to Climate 2A.

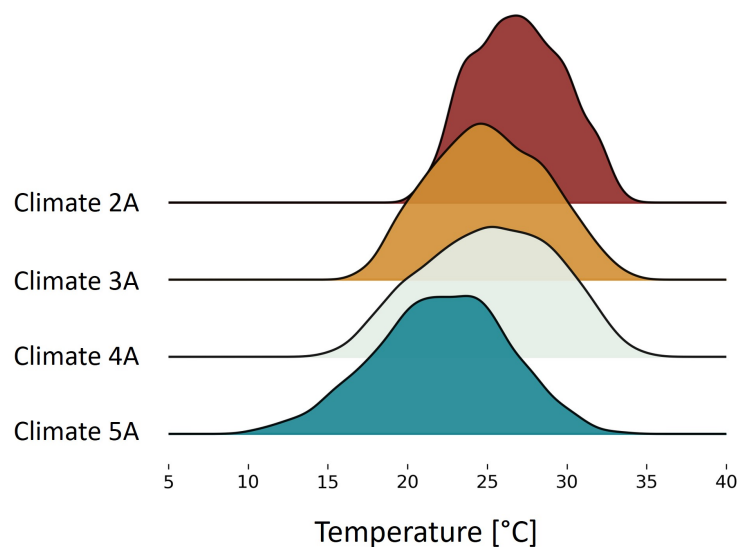


Fig. 3.8 Temperature distribution of the different deployment climates [35]

Table 3.5 Reward and KPI evolution over training period [35]

| | Episode 1 | Episode 2 | Episode 3 | Episode 4 | Episode 5 | Deployment |
|--------|-----------|-----------|-----------|-----------|-----------|------------|
| Reward | -343 | -337 | -297 | -271 | -271 | -265 |
| Cost | 1.1 | 1.1 | 1.06 | 0.98 | 0.97 | 0.96 |
| Peak | 1.07 | 1.29 | 0.96 | 0.96 | 0.96 | 0.88 |

3.2.4 Results

The section reports the results of the implemented framework. Firstly, a brief evolution of the control policy is presented. Then, a comparison between the two control strategies by analysing the results with a focus at single building scale and cluster level is provided. Eventually, the analysis focuses on the load curve and the role of storage devices in grid stability. To this purpose, a further comparison is performed by computing the load duration curve for the cluster of buildings also considering the case without storage. Furthermore, to summarise the performance of the two control strategies a numeric comparison is provided.

3.2.4.1 Training results

The subsection presents the evolution of the DRL control strategy over the training period and compares it with the RBC. In particular, Table 3.5 reports the evolution of the reward function over the training period, together with the normalized values of cost and peak compared to the RBC (where a value smaller than 1 suggests better performance of the DRL). The first episode is used to store states and actions and after that it can be observed a quick convergence of both cost and peak term, which stabilize after episode 4. To prove this point, a sensitivity analysis was performed on the number of training episodes, spanning from 5 to 20, which showed little to no improvements with more than 5 episodes.

Comparison of controllers at single building level Figure 3.9 shows the charging and discharging patterns of both storage determined by the RBC and RL controller. In particular, the figure shows the days related to the maximum peak demand of the RBC, to highlight the difference with the DRL control strategy. Moreover, the figure shows the relation between the control process and the forcing variables

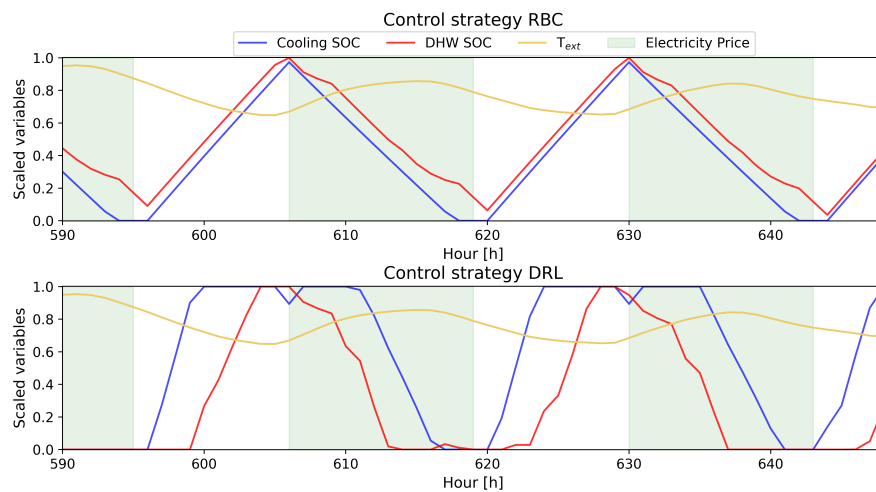


Fig. 3.9 State of charge of storage and forcing variables scaled between 0 and 1 [35]

(i.e., external temperature and the electricity price). To allow an easier comparison, all quantities were normalized on maximum values between 0 and 1, where the maximum temperature is 35 °C. It can be observed that RBC charges both storage mutually at a lower rate, exploiting off-peak tariff and the highest COP of the heat pump. However, to exploit off-peak tariff and avoid sudden peaks, RBC control strategy leads the heat pump to work at part load. Moreover, it has no information about outdoor temperature evolution and so on the efficiency of the heat pump. On the other hand, the DRL controller learns to charge the two storage as soon as the electricity price and the temperature tend to decrease. However, the main difference is related to the discharge pattern, since DHW is used as soon as needed to reduce electricity demand, while cooling storage is discharged when external temperature is high, avoiding using heat pump when the COP is low.

Comparison of controllers at cluster level Figure 3.10 shows a comparison between the aggregate electrical load obtained through the implementation of the RBC controller in the simulation environment, in which each building is optimised to minimize its costs, and the DRL controller, which optimises cluster behaviour. In particular, Figure 3.10 shows three days during which the RBC determined the occurrence of load peak at cluster level that could cause stress on the grid. As shown in Figure 3.10, the DRL controller is capable to better flat the aggregate load profile and to diversify the charge time of the storage among the buildings

in the cluster. As a result, the cluster profile is more homogeneous and, in this particular situation, a great reduction can be observed by looking at the two peaks (hour 605 and 655). This result does not represent the average performance of the DRL controller but highlights the potential of buildings coordination in increasing grid stability in specific situations. To understand how these results have been

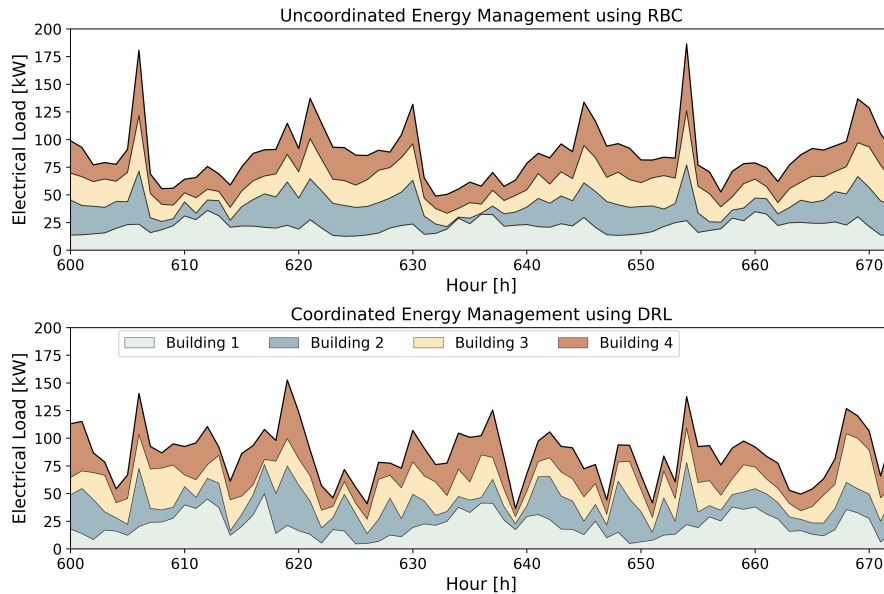


Fig. 3.10 Comparison between uncoordinated and coordinated energy management [35]

achieved, Figure 3.11 shows the average evolution of the state of charge related to the storage device. As can be seen, the cooling and DHW storage is charged during the night homogeneously, as the RBC. The main difference is related to the storage discharges. In particular, as soon as the electricity price increases, the DHW storage starts the discharge phase almost simultaneously, since they are only influenced by the electricity price. On the other hand, the agent learned the dependency between external temperature and heat pump COP (the higher the temperature lower the COP). As a result, the optimal policy discharges the cooling storage during the hottest hour, maximising heat pump efficiency.

Comparison of controllers at grid level To highlight the flexibility provided by the introduced framework in terms of load profile flattening, the load duration curves resulting from the application of RBC and DRL control and the case without storage were compared in Figure 3.12. As can be seen, the baseload increases with both

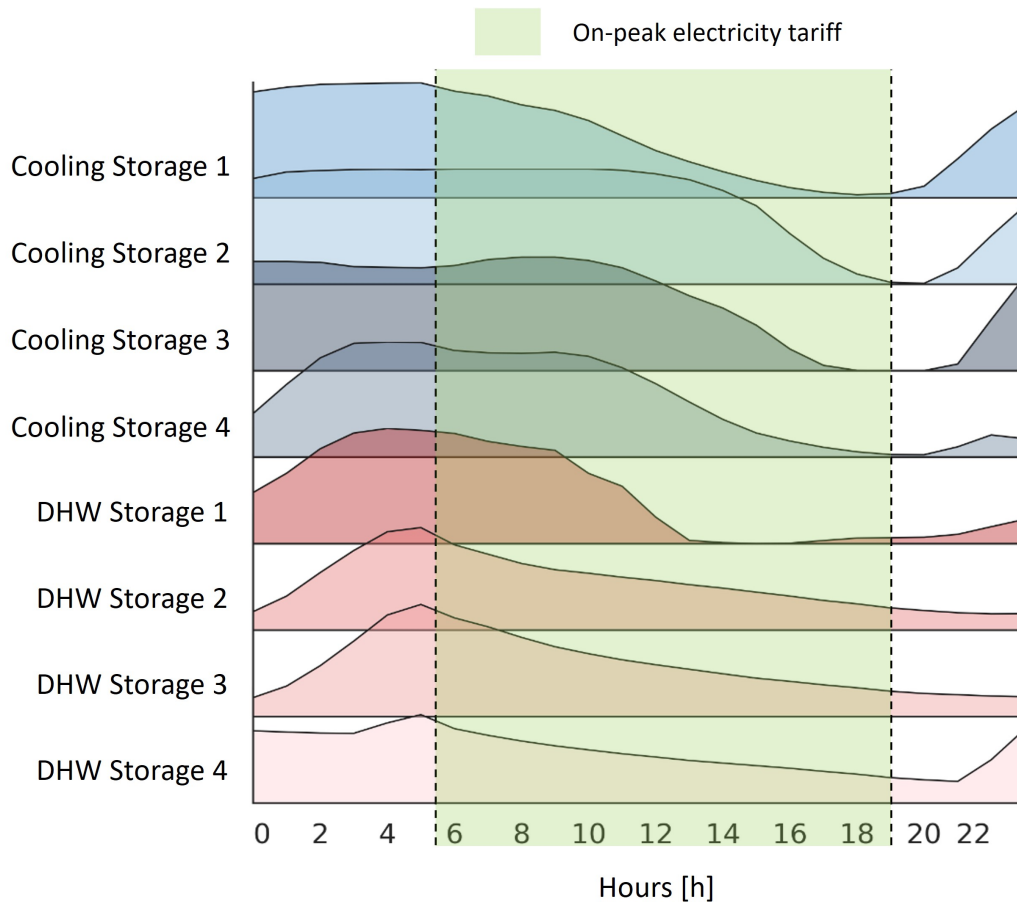


Fig. 3.11 State of charge of storage averaged over a day [35]

RBC and DRL, underlying the importance of the storage in increasing buildings energy flexibility. However, RBC leads to the creation of new undesirable peaks (as shown inside the “zoom” area) while DRL algorithm, thanks to the coordinated approach, can reduce them.

Eventually, to provide a comprehensive analysis of the results different KPIs are introduced to compare the performances of the control strategies. In particular, the KPIs chosen are the total energy consumption, the total energy cost, maximum peak, average daily peak, peak-to-average ratio (PAR) and daily peak-to-average ratio. These KPIs have been chosen to summarise the advantages of DRL control strategies at cluster level (energy consumption, costs, maximum peak and peak-to-average ratio) and the effect on the grid (average daily peak and daily peak-to-average ratio). Table 3.6 shows the performance of the two control strategies with respect to the main criteria selected. to allowan easier comparison, the values of KPIs are normalized on

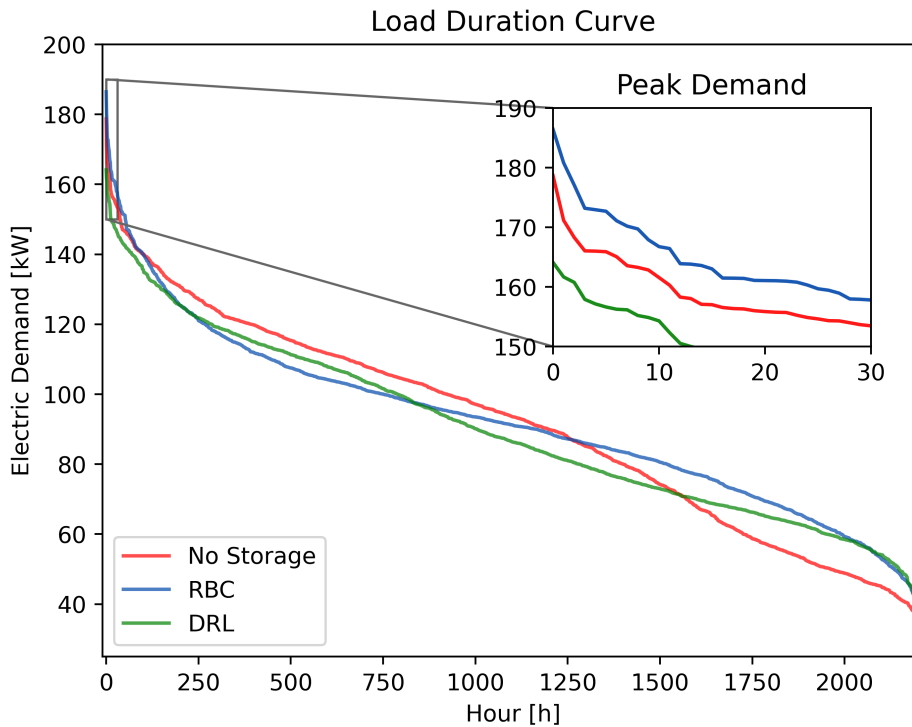


Fig. 3.12 Load duration curve for the base case without energy storage in buildings and the two control strategies [35]

the RBC values. DRL outperforms the manually optimised RBC. In particular, DRL controller exploits the storage charge and discharge to increase heat pump efficiency, while slightly reducing electricity cost.

Table 3.6 Comparison between performances of the two control strategies [35]

| | Energy Consumption | Electricity Cost | Maximum Peak | Peak-to-average ratio (PAR) | Average daily peak | Average daily PAR |
|------------------------|--------------------|------------------|--------------|-----------------------------|--------------------|-------------------|
| Manually Optimised RBC | 1 | 1 | 1 | 1 | 1 | 1 |
| DRL | 1.03 | 0.96 | 0.88 | 0.96 | 0.90 | 0.94 |

Nevertheless, it must be noticed that the manually optimised RBC already took full advantage of off-peak electricity tariff, therefore the economic improvement of DRL over RBC is closely related to the more efficient use of energy. On the other hand, coordinated approach showed good results at cluster level, reducing maximum

peak of 12% and average daily peak of 10%. Moreover, the PAR and average daily PAR reduction of 4 and 6% respectively highlight the benefits of building coordination which can be translated into more homogenous energy consumption. Furthermore, the advantage provided by the increased grid stability could be translated into reduced electricity tariff, with additional advantages for users. The DRL approach can reduce peaks of 12% with respect to the RBC and 8% with respect to the no storage case, but more importantly, the peak demand rapidly decreases, resulting in a more homogeneous profile.

3.2.4.2 Deployment of deep reinforcement learning controller in different climatic conditions

The last section analyses the deployment of the agent in the other 3 climates described in 4.5.2. After the training and deployment of the agent in Climate 2A, a simulation of 3 months was run using the trained agent with climate 3A, 4A and 5A. To evaluate the performances of the agent in the new climates, as done before, the DRL controller was compared with the RBC controller, analysing the previously introduced KPIs and normalizing them on the RBC values. Figure 3.13 summarizes the results of the deployment phase, where 100 represents the RBC performance.

It can be seen how the controller is able, also in Climate 3A, to flatten the load pattern. This is highlighted by the peak and PAR reduction, looking both at maximum and daily values. These results are achieved consuming slightly more electricity with respect to the RBC, but with the same energy cost. Looking at Climate 4A, it can be noticed a peak reduction of around 5%, but with negligible effect on the PAR. On the other hand, looking at the average daily values, it can be noticed that the daily PAR is 10% lower with respect to the RBC, highlighting the more homogeneous consumption. Eventually, analysing climate 5A, the coldest one, it can be seen how the cost slightly increases, around 3%, however there are great improvements at district level, with a peak reduction of 11% and an average daily PAR 22% lower with respect to the RBC case.

3.2.5 Discussion

The presented work aims to exploit model-free DRL controller to coordinate the energy management of a cluster of buildings. The analysis is performed with the

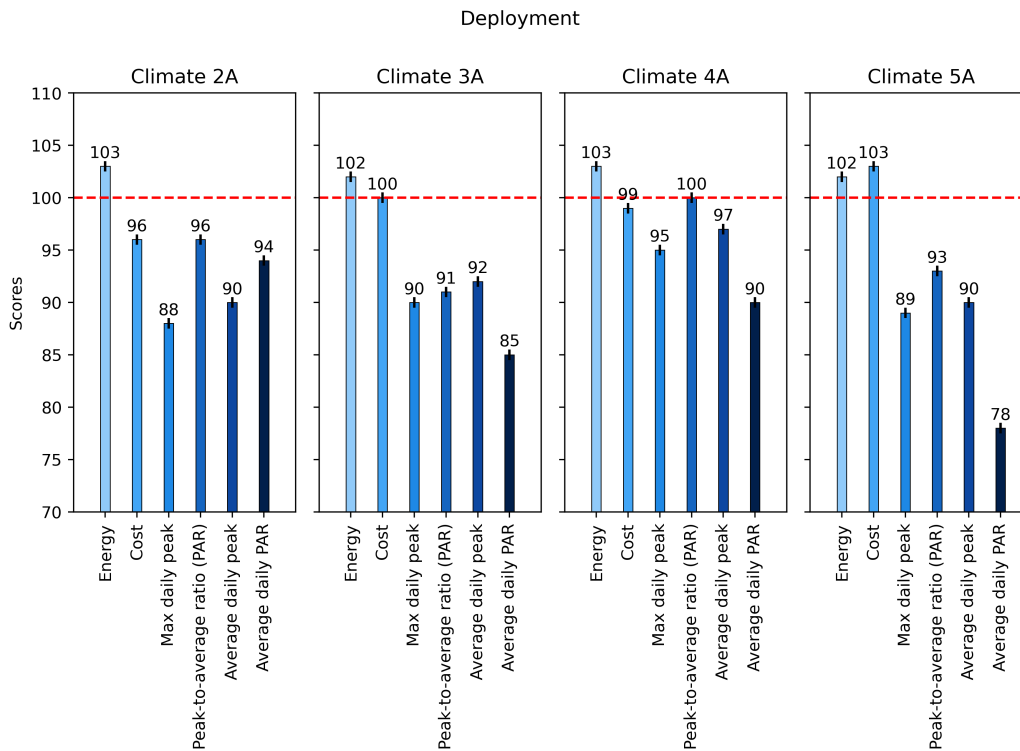


Fig. 3.13 KPI comparison for the four-deployment case [35]

CityLearn environment, an openAI gym environment where a detailed representation of the heat pump and a variable electricity price has been implemented. The DRL controller was designed to act on the DHW and cold storage of 4 buildings to optimise both energy costs and peak demand at cluster level. The control problem analysed involved renewable energy sources, variable electricity price and building coordination. To compare the DRL performances and underline the effect of a coordinated energy management versus a single building optimization, a manually optimised RBC controller baseline was introduced. Despite the complex environment, the DRL controller found the optimal policy to exploit environment behaviour, consuming energy more efficiently and charging and discharging storage to optimise the cluster load profile. Additionally, due to the problem nature, the solution was found with a very short training period of 5 episodes. The analysis highlights how the real-world implementation could be done with a relatively small amount of data for the training, proving the versatility of the proposed approach. However, to study the interaction among states, actions and rewards it is still necessary a simulation environment

when dealing with a district scale. Looking at the problem formulation, forecast information on electricity price and weather helped to rapidly find an optimal policy, highlighting how important is the proper design of the state-space. Moreover, the design of the reward function plays a key role for the DRL controller behaviour. It is therefore necessary to find an optimal trade-off between the advantages of single users and cluster that are bounded to the case study. During the work, the adoption of the square minimization was found to be effective at both single building and cluster level, proving to be scalable independently by the number of buildings. To test its adaptability, the controller was deployed considering four different climatic conditions. The results highlight that the controller flattened the cluster load profile, almost independently from the external conditions, while the economic performances varied with the different cases. Even with the same (or slightly higher) electricity costs, the services provided to the grid, such as peak reduction and load shaping, justify the adoption of the DRL controller with respect to the RBC. Eventually, the strength of the proposed approach is not only the mere improvement of energy performances, but the opportunity provided by its adaptive nature to account the cluster environment evolution. A large environment may involve rapid changes, such as consumption pattern modification and demand response programs.

3.3 A comparison among coordinated and cooperative deep reinforcement learning architectures in buildings

The use of MARLs has attracted increasing interest, due to the possibility of controlling complex and distributed systems, optimising often conflicting objective functions. Depending on the scale of application, the type of multi-agent system directly influences its applicability, scalability and robustness. Although these analyses have been carried out at larger building scales, there are still many questions about their use for scales ranging from individual buildings to microgrid.

The next section presents the research questions analyzed and introduces the motivations and novelty of the proposed methodological approach.

3.3.1 Motivations and novelty of the proposed approach

Multi-agent systems (MAS) represent a viable alternative to enhance the demand-side-management of multiple entities. Multi-agent systems find their natural use in microgrid applications, where they are mainly used in power market scenarios [185, 186] and microgrid management [187, 188]. MAS leverage several methods, including mathematical methods [189], meta-heuristic methods [190], and heuristic methods, that can be further divided into game-theory based [191] and reinforcement learning based [192–194]. In microgrid applications, MAS often considers the entire demand as aggregated by a cognitive agent, as done in [195], in which the cognitive agent represents the entire microgrid demand and coordinates its operation with the generation agents (reactive) to optimise several objective functions including cost, emissions and grid stability. To fully exploit the flexibility associated with buildings, the scale of analysis should be between single buildings and aggregated demand, in the so-called neighborhood, communities, districts or integrated microgrid. From this perspective, Labeodan et al.[196] analysed the role of MAS in smart-grid integration, while [51] reviewed the different kinds of MAS applications for smart homes, highlighting the role of MAS architectures.

As explained in the previous chapter, multi-agent reinforcement learning has been chosen as a grid-interactive building control framework for districts, due to the advantages provided by its applications in buildings. In their review of RL for building controls, Wang et al.[87] highlight the growing research interest of RL and the potential of MARL to address some of the limitations of other advanced control strategies such as model predictive control. However, the application of MARL for building energy management is relatively new. While some pioneering studies proved the effectiveness of MARL [197–199], further studies need to be performed to explicitly address the advantages deriving from the combination of the different algorithms and architectures in buildings.

Among the most recent RL algorithms, the Soft Actor Critic (SAC) algorithm [171] emerges for its ability to handle a continuous action space and it has gained significant interest since its first publication. The effectiveness of the SAC algorithm has been proven in the energy environment [200] and for the energy management of single buildings [201]. Biemann et al. [202] compared SAC algorithm with other three actor-critic algorithms to control the HVAC of a data centre, finding that SAC algorithm showed substantial improvement in both performance and sample

efficiency. In this framework, it may be useful to analyse the effectiveness of different RL architectures. An initial attempt to compare multiple SAC architectures for buildings control was performed by Dhamankar et al. [203]. The authors provided an empirical comparison of independent learners (distributed architecture), centralised critics with decentralised execution (centralised architecture) and value factorisation learners (hybrid architecture). The main limitation of that work is related to the comparison of an average metric, which does not allow to understand the strengths and weaknesses of each approach, especially shifting the attention from the district to single buildings.

The literature review presented revealed the following research gaps: the application of advanced control strategies for DSM to date has largely been confined to single buildings. Among the few studies that analysed multiple grid-interactive buildings, the focus has been on microgrid applications with appliance scheduling or electric vehicles, requiring further analysis on the role of thermostatically controlled loads and thermal storage for grid-interaction. Moreover, there is a lack of studies aimed at comparing different control architectures when dealing with heterogeneous energy systems. Indeed, individual buildings may have their independent objectives and the way by which such individual objectives, when part of a district, influence control design problems, needs to be further investigated. Lastly, considering the multi-objective nature of the grid-interactive DSM problem, a detailed analysis of the advantages and disadvantages of each architecture/algorithm is required.

With these research gaps in mind, this work provides the following contributions and novelty by:

1. Comparing the performance of a coordinated (centralised) and cooperative (decentralised) MARL architecture for the provision of DSM in a district of heterogeneous buildings.
2. Analysis of a DRL controlled grid-interactive district at different scales and times. Assessment of advantages and limitations of the proposed architectures for specific buildings and the entire district.
3. Studying the application of a multi-agent SAC RL algorithm to a district DSM problem with heterogeneous buildings, testing their robustness in different conditions and assessing the versatility of different controller architectures.

The presented work deals with the energy management of four buildings, equipped with thermal energy storage and PV systems and formulating the problem as a reinforcement learning based one. Two SAC-MARL algorithms are explored: a centralised (coordinated) controller and a decentralised (cooperative) controller, which are benchmarked against a rule-based controller (RBC) that aims at exploiting electricity tariffs to minimize the cost.

The paper is organised as follows: Section 3.3.3 describes the case study and control problem, followed by the baseline reference controller and KPIs used for comparison. Section 3.3.4 presents the design process of the two DRL controller architectures. Section 3.3.5 provides the results of the key findings with a focus on both the comparison of the various MARL architectures against the baseline controller and the robustness of the agents under different climate types. Lastly, Section 3.3.6 critically discuss these results.

3.3.2 Methodology

In this section, the methodological framework for the development and assessment of the performances of the two RL architectures (coordinated and cooperative) is presented. In particular, the methodology is structured in three steps, as represented in Figure 3.14 and described below in further detail.

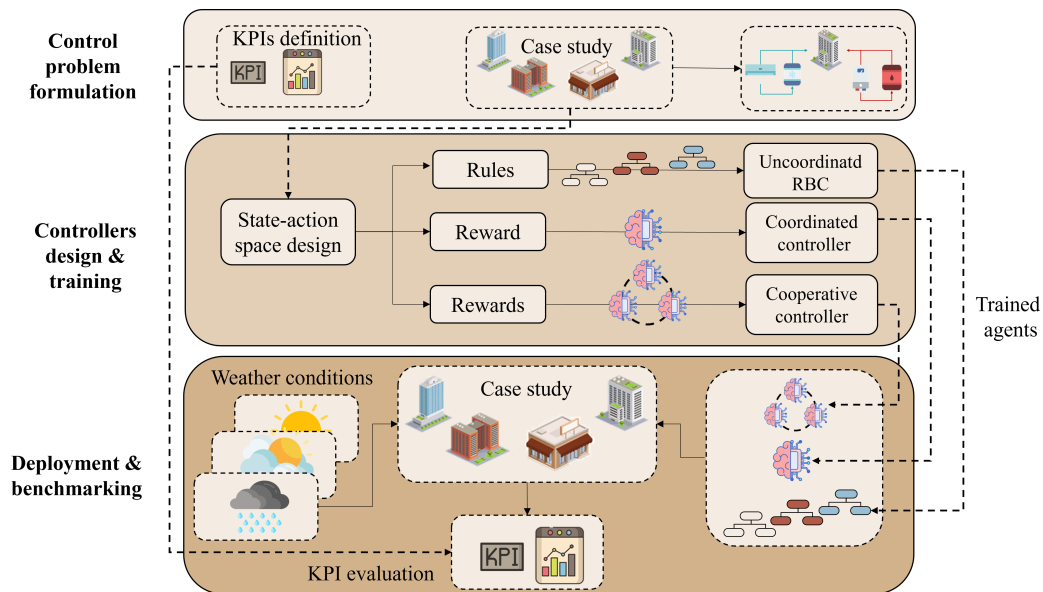


Fig. 3.14 Methodological framework overview [37]

Control Problem Definition: The first step describes the case study district (Section 3.3.3.1), with a focus on the controllable energy systems and the uncoordinated RBC, which is used as baseline. Lastly, it describes of weather data used to test the robustness of the proposed control strategies. Section 3.3.3.3 describes the control problem and outlines the electricity tariffs which support the more flexible operation of the energy systems and the reference baseline controller. To quantify controller performance and allow comparisons to be made between the controllers, Section 3.3.3.4 introduces the set of specific KPIs used in this study.

Controller Design & Training: The second step of the methodology analyses the main components of the RL problem. In particular, Section 3.3.4.1 presents the design of the action-space to analyse the possible control actions that can be taken by the agents. Section 3.3.4.2 similarly presents the state-space design which is the information provided by the environment to the RL agents. Section 3.3.4.3 formulates the reward functions for each architecture analysed to quantify controller performance with respect to control objectives.

Deployment & Benchmark: The last step focuses on the deployment and benchmarking of the trained agents. In particular, to test their robustness, the controllers are deployed in several climates, previously described in Section 3.3.3.1 and their performance is evaluated through several KPIs (Section 3.3.3.4).

3.3.3 Case study district & control problem

This section describes of the case study. In Section 3.3.3.1, the energy systems and weather climates used for the analysis are outlined. Next, the control problem is analysed in Section 3.3.3.3 and lastly, the KPIs used for the analysis are presented in Section 3.3.3.4.

3.3.3.1 District

The district includes four buildings: a restaurant, and three multi-family buildings, which can be further demarcated as prosumers that do not export electricity (Building 2 and Building 4) and prosumers that export electricity (Building 1 and Building 3).

Each building is equipped with PV panels, a reversible heat pump, an electric heater and two storage devices (chilled water and SHW). The control problem focuses on the energy management of the two storage devices per building, to optimise costs, profile shape and self-consumption. To quantify the effects of the control strategy, several KPIs, described in the next subsection, have been used.

The district electrical load is mainly influenced by the building cooling loads and, as a result, the analysis focuses only on the summer period (defined as the 1st June to 31st August), which represents the simulation period used in this study.

Moreover, as weather conditions influence the cooling load and control strategy, the effects of weather variation on the behaviour and robustness of the controllers were analysed. Whilst studies have investigated the ability of DRL to adapt to different operating conditions (e.g., weather conditions [204], occupancy and set point changes [205]), there is a necessity to study how multiple agents address these changes for each of the cooperative and coordinated environments, which may lead to a non-stationary problem. On the grounds of this, each agent is trained on one climate (2A) and further deployed in the other two climates (3A and 5A), as summarised in Table 3.7. The climate zones considered are diverse and are as per the ASHRAE standard definitions. This analysis aims to evaluate and compare the ability of the two controllers to adapt to different environmental conditions.

Table 3.7 Climate zones (per ASHRAE definitions) considered in this study [37]

| Climate Zones | Location | T_{min} [°C] | T_{mean} [°C] | T_{max} [°C] | T_{σ} [°C] |
|---------------|-------------|----------------|-----------------|----------------|-------------------|
| 2A | Houston, TX | 20.0 | 27.5 | 35.5 | 3.0 |
| 3A | Atlanta, GA | 16.0 | 25.5 | 36.0 | 4.0 |
| 5A | Chicago, IL | 8.5 | 22.0 | 35.0 | 4.5 |

3.3.3.2 Energy systems at building level

Figure 3.15 shows a schematic of the control architecture with details of the energy systems for a representative building of the district, while a comprehensive formulation of the mathematical problem can be found in [206]. In particular, the scheme highlights the controlled systems (chilled water and SHW storage) and their interaction with other energy systems. The heat pump can either charge the chilled water

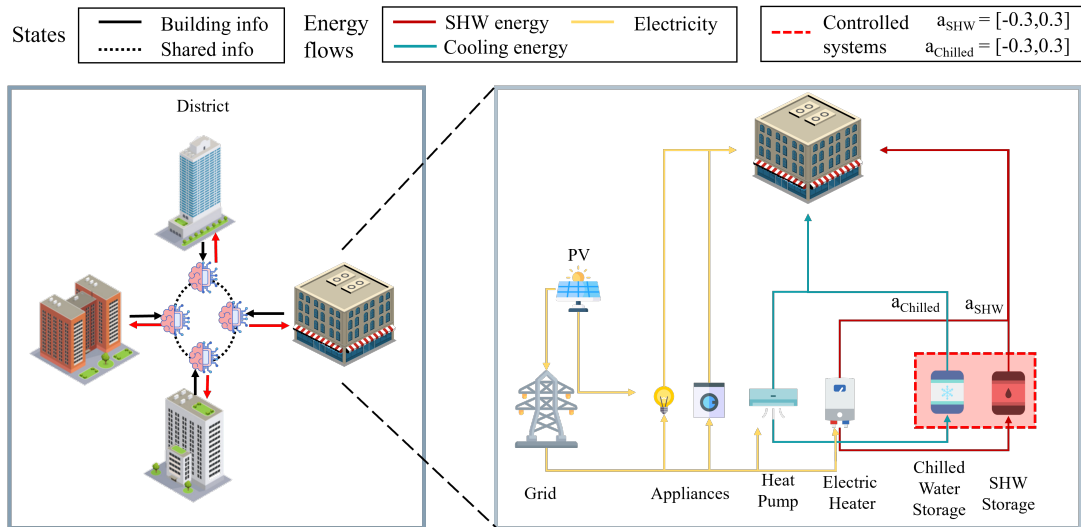


Fig. 3.15 Building energy management control scheme [37]

storage and satisfy the heating and cooling energy demand of the building, although the current analysis only focuses on the summer period. The electric heater is used to charge the SHW storage and to meet the SHW demand, while non-shiftable loads can be satisfied using electricity from PV or imported from the grid. Furthermore, Table 3.8 reports in detail the geometrical features of the buildings, together with the capacity of the two controlled systems (storage) and the PV size.

Table 3.8 Summary of building geometrical features and energy systems in district

| | Type | Floor Area [m^2] | Volume [m^3] | TES Capacity [kWh] | SHW Storage Capacity [kWh] | PV Capacity [kW] |
|------------|--------------|-------------------------|---------------------|--------------------------|-------------------------------------|------------------------|
| Building 1 | Restaurant | 230 | 710 | 235 | 50 | 50 |
| Building 2 | Multi-family | 3130 | 9550 | 150 | 75 | 20 |
| Building 3 | Multi-family | 3130 | 9550 | 200 | 70 | 60 |
| Building 4 | Multi-family | 3130 | 9550 | 185 | 105 | 20 |

It can be noticed that, despite having the same floor area, the three multi-family buildings are characterised by different cooling, heating and appliance loads. Indeed, to represent user stochasticity, probabilistic regression models were trained from different open source datasets [206] to create realistic instances of indoor temperature set point, SHW consumption and appliance schedules. Accordingly, the two storage devices are sized to satisfy three times the maximum hourly demand, of cooling and SHW loads respectively, while the heat pump and electric heater are sized

to always ensure the meeting of building loads [206]. Based on this information, an optimal control strategy should leverage PV electricity to partially offset non-shiftable, cooling and SHW loads, or even charging thermal storage during renewable overproduction periods, exploiting the energy multi-carrier nature of the control problem.

Lastly, to analyse the contribution of renewable electricity to the building load, Figure 3.16 displays PV self-consumption and export for each building, together with their net load for the first three days of the simulation period for climate 2A. As highlighted earlier, Building 1 and Building 3 are prosumers, exporting a certain quantity of energy. On the other hand, Building 2 and Building 4 self-consume renewable energy. It is crucial to notice that the building electrical demand, affected by climatic conditions, directly determines the ability of a prosumer to export electricity or not. The energy systems are managed with two controller architectures (Section 3.3.4), which aim to reduce operational costs and to flatten the aggregated load profile, exploiting the existing sources of flexibility.

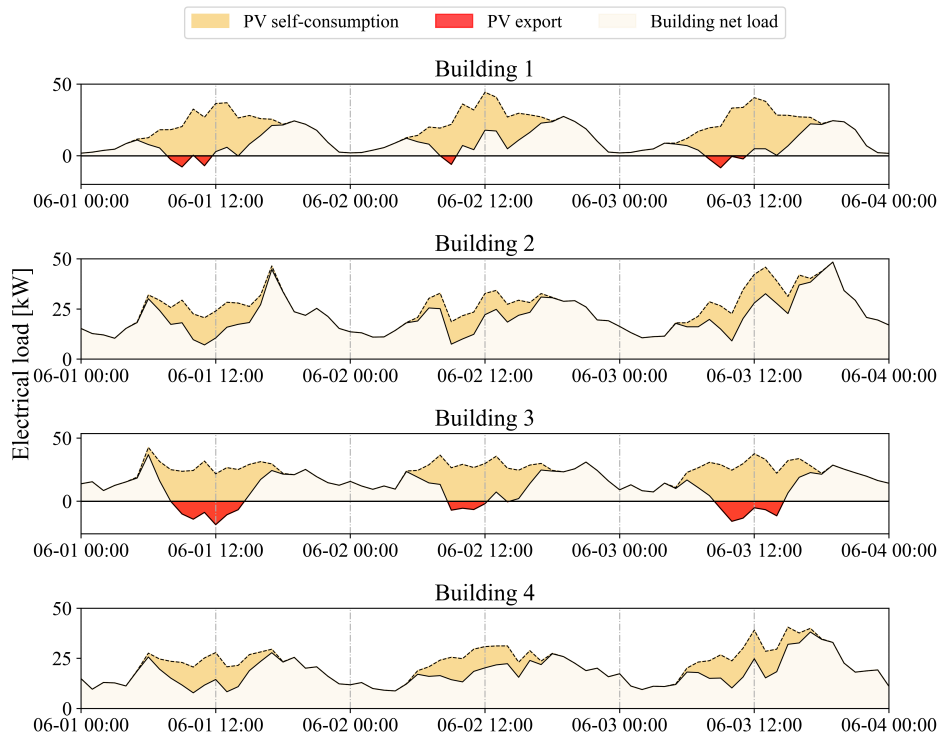


Fig. 3.16 Electrical load profile for each building in the district for Climate 2A [37]

3.3.3.3 Definition of the control problem

The controllers were designed to manage the charging and discharging of chilled water and SHW storage systems for the district of buildings, to minimise electricity costs, reduce cluster demand peaks and maximise self-consumption. The electricity price tariffs and the PV production are the main drivers of the district cost. In particular, the electricity price (c_{el}), chosen according to [207], varies from $c_{el,off-peak} = 0.01891$ \$/kWh during off-peak hours (21.00 – 12.00) to $c_{el,on-peak} = 0.05491$ \$/kWh during on-peak hours (12.00 – 21.00). Moreover, a cost related to the monthly peak load of the district was considered and defined below:

$$C_{Peak} = c_{Peak} * P_{Monthly,Peak} \quad (3.7)$$

Where $c_{Peak} = 11.02$ [\$/kW] is the tariff related to the monthly peak load $P_{Monthly,Peak}$ [kW], evaluated as the maximum district load for each month. In the context of coordinated energy management, if the cluster of buildings is managed by the same aggregator, it would face a cost related to the district monthly peak demand, that the controller should try to minimize, since it could represent a not negligible part of the total cost faced by the district. Furthermore, any electricity overproduction can be sold to the grid according to the following tariff: $c_{sell} = 0.01$ \$/kWh. The electricity tariffs are summarised in Table 3.9.

Table 3.9 Electricity tariff including energy terms and peak terms [207, 37]

| | On-peak [\$/kWh] | Off-peak [\$/kWh] | Sold [\$/kWh] | Peak [\$/kW] |
|--------------|------------------|-------------------|---------------|--------------|
| Price | 0.0549 | 0.0189 | 0.0100 | 11.02 |

To benchmark the performance of the two DRL architectures, a RBC was used as baseline. The RBC uses a distributed architecture, to minimise individual building energy cost. This is achieved by exploiting the electricity tariff, charging chilled water and SHW storage over the night period and discharging uniformly over the day to reduce electricity consumption during on-peak hours. In this configuration, the individual building controller does not share any information with the other buildings. To avoid a suddenly shifted peak that could lead to higher costs, both charging and discharging operations are uniform.

3.3.3.4 Key performance indicator design

Due to the multi-objective nature of the problem, the optimal control strategy needs to optimise multiple objectives, finding a trade-off between all. Several KPIs [208], shown in Table 3.10, are used to quantify the performance of the controller, considering: an economic KPI (*Cost*), grid-interaction KPIs (*Peak*, *Peak-to-average ratio (PAR)*, *Daily Peak* and *Daily PAR*) and flexibility KPIs (*Flexibility Factor*, *Self-sufficiency*). To analyse the effects of the proposed control strategies on a daily basis, this study calculates and investigates Peak and PAR during the entire simulation period and at daily granularity, to emphasise building interaction with the grid. Furthermore, the self-sufficiency indicator, defined as the ratio between self-consumption and total consumption, is introduced to quantify the impact of the control strategy on renewable electricity integration. Lastly, the flexibility factor, defined as the ratio between off-peak imported electricity consumption and total imported electricity consumption, is used to analyse the amount of electricity consumed during each tariff period. The mathematical definition of these KPIs is provided in Table 3.10.

3.3.4 Design of multi-agent reinforcement learning control strategies

This section describes the design of the two DRL architectures, denoted as coordinated (centralised) and cooperative (distributed) approaches. Section 3.3.4.1 describes of the state-space, Section 3.3.4.2 outlines the action-space and finally Section 3.3.4.3 details the reward functions utilised by each approach. Together, these characterise the MARL approach utilised.

Figure 3.17 shows the framework of the two proposed DRL architectures. The image on the left describes the coordinated architecture. System level information is shared with the control level, which coordinates actions using all the information available for the cluster of buildings, to find the optimal coordination. On the other hand, cooperative management (image on the right) exploits multiple controllers, that share only common information such as weather forecast, grid information or district total electrical load, to find the best policy for each building.

Table 3.10 KPIs Used in MARL controller comparisons [37]

| KPI | Formula | Units |
|-----------------------------|---|-------|
| Cost | $\sum_i^n e_i * c_i$ | [\$] |
| Peak | $\max \sum_i^n \frac{e_i}{\Delta t}$ | [kW] |
| Daily-Peak | $\frac{\sum_i^{n_{day}} Peak_{day}}{n_{day}}$ | [kW] |
| Peak-to-average ratio (PAR) | $\frac{Peak}{\sum_i^n e_i / n_{day}}$ | [-] |
| Daily-PAR | $\frac{\sum_i^{n_{day}} PAR_{day}}{n_{day}}$ | [-] |
| Self-sufficiency (SF) | $\frac{\sum_i^n \sum_{j=1}^T \min(PV_{i,j}, e_{i,j})}{\sum_i^n e_i}$ | [%] |
| Flexibility factor (FF) | $\frac{\sum_i^n e_{i,off-peak}}{\sum_i^n (e_{i,off-peak} + e_{i,on-peak})}$ | [-] |

3.3.4.1 Design of action-space

The case study considers the problem of optimising a cluster of buildings composed of prosumers, by acting on the charging and discharging processes of the thermal storage in the buildings. More specifically, the control actions are related to chilled water storage, that can be charged with a heat pump and discharged to meet building cooling demand, and a SHW storage, that can be charged by an electric heater. Therefore, each building has two control actions and, depending on the type of architecture considered, the number of controller actions is two (cooperative RL) or eight (coordinated RL).

For each control time step (with a resolution of one hour), the DRL agent selects actions between [-1,1], where -1 represents a complete discharge of the storage system and 1 represents a complete charge. The action-space is then constrained between [-1/3,1/3], to facilitate realistic charging and discharging time, according to

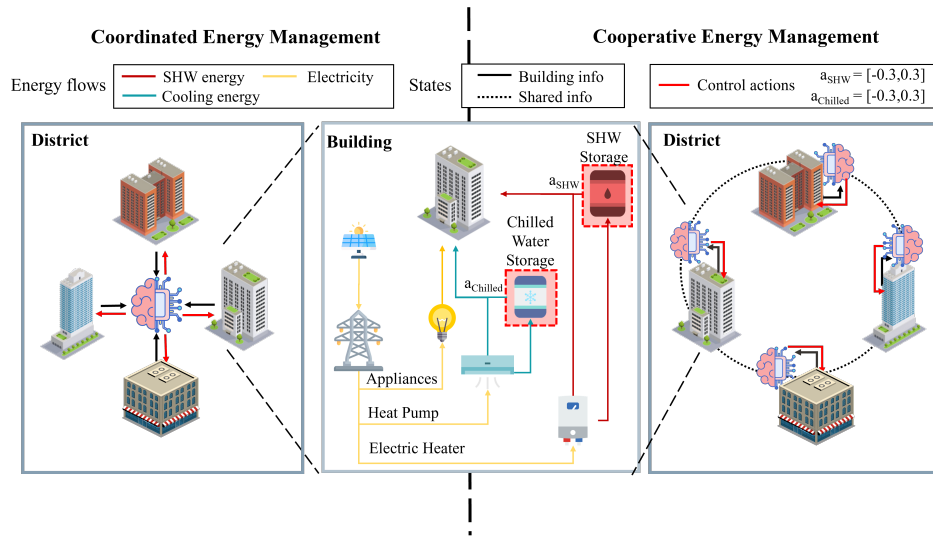


Fig. 3.17 Coordinated and cooperative control architectures [37]

[209]. The action-space is represented by a tuple of eight values for the coordinated controller and four tuples of two values for each of the four cooperative controllers.

3.3.4.2 Design of state-space

The agents learn the optimal control policy, observing the effects of its actions on the environment states. Therefore, the definition of the state-space, together with the reward function, is crucial to help the learning process of the controller and represents one of the points of differentiation between the two architectures. The variables selected by both architectures are reported in Table 3.11 with further commentary below.

The variables can be categorised into weather, district and building states. Both architectures use weather and district variables, while the main difference is related to the building variables. In particular, the coordinated architecture has access to information for all buildings, e.g., by collecting the State of Charge (SoC) of the eight storage devices, exploiting the information to optimally control the buildings. On the other hand, cooperative architecture exploits only the information related to the controlled building, being unaware of the information from other buildings.

Weather variables, such as outdoor air temperature, direct and diffuse solar radiation, were included to account for their influence on the cooling load. For outdoor air temperature and direct solar radiation, both short (1 and 2 hours ahead) and

Table 3.11 State-space description for coordinated and cooperative DRL agents [37]

| Variable | Unit |
|---|---------------------|
| Weather | |
| Outdoor air temperature | [°C] |
| Outdoor air temperature forecast (1, 2, 6 hr ahead) | [°C] |
| Direct solar radiation | [W/m ²] |
| Direct solar radiation forecast (1, 2, 6 hr ahead) | [W/m ²] |
| Diffuse solar radiation | [W/m ²] |
| District | |
| District total load | [kW] |
| Electricity price (c_{el}) | [\$/kWh] |
| Electricity price forecast (1, 2 hr ahead) | [\$/kWh] |
| Hour of day | [h] |
| Building | |
| Non-shiftable load | [kW] |
| Solar generation | [kW] |
| Chilled water storage SoC (state of charge) | [-] |
| SHW storage SoC | [-] |

medium (6 hours ahead) term forecasts were used to exploit the potential predictive capabilities of the controllers. CityLearn considers the weather as estimated with a generic model with a pre-calculated prediction error. In detail, the prediction error increases with the forecast time-horizon for both temperature and solar radiation. The errors start from 2.5% for 6 hours ahead predictions and they increase up to 10% for 24 h ahead. Therefore, the PV generation is evaluated considering a prediction model of solar radiation with a pre-determined accuracy.

Common variables amongst the buildings were included in district states, such as hour of day, electricity price and electricity price forecasts, with a time horizon of 1 and 2 hours ahead, together with the district total electrical load. The states involving information on the specific energy system were categorised as building variables, such as the appliance electrical load (non-shiftable load), the PV electricity production, and the SOC of the cooling and SHW storage devices. As previously explained, for the coordinated architecture, the centralised controller collects these four variables for each building, together with district and weather variables, to find the control strategy. The cooperative architecture, however, exploits only the four states of the controlled building.

3.3.4.3 Design of reward functions

The reward function must be representative of the defined control problem and it assesses the effectiveness of the control policy. In this work, comparable reward functions were defined for the coordinated (Section 3.3.4.3) and cooperative (3.3.4.3) RL controllers, to benchmark their respective performance. Reward function definition is indirectly related to the previously defined KPIs. The KPIs were defined according to the objective functions that the controller had to achieve. However, the results of the training process are only affected by the cumulative values of the reward function, and not by the evolution of the single KPIs. KPIs were evaluated to assess the performance of the control policy in a post-processing phase, after the reward (which includes different contrasting objectives) reached convergence.

Reward function for coordinated RL controller For the coordinated DRL controller, the reward (R) was formulated as a linear combination of three different contributions: the profile flattening term, a cost term and an overproduction term. This is defined as follows:

$$R = \sum_{i=1}^n e_i^2 \times k_1 + \sum_{i=1}^n |\min(e_i, 0)| \times c_{el} \times k_2 + \sum_{i=1}^n |\max(e_i, 0)| \times c_{sell} \times k_3 \quad (3.8)$$

The formulation of the flattening term employs a square factor, that leads to a more homogeneous consumption of the cluster. On the other hand, second and third terms are related to the electricity used/produced from the cluster, with the final goal of reducing operative costs. particular, e_i is negative if the building imports electricity from the grid and positive if the building sells electricity to the grid. For a specific building, these two terms are mutually exclusive, as the last term assumes electricity overproduction, whereas the second term assumes electricity import from the grid. This formulation is used to reduce electricity costs for the buildings (second term) and to increase self-consumption in buildings (third term), penalising it when selling electricity rather than trying to exploit renewable overproduction. Considering that in the SAC framework the magnitude of the reward has effects on the results, the terms k_1 , k_2 and k_3 were defined to maximize the reward, while balancing the flattening term and the economic results. Therefore, these terms were varied, to

achieve optimal trade-off between performance at single building and district scale. The values chosen for the coordinated DRL coefficients (k_1, k_2, k_3) are reported in Table 3.12.

Reward function for cooperative RL controller To allow a fair comparison among the two architectures, the reward of the cooperative DRL controller was formulated as for the coordinated case, using a linear combination of three terms related to the profile flattening, the imported electricity and the self-consumption. The general formulation of the reward (R_i) for building i is as follows:

$$R_i = \sum_{i=1}^n e_i^2 \times k_1 + |\min(e_i, 0)| \times c_{el} \times k_2 + |\max(e_i, 0)| \times c_{sell} \times k_3 \quad (3.9)$$

The main difference with respect to the previous architecture (coordinated RL controller) is related to the self-consumption and cost. While the profile flattening term is similar, the imported electricity and self-consumption terms consider only the controlled building, with the same aim of equation 4 previously described. For example, Building 2 and 4 will never experience the overproduction term, due to the lower capacity of PV panels. The values of the three coefficients (k_1, k_2, k_3) are reported in Table 3.12. It is important to notice that the k_1 term for the two architectures is different. This is because in the coordinated approach, the entire electricity consumption of the district is squared, while for the cooperative approach the electricity consumption of each building is first squared and then summed up for all the buildings. Analysing the two quantities previously described (average district power squared and the sum of squared power of each building) a suitable value of k_1 for each architecture was set.

Table 3.12 Reward function hyperparameter values [37]

| Variable | Coordinated Controller | Cooperative Controller |
|----------|------------------------|------------------------|
| k_1 | -10^{-5} | -10^{-4} |
| k_2 | -5 | -5 |
| k_3 | -350 | -350 |

As mentioned in Section 3.3.2, DRL algorithms are characterised by several hyperparameters, that directly influence controller performance. These parameters

need to be tuned according to the specific control problem and they can be further divided into RL hyperparameters and control problem related hyperparameters. To obtain a fair benchmark among the two controllers, RL hyperparameters (decay rate, temperature coefficient, learning rate) were subjected to a hyperparameter optimization, the results of which are reported in A.2 to promote the reproducibility of the analysis. To perform hyperparameter optimization, a grid-search process was used, exploring the search space completely. However, prior to that, domain expertise knowledge and previous experiences were used to constrain the possible search space. Moreover, control problem hyperparameters include the episode length, the starting period of learning and the training episodes, on which a specific analysis was performed. Figure A.1, reported in A.2, shows the evolution of the reward function with the number of episodes. To account for stochasticity the mean and standard deviation of 15 simulations were used. It is possible to notice that after the environment initialization and around 3 episodes the reward function stabilizes. Furthermore, as the number of episodes grows, the standard deviation of the coordinated architecture tends to increase, while the standard deviation of the cooperative architecture remains stable. Therefore, the analysis of the results suggested that a trade-off between simulation period and variance can be found at around 5 episodes, selected for the work. The controllers were then tested over 3 months (an episode) using the three climates described in Table 3.7.

3.3.5 Results

This section describes and analyses the results obtained from the implementation of the two DRL architectures, comparing them with the benchmark RBC strategy. Section 3.3.5.1 describes the results of the deployment of both controllers for climate zone 2A (Table 3.7). More specifically, the financial cost accruing to single users and to the district are described, highlighting how a part of the total cost is related to the district peak, and how the different architectures influence the latter. Following this, the attention is shifted towards the district load, with a focus on storage operation and self-consumption, quantifying the results of cooperative and coordinated approaches at this level. Moreover, the section compares the different use of energy under the control strategies and quantifies the advantages based on several KPIs. Lastly, Section 3.3.5.2 presents a summary of the results for deployment for other climates than that outlined in (Table 3.7) based on the same KPIs.

3.3.5.1 Comparison with baseline RBC

Figure 3.18 shows the energy consumption costs for each building (left) and the energy consumption and peak load (penalty) costs for the district (right). Results are presented for the 3 months simulation period. As it can be seen, both coordinated and cooperative RL result in a lower cost at the district level, namely 3% and 7% savings, respectively. However, the main difference between the two approaches is related to single building costs. For the coordinated approach, Building 2 and 4 experience a cost increase in comparison to RBC strategy of 4% and 3%, compensated by the reduction of the peak term. On the other hand, cooperative architecture shows a cost reduction for each building, leading to greater overall savings at the district scale.

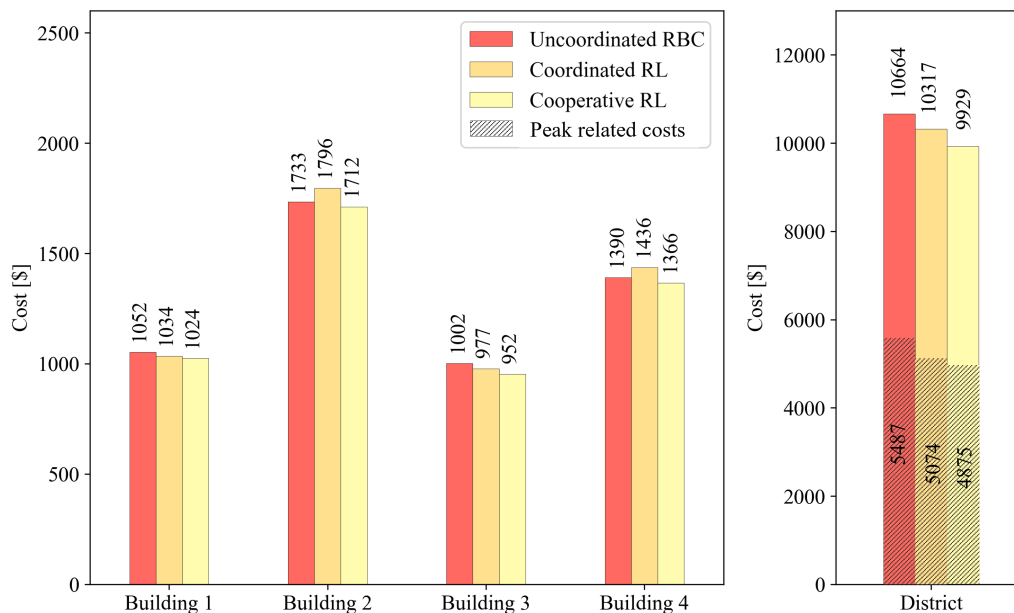


Fig. 3.18 Cost related to the energy term for each building (left) and total district cost, sum of energy and peak terms (right), for the different control strategies over the entire simulation period [37]

To analyse further the basis for these results, Figure 3.19 shows the district electrical load evolution with the three control strategies for three days during the first week of June. This figure highlights the contribution of both PV self-consumption and PV export. Figure 3.19 a) shows how the uncoordinated RBC approach leads to demand peaks during the night due to the charging of the storage devices during these periods, while discharging them during the day, exporting the overproduction

of renewable electricity around 12 p.m., June 1. On the other hand, Figure 3.19 b) shows the coordinated approach, which tries to exploit PV production as well as flatten the load profile. Lastly, Figure 3.19 c) displays the cooperative approach, in which buildings try to reduce peak consumption, as at around 6 a.m., June 1, and maximize self-consumption, which can be attributed to a reduction of electricity export of Building 1 and Building 3 around 12 p.m., June 1.

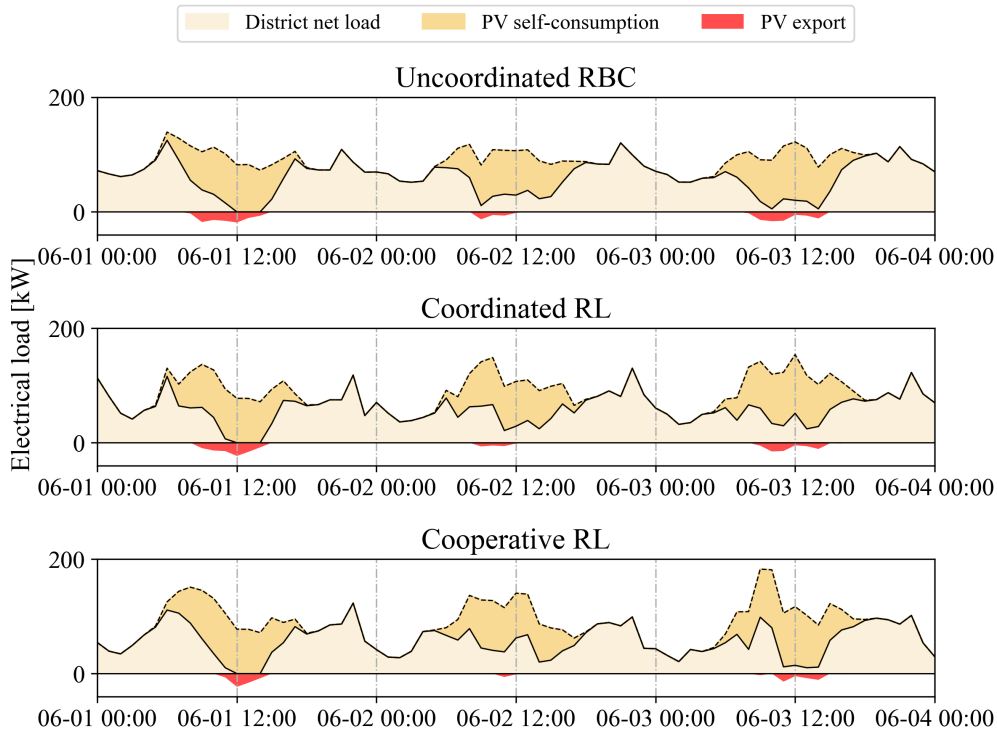


Fig. 3.19 District electrical load profile for each control strategy during a a three-days period [37]

To understand the differences between the two proposed control strategy and the baseline, a detailed comparison is provided for Building 1 in Figure 3.20 for three days. The plotted variables are normalised with respect to their maximum values and include: the state of charge (SOC) for the cooling and SHW storage, the solar radiation, the outdoor temperature and the electricity price. These variables have been selected to highlight the behaviour of an optimal control strategy. Indeed, the best control policy for a prosumer aims at maximising self-consumption, exploiting the lower electricity price and resulting in the minimum district peak demand. To achieve such objectives, both coordinated and cooperative controller shift the charge between the two storage (TES (cooling) and SHW) devices, thereby flattening

building electrical load. In particular, they tend to charge the chilled water storage during the night, exploiting the lower ambient temperatures (higher COP) and the SHW TES during the day, to use possible PV over-production. The two storage devices are discharged during high-electricity price periods and low periods of PV production, to obtain a flatter profile.

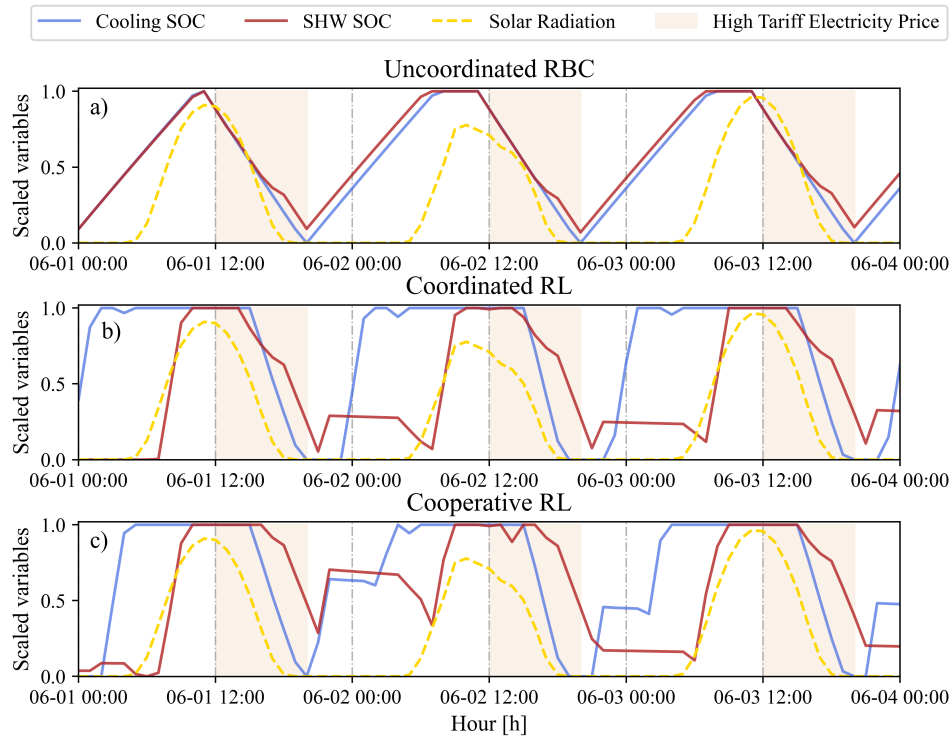


Fig. 3.20 Comparison of control strategies for Building 1 [37]

Furthermore, to assess the ability of the controller to adapt to weather conditions and grid requirements, the mean values and standard deviations of SOC for storage devices for a single day period, averaged over the entire season have been shown in Figure 3.21. It can be noticed how both RL controllers have a higher standard deviation for TES state of charge compared to RBC, explained by noticing the variability of SHW and weather conditions, that strongly affect cooling demand. It is important to highlight that the optimal control strategy should not be searched by looking at mean values, since the control actions of a specific building depend on weather conditions, electricity price, building load and grid requirements, which in turn are influenced by other building control actions. However, Figure 3.21 can be

used to understand how much optimal control actions can be influenced by external factors.

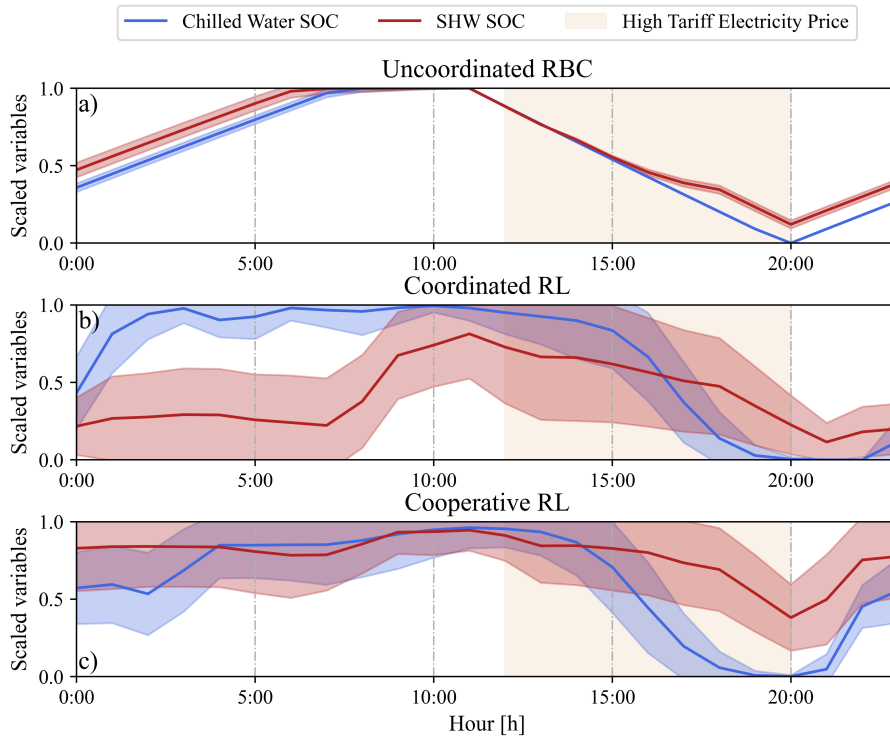


Fig. 3.21 Daily average hourly scale profiles of SOC with relative standard deviations for the three control strategies in Building 1 [37]

Figure 3.22 reports the evolution of exported electricity at district level for the two controllers and the baseline over the entire simulation period (3 months). Although the absolute exported quantities only represent a small percentage of the total district consumption, their comparison can provide insights into the effectiveness of the control strategies, since minimization of exported electricity is one of the most effective ways to reduce costs. Given that one of the objectives of the RL controllers is to minimize exported electricity and considering Figure 3.22, it can be observed how the uncoordinated RBC is outperformed by the two proposed control strategies. In particular, the cooperative RL reduces the electricity sold to the grid by approximately a quarter compared to the coordinated controller, consequently increasing savings.

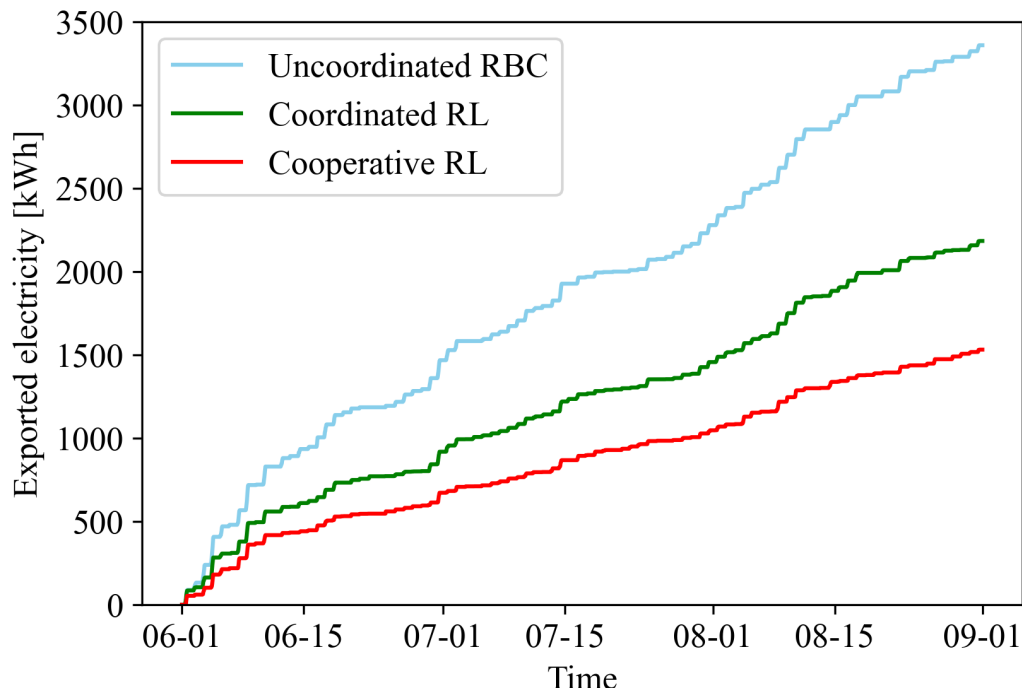


Fig. 3.22 Comparison of district cumulative exported electricity between control strategies over the entire simulation period (3 months) [37]

To relate the storage operation with the controller performance benefits, Figure 3.23 shows the electricity consumption at a district level for the entire simulation period (3 months) as follows: i) on-peak periods with direct building consumption; ii) off-peak periods with direct building consumption; iii) PV production and associated self-consumption, and; iv) storage discharge (either from the grid or PV) and used to charge either the cooling and SHW storage. Furthermore, to assess the contribution of storage for the integration of renewable energy sources, the bar plot also includes results considering the absence of storage (No Storage), which results in self-consumption from PV being halved compared to the RBC case. The effectiveness of the RBC strategy is evident by examining the consumption reduction during on-peak periods with respect to the No Storage scenario. Despite slightly increasing the amount of on-peak period electricity consumption, the coordinated controller further increases the advantages with respect to RBC, reducing off-peak electricity optimally using thermal storage, leading to cost savings. These advantages are even greater for the cooperative controller, which shows a slight increase of

self-consumption and the highest storage operation, emphasising the role of storage towards the optimal energy management of the district.

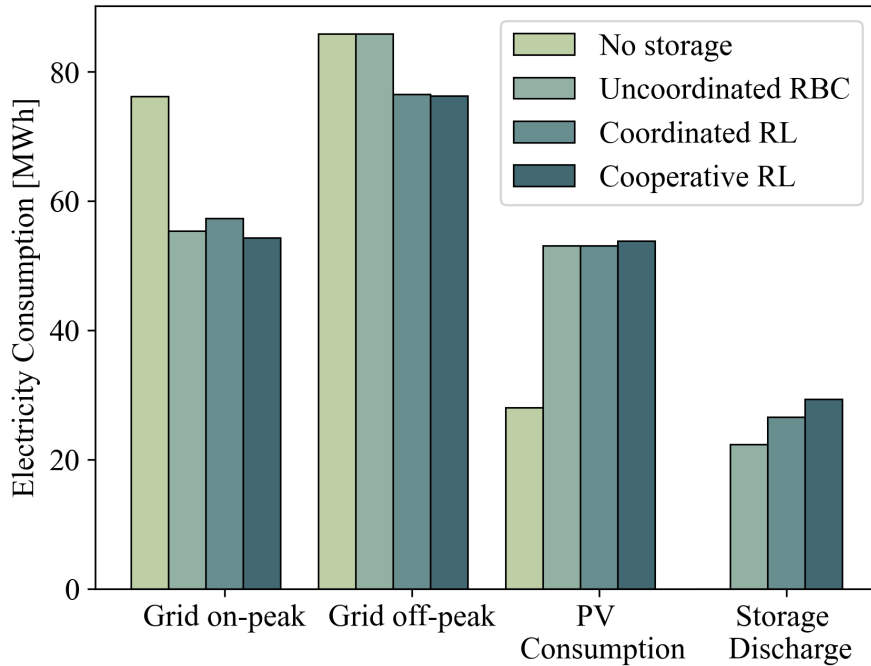


Fig. 3.23 District energy disaggregation comparison over the entire simulation period (3 months) [37]

Lastly, to analyse the performance of the three controllers, Table 3.13 summarizes the values assumed by the KPIs to assess the benefits provided by the two RL architectures for the entire simulation period (3 months). The table also shows different KPIs for the RBC, used as a benchmark, while displaying the same KPIs for the coordinated and cooperative architectures with relative improvement (or worsening) in square brackets. For all but the last two KPIs, a lower value indicates a better control policy. Therefore, it is clear that the cooperative architecture outperforms the coordinated architecture especially for the daily-Peak and daily-PAR, where the coordinated controller performs worse than RBC. The coordinated controller can reduce costs and peaks with respect to the RBC of around 3% and 10%, respectively. Examining the flexibility factor, it can be seen how both RL controllers perform worse than the RBC. However, the flexibility factor KPI was lower for the DRL controllers because of the decreasing use of off-peak tariff energy consumption.

Table 3.13 Results of the MARL controllers deployed on Climate 2A (performance improvement in brackets) [37]

| KPI | Climate 2A | | |
|---------------------------------|------------|---------------|---------------|
| | RBC | Coordinated | Cooperative |
| Cost [\$] | 10663 | 10311 [-3.3%] | 9927 [-6.9%] |
| Peak [kW] | 171 | 154 [-9.7%] | 147 [-13.8%] |
| Daily-Peak [kW] | 123 | 125 [+2.0%] | 109 [-11.2%] |
| Peak-to-average ratio (PAR) [-] | 2.31 | 2.13 [-7.7%] | 2.05 [-11.2%] |
| Daily-PAR [-] | 1.66 | 1.72 [+4.2%] | 1.51 [-8.5%] |
| Self-sufficiency [%] | 0.240 | 0.243 [+1.6%] | 0.248 [+3.5%] |
| Flexibility Factor (FF) [%] | 0.66 | 0.62 [-5.7%] | 0.64 [-2.0%] |

3.3.5.2 Deployment of RL controllers for different climates

Table 3.14 and Table 3.15 report the results of the deployment in climates 3A and 5A, comparing the performance of the two RL controllers with respect to RBC. These two climates are characterised by a lower temperature and solar radiation, thus requiring less cooling energy in the summer period, as highlighted by the lower costs. The analysis of the three tables presented has the role to study the robustness of the controllers, here highlighted by the use of KPIs at a different time horizon (Peak, Daily-Peak) as well as examining the adaptability of the architectures to different climates. It can be observed that the cooperative approach exhibits better performance in climate zones 2A and 3A, achieving significant advantages (7% and 4%, respectively) in terms of economic costs, while the coordinated architecture performs slightly better in climate zone 5A. These results can be explained noticing that climate 5A is characterised by lower external temperature and solar radiation, which strongly reduce the need for cooling energy, limiting the flexibility provided by the chilled water storage and the RL control strategy. In general, both architectures achieve better performance with respect to RBC, despite their effectiveness being greatly influenced by climatic conditions. In particular, the main drivers of the problem, district costs and peak, are similar to RBC values shifting the controller from climate 2A to climate 5A, while the controller retains substantial improvement for daily-values, highlighting its stability.

Table 3.14 Results of the MARL controllers deployed on Climate 3A (performance improvement in brackets) [37]

| KPI | Climate 3A | | |
|---------------------------------|------------|---------------|---------------|
| | RBC | Coordinated | Cooperative |
| Cost [\$] | 10258 | 10237 [-0.2%] | 9806 [-4.4%] |
| Peak [kW] | 179 | 174 [-2.4%] | 156 [-12.5%] |
| Daily-Peak [kW] | 121 | 117 [-2.6%] | 106 [-11.7%] |
| Peak-to-average ratio (PAR) [-] | 2.61 | 2.6 [-0.3%] | 2.34 [-10.1%] |
| Daily-PAR [-] | 1.77 | 1.76 [-0.5%] | 1.60 [-9.4%] |
| Self-sufficiency [%] | 0.250 | 0.255 [+2.2%] | 0.258 [+3.5%] |
| Flexibility Factor (FF) [%] | 0.65 | 0.616 [-5.2%] | 0.623 [4.1%] |

Table 3.15 Results of the MARL controllers deployed on Climate 5A (performance improvement in brackets) [37]

| KPI | Climate 5A | | |
|---------------------------------|------------|---------------|---------------|
| | RBC | Coordinated | Cooperative |
| Cost [\$] | 8946 | 8856 [-1%] | 8874 [-0.8%] |
| Peak [kW] | 150 | 145 [-3.3%] | 145 [-2.7%] |
| Daily-Peak [kW] | 111 | 98 [-11.7%] | 99 [-10.6%] |
| Peak-to-average ratio (PAR) [-] | 2.42 | 2.4 [-0.6%] | 2.42 [+0.2%] |
| Daily-PAR [-] | 1.8 | 1.63 [-9.3%] | 1.65 [-7.9%] |
| Self-sufficiency [%] | 0.270 | 0.275 [+2.1%] | 0.275 [+1.9%] |
| Flexibility Factor (FF) [%] | 0.69 | 0.645 [-6.4%] | 0.649 [-5.9%] |

3.3.6 Discussion

Grid-interactive buildings can exploit energy flexibility to increase the energy efficiency of each building and provide advantages to the grid, with a key role in the energy transition. This research aims to exploit different DRL architectures to enhance the energy grid-interaction for a district of buildings. The DRL controllers were designed to act on building active thermal storage systems, to exploit energy flexibility, minimising the energy cost for both the individual buildings and the entire district. Moreover, the problem involved time-varying electricity tariffs, including a peak-related term, to incentivise a rational use of electricity amongst the different buildings and to favour cooperation and coordination. To assess the performances of the two DRL control architectures, an uncoordinated RBC was introduced as a baseline, due to the widespread use of this strategy for thermal storage control and to provide a fair comparison between the different RL architectures. The same

information (state-space) was provided to each controller (with the only exception of information specifically related to that architecture). Moreover, the reward function formulation was also conceived with the same objectives, reducing imported electricity and demand peaks and increasing self-consumption.

The control problem was formulated allowing the DRL controllers to exploit forecast information about electricity price and weather for searching the optimal policy. However, despite SAC DRL use of historical data to speed up the training process, the interaction between different buildings, requires a simulation environment for the training of the controllers. Some key observations for the application and scalability of DRL controllers are related to their computational cost and robustness. Considering that the coordinated architecture scales exponentially with the number of buildings, while the cooperative architecture scales linearly, a cooperative architecture may represent the best solution, but as the number of buildings increases, the non-stationarity of the environment can influence the stability of the cooperative control policy. The present work tried to reduce some of the variability associated with DRL controllers performing hyperparameter optimization, adopting a similar reward function and studying the evolution of the cumulative reward with episodes. However, as highlighted by Figure A.1, the inherent stochasticity of the coordinated architecture is higher with respect to cooperative architecture. After the training period, the two controllers achieved superior performance compared to the RBC and took advantage of their predictive nature to flatten the load profile, reducing maximum peak and consequently cost. Table 3.13 demonstrates the advantages of the cooperative controller over the coordinated controller, particularly when considering daily peaks (11% reduction of the cooperative controller compared to a 2% increase of the coordinated controller) and the reduction of energy costs (7% reduction compared to 3%). Moreover, the two RL controllers differ due to PV self-consumption, which is slightly higher for the cooperative controller.

The reward function formulation plays a crucial role to achieve specific objectives, therefore trade-offs between different terms should be carefully considered. In this perspective, the cooperative architecture is more flexible to the formulation of the reward function, designed to represent user needs in particular, while the coordinated architecture should be defined to achieve high-level performance, averaging over single building requirements. For the specific application considered in this paper, cooperative controller proved to perform better since it was able to search for a better control policy oriented to the maximization of self-consumption. On the

other hand, in a coordinated architecture, the results obtained for Building 2 and 4 suggest that, despite reducing district costs, some users may experience increased costs, discouraging them from participating in this type of control. Based on this result, we concluded that in heterogeneous context, with different energy systems and users' needs, a cooperative architecture can be more flexible. Furthermore, the work highlighted that despite the relation between the reward function with some of the KPIs, the multi-objective nature of the problem and the different scales analysed makes important to analyze KPIs in addition to the cumulative reward. Indeed, looking only at the reward function as a performance indicator, the analysis could lack information about the costs faced by individual buildings, as in the case of the coordinated controller. To test the robustness of the learned optimal policy for both architectures, the controllers were deployed in two other climates. Table 3.14 and Table 3.15 highlight that, despite both controllers performing better than the RBC, the deployment conditions can highly affect maximum peak and PAR, while they do not influence daily controller performances on average (Daily-Peak and Daily-PAR).

3.3.6.1 Limitations

A key concern about the comparison between the architectures is whether the conclusions drawn from the current case study can be generalised. For instance, it should be noticed that for Climate 5A, the performance of the coordinated controller is marginally better than that of the cooperative controller, not allowing to identify a superior alternative among cooperative and coordinated architectures. Furthermore, the comparison between the two architectures is influenced by the hyperparameter settings, the number of training episodes, the formulation of the reward function, the inherent stochasticity of DRL and the case study itself. As a consequence, the findings can not be considered generalised and thus need further investigations. The study had the aim to produce a fair comparison among the architectures, using the same hyperparameters, except for a number of neurons related to the state-action space. Moreover, also the reward function was conceived to have the same structure, despite the different information the controller exploits. Lastly, the work aimed to analyse the effect of the two control strategies for the buildings in the district and the district itself. The computational comparison of the two algorithms was beyond the scope of the thesis and may represent a limitation that will be addressed in a future work. However, the influence of the number of buildings on the computational cost

and the effectiveness of the control strategies is important to be assessed especially when different architectures are compared.

Chapter 4

3DEM: A methodology to combine data-driven models and controllers

Previous chapters described the effectiveness of data-driven controllers, explaining how to scale them. The previously introduced applications also highlighted the importance of exploiting energy flexibility, but neglected the control of HVAC systems, which represent one of the highest energy consumption in buildings. This chapter firstly reviews how data-driven models can support energy management, and then introduces an application that makes use of data-driven models to represent building thermal dynamics in multiple buildings, coupling them with a DRL controller to leverage the different sources of flexibility within the district. This application aims to summarize the lessons learned during the previous chapter to demonstrate how both controllers and models can be integrated and used at scale to perform data-driven energy management in buildings.

Portions of the present Chapter were already published in the following scientific papers:

- Giuseppe Pinto, Davide Deltetto, and Alfonso Capozzoli. Data-driven district energy management with surrogate models and deep reinforcement learning. *Applied Energy*, 304:117642, 2021 [36]
- Giuseppe Pinto, Riccardo Messina, Han Li, Tianzhen Hong, Marco Savino Piscitelli, and Alfonso Capozzoli. Sharing is caring: An extensive analysis

of parameter-based transfer learning for the prediction of building thermal dynamics. *Energy and Buildings*, page 112530, 2022 [210]

4.1 Motivations and novelty of the proposed approach

The next subsections provide a literature review on machine learning techniques for load and thermal dynamic prediction models, that can be used to support energy management in buildings. Then, it discusses the motivations and the novelty of the proposed approach.

4.1.1 Building load prediction models

Building load prediction has received a lot of interest as it is used in different ways for increasing building energy efficiency. Load prediction is particularly important in grid-interactive and energy efficient building models for two main reasons:

- It is a crucial component for advanced controllers such as MPC and RL controllers since information on load evolution can be leveraged to optimize energy systems.
- Load prediction is critical in building-grid integration such as demand response and transactive load control, facilitating the interaction between the grid and the building-side or demand-side.

The data-driven load prediction is a regression problem, therefore machine learning techniques have been widely applied in this field, with neural networks standing out since the beginning of 2000s [211].

Among machine learning techniques, artificial neural networks (ANN) and Support Vector Machine (SVM) were the two most widely used techniques for building load prediction [212, 213]. He [214] used Convolutional Neural Network components to extract rich features from historical load sequences and use Recurrent Components to model the implicit dynamics. Results showed good prediction performance on large building datasets. Marino et al. [215] used Long Short Term Memory (LSTM) neural networks to predict a residential building load at two granularities, one-minute and one-hour timestep. Furthermore, datasets like Building

Data Genome Project [216] have been used as a base to study the performances of different machine learning algorithms, including several NN architectures.

In thermal energy prediction, numerous researchers employed neural networks. Li et al. [217] compared ANN and SVM to predict space cooling load in an office building. Mihalakakou et al. [218] studied the prediction of space heating and cooling loads using different neural network architectures and assessed the importance of lagged inputs for NN performance. Aydinalp et al. [219] used ANN to predict the domestic hot water consumption in a stock of Canadian residential buildings. Wang et al. [220] applied 12 data-driven models with the aim to predict the heat load of a single building. LSTM and eXtreme Gradient Boosting (XGBoost) resulted the best, respectively for short-term load prediction and long-term load prediction. Ben-Nakhi and Mahmoud [221] used a NN to predict next-day building cooling load to optimise the HVAC thermal energy storage system operation.

4.1.2 Building thermal dynamic models

Among the first applications, Ruano et al. [222] explored the use of a radial basis function neural network to predict the indoor air temperature of a public building. Sun et al. [223] proposed a multiple linear regression model to predict the supply temperature of a district heating network, adjusting it according to actual indoor temperature deviation. Shi et al. [224] used a back-propagation neural network to predict indoor relative humidity and air temperature with different forecasting horizons. Kusiak and Xu [225] proposed a dynamic neural network to relate HVAC energy consumption with indoor temperature evolution, optimizing the control strategy of the HVAC system with a data-driven approach. Similarly, nonlinear [226, 227] autoregressive neural networks for indoor temperature prediction were integrated with controllers to optimize the HVAC systems. More recently Huang et al. [228] implemented a predictive controller coupled with a neural network able to estimate the indoor air temperature of a multi-zone building, to optimize the start and stop of an HVAC system, while Drgona et al. [229] exploited neural networks and regression trees to construct an approximate model predictive controller.

Recently, a large interest was devoted to the application of LSTM for thermal dynamic prediction. In [230] an LSTM neural-network was employed to predict the indoor air temperature in a multi zone building. Xu et al. [231], compared two

LSTM models to predict indoor temperature evolution one step ahead and multi-time step ahead, studying the advantages of using an error correction for multi-time step ahead. Ellis et al. [232] used an Encoder-Decoder LSTM to describe the dynamic of an air-handling unit with variable air volume relating it to the indoor temperature evolution, coupling the information with a model predictive controller (MPC) to reduce energy costs. Fang et al. [233] proposed three LSTM-based sequence to sequence model architectures to perform a multi-step ahead indoor air temperature forecasting: a LSTM-Dense model, a LSTM-LSTM model and a LSTM-dense-LSTM model, evaluating the performance under different forecast horizons. The results and analyses showed that the LSTM-dense model performs better for shorter forecast horizons, while the other two are more suitable for longer forecast horizons. Recently, a new paradigm in neural network was introduced with physics-informed neural networks (PINNs) [234]. These neural networks are trained to solve supervised learning tasks while respecting any given laws of physics described by general nonlinear partial differential equations, combining the advantages of white-box modeling with black-box modeling. However, despite the interest in this field, only few works explored PINNs in the domain of building energy control. Bunning et al. [235] compared physics-informed Autoregressive-Moving-Average with Exogenous Inputs (ARMAX) models to Machine Learning models based on Random Forests and Input Convex Neural Networks. In [236] a physics-informed neural network was used to predict temperature evolution in a building, increasing sample-efficiency of neural network and performances for longer prediction horizons. The authors of [237] introduced a physics-constrained recurrent neural network (RNN) to model the thermal dynamics of buildings constraining the eigenvalues of the model, and using penalty methods to impose physically meaningful boundary conditions to the learned dynamics. Di Natale et al. [238] proposed a physics-informed NN that predicts indoor air temperature with respect to different control inputs, zone-zone, and outdoor-zone air temperature differences. However, despite [236] proving a greater sample efficiency of PINNs, they still need a lot of data and physics knowledge.

The previous chapters reviewed the literature of both data-driven models and controllers, identifying the following gaps:

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 the local production side, considering pre-computed ideal building energy demand. This approach disregards assessing user comfort and exploiting 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 environments. To overcome the current limitations of district energy management, this chapter proposes a fully data-driven framework to coordinate multiple energy systems (heat pumps and thermal storage) for a group of four buildings modeling the building thermal dynamics and the indoor temperature evolution using deep neural networks (DNN).

To this purpose CityLearn, previously described, 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 chapter can be summarized as follows:

1. Several LSTM neural networks were developed to predict the indoor temperature evolution of different buildings to reduce 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 chapter is organised as follows: Section 4.2 describes the case study and the control problem. Section 4.3 introduces the proposed methodological framework, while Section 4.4. describes the implementation of the methodology, with particular attention to the training process and controller design. Section 4.5 presents the results of the training and deployment phase, while a discussion of the results is given in Section 4.6.

4.2 Case study and control problem

The proposed methodology, described in detail in the 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 4.1 shows a schematic of the control architecture with detail on energy systems for a representative building. A heat pump serves for both space heating and cooling, with the possibility to charge the cold storage and/or to directly supply cooling energy to control indoor temperature, while electric heater and hot storage meet DHW demand.

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 (8p.m. - 7 a.m.) to 0.06605 \$/kWh during on-peak day-time period (7 a.m. - 8p.m.). Table 4.1 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.

Figure 4.2 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

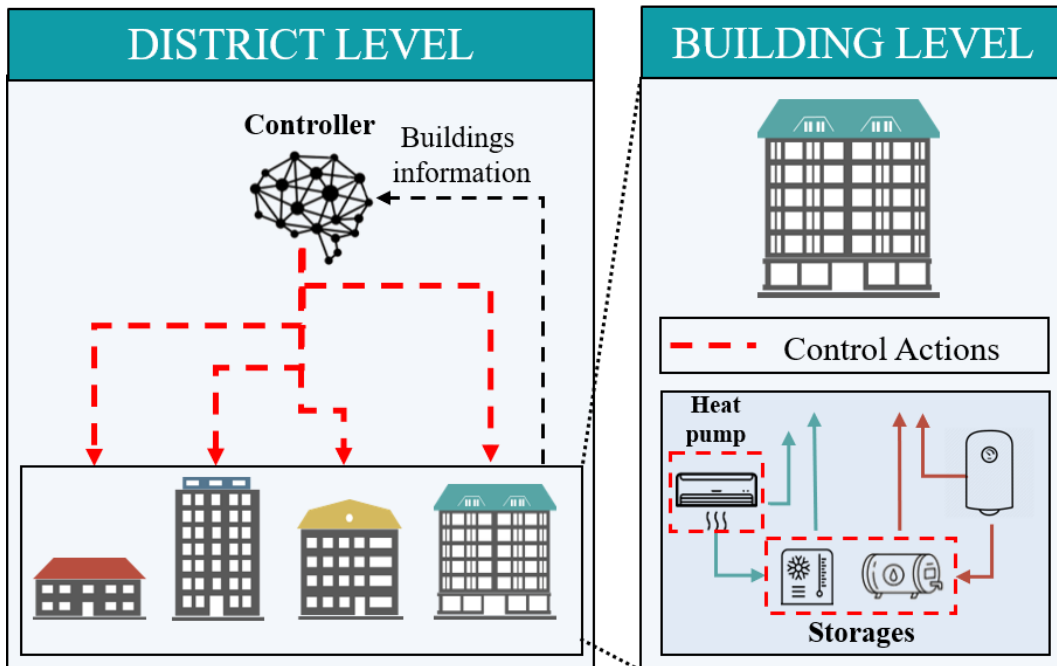


Fig. 4.1 Schematic of the district energy management and controlled energy systems [36]

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).

4.3 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 4.3 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 modeling 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 it to work only with a pre-computed building energy

Table 4.1 Building and energy systems properties [36]

| | Type | Surface [m^2] | Volume [m^3] | Heat Pump | Cold Storage | Hot Storage | PV Capacity |
|------------|------------------|----------------------|---------------------|------------------|-------------------|-------------------|----------------|
| | | | | Capacity [kW] | Capacity [kWh] | Capacity [kWh] | [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 |

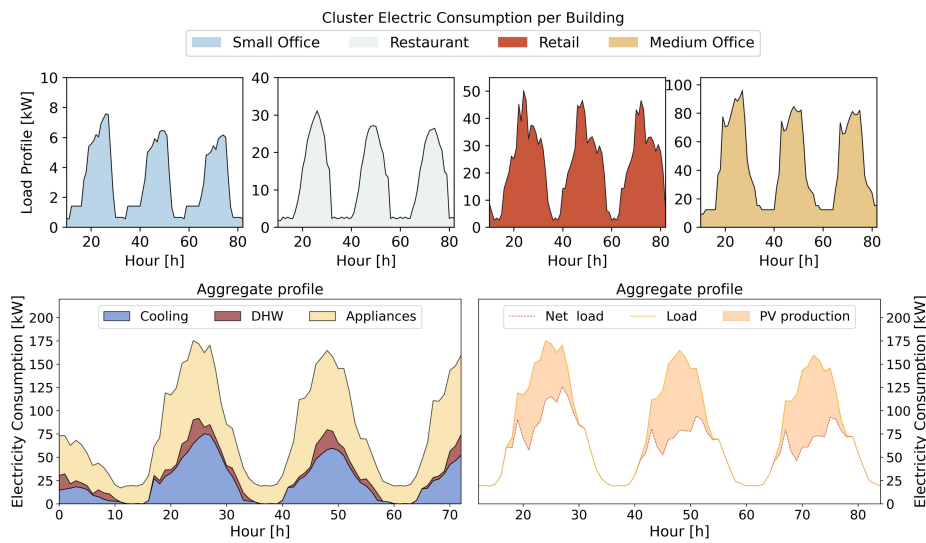


Fig. 4.2 Electrical load profile for each building (up) and electrical load profile and PV production for the cluster of buildings (down) [36]

demand. 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 learned how to control indoor temperature in the different buildings, coordinating heat pump and storage operations to reduce costs and peak demand.

4.3.1 Development of artificial neural networks

To generate a labelled dataset for training and testing LSTM models, the four buildings of the cluster were preliminary modeled and simulated through EnergyPlus.

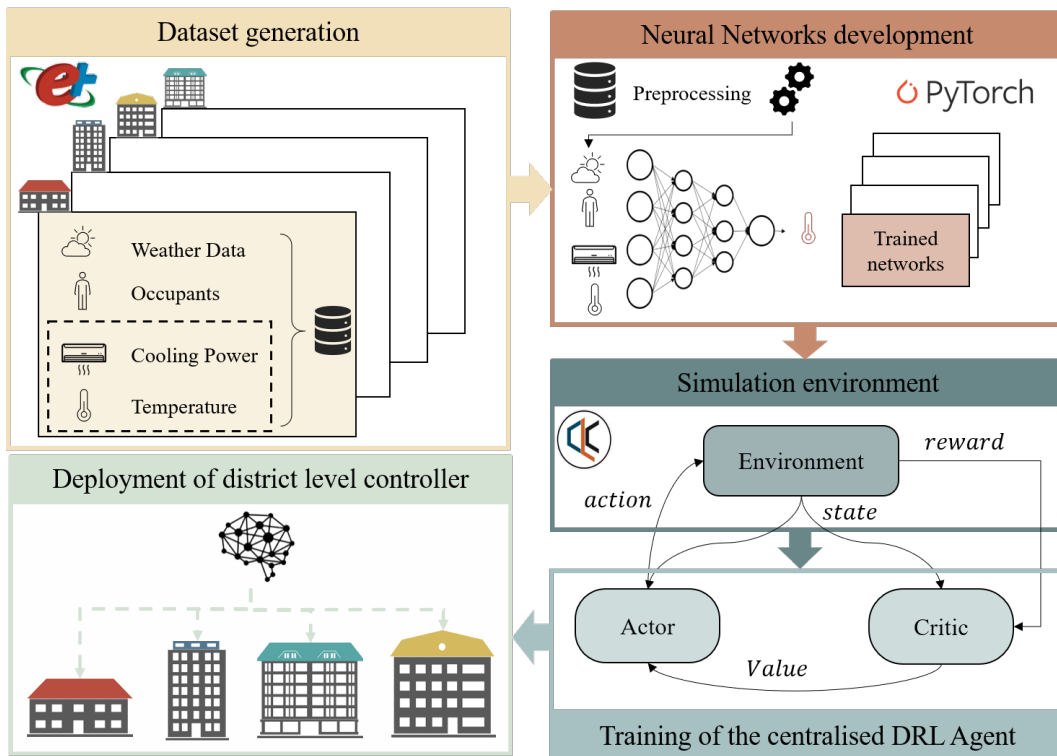


Fig. 4.3 Proposed framework for the district energy management [36]

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 11,520 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 4.4 were used as input variables of the DNNs to predict indoor temperature for each building. In detail, 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. The 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. To select the best architecture for each LSTM model, different hyperparameters were analysed. A sensitivity analysis

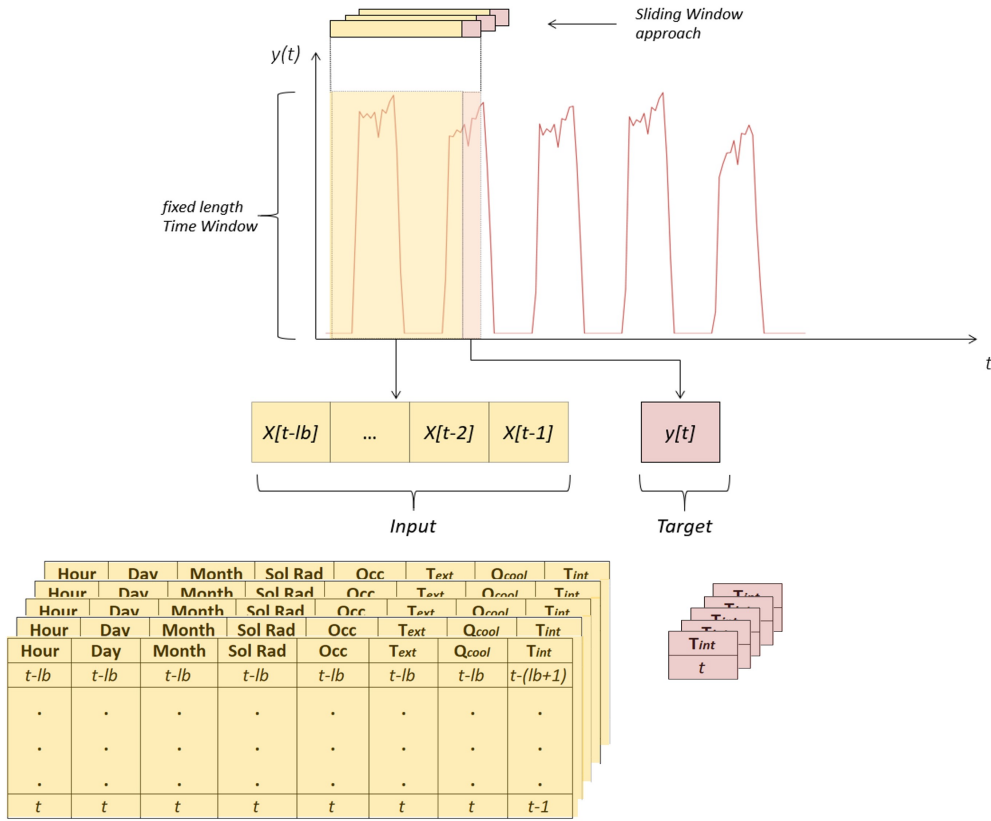


Fig. 4.4 Proposed framework for the district energy management [36]

was performed changing batch size, number of hidden neurons and layers, lookback and learning rate iteratively and finally selecting the best set of parameters for each building that led to the highest accuracy during testing after a training period of 100 epochs. The accuracy was evaluated by computing the following metrics:

$$RMSE = \sqrt{\left(\frac{1}{n}\right) \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (4.1)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (4.2)$$

$$CV - RMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|^2}}{\bar{y}_i} \quad (4.3)$$

Table 4.2 DNN hyperparameters for each building model [36]

| | 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 |

The selected parameters resulted from the sensitivity analysis for each neural network are reported in Table 4.2. It can be seen how all the DNNs share the same value for batch size and lookback period, while according to the specific building the number of hidden neurons, layers, and learning rate changes. For example, the medium office building, which has multiple zones and more complex dynamics, has only 1 LSTM layer with a higher number of neurons and a lower learning rate, while the other 3 buildings share the same DNN architecture, with a different learning rate for the retail.

4.3.2 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 4.5.

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. 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. All the information on the data, the code and the

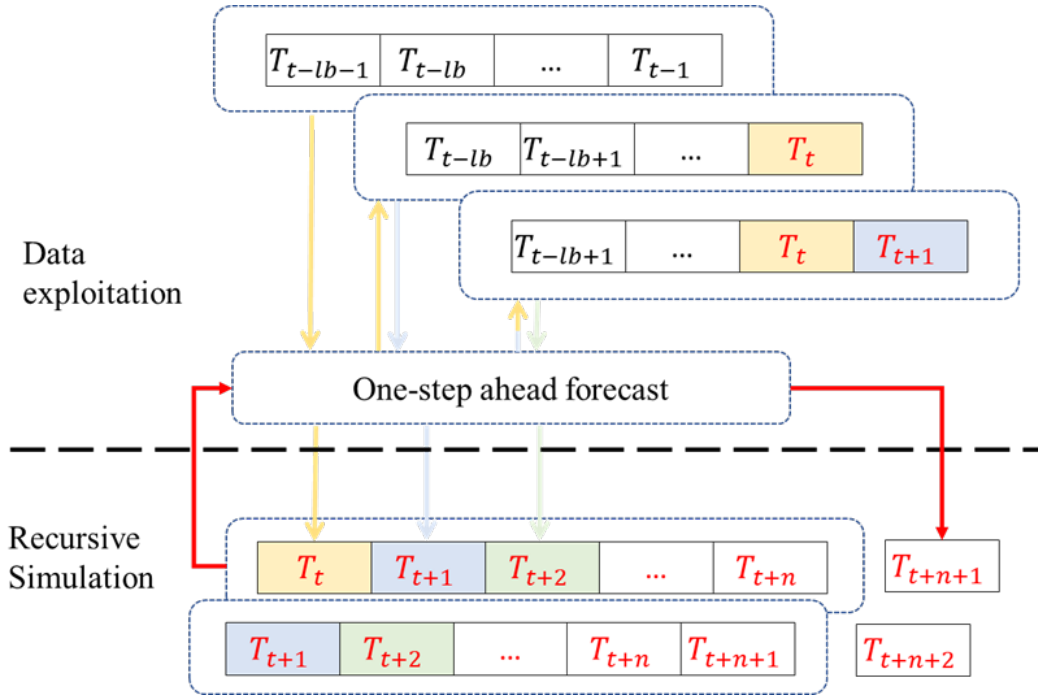


Fig. 4.5 Proposed framework for the district energy management [36]

results produced by the novel framework introduced within the thesis are open-source and available at the following link: <https://github.com/baeda-polito/3DEM>.

4.3.3 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.

4.3.4 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 [59]. 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).

4.4 Implementation

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

4.4.1 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.

4.4.2 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.

4.4.2.1 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 h according to [60].

4.4.2.2 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 4.3 and further described below.

The variables are classified as weather, district and building related variables. Weather information such as 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

Table 4.3 State-space variables [36]

| 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 | [-] |

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 through the LSTM model during the previous hour (ΔT Setpoint - LSTM indoor temperature @t-1) were introduced, together with occupancy information. Figure 4.6 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.

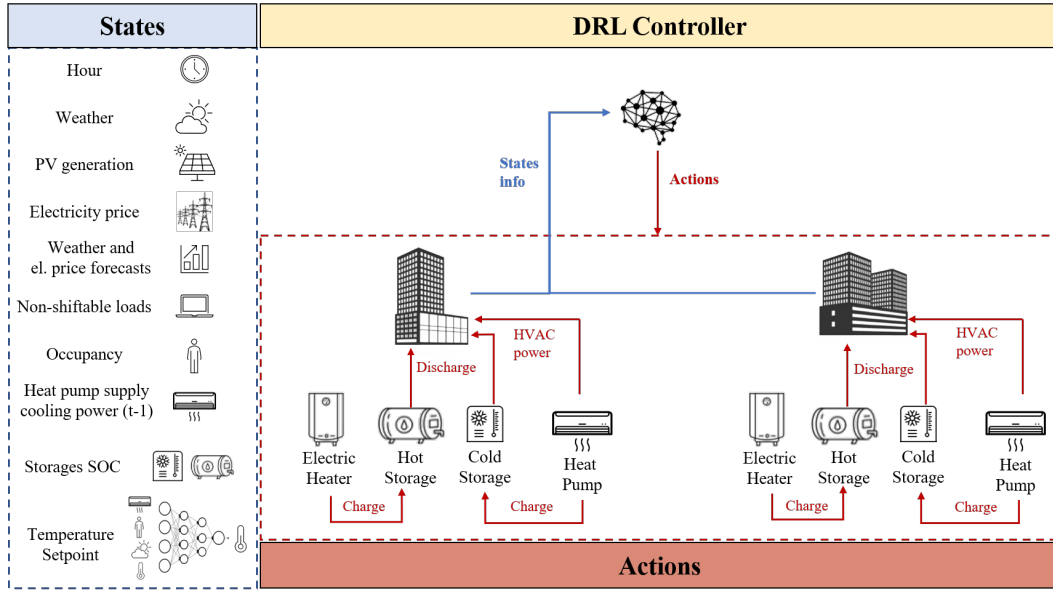


Fig. 4.6 State and action spaces of the DRL control strategy [36]

4.4.2.3 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, the 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:

$$R = \sum_{i=1}^n Comf_p + \sum_{i=1}^n Storage_p + Peak_p \quad (4.4)$$

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 was structured as follows:

$$Comf_p = \begin{cases} -m(SP - T_{in})^3, T_{in} < T_{low} \\ -m(SP - T_{in}), T_{low} \leq T_{in} < SP \\ 0, SP \leq T_{in} < T_{up} \\ -m(SP - T_{in})^2, T_{in} \geq T_{up} \end{cases} \quad (4.5)$$

The comfort term of the reward, shown in Figure 4.7, 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.

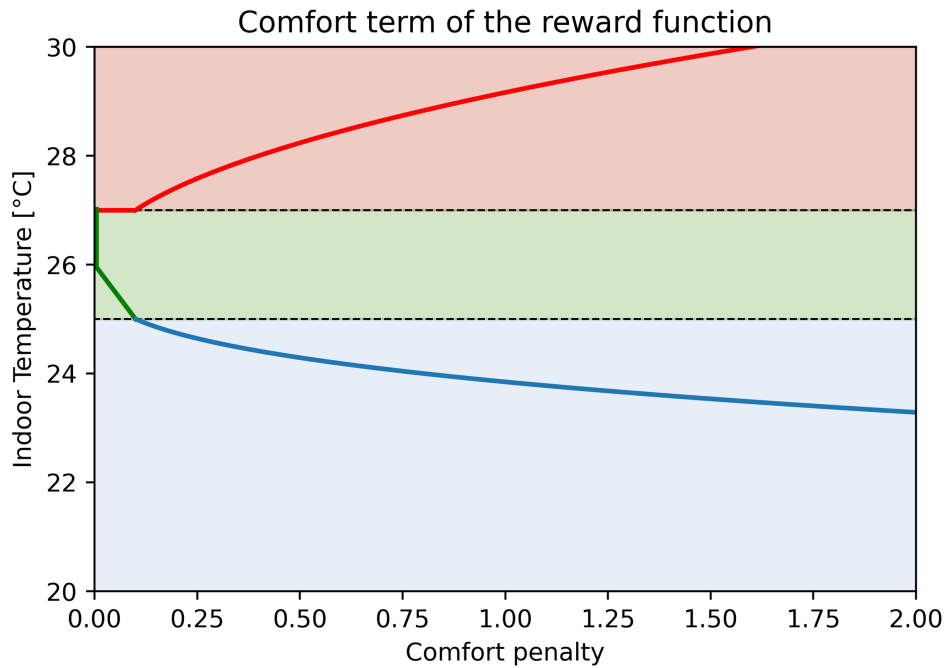


Fig. 4.7 Comfort term of the reward function [36]

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:

$$Storage_p = \max(0, \Delta SOC_{DHW}) * K_1 + \max(0, \Delta SOC_{chilled}) * K_2 \quad (4.6)$$

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:

Table 4.4 Reward function coefficients [36]

| Coefficient | Value |
|-------------|-------|
| m | 0.12 |
| K_1 | 3 |
| K_2 | 2 |
| K_p | 0.6 |

$$Peak_p = \begin{cases} c_{el} = \max c_{el}, -\max(0, e - th_1) * K_p \\ c_{el} < \max c_{el}, -[\max(0, e - th_2) * K_p + \max(0, th_3 - e) * K_p] \end{cases} \quad (4.7)$$

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: 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 1.6 weight the relative importance of temperature violations and peak shaving actions. Moreover, since the reward magnitude influences the behaviour of SAC algorithm [54], these coefficients were used to tune exploration–exploitation trade-off of the agent. Their values are shown in Table 4.4:

4.4.2.4 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 4.5. As explained in section 2.2.3, 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 4.5 Hyperparameter settings [36]

| Variable | Value |
|--------------------------|-----------------------------|
| DNN architecture | 2 Layers |
| Neurons per hidden layer | 256 |
| DNN optimiser | Adam |
| Batch size | 512 |
| Learning rate λ | 0.001 |
| Discount rate γ | 0.9 |
| Decay rate τ | 0.005 |
| Temperature* α | Starting = 1, Final = 0.1 |
| Entropy coefficient* H | Starting = 8, Final = 5 |
| Target model update | 1 |
| Eposide length | 2196 Control Steps (92 day) |
| Training Episodes | 30 |

4.5 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.

4.5.1 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: As shown in Table 4.6 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 and a CV-RMSE of around 1%.

Figure 4.8 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 accurate forecasting.

Table 4.6 Evaluation metrics [36]

| | MAPE [%] | RMSE [°C] | CV-RMSE [%] |
|---------------|----------|-----------|-------------|
| Small Office | 0.80 | 0.28 | 1.08 |
| Retail | 0.45 | 0.15 | 0.58 |
| Restaurant | 0.78 | 0.26 | 1.02 |
| Medium Office | 0.81 | 0.28 | 1.04 |

4.5.2 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 system operation. Finally, the section includes the results obtained at grid level.

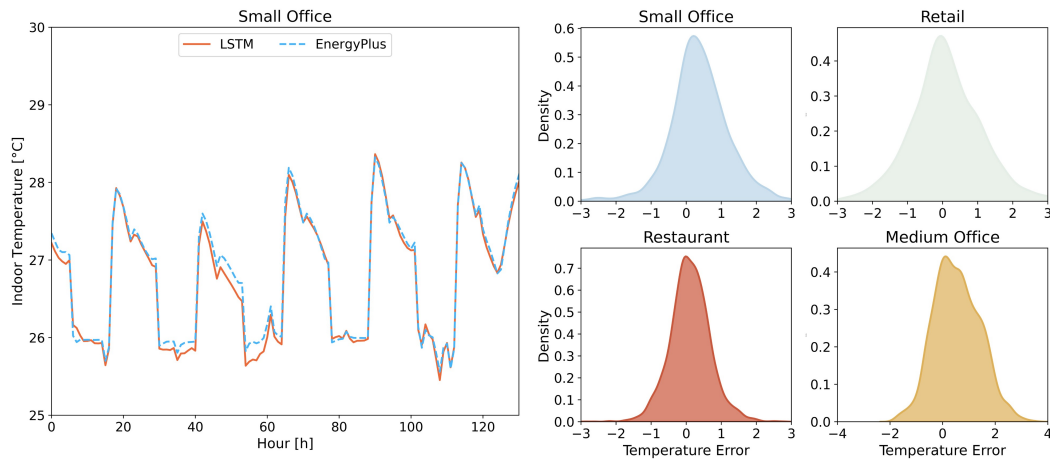


Fig. 4.8 Comparison between indoor temperature predicted with LSTM model and simulated with EnergyPlus (left) and relative error distribution of indoor temperature predicted with LSTM models (right) [36]

4.5.2.1 Comparison at district level

The carpet plots in Figure 4.9 shows a comparison between the aggregate energy consumption at cluster level with the RBC and the DRL controller. 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.

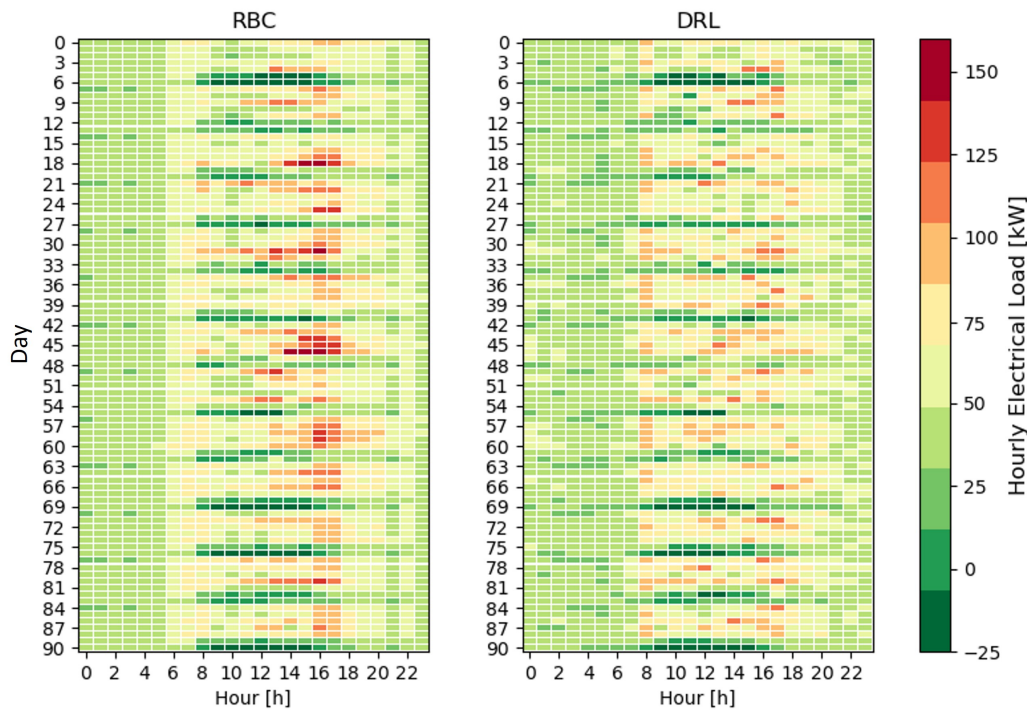


Fig. 4.9 Carpet plot of electrical load at cluster of buildings level with RBC and DRL strategy [36]

To understand how these results have been achieved, Figure 4.10 shows the state of charge profile (SOC) for 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 peaks 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 4.11 shows the distribution of the indoor temperature for the four controlled buildings during occupancy period. As can be seen, 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,

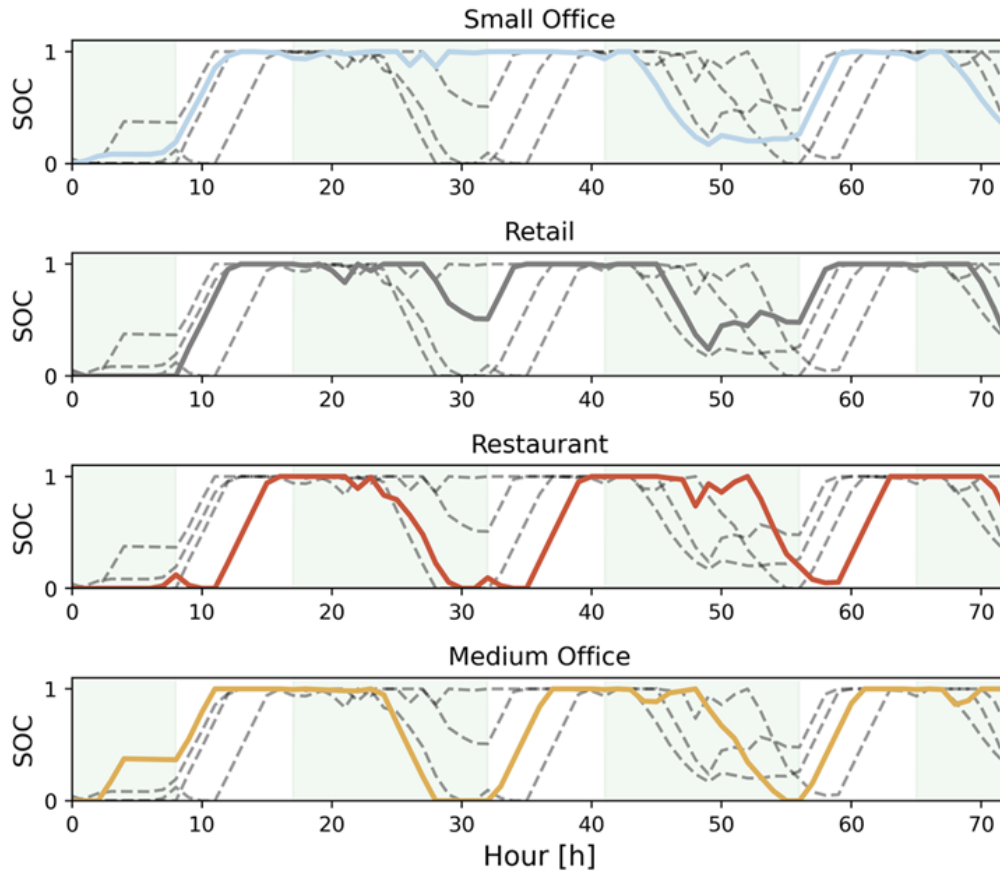


Fig. 4.10 State of charge profile of thermal storage for each building of the cluster [36]

influenced by the external temperature during the early morning hours, when it is open.

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 4.7.

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

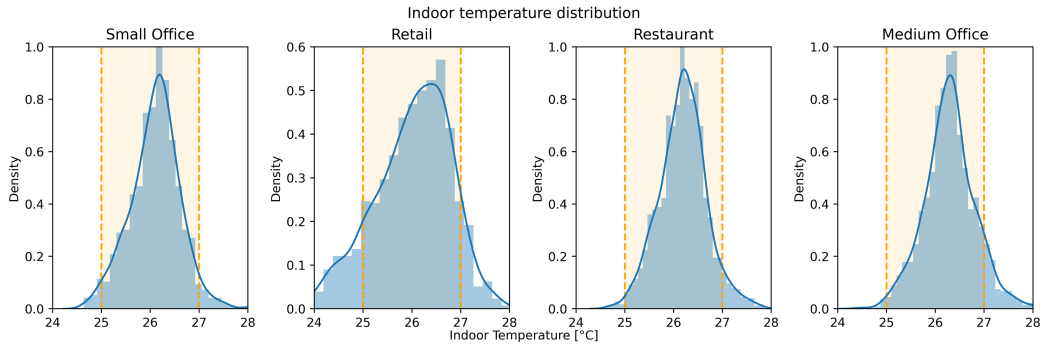


Fig. 4.11 Indoor temperature distribution for each building of the cluster [36]

Table 4.7 Metrics related to indoor temperature control [36]

| | 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 |

between thermal comfort and energy consumption. Figure 4.12 reports internal temperature evolution and cooling energy supplied (i.e., heat pump-to-building or storage-to-building) 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 4.12 a) 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 in energy consumption. Figure 4.12 b) 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 4.12 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.

4.5.2.2 Analysis at grid level

The analysis was then shifted towards the benefits provided by the coordinated control strategy on the grid. In particular, Figure 4.13 shows the load duration curve

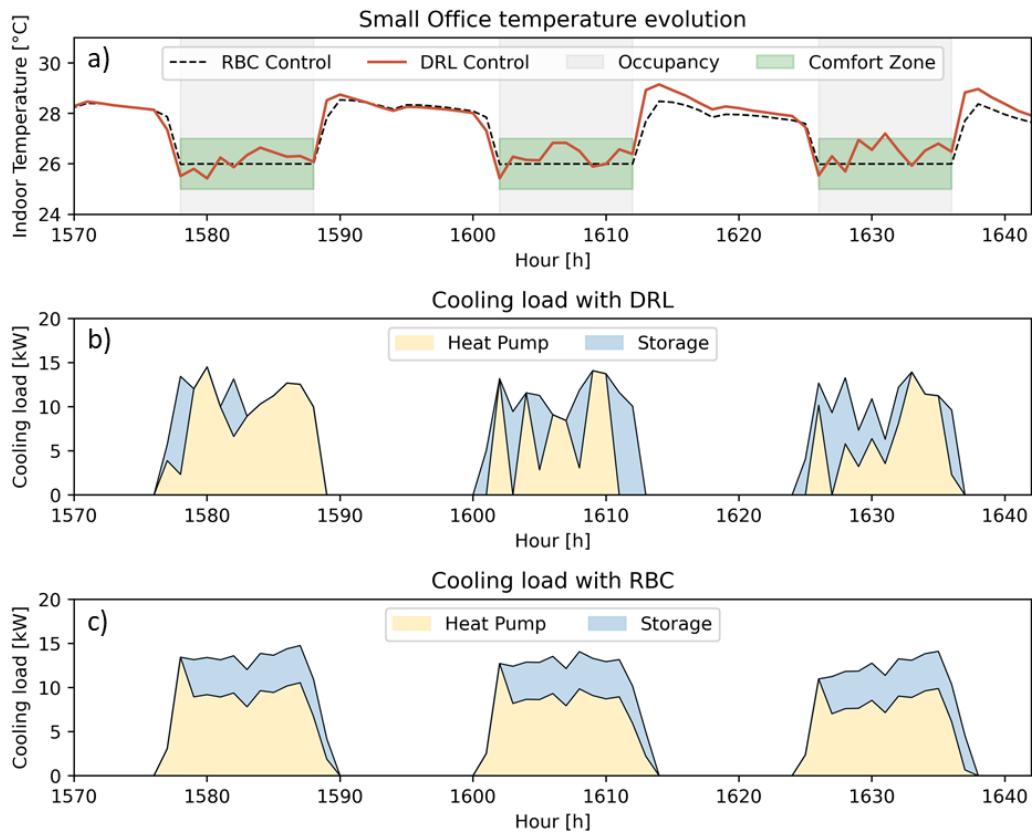


Fig. 4.12 Profiles of indoor temperature and cooling load for the small office building [36]

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. 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 resulting from the implementation of the RBC strategy. DRL controller exploits the possibility to modulate the heat pump cooling power to avoid peak load and takes advantage of the storage charge and discharge process to increase heat pump efficiency, while slightly reducing electricity costs. As pointed out in Table 4.8 and Figure 4.12, 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 into a more uniform baseload. Finally, the controller also shows the ability to better exploit energy the flexibility of the multiple energy systems highlighted by a 4% increase in flexibility factor.

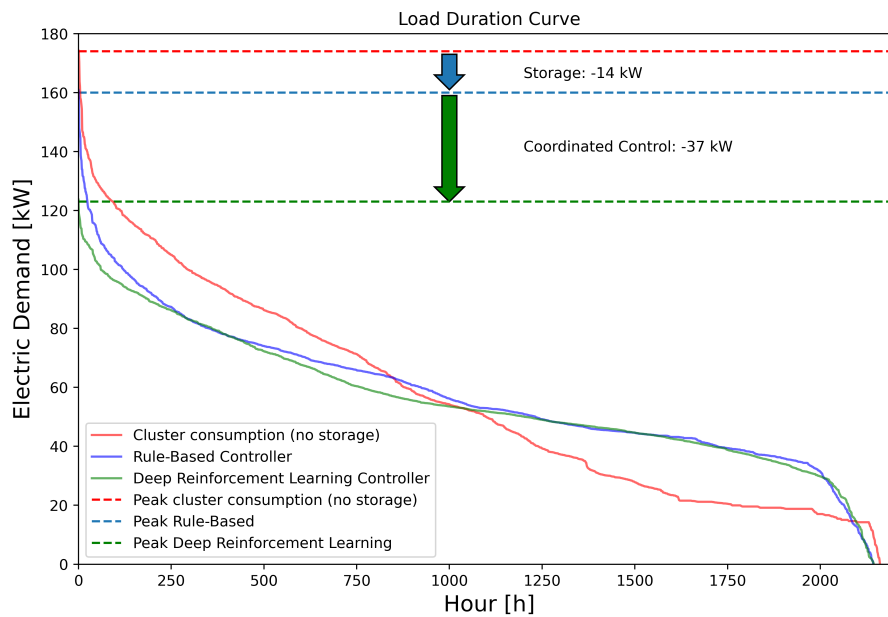


Fig. 4.13 Load duration curve for the different control strategies [36]

Table 4.8 Comparison between performances of the two control strategies [36]

| | 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 |

4.6 Discussion

The presented chapter 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 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 receive 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 thermal storage, consuming energy more efficiently and flattening the electrical load profile. The chapter analyses the role of the state-space and the reward function in the optimal control strategy. The reward function was designed to search 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 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 s, while the forward simulation run for 2300 s. During the deployment of the trained DRL agent, episode was run within 60 s by the proposed approach and 87 s 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.

Chapter 5

Scale-out energy management in buildings with data-driven models

The previous chapter has highlighted the opportunity provided by the use of data-driven models to simulate building thermal dynamics. In particular, the thesis analysed the role of these models in supporting building energy management. However, collecting and preparing a large amount of high quality data to train machine learning algorithms is time consuming and not always feasible, as most buildings lack reliable sensing or metering systems or lack the IT infrastructure to collect and store the data. To address this gap, one key technique needed is to transfer machine learning models trained and validated for buildings with rich data to buildings with limited or poor data. With this motivation in mind, this provides a mathematical background and conducts a comprehensive and structured review on transfer learning applications that supports energy management in the following section. Then, after identifying potential applications and challenges, the chapter discusses an application of transfer learning for building thermal dynamic models, that can be used to scale-out data-driven models in buildings.

Portions of the present Chapter were already published in the following scientific papers:

- Giuseppe Pinto, Zhe Wang, Abhishek Roy, Tianzhen Hong, and Alfonso Capozzoli. Transfer learning for smart buildings: A critical review of algorithms, applications, and future perspectives. *Advances in Applied Energy*, 5:100084, 2022 [239]

- Giuseppe Pinto, Davide Deltetto, and Alfonso Capozzoli. Data-driven district energy management with surrogate models and deep reinforcement learning. *Applied Energy*, 304:117642, 2021 [36]
- Giuseppe Pinto, Riccardo Messina, Han Li, Tianzhen Hong, Marco Savino Piscitelli, and Alfonso Capozzoli. Sharing is caring: An extensive analysis of parameter-based transfer learning for the prediction of building thermal dynamics. *Energy and Buildings*, page 112530, 2022 [210]

5.1 Theoretical background on transfer learning

The following section describe the methods used within the thesis, starting with the theoretical background of transfer learning, followed by a short literature review that aims to describe the potential of the proposed method in the built environment. For a more detailed review of the applications of transfer learning in smart buildings, please refer to Pinto et al. [239].

5.1.1 Transfer learning

The transfer learning concept is based on that of “domain” and “task,” whose definitions are reported below according to Pan et al. [240].

Definition 1. *Domain: a domain \mathcal{D} consists of two components: a feature space \mathcal{X} and a marginal probability distribution $P(X)$, where $X = \{x_1, \dots, x_n\} \in \mathcal{X}$.*

The prediction of building thermal dynamics can be modelled as a regression task. As a result, \mathcal{X} is the space of all influencing variables, (e.g., external temperature, occupancy, HVAC load), while x_i represents the i^{th} influencing variables and X a specific learning sample.

Definition 2. *Task: a task consists of two components: a label space Y and an objective predictive function $f(\cdot)$ (denoted by $\mathcal{T} = \{Y, f(\cdot)\}$), which is not observed but can be learned from the training data, represented by a pair $\{x_i, y_i\}$, where $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$. The function $f(\cdot)$ is used to approximate the conditional probability $P(y|x)$ and predict the corresponding label of a new instance x .*

Analyzing the same application, Y is a continuous space with the possible values of the internal (indoor air) temperature.

We denote the source domain data as $D_S = \{(x_{S1}, y_{S1}), \dots, (x_{Sn_S}, y_{Sn_S})\}$, where $x_{Si} \in X_S$ is the data instance and $y_{Si} \in Y_S$ is the corresponding output. Similarly, the target domain data are denoted as $D_T = \{(x_{T1}, y_{T1}), \dots, (x_{Tn_T}, y_{Tn_T})\}$, where $x_{Ti} \in X_T$ and $y_{Ti} \in Y_T$ are the corresponding outputs. In many cases, transfer learning provides advantages where $0 \leq n_T \ll n_S$.

Definition 3. *Transfer Learning:* Given a source domain \mathcal{D}_S and learning task \mathcal{T}_S , a target domain \mathcal{D}_T , and a learning task \mathcal{T}_T , transfer learning aims to help improve the learning of the target predictive function $f(\cdot)$ in \mathcal{D}_T using the knowledge in \mathcal{D}_S and \mathcal{T}_S , where $\mathcal{D}_S \neq \mathcal{D}_T$, or $\mathcal{T}_S \neq \mathcal{T}_T$.

A schematic representation of the application of transfer learning in buildings is shown in Figure 5.1, highlighting the differences with respect to a classical machine learning problem.

Transfer learning can be classified according to label availability, domain and task similarity and technique used to transfer the knowledge.

Looking at label availability, there are three main categories: inductive, transductive and unsupervised transfer learning.

- In inductive transfer learning, both the source and target domains have labeled data, yet the source and target tasks are different.
- In transductive transfer learning, the source and target tasks are the same, yet the source and target domains are different. In this setting, the source domain has sufficient labeled data while the target domain has none.
- In unsupervised transfer learning the settings are similar to that in inductive learning, i.e., the source and target domains are the same with different but related tasks. However, there are no labeled data in both domains, and the aim is to explore the intrinsic data characteristics in different domains.

Moving to the domain and task similarity, there are mainly two cases: i.e., heterogeneous and homogeneous transfer learning. In the space classification, if the feature space and the label space of both source and target domain are the same, the scenario is classified as *homogeneous*. Otherwise, if they have a different feature space or

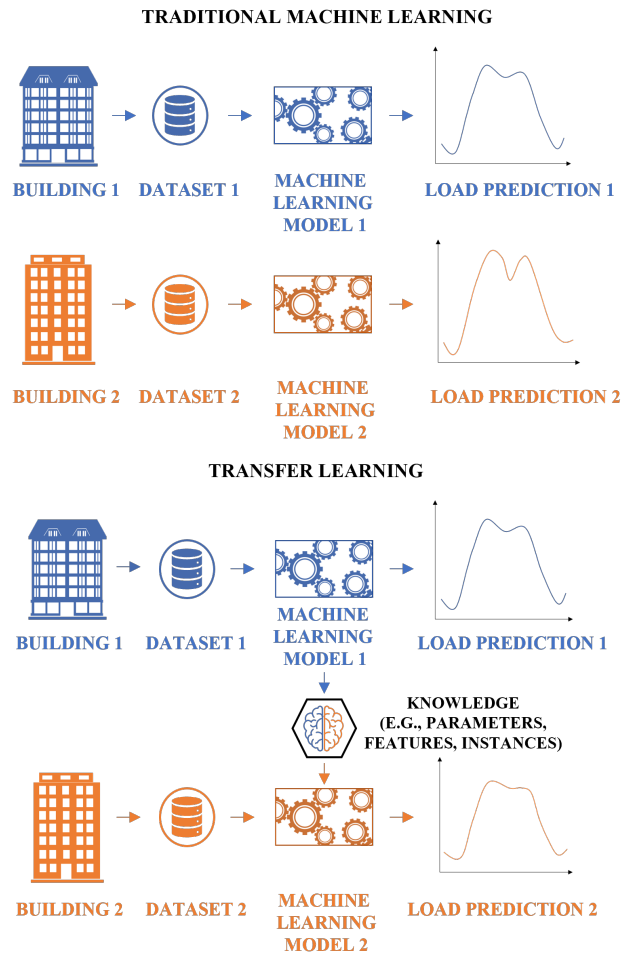


Fig. 5.1 Schematic representation of machine learning and transfer learning problem in buildings [239]

label space, the scenario is classified as *heterogeneous*. Lastly, transfer learning can also be categorized according to the strategy that is adopted to share the knowledge, i.e., data instance-based, model parameter-based, feature representation-based, and relational knowledge-based strategies; the classification based on the strategy adopted hereafter will be defined as *solution classification*.

- The instance-based TL exploits data from a source domain in a target domain. The reasoning behind this is that a subset of data from the source domain can be used to improve the task in the target domain. To incorporate source domain data into the target task training process, one common practice is to use re-weighting and importance sampling techniques. These techniques are

typically used when the domains share the same data variables, increasing the data availability for training purposes without changing the algorithm itself.

- The feature representation-based TL extracts and exploits features to map instances between the two domains (source and target) to increase the performance of the target model. A popular approach is to identify a latent feature space from the source domain, based on which the marginal distributions between two domains are minimized.
- Relational knowledge-based TL is based on the assumption that data have similar relations in the two domains. As a result, it is used in multi-relational datasets and the knowledge to be transferred is the relationship between the data.
- The model parameter-based TL shares some parameters or prior distributions of the hyper-parameters of the models (e.g., neural networks). The latter is built assuming that model parameters or hyper-parameters generated for similar tasks would be similar. In this situation, the information collected from the source task is sent to another task in the form of shared model weights. The increasing advancement in deep learning has inspired a new type of transfer learning — network-based transfer learning [241] — that falls within the parameter-based transfer learning category and may be further categorised based on the approach used to share model parameters:
 - The first method is to extract the features from the pretrained model. The weights of these layers are fixed in this scenario, with the exception of the input/output layer, which is domain dependent and must be fine-tuned using target data. The key benefit is the acquired ability to deal with limited quantity of data to generalize over different domains.
 - An alternative is to use the source model for initialization purposes. In this scenario, the source model is used as a starting point and further fine-tuned.

Figure 5.2 shows the two parameter-based TL approaches used, henceforth called *feature-extraction* and *weight-initialization*.

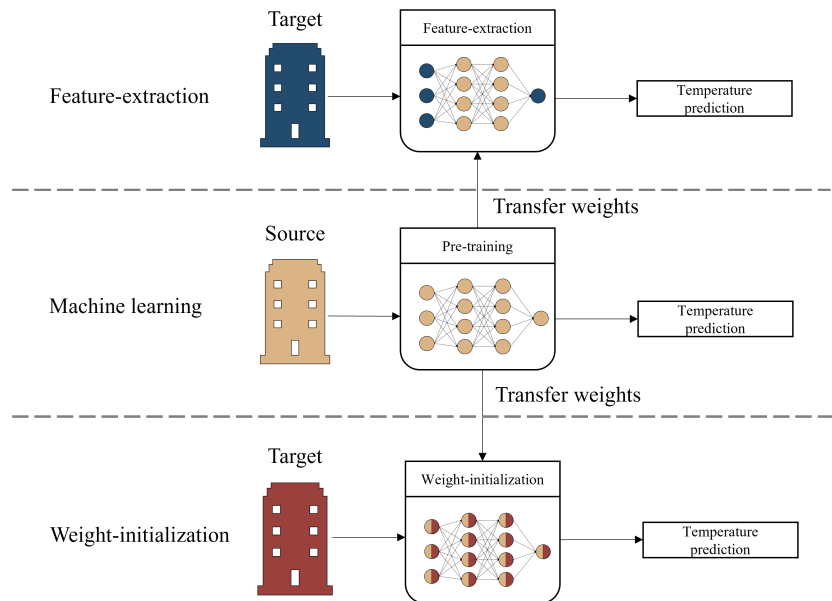


Fig. 5.2 Feature-extraction (top) and weight-initialization (bottom) transfer learning schematization [210]

5.1.2 Literature review on transfer learning applications

The main applications of TL in smart buildings are related to load prediction, fault detection and diagnosis, non-intrusive load monitoring and occupancy detection, while during recent years an increasing trend of works focused on building thermal dynamics, and systems control were observed. The thesis will further discuss the latter applications since they are still in their infancy and can help to scale energy management in buildings, while for a detailed literature review of transfer learning applications for smart buildings the reader can refer to Pinto et al. [239].

5.1.2.1 Energy systems control

BAS are computer-based automated systems that monitor and regulate all energy-related systems in buildings, including mechanical and electrical equipment. They are frequently used to automate all services and operations within a building in order to optimize its performance, efficiency, and energy usage. This technology enables the execution of essential energy management activities such as automating demand response techniques and supervising energy prices, favoring distributed energy resources exploitation and energy transition. However, as the energy systems

are unique for each building, advanced control strategies are usually tailored for each specific building. Very recently, transfer learning was used to enhance building systems control, favoring the information sharing between different energy systems. Some pioneering works exploited a policy-transfer approach [242] in combination with RL to optimize control at different scales: microgrid [243, 244], batteries [245], HVAC systems [246, 247], and appliances [248, 249]. A key pain point of applying RL controllers in buildings is the training process that is time- and data-demanding before it can converge. To address this problem, Zhang et al. [248] first identified several homes similar to the target home that have the same number and type of appliances. Then the RL controller was trained on the source home and fine-tuned for the target homes. The results showed that TL can effectively reduce the training time of a new policy if the target home is similar to the source homes. Tsang et al. [249] used transfer learning to train a DRL controller of Household Energy Management. The agents in the target domain are advised by the suggested actions of the existing model pretrained in the source domain.

Xu et al. [247] applied the same process, shifting the domain from appliances to HVAC systems, transferring the policy of DRL-based HVAC controllers from source buildings to target buildings with different materials and layouts, HVAC equipment, and weather conditions. Furthermore, they analysed a case with a different number of thermal zones, being the only work that used heterogeneous TL for control application, thanks to its ability to generalize over thermal zones. Similarly, Lissa et al. [246] studied the effect of transferring the policy of an HVAC controller from one room to another in the same building, performing several experiments to test the robustness of the controller and assessing the impact on occupant discomfort time, showing reductions using a TL approach. Looking at microgrid scale, Fan et al. [243] evaluated similarity between the production and generation of two different microgrids to find the optimal way to transfer knowledge, sharing the weight of the DRL neural networks and speeding the controller performance, paving the way for possible application at a large scale. Lissa et al. [244] proposed an inter-agent transfer, in which knowledge is shared with another agent with similar characteristics, and this agent can merge the transferred knowledge with its own experience. This concept is called *parallel transfer learning*, where the knowledge to be shared between agents does not need to wait until the end of the process to be available. Lastly, Mbuwir et al. [245] applied transfer learning to speed the convergence of the learning algorithm to optimize thermal and battery storage planning, improving also

its scalability. The results show that reinforcement learning coupled with transfer learning can represent a suitable alternative when few data are available, despite further studies are needed to demonstrate the ability of transfer learning to generalize across multiple buildings, especially when controlling different energy systems.

5.1.2.2 Building thermal dynamic models

Building thermal dynamic models (predicting how the building thermal state will evolve under different weather, disturbances, and other factors) have many applications, including but not limited to advanced controls such as Model Predictive Control (MPC) and DRL or the increased accuracy of load prediction models. Conventional building thermal models are developed through a physics-based approach, such as in EnergyPlus. The shortcomings of physics-based building modeling are the high time and expertise demand needed to develop such a model and the need for a great deal of information about the building and system features. An alternative approach to model building thermal dynamics is data driven modeling. However, a large amount of historical data may be needed to train such data-driven building thermal dynamics models, which is challenging, especially for buildings that are brand new or not yet commissioned [250]. This highlights how transfer learning can be leveraged for this application. Recently, few works tried to adopt transfer learning to develop building dynamics models, overcoming the presented limitations. Jiang et al. [250] pretrained an LSTM S2S model using a large amount of data from source buildings to study building temperature evolution. Then weight initialization was used to enhance the performance of the target building. In that case, the whole model was fine-tuned without freezing any hidden layers. Similarly, Chen et al. [251] applied transfer learning to predict not only internal temperature but also relative humidity. In other studies (such as [252]), the hidden layers have been frozen while only the last fully connected layers were fine-tuned. It was found that the deep supervised domain adaptation is effective to adapt the pretrained model from one building to another, and has better predictive performance than learning from scratch with only a limited amount of data [250].

Hossain et al. [253] trained a Bayesian neural network (BNN) to directly learn an RC model rather than estimating parameters. The work proved that at least several weeks of data are necessary to obtain good performance. The paper proposes a methodology on how to transfer these models in new buildings with only one day of

data, identifying and selecting the best RC model according to consumption patterns and outperforming time-series methods directly constructed on available data.

Additionally, data-driven models have been used to represent specific temperature evolution, as in Kazmi et al. [254], which applied TL to train a model to predict the thermal behaviors of hot water storage systems; or Hu et al. [255], which applied transfer learning to predict the thermal comfort state in buildings. Lastly, Grubinger et al. [256] present an interesting approach of online transfer learning coupling the resulting prediction with an MPC controller, paving the way for possible application of this technique.

5.2 Motivations and novelty of the proposed approach

Applying transfer learning in smart buildings is an emerging research topic that has attracted increasing research attention. The idea of transfer learning originated from machine learning, which was accelerated as more data and computing power became available in the past decade. However, research on this topic is still at a very early stage or, in the case of relation-based TL, still needs to be explored in smart buildings. Moreover, despite using real data, existing literature used such data in an offline fashion, without deploying them in real world applications. This approach tends to be simplified and may not reflect real world problems in real buildings. More in-field studies are needed to validate TL performance in real buildings. Collaboration and coordination between academia, industry, and policymakers are needed to apply TL to solve real-world problems and make true impacts. Despite the emerging interest for transfer learning in smart buildings, many research gaps still need to be addressed. Below are reported considerations and insights for a number of open questions based on the knowledge extracted during the thesis:

- **Why Transfer Learning for energy and buildings?** Higher data availability in buildings is leading more and more to a data-centric energy management with the opportunity of exploiting complex AI-based models. In this context, TL can support the penetration of ML for energy management in buildings by contributing to reduced implementation costs (i.e., pipeline of the machine learning frameworks) and time. The natural use of TL can be found in existing buildings recently equipped with monitoring infrastructure (i.e., no historical

data) or new buildings (with limited historical data). However, guidance on the type and number of sensors needed to fully exploit the benefits of TL are heavily dependent on applications, and are still not clear. While building thermal dynamic models often require physics-based approaches or a lot of data, limiting their adoption, transfer learning can speed-up and overcome data availability, allowing for an effective coupling with advanced control strategy. Looking at energy systems control, the application of transfer learning is crucial to broaden the adoption of advanced control strategies, which have a bottleneck of high effort to train and tune models. In fact, these approaches are too data intensive to be applied at scale. In general, the main advantages of TL use in smart buildings are the increase in performance and the potential to scale-up and speed-up ML processes. However, compared to computer vision applications, deep learning models used for buildings are not computationally demanding, therefore further analysis is needed to assess computational advantages when the scale of the analysis is larger (e.g., communities, districts, or cities).

- **When to use Transfer Learning in the built environment?** As previously said, TL finds its natural application when trying to apply ML models in existing buildings with few, poor, or no historical data, as well as in new buildings without historical data. However, its application is further complicated by the type of task to be completed. A prerequisite for TL applications is a certain degree of similarity among the two domains; however, except for a few studies in building load prediction that tried to quantify time-series analogies, no studies have quantified the specific features' importance of the similarity, and those studies are needed. In particular, looking at building load prediction and building dynamics estimation, similarity plays a key role, since buildings can have similar (or different) shapes, use, climatic condition, and equipment that, depending on the considered task, may have more or less influence. Therefore, to fully understand the advantages of transfer learning applications in building load prediction and dynamics estimation, a proper definition of similarity must be defined, contour the range of applications of transfer learning. Lastly, control applications may benefit from transfer learning when buildings are subject to a retrofit of energy systems and the optimal control strategy may obtain a significant jumpstart using the initial control policy from a similar building.

Summarizing the previous questions, the thesis highlights the main challenges related to the application of TL in buildings, described below. Some challenges are common to the different tasks and related to the models, while others are related to specific applications.

1. Further studies are necessary to propose robust methods on how to select the right source building, quantifying the similarity between buildings, thus avoiding negative transfer. Despite few attempts in literature, there are no well recognized standards or principles, and guidance is needed in this regard.
2. Looking at parameter-based TL, it is not yet clear which of the feature-extraction and weight-initialization brings the greatest benefits in smart building applications. In particular, feature-extraction is much more used for classification problems, while for regression problems there is not enough evidence, representing one of the challenges to be overcome to increase the effectiveness of transfer learning.
3. Another common question that still needs to be addressed is related to the amount of data necessary in the source building and the amount of data necessary to properly transfer knowledge in the target buildings. This becomes even more true when considering applications that can be highly dependent on seasonality, such as building dynamics and systems control.

To overcome the identified gaps of transfer learning, the thesis developed an application that leverages a synthetic dataset to create multiple energy models of a single building in different conditions, changing building features such as efficiency level, occupancy and climate. The dataset was then used to train and compare machine learning and transfer learning models. A machine learning model only leverages data available for the target building, while the transfer learning model reuses knowledge from a source building to reduce implementation costs, speed up the training and increase performance. The aim is to assess their performance, isolating the contribution of specific features and studying the effect of data availability on transfer learning performance. With this in mind, this study aimed to address the literature gaps, with the following contributions:

1. Isolating and evaluating the contribution of key features in determining machine learning and transfer learning effectiveness, using a synthetic building dataset gathered from a detailed physics-based building energy model.

2. Performing a statistical investigation by developing approximately 250 models to assess the feature importance and data availability impact.
3. Conducting a specific analysis of negative transfer to assess the limitations of transfer learning for building thermal dynamics, to identify guidelines for future research.
4. Assessing the effectiveness of transfer learning in an online deployment setting, supporting its real-world implementation

The work is organised as follows: Section 5.3 introduces the case study, explaining in detail the design of experiment. Then, Section 5.4 describes the methodological framework at the basis of the analysis. Section 5.5 presents the results of the comparison of ML and TL models, taking into account performances and computational costs, while a discussion of the results is given in Section 5.6.

5.3 Case study

The selected case study is an archetype building energy model developed from the U.S. Department of Energy (DOE). The model is a medium-sized office building with three floors and a total floor area of 4,890 square meters [257]. The building consists of 12 space types: open and enclosed office rooms, conference room, classroom, dining area, lobby, corridor, stair, storage, restroom, plenum and mechanical room. A schematic representation of a floor is shown in Figure 5.3.

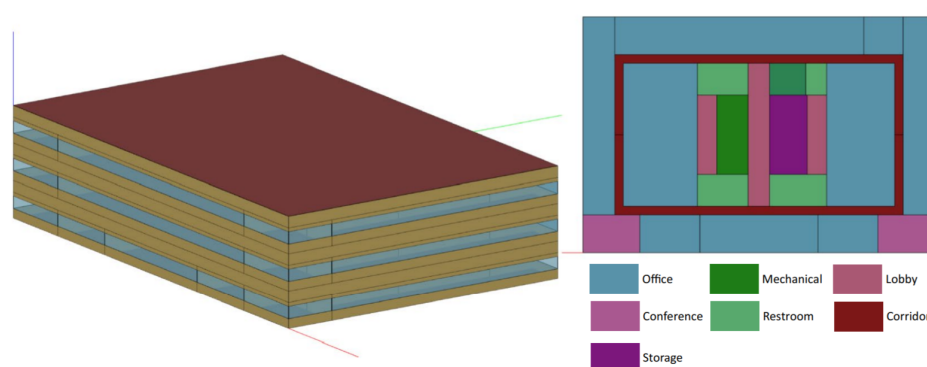


Fig. 5.3 A schematic representation of medium office geometry and thermal zones for a single floor [210]

The synthetic dataset includes simulations for the selected building model in different climates for multiple years, efficiency levels and occupancy patterns. The dataset includes three energy efficiency levels, obtained by changing building envelope properties, and the efficiencies of lighting, miscellaneous electric loads (MELs), and HVAC systems. Furthermore, three sets of schedules for zone-level occupancy, lighting, MELs, and thermostat setpoint, reflecting realistic building operations from stochastic occupancy simulations, were used [258]. The resulting configuration are reported in Table 5.1.

Table 5.1 Parameters and modified features used for the design of experiment [210]

| Parameter | Cases | Features involved |
|------------|---------------------|---|
| Efficiency | Low, Standard, High | Building envelope properties, efficiency of lighting, MELs and HVAC systems |
| Climate | 1A,3C,5A | Outdoor air temperature, solar radiation |
| Occupancy | 1,2,3 | Schedule of occupancy, MELs, lighting, setpoints |

To study the contribution of different weather conditions on model performance, three typical climate zones were selected: Miami (1A, hot and humid), San Francisco (3C, moderate/mild), and Chicago (5A, cold winter and hot summer). A synthetic dataset [259] was used with twofold advantages: (i) it can reflect the effects of different influencing variables on building operation, and (ii) it isolates the contribution of specific features on machine learning and transfer learning model accuracy.

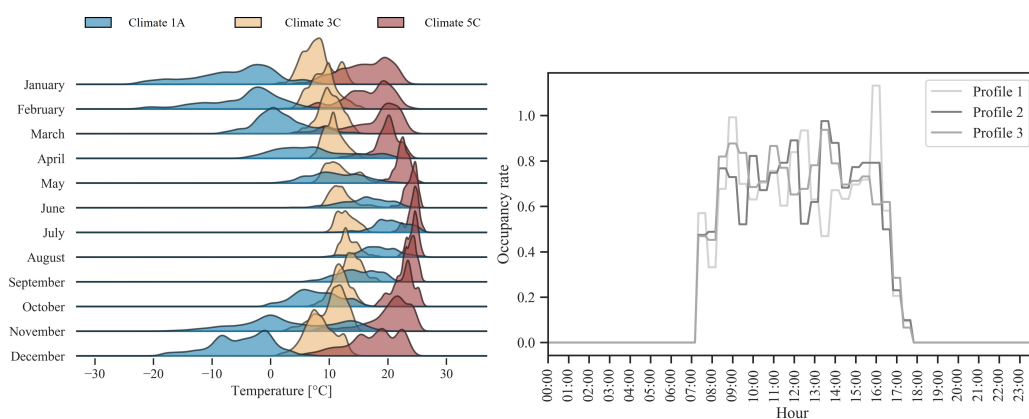


Fig. 5.4 Distribution of the outdoor air temperature for each month and climate considered during the analysis (left) and occupancy profile distribution (right) [210]

Figure 5.4 shows on the left part the outdoor air temperature distributions for the three climates selected. The selected climatic zones exhibit very different tempera-

ture patterns, with Climate 1A being cooling dominant, Climate 5A being heating dominant, and Climate 3C representing a mild climate. Furthermore, it shows the probability distribution of the three occupancy profile considered to highlight different users behaviour. For each combination between efficiency level, occupancy and climate, up to two years of meteorological data were used for training and testing purposes. The simulations yielded time-series data that included whole-building and end-use energy metering, indoor and outdoor environmental variables, and system and component variables (e.g., zone thermostat setpoints, VAV terminal supply air temperature). For a detailed description of how the synthetic dataset was obtained, refer to Li et. al [259].

5.4 Methodology

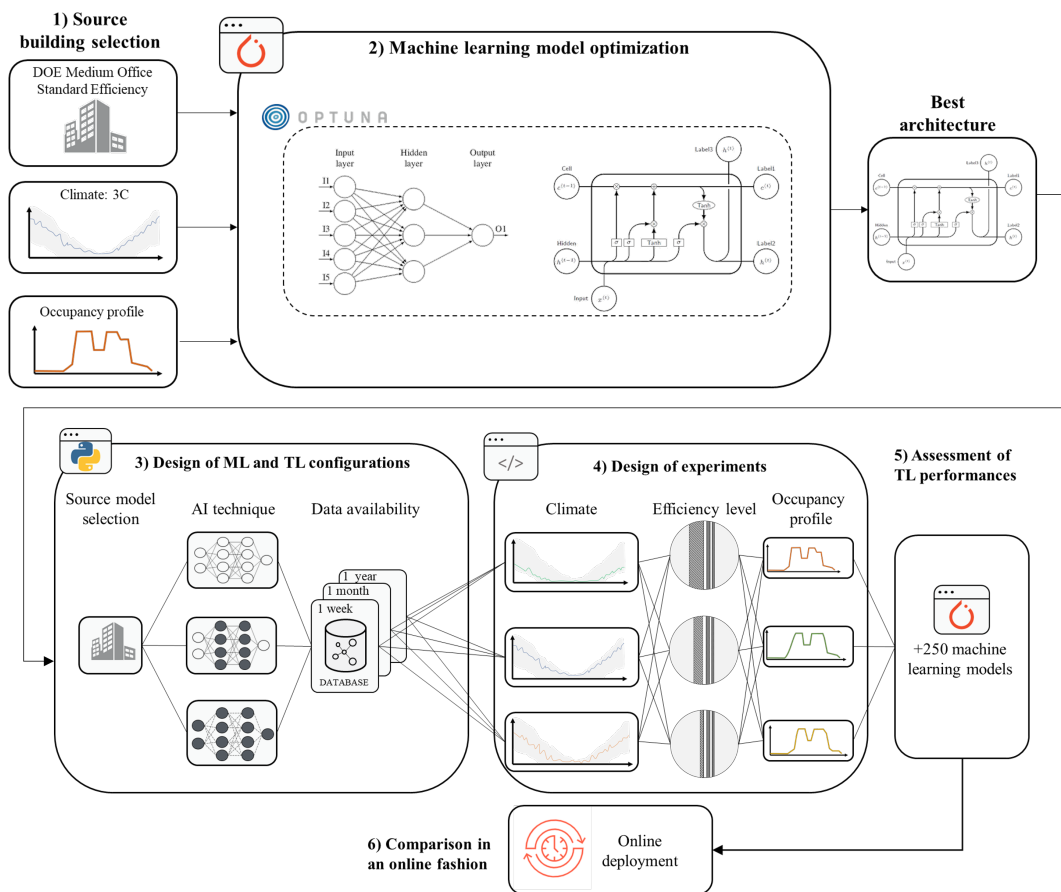


Fig. 5.5 Methodological framework [210]

This section reports the methodological framework adopted, as shown in Figure 5.5. The methodology unfolds in four main steps, described below.

5.4.1 Source building selection

The first step consists in the identification of the “source building,” used as a starting point for transfer learning. As pointed out in the previous section, the dataset analyzed refers to a medium-sized office building simulated in 3 climates, 3 energy efficiency levels, 3 stochastic occupancy schedules, for a total of 27 EnergyPlus models. The source building was conceived with a standard energy efficiency, the occupancy profile 1 (according to Table 5.1) and was simulated in Climate 3C. The climate and the energy efficiency level were chosen to represent an intermediate condition between the other two options, with the aim to further evaluate the potential of applying transfer learning. The dataset has a 10-minute granularity, with information related to whole building variables as well as zone variables.

5.4.2 Machine learning model optimization

The second step includes the model development, the selection of the architecture and the optimization of the related hyperparameters. The models aimed to predict the temperature evolution of a single zone (mid-office) one-hour ahead (six time-steps), exploiting information of the specific zone. This was necessary due to the impossibility of aggregating data at a higher level, since different zones may have different setpoints and occupancy schedules. The selected inputs for the machine learning models were the zone heating and cooling temperature setpoints, the outdoor air temperature, the previous internal (zone air) temperature, solar radiation, and information about hour, day and month. Figure 5.6 shows the input parameter together with the sliding window approach used to perform the predictions.

The architectures selected were MLP and LSTM. The developed models used 48 time-steps (8 hours) as a lookback period to predict the next 6 time-steps (1 hour). Each architecture was characterised by specific hyperparameters, therefore an optimization process was carried out using the Optuna [260] framework. The tool allows the optimal hyperparameter combination to be searched by performing an automatic grid-search. The work performed the grid-search using five values in

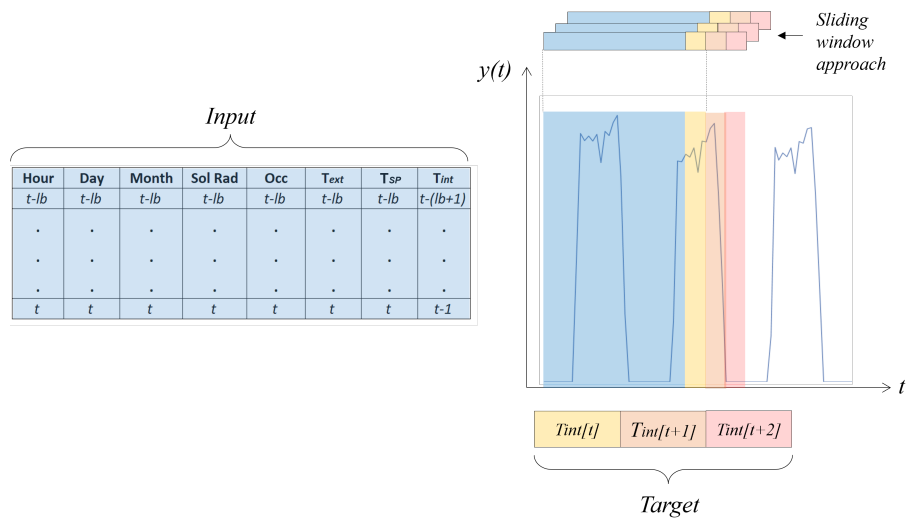


Fig. 5.6 Input of the neural networks and sliding window approach [210]

Table 5.2 Neural network hyperparameter optimization process [210]

| MLP | Range | Optimum | LSTM | Range | Optimum |
|-------------------|-------------------------|----------------------|---------------------|-------------------------|---------------------|
| # Neurons layer 1 | [50-200] | 100 | # LSTM layer | [3-7] | 3 |
| # Neurons layer 2 | [50-150] | 70 | # Neurons per layer | [70-300] | 175 |
| # Neurons layer 3 | [20-90] | 70 | Epochs | [80-120] | 90 |
| # Neurons layer 4 | [10-70] | 10 | Learning rate | $[7-8.5 \cdot 10^{-3}]$ | $7.7 \cdot 10^{-3}$ |
| Epochs | [80-200] | 120 | Batch size | [800-1000] | 900 |
| Learning rate | $[7-8.5 \cdot 10^{-3}]$ | $7.57 \cdot 10^{-3}$ | Optimizer | | Adam |
| Batch size | [800-1000] | 900 | | | |
| Optimizer | | Adam | | | |
| MAPE | | 1.096 | | | 0.535 |

the specified interval shown in Table 5.2 with a uniform distribution. The dataset included two years of data: one used for training and validation and the other one used for testing. Table 5.2 illustrates the hyperparameters subject to the grid-search optimization with their optimized values, as well as the value of the mean absolute percentage error (MAPE) evaluated in the testing period. Table 5.2 highlights the higher accuracy of the LSTM architecture, which was then selected to perform the experiments. Consequently, all the transfer learning models further described will share the same architecture, despite changing the learning rate.

5.4.3 Design of ML and TL configurations

The third step compares classical ML with two TL techniques to predict indoor air temperature evolution. A classical machine learning approach used the optimal hyperparameter identified in step 2 to train LSTM models on data available for the target building. The performance of the LSTM model was then compared with that resulting from the models trained using two transfer learning methods: weight-initialization and feature extraction. In weight-initialization, the whole network is fine-tuned using the data available in the target building and a lower learning rate with respect to the one used to train the source network, while in the feature extraction, the LSTM layers are frozen and only the last dense layer is fine-tuned. For both weight-initialization and feature-extraction, a learning rate equal to $2 * 10^{-3}$ was used to train the LSTM for 80 epochs.

Moreover, this step aims to analyse the impact of data availability on model performance. To this purpose, three cases were considered regarding the data availability for the target building: (i) 1 week of data, (ii) 1 month of data, and (iii) 1 year of data. The cases of one week and one month of data were used to represent a data-scarcity context and had the main purpose of highlighting in which conditions TL performs better than ML and the minimum amount of data necessary to develop an effective ML model. On the other hand, an ideal case that considered one year of data available in the target building was used to assess the generalizability of TL over ML, to assess if TL can provide additional advantages even in the presence of an extensive amount of data for the target building.

5.4.4 Design of experiments

The fourth step deals with the design of the scenarios resulting from the combination of the different features for the target building as reported in Section 5.3. Machine learning and the two transfer learning strategies were implemented to consider the combination of three climates, three energy efficiency levels, three occupancy patterns and three data availability periods. This led to 243 different models, including the one related to the source model used for transfer learning. These simulations were used to perform a statistical investigation on the most important features for the application of TL for building thermal dynamic models. All the information on the data, the code and the results produced by

the statistical investigation are open-source and available at the following link: https://github.com/baeda-polito/Transfer_learning_building_dynamics.

5.4.5 Assessment of TL performance

Lastly, model performance is compared using several metrics. In particular, model absolute performance was compared using metrics such as MAE , MAPE, MSE and CV-RMSE, the definition of which is provided below. Relative performance was quantified using the asymptotic performance and jumpstart.

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (5.1)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (5.2)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|^2 \quad (5.3)$$

$$CV - RMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|^2}}{\bar{y}_i} \quad (5.4)$$

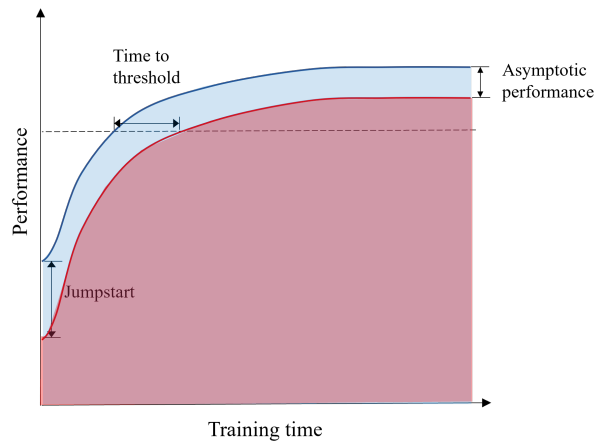


Fig. 5.7 Transfer learning metrics used to quantify the performances of the new model [210]

Figure 5.7 shows three metrics often used to assess the improvements after transfer learning application.

1. *Jumpstart*, which reflects the improvement in initial performance in the target task achieved by utilizing transferred information prior to any further learning.
2. *Time to threshold*, which compares the time it takes the model to acquire a specific level of performance in the target task given transferred knowledge to the time it takes to learn it from scratch.
3. *Asymptotic performance level*, which quantifies the ultimate performance level of the agent in the target task when transferred from the source.

Regardless of the specific ML task, jumpstart and asymptotic performances for the regression problem are evaluated using MAE, MAPE, MSE and CV-RMSE. However, in this work the metric time to threshold was not analysed due to the necessity to quantify a specific threshold (e.g., MAE = 0.5 °C), which may or may not ever be reached by machine learning models.

5.4.6 Comparison in an online fashion

To further demonstrate the effectiveness of transfer learning in an online fashion, this study compared an online machine learning approach (updating the weights of the neural network as new data become available) with an online transfer learning deployment strategy. Online transfer learning leverages one year of source data and updates the model in an online fashion each week as new data become available, performing a fine-tuning of the model. The comparison is helpful since real-world application often works with online data and building thermal dynamic models are used as a part of a model predictive control implementation, thus requiring it to be robust and fast.

5.5 Results

This section describes and analyses the results obtained from the proposed design of experiments. Section 5.5.1 describes the results obtained from both ML and TL

models, analysing the performance distribution and identifying the factors that most influence model performance. Furthermore, statistical analysis was performed to compare absolute and relative performance of the proposed approaches with respect to the different features. Section 5.5.2 focuses on negative transfer, describing the boundary conditions in which it occurs and assessing benefits and limitations. Lastly, Section 5.5.3 describes computational advantages related to the application of TL, analysing jumpstart and training asymptotic performance.

5.5.1 Machine learning and transfer learning performance

Figure 5.8 shows the average performance over the entire design of experiments of ML and TL models using one month of data to assess the previously introduced metrics (MAE, MSE, MAPE, CV-RMSE) over all the six time-steps. As can be seen, the ML algorithm error is almost constant over the time-steps, while both transfer learning techniques show a lower error for the first prediction time-step, reaching about the same accuracy at the last time-step (one hour). On average, both feature extraction and weight initialization techniques perform better than machine learning. The analysis of MAE, expressed in °C shows that for the first time-step the two TL techniques have a value of 0.17 °C smaller compared to standard ML, achieving a performance improvement of 50%. Similar considerations can be made for the other two metrics, that show substantial improvement with respect to ML performance for the first time-step and a better average performance.

For the sake of simplicity, the following analysis considers only the average performance of the mean absolute error over the entire prediction horizon, since it can be interpreted easily.

The first step aimed to assess the effectiveness of transfer learning between different zones of the building. To prove the effectiveness of TL in different zones, two target zones (highlighted in red in Figure 5.9a) were selected: a conference room on the second floor and an enclosed office on the second floor. The rationale behind zone selection was to test the neural networks with different orientation, area and floor, that represents the heterogeneity in terms of size, shape, and orientation of different buildings. The conference room on the second floor (MID_2) was selected to test the influence of a different exposure on the model (changing it from east to west), while the enclosed office on the second floor (BOT_2) was selected to test both

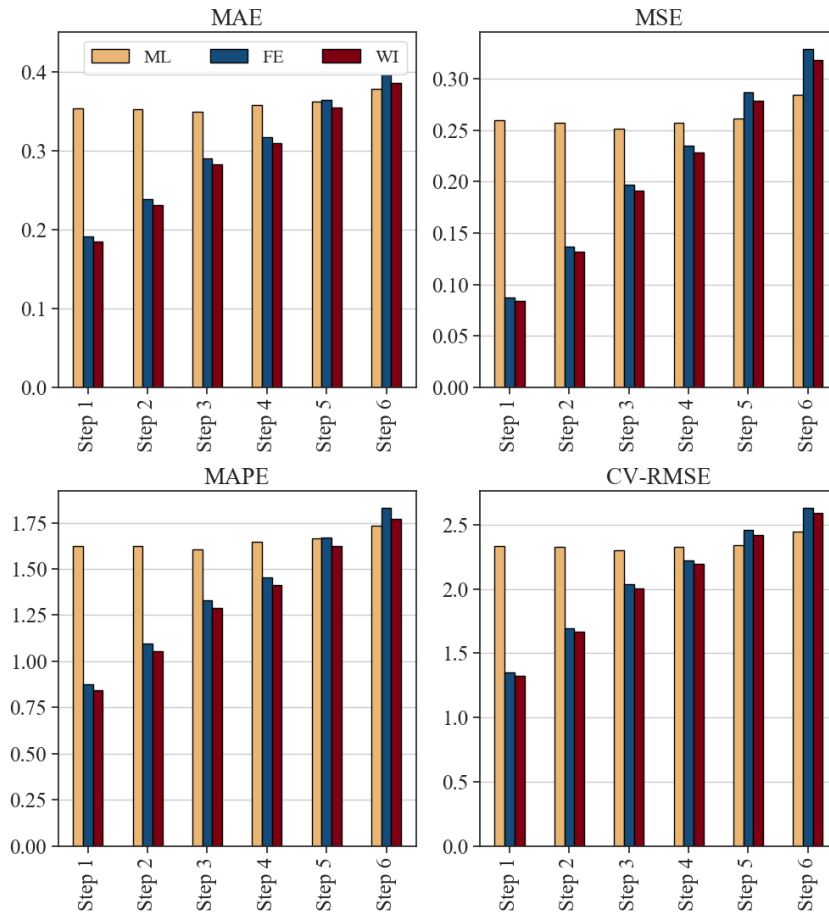


Fig. 5.8 Performance of the different techniques over the control horizon [210]

a different area and a different exposure. Once the zones were identified, the analysis was performed, considering different data availability (from one week to one year). Then several tests were performed that considered different data availability and compared the results of ML and TL models analysing the mean absolute error. Figure 5.9b shows that despite the different characteristics, TL was able to obtain better performance than standard ML independently from the data availability. Indeed, the ML model performance was heavily influenced by the amount of training data for the target building, while the TL model presented robust results over different data availability. After having assessed the ability of TL in different thermal zones, to isolate the effect of other variables, the following analysis was performed using the same thermal zone as a source.

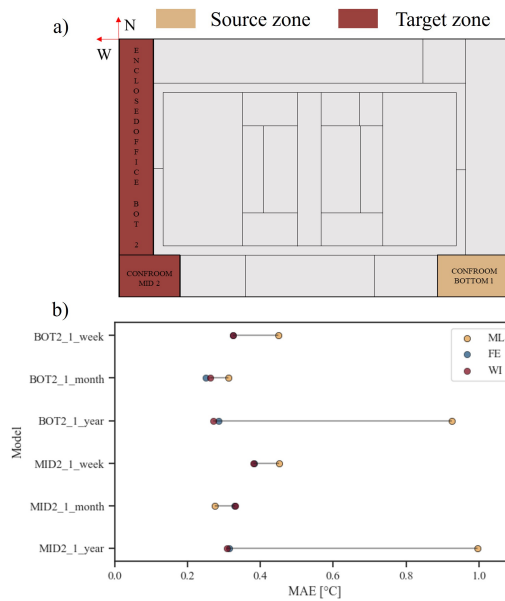


Fig. 5.9 Performance of the different techniques over different zones [210]

Then, to analyse the average performance of the three techniques on the whole design of the experiments, mean absolute error was used to aggregate results over different climate, data availability, efficiency and occupancy profiles. As a result, Figure 5.10 shows the average MAE distribution for the three proposed approaches over all the simulations performed. The analysis of the distributions showed that ML trained over a period of one year in Climate 5A had in many cases unacceptable errors. A specific analysis will be conducted later to understand the main factors related to the lower ML model performance. Furthermore, Figure 5.10a highlights the larger error distribution of the ML technique, which reaches values of more than 1 °C, while the TL maximum errors are below 0.7 °C. To better understand how the ML error is distributed, details for different data availability are shown in Figure 5.10b. The figure displays how one year of data led the ML model to a large error distribution, while one month of data showed the best performance, with an average error of 0.35 °C. As a result, the focus was shifted toward a one month training period. Figure 5.10c compares the error for each technique, assessing a slight performance improvement for both TL techniques over ML, with no particular differences between feature extraction and weight initialization.

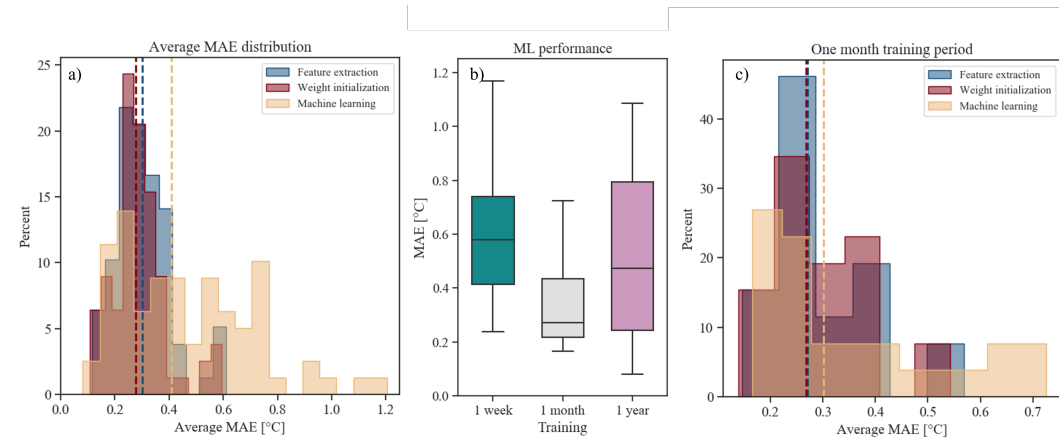


Fig. 5.10 MAE distribution over different periods and techniques [210]

To further study the effectiveness of transfer learning, average MAE distributions were divided in three ranges: low error ($MAE < 0.4 \text{ }^\circ\text{C}$), medium error ($0.4 \text{ }^\circ\text{C} < MAE < 0.7 \text{ }^\circ\text{C}$), and high error ($MAE > 0.7 \text{ }^\circ\text{C}$).

Figure 5.11 shows the error distribution by technique over all the influencing factors using a categorical plot. The ML technique is the only one with a high error, which mainly occurred with one week and one year of data. Furthermore, it shows how high errors are predominant in Climate 5A but are evenly distributed over the efficiency levels and occupancy runs. On the other hand, both feature extraction and weight initialization showed better performance; almost evenly distributed over different data availability, with lower error for Climate 3C, the same climate as that of the source building.

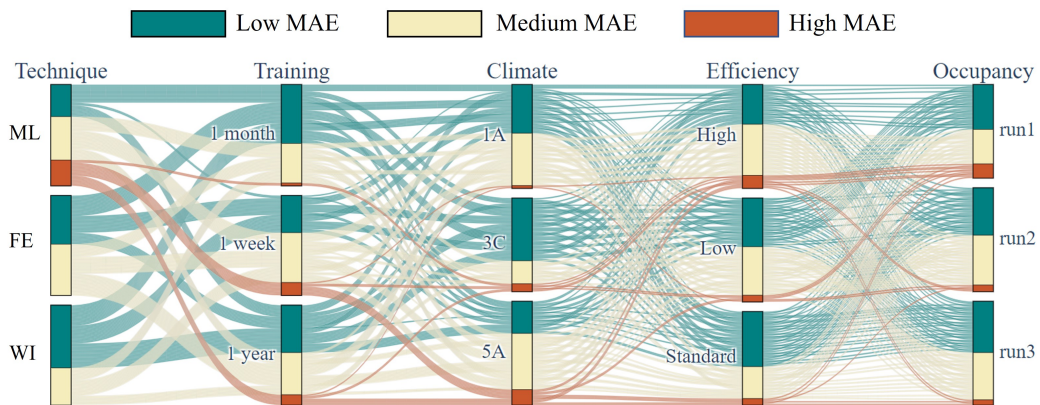


Fig. 5.11 Categorical plot of the error distribution for each technique over all the influencing factors [210]

Due to the co-occurrence of different features on the model (e.g, different climate, occupancy and efficiency levels), a specific analysis was performed by changing only one feature at a time, with the goal of isolating their effect on model performance. Figure 5.12 shows the MAE for different techniques for several cases. Furthermore, it shows how by changing only the efficiency level (same climate and same occupancy profile), transfer learning outperforms machine learning for every data availability, while negative transfer can occur when buildings across different climates are analysed, with very different results according to data availability. Looking at results with various occupancy profiles, a narrow performance improvement can be seen, with a negligible case of negative transfer learning, since both ML and TL techniques have an average error below 0.2 °C and very similar performance.

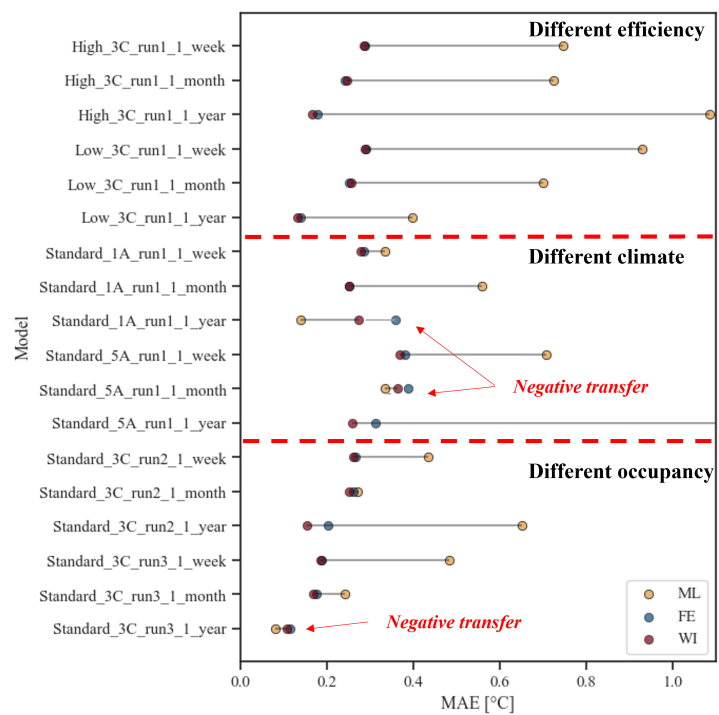


Fig. 5.12 Performance comparison with isolated effects of features [210]

Therefore, to assess the influence of climate and data availability on model performance, a specific analysis was conducted, as shown in Figure 5.13. In particular, Figure 5.13a shows the distribution of the mean absolute error for one week, one month and one year of data availability over the three different climates. For the sake of clarity, the error bar related to one year and Climate 5A, which exceeded 1.5 °C, has not been shown, while its lower outliers have been included in the figure.

Note that often an MAE of 0.5 °C is seen as threshold for the deployment of a model that predicts the internal air temperature. As a result, the figure highlights the inadequacy of ML models to be deployed for the specific combination of climate and time horizon. With increasing data availability, the median value of the ML models decreases. In general, TL approaches are more robust compared to ML approaches. Furthermore, the analysis showed how almost every TL model had an error below 0.5 °C, while ML often exceeded this threshold. Figure 5.13b uses the asymptotic performance improvement to compare the simulation point by point. It can be seen how, on average, the best performance improvements are achieved in Climate 3C (i.e., the climate selected for the source building). Note that performance improvements for climate 5A are highly influenced by the poor performance of ML models, increasing the advantages of using TL. The main reason may be related to the high temperature variation of Climate 5A, which makes it hard for the model to generalize over the entire year. However, Figure 5.13b also highlights the presence of negative transfer, especially with one month of data, a period in which ML already has good performance. As a result, a further analysis was conducted to identify the main driver of negative TL.

5.5.2 Negative transfer learning

Figure 5.14 shows the asymptotic performance of all the simulations, highlighting three particular areas: negative transfer learning, neutral transfer and effective transfer. Negative transfer occurs when the MAE is greater than 0.05 °C compared to ML models, neutral transfer is when it is smaller than 0.1 °C, and effective transfer reduces the MAE at least 0.1 °C. As can be seen, about 20% of cases have negative transfer, 20% have neutral transfer and 60% of cases show effective transfer. Figure 5.14b displays a detail of negative transfer, using different shapes and colors to highlight data availability and climate. The figure highlights how negative transfer occurred only 4 times out of 52 simulations, when one week of data was used (turquoise color), suggesting an effectiveness of TL in over 90% of the cases when one week of data is considered. It also can be noticed how negative transfer occurred only 4 times out of 52 simulations (diamond shape), with a performance increase in about 90% of the cases when the target building had the same climate as the source building. The figure shows that the highest amount of negative TL happened with

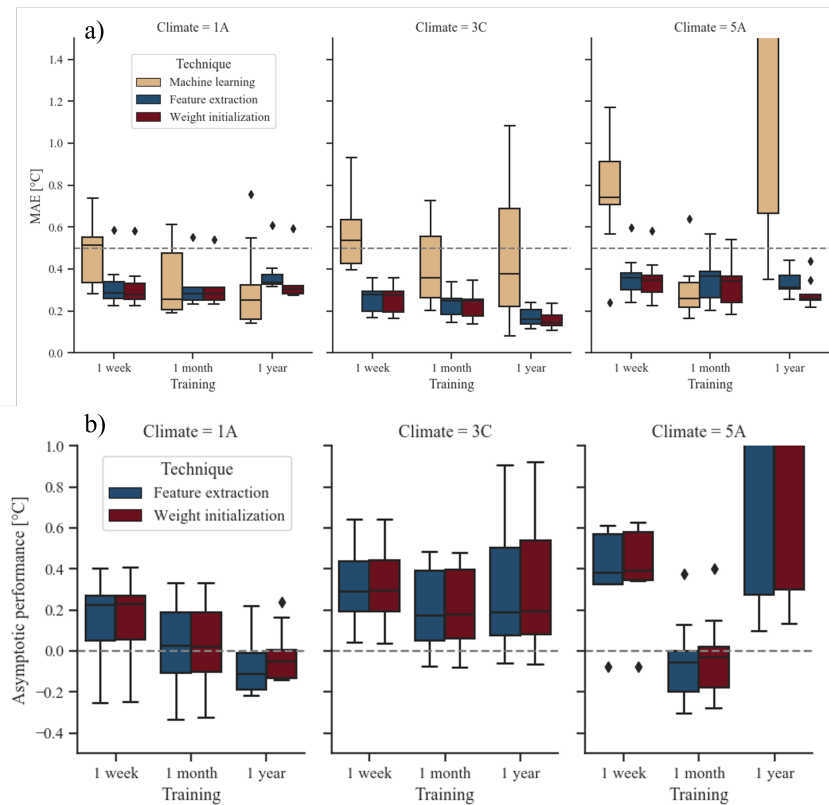


Fig. 5.13 Error distribution for each technique over different climate and data availability (top) Asymptotic performance for each technique over different climate and data availability (bottom) [210]

one month of data availability, identifying this amount of data as enough to obtain a good ML model performance.

Lastly, to provide a comparison of the model performance with effective and negative transfer, Figure 5.15 displays temperature evolution for the first predicted time-step over a random day for real values using ML and TL models. The figure on the left highlights how in this case ML was not able to properly describe the building dynamics, while both TL techniques follow the trend of the real value (green). On the other hand, the right figure shows a case of a negative TL, in which the performance of TL was still able to capture the building dynamic but perform worse than the classical ML approach.

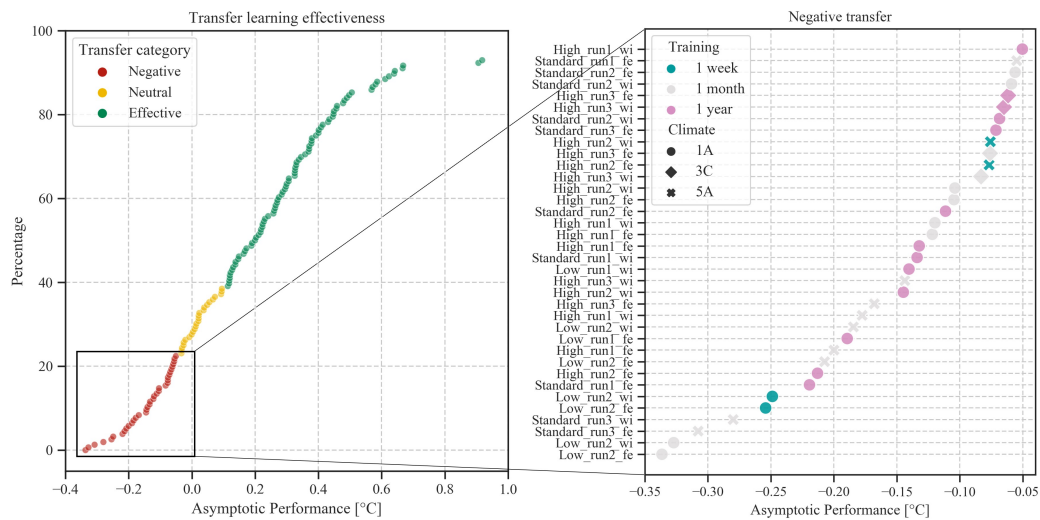


Fig. 5.14 Categorization of transfer learning effectiveness and negative transfer analysis [210]

5.5.3 Jumpstart performance

Figure 5.16 shows jumpstart performance with different data availability. As shown, the highest jumpstart occurred for one month and one week, reducing the MAE of the first epoch about 8°C . Despite this reduction, the final performance during training was comparable to an ML model. Moreover, the figure shows how the performance of transfer learning is almost constant, thus highlighting the possibility of great computational cost reduction when using transfer learning. Transfer learning also provided a computational advantage; however, the model complexity and the time required to train such models in this kind of problem is little. These advantages are usually more important when dealing with different applications, such as in computer vision. As a result, the jumpstart performance is a less reliable metric compared to the asymptotic performance, which is better suited to quantify the goodness of a model.

5.5.4 Online deployment

Figure 5.17 shows the MAE error distribution over each week of the year for the two techniques (ML and TL) deployed online. The configuration selected for the target building was characterized by Climate 3C (the same of the source building),

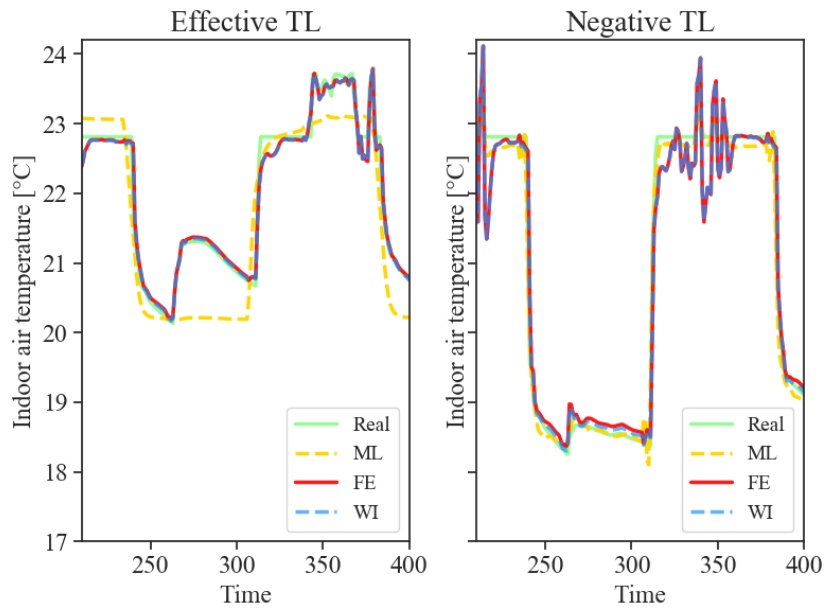


Fig. 5.15 Prediction evolution for the first time-step with different techniques for effective and negative TL [210]

occupancy pattern 2 and a high efficiency level. This configuration was selected on the basis of the outcome of the previous analysis. The transfer learning model (already trained on one year of source building data) was updated for the target building each week as new data became available following an anchored deployment configuration that employed existing data and a new week's data using the same learning rate of transfer learning configuration ($2 \cdot 10^{-3}$). The machine learning model used the same deployment strategy without leveraging pre-training data from the source building. To highlight both relative and absolute performance, Figure 5.17 reports a candlestick visualization. The green color of the box highlights the cases when the TL showed higher performance against ML, while the red box represents the opposite occurrence. The height of the box measures the difference in terms of performance between the two models, while the two extremes indicate the absolute value of MAE. The figure shows that especially during the first weeks of deployment, the ML had very poor performance when compared with TL. However, as training data became available for the ML, the performance difference between the two models tended to decrease, and after week 40, the performance of the two models were comparable.

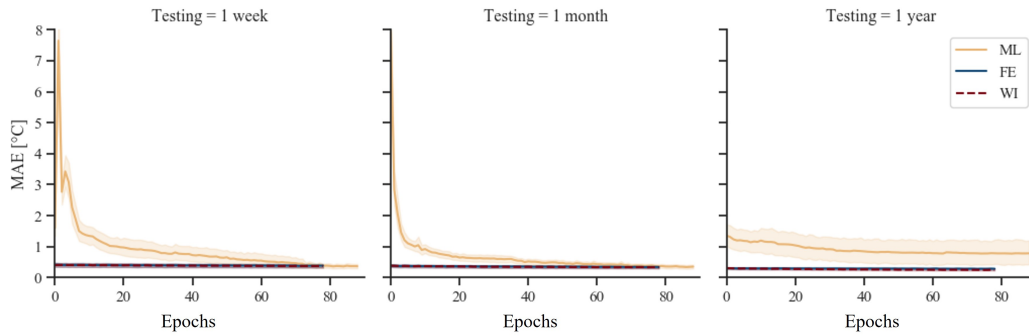


Fig. 5.16 Jumpstart comparison over different training time [210]

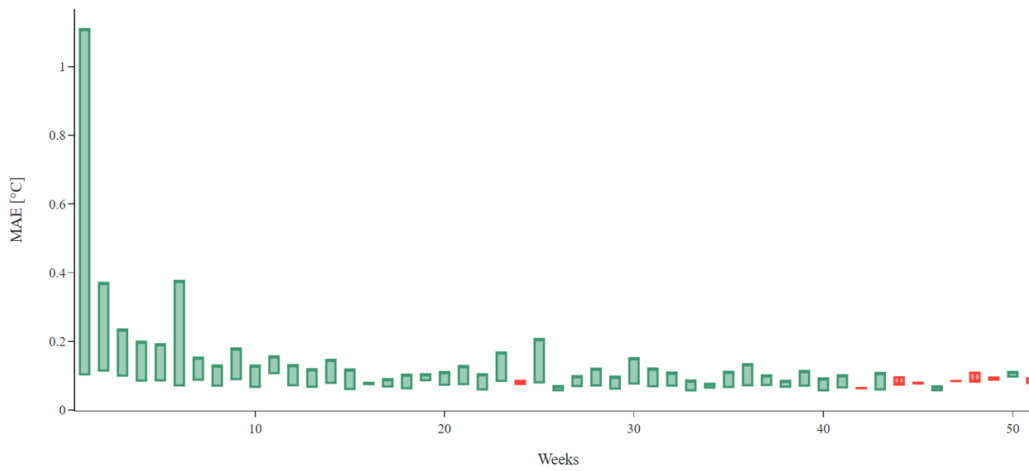


Fig. 5.17 Performance comparison between online ML and online TL [210]

5.6 Discussion

Building dynamics prediction proved to be effective to unlock the potential of advanced control strategies. However, the main bottleneck is represented by the data availability in most of the buildings, making the exploitation of data driven models a niche. TL promises to overcome this problem, but still requires further studies to quantify building similarity. This research aimed to quantify the feature importance of several variables in a TL setting. In particular, this study compared two TL techniques and assessed the effect of data availability and case specific features (e.g., climate, efficiency level, occupancy). To capture the effect of the different variables, an experiment design was conceived. Analysis of the results revealed several insights: first, unlike the ML models, the error performance over multiple time-steps is very small for the first time-step, increasing more steeply with the following time-steps,

while on average TL showed better performance. This information is particularly helpful for advanced predictive controllers, where an optimization process can be performed on the basis of the prediction. In particular, when the control time step is smaller than the prediction time horizon (e.g., control the energy system every 10 minute, while predicting the internal temperature for the next hour) the transfer learning approach can ensure higher performance especially during the first time steps. Additionally, analysis of the data availability aimed to assess how much data are necessary in TL and ML settings. The analysis showed that for ML a higher amount of data in the target building may be counterproductive, especially when the target building is located in climates with a great variation between the different seasons (Climate 5A). On the other hand, using a large amount of data helps to reduce the variance of TL models, obtaining more robust results.

Furthermore, the analysis confirmed the ability of TL to deal with different efficiency levels and occupancy, while limitations were observed for its effective applications across different climates, highlighting the role of external (outdoor air) temperature as the most important feature. Moreover, the focus on the asymptotic performance and the negative transfer allowed researchers to identify guidelines and constraints on the application of transfer learning for building dynamics prediction. In particular, the analysis showed how negative transfer mainly occurs when different climates are considered, identifying data-scarcity (one week) and the application on the same climate of the source building as the best case study to deploy TL. Furthermore, performance analysis suggest that for different features (e.g., climate), when new data are available the optimal solution consists in using online transfer learning and shifting to online machine learning when a robust dataset in the target building is available (e.g., one year). On the other hand, when the most important features are the same for the source and target building, transfer learning may achieve performance improvement independently from the amount of data used, highlighting its effectiveness in generalizing machine learning models. Lastly, a specific analysis was carried out on jumpstart performance; however, despite the computational advantages introduced by TL, the time needed to train such models is relatively low, due to the small dimension with respect to computer vision domains. This suggests the use of asymptotic performance as a key performance indicator to evaluate the effectiveness of TL.

Chapter 6

Conclusions

The dissertation aimed at proving the feasibility and effectiveness of data-driven controllers and models to support energy management in buildings at scale. Firstly, RL-based controllers have been used to control the energy systems of multiple buildings in two applications. The results showed that RL-based controllers can pursue multi-objective functions, increasing energy system performances while providing services to the grid. Despite their abilities, their application in energy management systems is still in its infancy, since it requires a high development and deployment cost, as well as heterogeneous background knowledge, that span from computer science to energy systems. The present dissertation leveraged an already existent control algorithm, trying to bridge the gap between computer science and its application in the building field, to highlight the additional advantages deriving from the deployment of these controllers at scale. In this context, the two applications designed faced different challenges, providing the following contributions:

- **Enhancing energy management in grid-interactive buildings**

The literature review described how advanced control strategies can enable energy flexibility in buildings by enhancing on-site renewable energy exploitation and storage operation, significantly reducing both energy costs and emissions. However, when energy management is faced shifting from a single building to a cluster of buildings, uncoordinated strategies may have negative effects on the grid reliability, causing undesirable new peaks. To overcome these limitations, the contribution explored the opportunity to take advantage of the mutual collaboration between single buildings by pursuing a coordinated

approach in energy management using deep reinforcement learning. The case study presented in Section 3.2 was composed of four buildings, whose thermal storage was controlled to reduce electricity cost and peak demand of the district. Despite the simplicity of the case study, which made use of precomputed demand, the work allowed to test a state-of-the-art control algorithm (SAC) on multiple buildings, leveraging its mixed entropy-reward maximization approach to obtain robust control strategies. The work also highlighted the role of the reward function, which should be conceived leveraging domain expertise to balance the multi-objective functions. Furthermore, the application showed the ability of a centralized controller to understand how to optimize the operation and coordination of multiple energy systems, achieving a reduction of operational costs of about 4%, together with a decrease of peak demand by up to 12%. Lastly, the control strategy allows for the reduction of the average daily peak and average peak-to-average ratio by 10% and 6% respectively, highlighting the benefits of a coordinated approach.

- **Comparing multi-agent architectures in grid-interactive buildings**

After having identified the potentialities of a coordinated approach for energy management in grid-interactive buildings using reinforcement learning, the work analysed in Section 3.3 different configurations of multi-agent reinforcement learning systems. In particular, it is considered a heterogeneous cluster of buildings with the presence of two prosumers able to export electricity, further complicating the energy management problem. Two multi-agent reinforcement learning methods were explored: a centralised (coordinated) controller and a decentralised (cooperative) controller, benchmarked against a rule-based controller. The two controllers were tested for three different climates, outperforming the rule-based controller by 3% and 7% respectively for cost, and 10% and 14% respectively for peak demand. It is interesting to notice that the work demonstrated how even if a coordinated approach may provide benefits for the sum of the users and the grid, specific users may be penalized. As a result, cooperative control strategies emerge as more suitable for districts with heterogeneous objectives within the individual buildings, also thanks to their ability to define and tune a reward function according to the needs of a specific user.

Furthermore, the work proposed an application that tried to combine data-driven models and controllers, to achieve a fully data-driven framework for district energy management, introducing the following innovations:

- **A methodology to combine data-driven models and data-driven controllers**

The proposed methodology makes use of a fully data-driven control scheme that exploits LSTM neural networks to simulate building thermal dynamics, allowing the exploitation of thermal mass, further used by DRL controllers to enhance energy management. The methodology leveraged synthetic data created in EnergyPlus, integrated into a simulation environment used to train and test the DRL controller. The controller managed the operation of heat pumps, chilled and domestic hot water storage for multiple buildings, comparing its performance with a manually optimized RBC. Results showed that the proposed approach was able to reduce the overall cluster electricity costs, while decreasing the peak energy demand by 23% and the peak to average ratio by 20%, without penalizing indoor temperature control. The application showed the potentiality for a district energy management fully based on data-driven approaches. The main advantage is related to a lower effort associated to the modelling phase and the possibility to create building thermal dynamic models exploiting building-related data. However, to optimize the control of the energy system within the district, a simplified simulation environment is still required to train the DRL controller.

Lastly, the thesis focused on the role of data-driven models to describe building thermal dynamics, highlighting the limitations related to their scalability. In this context, the thesis proposes an alternative, providing the following contributions:

- **Sharing building dynamic models to support energy management**

In recent years deep neural networks have been proposed as a lightweight data-driven model to capture complicated physical processes, supporting the deployment of advanced control strategies able to exploit building thermal mass. However, their reliance on a large amount of data needed for the training process clashes with the currently limited data availability in most buildings. To overcome this problem, transfer learning aims to improve the performance

of a target learner by exploiting knowledge from related environments, such as similar buildings. Nevertheless, there is a lack of approaches to evaluate building similarity to perform transfer learning. This thesis helped to quantify the feature importance of the most common variables adopted in a transfer learning setting, conducting a suite of experiments that leveraged 250 data-driven models to study the influence of data availability, building energy efficiency level, occupancy and climate on machine learning and transfer learning performances. The results of the analysis showed that climate and data availability are crucial factors for the application of transfer learning to building thermal dynamics models, suggesting the creation of archetypes for each climate, while showing that transfer learning can increase the performance when dealing with different occupancy schedules, efficiency level and low data availability, also highlighting its superiority with respect to a pure online machine learning approach.

The general purpose of the developed methodologies in the context of the thesis was to ease the implementation of energy management strategies at scale. Leveraging an energy engineer background, the work identified the main bottlenecks for their effective implementation, proposing an innovative solution that exploits artificial intelligence. However, the dissertation focused more on the development of a general framework for the application of data-driven models and controllers, rather than aiming to improve performance through the creation of new algorithms. Furthermore, the thesis analyzed the benefits provided by such methodologies at different scale, from single users to grid, aiming to identify potential advantages and limitations for different stakeholders and adopters of the technology. The technical outcomes of the present research were already discussed in the previous chapters, while the following provides an overview of the lessons learned throughout the creation of the proposed frameworks.

- **“The whole is greater than the sum of its part”: The advent of data-driven controllers**

While the potentialities of sharing distributed energy resources between multiple buildings are bursting with the creation of the so-called "energy communities", the extension of a shared ecosystem between multiple individual controllers still requires feasibility analysis to understand if additional advantages can offset the costs of development and implementation of advanced

controllers. The presented applications demonstrated the advantages of extending the scale of advanced energy management from single buildings to multiple buildings. Indeed, the cluster of building scale has enough flexibility to increase grid-interaction and renewable exploitation, also allowing peak and PAR reduction. This research aimed to quantify the benefit for the grid, which can be translated in additional revenue streams for the users, favoring the implementation of such control strategies. In this context, the adoption of advanced controllers faces some implementation barrier, including the detailed modeling of buildings and energy systems, that strongly limit their real-life use. The presented applications paved the way for a new concept of data-driven controller at cluster scale based on deep reinforcement learning, which strength is not only the mere improvement of performances, but the opportunity provided by their adaptive nature to account for the cluster environment evolution.

- **“The more the merrier?” The double role of the reward function: constraint and opportunity**

Despite the presented applications achieving better results with respect to standard RBC, it should be noticed that for their effective implementation a detailed study on the reward function was conducted, searching balances between grid advantages and single user objectives. Indeed, adding more building to the case study influences the magnitude of the reward function, changing the balance between its objective. Thus, domain expertise is fundamental to fine-tune the reward function, since without explicit constraint, reinforcement learning reward function maximization can penalize some users, as seen in some of the previous applications. Moreover, as the number of building increase, the possibility that they have different preferences increases, further complicating the problem. As a result, the thesis identified the sweet point for this kind of application in dozens of buildings that employs decentralized control, able to consider the needs of different users.

- **“Savings begin at the edge of your comfort zone”**

The application of data-driven controllers at multiple buildings scale has proven to be effective when a simplified model of the buildings is enough to control the energy systems. This situation is common in presence of storage or batteries, when it is possible to decouple demand and production, easily optimising the supply side. However, neglecting the intrinsic opportunity of the interaction

between demand and production and the role of the occupant and thermal mass leads to a substantial reduction in potential benefits. Indeed, exploiting the energy flexibility provided by the building thermal mass can result in additional savings, especially if involving the occupants in demand response programs. The application discussed in Chapter 4 evaluated the effectiveness of using data-driven models to support advanced energy management, leveraging building thermal mass. The dissertation presented a context in-between the more common model-based and model-free control, in which data can be used to create models further embedded in the simulation environment, allowing the exploitation of data-driven models at cluster scale.

- **“Sharing is caring”: transferring models from one building to another**

Current scientific literature highlighted the rising trend in both data-driven models and controllers that are leveraging machine learning for energy management in buildings. The thesis provided a precise overview of the main limitation of real-world implementation of these applications, including being too tailor-made to individual buildings. In this context, transfer learning is a potential approach for scaling up the use of machine learning models in real-world settings. This approach seeks to transfer a model learned for one system or task to a similar model or task with the least amount of modeling work. The application shown in Section 5.2 provided a detailed literature review on the application of transfer learning for smart buildings, with a particular interest in transferring building thermal dynamic models. In conclusion, transfer learning has been identified as an effective way to reuse building thermal dynamic models or control policies based on previously gained information, enhancing the scalability of advanced control strategies and decreasing their implementation cost.

- **“There is no need to reinvent the wheel”: Data-driven should not always mean model-free**

The dissertation aimed at proving the effectiveness of a “fully data-driven” approach for energy management in multiple buildings. However, as described within the application, the methodology leveraged a simplified simulation environment to describe energy systems using physical principles approaches, while modelling the building thermal dynamics employing neural networks. The combination of data-driven and engineering methods to describe the evo-

lution of the built environment should be conceived with the aim to speed-up energy simulation, while maintaining a certain level of accuracy and interpretability. Future research should be focused on studying how to integrate these two approaches, to obtain the best of both worlds.

- **“Knowing is not enough; we must apply”: business, as usual**

The applications developed in the context of the dissertation highlighted how artificial intelligence (data-driven models and controllers) can be leveraged to provide significant benefits to energy management in buildings. The higher the scale and the complexity of the energy systems, the greater the advantages for both the users and the grid. Despite so, their application is currently limited, due to the cost of development and implementation. However, as technology evolves and sensors and cloud services are becoming more affordable, economic savings justify the adoption of advanced control strategies. Furthermore, the thesis aimed to present how, once proving their effectiveness in real-world application, the adoption of data-driven controllers at scale can leverage additional economic revenues from grid services to increase their competitiveness over traditional control strategies.

References

- [1] A Capozzoli, T Cerquitelli, and M Piscitelli. Enhancing energy efficiency in buildings through innovative data analytics technologies. *Next Generation Platforms for Intelligent Data Collection*; Dobre, C., Xhafa, F., Eds, pages 353–389, 2016.
- [2] Miguel Molina-Solana, María Ros, M Dolores Ruiz, Juan Gómez-Romero, and M J Martin-Bautista. Data science for building energy management: A review. *Renewable and Sustainable Energy Reviews*, 70:598–609, 2017.
- [3] Ioannis Antonopoulos, Valentin Robu, Benoit Couraud, Desen Kirli, Sonam Norbu, Aristides Kiprakis, David Flynn, Sergio Elizondo-Gonzalez, and Steve Wattam. Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review. *Renewable and Sustainable Energy Reviews*, 130:109899, 2020.
- [4] Tianzhen Hong, Zhe Wang, Xuan Luo, and Wannan Zhang. State-of-the-art on research and applications of machine learning in the building life cycle. *Energy and Buildings*., 2020.
- [5] Kadir Amasyali and Nora M El-Gohary. A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, 81:1192–1205, 2018.
- [6] Anjukan Kathirgamanathan, Mattia De Rosa, Eleni Mangina, and Donal P Finn. Data-driven predictive control for unlocking building energy flexibility: A review. *Renewable and Sustainable Energy Reviews*, 135:110120, 2021.
- [7] Ján Drgoňa, Javier Arroyo, Iago Cupeiro Figueroa, David Blum, Krzysztof Arendt, Donghun Kim, Enric Perarnau Ollé, Juraj Oravec, Michael Wetter,

- Draguna L Vrabie, and Lieve Helsen. All you need to know about model predictive control for buildings. *Annual Reviews in Control*, 50:190–232, 2020.
- [8] Gianluca Serale, Massimo Fiorentini, Alfonso Capozzoli, Daniele Bernardini, and Alberto Bemporad. Model Predictive Control (MPC) for enhancing building and HVAC system energy efficiency: Problem formulation, applications and opportunities. *Energies*, 11(3), 2018.
- [9] Woohyun Kim and Srinivas Katipamula. A review of fault detection and diagnostics methods for building systems. *Science and Technology for the Built Environment*, 24(1):3–21, 2018.
- [10] Yassine Himeur, Khalida Ghanem, Abdullah Alsalemi, Faycal Bensaali, and Abbas Amira. Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives. *Applied Energy*, 287:116601, 2021.
- [11] Zhengwei Li, Yanmin Han, and Peng Xu. Methods for benchmarking building energy consumption against its past or intended performance: An overview. *Applied Energy*, 124:325–334, 2014.
- [12] Alfonso Capozzoli, Marco Savino Piscitelli, Francesco Neri, Daniele Grassi, and Gianluca Serale. A novel methodology for energy performance benchmarking of buildings by means of Linear Mixed Effect Model: The case of space and DHW heating of out-patient Healthcare Centres. *Applied Energy*, 171:592–607, 2016.
- [13] Gianfranco Chicco. Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy*, 42(1):68–80, 2012.
- [14] Yi Wang, Qixin Chen, Chongqing Kang, Mingming Zhang, Ke Wang, and Yun Zhao. Load profiling and its application to demand response: A review. *Tsinghua Science and Technology*, 20(2):117–129, 2015.
- [15] Marco Savino Piscitelli, Silvio Brandi, and Alfonso Capozzoli. Recognition and classification of typical load profiles in buildings with non-intrusive learning approach. *Applied Energy*, 255:113727, 2019.

- [16] Zhiyuan He, Tianzhen Hong, and S K Chou. A framework for estimating the energy-saving potential of occupant behaviour improvement. *Applied Energy*, 287:116591, 2021.
- [17] Hannah Kramer, Claire Curtin, Guanjing Lin, Eliot Crowe, and Jessica Granderson. Proving the Business Case for Building Analytics, 2020.
- [18] Marco Piscitelli. *Enhancing energy management in buildings through data analytics technologies*. PhD thesis, Politecnico di Torino, 2020.
- [19] Dasheng Lee and Chin-Chi Cheng. Energy savings by energy management systems: A review. *Renewable and Sustainable Energy Reviews*, 56:760–777, 2016.
- [20] Andrew Satchwell, Mary Ann Piette, Aditya Khandekar, Jessica Granderson, Natalie Mims Frick, Ryan Hledik, Ahmad Faruqui, Long Lam, Stephanie Ross, Jesse Cohen, and Others. A National Roadmap for Grid-Interactive Efficient Buildings. 2021.
- [21] Sonia Aggarwal and Robbie Orvis. Grid flexibility: Methods for modernizing the power grid. *Energy Innovation San Francisco, California March*, 2016.
- [22] Selina Kerscher and Pablo Arboleya. The key role of aggregators in the energy transition under the latest European regulatory framework. *International Journal of Electrical Power & Energy Systems*, 134:107361, 2022.
- [23] Merlinda Andoni, Valentin Robu, David Flynn, Simone Abram, Dale Geach, David Jenkins, Peter McCallum, and Andrew Peacock. Blockchain technology in the energy sector: A systematic review of challenges and opportunities. *Renewable and Sustainable Energy Reviews*, 100:143–174, 2019.
- [24] Silvio Brandi. *Deep Reinforcement Learning-based Control Strategies for Enhancing Energy Management in HVAC Systems*. PhD thesis, Politecnico di Torino, 2022.
- [25] Timothy I. Salsbury. A survey of control technologies in the building automation industry. *IFAC Proceedings Volumes*, 38(1):90–100, 2005. 16th IFAC World Congress.

- [26] Farinaz Behrooz, Norman Mariun, Mohammad Hamiruce Marhaban, Mohd Amran Mohd Radzi, and Abdul Rahman Ramli. Review of control techniques for hvac systems—nonlinearity approaches based on fuzzy cognitive maps. *Energies*, 11(3), 2018.
- [27] D. Subbaram Naidu and Craig G. Rieger. Advanced control strategies for heating, ventilation, air-conditioning, and refrigeration systems—an overview: Part i: Hard control. *HVAC&R Research*, 17(1):2–21, 2011.
- [28] D. Subbaram Naidu and Craig G. Rieger. Advanced control strategies for hvac&r systems—an overview: Part ii: Soft and fusion control. *HVAC&R Research*, 17(2):144–158, 2011.
- [29] Zhe Wang and Tianzhen Hong. Reinforcement learning for building controls: The opportunities and challenges. *Applied Energy*, 269:115036, 2020.
- [30] Rasmus Halvgaard, Niels Kjølstad Poulsen, Henrik Madsen, and John Bagterp Jørgensen. Economic model predictive control for building climate control in a smart grid. In *2012 IEEE PES Innovative Smart Grid Technologies (ISGT)*, pages 1–6, 2012.
- [31] Georgios D Kontes, Georgios I Giannakis, Víctor Sánchez, Pablo De Agustincamacho, Ander Romero-amorrortu, and Natalia Panagiotidou. Simulation-Based Evaluation and Optimization of Control Strategies in Buildings. *Energies*, 11:1–23, 2018.
- [32] Richard S Sutton and Andrew G Barto. Reinforcement Learning: An Introduction. *MIT press Cambridge*, 1998.
- [33] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dhharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, February 2015.
- [34] Giuseppe Pinto, Silvio Brandi, Josè Ramòn Vazquez-Canteli, Zoltán Nagy, and Alfonso Capozzoli. Towards Coordinated Energy Management in Buildings via Deep Reinforcement Learning.pdf. pages 1–14, 2020.

- [35] Giuseppe Pinto, Marco Savino Piscitelli, José Ramón Vázquez-Canteli, Zoltán Nagy, and Alfonso Capozzoli. Coordinated energy management for a cluster of buildings through deep reinforcement learning. *Energy*, 229:120725, 2021.
- [36] Giuseppe Pinto, Davide Deltetto, and Alfonso Capozzoli. Data-driven district energy management with surrogate models and deep reinforcement learning. *Applied Energy*, 304:117642, 2021.
- [37] Giuseppe Pinto, Anjukan Kathirgamanathan, Eleni Mangina, Donal P. Finn, and Alfonso Capozzoli. Enhancing energy management in grid-interactive buildings: A comparison among cooperative and coordinated architectures. *Applied Energy*, 310:118497, 2022.
- [38] Davide Deltetto, Davide Coraci, Giuseppe Pinto, Marco Savino Piscitelli, and Alfonso Capozzoli. Exploring the Potentialities of Deep Reinforcement Learning for Incentive-Based Demand Response in a Cluster of Small Commercial Buildings. *Energies*, 14(10), 2021.
- [39] Han Li, Zhe Wang, Tianzhen Hong, and Mary Ann Piette. Energy flexibility of residential buildings: A systematic review of characterization and quantification methods and applications. *Advances in Applied Energy*, 3:100054, 2021.
- [40] Dimitris Lazos, Alistair B. Sproul, and Merlinde Kay. Optimisation of energy management in commercial buildings with weather forecasting inputs: A review. *Renewable and Sustainable Energy Reviews*, 39:587–603, 2014.
- [41] IEA EBC Annex 67. Energy Flexible Buildings. 2014.
- [42] A. Fattahi Meyabadi and M.H. Deihimi. A review of demand-side management: Reconsidering theoretical framework. *Renewable and Sustainable Energy Reviews*, 80:367–379, 2017.
- [43] Ioannis Antonopoulos, Valentin Robu, Benoit Couraud, Desen Kirli, Sonam Norbu, Aristides Kiprakis, David Flynn, Sergio Elizondo-Gonzalez, and Steve Wattam. Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review. *Renewable and Sustainable Energy Reviews*, 130(June):109899, 2020.

- [44] John S Vardakas, Nizar Zorba, and Christos V Verikoukis. A survey on demand response in smart grids: Mathematical models and approaches. *IEEE Transactions on Industrial Informatics*, 11(3):570–582, 2015.
- [45] Karen Herter, Patrick McAuliffe, and Arthur Rosenfeld. An exploratory analysis of California residential customer response to critical peak pricing of electricity. *Energy*, 32(1):25–34, 2007.
- [46] Fariba Mousavi, Morteza Nazari-Heris, Behnam Mohammadi-Ivatloo, and Somayeh Asadi. Chapter 1 - energy market fundamentals and overview. In Behnam Mohammadi-Ivatloo, Amin Mohammadpour Shotorbani, and Amjad Anvari-Moghaddam, editors, *Energy Storage in Energy Markets*, pages 1–21. Academic Press, 2021.
- [47] Limei Shen, Zhengwei Li, and Yongjun Sun. Performance evaluation of conventional demand response at building-group-level under different electricity pricings. *Energy and Buildings*, 128:143–154, 2016.
- [48] Yanxin Chai, Yue Xiang, Junyong Liu, Chenghong Gu, Wentao Zhang, and Weiting Xu. Incentive-based demand response model for maximizing benefits of electricity retailers. *Journal of Modern Power Systems and Clean Energy*, 7(6):1644–1650, 2019.
- [49] Madhur Behl, Francesco Smarra, and Rahul Mangharam. Dr-advisor: A data-driven demand response recommender system. *Applied Energy*, 170:30–46, 2016.
- [50] Helen Stopps. *Smart Thermostat Use in Multi-Unit Residential Buildings: Impacts on Occupant Behaviour, Thermal Comfort, and Energy Performance*. PhD thesis, 2021. Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Ultimo aggiornamento - 2022-01-05.
- [51] Farshad Etedadi Aliabadi, Kodjo Agbossou, Souso Kelouwani, Nilson Henao, and Sayed Saeed Hosseini. Coordination of Smart Home Energy Management Systems in Neighborhood Areas: A Systematic Review. *IEEE Access*, 9:36417–36443, 2021.

- [52] Maomao Hu, Fu Xiao, and Shengwei Wang. Neighborhood-level coordination and negotiation techniques for managing demand-side flexibility in residential microgrids. *Renewable and Sustainable Energy Reviews*, 135:110248, 2021.
- [53] Adam Hirsch, Yael Parag, and Josep Guerrero. Microgrids: A review of technologies, key drivers, and outstanding issues. *Renewable and Sustainable Energy Reviews*, 90:402–411, 2018.
- [54] Kathryn Kaspar, Mohamed Ouf, and Ursula Eicker. A critical review of control schemes for demand-side energy management of building clusters. *Energy and Buildings*, 257:111731, 2022.
- [55] Andong Wang, Rongling Li, and Shi You. Development of a data driven approach to explore the energy flexibility potential of building clusters. *Applied Energy*, 232:89–100, 2018.
- [56] Ilaria Vigna, Roberta Perneti, Wilmer Pasut, and Roberto Lollini. New domain for promoting energy efficiency: Energy flexible building cluster. *Sustainable Cities and Society*, 38:526–533, 2018.
- [57] Hussain Kazmi, Merel Keijsers, Fahad Mehmood, and Clayton Miller. Energy balances, thermal performance, and heat stress: Disentangling occupant behaviour and weather influences in a dutch net-zero energy neighborhood. *Energy and Buildings*, 263:112020, 2022.
- [58] Ayako Taniguchi, Takuya Inoue, Masaya Otsuki, Yohei Yamaguchi, Yoshiyuki Shimoda, Akinobu Takami, and Kanako Hanaoka. Estimation of the contribution of the residential sector to summer peak demand reduction in japan using an energy end-use simulation model. *Energy and Buildings*, 112:80–92, 2016.
- [59] Cristian Perfumo, Ernesto Kofman, Julio H. Braslavsky, and John K. Ward. Load management: Model-based control of aggregate power for populations of thermostatically controlled loads. *Energy Conversion and Management*, 55:36–48, 2012.
- [60] Maomao Hu and Fu Xiao. Quantifying uncertainty in the aggregate energy flexibility of high-rise residential building clusters considering stochastic occupancy and occupant behavior. *Energy*, 194:116838, 2020.

- [61] Olaf van Pruissen, Vincent Kamphuis, Armin van der Togt, and Ewoud Werkman. A thermal grid coordinated by a multi agent energy management system. In *IEEE PES ISGT Europe 2013*, pages 1–5, 2013.
- [62] Gabriele Comodi, Andrea Giantomassi, Marco Severini, Stefano Squartini, Francesco Ferracuti, Alessandro Fonti, Davide Nardi Cesarini, Matteo Morodo, and Fabio Polonara. Multi-apartment residential microgrid with electrical and thermal storage devices: Experimental analysis and simulation of energy management strategies. *Applied Energy*, 137:854–866, 2015.
- [63] Pei Huang, Cheng Fan, Xingxing Zhang, and Jiayuan Wang. A hierarchical coordinated demand response control for buildings with improved performances at building group. *Applied Energy*, 242(March):684–694, 2019.
- [64] Farhad Angizeh, Ali Ghofrani, Esmat Zaidan, and Mohsen A. Jafari. Adaptable scheduling of smart building communities with thermal mapping and demand flexibility. *Applied Energy*, 310:118445, 2022.
- [65] Xiaolong Jin, Jianzhong Wu, Yunfei Mu, Mingshen Wang, Xiandong Xu, and Hongjie Jia. Hierarchical microgrid energy management in an office building. *Applied Energy*, 208:480–494, 2017.
- [66] Mehdi Ganji and Mohammad Shahidehpour. 5 - development of a residential microgrid using home energy management systems. In Lisa Ann Lamont and Ali Sayigh, editors, *Application of Smart Grid Technologies*, pages 173–192. Academic Press, 2018.
- [67] Christie Etukudor, Benoit Couraud, Valentin Robu, Wolf-Gerrit Früh, David Flynn, and Chinonso Okereke. Automated negotiation for peer-to-peer electricity trading in local energy markets. *Energies*, 13(4), 2020.
- [68] Duong Tung Nguyen and Long Bao Le. Joint optimization of electric vehicle and home energy scheduling considering user comfort preference. *IEEE Transactions on Smart Grid*, 5(1):188–199, 2014.
- [69] Mosaddek Hossain Kamal Tushar, Chadi Assi, Martin Maier, and Mohammad Faisal Uddin. Smart microgrids: Optimal joint scheduling for electric vehicles and home appliances. *IEEE Transactions on Smart Grid*, 5(1):239–250, 2014.

- [70] Ahmed Ouammi. Optimal power scheduling for a cooperative network of smart residential buildings. *IEEE Transactions on Sustainable Energy*, 7(3):1317–1326, 2016.
- [71] Thillainathan Logenthiran, Dipti Srinivasan, and Tan Zong Shun. Demand side management in smart grid using heuristic optimization. *IEEE Transactions on Smart Grid*, 3(3):1244–1252, 2012.
- [72] Anna Magdalena Kosek, Giuseppe Tommaso Costanzo, Henrik W. Bindner, and Oliver Gehrke. An overview of demand side management control schemes for buildings in smart grids. In *2013 IEEE International Conference on Smart Energy Grid Engineering (SEGE)*, pages 1–9, 2013.
- [73] Christoph Molitor, Milahi Marin, Lisette Hernández, and Antonello Monti. Decentralized coordination of the operation of residential heating units. In *IEEE PES ISGT Europe 2013*, pages 1–5, 2013.
- [74] Wesley J. Cole, Joshua D. Rhodes, William Gorman, Krystian X. Perez, Michael E. Webber, and Thomas F. Edgar. Community-scale residential air conditioning control for effective grid management. *Applied Energy*, 130:428–436, 2014.
- [75] Amir Safdarian, Mahmud Fotuhi-Firuzabad, and Matti Lehtonen. Optimal residential load management in smart grids: A decentralized framework. *IEEE Transactions on Smart Grid*, 7(4):1836–1845, 2016.
- [76] Phani Chavali, Peng Yang, and Arye Nehorai. A distributed algorithm of appliance scheduling for home energy management system. *IEEE Transactions on Smart Grid*, 5(1):282–290, 2014.
- [77] Robin Roche, Siddharth Suryanarayanan, Timothy M. Hansen, Sila Kiliccote, and Abdellatif Miraoui. A multi-agent model and strategy for residential demand response coordination. In *2015 IEEE Eindhoven PowerTech*, pages 1–6, 2015.
- [78] Amir-Hamed Mohsenian-Rad, Vincent W. S. Wong, Juri Jatskevich, Robert Schober, and Alberto Leon-Garcia. Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid. *IEEE Transactions on Smart Grid*, 1(3):320–331, 2010.

- [79] M. Reyasudin Basir Khan, Razali Jidin, and Jagadeesh Pasupuleti. Multi-agent based distributed control architecture for microgrid energy management and optimization. *Energy Conversion and Management*, 112:288–307, 2016.
- [80] T Chang, M Alizadeh, and A Scaglione. Real-Time Power Balancing Via Decentralized Coordinated Home Energy Scheduling. *IEEE Transactions on Smart Grid*, 4(3):1490–1504, 2013.
- [81] Javier Arroyo, Carlo Manna, Fred Spiessens, and Lieve Helsen. Reinforced model predictive control (rl-mpc) for building energy management. *Applied Energy*, 309:118346, 2022.
- [82] Timothy I. Salsbury. A survey of control technologies in the building automation industry. *IFAC Proceedings Volumes*, 38:90–100, 2005.
- [83] Peyton Young and Shmuel Zamir. *Handbook of game theory*. Elsevier, 2014.
- [84] Walid Saad, Zhu Han, H. Vincent Poor, and Tamer Basar. Game-theoretic methods for the smart grid: An overview of microgrid systems, demand-side management, and smart grid communications. *IEEE Signal Processing Magazine*, 29(5):86–105, 2012.
- [85] Z. Beheshti. A review of population-based meta-heuristic algorithm. In *SOCO 2013*, 2013.
- [86] Jun Zhang, Zhi-hui Zhan, Ying Lin, Ni Chen, Yue-jiao Gong, Jing-hui Zhong, Henry S.H. Chung, Yun Li, and Yu-hui Shi. Evolutionary computation meets machine learning: A survey. *IEEE Computational Intelligence Magazine*, 6(4):68–75, 2011.
- [87] Zhe Wang and Tianzhen Hong. Reinforcement Learning for Building Controls: The problem, opportunities and challenges. *Applied Energy*, 269(1):300, 2020.
- [88] José R. Vázquez-Canteli and Zoltán Nagy. Reinforcement learning for demand response: A review of algorithms and modeling techniques. *Applied Energy*, 235(April 2018):1072–1089, 2019.
- [89] Eduardo F Camacho and Carlos Bordons Alba. *Model predictive control*. Springer science & business media, 2013.

- [90] Ji Hoon Yoon, Ross Bladick, and Atila Novoselac. Demand response for residential buildings based on dynamic price of electricity. *Energy and Buildings*, 80:531–541, 2014.
- [91] Fatma Mtibaa, Kim-Khoa Nguyen, Vasken Dermardiros, and Mohamed Cheriet. Context-aware model predictive control framework for multi-zone buildings. *Journal of Building Engineering*, 42:102340, 2021.
- [92] Young Ran Yoon and Hyeun Jun Moon. Performance based thermal comfort control (ptcc) using deep reinforcement learning for space cooling. *Energy and Buildings*, 203:109420, 2019.
- [93] David P. Chassin, Jakob Stoustrup, Panajotis Agathoklis, and Nedjib Djilali. A new thermostat for real-time price demand response: Cost, comfort and energy impacts of discrete-time control without deadband. *Applied Energy*, 155:816–825, 2015.
- [94] Maomao Hu, Fu Xiao, John Bagterp Jørgensen, and Shengwei Wang. Frequency control of air conditioners in response to real-time dynamic electricity prices in smart grids. *Applied Energy*, 242:92–106, 2019.
- [95] Bingqing Chen, Jonathan Francis, Marco Pritoni, Soumya Kar, and Mario Bergés. Cohort: Coordination of heterogeneous thermostatically controlled loads for demand flexibility. In *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, BuildSys '20, page 31–40, New York, NY, USA, 2020. Association for Computing Machinery.
- [96] Wei Gu, Haojun Yu, Wei Liu, Junpeng Zhu, and Xiaohui Xu. Demand response and economic dispatch of power systems considering large-scale plug-in hybrid electric vehicles/electric vehicles (phevs/evs): A review. *Energies*, 6(9):4394–4417, 2013.
- [97] Corey D. White and K. Max Zhang. Using vehicle-to-grid technology for frequency regulation and peak-load reduction. *Journal of Power Sources*, 196(8):3972–3980, 2011.
- [98] G. Barone, A. Buonomano, F. Calise, C. Forzano, and A. Palombo. Building to vehicle to building concept toward a novel zero energy paradigm: Modelling

- and case studies. *Renewable and Sustainable Energy Reviews*, 101:625–648, 2019.
- [99] Raul Martinez Oviedo, Zhong Fan, Sedat Gormus, and Parag Kulkarni. A residential phev load coordination mechanism with renewable sources in smart grids. *International Journal of Electrical Power & Energy Systems*, 55:511–521, 2014.
- [100] Francesco Calise, Francesco Liberato Cappiello, Massimo Dentice d’Accadia, and Maria Vicidomini. Smart grid energy district based on the integration of electric vehicles and combined heat and power generation. *Energy Conversion and Management*, 234:113932, 2021.
- [101] Nikolaos G. Paterakis, Ozan Erdinç, Iliana N. Pappi, Anastasios G. Bakirtzis, and João P. S. Catalão. Coordinated operation of a neighborhood of smart households comprising electric vehicles, energy storage and distributed generation. *IEEE Transactions on Smart Grid*, 7(6):2736–2747, 2016.
- [102] B. Svetozarevic, C. Baumann, S. Muntwiler, L. Di Natale, M.N. Zeilinger, and P. Heer. Data-driven control of room temperature and bidirectional ev charging using deep reinforcement learning: Simulations and experiments. *Applied Energy*, 307:118127, 2022.
- [103] Frauke Oldewurtel, Andreas Ulbig, Manfred Morari, and Göran Andersson. Building control and storage management with dynamic tariffs for shaping demand response. In *2011 2nd IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies*, pages 1–8, 2011.
- [104] Jason Leadbetter and Lukas Swan. Battery storage system for residential electricity peak demand shaving. *Energy and Buildings*, 55:685–692, 2012. Cool Roofs, Cool Pavements, Cool Cities, and Cool World.
- [105] Ozan Erdinc. Economic impacts of small-scale own generating and storage units, and electric vehicles under different demand response strategies for smart households. *Applied Energy*, 126:142–150, 2014.
- [106] Jianing Li, Zhi Wu, Suyang Zhou, Hao Fu, and Xiao-Ping Zhang. Aggregator service for pv and battery energy storage systems of residential building. *CSEE Journal of Power and Energy Systems*, 1(4):3–11, 2015.

- [107] Faeze Brahman, Masoud Honarmand, and Shahram Jadid. Optimal electrical and thermal energy management of a residential energy hub, integrating demand response and energy storage system. *Energy and Buildings*, 90:65–75, 2015.
- [108] Behrang Alimohammadisagvand, Juha Jokisalo, Simo Kilpeläinen, Mubbashir Ali, and Kai Sirén. Cost-optimal thermal energy storage system for a residential building with heat pump heating and demand response control. *Applied Energy*, 174:275–287, 2016.
- [109] Massimo Fiorentini, Josh Wall, Zhenjun Ma, Julio H. Braslavsky, and Paul Cooper. Hybrid model predictive control of a residential hvac system with on-site thermal energy generation and storage. *Applied Energy*, 187:465–479, 2017.
- [110] W.J.N. Turner, I.S. Walker, and J. Roux. Peak load reductions: Electric load shifting with mechanical pre-cooling of residential buildings with low thermal mass. *Energy*, 82:1057–1067, 2015.
- [111] J. Le Dréau and P. Heiselberg. Energy flexibility of residential buildings using short term heat storage in the thermal mass. *Energy*, 111:991–1002, 2016.
- [112] G. Reynders, T. Nuytten, and D. Saelens. Potential of structural thermal mass for demand-side management in dwellings. *Building and Environment*, 64:187–199, 2013.
- [113] D.F. Dominković, P. Gianniou, M. Münster, A. Heller, and C. Rode. Utilizing thermal building mass for storage in district heating systems: Combined building level simulations and system level optimization. *Energy*, 153(C):949–966, 2018.
- [114] Amjad Anvari-Moghaddam, Ashkan Rahimi-Kian, Maryam S. Mirian, and Josep M. Guerrero. A multi-agent based energy management solution for integrated buildings and microgrid system. *Applied Energy*, 203:41–56, 2017.
- [115] Naren Srivaths Raman, Ninad Gaikwad, Prabir Barooah, and Sean P. Meyn. Reinforcement learning-based home energy management system for resiliency. In *2021 American Control Conference (ACC)*, pages 1358–1364, 2021.

- [116] Akın Taşcıkaraoğlu, Ali Rıfat Boynueğri, and Mehmet Uzunoglu. A demand side management strategy based on forecasting of residential renewable sources: A smart home system in turkey. *Energy and Buildings*, 80:309–320, 2014.
- [117] Matteo Bilardo, Maria Ferrara, and Enrico Fabrizio. Performance assessment and optimization of a solar cooling system to satisfy renewable energy ratio (rer) requirements in multi-family buildings. *Renewable Energy*, 155:990–1008, 2020.
- [118] Osman A. Hamed, Hamed A. Al-Washmi, and Holayil A. Al-Otaibi. Thermoeconomic analysis of a power/water cogeneration plant. *Energy*, 31(14):2699–2709, 2006.
- [119] Cuo Zhang, Yan Xu, Zhengmao Li, and Zhao Yang Dong. Robustly coordinated operation of a multi-energy microgrid with flexible electric and thermal loads. *IEEE Transactions on Smart Grid*, 10(3):2765–2775, 2019.
- [120] Elisa Guelpa, Giulia Barbero, Adriano Sciacovelli, and Vittorio Verda. Peak-shaving in district heating systems through optimal management of the thermal request of buildings. *Energy*, 137:706–714, 2017.
- [121] Giuseppe Pinto, Elnaz Abdollahi, Alfonso Capozzoli, Laura Savoldi, and Risto Lahdelma. Optimization and multicriteria evaluation of carbon-neutral technologies for district heating. *Energies*, 12(9), 2019.
- [122] Dorota A. Chwieduk. Towards modern options of energy conservation in buildings. *Renewable Energy*, 101:1194–1202, 2017.
- [123] Hans Auer and Reinhard Haas. On integrating large shares of variable renewables into the electricity system. *Energy*, 115:1592–1601, 2016.
- [124] Haider Tarish Haider, Ong Hang See, and Wilfried Elmenreich. A review of residential demand response of smart grid. *Renewable and Sustainable Energy Reviews*, 59:166–178, 2016.
- [125] Pierluigi Siano. Demand response and smart grids—a survey. *Renewable and Sustainable Energy Reviews*, 30:461–478, 2014.

- [126] Yi Liu, Chau Yuen, Shisheng Huang, Naveed Ul Hassan, Xiumin Wang, and Shengli Xie. Peak-to-average ratio constrained demand-side management with consumer's preference in residential smart grid. *IEEE Journal of Selected Topics in Signal Processing*, 8(6):1084–1097, 2014.
- [127] Hassan Bevrani, Arindam Ghosh, and Gerard Ledwich. Renewable energy sources and frequency regulation: survey and new perspectives. *IET Renewable Power Generation*, 4(5):438–457, 2010.
- [128] Luis Pérez-Lombard, José Ortiz, and Christine Pout. A review on buildings energy consumption information. *Energy and Buildings*, 40(3):394–398, 2008.
- [129] Abhinandana Boodi, Karim Beddiar, Malek Benamour, Yassine Amirat, and Mohamed Benbouzid. Intelligent systems for building energy and occupant comfort optimization: A state of the art review and recommendations. *Energies*, 11(10), 2018.
- [130] Paul Centolella, Mindi Farber-DeAnda, LA Greening, and T Kim. Estimates of the value of uninterrupted service for the mid-west independent system operator. *Science Applications International Corporation, McLean*, 2006.
- [131] Zakia Afroz, GM Shafiullah, Tania Urmee, and Gary Higgins. Modeling techniques used in building hvac control systems: A review. *Renewable and Sustainable Energy Reviews*, 83:64–84, 2018.
- [132] í Ciglera, Dimitrios Gyalistrasb, Vinh-Nghi Tietd, Luká, and Ferkla. Beyond theory : the challenge of implementing model predictive control in buildings ji ř. 2013.
- [133] Krzysztof Arendt, Muhyiddine Jradi, Hamid Reza Shaker, and Christian Veje. Comparative analysis of white-, gray-and black-box models for thermal simulation of indoor environment: Teaching building case study. In *Proceedings of the 2018 Building Performance Modeling Conference and SimBuild co-organized by ASHRAE and IBPSA-USA, Chicago, IL, USA*, pages 26–28, 2018.
- [134] David H. Blum, Nora Xu, and Leslie K. Norford. A novel multi-market optimization problem for commercial heating, ventilation, and air-conditioning

- systems providing ancillary services using multi-zone inverse comprehensive room transfer functions. *Science and Technology for the Built Environment*, 22(6):783–797, 2016.
- [135] Xiwang Li and Jin Wen. Review of building energy modeling for control and operation. *Renewable and Sustainable Energy Reviews*, 37:517–537, 2014.
- [136] Shengwei Wang, Xue Xue, and Chengchu Yan. Building power demand response methods toward smart grid. *HVAC&R Research*, 20(6):665–687, 2014.
- [137] Rodrigo Verschae, Hiroaki Kawashima, Takekazu Kato, and Takashi Matsuyama. Coordinated energy management for inter-community imbalance minimization. *Renewable Energy*, 87:922–935, 2016. Optimization Methods in Renewable Energy Systems Design.
- [138] Elena Mocanu, Decebal Constantin Mocanu, Phuong H. Nguyen, Antonio Liotta, Michael E. Webber, Madeleine Gibescu, and J. G. Slootweg. On-line building energy optimization using deep reinforcement learning. *IEEE Transactions on Smart Grid*, 10(4):3698–3708, 2019.
- [139] Renzhi Lu and Seung Ho Hong. Incentive-based demand response for smart grid with reinforcement learning and deep neural network. *Applied Energy*, 236:937–949, 2019.
- [140] Xiaodi Wang, Youbo Liu, Junbo Zhao, Chang Liu, Junyong Liu, and Jinyue Yan. Surrogate model enabled deep reinforcement learning for hybrid energy community operation. *Applied Energy*, 289:116722, 2021.
- [141] Thomas Schreiber, Sören Eschweiler, Marc Baranski, and Dirk Müller. Application of two promising reinforcement learning algorithms for load shifting in a cooling supply system. *Energy and Buildings*, 229:110490, 2020.
- [142] Silvio Brandi, Marco Savino Piscitelli, Marco Martellacci, and Alfonso Capozzoli. Deep reinforcement learning to optimise indoor temperature control and heating energy consumption in buildings. *Energy and Buildings*, 224:110225, 2020.
- [143] Gregor P Henze and Moncef Krarti. Predictive optimal control of active and passive building thermal storage inventory. 1 2003.

- [144] Eric O'Shaughnessy, Dylan Cutler, Kristen Ardani, and Robert Margolis. Solar plus: Optimization of distributed solar pv through battery storage and dispatchable load in residential buildings. *Applied Energy*, 213:11–21, 2018.
- [145] Lei Yang, Zoltan Nagy, Philippe Goffin, and Arno Schlueter. Reinforcement learning for optimal control of low exergy buildings. *Applied Energy*, 156:577–586, 2015.
- [146] Jose R Vazquez-Canteli, Gregor Henze, and Zoltan Nagy. Marlisa: Multi-agent reinforcement learning with iterative sequential action selection for load shaping of grid-interactive connected buildings. In *Proceedings of the 7th ACM international conference on systems for energy-efficient buildings, cities, and transportation*, pages 170–179, 2020.
- [147] Pei Huang, Cheng Fan, Xingxing Zhang, and Jiayuan Wang. A hierarchical coordinated demand response control for buildings with improved performances at building group. *Applied Energy*, 242:684–694, 2019.
- [148] Abigail D. Ondeck, Thomas F. Edgar, and Michael Baldea. Impact of rooftop photovoltaics and centralized energy storage on the design and operation of a residential chp system. *Applied Energy*, 222:280–299, 2018.
- [149] Rui Tang and Shengwei Wang. Model predictive control for thermal energy storage and thermal comfort optimization of building demand response in smart grids. *Applied Energy*, 242:873–882, 2019.
- [150] M. Robillart, P. Schalbart, F. Chaplais, and B. Peuportier. Model reduction and model predictive control of energy-efficient buildings for electrical heating load shifting. *Journal of Process Control*, 74:23–34, 2019. Efficient energy management.
- [151] Sebastian Gonzato, Joseph Chimento, Edward O'Dwyer, Gonzalo Bustos-Turu, Salvador Acha, and Nilay Shah. Hierarchical price coordination of heat pumps in a building network controlled using model predictive control. *Energy and Buildings*, 202:109421, 2019.
- [152] Karl Mason and Santiago Grijalva. A review of reinforcement learning for autonomous building energy management. *Computers & Electrical Engineering*, 78:300–312, 2019.

- [153] Davide Coraci, Silvio Brandi, Marco Savino Piscitelli, and Alfonso Capozzoli. Online implementation of a soft actor-critic agent to enhance indoor temperature control and energy efficiency in buildings. *Energies*, 14(4), 2021.
- [154] F. Ruelens, B. J. Claessens, S. Quaiyum, B. De Schutter, R. Babuška, and R. Belmans. Reinforcement learning applied to an electric water heater: From theory to practice. *IEEE Transactions on Smart Grid*, 9(4):3792–3800, 2018.
- [155] H. Kazmi, S. D’Oca, C. Delmastro, S. Lodeweyckx, and S.P. Corgnati. Generalizable occupant-driven optimization model for domestic hot water production in nzeb. *Applied Energy*, 175:1–15, 2016.
- [156] José Vázquez-Canteli, Jérôme Kämpf, and Zoltán Nagy. Balancing comfort and energy consumption of a heat pump using batch reinforcement learning with fitted q-iteration. *Energy Procedia*, 122:415–420, 2017. CISBAT 2017 International Conference Future Buildings & Districts – Energy Efficiency from Nano to Urban Scale.
- [157] José R. Vázquez-Canteli and Zoltán Nagy. Reinforcement learning for demand response: A review of algorithms and modeling techniques. *Applied Energy*, 235:1072–1089, 2019.
- [158] P. Kofinas, A.I. Dounis, and G.A. Vouros. Fuzzy Q-Learning for multi-agent decentralized energy management in microgrids. *Applied Energy*, 219(C):53–67, 2018.
- [159] Dawei Qiu, Yujian Ye, Dimitrios Papadaskalopoulos, and Goran Strbac. Scalable coordinated management of peer-to-peer energy trading: A multi-cluster deep reinforcement learning approach. *Applied Energy*, 292:116940, 2021.
- [160] Flora Charbonnier, Thomas Morstyn, and Malcolm D. McCulloch. Scalable multi-agent reinforcement learning for distributed control of residential energy flexibility. *Applied Energy*, 314:118825, 2022.
- [161] Vijaykumar Gullapalli. A stochastic reinforcement learning algorithm for learning real-valued functions. *Neural Networks*, 3(6):671–692, 1990.
- [162] Michael L. Littman, Thomas L. Dean, and Leslie Pack Kaelbling. On the complexity of solving markov decision problems. In *Proceedings of the*

- Eleventh Conference on Uncertainty in Artificial Intelligence*, UAI'95, page 394–402, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc.
- [163] Ofir Nachum, Mohammad Norouzi, Kelvin Xu, and Dale Schuurmans. Bridging the gap between value and policy based reinforcement learning. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 2772–2782, Red Hook, NY, USA, 2017. Curran Associates Inc.
- [164] Michael L. Littman. Markov games as a framework for multi-agent reinforcement learning. In William W. Cohen and Haym Hirsh, editors, *Machine Learning Proceedings 1994*, pages 157–163. Morgan Kaufmann, San Francisco (CA), 1994.
- [165] Daniel S. Bernstein, Shlomo Zilberstein, and Neil Immerman. The complexity of decentralized control of markov decision processes. In *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, UAI'00, page 32–37, San Francisco, CA, USA, 2000. Morgan Kaufmann Publishers Inc.
- [166] Eric A. Hansen, Daniel S. Bernstein, and Shlomo Zilberstein. Dynamic programming for partially observable stochastic games. In *Proceedings of the 19th National Conference on Artificial Intelligence*, AAAI'04, page 709–715. AAAI Press, 2004.
- [167] Christopher J. C. H. Watkins and Peter Dayan. Technical note: q-learning. *Mach. Learn.*, 8(3–4):279–292, may 1992.
- [168] Liangpeng Zhang, Ke Tang, and Xin Yao. Explicit planning for efficient exploration in reinforcement learning. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.
- [169] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. 2013. cite arxiv:1312.5602Comment: NIPS Deep Learning Workshop 2013.
- [170] Satinder P. Singh, Tommi S. Jaakkola, and Michael I. Jordan. Learning without state-estimation in partially observable markovian decision processes.

- In *Proceedings of the Eleventh International Conference on International Conference on Machine Learning*, ICML'94, page 284–292, San Francisco, CA, USA, 1994. Morgan Kaufmann Publishers Inc.
- [171] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *35th International Conference on Machine Learning, ICML 2018*, volume 5, pages 2976–2989, 2018.
- [172] Samuel L. Smith, Pieter-Jan Kindermans, and Quoc V. Le. Don't decay the learning rate, increase the batch size. *CoRR*, abs/1711.00489, 2017.
- [173] Tuomas Haarnoja, Aurick Zhou, Kristian Hartikainen, George Tucker, Sehoon Ha, Jie Tan, Vikash Kumar, Henry Zhu, Abhishek Gupta, Pieter Abbeel, and Sergey Levine. Soft Actor-Critic Algorithms and Applications. 2018.
- [174] Abien Fred Agarap. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018.
- [175] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.
- [176] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. Openai gym, 2016.
- [177] Guido Van Rossum and Fred L Drake Jr. *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.
- [178] Pecan Street Dataport. Pecanstreet.
- [179] NREL. Solarrow.
- [180] NREL. Resstock.
- [181] 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". Technical report, Standard, UNI—Ente Nazionale Italiano di Unificazione, Italy, 2019.

- [182] Andrei Marinescu, Ivana Dusparic, and Siobhán Clarke. Prediction-based multi-agent reinforcement learning in inherently non-stationary environments. *ACM Trans. Auton. Adapt. Syst.*, 12(2), may 2017.
- [183] Gregor P. Henze and Jobst Schoenmann. Evaluation of reinforcement learning control for thermal energy storage systems. *HVAC&R Research*, 9(3):259–275, 2003.
- [184] Austin Energy. Austin energy electricity tariff, 2020. Available at <https://austinenergy.com/ae/>.
- [185] Y S Foo, Eddy, H B Gooi, and S X Chen. Multi-Agent System for Distributed Management of Microgrids. *IEEE Transactions on Power Systems*, 30(1):24–34, 2015.
- [186] Gabriel Santos, Tiago Pinto, Hugo Morais, Tiago M Sousa, Ivo F Pereira, Ricardo Fernandes, Isabel Praça, and Zita Vale. Multi-agent simulation of competitive electricity markets: Autonomous systems cooperation for European market modeling. *Energy Conversion and Management*, 99:387–399, 2015.
- [187] Muhammad Waseem Khan, Jie Wang, Meiling Ma, Linyun Xiong, Penghan Li, and Fei Wu. Optimal energy management and control aspects of distributed microgrid using multi-agent systems. *Sustainable Cities and Society*, 44:855–870, 2019.
- [188] Christos-Spyridon Karavas, George Kyriakarakos, Konstantinos G Arvanitis, and George Papadakis. A multi-agent decentralized energy management system based on distributed intelligence for the design and control of autonomous polygeneration microgrids. *Energy Conversion and Management*, 103:166–179, 2015.
- [189] Mohamed A Mohamed, Tao Jin, and Wencong Su. Multi-agent energy management of smart islands using primal-dual method of multipliers. *Energy*, 208:118306, 2020.
- [190] Cunbin Li, Xuefeng Jia, Ying Zhou, and Xiaopeng Li. A microgrids energy management model based on multi-agent system using adaptive weight and

- chaotic search particle swarm optimization considering demand response. *Journal of Cleaner Production*, 262:121247, 2020.
- [191] Shunping Jin, Shoupeng Wang, and Fang Fang. Game theoretical analysis on capacity configuration for microgrid based on multi-agent system. *International Journal of Electrical Power & Energy Systems*, 125:106485, 2021.
- [192] Dawei Qiu, Yujian Ye, Dimitrios Papadaskalopoulos, and Goran Strbac. Scalable coordinated management of peer-to-peer energy trading: A multi-cluster deep reinforcement learning approach. *Applied Energy*, 292:116940, 2021.
- [193] Esmat Samadi, Ali Badri, and Reza Ebrahimpour. Decentralized multi-agent based energy management of microgrid using reinforcement learning. *International Journal of Electrical Power & Energy Systems*, 122:106211, 2020.
- [194] Renzhi Lu, Yi-Chang Li, Yuting Li, Junhui Jiang, and Yuemin Ding. Multi-agent deep reinforcement learning based demand response for discrete manufacturing systems energy management. *Applied Energy*, 276:115473, 2020.
- [195] Linyun Xiong, Penghan Li, Ziqiang Wang, and Jie Wang. Multi-agent based multi objective renewable energy management for diversified community power consumers. *Applied Energy*, 259:114140, 2020.
- [196] Timilehin Labeodan, Kennedy Aduda, Gert Boxem, and Wim Zeiler. On the application of multi-agent systems in buildings for improved building operations, performance and smart grid interaction – A survey. *Renewable and Sustainable Energy Reviews*, 50:1405–1414, 2015.
- [197] Hussain Kazmi, Johan Suykens, Attila Balint, and Johan Driesen. Multi-agent reinforcement learning for modeling and control of thermostatically controlled loads. *Applied Energy*, 238:1022–1035, 2019.
- [198] Srinarayana Nagarathinam, Vishnu Menon, Arunchandar Vasan, and Anand Sivasubramaniam. MARCO - Multi-Agent Reinforcement Learning Based CONTROL of Building HVAC Systems. In *Proceedings of the Eleventh ACM International Conference on Future Energy Systems*, e-Energy '20, pages 57–67, New York, NY, USA, 2020. Association for Computing Machinery.

- [199] P Kofinas, A I Dounis, and G A Vouros. Fuzzy Q-Learning for multi-agent decentralized energy management in microgrids. *Applied Energy*, 219:53–67, 2018.
- [200] Bin Zhang, Weihao Hu, Di Cao, Tao Li, Zhenyuan Zhang, Zhe Chen, and Frede Blaabjerg. Soft actor-critic –based multi-objective optimized energy conversion and management strategy for integrated energy systems with renewable energy. *Energy Conversion and Management*, 243:114381, 2021.
- [201] Davide Coraci, Silvio Brandi, Marco Savino Piscitelli, and Alfonso Capozzoli. Online implementation of a soft actor-critic agent to enhance indoor temperature control and energy efficiency in buildings. *Energies*, 14(4), 2021.
- [202] Marco Biemann, Fabian Scheller, Xiufeng Liu, and Lizhen Huang. Experimental evaluation of model-free reinforcement learning algorithms for continuous HVAC control. *Applied Energy*, 298:117164, 2021.
- [203] Gauraang Dhamankar, Jose R Vazquez-Canteli, and Zoltan Nagy. Benchmarking Multi-Agent Deep Reinforcement Learning Algorithms on a Building Energy Demand Coordination Task. *RLEM 2020 - Proceedings of the 1st International Workshop on Reinforcement Learning for Energy Management in Buildings and Cities*, pages 15–19, 2020.
- [204] Anjukan Kathirgamanathan, Eleni Mangina, and Donal P Finn. Development of a Soft Actor Critic Deep Reinforcement Learning Approach to a Virtual Large Office Building for Harnessing Energy Flexibility. *Energy and AI (under review)*, pages 1–32, 2021.
- [205] Silvio Brandi, Marco Savino Piscitelli, Marco Martellacci, and Alfonso Capozzoli. Deep reinforcement learning to optimise indoor temperature control and heating energy consumption in buildings. *Energy and Buildings*, 224:110225, 2020.
- [206] Jose R Vazquez-Canteli, Sourav Dey, Gregor Henze, and Zoltan Nagy. CityLearn: Standardizing Research in Multi-Agent Reinforcement Learning for Demand Response and Urban Energy Management, 2020.
- [207] Entergy. Entergy electricity tariff, 2020.

- [208] John Clauß, Christian Finck, Pierre Vogler-finck, and Paul Beagon. Control strategies for building energy systems to unlock demand side flexibility – A review. In *Proc. of BS2017: 15th Conference of International Building Performance Simulation Association, San Fransisco, USA, Aug 7-9, San Fransisco, 2017*.
- [209] Gregor P. Henze and Jobst Schoenmann. Evaluation of reinforcement learning control for thermal energy storage systems. *HVAC&R Research*, 9(3):259–275, 2003.
- [210] Giuseppe Pinto, Riccardo Messina, Han Li, Tianzhen Hong, Marco Savino Piscitelli, and Alfonso Capozzoli. Sharing is caring: An extensive analysis of parameter-based transfer learning for the prediction of building thermal dynamics. *Energy and Buildings*, page 112530, 2022.
- [211] H.S. Hippert, C.E. Pedreira, and R.C. Souza. Neural networks for short-term load forecasting: a review and evaluation. *IEEE Transactions on Power Systems*, 16(1):44–55, 2001.
- [212] A.S. Ahmad, M.Y. Hassan, M.P. Abdullah, H.A. Rahman, F. Hussin, H. Abdullah, and R. Saidur. A review on applications of ann and svm for building electrical energy consumption forecasting. *Renewable and Sustainable Energy Reviews*, 33:102–109, 2014.
- [213] Liang Zhang, Jin Wen, Yanfei Li, Jianli Chen, Yunyang Ye, Yangyang Fu, and William Livingood. A review of machine learning in building load prediction. *Applied Energy*, 285(January):116452, 2021.
- [214] Wan He. Load forecasting via deep neural networks. *Procedia Computer Science*, 122:308–314, 2017. 5th International Conference on Information Technology and Quantitative Management, ITQM 2017.
- [215] Daniel L. Marino, Kasun Amarasinghe, and Milos Manic. Building energy load forecasting using deep neural networks. In *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*, pages 7046–7051, 2016.
- [216] Clayton Miller and Forrest Meggers. The building data genome project: An open, public data set from non-residential building electrical meters. *Energy*

- Procedia*, 122:439–444, 2017. CISBAT 2017 International Conference Future Buildings & Districts – Energy Efficiency from Nano to Urban Scale.
- [217] Qiong Li, Qinglin Meng, Jiejun Cai, Hiroshi Yoshino, and Akashi Mochida. Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks. *Energy Conversion and Management*, 50(1):90–96, 2009.
- [218] G Mihalakakou, M Santamouris, and A Tsangrassoulis. On the energy consumption in residential buildings. *Energy and Buildings*, 34(7):727–736, 2002.
- [219] Merih Aydinalp, V. Ismet Ugursal, and Alan S. Fung. Modeling of the space and domestic hot-water heating energy-consumption in the residential sector using neural networks. *Applied Energy*, 79(2):159–178, 2004.
- [220] Zhe Wang, Tianzhen Hong, and Mary Ann Piette. Building thermal load prediction through shallow machine learning and deep learning. *Applied Energy*, 263:114683, 2020.
- [221] Abdullatif E. Ben-Nakhi and Mohamed A. Mahmoud. Cooling load prediction for buildings using general regression neural networks. *Energy Conversion and Management*, 45(13):2127–2141, 2004.
- [222] A.E. Ruano, E.M. Crispim, E.Z.E. Conceição, and M.M.J.R. Lúcio. Prediction of building's temperature using neural networks models. *Energy and Buildings*, 38(6):682–694, 2006.
- [223] Chunhua Sun, Jiali Chen, Shanshan Cao, Xiaoyu Gao, Guoqiang Xia, Chengying Qi, and Xiangdong Wu. A dynamic control strategy of district heating substations based on online prediction and indoor temperature feedback. *Energy*, 235:121228, 2021.
- [224] Xin Shi, Weiding Lu, Ying Zhao, and Pengjie Qin. Prediction of indoor temperature and relative humidity based on cloud database by using an improved bp neural network in chongqing. *IEEE Access*, 6:30559–30566, 2018.
- [225] Andrew Kusiak and Guanglin Xu. Modeling and optimization of hvac systems using a dynamic neural network. *Energy*, 42(1):241–250, 2012. 8th World Energy System Conference, WESC 2010.

- [226] G. Mustafaraj, G. Lowry, and J. Chen. Prediction of room temperature and relative humidity by autoregressive linear and nonlinear neural network models for an open office. *Energy and Buildings*, 43(6):1452–1460, 2011.
- [227] Zakia Afroz, Tania Urmee, G.M. Shafiullah, and Gary Higgins. Real-time prediction model for indoor temperature in a commercial building. *Applied Energy*, 231:29–53, 2018.
- [228] Hao Huang, Lei Chen, and Eric Hu. A neural network-based multi-zone modelling approach for predictive control system design in commercial buildings. *Energy and Buildings*, 97:86–97, 2015.
- [229] Jan Drgona, Damien Picard, Michal Kvasnica, and Lieve Helsen. Approximate model predictive building control via machine learning. *Applied Energy*, 218(February):199–216, 2018.
- [230] Fatma Mtibaa, Kim-Khoa Nguyen, Muhammad Azam, Anastasios Papachristou, Jean-Simon Venne, and Mohamed Cheriet. Lstm-based indoor air temperature prediction framework for hvac systems in smart buildings. *Neural Comput. Appl.*, 32(23):17569–17585, dec 2020.
- [231] Chengliang Xu, Huanxin Chen, Jiangyu Wang, Yabin Guo, and Yue Yuan. Improving prediction performance for indoor temperature in public buildings based on a novel deep learning method. *Building and Environment*, 148:128–135, 2019.
- [232] Matthew J. Ellis and Venkatesh Chinde. An encoder–decoder lstm-based empc framework applied to a building hvac system. *Chemical Engineering Research and Design*, 160:508–520, 2020.
- [233] Zhen Fang, Nicolas Crimier, Lisa Scanu, Alphanie Midelet, Amr Alyafi, and Benoit Delinchant. Multi-zone indoor temperature prediction with lstm-based sequence to sequence model. *Energy and Buildings*, 245:111053, 2021.
- [234] M. Raissi, P. Perdikaris, and G.E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

- [235] Felix Bünnig, Benjamin Huber, Adrian Schalbetter, Ahmed Aboudonia, Mathias Hudoba de Badyn, Philipp Heer, Roy S. Smith, and John Lygeros. Physics-informed linear regression is competitive with two machine learning methods in residential building mpc. *Applied Energy*, 310:118491, 2022.
- [236] Gargya Gokhale, Bert Claessens, and Chris Develder. Physics informed neural networks for control oriented thermal modeling of buildings. *ArXiv*, abs/2111.12066, 2021.
- [237] Ján Drgoňa, Aaron R. Tuor, Vikas Chandan, and Draguna L. Vrabie. Physics-constrained deep learning of multi-zone building thermal dynamics. *Energy and Buildings*, 243:110992, 2021.
- [238] Loris Di Natale, Bratislav Svetozarevic, Philipp Heer, and Colin N Jones. Physically consistent neural networks for building thermal modeling: theory and analysis. *arXiv preprint arXiv:2112.03212*, 2021.
- [239] Giuseppe Pinto, Zhe Wang, Abhishek Roy, Tianzhen Hong, and Alfonso Capozzoli. Transfer learning for smart buildings: A critical review of algorithms, applications, and future perspectives. *Advances in Applied Energy*, 5:100084, 2022.
- [240] Sinno Jialin Pan and Qiang Yang. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.
- [241] Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, and Chunfang Liu. A survey on deep transfer learning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11141 LNCS:270–279, 2018.
- [242] Zhuangdi Zhu, Kaixiang Lin, and Jiayu Zhou. Transfer Learning in Deep Reinforcement Learning: A Survey. pages 1–23, 2020.
- [243] Luqin Fan, Jing Zhang, Yu He, Ying Liu, Tao Hu, and Heng Zhang. Optimal scheduling of microgrid based on deep deterministic policy gradient and transfer learning. *Energies*, 14(3):1–15, 2021.
- [244] Paulo Lissa, Michael Schukat, Marcus Keane, and Enda Barrett. Transfer learning applied to DRL-Based heat pump control to leverage microgrid energy efficiency. *Smart Energy*, page 100044, 2021.

- [245] Brida V Mbuwir, Kaveh Paridari, Fred Spiessens, Lars Nordström, and Geert Deconinck. Transfer learning for operational planning of batteries in commercial buildings. In *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pages 1–6. IEEE, 2020.
- [246] Paulo Lissa, Michael Schukat, and Enda Barrett. Transfer Learning Applied to Reinforcement Learning-Based HVAC Control. *SN Computer Science*, 1, 2020.
- [247] Shichao Xu, Yixuan Wang, Yanzhi Wang, Zheng O’Neill, and Qi Zhu. One for many: Transfer learning for building hvac control. In *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, pages 230–239, 2020.
- [248] Xiangyu Zhang, Xin Jin, Charles Tripp, David J Biagioni, Peter Graf, and Huaiguang Jiang. Transferable reinforcement learning for smart homes. In *Proceedings of the 1st International Workshop on Reinforcement Learning for Energy Management in Buildings & Cities*, pages 43–47, 2020.
- [249] Nathan Tsang, Collin Cao, Serena Wu, Zilin Yan, Ashkan Yousefi, Alexander Fred-Ojala, and Ikhlaz Sidhu. Autonomous household energy management using deep reinforcement learning. In *2019 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, pages 1–7. IEEE, 2019.
- [250] Zhanhong Jiang and Young M Lee. Deep transfer learning for thermal dynamics modeling in smart buildings. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 2033–2037. IEEE, 2019.
- [251] Yujiao Chen, Zheming Tong, Yang Zheng, Holly Samuelson, and Leslie Norford. Transfer learning with deep neural networks for model predictive control of hvac and natural ventilation in smart buildings. *Journal of Cleaner Production*, 254:119866, 2020.
- [252] Yujiao Chen, Yang Zheng, and Holly Samuelson. Fast Adaptation of Thermal Dynamics Model for Predictive Control of HVAC and Natural Ventilation Using Transfer Learning with Deep Neural Networks. In *2020 American Control Conference (ACC)*, pages 2345–2350. IEEE, 2020.

- [253] Md Monir Hossain, Tianyu Zhang, and Omid Ardakanian. Evaluating the feasibility of reusing pre-trained thermal models in the residential sector. *UrbSys 2019 - Proceedings of the 1st ACM International Workshop on Urban Building Energy Sensing, Controls, Big Data Analysis, and Visualization, Part of BuildSys 2019*, pages 23–32, 2019.
- [254] Hussain Kazmi, Johan Suykens, and Johan Driesen. Large-scale transfer learning for data-driven modelling of hot water systems. *Building Simulation Conference Proceedings*, 4:2611–2618, 2019.
- [255] Weizheng Hu, Yong Luo, Zongqing Lu, and Yonggang Wen. Heterogeneous transfer learning for thermal comfort modeling. In *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, pages 61–70, 2019.
- [256] Thomas Grubinger, Georgios C. Chasparis, and Thomas Natschläger. Generalized online transfer learning for climate control in residential buildings. *Energy and Buildings*, 139:63–71, 2017.
- [257] Department of Energy. Commercial reference buildings, 2022.
- [258] Yixing Chen, Tianzhen Hong, and Xuan Luo. An agent-based stochastic occupancy simulator. *Building Simulation*, 11:37–49, 2017.
- [259] Han Li, Zhe Wang, and Tianzhen Hong. A synthetic building operation dataset. *Scientific Data*, 8:213, 2021.
- [260] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework, 2019.

Appendix A

A.1 CityLearn Documentation

A.1.1 Input Attributes

- *data_path* - path indicating where the data is
- *building_attributes* - name of the file containing the characteristics of the energy supply and storage systems of the buildings
- *weather_file* - name of the file containing the weather variables
- *solar_profile* - name of the file containing the solar generation profile (generation per kW of installed power)
- *electricity_profile* - name of the file containing the electricity price profile (per kWh of electricity)
- *building_ids* - list with the building IDs of the buildings to be simulated
- *buildings_states_actions* - name of the file containing the states and actions to be returned or taken by the environment
- *simulation_period* - hourly time period to be simulated. (0, 8759) by default: one year.
- *cost_function* - list with the cost functions to be minimized.
- *central_agent* - allows using CityLearn in central agent mode or in decentralized agents mode. If True, CityLearn returns a list of observations, a single

reward, and takes a list of actions. If False, CityLearn will allow the easy implementation of decentralized RL agents by returning a list of lists (as many as the number of building) of states, a list of rewards (one reward for each building), and will take a list of lists of actions (one for every building).

- *verbose* - set to 0 if you don't want CityLearn to print out the cumulated reward of each episode and set it to 1 if you do

A.1.2 Internal Attributes

- *net_electric_consumption* - district net electricity consumption
- *net_electric_consumption_no_storage* - district net electricity consumption if there were no cooling storage and DHW storage
- *net_electric_consumption_no_pv_no_storage* - district net electricity consumption if there were no cooling storage, DHW storage and PV generation
- *electric_consumption_dhw_storage* - electricity consumed in the district to increase DHW energy stored (when > 0) and electricity that the decrease in DHW energy stored saves from consuming in the district (when < 0).
- *electric_consumption_cooling_storage* - electricity consumed in the district to increase cooling energy stored (when > 0) and electricity that the decrease in cooling energy stored saves from consuming in the district (when < 0).
- *electric_consumption_dhw* - electricity consumed to satisfy the DHW demand of the district
- *electric_consumption_cooling* - electricity consumed to satisfy the cooling demand of the district
- *heat_pump_performance* – output the heat pump coefficient of performance and capacity as function of partial load ratio and external temperature
- *heat_pump_temperature_perfromance* - output the heat pump coefficient of performance and capacity as function of external temperature only
- *electric_consumption_appliances* - non-shiftable electricity consumed by appliances

- *electric_generation* - electricity generated in the district

A.1.3 CityLearn Methods

- *get_state_action_spaces()* - returns state-action spaces for all the buildings
- *next_hour()* - advances simulation to the next time-step
- *get_building_information()* - returns attributes of the buildings that can be used by the RL agents (i.e. to implement building-specific RL agents based on their attributes, or control buildings with correlated demand profiles by the same agent)
- *get_baseline_cost()* - returns the costs of a Rule-based controller (RBC), which is used to divide the final cost by it.
- *cost()* - returns the normalized cost of the environment after it has been simulated. $cost < 1$ when the controller's performance is better than the RBC.

A.1.4 Methods inherited from OpenAI Gym

- *step()* - advances simulation to the next time-step and takes an action based on the current state
- *_get_ob()* - returns all the states
- *_terminal()* - returns True if the simulation has ended
- *seed()* - specifies a random seed

A.1.5 States

- *month* - 1 (January) through 12 (December)
- *day* - type of day as provided by EnergyPlus (from 1 to 8). 1 (Sunday), 2 (Monday), ..., 7 (Saturday), 8 (Holiday)
- *hour* - hour of day (from 1 to 24).

- *t_out* - outdoor temperature in Celcius degrees.
- *t_out_pred_6h* - outdoor temperature predicted 6h ahead (accuracy: +-0.3C)
- *t_out_pred_12h* - outdoor temperature predicted 12h ahead (accuracy: +-0.65C)
- *t_out_pred_24h* - outdoor temperature predicted 24h ahead (accuracy: +-1.35C)
- *direct_solar_rad* - direct solar radiation in W/m^2 .
- *direct_solar_rad_pred_6h* - direct solar radiation predicted 6h ahead (accuracy: +-2.5%)
- *direct_solar_rad_pred_12h* - direct solar radiation predicted 12h ahead (accuracy: +-5%)
- *direct_solar_rad_pred_24h* - direct solar radiation predicted 24h ahead (accuracy: +-10%)
- *electricity_price* – electricity price in \$/kWh
- *electricity_price_pred_1h* – electricity price predicted 1h ahead
- *electricity_price_pred_2h* – electricity price predicted 2h ahead
- *electricity_price_pred_3h* – electricity price predicted 3h ahead
- *t_in* - indoor temperature in Celcius degrees.
- *non_shiftable_load* - electricity currently consumed by electrical appliances in kWh.
- *solar_gen* - electricity currently being generated by photovoltaic panels in kWh.
- *cooling_storage_soc* - state of the charge (SOC) of the cooling storage device. From 0 (no energy stored) to 1 (at full capacity).
- *dhw_storage_soc* - state of the charge (SOC) of the domestic hot water (DHW) storage device. From 0 (no energy stored) to 1 (at full capacity).

- *net_electricity_consumption* - net electricity consumption of the building (including all energy systems) in the current time step in kWh
- *total_load* – power withdrawn from the grid by the cluster of buildings in kW.

A.1.6 Actions

- *cooling_storage* - increase (action > 0) or decrease (action < 0) of the amount of cooling energy stored in the cooling storage device. $-1.0 \leq \text{action} \leq 1.0$ (attempts to decrease or increase the cooling energy stored in the storage device by an amount equal to the action times the storage device's maximum capacity). In order to decrease the energy stored in the device (action < 0), the energy must be released into the building's cooling system. Therefore, the state of charge will not decrease proportionally to the action taken if the demand for cooling of the building is lower than the action times the maximum capacity of the cooling storage device.
- *dhw_storage* - increase (action > 0) or decrease (action < 0) of the amount of DHW stored in the DHW storage device. $-1.0 \leq \text{action} \leq 1.0$ (attempts to decrease or increase the DHW stored in the storage device by an amount equivalent to action times its maximum capacity). In order to decrease the energy stored in the device, the energy must be released into the building. Therefore, the state of charge will not decrease proportionally to the action taken if the demand for DHW of the building is lower than the action times the maximum capacity of the DHW storage device.

A.1.7 Rewards

For a central single-agent (if CityLearn class attribute `central_agent = True`):

- *reward_function_sa* – it takes the total net electricity consumption of each building (< 0 if generation is higher than demand) at every time-step as input and returns a single reward for the central agent. For a decentralized multi-agent controller (if CityLearn class attribute `central_agent = False`):
- *reward_function_ma* - class that can take building information and the number agents when instantiated. It contains a “`get_rewards()`” method that takes

the total net electricity consumption of each building (< 0 if generation is higher than demand) at every time-step as input and returns a list with as many rewards as the number of agents.

A.1.8 Evaluation metrics

There are multiple KPIs available, which are all defined as a function of the total nonnegative net electricity consumption of the whole neighborhood:

- *electrical_cost* – total costs for the cluster of buildings
- *average_daily_peak* - average daily peak net demand.
- *peak_demand* - maximum peak electricity demand
- *net_electricity_consumption* - total amount of electricity consumed
- *PAR* – Peak-to-average ratio
- *average_daily_PAR* – average daily PAR
- *Flexibility_factor* – ratio between energy consumed during on peak period and total energy

A.2 Deep reinforcement learning hyperparameters

Table A.1 list the SAC hyperparameters for the two architectures, along with the optimization space analysed. Table A.2 shows the hyperparameter of the control problems, along with the final configuration selected to perform the analysis, while Figure A.1 displays the evolution of the reward function with the number of episodes.

Table A.1 Settings of the DRL hyperparameters for coordinated and cooperative architectures

| Hyperparameter | Coordinated controller | Cooperative controller | Search Space |
|--------------------------------------|------------------------|------------------------|--------------------|
| DNN architecture | 4 Layers (2 hidden) | 4 Layers (2 hidden) | - |
| Neurons per hidden layer | 256 | 64 | [64,128,256] |
| DNN Optimiser | Adam | Adam | - |
| Batch size | 512 | 512 | - |
| Learning rate (λ) | 0.001 | 0.001 | [0.001,0.005,0.01] |
| Discount rate (γ) | 0.99 | 0.99 | [0.9,0.95,0.99] |
| Decay rate (τ) | 0.005 | 0.005 | [0.001,0.005,0.01] |
| Temperature coefficient (α) | 0.05 | 0.05 | [0.01,0.05,0.1] |

Table A.2 Settings of the control problem hyperparameters for coordinated and cooperative architectures

| Hyperparameter | Coordinated controller | Cooperative controller |
|---------------------|------------------------|------------------------|
| Learning starts | 2208 | 2208 |
| Target model update | 1 | 1 |
| Episode Length | 2208 Control steps | 2208 Control steps |
| Training Episodes | 5 | 5 |

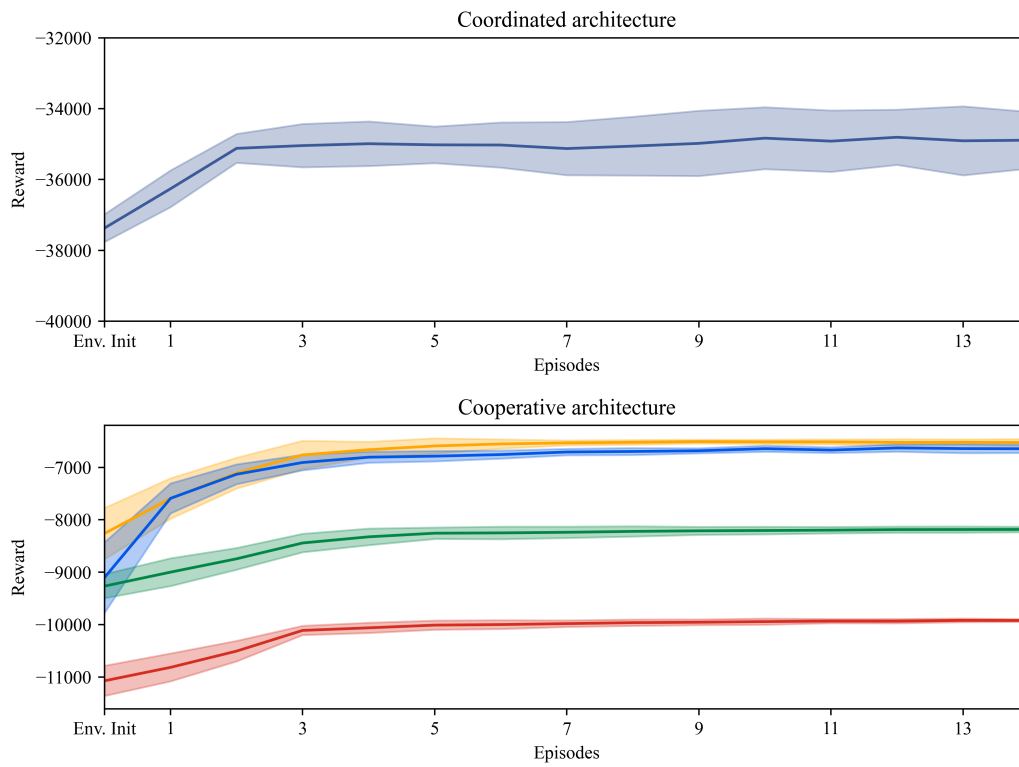


Fig. A.1 Evolution of the reward function with episodes

A.3 List of articles



This Appendix lists the papers published by the author that have been included/partially included in this dissertation.

Energy 229 (2021) 120725

Contents lists available at ScienceDirect

Energy

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

Coordinated energy management for a cluster of buildings through deep reinforcement learning

Giuseppe Pinto ^a, Marco Savino Piscitelli ^a, José Ramón Vázquez-Canteli ^b, Zoltán Nagy ^b, Alfonso Capozzoli ^{a,*}

^a Politecnico di Torino, Department of Energy, TEBE research group, BAEDA Lab, Corso Duca degli Abruzzi 24, 10129, Torino, Italy
^b Intelligent Environment Laboratory, Department of Civil, Architectural and Environmental Engineering, The University of Texas, Austin, TX, 78712, USA

ARTICLE INFO

Article history:
 Received 30 November 2020
 Received in revised form 18 March 2021
 Accepted 22 April 2021
 Available online 27 April 2021

Keywords:
 Coordinated energy management
 Deep reinforcement learning
 Building energy flexibility
 Peak demand reduction
 Grid interaction

ABSTRACT

Advanced control strategies can enable energy flexibility in buildings by enhancing on-site renewable energy exploitation and storage operation, significantly reducing both energy costs and emissions. However, when the energy management is faced shifting from a single building to a cluster of buildings, uncoordinated strategies may have negative effects on the grid reliability, causing undesirable new peaks.

To overcome these limitations, the paper explores the opportunity to enhance energy flexibility of a cluster of buildings, taking advantage from the mutual collaboration between single buildings by pursuing a coordinated approach in energy management.

This is achieved using Deep Reinforcement Learning (DRL), an adaptive model-free control algorithm, employed to manage the thermal storages of a cluster of four buildings equipped with different energy systems. The controller was designed to flatten the cluster load profile while optimizing energy consumption of each building. The coordinated energy management controller is tested and compared against a manually optimised rule-based one.

Results shows a reduction of operational costs of about 4%, together with a decrease of peak demand up to 12%. Furthermore, the control strategy allows to reduce the average daily peak and average peak-to-average ratio by 10 and 6% respectively, highlighting the benefits of a coordinated approach.

© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

The current energy transition is deeply changing the way energy is used and generated. The need of a further decarbonisation of the building sector [1], together with the rapid growth of urban areas, has fostered the use of distributed renewable energy resources. Nonetheless, the rapid penetration of renewable energy sources, characterised by their stochastic behaviour, represents the main cause of an intermittent injection of electricity into the power grid, which can jeopardize grid stability [2]. A recent solution lies in a new paradigm of energy management, which shifts from the supply-side to building demand-side control. The latter exploits the novel concept of building energy flexibility, that represents the ability of adapting energy consumption and storage operation without compromising technical and comfort constraints, to

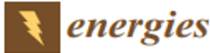

increase on-site renewable energy consumption, reduce costs and provide services to the grid (i.e. load shifting, peak shaving) [3]. Among the different strategies aimed at increasing grid stability arises demand response (DR). DR programs are designed to control power demand through different mechanisms that can be classified as i) price-based mechanisms, which aim to encourage consumption in specific periods of the day by reducing tariffs, and ii) event-based mechanisms, such as load curtailment, which are used to preserve network reliability. However, the adoption of price-based programs in some circumstances could be a double-edged sword, causing new undesirable peaks of demand during periods with low electricity prices [4].

In this framework, building energy management should leverage automated algorithms capable to adapt to a changing environment and to learn from user's behaviour and historical building-related data to optimise, coordinate and control the different actors of the smart grids (e.g., producers, service providers, consumers) [5].

* Corresponding author.
 E-mail address: alfonso.capozzoli@polito.it (A. Capozzoli).



<https://doi.org/10.1016/j.energy.2021.120725>
 0360-5442/© 2021 Elsevier Ltd. All rights reserved.

Giuseppe Pinto, Marco Savino Piscitelli, José Ramón Vázquez-Canteli, Zoltán Nagy, and Alfonso Capozzoli. Coordinated energy management for a cluster of buildings through deep reinforcement learning. *Energy*, 229:120725, 2021 [35]

Article


Exploring the Potentialities of Deep Reinforcement Learning for Incentive-Based Demand Response in a Cluster of Small Commercial Buildings

Davide Deltetto, Davide Coraci, Giuseppe Pinto, Marco Savino Piscitelli  and Alfonso Capozzoli * 

TEBE Research Group, BAEDA Lab, Department of Energy “Galileo Ferraris”, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy; davide.deltetto@polito.it (D.D.); davide.coraci@polito.it (D.C.); giuseppe.pinto@polito.it (G.P.); marco.piscitelli@polito.it (M.S.P.)
* Correspondence: alfonso.capozzoli@polito.it

Abstract: Demand Response (DR) programs represent an effective way to optimally manage building energy demand while increasing Renewable Energy Sources (RES) integration and grid reliability, helping the decarbonization of the electricity sector. To fully exploit such opportunities, buildings are required to become sources of energy flexibility, adapting their energy demand to meet specific grid requirements. However, in most cases, the energy flexibility of a single building is typically too small to be exploited in the flexibility market, highlighting the necessity to perform analysis at a multiple-building scale. This study explores the economic benefits associated with the implementation of a Reinforcement Learning (RL) control strategy for the participation in an incentive-based demand response program of a cluster of commercial buildings. To this purpose, optimized Rule-Based Control (RBC) strategies are compared with a RL controller. Moreover, a hybrid control strategy exploiting both RBC and RL is proposed. Results show that the RL algorithm outperforms the RBC in reducing the total energy cost, but it is less effective in fulfilling DR requirements. The hybrid controller achieves a reduction in energy consumption and energy costs by respectively 7% and 4% compared to a manually optimized RBC, while fulfilling DR constraints during incentive-based events.

Keywords: demand response; energy flexibility; cluster of buildings; energy management; deep reinforcement learning


 **check for updates**

Citation: Deltetto, D.; Coraci, D.; Pinto, G.; Piscitelli, M.S.; Capozzoli, A. Exploring the Potentialities of Deep Reinforcement Learning for Incentive-Based Demand Response in a Cluster of Small Commercial Buildings. *Energies* 2021, 14, 2933. <https://doi.org/10.3390/en14102933>

Academic Editor: Ricardo J. Bessa

Received: 9 April 2021
Accepted: 15 May 2021
Published: 19 May 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The path towards the decarbonization of the energy and building sector paved the way for the integration of Renewable Energy Sources (RES), seen as key actors to tackle climate change.

However, the high volatility of renewable electricity sources can jeopardize grid reliability [1]. In this scenario, system flexibility can be exploited to guarantee the stability of the electricity grid [2]. Energy flexibility can be provided by three main sources: flexible generators (e.g., cogeneration units), energy storages (e.g., batteries, thermal storages), and flexible demand (e.g., industrial or commercial buildings) [2].

However, due to the high cost of operating and maintaining flexible sources on the supply-side [3], the last few years have seen building Demand Side Flexibility (DSF) as one of the most explored and promising opportunities. In fact, buildings account for around 40% of global energy demand, thus representing a valuable opportunity for the design of advanced strategies oriented to provide demand flexibility services. According to [4], the energy flexibility of a building depends on the possibility to manage its energy demand and local generation according to climate conditions, user needs, and grid requirements. This can be achieved by implementing Demand Side Management (DSM) [5] and load control strategies, which also include Demand Response (DR) programs [6]. DR programs

Energies 2021, 14, 2933. <https://doi.org/10.3390/en14102933>

<https://www.mdpi.com/journal/energies>



Davide Deltetto, Davide Coraci, Giuseppe Pinto, Marco Savino Piscitelli, and Alfonso Capozzoli. Exploring the Potentialities of Deep Reinforcement Learning for Incentive-Based Demand Response in a Cluster of Small Commercial Buildings. *Energies*, 14(10), 2021 [38]


Applied Energy 310 (2022) 118497

Contents lists available at ScienceDirect

Applied Energy

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



Enhancing energy management in grid-interactive buildings: A comparison among cooperative and coordinated architectures

Giuseppe Pinto^a, Anjukath Kathirgamanathan^{b,c}, Eleni Mangina^{c,d}, Donal P. Finn^{b,c}, Alfonso Capozzoli^{a,*}

^a Department of Energy, TIBB Research Group, BAETA Lab, Politecnico di Torino, Italy
^b School of Mechanical and Materials Engineering, University College Dublin, Ireland
^c UCD Energy Institute, O'Brien Centre for Science, University College Dublin, Ireland
^d School of Computer Science, University College Dublin, Ireland

ARTICLE INFO

Keywords:
 Deep Reinforcement Learning (DRL)
 Building energy flexibility
 Soft Actor Critic (SAC)
 Multi Agent Reinforcement Learning (MARL)
 Grid-interactive buildings

ABSTRACT

The increasing penetration of renewable energy sources has the potential to contribute towards the decarbonisation of the building energy sector. However, this transition brings its own challenges including that of energy integration and potential grid instability issues arising due to the stochastic nature of variable renewable energy sources. One potential approach to address these issues is demand side management, which is increasingly seen as a promising solution to improve grid stability. This is achieved by exploiting demand flexibility and shifting peak demand towards periods of peak renewable energy generation. However, the energy flexibility of a single building needs to be coordinated with other buildings to be used in a flexibility market. In this context, multi-agent systems represent a promising tool for improving the energy management of buildings at the district and grid scale. The present research formulates the energy management of four buildings equipped with thermal energy storage and PV systems as a multi-agent problem. Two multi-agent reinforcement learning methods are explored: a centralised (coordinated) controller and a decentralised (cooperative) controller, which are benchmarked against a rule-based controller. The two controllers were tested for three different climates, outperforming the rule-based controller by 3% and 7% respectively for cost, and 10% and 14% respectively for peak demand. The study shows that the multi-agent cooperative approach may be more suitable for districts with heterogeneous objectives within the individual buildings.

1. Introduction

As stated in the European Green Deal, the European Commission has set net-zero carbon emission ambitions for 2050 in response to the emerging climate challenge [1]. Significant progress has been made in decarbonising the electricity sector in recent years, with solar photovoltaic (PV), onshore and offshore wind showing evidence of being promising contributors towards a fully decarbonised energy system [2]. However, renewable solar and wind energy sources are intrinsically variable by nature and this has the potential to create stability issues for the electricity grid with the fluctuating supply needing to be balanced with demand [3]. Villar et al. [4] summarise some of the challenges faced by the new power system paradigm, that is transitioning from a centralised power production, to a decentralised production, thus requiring the need for new flexibility products and markets. The flexibility to manage supply-demand mismatches can come from the supply side (through the use of dedicated standby conventional power plants or storage), through reinforcing interconnections between neighbouring countries or electrical grids [2] or from the demand side [3,5]. Analysing the latter, Demand Side Management (DSM) can be defined as a set of actions that influence the quantity, patterns of use or the primary source of energy consumed by end-users [6]. Demand Response (DR) is one promising pillar of DSM, where consumers curtail or shift their electricity usage in response to financial or other incentives [7]. As buildings represent about 40% of the total primary energy consumption in Europe [8], they are very relevant to participation in DR. A significant portion of building energy demand is towards conditioning the interior spaces for human thermal comfort through the use of Heating, Ventilation and Air Conditioning (HVAC) systems [9]. These loads can often be shifted through the use of active thermal energy storage such as water tanks, and passive thermal mass of the building [10], thus playing an expanding role in the future smart grid [11,12].

* Corresponding author.
 E-mail address: alfonso.capozzoli@polito.it (A. Capozzoli).

<https://doi.org/10.1016/j.apenergy.2021.118497>
 Received 30 September 2021; Received in revised form 3 December 2021; Accepted 28 December 2021
 Available online 28 January 2022
 0306-2619/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license
<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Giuseppe Pinto, Anjukan Kathirgamanathan, Eleni Mangina, Donal P. Finn, and Alfonso Capozzoli. Enhancing energy management in grid-interactive buildings: A comparison among cooperative and coordinated architectures. *Applied Energy*, 310:118497, 2022 [37]

Applied Energy 304 (2021) 117642

Contents lists available at ScienceDirect

Applied Energy

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






Data-driven district energy management with surrogate models and deep reinforcement learning

Giuseppe Pinto, Davide Deltetto, Alfonso Capozzoli^{*}

Politecnico di Torino, Department of Energy, TEEB research group, BAEDA Lab, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

HIGHLIGHTS

- LSTM models and DRL provide an effective data-driven district energy management.
- The proposed approach reduces computational cost compared to a forward modelling.
- The coordinated management achieves 23% of peak reduction compared to baseline RBC.
- The DRL controller is capable to optimize comfort, cost and peaks at district level.

ARTICLE INFO

Keywords:
 Coordinated energy management
 Deep reinforcement learning
 Long short-term memory neural network
 Data-driven modelling
 Building energy flexibility

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 is 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.

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

^{*} Corresponding author.
 E-mail address: alfonso.capozzoli@polito.it (A. Capozzoli).

<https://doi.org/10.1016/j.apenergy.2021.117642>
 Received 13 April 2021; Received in revised form 30 July 2021; Accepted 15 August 2021
 Available online 10 September 2021
 0306-2619/© 2021 Elsevier Ltd. All rights reserved.

Giuseppe Pinto, Davide Deltetto, and Alfonso Capozzoli. Data-driven district energy management with surrogate models and deep reinforcement learning. *Applied Energy*, 304:117642, 2021 [36]



Giuseppe Pinto, Zhe Wang, Abhishek Roy, Tianzhen Hong, and Alfonso Capozzoli. Transfer learning for smart buildings: A critical review of algorithms, applications, and future perspectives. *Advances in Applied Energy*, 5:100084, 2022 [239]



ELSEVIER

Energy & Buildings 276 (2022) 112530

Contents lists available at ScienceDirect

Energy & Buildings

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



Sharing is caring: An extensive analysis of parameter-based transfer learning for the prediction of building thermal dynamics

Giuseppe Pinto^{a,b}, Riccardo Messina^a, Han Li^b, Tianzhen Hong^b, Marco Savino Piscitelli^a, Alfonso Capozzoli^{a,*}

^aDepartment of Energy, TEBE Research Group, BAEDA Lab, Politecnico di Torino, Italy
^bBuilding Technology and Urban Systems Division, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, United States



ARTICLE INFO

Article history:
Received 16 June 2022
Revised 21 September 2022
Accepted 27 September 2022
Available online 1 October 2022

Keywords:
Transfer learning
Building thermal dynamics
Deep neural network
Data-driven models
Grid-interactive buildings

ABSTRACT

In recent years deep neural networks have been proposed as a lightweight data-driven model to capture high-dimensional, nonlinear physical processes to predict building thermal responses. However, the need of a large amount of data for the training process of deep neural networks clashes with the potential limited data availability in most existing or new buildings. Transfer learning aims to enhance the performance of a target learner exploiting knowledge from related and similar environments. This study conducted a suite of experiments that leveraged 250 data-driven models based on a synthetic dataset of a building archetype to study the influence of data availability, energy efficiency level, occupancy and climate for the transfer process of thermal dynamics. The performance of the transfer learning process was compared against a classical machine learning approach. The results suggest that building thermal dynamics can be effectively transferred under the same climatic conditions, increasing performance when dealing with different occupancy schedules, efficiency levels and low data availability. Furthermore, the paper compares the performance of both transfer learning and machine learning approaches in an online fashion, to support the implementation in real-world deployment.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

The current energy transition is deeply changing the way energy is used and generated. The need of a further decarbonisation of the building sector [1] has fostered the use of distributed renewable energy resources. In this framework, Grid-interactive Efficient Buildings (GEB) [2] are crucial in the energy transition process, exploiting advanced control strategies, identified as a way to increase energy savings up to 28% [3] while providing benefits for the electric grid [4]. However, the main bottleneck for their widespread implementation is that these control strategies often rely on predictive-based optimization [5], which requires the development of a fast and accurate building thermal dynamic model [6]. A thermal dynamic model of a building (usually built at the resolution of a thermal zone or room/space) predicts how the indoor environmental conditions (e.g., indoor air temperature and humidity) respond to the internal disturbances (e.g., heat gains from occupants, lighting and plug-in equipment), external factors (e.g., outdoor air temperature, humidity, solar irradiance), and

HVAC operations (zone temperature setpoint, supplied cooling or heating energy) both in normal and faulty conditions [7]. The thermal dynamics of a building is governed by high-dimensional, nonlinear and discontinuous dynamics, which require effort and expertise to be properly modeled [8].

In particular, three main techniques are used to model building thermal dynamics: white-box modeling, gray-box modeling and black-box modeling. White-box models use physical knowledge to describe building dynamics [6] and are based on the concept of heat transfer and energy and mass conservation. The major barrier of white-box modeling is represented by the time and effort necessary to define and collect reliable building features. The gray-box category covers a wide range of models that exploit simplified physical relationships but also require parameter estimation based on measured data. A typical concept in gray-box models consists in the resistor-capacitor analogy with electrical circuits [9], and their development is related to the robust estimation of R-C parameters. Black-box models learn the building thermal dynamics directly from the measured historical data, without assuming prior hypothesis regarding any physical relationships [10]. The main advantages of the black-box models are the lower development cost and the flexibility in using any

* Corresponding author.
E-mail address: alfonso.capozzoli@polito.it (A. Capozzoli).

<https://doi.org/10.1016/j.enbuild.2022.112530>
0378-7788/© 2022 Elsevier B.V. All rights reserved.

Giuseppe Pinto, Anjukan Kathirgamanathan, Eleni Mangina, Donal P. Finn, and Alfonso Capozzoli. Enhancing energy management in grid-interactive buildings: A comparison among cooperative and coordinated architectures. *Applied Energy*, 310:118497, 2022 [37]