Data-Driven Methodology for Model Order Reduction to Predict and Manage Building Energy Flexibility in Smart Grids

Anthony Maturo

A Thesis

In the

Department of

Building, Civil and Environmental Engineering

Presented in Partial Fulfillment of the Requirements
For the Degree of

Doctor of Philosophy (Building Engineering) at

Concordia University

Montréal, Québec, Canada

March 2025

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: *Anthony Maturo*

Entitled: *Data-Driven Methodology for Model Order Reduction to Predict and Manage Building Energy Flexibility in Smart Grids*

and submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY (Building Engineering)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

______________________________ Chair
Dr. Rabin Raut

______________________________ External Examiner
Dr. Per Kvols Heiselberg

______________________________ Examiner
Dr. Ursula Eicker

______________________________ Examiner
Dr. Liangzhu (Leon) Wang

______________________________ Examiner
Dr. Bruno Lee

______________________________ Examiner
Dr. Adolfo Palombo

______________________________ Thesis Supervisor
Dr. Andreas K Athienitis

______________________________ Thesis Supervisor
Dr. Annamaria Buonomano

Approved by ______________________________
Graduate Program Director

March 17, 2025 ______________________________
Dean of the Gina Cody School of Engineering and Computer Science

ABSTRACT

Data-Driven Methodology for Model Order Reduction to Predict and Manage Building Energy Flexibility in Smart Grids

Anthony Maturo, Ph.D.
Concordia University, 2025

The evolving energy landscape, driven by rising demand, electrification, and renewable energy integration, necessitates a shift from traditional "follow-the-load" model to demand-side management. This transition requires accurate prediction of building energy demand, effective demand response participation, and quantification of energy flexibility.

This thesis develops a methodology for predicting and optimizing building thermal energy demand using data from smart thermostats and monitoring infrastructures. Multi-zone buildings and schedule-based operations are modelled using resistance-capacitance (RC) thermal networks. An automated model order reduction approach identifies dominant thermal zones in multi-zone buildings, while control-oriented RC archetypes capture key dynamics in schedule-based operations. Calibration follows a Model Predictive Control Relevant Identification (MRI) process, ensuring models accurately predict thermal dynamics up to 24 hours ahead.

Weather variability is managed through clustering techniques that identify representative days, reducing computational complexity while enabling scenario-driven analysis. This approach bridges the gap between operational and design studies by integrating energy flexibility considerations early in building and community planning.

A distributed economic Model Predictive Control (e-MPC) framework optimizes thermal load management while maintaining occupant comfort and system constraints. It supports applications at both single-building and community scales, such as virtual power plants. Performance is assessed using energy flexibility Key Performance Indicators (efKPIs) against a reference scenario.

The methodology is validated through three case studies: (1) Residential buildings: 30 detached homes equipped with smart thermostats (data from Hydro-Québec); (2) Institutional building: The Varennes Net-Zero Energy Library, Canada's first net-zero energy institutional building; (3) Community-scale system: A simulated hybrid photovoltaic-battery microgrid in Varennes serving residential and institutional buildings.

Findings highlight how varying building participation in demand response influences aggregated demand profiles, utility metrics (load shifting, peak shaving), and the sizing of grid-supportive technologies. At the single-building level, insights are provided for optimizing thermal load management across convective, radiant, and mixed heating systems. By integrating data-driven modelling, advanced control, and scalable design, this thesis provides actionable solutions for energy efficiency, flexibility, and resilience, supporting a sustainable energy transition.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisors, Prof. Andreas Athienitis and Prof. Annamaria Buonomano, for their unwavering support and guidance throughout my PhD journey. Their encouragement to participate in conferences and travel opportunities has significantly enriched my academic and personal growth.

I am sincerely grateful for the financial support that made this research possible. I would like to thank the people from Hydro-Québec, Hilo, Camnet, and Regulvar for their invaluable contributions and collaboration. It is truly encouraging and inspiring to see such important companies actively supporting research and innovation in this field, reaffirming the value of the work we do and its potential impact.

To my office colleagues, thank you for your support and for the good times we shared. The camaraderie and encouragement from all of you made this experience much more enjoyable.

To my family and friends, despite the physical distance, your unwavering love and support have always kept me close. No matter how far I am, you make it feel like I never left. Your presence in my life is a source of strength and comfort.

To my girlfriend, thank you for standing by me through difficult times, for your unwavering support, and for making me feel like home.

Thank you all for being part of this significant chapter of my life.

“Memento audere semper”

G. D’Annunzio

Table of Contents

List of figures

List of tables

Nomenclature

Variables

A	State matrix
B	Input to state matrix
ch	Control horizon [h]
C	State to output matrix
d	Disturbance variables
D	Feedthrough matrix
DBI	Daily Boolean index
F	Filter
G	Transfer function
n	Number of cycles
ph	Prediction horizon [h]
p	Number of poles in a transfer function
P	Electricity input [W]
q	Shift operator
Q	Heating input [W]
s	Laplace transfer variable
S	Solar radiation [W/m^2]
SoC	State of charge [-]
t	Time step or sampling time [s]
T	Temperature [°C]
u	Input variable
w	Frequency [s^{-1}]
y	Output variable
z	Number of zeros in transfer functions

Abbreviations

ADR	Automated demand response
AI	Artificial intelligence
AHU	Air handling unit
ANN	Artificial neural network
ASHP	Air-source heat pump
ARMA	Autoregressive moving average
ARIMA	Autoregressive integrated moving average
BAS	Building automation system
BEFI	Building energy flexibility index
BIPV	Building integrated photo-voltaic
BIPV/T	Building integrated photo-voltaic thermal
CVI	Cluster validity index
CPP	Critical peak pricing

DR	Demand response
DSM	Demand side management
DSO	Distribution system operator
DTW	Dynamic time warping
efKPI	Energy flexibility key performance indicator
eMPC	Economic model predictive control
EDP	Extreme day pricing
FF	Flexibility factor
FIT	Fitness function
FD	Frequency domain
GHG	Greenhouse gas
HX	Heat exchanger
HVAC	Heating ventilation and air conditioning
KPI	Key performance indicator
IEA	International energy agency
LF	Load factor
LLP	Loss of load probability
M	Available samples/measurements
ML	Machine learning
MM	Moment matching
MOR	Model order reduction
MPC	Model predictive control
MRI	MPC relevant identification
MI	Mixed integer
ML	Machine learning
NRMSE	Normalized Root Mean Square Error
NZEB	Net zero energy buildings
NZEC	Net zero energy community
PCA	Principal component analysis
PCM	Phase change material
PCM HX	Phase-change material heat exchanger
PSO	Particle swarm optimization
PV	Photovoltaic
R	Thermal resistance
RR	Radiative ratio
RC	Resistance capacitance
RMSE	Root mean square error
ROM	Reduced-order model
RTP	Real time pricing
STLF	Short term load forecasting
SVD	Singular value decomposition
SVM	Support vector machine
SVR	Support vector regression

TABS	Thermally activated building systems
TEC	Transactive energy control
TES	Thermal energy storage
TOU	Time of Use
TSO	Transmission system operator

Subscripts

e	Envelope mass
f	Filtered
flex	Flexible
hor	Horizon
i	Interior mass
lb	Lower bound
o	Outdoor
opt	Optimal
p	Peak
ref	Reference
th	Thermal
ub	Upper bound

Greek symbols

δ	Threshold
Δ	Variation
ε	Cost function penalty
γ	Slack variable
λ	Thermal conductivity [W/mK]
η	Efficiency
ρ	Differential operator
τ	Time constant

Chapter: 1 Introduction

1.1 Problem statement

The increasing energy demand, integration of renewable energy sources and electrification poses significant challenges, including grid stability, efficient energy use, and cost-effective operation.

Buildings, as major energy consumers, are key players in this transformation. The building sector accounts for approximately 40% of global energy and produces 36% of CO_2 emissions in developed countries [1]. This high consumption stems from heating, cooling, lighting, and powering electronic devices in residential, institutional, and commercial buildings. With ongoing urbanization and economic growth, energy demand of buildings is expected to increase, further straining energy supply systems [2].

The rising energy demand in the building sector challenges the electrical grid, especially during peak consumption periods [3]. As shown in Figure 1.1, in cold Quebec winters, heating demand peaks stress the grid ultimately increasing the risk of blackouts and higher operational costs. Similarly, hot California summers increase air conditioning use, potentially overloading the grid and causing rolling blackouts.

Transitioning to low-carbon energy systems adds complexity, as traditional energy production and distribution are ill-equipped for higher demand and the integration of renewable energy sources [4, 5]. Addressing these challenges requires major investments in grid modernization and expansion.

While integrating renewables like wind and solar is essential for low-carbon energy production, their variability poses stability challenges [6, 7]. As shown in Figure *1.2* and Figure 1.3, California's ambitious renewable energy targets have boosted solar power adoption. However, the variability in solar power generation and periods of low demand have led to overgeneration and required curtailment of renewable production. Other countries face similar challenges:

- **Germany**: The 'Energiewende' initiative has increased renewables but caused grid stability issues. This mismatch between peak renewable generation and consumption necessitates advanced grid management and greater storage capacity. The Renewable Energy Sources Act (EEG) has driven this shift with incentives and feed-in tariffs.
- **Netherlands**: Heavy investment in wind power aims to cut CO_2 emissions by 49% by 2030. Large-scale wind farms have created grid challenges during high wind generation and low consumption. Solutions include grid expansion, demand response, and increased interconnections with neighbouring countries to manage these fluctuations.
- **Denmark**: Reliance on wind power sometimes results in excess electricity. The Danish Energy Agreement targets 55% renewables by 2030, with surplus energy exported to neighbouring countries. Managing exports and maintaining grid stability are ongoing

challenges, addressed through smart grid innovations and increased storage.

- **Australia**: Rapid growth in rooftop solar photovoltaics, driven by the Renewable Energy Target (RET), has led to grid stability issues, especially during sunny periods with high solar output and low demand. Strategies being explored include virtual power plants and improved grid management by the Australian Energy Market Operator (AEMO).

Poor renewable energy management and peak consumption periods challenge global energy systems by creating steep ramps [8, 9] and increasing reliance on fossil fuel-based power plants [10], leading to higher greenhouse gas emissions and undermining the environmental benefits of renewable energy. Therefore, addressing these issues requires a multifaceted approach that integrates energy storage [11, 12], improved grid infrastructure [13, 14], and demand-side management (DSM) [15-17], alongside a comprehensive understanding of consumption patterns [18-20]. Additionally, strategies to reduce peak demand [21] must be complemented with predictive analytics and advanced control mechanisms to optimize energy flows, enhance system flexibility, and support the seamless integration of renewables into the grid. Finally, coordinated efforts in technology, policy, and user engagement are essential to achieving a sustainable energy transition.

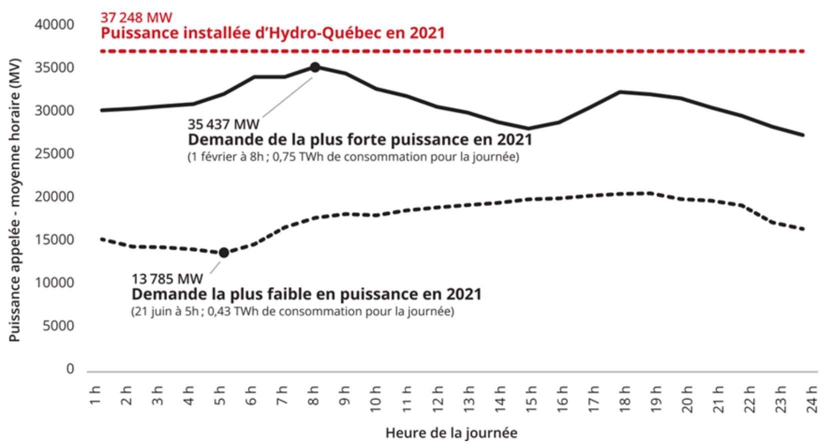


Figure 1.1: Québec daily energy consumption during winter and summer.

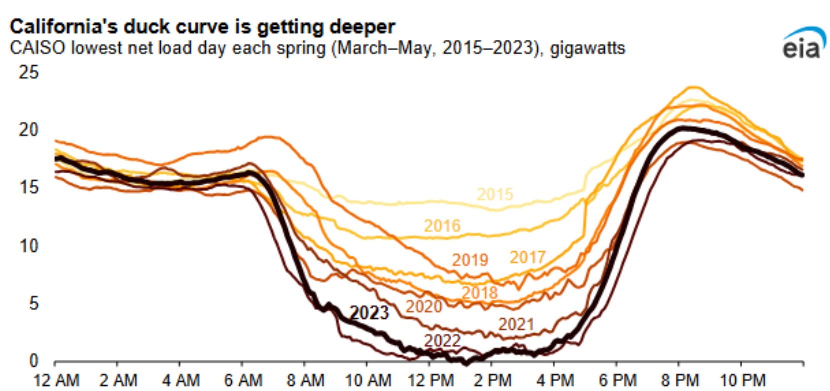


Figure 1.2: California's duck curve (during spring from March-May, 2015-2023).

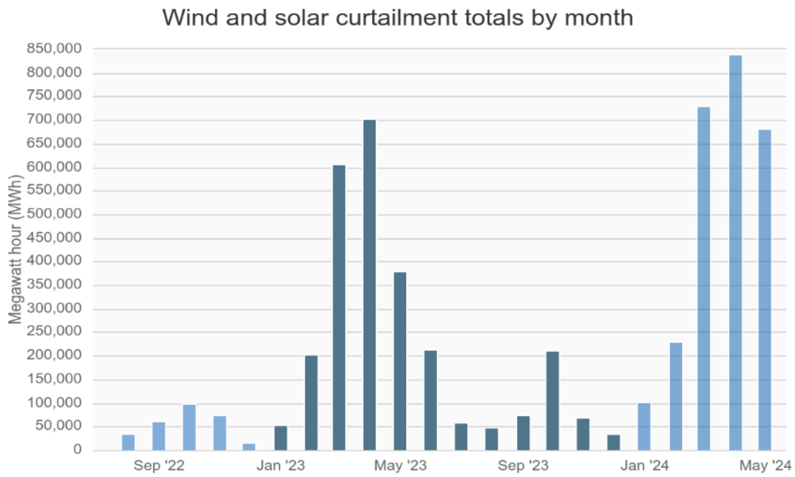


Figure 1.3: Renewable curtailment by month (from September 2022 to May 2024).

In this context, *energy flexibility* in buildings is crucial for enhancing grid stability and integrating renewable energy resources [22, 23]. It involves adjusting consumption patterns based on external signals like grid demand or price changes [24]. The energy flexibility is linked to the *smart building* concept and it is supported by International Energy Agency (IEA) research through Annexes 67, 81, and 82 on flexibility definition, demand response, and integrated building energy solutions [25-27].

Flexibility can be achieved with advanced control strategies, such as Model Predictive Control (MPC) [28], by exploiting building thermal mass [29], or by optimally controlling Heating Ventilation and Air Conditioning (HVAC) [30, 31], electric vehicles [32, 33], and energy storage systems [34, 35]. Defining and quantifying energy flexibility is a complex task due to the interactions between building systems, occupant behaviour, and external factors. It requires a comprehensive understanding of the dynamic interactions between building energy systems and the electrical grid through advanced modelling, predictive analytics, and the development of key performance indicators (KPIs) to assess the performance and impact of flexibility measures [36].

Energy flexibility must be addressed at various scales, ranging from individual buildings to entire *communities*, with tailored approaches for each [37, 38]. In this framework, the interconnection between design and control is crucial, as building technologies should be sized and selected based on their intended use and control strategies [39-41]. This highlights the need for integrated planning and optimization, where design and control are interdependent for effective energy flexibility. In this regard, *guidelines* for exploiting energy flexibility are essential, encompassing concepts, applications, and methodologies to ensure effective implementation [42].

1.2 Objectives

The building sector presents significant challenges and opportunities for improving energy efficiency, integrating renewable energy, and transitioning to a more flexible and resilient smart grid. By leveraging advanced technologies, innovative control strategies, and supportive policies, it is possible to enhance energy flexibility and contribute to a low-carbon future. The continued evolution of the energy system requires a comprehensive understanding of the interactions between production, distribution, and consumption, as well as the engagement of all stakeholders in this transformative process. This thesis addresses these challenges by developing advanced methodologies and strategies to support data-driven energy modelling, optimize thermal load management, connect optimal control with design, and implement key performance indicators for evaluating energy flexibility measures. Through these contributions, the research aims to provide actionable insights for enhancing the sustainability and resilience of building energy systems.

Objective 1: Automated Model Order Reduction for Thermal Load Prediction

The first objective focuses on developing automated energy modelling approaches that leverage data from smart thermostats to predict building energy demand. The methodology emphasizes model order reduction to balance accuracy and computational efficiency, enabling real-time predictive control. These models simulate the interactions between building systems and the grid, considering variables such as occupancy, weather, and energy pricing. By providing accurate and scalable predictive tools, this objective lays the foundation for informed decision-making in building energy management.

Why it is important: Automated, data-driven approaches ensure scalability and adaptability, supporting wide-scale adoption in diverse building types. The ability to customize model accuracy according to application needs makes this approach highly practical for a range of use cases, from model predictive control to operational optimization to strategic planning. This methodology is critical for improving energy management systems and enabling grid-interactive buildings.

Objective 2: Control-Oriented Model Archetypes for Buildings Operating on a Schedule

This objective involves the development of control-oriented model archetypes to study and optimize building operations based on predefined schedules. These archetypes will provide a framework for evaluating the impact of operational changes on energy consumption and flexibility. Similar to the previous methodology also this one aligns with global initiatives, such as IEA EBC Annex 81: Data-Driven Smart Buildings, which emphasize the importance of leveraging high-quality data and real-time analytics to optimize building energy use.

Why it is important: Control-oriented models allow energy systems to respond dynamically to changing grid conditions and user needs. These archetypes are instrumental for integrating flexibility into daily operations, minimizing energy waste, and reducing peak demand.

Objective 3: Optimization of Energy Flexibility through Thermal Load Management

A key objective of this thesis is to optimize energy flexibility by using electricity price responsiveness and thermal load management in buildings equipped with full-convective, full-radiant, or mixed convective-radiant heating. This involves exploiting building thermal mass to shift and reduce energy demand during peak periods. Proposed strategies will include price-based demand response mechanisms and load-shifting techniques, enhancing both grid stability and economic performance. KPIs will be introduced to monitor the effectiveness of these strategies and identify areas for refinement.

Why it is important: Effective thermal load management reduces the need and activation of fossil fuel-based peaking power plants, lowers operational costs, and minimizes the environmental impact of buildings. The introduction of KPIs ensures measurable and transparent evaluation of flexibility measures, fostering continuous improvement.

Objective 4: Design and Control of Energy Flexibility at Community Scale

The fourth objective is to develop an integrated approach for the design and optimal control of grid-supportive technologies within microgrids, enhancing energy flexibility at various scales. By examining the dynamic interactions between energy demand and supply, this research evaluates how optimal management strategies influence the sizing of the involved systems, ensuring efficient and resilient operation. The methodology, applied in this thesis to specific case studies, is generalizable and can accommodate a variety of technologies beyond the initial applications.

Why it is important: Microgrids play a critical role in modern energy systems by enabling distributed energy generation, storage, and management. Understanding the interplay between operational strategies and system design ensures that the adopted technologies are not only cost-effective but also capable of providing energy flexibility and supporting grid stability. This objective is essential for advancing the planning and implementation of microgrids as a cornerstone of sustainable energy systems.

By addressing these four main objectives, the thesis aims to provide actionable insights and innovative solutions for transforming the building sector into a key enabler of energy system decarbonization. The integration of automated modelling, optimal control, and scalable design strategies supports the development of grid-interactive buildings that align with the demands of a sustainable and resilient energy future.

1.3 Thesis structure

The thesis is structured in chapters. The chapters are published or under review in well-established journals, however the chapters herein include more details and supplemental information.

Chapter: 1 Introduction. The chapter introduces the problem and motivation behind the thesis. It shows the current state of the grid, how renewable energy sources and electrification are affecting it and the role of buildings and building control.

Chapter: 2 Literature Review. This chapter provides a comprehensive review of the existing literature relevant to building energy modelling, energy flexibility, and the integration of grid-supportive technologies. The chapter begins by introducing the concept of building energy flexibility, highlighting its significance in transitioning from conventional energy demand patterns to grid-interactive profiles. Next, the chapter delves into the available approaches for building energy modelling, emphasizing the critical role of automated modelling and model order reduction techniques. These approaches are explored in the context of their ability to balance modelling accuracy and computational efficiency, making the developed models suitable for real-time applications and predictive control strategies. The chapter further reviews the integration of thermal and electrical technologies, addressing advances in design strategies and operational frameworks. Special attention is given to methodologies that connect design and control strategies. Finally, the chapter reviews the development and application of KPIs for assessing energy flexibility measures. It evaluates the role of these indicators in quantifying performance, identifying trade-offs, and guiding the implementation of effective management strategies.

Chapter: 3 Automated Model Order Reduction for Building Thermal Load Prediction using Smart Thermostats Data. This chapter presents a methodology for automating the development of grey-box models for model predictive control, energy efficiency, and energy flexibility applications in buildings. The proposed approach combines model order reduction and system identification techniques, featuring enhanced data pre-processing, multistage order reduction, and parameter estimation. By employing a cascade methodology, the structure of thermal models is optimized through frequency domain analysis to aggregate adjacent thermal zones with similar set point temperatures. The developed models are then calibrated with smart thermostat data to predict the indoor air temperature up to 24 hours ahead. This chapter has been published as a manuscript[1].

Chapter: 4 Optimizing Energy Flexibility through Electricity Price-Responsiveness and Thermal Load Management in Buildings with Convective and Radiant Heating. This chapter presents a data-driven, control-oriented methodology using Resistance-Capacitance model archetypes to forecast and optimize building thermal loads in buildings operating on a schedule. It integrates weather forecasts, dynamic tariffs, and model predictive control to coordinate convective and radiant heating systems and to optimize building energy flexibility. A case study on the Varennes Library, a Net Zero Energy institutional building in Québec, demonstrates that the proposed approach effectively shifts thermal loads during peak price periods, enhancing grid interaction. This chapter has been published as a manuscript[2].

Chapter: 5 Clustering-driven Design and Predictive Control of Hybrid PV-Battery Storage Systems for Demand Response in Energy Communities. This chapter introduces a methodology developed to select representative periods for analysis and evaluate the influence of controllable building loads on the design and operation of grid-supportive technologies. Using clustering techniques, characteristic periods are identified, and a distributed Model Predictive Control manages individual building thermal loads during demand response events while a supervisory MPC coordinates technology operations to meet flexibility targets. Applied to a virtual community in Varennes, Québec, the approach achieves over 40% peak demand reduction and reduces grid-supportive system sizes by up to 26%, demonstrating the benefits of integrated thermal load management for energy flexibility and system design optimization.

Chapter: 6 Conclusions and Future Works. The chapter concludes the thesis and discusses the potential for future work. The research contributions are summarized, along with a list of all published journal articles, presented conference papers, and other non-refereed work.

---

[1] Maturo, Anthony, et al. «Automated model order reduction for building thermal load prediction using smart thermostats data». Journal of Building Engineering, vol. 96, November 2024, p. 110492. ScienceDirect, https://doi.org/10.1016/j.jobe.2024.110492.

[2] Maturo, Anthony, et al. «Optimizing energy flexibility through electricity price-responsiveness and thermal load management in buildings with convective and radiant heating systems». Energy and Buildings, January 2025, p. 115355. ScienceDirect, https://doi.org/10.1016/j.enbuild.2025.115355.

Chapter: 2 Literature Review

This chapter presents an overview of the current research on the topics explored in this thesis is presented in this chapter. The workflow begins with the description of the building energy flexibility concept, then moves to the importance of building energy modelling and advances to the integration of technologies, ultimately aiming for energy management, interconnection with design and development of key performance indicators. The chapter is divided into eight sections: Section 2.1, explores the concept of building energy flexibility, Section 2.2 explores the main approaches in building energy modelling, Section 2.3 focus on advanced technologies distinguishing thermal and electrical systems. Section 2.4 deals with the energy management strategies, Section 2.5 describes the interconnection between design and control, Section 2.6 deals with the key performance indicators and Section 2.7 defines the research question of the thesis.

2.1 Building energy flexibility

The progression towards a grid-interactive smart building involves several stages. As depicted in Figure 2.1, these are: improving energy efficiency to reduce energy consumption, becoming a prosumer to self-balance energy needs, and accessing smart building-to-grid services. Smart buildings, synonymous with energy-flexible buildings, are characterized by four main features: climate response, grid response, user response, and monitoring and supervision [43]. These buildings must effectively respond to external climate conditions, grid signals, real-time user interactions with implemented technologies, and perform continuous monitoring of building operations.

Energy flexibility in buildings has gained growing research momentum in recent years. Numerous national and international collaborations and initiatives, including the IEA EBC Annexes [25-27, 44], and the GEB initiative by the U.S. DOE [45], have been trying to bring building energy flexibility to the next level of maturity. The IEA EBC Annex 67 developed a common definition of building energy flexibility as "*the ability to manage its demand and generation according to local climate conditions, user needs, and energy network requirements without jeopardizing the technical capabilities of the building systems and the comfort of occupants. Energy Flexibility of buildings will thus allow for DSM/load control and thereby DR based on the requirements of the surrounding energy networks*" [25]. Buildings can provide grid services via flexible operations (e.g., adjusting its demand and behind-the-meter power generation and storage) [46].

These capabilities are integral to the concepts of demand side management (DSM) and demand response (DR), which are becoming increasingly important in addressing the challenges of rising energy demand, renewable power generation, electrification penetration, and global warming to maintain a safe and cost-effective power system [47].

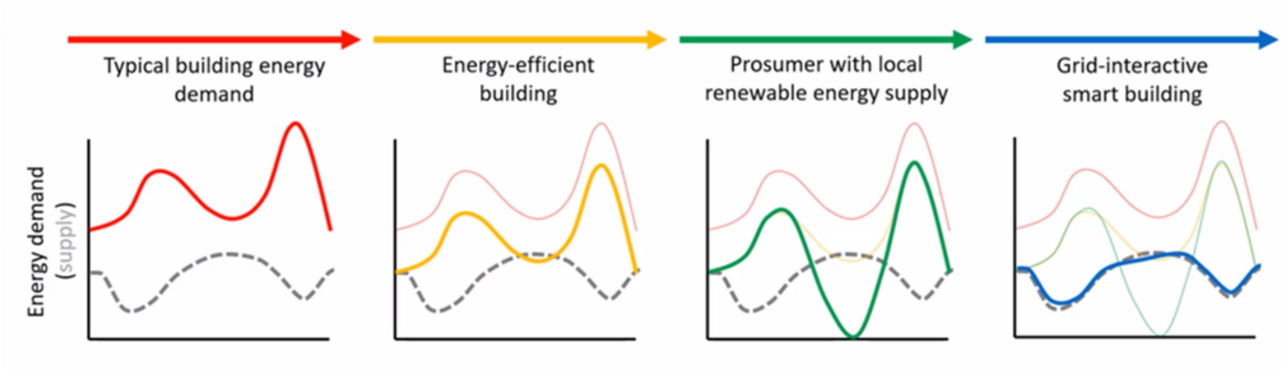


Figure 2.1: Type of energy demand according to the building "smartness" (Image from [48]).

There are two main categories of DR: *direct control* and *indirect control*. Direct control strategies regulate end-user systems through a two-way communication link, directly instructing devices on power usage based on technical limits or service levels. Indirect control strategies influence these systems via incentive or penalty signals broadcasted by aggregators or grid operators, to which local controllers respond [49]. These signals can include energy spot prices, energy price forecasts, or CO_2 intensity levels. Buildings and end-users must integrate these signals with weather forecasts, occupancy predictions, and energy demand modulation estimates to adjust their operations, aiming to minimize total energy costs or CO_2 emissions over a short period [50, 51]. A direct control service is the Automated demand response (ADR) event [52]. This allows electric devices to be turned off during periods of high demand via an internal control signal from the building control system or via an external control signal from the grid.

To effectively manage flexible load modulations through direct control, seamless two-way communication is required among grid-interactive smart buildings (which use, store, or supply energy), smart energy grids (local distribution and transmission networks for electricity, gas, and district heating/cooling), and building occupants, owners, and managers. This interaction is facilitated by smart home automation systems and smartphone apps. For indirect control demand response, broadcasting an incentive signal, such as a price signal, is sufficient. For example, in the Smart Energy Operating System setup [53], the price signal comprises composite elements derived from solving specific grid challenges.

Flexibility services can be provided at various levels of aggregation: device, zone, building, and building cluster. Demand flexibility helps achieve local objectives, such as enhancing energy efficiency, addressing capacity constraints, or increasing local renewable energy self-supply. At the distribution grid level, it helps manage congestion and voltage issues exacerbated by intermittent renewables [54]. On a broader scale, flexibility supports national goals, like reducing peak power demand to prevent blackouts and avoid curtailments due to insufficient peak supply.

Building-to-grid services depend on smart buildings that can both provide monitoring data and interpret smart grid signals for energy profile adjustments using smart home technologies or building automation systems (BAS). To participate in those services, continuous assessment of efficiency and effectiveness is necessary, requiring data-driven methods that generate energy demand baselines, optimal load profiles and KPIs for energy flexibility quantification [55]. In this framework, the absence of standard methodologies to study the effectiveness of demand response actions hinders the decisions to perform investments [48].

2.2 Building energy modelling

Technological advancements in building automation and smart monitoring are facilitating the development of design and operation guidelines for high performance buildings. These strategies are supported by appropriate models that predict the load dynamics of buildings and their associated energy systems [56]. Energy modelling plays a crucial role in studying the effect of various variables on building energy performance, providing valuable insights when establishing new energy efficiency standards [57-59].

2.2.1 Building thermal modelling

Building thermal energy models are used to predict the thermal dynamics of a building. At the design stage, these models can be used to simulate the behavior of the building for different design parameters and observe their effect on energy demand, thermal comfort and even control strategies. Depending on the objective and the problem at hand, the whole building may be modelled, or just the section of interest. Thermal models can vary in complexity, accuracy and computational demand. According to the ASHRAE Handbook of Fundamentals, there are two distinct approaches to modelling: forward (classical) and data-driven (inverse). The models created with the forwards approach, also called white-box models, use detailed physics-based mathematical equations to model the building and its components. On the other hand, data-driven models rely completely on available data.

From the literature it is possible to classify models into three broad categories: white-box, black-box, and grey-box models [60]. White-box models, often referred as physics-based or forward models, are constructed from sets of differential equations designed to capture the energy flows within the building. They are funded upon first principles modelling of the physics of the processes involved. Such models are used in detailed building energy simulation software, available commercially (such as EnergyPlus and TRNSYS), as well as in in-house tools increasingly developed by researchers and scientists for tailored analyses [61, 62]. They are fundamentally based on conservation of mass, energy, and momentum. The white-box models enable good prediction accuracy of outputs over a wide range of operating conditions and are ideal for design applications, for capturing the transient physical phenomena, also within building elements, and for assessing comfort analyses [63]. On the other hand, the significant input data requirements and the high computational burden make white-box models not suitable for model-based control applications [64].

Black-box models are purely data-driven, they employ linear or nonlinear mathematical regressors to approximate system behavior under specific or standard operating conditions; they provide input-output relationships, without offering physical interpretations. This characteristic reduces engineering cost and demands less domain knowledge. Their development typically requires sufficient clean data and machine learning algorithms. However, they have drawbacks, including a high demand for data quality. Missing, wrong, or biased data often lead to low quality models [65]. Additionally, black-box models lack interpretability and may require intense computation, especially with deep learning–based algorithms. One form of black-box models is single-variable or multi-variable regression analysis which is performed between measured output variables (e.g., energy consumption or temperature) and parameters of occupancy and weather forecast. These regression models may take the form of purely statistical approaches or be loosely based on some basic engineering formulation of energy use in buildings [66]. In several papers available in literature, the parameters of these models are estimated using neural networks [67], support vector machines [68] and other techniques.

The transfer function method, recommended by American Society of Heating, Refrigerating and Air-conditioning Engineers (ASHRAE), is one of the most useful techniques available to solve heat transfer problems in building envelopes and environments [69]. Energy simulation software, like EnergyPlus, rely on the finite difference method and transfer function method. This last approach guarantees computational time advantage compared to detailed numerical methods. By means of frequency domain modelling techniques important building transfer functions relating inputs to outputs can be obtained and studied without the need for simulation. Also, it has been shown to be a useful approach to support building simulation and develop simplified physical building models from measured data [70, 71]. Candanedo and Athienitis [72] developed simplified transfer function models using system identification techniques available in MatLab to study the predictive control of solar homes with passive and active thermal storage. Their methodology provided reasonably accurate models while offering advantages of simplicity and fast computation. In another paper, Candanedo et al. [73] studied the synchronized control of radiant floor heating and the fenestration system shading using a low order transfer function model. Chen et al. [74] presented a charge control strategy using frequency domain models and room air temperature set-point profile as input. Frequency domain models are used to predict the required charging rates of a thermal energy storage based on the desired room air temperature set-point profile and the corresponding space conditioning load. Zhuang et al. [75] proposed a new simplified modelling method, relying on transfer functions, for model predictive control application in a medium-sized commercial building. Athienitis [76, 77] showed how studying the transfer functions gives substantial insight into the building thermal behavior and supports the creation of reduced order models. Other studies show how frequency domain techniques have the ability to detect patterns [78], select features or states of a model, and classify building load profiles [79].

Finally, grey-box models represent a good trade-off between white-box and black-box models, overcoming the limits of black-box models that lack of knowledge about physical structure of the modelled system. Grey-box models aim to capture the dominant building physics that are essential for optimal operating algorithm development and implementation [22]. This approach begins by creating a physical model that represents the structure or layout of the building. Then, through statistical analysis, it identifies important parameters that capture key aggregated physical aspects and other relevant characteristics. When data are available, simplified mathematical energy building models are, thus, derived via reverse engineering methods [80]. However, this requires a high level of user expertise both in setting up the appropriate modelling equations and in estimating these parameters. There are three major benefits of grey-box models: (1) faster space conditioning load calculation with major physical dynamics, which is especially helpful for controls and building-grid integration; (2) possibility to include some time-varying and nonlinear parameters, and (3) online predictive controls.

To develop efficient and accurate semi-physical or grey-box models suitable for predictive control design and implementation, as well as for energy performance simulation tools at individual building level and on a larger scale, lumped-parameter models are more and more used. The lumped-parameter model, which is equivalent to the Resistance Capacitance (RC) thermal network representation, belongs to the white and grey-box categories. It enables achieving comparable accuracy to advanced building energy simulation tools that rely on physical models, while keeping the mathematical complexity within the model structure minimal. To accomplish this, the methodology entails both simplifying the physical aspects (breaking down the larger dynamic system into several smaller and simpler sub-systems) and utilizing model order reduction techniques. Specifically, this model consists of discretizing the temperature field of a thermodynamic system, by identifying a certain number of representative nodes where an energy balance is computed [81]. Each node is connected to the adjacent nodes by means of thermal resistances, and thermal capacities are assigned to all elements that can store internal energy, such as walls and relevant volumes of air. The simplicity of this method is associated with the number of nodes, or state variables, that defines the order of the model.

2.2.2 Model order reduction

Model order reduction (MOR) techniques have been introduced to find the optimal number of states to describe a certain dynamic. The purpose of MOR is to derive a simplified model of a detailed and complex system, reducing computational effort while preserving the key dynamic characteristics [82]. MOR techniques can be grouped in multiple categories:

- Polynomial reduction methods: generally based on frequency domain theory and some matrix operations.
- Matrix or state-space transformation-based techniques: based on the evaluation of the stability of different state coordinate selections in the high order model and on the selection of a reduced order model maintaining the original model properties such as time response, observability, controllability, closed-loop performance.
- Network-theory reduction approaches: particularly involving the creation of Norton/Thevenin equivalent networks and the utilization of the principle of superposition.
- Optimization or parameter estimation approaches: based on sequential parametric optimization procedures which aim at the optimization of some indices, generally corresponding to the accuracy and detail of the model. The optimization approach can be based on time or frequency domain and uses parameter estimation to build the reduced-order models. During the optimization routine, the best set of state variables and parameters is selected to better capture the system dynamic.

Polynomial reduction methods are divided in Singular Value Decomposition (SVD) and Moment Matching (MM) methods. The latter always result in higher error norms than the SVD-based methods, but they significantly reduce computational costs and storage requirements. Antoulas et al. [83] compared several polynomial reduction methods to six different dynamical systems. The results showed that balanced reduction and approximate balanced reduction, two SVD methods, provide models with good accuracy over the whole frequency range. Kim et al. [84] used the Krylov subspace method, a moment matching-based method, to address the issues of computational and data storage requirements when applying MOR techniques to large buildings. They compared the reduced-order model with a high-fidelity model developed in TRNSYS for a 60-zone case study, showing relative errors of the annual heating and cooling load under 5 %. Siddhart et al. [85] used a balanced truncation method to determine the minimum model order for investigating the thermal response in multi-zone buildings. The authors acknowledged that the resulting model has a high order and suggested that new techniques should focus on lumping thermal zones that exhibit strong interaction into a smaller number of “super zones”.

Network-theory reduction approaches are based on the circuit theory and can be applied not only to electrical circuits but also to thermal RC networks. According to these principles, a complex RC network can be simplified to a single distributed admittance, heat source and thermal resistance, accurately representing its thermal behavior [71, 86, 87].

In the literature, papers discussing the utilization of RC thermal models often overlook explanations regarding the selection of network structures for envelope models, internal mass, and zone models. There is a need for further research to develop a generic methodology that can be adapted to various conditions and types of buildings [88]. The selection of these models should be guided by an optimization algorithm and specific indices, or cost functions, derived from the optimization process. These indices serve as criteria for determining the most suitable model order, ensuring the accuracy and efficiency of the representation of RC thermal networks. This will finally provide a generalizable methodology applicable to a large range of buildings.

Optimization or parameter estimation approaches have the potential to guide towards the selection of the most representative RC thermal network for a specific building. During the optimization routine, optimization approaches consider the option of state variables aggregation or capacitances aggregation in case of lumped parameter models. Deng et al. [89] suggested an aggregation-based approach to streamline thermal models of multi-zone buildings. The proposed methodology relies on an optimization routine that utilizes the aggregation of Markov chains. The objective of their paper is to reduce the complexity of a RC network by creating a scalable representation with a reduced number of states. As stated by the authors, the choice of the aggregation-based methodology is supported by the prevalence of zone-based models in the HVAC community [90] which refer to the so called zoning approach [91]. Banihashemi et al. [92] introduced a novel approach to reduce the order of building energy simulation models. The approach utilizes a deep learning-based unsupervised convolutional neural network autoencoder. The method decomposes complex time series data derived from detailed simulations into lower dimension features. The low-dimension representations can be grouped through clustering algorithms to build a reduced-order model (ROM), exploiting the model-cluster-reduce pipeline [93]. Shin et al. [94] developed a procedure for automating the aggregation of thermal zones based on a grid-to-cluster method. To cluster different building sections in multiple thermal zones, they adopted a two-step process that first matches the annual energy use of each building section and then applies a linear correlation coefficient of the simulated 24-hrs indoor temperature profiles during peak days. This will serve as a final indicator for clustering. Vallianos et al. [95] developed a methodology to create RC thermal network models of residential buildings with data obtained from smart thermostats. The methodology is based on an optimization routine which starts with a very simple model and iteratively adds one parameter at a time. This parameter corresponds to the one that increases the quality of the model the most, as measured with the Bayesian Information Criterion. When the quality of the model cannot be improved any further, the procedure is reversed to delete any redundant parameters. Although, there is no aggregation of nodes, which correspond to the thermal zones of the building, since the data is recorded by smart thermostats, the estimated parameters of the RC networks show how in some cases the capacitances of specific nodes are equal to a null value. This effect corresponds to a form of aggregation of two adjacent thermal zones.

Optimization MOR approaches can facilitate the development of *automated* methodologies for RC thermal models of buildings. Traditionally, MOR techniques rely on detailed models as a starting point for the reduction process. However, the availability of data from smart thermostats in buildings has led to a shift towards practical approaches that generate energy models without the need for detailed initial models [96, 97]. This shift provides an opportunity for automation in the selection of the structure of the model that describe the building dynamic, prioritizing accuracy and simplicity for control-oriented applications. This methodology must deal with scarce data, while still being able to provide accurate models, and be less computationally intensive, providing results in acceptable time frames. This modelling automation, in turn, requires a combination of data extraction, preparation, selection of the state variables and structure of the model, and calibration [92, 98].

2.2.3 Thermal zoning and model archetypes

Thermal zoning and model archetypes are fundamental concepts in building energy modelling, each with distinct strengths and limitations that shape their practical application. Thermal zoning involves dividing a building into separate zones, each with unique thermal properties and control parameters [91]. This approach offers several advantages. First, it simplifies the modelling process by allowing each zone to be treated independently, making the implementation easier and the analysis more straightforward. Additionally, thermal zoning is particularly useful for buildings with complex architectural designs and mechanical system, as it provides a detailed understanding of the different thermal zone dynamics. Through this approach, the transition occurs from predictive modelling only for temperature forecasting to developing models that enhance the optimization of the mechanical system's performance. Finally, this approach is very powerful to optimize building operations based on predefined schedules since thermal zoning supports the aggregation of zones with similar usage patterns.

Model archetypes are generally associated with the zoning concept, and represent standardized building models for generalizing and predicting the behaviour of similar types of buildings [99]. Model archetypes provide a standardized framework for building energy modelling, enabling scalability across multiple buildings with similar characteristics. This standardization simplifies the process of energy analysis and comparison across different buildings or building types. Moreover, using archetypes reduces the need for detailed data collection and model calibration for each individual building, significantly cutting down on the time and cost associated with creating and validating energy models. Model archetypes are particularly invaluable for benchmarking energy performance and developing energy policies. They provide a reference point against which the performance of individual buildings can be measured, aiding in identifying best practices and areas for improvement.

The concept of model archetype has been associated with RC thermal models in [100], where Candanedo et al. present a versatile approach using low-order control-oriented thermal network RC archetypes. This streamlines the development and testing of scalable building control solutions, specifically assessing control strategy effects on energy efficiency and load management.

However, the primary drawback of this approach is the increased complexity when dealing with multizone buildings. Managing the interactions between numerous zones can become challenging, requiring sophisticated models and extensive computational resources. Nevertheless, the primary limitation of model archetypes is their generalization. While they offer a useful starting point, they may not capture the specific nuances and unique characteristics of individual buildings, leading to less accurate predictions. Archetypes rely on assumptions and average data, which may not always reflect actual conditions and behaviours of specific buildings, resulting in discrepancies between model predictions and real-world performance. Additionally, as buildings evolve and new technologies are introduced, model archetypes may require regular updates to remain relevant, posing an adaptability challenge that can limit their long-term applicability and accuracy.

2.2.4 Model selection for single-building versus large scale applications

When defining the structure and detail of a model, the importance of its end-use cannot be overstated. The specificity of a model application dictates the level of detail, and the structural complexity required to achieve accurate and reliable results. Various researchers have developed methodologies tailored to specific applications, ensuring that the models they produce are optimized for their intended purposes. For instance, in energy modelling for buildings, a model designed to predict hourly energy consumption for an individual building might differ significantly in structure and granularity from a model aimed at forecasting regional energy demand. The former would require detailed inputs on individual building characteristics, occupant behaviour, and specific HVAC system performance, whereas the latter would aggregate data across many buildings and possibly integrate broader climatic and socio-economic factors.

In literature, Privara et al. [101] developed a methodology for grey-box models in predictive control applications, emphasizing that identifying an appropriate model for MPC is a critical challenge. They state that to maximize performance over the prediction horizon, parameters must be calibrated to minimize prediction error throughout. Arroyo et al. [102] demonstrates the feasibility and benefits of using multi-zone grey-box building models for predictive control. By splitting the parameter estimation process by individual zones, the complexity is reduced, allowing for more accurate and effective models. Results from a virtual test case in the BOPTEST framework highlight the importance of accounting for thermal interactions between zones to enhance simulation and control performance. Vallianos et al. [95] presented an automated methodology for generating multi-zone RC thermal network models of residential buildings, essential for advanced control strategies like Model Predictive Control and building energy flexibility estimation. In [103], the authors evaluate the impact of model resolution and structure on the performance of MPC in an unoccupied research house in Québec, equipped with smart thermostats. Two low-order models are compared with the high-order multi-zone model and were calibrated using measured data, with the multi-zone model structure being generated automatically during calibration. These models were used to apply real-time MPC to the experimental house, utilizing dynamic tariffs for morning and evening peaks. All three models enabled successful preheating before demand-response events, outperforming a reference reactive controller by reducing cost and thermal discomfort. The high-order multi-zone model achieved the best performance, cutting electricity costs by 55% and high-price energy consumption by 71%. In comparison, the low-order models reduced costs by 40% and 44% and high-price energy consumption by 48% and 54%, respectively.

Concerning large scale applications, Ouf et al. [104] used smart thermostat data to investigate the thermal performance of 11,000 Canadian houses. They represented each house with a simple 1R1C model and used two methods to identify the model parameters: least-squares fitting of exponential decay curves and of the overall energy balance. Huchuk, Sanner, and O'Brien [105] used smart thermostat data from 1,000 houses in the United States to evaluate data-driven thermal models for multi-hour predictions. They compared simple 1R1C grey-box models and various black-box models, including Lasso and Ridge Regression, Random Forest and

Autoregressive models with exogenous variables, to a baseline model that consisted of a constant temperature. They showed that the first-order models were outperformed by the other models, and they suggested the order and structure of the model as the most likely explanation. Vallianos et al. [106] utilized smart thermostat data from 60,000 houses in North America to create single-zone models using an automated forward selection procedure. Results showed that 61% of the final models were good fits, with 80% being 5th-order models (five thermal capacitances). The 24-hours prediction error analysis confirmed that the good-fit models were accurate enough for day-ahead predictions and MPC applications. Despite no strong correlation found between model parameters and available metadata, the estimated time constants provided valuable insights into the houses' thermal inertia. These models, capable of accurately capturing and leveraging thermal inertia, are ideal for MPC and energy flexibility applications for large scale deployment.

These works show the impact of different research objectives, i.e., predictive control, flexibility potential, energy efficiency, individual building versus large scale, on the choice of specific properties which affects the complexity and performance of the model.

Developing energy modelling methodologies that offer adjustable levels of granularity is critical to standardizing practices across scales. These approaches must allow for seamless customization of model detail to suit specific applications, avoiding the need for complete redevelopment. This adaptability enables researchers to balance model complexity with computational efficiency while maintaining robustness. Flexible modelling methods are essential for applications ranging from detailed building system optimization to strategic energy planning at urban or national levels, ensuring the methodologies remain relevant and scalable.

2.2.5 Application of reduced order thermal models

The development of thermal models gives the possibility to study the building load response under different conditions. Several works adopted RC thermal networks to (i) study the design of new buildings or retrofitting of existing ones, (ii) perform analysis on the integration of energy efficient technologies, and (iii) apply advanced control strategies.

The ISO 13790 standard [107] introduced a monthly method and a simplified RC-based hourly method that can be used to calculate the heating and cooling needs for buildings. This standard is now replaced by the ISO 52016-1 [108] which includes a more detailed RC network. This proposes a simplified methodology to calculate the building thermal load using RC models. The calculation methods have been devised to compute both basic energy loads and system-specific needs independently from technical building systems. Additionally, the hourly calculation procedures serve as a foundation for more complex calculations involving advanced system control options.

In literature, several works show the potential of different RC thermal network architectures for design and control applications [61]. Ravelo et al. [58] uses RC thermal networks to model multilayer walls and study the effects on retrofit. Shen et al. [109] created a tool based on the RC modelling method and showed the performance for rapid modelling and assessment of building energy for a variety of energy conservation options. Andrade-Cabrera et al. [110] provided lumped parameter models and show their suitability for integrated analysis of building retrofits and electricity grid models. Sigounis et al. [111] studied the integration of photovoltaic thermal collectors in buildings using RC thermal networks. They show that the optimal design and control of this technology can reduce the building energy consumption by 40%. Ioannidis et al.

[112] exploited the RC thermal network approach to evaluate the energy performance of double skin facades integrating PV systems. Similar analyses conducted by considering the same modelling techniques are available in the literature [113, 114]. Petrucci et al. [115] studied the interconnection between design and control when choosing the size of heat pumps and solar technologies in a virtual community. The results showed a reduction of 40% in total energy consumption and a peak reduction of around 32% for the optimal design-control scenario. Maturo et al. [39, 55] used a high order RC thermal network to study the design and operation of integrated solar and storage technologies in an office building, using genetic algorithm and model predictive control respectively. Other studies combined the use of model predictive control with grey box model [116-118], showing reduction in the energy consumption and greenhouse gas emissions over 50%. Building models based on RC thermal network approach have been also increasingly used to perform comfort analyses [119, 120], to evaluate the performance of integrated technologies and control techniques [62, 70].

The use of grey-box models, specifically RC thermal network models, holds vast potential. Over recent years, these models have garnered significant interest within the research community due to their ability to balance complexity and computational efficiency while capturing the essential thermal dynamics of buildings. This growing focus reflects the pressing need for scalable, model-based approaches to address challenges in energy management and building performance optimization. Integrating design and control strategies with automated methodologies capable of generating accurate models of building thermal dynamics enables the broader adoption of model-based approaches. This advancement paves the way for developing comprehensive guidelines and fostering a more automated, efficient, and sustainable building environment.

This approach is further strengthened by leveraging artificial intelligence [121], which facilitates autonomous management through tasks such as monitoring, analysis, and decision-making to enhance energy efficiency and resilience [122]. Additionally, the integration of RC models supports the evaluation of energy optimization scenarios, spanning individual buildings to urban scales, and extends to applications in control of smart grid environments [123].

2.2.6 Base load modelling

Base load modelling is essential to fully represent building energy consumption which includes both controllable and uncontrollable components. While thermal energy modelling relies on heat transfer equations and HVAC system performance, base loads—often considered static—account for energy uses like lighting, domestic hot water, and miscellaneous equipment.

The advance in smart metering infrastructure has provided access to high-resolution electricity consumption data, enabling researchers to analyse consumer load profiles with greater precision. This development facilitates a more nuanced characterization of base loads and their integration into comprehensive energy models. Transitioning from raw data to reliable models involves several key steps, including data cleaning, instance selection, feature extraction, and the application of appropriate modelling algorithms [79]. Feature extraction techniques such as Principal Component Analysis (PCA) and Autoencoders are widely employed in this context. These methods reduce the dimensionality of large datasets while retaining critical information, streamlining the modelling process and enhancing the accuracy of base load representation [124, 125].

The literature on electric consumption forecasting continues to expand, driven by advancements in data analytics, artificial intelligence (AI), and machine learning (ML). Researchers are leveraging these technologies to develop more efficient and generalized models for predicting energy consumption. A critical aspect of this endeavor is the use of clustering techniques, which extract representative load patterns from historical data, thereby improving the accuracy and robustness of forecasts.

Various clustering methodologies are applied to process and classify load profiles, each with distinct strengths and applications. Division-based clustering algorithms, such as K-means and fuzzy C-means, are widely used for their simplicity and effectiveness in handling large datasets. Hierarchical clustering offers insights into data structure by creating nested groupings based on similarity measures. Self-organizing maps (SOM), a type of network-based clustering algorithm, are particularly adept at visualizing high-dimensional data and identifying consumption trends. Density-based clustering algorithms, like DBSCAN, excel in discovering arbitrarily shaped clusters and outlier detection. Model-based clustering approaches, which assume data is generated by a mixture of underlying probabilistic distributions, provide a statistically robust framework for classifying load profiles [126, 127].

Short-term Load Forecasting (STLF) models traditionally rely on statistical approaches such as Support Vector Regression (SVR), Auto-Regressive Moving Average (ARMA), and Auto-Regressive Integrated Moving Average (ARIMA). While these methods can provide satisfactory results in linear scenarios, they require extensive datasets for training and often suffer from limitations, such as collinearity among predictor variables, which complicates the prediction of dependent variables and affects the statistical significance of the models. Furthermore, conventional statistical models struggle to capture the complexity of electric consumption patterns, particularly under conditions involving noisy data or concept drift—where the statistical properties of the target variable change over time. To overcome these challenges, researchers are increasingly turning to advanced ML techniques, which offer greater flexibility and capability to model complex, nonlinear relationships in energy consumption data. Methods such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM) have gained prominence due to their robustness in handling intricate patterns and variability in electric consumption. ANN is particularly valued for its ability to learn and generalize from large datasets, capturing intricate temporal dependencies [128], while SVM excels in scenarios with limited datasets and high-dimensional spaces [129].

2.3 Integrated technologies

As the demand for energy-efficient buildings increases, to reduce energy consumption and enhance indoor comfort it is crucial to optimize heating, ventilation, and air conditioning (HVAC) systems. Among the various technologies employed in buildings, ventilation systems and radiant heating systems stand out for their effectiveness in maintaining indoor air quality and thermal comfort. Advanced thermal storages also represent an important way to improve thermal comfort and energy efficiency by exploiting their thermal inertia.

The need of sustainable energy solutions has also driven significant advancements in renewable technologies and battery storage systems for the implementation in buildings. Among the most prominent renewable energy sources are photovoltaic (PV) systems and building-integrated photovoltaic-thermal (PV/T) systems, which harness solar energy to generate electricity and thermal energy, respectively. Complementing these renewable technologies are battery storage systems, playing a crucial role in managing the intermittent nature of renewable energy sources.

This section explores the above-mentioned technologies and their primary aspects, serving as a foundation for the methodologies and applications investigated in this thesis.

2.3.1 Synergies between ventilation and radiant heating

Ventilation systems are crucial for maintaining indoor air quality by supplying fresh air and removing stale air, pollutants, and excess moisture. Effective ventilation is essential for ensuring a healthy indoor environment, reducing the concentration of indoor pollutants, and preventing the buildup of humidity that can lead to mold growth and structural damage. Traditional ventilation systems operate based on predefined schedules or simple on-off controls, leading to inefficiencies and energy waste. Advancements have introduced efficient energy recovery systems and sophisticated control strategies that adjust ventilation rates based on real-time occupancy, indoor air quality sensors, and predictive models of indoor conditions. Yang et al. [130] demonstrate that efficient control of ventilation heating can lead to electricity savings ranging generally between 15 and 20%. Zhao et al. [30] show that HVAC systems can also provide frequency regulation services to the electric grid by adjusting the power consumption in response to a signal sent by the electric grid operator. Despite these improvements, the primary focus of most studies has been on the independent optimization of ventilation systems without considering their interaction with other HVAC components or building elements [131, 132].

Radiant heating systems, particularly those integrated into building envelopes (also called *active* envelope systems), provide an efficient means of heating spaces by emitting infrared radiation that directly warms occupants and objects. This method of heat delivery is often more energy-efficient than traditional convective heating systems because it reduces the need for high air temperatures and minimizes heat losses [133]. Radiant heating systems can be categorized into high-temperature systems, such as radiant panels, and low-temperature systems, like underfloor heating. Each type of system has distinct thermal dynamics: high-temperature systems provide rapid heat delivery but can create uneven temperature distributions, whereas low-temperature systems offer more uniform heating but have slower response times. The complexity of these dynamics poses challenges for integrating radiant heating with other building systems, such as ventilation, where synchronization and control become critical.

The integration of ventilation heating and radiant heating systems requires careful consideration of their differing thermal dynamics. Ventilation systems typically exhibit fast-reacting dynamics, responding quickly to changes in setpoint temperature [134], occupancy and indoor air quality [135]. In contrast, radiant heating systems, especially those with high thermal mass like underfloor heating, have slower dynamics due to the time required for heat to transfer through materials and affect room temperatures [136]. To overcome this disparity and to achieve a proper integration of ventilation and radiant heating systems, advanced control strategies are needed to ensure that the systems operate synergistically, maximizing energy efficiency and indoor comfort. Predictive control algorithms, particularly MPC, offer a robust solution to this integration challenge [137]. MPC leverages predictive models to anticipate future conditions and optimize control actions accordingly. Control-oriented grey-box models, which combine physical principles with empirical data, are particularly well-suited for this purpose [138]. These models can accurately capture the thermal behavior of both ventilation and radiant heating systems, enabling precise control that accounts for their interactions and dynamic responses. By integrating these systems through MPC, it is possible to achieve coordinated operation that minimizes energy consumption and enhances indoor comfort. Examples of the proper development and use of this technique are provided by Maturo et al. [55] that proposed a methodology, combining particle swarm optimization search and MPC, to deal with the control of a high mass thermal energy storage system. Similarly, Chen et al. [139] developed a state-space model of a variable-flow radiant heating system and used an MPC controller to reduce the response time of the technology by about 56% compared with a PID controller.

Nevertheless, despite the potential benefits, the integration of ventilation and radiant heating systems has received limited attention in the literature. Several studies have explored the design and control of these systems, but comprehensive approaches that address their mutual integration are scarce. Research has predominantly focused on single-zone office buildings or test-rooms, which do not necessarily provide a data-driven and scalable methodology applicable to other building types and configurations.

For instance, studies by Zhang et al. [140-142] investigated the integration of ventilation and radiant heating systems using experimental setups in a test room. These studies adopted MPC and were primarily aimed to enhance occupant comfort by optimizing air quality and thermal conditions. However, they did not delve into the differentiation between thermal energy contributions from ventilation and radiant heating, nor did they propose replicable methods for broader application. In another study, Joe and Karava [143] employed a MPC approach to coordinate the operation of ventilation and radiant floor heating in a single-zone office building. While the results demonstrated improved energy efficiency and comfort, the study’s scope was limited to a specific building type and did not address the challenges of extending the methodology to other building types or multi-zone environments. Additionally, the primary focus was on maintaining comfort levels rather than optimizing the distinct thermal contributions of the two systems. Experimental studies by Viot et al. [144] highlighted the importance of integrating HVAC components for energy savings. These studies utilized real-time data and predictive algorithms to manage the interaction between ventilation and radiant heating. However, similar to previous research, the emphasis was on occupant comfort and overall energy use, with limited exploration of the specific thermal dynamics and energy contributions of each system.

Moreover, there are limited studies assessing the benefits of integrating ventilation and radiant heating systems in terms of energy flexibility [25] and response to price signals [145]. The coordination of ventilation and radiant heating can potentially increase a building's energy flexibility by leveraging the thermal mass of radiant systems to shift energy consumption to off-peak periods and by activating convective-based systems when reactive adjustments are needed. However, most existing studies do not address this potential, they instead focus on the two technologies separately [146], or they focus on the issues of comfort and energy efficiency without considering the broader implications for energy flexibility and cost savings in response to dynamic pricing [147]. Furthermore, the literature reveals a significant gap in methodologies that can be generalized across different building types and operational scenarios. The existing research largely focuses on comfort optimization and does not provide a detailed analysis of the thermal energy supplied by ventilation versus radiant heating. This gap underscores the need for comprehensive studies that address both energy performance and comfort, utilizing predictive control strategies and energy modelling approaches that are data-driven and that can be adapted to various building configurations.

2.3.2 Advanced thermal storage systems

Energy storage systems mitigate building load fluctuations and offer advantages in terms of energy efficiency and operational cost. As affirmed by Ibrahim et al. [148], energy storages strengthen the power network as they allow energy usage during peak hours, enhancing on-site generation-demand alignment. Renewable technologies are utilized to their fullest when integrated with batteries, as stated by Mariaud et al. [149], or thermal storage systems, *sensible* or *latent* based (by using phase-change materials, PCMs [150]). PCMs have gained much attention as they offer a high storage capacity compared to sensible thermal storage systems in a very wide range of possible storage temperatures [151]. As stated by the International Energy Agency (IEA) [152], the main advantages of such an application include a higher thermal energy storage capacity, which results in a smaller storage unit, and the ability to perform both charging and discharging processes at a nearly fixed temperature. Additionally, burner cycles for the backup generation unit can be reduced, leading to lower carbon monoxide and hydrocarbon emissions. However, there are also notable disadvantages. Investment expenses are high, and the finite heat conduction of the solid-state PCM limits the discharge power. Furthermore, long-term charge-discharge cycles are restricted, and there are risks associated with material stability loss and deterioration of the encapsulation material.

Researchers have proposed several configurations to demonstrate the financial and grid benefits of PCM thermal energy storage for heating and cooling applications in buildings. Two main approaches can be introduced: passive and active utilization.

Studies on passive usage of this technology explore integrating PCM in building envelopes or furniture. Athienitis et al. [153] proposed an experimental and numerical simulation study on the application of PCM in building envelope components for passive solar design. The results showed that the PCM gypsum board may reduce the maximum room temperature by about 4 °C during daytime and can reduce the heating load at night significantly. Hu et al. [154] proposed a passive envelope solution that integrates PCM and hemp concrete for energy, thermal and hygrometric purposes. They simulated the behaviour of PCM closest to the interior during summer, resulting in reductions of 8.2% and 46.3% for heat load and temperature fluctuation respectively. Buonomano et al. [155] investigated the impact of PCMs integrated in building wallboards on the thermal behaviour of an office building located in a Mediterranean climate. The authors showed that the additional heat capacity obtained by PCMs reduces the internal surface temperature fluctuation up to 3 °C, with energy savings (from 2.3 to 4.1%) mostly obtained during the mild heating season. These studies indicate that a passive integration of PCMs positively affects thermal comfort. Therefore, their effects on energy efficiency are limited since the charging and discharging processes are guided by the natural convection [156].

Active configurations allow a better control of the PCMs and are proven to be more effective for load management. Active installations are particularly promising for retrofit applications where it is necessary to increase storage capabilities in conjunction with renewable systems and advanced control [55]. PCM technology can add significant controllable and dispatchable thermal energy storage in very low volume and low weight, thus eliminating the need for structural reinforcement during retrofit. Several studies analyzes the active integration of PCM in a Trombe wall [157] or placed on the ceiling of a thermal zone. Candanedo et al. [158] studied the design and operation of an energy storage device which can store and release energy with an air stream passing between PCM panels, operating as a PCM-air heat exchanger (PCM-HX). The results showed a peak heating load reduction under different control scenarios of up to 41% compared to a benchmark case without the technology.

During active utilization of PCM it is possible to distinguish a charging from a discharging mode. This poses challenges for both modelling and control stages, increased by the variability of the PCM properties according to its temperature. These features overcomplicate the system especially when linking the PCM with other technologies, further increasing the computational times and complicating numerical convergence. Maturo et al. [55] provided a methodology to optimize the energy performance of a PCH-HX installed between two thermal zones in combination with a building integrated photovoltaic thermal (BIPV/T) system. The results demonstrate an increase in energy efficiency, with savings ranging from 9% to 28% compared to a suitable baseline scenario, and a significant energy shift from on-peak to off-peak periods, potentially accounting for up to 46% of the total building load.

2.3.3 Solar energy integration: PV and PV/T Solutions

Renewable energy technologies, particularly PV and PV/T systems, offer numerous advantages in the quest for sustainable energy solutions. Solar PV systems convert sunlight directly into electricity, providing a clean and renewable energy source with low operational costs. Their scalability makes them suitable for a range of applications, from residential to large-scale utility installations. Therefore, PV systems face challenges such as intermittency due to varying sunlight conditions and efficiency limitations. Those systems are in fact characterized by a very small inertia, even if compared to wind farms, causing high frequency variation and compromising grid stability [159]. Additionally, most panels achieve around 15-20% efficiency and large-scale installations require substantial land area, which can be a constraint in densely populated regions.

PV/T systems enhance the efficiency of solar energy utilization by simultaneously generating electricity and thermal energy [160]. This dual functionality improves overall energy conversion efficiency and reduces the need for additional space by integrating the system into building structures. Several works enhance PV/T in building applications, especially for reaching the NZEB goal [161]. Li et al. [162] studied different configurations of building integrated PV/T, showing benefits in terms of energy production. Tomah et al. [163] demonstrated how PV/T can provide reduction of carbon emissions in the building sector and reduce the urban heat island effect. Despite these benefits, building integrated PV/T systems involve complex design and installation processes, higher initial costs, and require effective thermal management to prevent overheating. Effective orientation, shading considerations, and temperature effects are crucial factors that impact the operational efficiency of both PV and BIPV/T systems.

2.3.4 Battery storage: adaptation and application at various scales

Advancements in battery storage technologies, bolstered by improvements in reliability, efficiency, and cost-effectiveness, are expanding their role in the electrical power system. Battery storage systems are now being deployed for a variety of new functionalities, including enhancing power quality, optimizing energy management, and supporting grid stability. They facilitate better integration of non-dispatchable renewable energy sources, enabling innovative operational modes such as microgrid island operation [164]. Additionally, these systems contribute to reducing peak loads [21], providing frequency regulation [165], and offering fault detection. For these reasons, hybrid wind and photovoltaic power plants integrated with battery storage systems are gaining traction in research and practical implementation.

Delfanti et al. [166] indicate that while battery storage systems may not yet be cost-effective, anticipated price reductions could change this scenario. Despite current cost challenges, battery storage systems significantly enhance the penetration of renewable energy sources and offer various market-oriented services. With the continued reduction in investment costs and improvements in battery storage capabilities—such as extended lifespan despite degradation over multiple charge-discharge cycles—their economic feasibility is becoming increasingly viable.

Different types of batteries, such as lithium-ion, lead-acid, and flow batteries, offer various benefits and drawbacks. Lithium-ion batteries are renowned for their high energy density, efficiency, and longer cycle life, making them a popular choice for both residential and commercial applications. Lead-acid batteries, although more affordable, have lower energy density and shorter lifespans, limiting their use in more demanding applications. Flow batteries, though less common, offer scalable storage capacity and longer cycle life, making them suitable for large-scale energy storage solutions. In their review, Bullich-Massagué et al. [159] demonstrate the importance of selecting batteries based on specific applications. They state that (a) current grid codes require high power and medium energy storage, for which lithium-ion batteries are the most suitable technology, (b) future grid code requirements will demand high power, low energy, and fast response storage, with supercapacitors being the preferred option, (c) lead-acid batteries are adequate for supporting black start services, and (d) both flow batteries and lithium-ion technologies can be utilized for market-oriented services.

In their review, Parra et al. [167] differentiate between different levels of application for battery storage systems (Figure *2.2*): single home (for end-user, kWh scale), community (for end-user and the network, kWh-MWh scale), grid (for the regional network, MWh scale), and bulk storage (for the generators, GWh). Single home and community applications are associated with low-voltage distribution networks, while grid-level applications pertain medium-voltage distribution networks, and bulk storage is relevant to high-voltage transmission networks.

Battery storage systems near high-voltage generators, especially variable ones like large-scale photovoltaic power plants (LS-PVPPs), play a crucial role in both active and reactive power regulation [159]. In terms of active power, these systems assist in power curtailment by absorbing excess generation, provide frequency regulation by balancing supply and demand, and control ramp rates to smooth power output variability. Reactive power regulation includes voltage support and power factor correction, ensuring stable voltage levels and reducing transmission system losses [168]. The sizing of battery storage systems is regulated in various countries to ensure optimal functionality in delivering these services [169, 170]. Grid codes are also mandating capabilities for fast frequency response and inertia emulation [171], which can be limiting for LS-PVPPs due to their low inertia levels, as well as black-start capabilities (with examples from the California ISO) and power oscillation damping related to low-frequency electromechanical oscillations. These services will accommodate the evolving energy landscape.

When battery storage systems are deployed at lower voltage levels, closer to the distribution grid or end-users, different challenges and needs arise. Load leveling and load balancing become critical to manage variable demand and supply at these points, reducing stress on the grid infrastructure. Energy arbitrage allows for economic optimization by storing energy during low demand/price periods and discharging during peak demand/price periods, improving overall system efficiency and economy. At this level, and as defined in Figure 2.3, energy storage systems can also be categorized based on their ownership and operational structure: private, interconnected, common, and independent energy storage operators [172, 173].

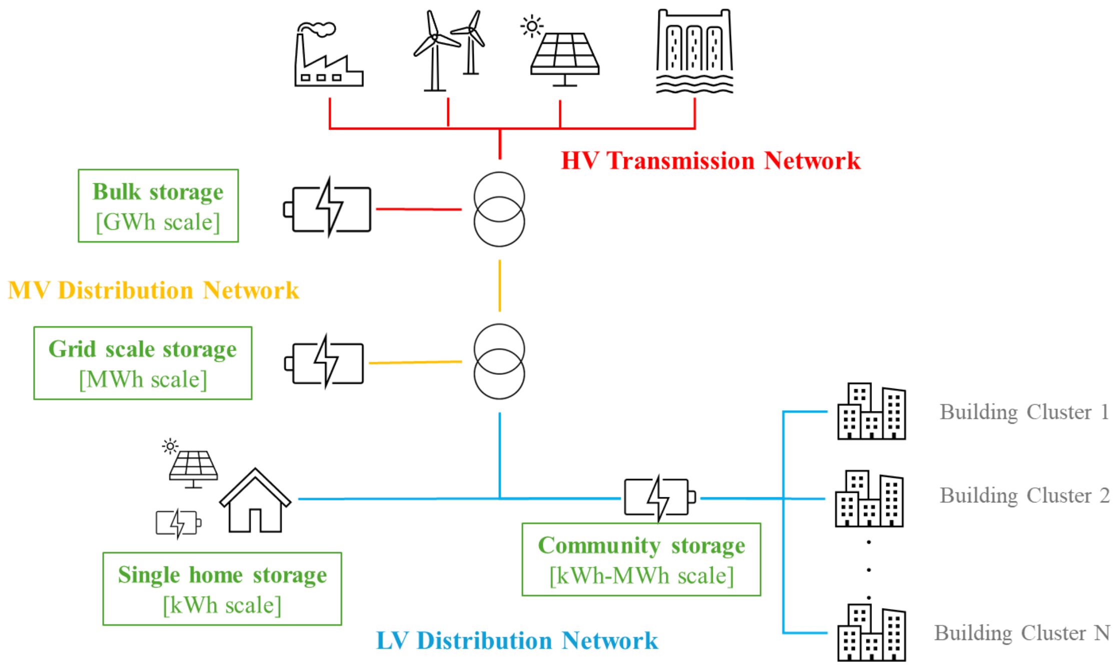


Figure 2.2: Simplified schematic representation of large-scale battery storage system application.

Private storage systems are typically used by individual households or businesses to manage their energy use and reduce reliance on the grid. Interconnected storage systems are integrated with the grid, allowing for the exchange of stored energy between different users and contributing to grid stability. Common storage systems serve multiple users within a community or building, often managed collectively to optimize energy use and reduce costs. Independent energy storage operators (e.g., aggregators, virtual power plants, microgrids) manage large-scale storage facilities that provide services to various stakeholders, including utilities and grid operators, enhancing overall energy resilience and reliability.

Maximizing self consumption and time-of-use (TOU) tariff strategies are the two most common battery operations at individual level. However, these strategies have some limitations. The first strategy does not account for battery degradation and increased dwell time at high state of charge level [174]. The TOU strategy involves buying energy from the grid at a low tariff and using it during high-price periods. Talent et al. [175] noted that the optimal battery size is not affected by the tariff structure, but the tariffs significantly impact on the system's economic efficiency. The TOU strategy tends to have significant advantages over other strategies in terms of economic efficiency, but it is only viable for specific users whose TOU tariff and load profile are well suited. In their review, Li et al. [172] emphasized the necessity of exploring the impact of different tariff structures and grid signals on users' load profiles and distribution network demand profiles, analyzing the combined effect of different tariff structures on battery system performance.

Applications in energy communities yield higher economic returns and operational benefits. Aggregating demands across various buildings can enhance battery performance and size, reducing capital expenditure (CAPEX) due to economies of scale [176-178]. Among the different battery structures, a stand-alone energy storage system has emerged, where the battery is owned by "Independent Operators" (e.g., aggregators, virtual power plants, microgrids) but used by multiple users. Despite the high economic benefits of shared battery systems, there are still limited studies focusing on exploring the potential of the energy storage in terms of peak shaving, resiliency [179, 180], and other ancillary services [173].

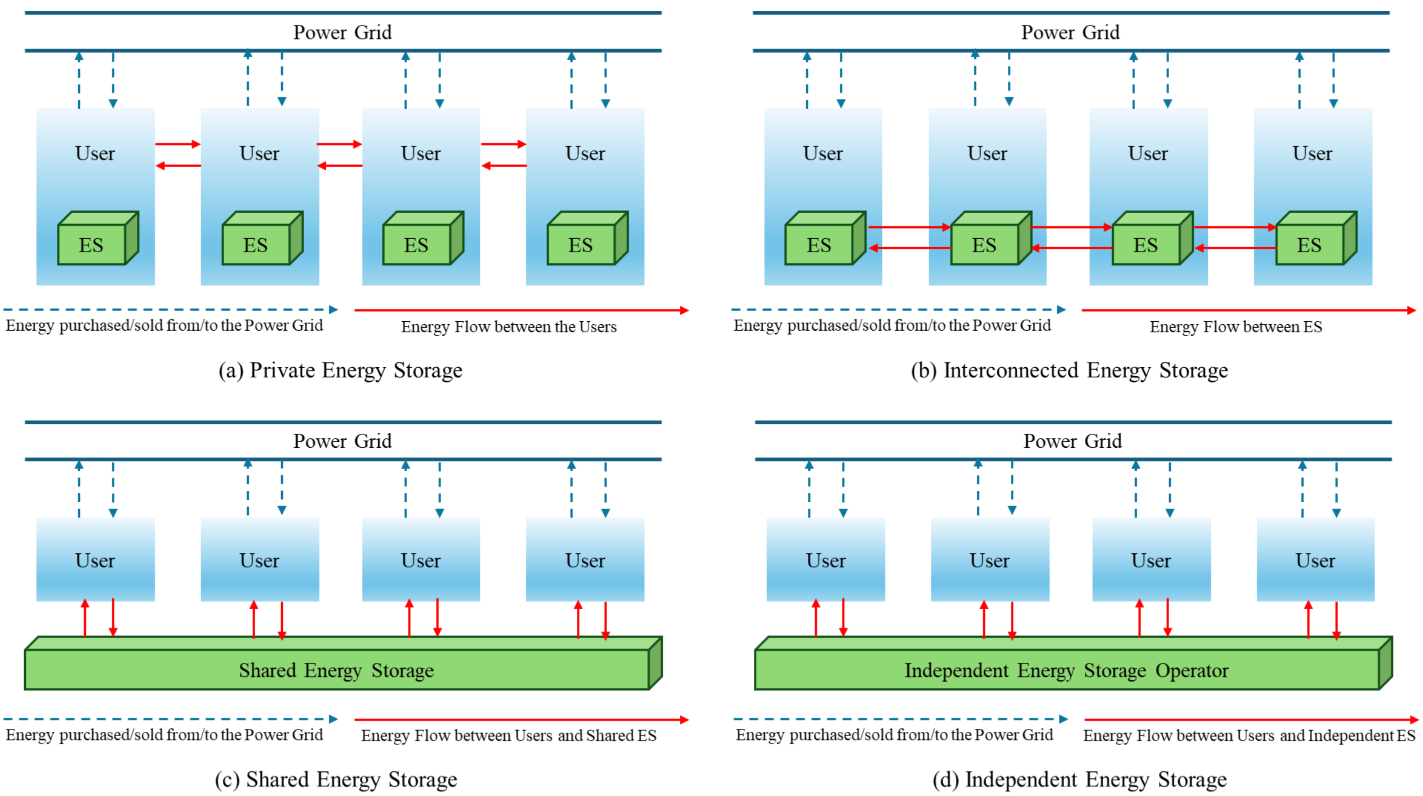


Figure 2.3: Structure of building energy storage systems in energy sharing communities.

2.4 Architectures and strategies for building energy management

Energy management strategies can be classified according to the "smart energy management matrix" [181]. As described in Figure 2.4, this matrix focuses on two main features: *communication*, distinguishing one-way or two-way, and *decisions*, which can be local and central. Hence, four different quadrants/approaches can be introduced [181].

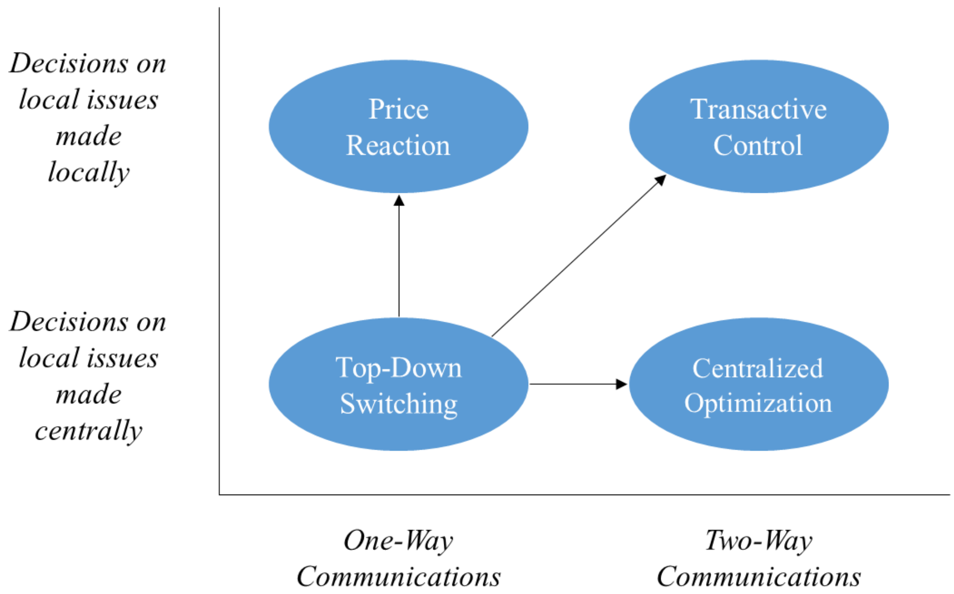


Figure 2.4: Smart Energy management matrix.

The “Top-Down Switching” or “Direct Control” quadrant encompasses traditional demand response programs. In a defined electrical grid area, a group of devices is simultaneously switched based on a broadcasted signal. Typically, the local utility company sends a signal through the power grid to deactivate systems like water heaters and air-conditioning during peak load times. Although the approach is simple and effective, it does not unlock the full response potential of devices, as the device state is not considered. Most of all, the method ignores the consumer altogether. It does not take user preferences into account and interferes with the autonomy of energy consumers.

In the “Centralized Optimization” quadrant, decisions are centralized, but communication is bidirectional. A sophisticated optimization engine oversees all flexible demand and supply within the smart grid cluster. It searches for the best solution for the entire system based on available information and considering global and local control goals. All pertinent local data must be communicated to the optimizer, which then informs the central controller for issuing control signals or schedules to the field. However, centralizing all local information limits accuracy and scalability. Updates are required in the central system with local equipment changes. As the number of responsive entities (houses, buildings, installations) increases, communication and optimization times grow nonlinearly. The approach also lacks resilience in case of communication or central optimizer failures.

The “Price-Reactive Systems” approach relies on one-way signaling of dynamic prices to end users. This price signal enables the efficient optimization of responsive devices through a local intelligent controller, owned or controlled by the consumer. Consequently, the controller can adjust consumer loads during low-priced periods and increase generation during high-priced periods, considering device states and user preferences. While advantageous compared to central optimization, predicting the demand-response pool's reaction to each price signal is challenging without knowledge of individual device states and end user preferences. This framework is generally associated, in literature, to a decentralized control architecture [182, 183].

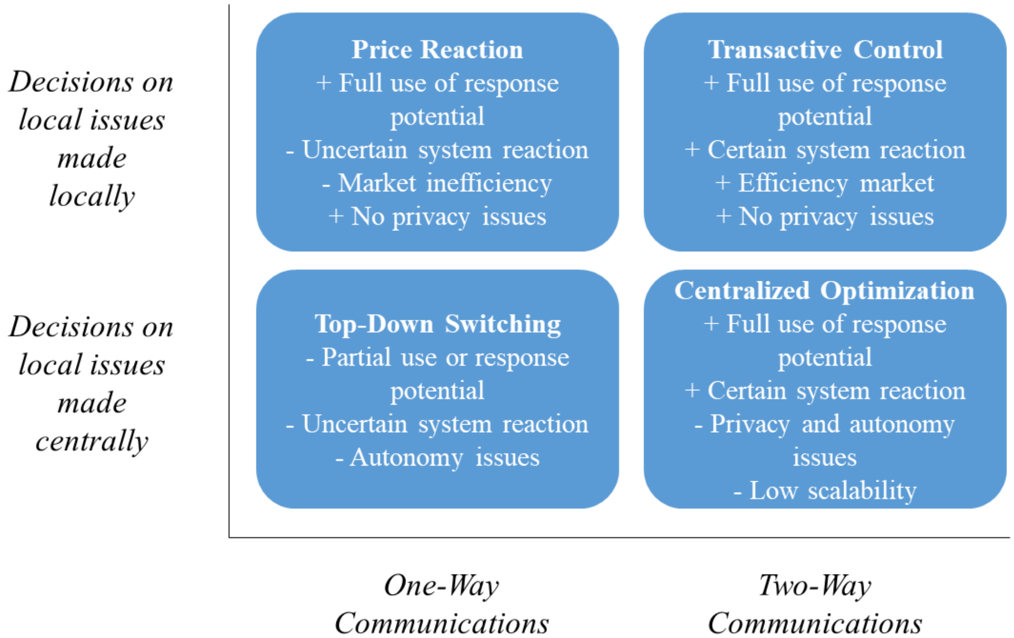


Figure 2.5: Advantages and disadvantages of the several management strategies.

The "Transactive Control and Coordination" quadrant offers significant advantages for integrating flexible devices into electricity operations. Technically, Transactive Energy Control (TEC) is defined by the GridWise Architecture Council as "a system of economic and control mechanisms that allows the dynamic balance of supply and demand across the entire electrical infrastructure using value as a key operational parameter" [184]. TEC can be implemented using hierarchical and non-hierarchical distributed control architectures. In [183], a non-hierarchical distributed control, also known as peer-to-peer control, involves prosumers directly in the control process [185]. Here, agents or prosumers manage the system through communication with other agents without any coordinator or system operator. In a hierarchical distributed control, like in non-hierarchical, agents manage the system, but the communication occurs only between agents and a coordinator or system operator.

On the other hand, in centralized control, the coordinator centrally manages the system, while agents can send responsive signals to the coordinator based on the control signals. Transactive energy control approaches unlock the full response potential of flexible devices, provide greater certainty about system reactions, can establish an efficient market with proper incentives, and can protect the privacy of end users participating in the energy management task.

However, the presence of multiple microgrids, virtual power plants, or energy aggregators controlling the demand of a building cluster — each receiving signals from the electrical grid — can create hierarchical frameworks [186], which combine the previously mentioned strategies. Selecting the proper control architecture is crucial, as it must align with the specific technologies' configuration, electrical grid targets, the nature of controllable and uncontrollable loads, and whether buildings are participating in demand response events.

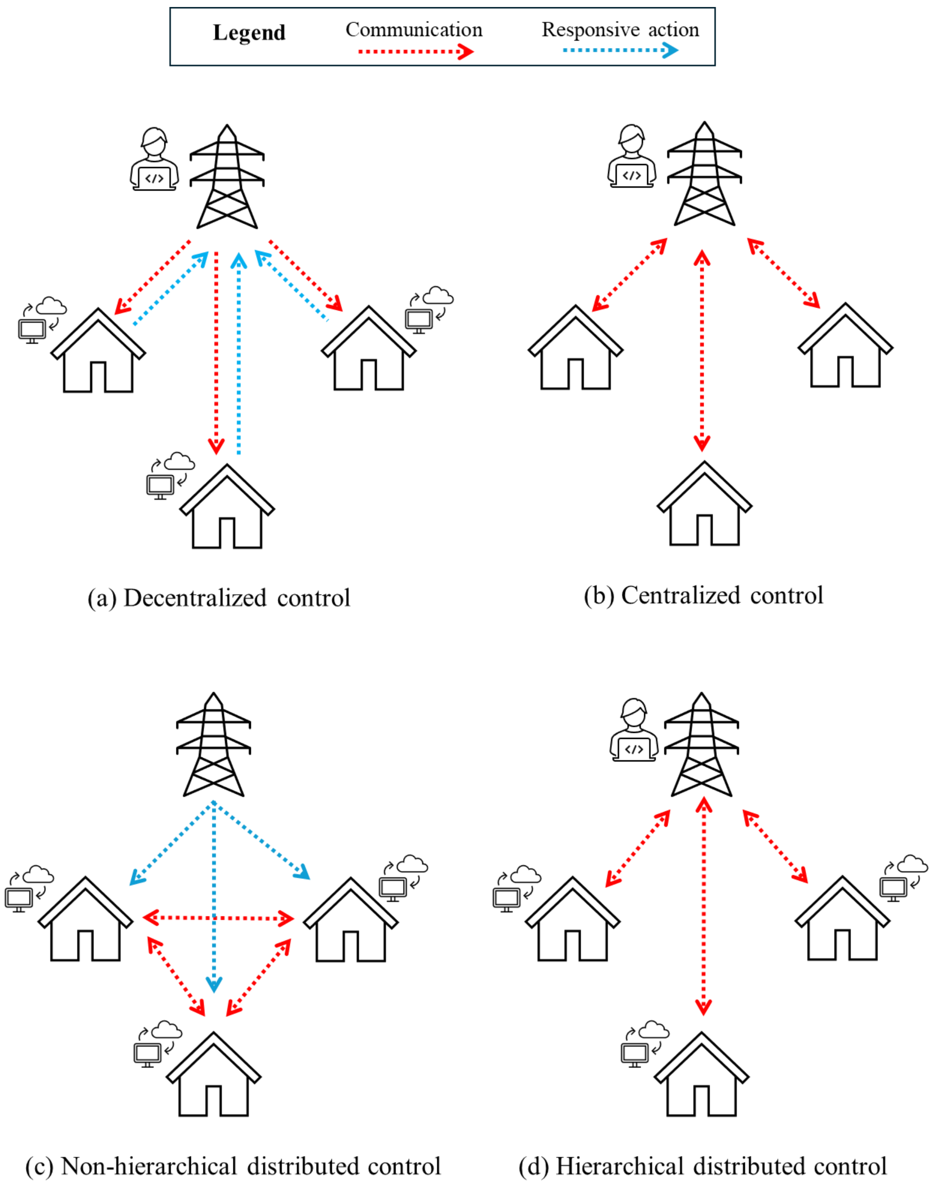


Figure 2.6: Controller architectures for transactive energy control.

2.4.1 Price responsiveness: grid signals and demand response events

The local utility can send two types of signals to customers which are generally categorized in involuntary and voluntary signals. The involuntary signals refer to planned rolling power outages during extreme peak periods, happening when the voluntary demand response fails to reduce power demand to match the maximum power generation. Instead, the voluntary signals refer to (1) incentive-based and (2) price-based programs that encourage participants to adjust their electricity consumption for a short period of time [187]. Incentive-based programs are further divided into classical programs and market-based programs. Classical programs include Peak Time Rebates, Direct Load Control and Interruptible/Curtailable programs. Market-based programs include Emergency DR programs, Demand Bidding, Capacity Market, Ancillary services market. In classical incentive-based programs, participating customers receive participation payments usually as a bill credit or discount rate for their participation in the programs. In market-based programs, participants are rewarded money for their performance depending on the amount of load reduction during critical conditions.

Price-based programs are based on dynamic pricing rates in which electricity tariffs are not flat, so the rates are fluctuating following the real time cost of electricity. The ultimate objective of these programs is to flatten the demand curve by offering a high price during peak periods and lower prices during off-peak periods. These rates include Time of Use (TOU) rate, Critical Peak Pricing (CPP), Extreme Day Pricing (EDP), Extreme Day CPP (ED-CPP), Real Time Pricing (RTP) and Variable Peak Pricing. Table 2.1 describes in detail all the demand response programs defined above. From 2003, several pilot studies were conducted to assess the effects of dynamic tariffs at the household level. Faruqui and Sergici [188] conducted surveys involving multiple pilot studies, with two of them examining the RTP tariff. These RTP pilot studies showcased the superiority of RTP tariffs over flat tariffs in reducing peak demand and decreasing household electricity bills. Nonetheless, due to the risk aversion of most households, they are generally unwilling to face fluctuating electricity prices [189, 190]. Currently retailers usually offer flat tariffs that charge a fixed rate for every kilowatt-hour of electricity purchased.

Dynamic tariffs hold appeal from the retailer's viewpoint. This occurs because they allow retailers to align retail prices with spot market prices, transferring some of the risks from the retailer to the customers. Nilsson et al. [191] noted that dynamic tariffs constitute the most effective strategy to increase demand flexibility. Nojavan et al. [192], in studying the impact of different price schemes on retail profit, suggested that RTP tariffs can result in higher retail profit. The penetration of distributed renewable energy sources led to increased imbalances and associated expenses, and has generated significant opportunities for DR coordinated with dynamic tariffs [193, 194]. Using load, price, and survey data from 119 large customers on dynamic tariffs, Boisvert et al. [195] found 18% of the most elastic consumers provides 75% of the aggregated price response. They also argued that when peak prices are substantially higher than off-peak prices, “Commercial/Retail and Government/Education customers are more price responsive than others.” Several studies speak about the benefits of dynamic tariffs in terms of retail profit, like Nojavan et al. [192]. Doostizadeh and Ghasemi [196] illustrated that implementing day-ahead RTP strategies that optimize retail profit while adhering to specific constraints would be more advantageous for both energy retailers and consumers.

Conversely, there are still some studies suggesting that dynamic tariffs will generate new problems. Dagoumas and Polemis [197] combined a unit-commitment dispatch model and an econometric model and argue that DR will result in changes in the wholesale price. If the retailer is not able to anticipate these changes, they will be exposed to wholesale risks. This conclusion justifies the need for an accurate prediction algorithm that studies the effect of different price signals on the building load at different scales. This effect needs to be accurately quantified through key performance indicators which provide information to the utility in forecasting and reducing the peak power production during certain hours of the day. The wholesale risk can be also addressed by means of aggregators [115] or energy communities [198], complying with inaccurate predictions of power demand and forecasting errors of the distributed energy sources. In France, with the EcoWatt program [199], utilities are proposing an application that alerts the users, in different areas, during challenging periods of electrical demand on the grid. This supports the idea of issuing diverse “action plans” or “directives” to different regions: leading to different zonal pricing and control.

Table 2.1: Classification of demand response programs.

Type		Name	Description	Participants
Incentive-based	Classical	Peak Time Rebates	Utilities provide refunds to customers for reducing demand during specified peak time periods, compared to the expected consumption.	Residential, commercial, industrial
		Direct Load Control	Utilities can remotely deactivate customers equipment with short notice.	Residential, small commercial customers
		Interruptible/Curtailable Programs	Customers must lower their load to predefined levels. Non-responsive customers may incur penalties per program terms.	Commercial, industrial
	Market-based	Demand Bidding	Customers bid a specific load reduction in the wholesale electricity market. Bids are accepted if they are below the market price. Upon acceptance, the customer must reduce the load by the specified bid amount, or penalties may apply.	Commercial, industrial
		Emergency DR	Customers receive incentives for measured load reductions during emergencies.	Commercial, industrial
		Capacity Market	Offered to customers committing to pre-specified load reductions during system contingencies, usually with a day-ahead notice. Participants face penalties for non-compliance with load reduction calls.	Commercial, industrial
		Ancillary Services Market	Customers can bid load curtailment in the spot market for operating reserves. Accepted bids result in participants receiving the spot market price for standby commitment and, if needed, the spot market energy price for load curtailment.	Commercial, industrial
Price-based		Time of Use (TOU)	This rate comprises two-time blocks: peak and off-peak. The design aims to mirror the average electricity cost during distinct periods.	Residential, commercial, industrial
		Critical Peak Pricing (CPP)	This rate incorporates a pre-specified higher electricity usage price added to TOU rates or standard flat rates. CPP prices are triggered during contingencies or periods of high wholesale electricity prices, limited to a specific number of days or hours per year.	Residential, commercial, industrial
		Extreme Day CPP (ED-CPP)	CPP rates for both peak and off-peak periods are applied during extreme days, while a flat rate is used for regular days.	Residential, commercial, industrial
		Extreme Day Pricing (EDP)	Similar to CPP in having elevated electricity prices, ED-CPP differs in that the higher prices are in effect for the entire 24 hours of the unknown extreme day, which is determined a day ahead.	Residential, commercial, industrial
		Real Time Pricing (RTP)	Customers are billed with hourly fluctuating prices that reflect the actual cost of electricity in the wholesale market.	Residential, commercial, industrial
		Variable Peak Pricing	A combination of TOU and RTP, where distinct pricing periods are predefined, but the price for on-peak periods fluctuates based on grid conditions.	Residential, commercial, industrial

2.4.2 Advanced control algorithms: Model Predictive Control

The integration of building energy systems with the electrical grid requires sophisticated control strategies to achieve high levels of flexibility and efficiency. Advanced control algorithms optimize control parameters, enabling seamless interaction between buildings and the grid [200]. Robust communication routines are essential for effective control, facilitating real-time data exchange between building systems and grid operators.

Model Predictive Control (MPC) is particularly effective due to its reliance on models to forecast future conditions and optimize control actions over a defined time horizon. MPC dynamically adjusts control parameters by incorporating factors such as weather forecasts, occupancy patterns, and electricity prices [138]. Key parameters optimized by MPC include HVAC temperature setpoints, load shifting to periods with lower electricity prices or higher renewable energy availability, battery management for stable grid support and cost reduction, and the coordination of renewable energy sources to maximize their contribution. MPC is based on the minimization of a cost function which describes the energy and cost dynamics of the considered system and varies according to the application. A general MPC framework consists of four parts: cost function, constraints, system dynamics, and the current state. The objective of the MPC controller is to minimize a certain cost function over a fixed prediction horizon, ph, while meeting the constraints. For building applications, the strength of MPC lies in the use of a mathematical model of the building to predict its future response. By using these predictions, MPC can optimally choose the control actions based on a given objective while considering occupants' comfort-related conditions, technological constraints, weather forecasts and grid signals in a systematic and flexible way [28]. The optimization thus considers the effects in the future of the disturbance variables that affect the system dynamics to generate the optimal control variables over a fixed control horizon, ch. The framework can be described by the following equations:

$$\min_{u(\kappa),\ldots,u(\kappa+ph)} \sum_{t=\kappa}^{\kappa+ph} \left(J\left(x(t),u(t)\right) + \lambda(t)\varepsilon(t) \right)$$

$$\text{Subject to} \quad \begin{cases} x(t) = Ax(t-1) + Bu(t) \\ y(t) = Cx(t) + Du(t) \\ y_{lb} \le y(t) \le y_{ub} \\ u_{lb} \le u(t) \le u_{ub} \end{cases} \tag{2.1}$$

where J is the cost function, x are the state variables, y are the output variables, u are the input variables, λ is the weight and ε is a slack variable. The constraints refer to the dynamic of the system, represented by A, the state matrix, B, the input to state matrix, C, the state to output matrix, and D, the feedthrough matrix. The inputs and output variables can be subjected to lower and upper bounds.

Several studies demonstrate that MPC can notably reduce energy use and mitigate GHG emissions [201]. Several control frameworks have been developed to deal with multi-energy systems and complex applications at different building levels (i.e., single building, community, microgrid, aggregator). Yu et al. [202] developed a multi-scale framework to control multi-energy systems. Their work demanded for complexity on the control strategy, proposing a composite optimal control approach based on time-scale separation. Maturo et al. [55] proposed a novel cascade methodology, combining particle swarm optimization search with MPC, to deal with mixed integer optimization problems. Li et al. [61] proposed a MPC based robust scheduling strategy to optimize the system performance under uncertainty. Lefebure et al. [203] compared the performances of a centralized and decentralized MPC in a energy hub.

Efficient control of combined energy systems poses challenges due to the diverse range of operation of technologies and applications. This diversity makes the development of a universally applicable and practical methodology challenging. Complex systems often require hierarchical structures for effective optimization, where each level manages specific subsystems and tasks. Accurately defining the cost function for each sublayer and justifying the selection of optimization hyperparameters can significantly enhance results. In MPC, these hyperparameters, such as prediction horizon length and cost function weightings, are critical for balancing objectives like energy efficiency, cost savings, and occupant comfort. Properly tuning these parameters ensures robust and efficient system performance.

2.5 Design-control interconnection: a path to efficiency and flexibility

The energy efficiency of buildings and their energy systems results from their design and the control routines. The choice of optimal design and scheduling is influenced by the technologies involved, which affect the building's performance and potential flexibility. Two key concepts are crucial when selecting and designing these technologies: (a) design and control are interdependent, and (b) the use of multiple energy systems requires special attention during building operation.

The design of thermal energy technologies in buildings is well-studied, with comprehensive guidelines provided by organizations such as ISO, UNI, and ASHRAE. These standards offer detailed recommendations for designing efficient thermal systems, ensuring optimized energy use and maintaining satisfactory comfort conditions. However, substantial opportunities remain to enhance the operation of these systems. Advanced control strategies, particularly when integrating multiple technologies, hold significant potential for improving performance and efficiency. By leveraging intelligent control systems, thermal technologies can be dynamically optimized (both in control and size), by studying their response to real-time inputs and predictive analytics.

For electrical systems, established guidelines exist, particularly for individual buildings equipped with PV and batteries, providing a foundation for their effective design and implementation. Batteries can be sized according to the desired number of autonomy days, ensuring sufficient energy storage to meet demand during periods without generation [204]. Studies show that adding batteries can increase the self-sufficiency ratio of PV-equipped homes by over 70% [205]. Home energy management system, which controls these technologies, can reduce electricity bills by 27.8% through TOU tariff strategies [206]. Combining batteries with flexible loads can reduce battery size requirements by up to 30% via DSM and peak load shifting [207]. PV systems are particularly cost-effective with large load demands that can be directly consumed [208], and large PV capacities with small battery additions are often recommended [175]. While PV-only systems are profitable for most consumers, adding batteries becomes economically viable with higher electricity prices, favorable feed-in-tariff rates, lower battery costs and more economic incentives [205, 209].

Furthermore, optimizing battery action through *co-planning* with other technologies can improve the optimal system size. Zhang et al. [210] highlight that while parametric and separate size design for basic PV-battery systems is mature, further innovation in system configuration and scale variation, as well as deeper model development such as battery degradation models and load estimation, is needed for optimizing system capacity.

2.5.1 Typical days for enhanced energy analysis

The use of typical energy consumption days offers a practical means to evaluate the interplay between system design and advanced control strategies. By employing representative days that capture the variability in energy use and weather conditions, researchers can effectively explore optimal interactions between buildings and their systems. This approach facilitates the analysis of varying tariff structures, weather forecasts, and control strategies within a manageable framework. By simplifying the evaluation process without compromising on detail, this approach enables more efficient and comprehensive assessments of (i) building and technologies energy performance, and (ii) the effectiveness of control mechanisms.

In literature, Schütz et al. [211] compare six aggregation methods for reducing full year input data and simulations to typical demand days for energy system synthesis. The results show that all the clustering methods, including k-means and k-medoids, provide good performance. Wakui et al. [212-214] model a full year of building energy consumption using typical days for winter, summer and transition periods. Furthermore, they consider a peak day in winter and summer, leading to a total number of five typical demand days with hourly resolution. Moradi et al. [215] optimize combined heat and power systems considering battery storage and thermal energy storages. They represent input profiles with four seasonal representatives that distinguish spring and fall, and differentiate between weekdays and weekends. Schütz et al. [216, 217] use a month-based clustering for reducing the inputs for simultaneously optimizing a building energy system and passive building components. Tostado-Véliz et al. [218] deal with the optimal sizing of a hybrid PV-battery storage system for home energy management considering reliability against grid outages and demand response. The adopted k-means clustering for reducing the building consumption dataset to those most characteristic profiles and manage with the unpredictable behaviour of the outage events. Other studies that employ aggregation methods based on fixed period of the year to optimize individual energy systems include Merkel et al. [219] and Buoro et al. [220]. In contrast to the other works mentioned, these studies use typical weeks instead of typical days. Merkel et al. [219] use a quarter-hourly time discretization and consider three typical weeks for each season (winter, summer, transition). Buoro et al. [220] use one typical week for each month of the year. Therefore, these studies primarily focus on extracting typical load profiles. To study load variations dynamically, clustering should focus on identifying trends in key parameters that affect building loads, such as weather conditions, occupancy dynamics, and grid signals, rather than directly clustering the load profiles themselves. Such an approach could improve the representation of variations critical to optimizing energy systems.

The definition of typical days derived from the obtained results plays a pivotal role in bridging the gap between advanced control frameworks and practical, easily implementable control routines in buildings [221, 222]. By distilling the complexities of dynamic energy consumption and weather patterns into representative profiles, typical days enable the design of generalized control strategies that retain the effectiveness of advanced methodologies while reducing computational and implementation burdens. This approach allows for the development of simplified yet robust control routines that can adapt to a range of building types and operational conditions. As a result, the integration of typical days into control design not only facilitates scalability across diverse applications but also supports the wider adoption of energy-efficient practices in building management systems, contributing to the sustainability and resilience of the built environment.

2.6 Key metrics for system optimization and energy flexibility

To assess the performance of optimal control algorithms and compare the results with the reference scenario, it is essential to establish key performance indicators (KPIs). A central research question emerging from the concept of energy flexibility concerns quantifying buildings' potential flexibility [223]. Various metrics have been developed for this purpose, enabling occupants, building owners, managers, as well as DSO/TSO and utility companies, to gauge their ability to manage their loads and, for grid operators, production levels. The insights provided by flexibility indicators are critical for policymakers and planners [55], shaping strategies to enhance energy management at both grid and user levels.

Drawing on literature and practical insights, evaluating the effectiveness of specific strategies involves a comprehensive analysis of their impacts. This includes examining their potential for peak shaving, energy reduction, energy shifting, and load variability throughout the day. Furthermore, this section categorizes KPIs into three main areas: load matching and grid interaction, thermal energy monitoring and energy flexibility, providing a structured approach to assessing system performance and operational optimization. To build on this framework, the section outlines some of the primary KPIs discussed in the literature, highlighting their relevance and application. These KPIs will serve as foundational metrics for the analyses and methodologies presented in this manuscript.

2.6.1 Metrics for load matching and grid interaction

Load matching and grid interaction indicators are critical for evaluating the performance of energy systems, particularly in the context of building-grid interaction. The journal paper [224] provides further details on load matching metrics, focusing on the extent to which on-site generation meets the building's energy demand and the excess energy that is sold back to the grid. Among the key metrics is the *Loss of Load Probability* (LLP), represents the frequency or magnitude of a system inability to meet load demand. Therefore, it can also indicate the mean percentage of the load unmet by the installed system. LLP is calculated as the ratio of the total energy deficit – which the grid should supply – to the total load demand over a defined period [225, 226]. It can be expressed as:

$$LLP = \frac{\sum_{t=\kappa}^{\kappa+\Delta\kappa} E_{el,grid}(t)}{\sum_{t=\kappa}^{\kappa+\Delta\kappa} E_{el,load}(t)} \tag{2.2}$$

In addition to load matching, grid interaction indicators are essential for assessing the power flows between buildings and the grid. These indicators examine the magnitude of the power injected into or consumed from the grid, as this directly influences the power flow through grid components, as well as associated parameters like currents and voltage levels. While the variability of these values is important, it is the coincidence of energy consumption and generation that determines the frequency and scale of power fluctuations. Relevant grid interaction factors can be computed using actual power values or presented in normalized form. Metrics such as the capacity factor and the dimensioning rate are commonly used to characterize these interactions, offering a clearer picture of the building's role in maintaining grid stability and efficiency [224].

2.6.2 Metrics for thermal monitoring

Thermal monitoring metrics are essential for understanding the dynamic thermal behavior of buildings, enabling precise control over HVAC systems and thermal energy storage solutions. The following indicators provide a clear perspective on the performance and effectiveness of thermal technology monitoring.

Temperature State of Charge ($TSoC_{n,i}$): this metric considers the temperature of the *active* envelope (i.e., radiant slab of a specific thermal zone n) or the thermal storage system, $T_{n,i}$, as a state of charge, taking into account an upper and lower bound. The temperature can be substituted by thermal capacitance in the case of phase change materials [39].

$$TSoC_{n,i}(t) = \frac{T_{n,i}(t) - T_{lb}}{T_{ub} - T_{lb}} \tag{2.3}$$

Generally, for radiant heating systems, the values of T_{lb} and T_{ub} can be fixed to 20 and 26 °C respectively.

Radiative Ratio (RR): this metric evaluates the percentage of thermal energy provided to the *active* envelope nodes versus the total amount of thermal energy required by the building. This metric distinguishes how thermal energy is provided to the building through different heating terminals.

$$RR(t) = \frac{\sum_{n=1}^{N_1} Q_{n,i}(t)}{\sum_{n=1}^{N} Q_n(t) + \sum_{n=1}^{N_1} Q_{n,i}(t)} \cdot 100 \tag{2.4}$$

In this formulation, Q_n represents the convective heating of a specific zone n, $Q_{n,i}$ represents the radiant heating of the *active* envelope of a specific zone n. Instead, N_1 is the number of thermal zones equipped with an *active* envelope, and N is the total number of thermal zones in the considered building.

Ramp Index ($Ramp$): this metric quantifies the maximum and minimum variation in the power consumption through a defined period (day). Unlike the previous indicators, it can be applied to any type of variable.

$$RampUp = \left(\Delta Q_{th}(t)\right)_{\max} = \left(\sum_{n=1}^{N}\left(Q_n(t) - Q_n(t-1)\right) + \sum_{n=1}^{N_1}\left(Q_{n,i}(t) - Q_{n,i}(t-1)\right)\right)_{\max} \tag{2.5}$$

$$RampDown = \left(\Delta Q_{th}(t)\right)_{\min} = \left(\sum_{n=1}^{N}\left(Q_n(t) - Q_n(t)\right) + \sum_{n=1}^{N_1}\left(Q_{n,i}(t) - Q_{n,i}(t-1)\right)\right)_{\min} \tag{2.6}$$

This metric is linked to the index that characterizes the price elasticity of the demand. The price elasticity is evaluated by dividing the $Ramp$ index by the maximum variation of price signal. The price elasticity index is important to capture the user behaviour effect on the energy demand due to a variation of the price signal [227]. Instead, the $Ramp$ index is suggested to monitor technology performance.

2.6.3 Metrics for energy flexibility

An international group of researchers under the IEA's EBC Annex 82, Energy Flexible Buildings Towards Resilient Low Carbon Energy Systems, recently published an article identifying ten key questions on energy flexibility of buildings [36]. In their second question, how can energy flexibility be quantified, they describe key performance indicators as essential to quantify energy flexibility performance considering available flexible resources, building demand, grid signals and occupant comfort needs or constraints. In their review, Li et al. [55] tried to answer to this question by providing a holistic review of data-driven energy flexibility KPIs for buildings in the operational phase. They distinguished baseline-free from baseline required indicators and are here summarized some of the most relevant indicators adopted in important journal publications.

Load Factor (LF): this metric measures the ratio between the average daily energy consumption and the daily peak of energy consumption [228, 229].

$$LF = \frac{\left(E_{el}\right)_{\text{mean}}}{\left(E_{el}\right)_{\max}} \tag{2.7}$$

Flexibility Factor (FF): this metric is evaluated as the difference between the energy consumption during low price and high price periods divided by the total energy consumption [230].

$$FF = \frac{\sum_{\text{low price}}\left(E_{el}\right) - \sum_{\text{high price}}\left(E_{el}\right)}{\sum_{\text{low price}}\left(E_{el}\right) + \sum_{\text{high price}}\left(E_{el}\right)} \tag{2.8}$$

Building Energy Flexibility Index ($BEFI$): this metric measures the difference between the energy consumption during the reference or baseline scenario and the flexible scenario for a specific period of time [231].

$$BEFI_{kW} = \frac{\sum_{k}^{k+\Delta k} \left(E_{el}\right)_{ref} - \sum_{k}^{k+\Delta k} \left(E_{el}\right)_{flex}}{\Delta k} \tag{2.9}$$

$$BEFI_{kWh} = BEFI_{kW} \cdot \Delta \kappa \tag{2.10}$$

$$BEFI_{\%} = \frac{BEFI_{kWh}}{\sum_{k}^{k+\Delta k} \left(E_{el}\right)_{ref}} \cdot 100 \tag{2.11}$$

2.7 Research objectives

Through careful review of the literature, certain research gaps and research objectives have been identified. They are summarized here, and this thesis will try to address them to the best of its ability.

- **Selection of building thermal models and model reduction**: Existing literature on grey-box models often lacks detailed explanations regarding the selection of appropriate RC thermal network structures. The thesis aims to address this gap by proposing methodologies that can be adapted to various building types and conditions. In this thesis, an optimization approach is proposed to automate the creation of these models, directly linking data acquisition to model development.

- **Development of automated algorithms**: A key objective is to develop automated algorithms capable of generating multiple models with varying levels of detail, based on adjustable hyperparameters. This approach will improve the efficiency of model generation while maintaining flexibility across different building configurations, system architectures and control needs.

- **Control-oriented model archetypes**: In cases where automated methods may fail due to complex building configurations or specific mechanical system requirements, control-oriented model archetypes can provide a viable alternative. These archetypes will be selected to simplify the application of modelling techniques in such cases, enabling more consistent and reliable system optimization.

- **Coordination of different building elements through model predictive control**: The thesis will explore the importance of coordinating building elements (i.e., convective heating and radiant heating) with different thermal dynamics to improve operational efficiency, reduce energy waste, and enhance flexibility in responding to varying electricity pricing signals.

- **Interdependence of control and design**: Another objective is to investigate the effect of grid signals on load profiles at both individual building and community levels, specifically examining how these changes affect the aggregated load and grid metrics. The thesis examines the relationship between advanced modelling, control, and design to optimize grid-supportive systems. It explores how predicting buildings participation in demand response is essential for enabling the grid to adapt to predicted energy consumption patterns while enhancing the efficiency and design of grid-supportive energy systems.

- **Key performance indicators and typical days**: The thesis will also address the importance of specific KPIs in evaluating system performance and optimizing operation. By using clustering approaches to define typical days, the work will generalize results and establish guidelines for the design and operation of energy systems. Developing a methodology based on typical scenarios will enable more effective decision-making and enhance building performance across different contexts.

Figure *2.7* provides a comprehensive mapping of the research objectives described above to the corresponding chapters of this thesis. Chapter: 3 and Chapter: 4 primarily focus on the development of innovative modelling strategies, addressing key aspects of building energy performance. Chapter: 5 integrates the modelling results obtained from these two chapters, applying them to control and design applications that align with the overarching goals of this research.

Through these objectives, this thesis aims to fill significant gaps in the existing literature by developing novel methodologies and frameworks that integrate advanced modelling, control, and design principles. By addressing the interdependencies between system operation, grid interaction, and energy optimization, this work contributes to the ongoing evolution of building energy management systems, offering practical and advanced solutions for improving real-time energy efficiency, flexibility, and sustainability.

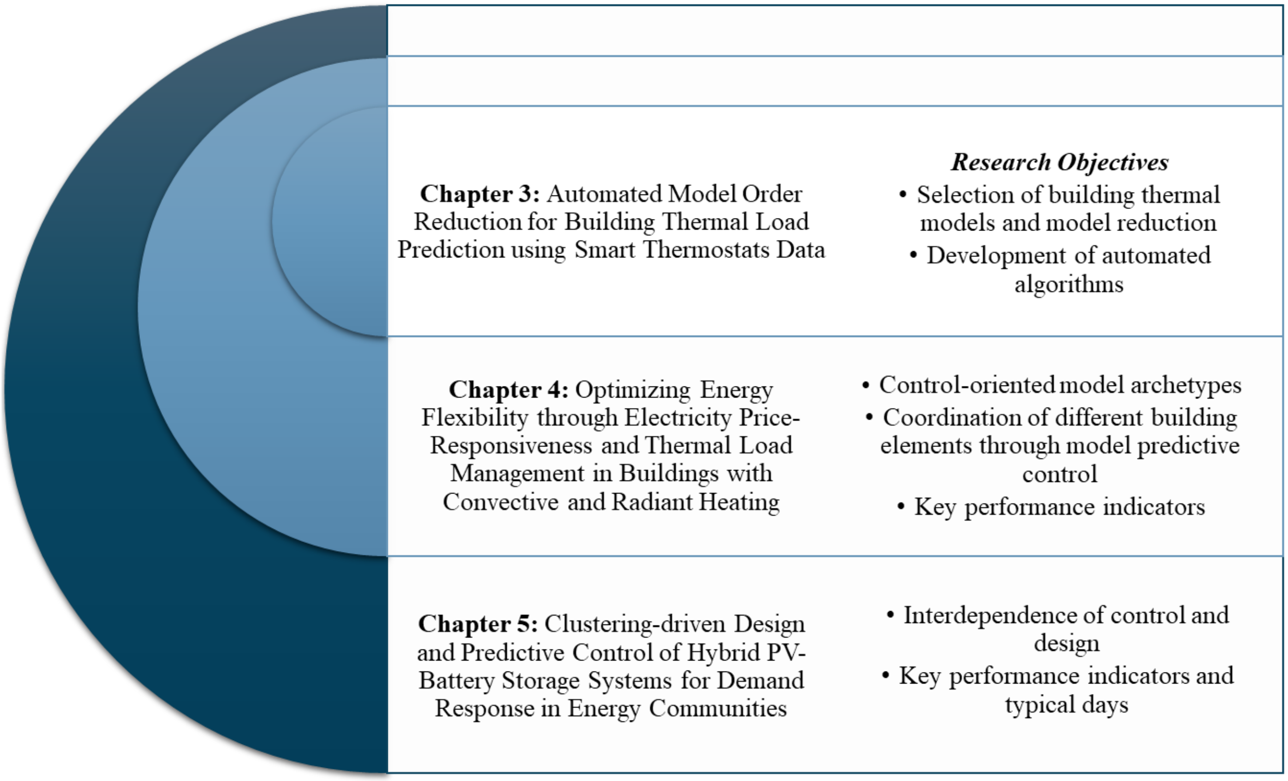


Figure 2.7: Mapping research objectives to thesis chapters.

Chapter: 3 Automated Model Order Reduction for Building Thermal Load Prediction using Smart Thermostats Data[3]

3.1 Introduction

This Chapter presents a methodology to automatically determine the structure of sufficiently accurate grey-box models for model predictive control, energy efficiency and flexibility applications in buildings. The methodology is based on model reduction and system identification techniques, with a path that enhances data pre-processing, a multistage order reduction, and parameter estimation. The model structure is determined with a cascade approach that either neglects, keeps, or aggregates thermal zones by using discrete and continuous frequency domain techniques. Once the optimal structure is identified, the parameters are calibrated with the measured data from smart thermostats, using the model predictive control relevant identification method. The methodology is applied to a monitored house located in Québec, Canada. The developed algorithm identifies adjacent zones, even when the building layout is unknown, by studying indoor temperature fluctuations. The results concerning the model creation suggest that, for this specific building, the aggregation by floor is the most efficient way for creating reduced order thermal models, limiting uncertainty due to thermal zone interaction. This methodology provides control-oriented models that accurately predict response up to 24-hours ahead with RMSE less than 0.5 °C and acceptable FIT values for the minimum number of selected parameters. Finally, several scenarios demonstrate the insights gained from using grey box building thermal models for design, control, and retrofitting applications.

The imperative to develop models for controlling and predicting building loads is evident. While many studies adopt specific RC model structures, there is a lack of justification regarding the selected model structure. Moreover, only a handful of papers propose automated methodologies for creating building thermal models using data from smart thermostats. There is a pressing need to justify the choice of model structure and order based on the number of "dominant" zones within a specific building, with the aim of providing methodologies applicable at a larger scale.

This paper introduces a methodology that bridges data acquisition from thermostats to model order determination, resulting in (i) a justification for the chosen model order in prior studies available in literature, and (ii) the development of a methodology applicable to buildings equipped with smart thermostats. This approach aims to ensure the creation of a model structure that accurately represents the building's layout, even in cases where such layout details are unknown. Subsequently, through statistical analysis, important parameters capturing key aggregated physical aspects and other relevant characteristics are identified. Hence, the paper

[3] This Chapter is based on the following journal paper: Maturo, Anthony, et al. «Automated model order reduction for building thermal load prediction using smart thermostats data». Journal of Building Engineering, vol. 96, November 2024, p. 110492. ScienceDirect, https://doi.org/10.1016/j.jobe.2024.110492.

aims to provide valuable insights into the effective design and development of reduced order building thermal models. The selected methods prioritize model simplicity and computational efficiency, which play a crucial role in expediting the model creation process. Since the order of a building thermal model is influenced by the number of thermal zones, this paper relies on frequency domain techniques to select and aggregate the most important thermal zones. These thermal zones, identified with the word "dominant zones", corresponds to the state variables of a state-space model or to the capacitances/nodes of an RC thermal network. Once the state variables and the required parameters in input that maximises the model accuracy are identified, the model, specifically an RC thermal network, is calibrated in the time domain. The developed models ensure accurate predictions up to 24-hrs ahead and are suitable for design and control studies. To test the potential of the proposed modelling approach and its applicability, a suitable case study analysis is conducted. The results show the prediction capabilities of the model and the effects of retrofitting involving the integration of photovoltaic and heat pump technologies.

The Chapter is divided in several sections. Section 3.2 focuses on the methodology, by underlying the multiple steps used to automatically produce building energy models. Section 3.3 describes the case study and how the integrated technologies are modelled. Section 3.4 shows the results of the model development. Section 3.5 describes the applications for indoor comfort and retrofitting, and Section 3.6 shows the conclusions.

3.2 Methodology

The proposed modelling methodology is divided in three axes that respectively deal with data pre-processing, selection of state variables and structure of the model, and model calibration (Figure *3.1*). The first axis deals with the processes of data treatment and selection, ensuring that the data used in the following steps are informative enough to capture the building thermal dynamic.

The second axis corresponds to the model reduction section and deals with the selection of the state variables and the structure of the model. This section relies on frequency domain identification algorithms. At the beginning, the selection of state variables (thermal zones) is assessed by ranking them according to the effect of the outdoor temperature on their thermal dynamic. This influence is evaluated by means of an index based on discrete transfer functions. This stage serves to filter thermal zones that are not necessarily affecting the building thermal load.

Subsequently, the aggregation or lump of thermal zones is conducted. This procedure is performed using an optimization routine. During this routine, transfer functions models are created for each thermal zone. These models are developed separately for each thermal zone, considering the effect of weather and the interaction with other building zones. For a specific state variable (thermal zone), the related models demonstrating high accuracy are those that either (i) capture the interaction of that specific zone with an adjacent thermal zone or (ii) account for the interaction with a thermal zone exhibiting a similar temperature pattern. The lumping of adjacent thermal zones with similar set point temperatures results in the formation of a "dominant zone," hence, this lumping procedure is also supervised by the temperature setpoint.

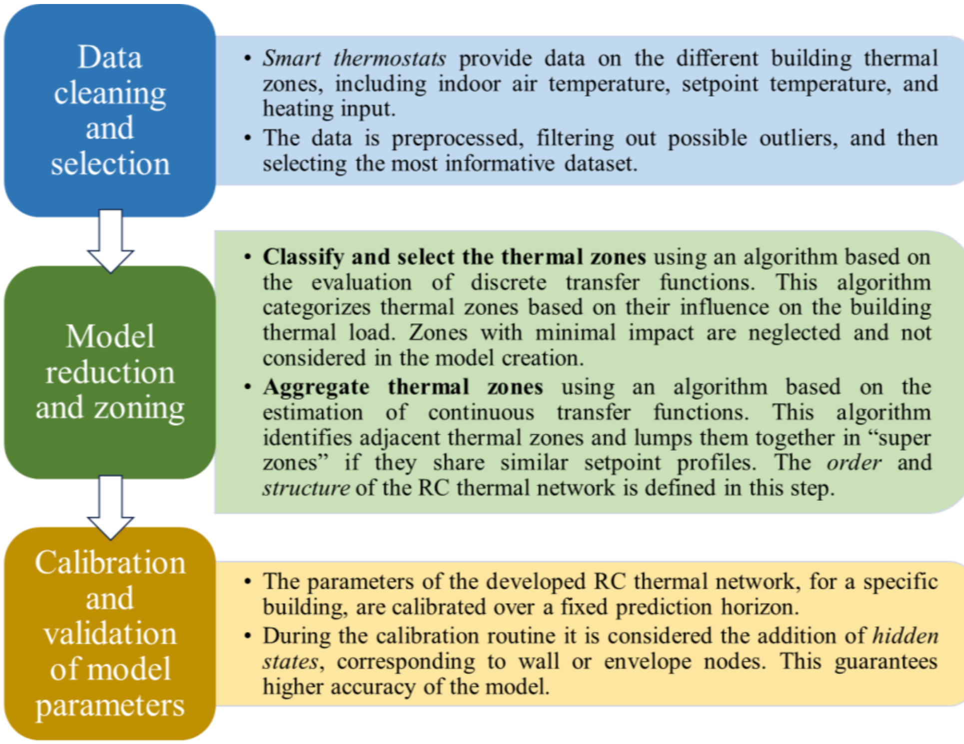


Figure 3.1: Flow chart schematic of the building modelling methodology, distinguishing data preprocessing, model order reduction and calibration.

When no more aggregation is possible, the algorithm stops and provides the final number of state variables ("dominant zones") with the necessary variables in input (e.g., outdoor temperature, solar radiation). The information on the identified state and input variables is turned into an RC thermal network structure, where the identified "dominant zones" represent its nodes. Figure 3.2 shows an example of these two steps of the methodology for a certain building characterized by ten smart thermostats and so, thermal zones. Using a simplified representation, Figure 3.2 displays (i) the classification of thermal zones according to their effect on the building thermal load. Here, a "neglect threshold" is introduced to neglect zones with null heating input and low temperature fluctuations. In this example, one zone is neglected while the other nine will become the input to the next step, which is (ii) aggregation of thermal zones. In this step, the algorithm finds the adjacent thermal zones by studying their dynamics and interactions. The algorithm iteratively studies the mutual interaction between the different zones, and aggregates or not the zones together. For this example, during the first iteration the algorithm finds the links between the zones and spots that eight zones can be aggregated in four different "dominant zones" plus the one remaining zone that will represent another "dominant zone". After the second iteration the algorithm spots the possibility to aggregate four thermal zones in two different "dominant zones" and so on. In this example, the algorithm stops at the third iteration. The identified three "dominant zones" will dictate the structure of the RC thermal network. The final number of state variables ("dominant zones") with the necessary variables in input (e.g., outdoor temperature, solar radiation) will become the input to the third axis. The third axis corresponds to the calibration routine. The parameters of the developed RC network are calibrated in the time domain over the selected period and prediction horizon by also considering the addition of hidden states (i.e., envelope nodes). This calibration process ensures that the model accurately represents the observed data and achieves the desired performance.

The following section provide more detail on the methodology adopted with all the algorithms involved. As mentioned before, the methodology for building thermal modelling is divided into three primary axes (Figure *3.1*). Each axis focuses on specific aspects, starting with data cleaning and selection (Section 2.1), followed by model reduction and zoning (Section 2.2) and model calibration and validation (Section 2.3). Finally, Section 3.3.1 is introduced to model the technologies adopted in a retrofitting scenario.

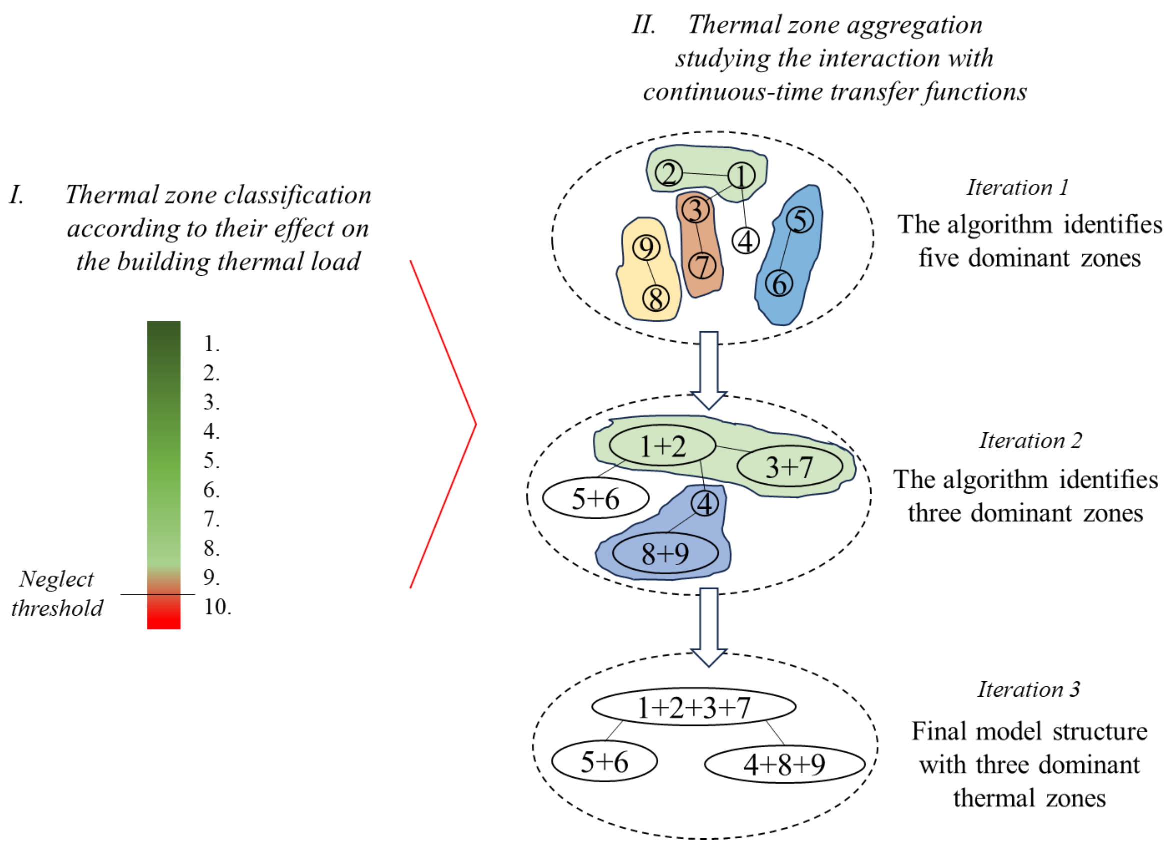


Figure 3.2: Example of the model order reduction and zoning axis using a simplified schematic.

3.2.1 Data cleaning and selection

The extracted data and its pre-processing affect the performance of the calibration algorithm to accurately capture the thermal dynamics of buildings. The data is preprocessed, filtering out possible outliers and then selecting the most informative dataset.

Incomplete data poses an inevitable challenge when working with most measured data sources. Several approaches, which generally statistically infer missing datapoints from the available ones, have been used by experts to deal with this issue. In this work the mean substitution approach is used for missing temperature measurements. This method imputes missing data values by substituting them with the mean of the previous and next measurement [232]. An index is then introduced to avoid the selection of days with excessive missing points.

Smoothing techniques are valuable for reducing fluctuations in measurements caused by unpredictable factors. They help mitigate the impact of noisy data, improve data quality, and enhance both predictive performance and model calibration [233]. This paper uses the Savitzky-Golay filter approach [98]. Savitzky and Golay demonstrated that least-squares smoothing reduces noise while preserving the shape and height of waveform peaks [234]. In this paper, this filter is applied when significant and anomalous temperature variations are observed, avoiding undue influence on the measurements.

The selection of the dataset in input has a significant impact on model calibration [98]. During experiments, the choice of input signals determines the operating point of the system and the specific parts and modes of the system dynamics that are stimulated. Therefore, it is important to carefully select data that provide sufficient information on the studied dynamics [235]. For this reason, are selected only consecutive periods of at least one week, necessary for calibrating grey-box models, with temperature data exhibiting significant fluctuations (and dynamics). To achieve so, the standard deviation, σ, of each zone temperature is computed and compared to a predetermined threshold value fixed at 0.4 [236]. Only periods with σ higher than the selected threshold are chosen. This enables the identification of periods when the heating/cooling system is operating (set-back periods) and with sufficient dynamic.

3.2.2 Model reduction and zoning

This section describes the methodology used to identify the structure of grey-box models (RC-networks) for different houses. The methodology, implemented in MatLab, is divided into two main layers, as depicted in Figure 3.3. These layers primarily involve the elimination of irrelevant state variables and then the aggregation of relevant ones. As stated in the introduction, when referring to buildings, the number of states is generally linked to the number of thermal zones [22]. The need for neglecting thermal zones is related to the high computational time for the creation of a detailed building model and the possibility to eliminate negligible loads. Consequently, the first step of the methodology aims to identify thermal zones with high temperature fluctuations and heating inputs. This helps in identifying zones that exhibit significant dynamics. Additionally, aggregating similar thermal zones allows for the creation of reduced-order models, thereby reducing the complexity and uncertainty during the model development.

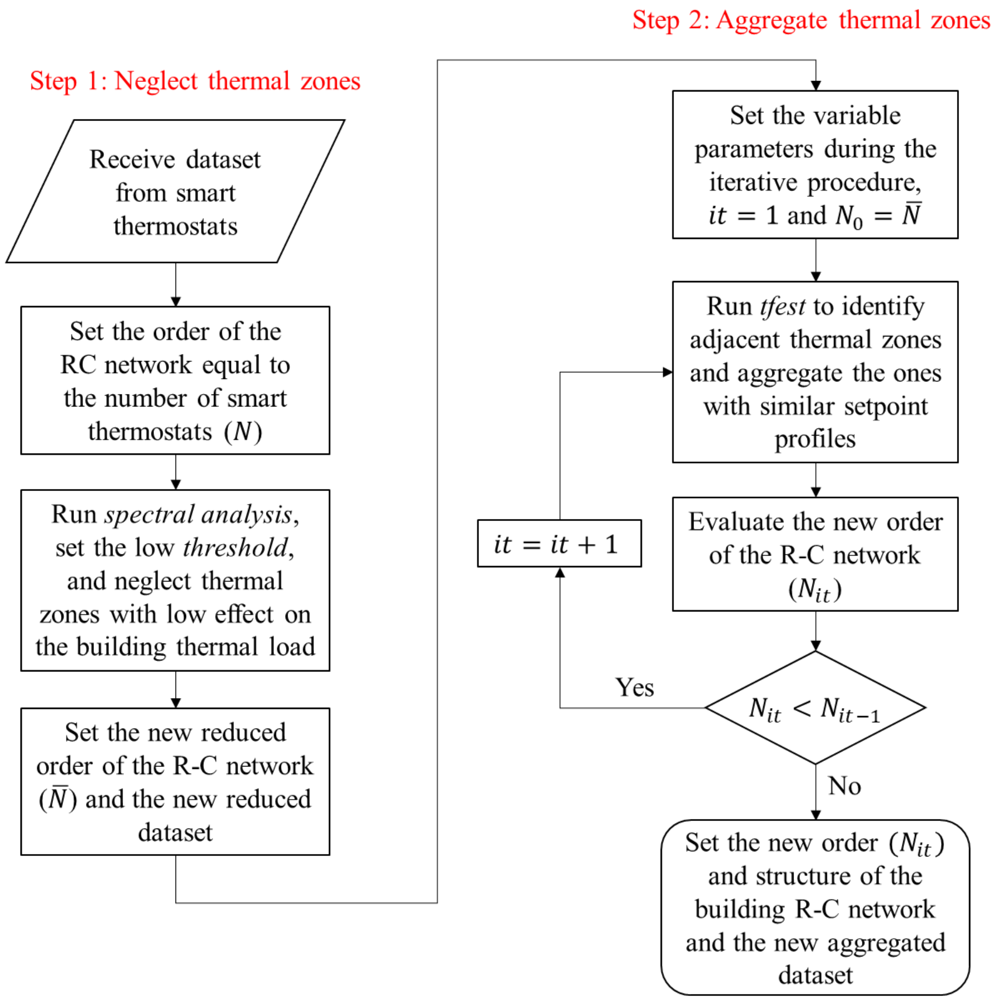


Figure 3.3: Schematic of the methodology for defining the structure of the RC building model.

3.2.3 Classify and select the thermal zones of the model

The first stage of this methodology refers to the classification and selection of the thermal zones with high effect on the building thermal load. A data-driven frequency domain approach based on spectral analysis of measured data [237] is used to decide which thermal zones can be neglected from the model creation. The algorithm defines the relationship between inputs and outputs by focusing on fixed frequency values, considering a discrete approach for faster computations. The model has the following structure:

- Model inputs $u(t)$: outdoor temperature, solar radiation, and heating input.
- Model outputs $y(t)$: indoor zone temperature.

The MatLab function *spa* is used to evaluate the transfer functions $G(q)$ of the model with a fixed frequency resolution:

$$y(t) = G(q) \cdot u(t) + v(t) \tag{3.1}$$

where q is the "shift operator", $u(t)$ and $y(t)$ are the input and output signals. $G(q)=\sum_{k=0}^{\infty} g(k)\cdot q^{-k}$, an infinite polynomial in q^{-1}, represents the transfer function of a discrete time LTI system and captures the system dynamics that take the input to the output. According to the application [71, 238], it is possible to neglect the noise parameter, $v(t)$, and Equation (3.1) becomes:

$$y(t)=G(q)\cdot u(t) \tag{3.2}$$

In detail, by considering the above-mentioned inputs and output, Equation (3.2) becomes:

$$T_{Z_i}(t)=G_{Z_i,T_{out}}(q)\cdot T_{out}(t)+G_{Z_i,S}(q)\cdot S(t)+G_{Z_i,E_{th}}(q)\cdot E_{th}(t) \tag{3.3}$$

The values of the transfer functions are estimated using the Blackman-Tuckey approach [239] based on the following equation:

$$\widehat{G}\left(e^{jw}\right)=\frac{\widehat{\phi}_{yu}(w)}{\widehat{\phi}_{u}(w)} \tag{3.4}$$

where the hat symbol represents approximate quantities and $\widehat{\phi}_{u}(w)$ and $\widehat{\phi}_{yu}(w)$ are the Fourier transforms of the covariance and the cross-variance related to the considered input and output. More details on their evaluation can be found in [239]. The value w represents the frequencies, defined as $w=\frac{2\pi n}{Period}$, where n represents the number of cycles [70]. In conclusion, with the provided inputs and outputs, the main transfer functions are defined as follows:

- $G_{Z_i,T_{out}}(n)$: transfer function for different frequencies (or also n cycles per day), representing the frequency response of the indoor temperature as a function of the outdoor temperature.
- $G_{Z_i,E_{th}}(n)$: transfer function for different frequencies, representing the frequency response of the indoor temperature as a function of the heating sources.
- $G_{Z_i,S}(n)$: transfer function for different frequencies, representing the frequency response of the indoor temperature as a function of the solar radiation.

The values of the transfer functions will vary with the number of cycles, n, distinguishing slow from fast dynamics (i.e., for convective loads, values of n over 35 cycles per day refer to fast dynamic). Moreover, with this simplified approach, each thermal zone is studied independently from the others, only considering the effects of external disturbances and heating inputs. By isolating individual zones, this first step of the methodology can discern their contributions to the building thermal load and comprehend their responses to external factors. Therefore, as the approach is data-driven and involves simplifications, various indices can be introduced to effectively present the key results obtained from the aforementioned transfer functions.

The outdoor temperature is the most influencing variable in terms of heat losses, infiltration, thermal load, and, in an indirect way, thermal comfort. This influence is further accentuated by the analytical framework employed. To compare the influence of the outdoor temperature on different thermal zones, $U_{ratio}(i)$ is introduced:

$$U_{ratio}(i) = \frac{M_i}{\max\left(M_i\right)} \qquad \text{with} \qquad M_i = \sum_{n=0}^{n^*} \left| \frac{G_{Z_i,T_{out}}(n)}{n^*} \right| \tag{3.5}$$

$U_{ratio}(i)$ assumes values between 0 and 1, and it is defined as the ratio between:

- M_i which is the magnitude of transfer function $G_{Z_i,T_{out}}(n)$, related to a certain zone, i, averaged over the selected frequency spectrum $\left[0, n^*\right]$, and
- $\max\left(M_i\right)$ which is the maximum among the values M_i. This indicates the value of the thermal zone subjected to the highest influence of the outdoor temperature.

$U_{ratio}(i)$ values approach 0 when the outdoor temperature has a lower impact on the temperature trend and fluctuations. Opposite case for values closer to 1. In this way, it is possible to classify the thermal zones according to the effect of the outdoor temperature, defining the ones with dominant effects on the building thermal dynamic and load.

For this reason, once the values of $U_{ratio}(i)$ are evaluated for each thermal zone i, the following criterion is established to identify and neglect the zones with a high percentage of heating input requested equal to a low value. The thermal zones with:

$$U_{ratio}(i) < mean\left(U_{ratio}\right) - std\left(U_{ratio}\right) \tag{3.6}$$

are then neglected from the modelling procedure, where $mean$ and std refer to the mean value and the standard deviation. The chosen threshold identifies with acceptable accuracy the zones with low influence on the whole thermal energy load of the studied building.

3.2.4 Aggregate thermal zones

In this section, the primary focus is on identifying the adjacent thermal zones, ultimately defining the structure of the model. The goal is to determine the relationships and dependencies between different components and subsystems within the building. The methodology, implemented in MatLab, is based on the transfer function estimation method, which estimates continuous-time transfer functions using time-domain data [240]. The process considers input data in time domain and generates the following model:

$$\begin{cases} y_u(t_k) = G(s,\vartheta) \cdot u(t_k) \\ y(t_k) = y_u(t_k) + v(t_k) \end{cases} \tag{3.7}$$

where $y(t_k)$ denotes the sample of the continuous-time signal $y(t)$ at time-instant $t_{ks} = k \cdot t_s$, with t_s the sampling time and s the differential operator $\left(s \cdot x(t) = {}^{dx(t)}\!/_{dt}\right)$, while $u(t_k)$ denotes the input signal including weather and zone interaction.

$G(s,\vartheta)$ is the plant model transfer function given by Equation (3.8):

$$G(s,\vartheta)=\frac{B(s)}{A(s)}=\frac{b_0+b_1s+...+b_zs^z}{a_0+a_1s+...+a_ps^p} \tag{3.8}$$

where z is the number of zeros, p is the number of poles $(p \geq z)$, and $\vartheta = \left[a_{p-1},\ ...\ ,\ a_0\ b_z,\ ...\ ,\ b_0\right]^T$ [239]. The algorithm uses direct methods to evaluate the parameters of the transfer function and allows the selection of the variables p and z. The transfer function estimation process is divided into two stages: (i) the primary stage which consists of finding a preprocessing method to generate some measures of the process signals and their time derivatives, and (ii) the secondary stage or estimation stage in which the continuous time parameters are estimated. The parameters are estimated using the instrumental variable (IV) methods [241]. More details on the transfer function estimation process can be found in [240].

Table 3.1 explains the proposed iterative approach to obtain the simplest acceptable model for each thermal zone. During the process, various models are generated based on selected inputs and state variables, such as weather and thermal zone interactions (from line 5 to 9 in Table 3.1). The accuracy of the models is evaluated using the "Normalized Root Mean Square Error" (NRMSE) (line 10 in Table 3.1). Especially in transfer function models, this accuracy can be influenced by the similar temperature trend of the thermal zones: it may happen that the developed models have good fit because the temperature trends of the thermal zones are similar. For this reason, an aggregation routine is necessary. The parameter that is used to supervise the aggregation of different thermal zones is the setpoint temperature. So, once the algorithm finds a big correlation by comparing the transfer function models developed for each thermal zone, a possible aggregation is examined by looking at their setpoint profiles difference.

For example, the algorithm finds out that the best models developed for the thermal zone i and j, described by the set $\{T_i, E_{th,i}\}$ and $\{T_j, E_{th,j}\}$ respectively, are the ones that consider their mutual interaction. So, the most accurate transfer function model developed for the thermal zone i has the thermal zone j as input, and vice-versa (lines 13 and 14 in Table 3.1). Subsequently, by looking at the difference of setpoint between the two thermal zones, $\Delta T_{set}(i,j)$, it is stated if these two zones can be aggregated (line 15 in Table 3.1). This $\Delta T_{set}(i,j)$ is compared to a fixed threshold value, $\Delta\tilde{T}_{set}$. If a condition is met, a "dominant zone" is generated by considering a temperature and thermal energy evaluated in Equation (3.9), otherwise, the correlation between the two thermal zones is kept without aggregation.

Equation (3.9) provides a schematic of the formulation behind the aggregation process and highlights the new temperature value and the heating input. The temperature of this new zone is generated by averaging the considered zone temperatures with their maximum heating installed capacity. The size of the heating terminal, and thus, indirectly, the size of a specific thermal zone, are involved in the formulation behind the aggregation procedure. With this approach, during the aggregation procedure and model creation, importance is always given to the zones with higher energy consumption.

$$\{T_i, E_{th,i}\} + \{T_j, E_{th,j}\} = \begin{cases} \left\{ \dfrac{T_i \cdot E_{th,i,\max} + T_j \cdot E_{th,j,\max}}{E_{th,i,\max} + E_{th,j,\max}}, E_{th,i} + E_{th,j} \right\} & if\ \Delta T_{set}(i,j) < \Delta \tilde{T}_{set} \\ \text{no aggregation} & if\ \Delta T_{set}(i,j) > \Delta \tilde{T}_{set} \end{cases} \quad (3.9)$$

The algorithm is executed iteratively until no more aggregation is possible $(N = N_0)$. A more detailed description of this process is reported in Table 3.1. It is important to notice that different values of $\Delta \tilde{T}_{set}$ can lead to models with different level of aggregation and so, different complexity. The results after this aggregation routine are (i) the model order with the number of "dominant zones", and (ii) the set of parameters corresponding to the structure of the RC thermal network to describe the thermal dynamic of the building.

Table 3.1: Algorithm for aggregation of thermal zones and extraction of the best set of parameters to model the building thermal dynamic.

1.	Define initial model order N and data after the spectral analysis
2.	Initialize $N_0 = N + 1$
3.	**while** $N < N_0$
4.	$N_0 = N$
5.	**for** $i = 1:N$ **do**:
6.	Initialize the set of input variables describing the thermal zone i: $\{T_i, E_{th,i}\}$
7.	**for** $j = 1:N$ with $j \neq i$ **do**:
8.	Initialize the set of input variables: $\{T_i, E_{th,i}, T_j\}$
9.	Run the transfer function estimation algorithm under different weather parameters
10.	Compare the transfer function models using the NRMSE
11.	Select the best set of parameters that maximises the accuracy for the transfer function model developed for every thermal zone
12.	**for** $i = 1:N$
13.	**if** the best set of parameters for thermal zone i is $\{T_i, E_{th,i}, T_j\}$
14.	**if** the best set of parameters for thermal zone j is $\{T_j, E_{th,j}, T_i\}$
15.	**if** $\Delta T_{set}(i,j) < \Delta \tilde{T}_{set}$ **do**:
16.	Evaluate new T_i, corresponding to a new "dominant zone" using Equation (3.9) and remove T_j from the dataset
17.	Set the new order: $N = N - 1$
18.	**Output**: Model order and number of "dominant zones". Set of parameters that describe the thermal dynamic of the building

3.2.5 Calibration and validation of the RC network parameters

In this section, the calibration process of the identified parameters of the RC model is explained. The aim is to fine-tune and optimize these parameters to ensure that the model accurately represents the observed data and exhibits desired performance.

The result of the previous steps is the model structure with the considered state variables or "dominant zones", the mutual influence between these zones, and the other input variables (e.g. outdoor temperature and solar radiation influence). Therefore, with this approach the air and mass capacitances of the identified thermal zones are generally lumped on the same node which establish the order of the model. The development of very low order models can require the addition of an extra node, called hidden state, to increase the prediction performance. During the calibration process, the potential addition of an extra capacitance to a specific thermal zone, corresponding to a "zone envelope node", is taken into consideration, ultimately increasing the order of the model. During this procedure, several models with different numbers of "zone envelope mass" nodes are developed. In the end, only the one which best describes the thermal dynamic of the building is selected. This procedure necessitates periods of warm-up before starting the calibration routine to evaluate the initial conditions of these hidden states, causing a possible increase in the computational cost.

Given the model structure, it is possible to write the system in a state-space formulation.

$$T^{t+1} = A \cdot T^t + B \cdot U^t \tag{3.10}$$

The hyperparameters of the model to be evaluated are the ones mentioned in the following expression.

$$\begin{bmatrix} T_1^{t+1} \\ T_2^{t+1} \\ \dots \\ T_{\tilde{N}}^{t+1} \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,N} \\ a_{2,1} & a_{2,2} & \dots & a_{2,N} \\ \dots & \dots & \dots & \dots \\ a_{\tilde{N},1} & a_{\tilde{N},2} & \dots & a_{\tilde{N},\tilde{N}} \end{bmatrix} \cdot \begin{bmatrix} T_1^t \\ T_2^t \\ \dots \\ T_{\tilde{N}}^t \end{bmatrix} + \begin{bmatrix} b_{1,1} & b_{1,2} & b_{1,3} & 0 & \dots & 0 \\ b_{2,1} & b_{2,2} & 0 & b_{2,4} & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & 0 \\ b_{\tilde{N},1} & b_{\tilde{N},2} & 0 & \dots & 0 & b_{\tilde{N},\tilde{N}+2} \end{bmatrix} \cdot \begin{bmatrix} T_{out}^t \\ S_t \\ E_{th,1}^t \\ E_{th,2}^t \\ \dots \\ E_{th,\tilde{N}}^t \end{bmatrix} \tag{3.11}$$

where $\tilde{N} = N + N_a$, with N corresponding to the number of "dominant zones", and N_a the number of additional capacitances related to the hidden states. The parameters are subjected to the following constraints to enforce stability [242], and give physical meaning to the model:

- $\det(A) \neq 0$ and value of the condition number of matrix A under a bound defined by Merikoski et al. [243].
- Controllability constraint on the state space matrices.
- $sum\left(\sum_j a_{i,j} + b_{i,1} - 1\right) = 0$ for $\forall i$, subjected to the issue of the critical time step in the explicit discretization method, giving numerical stability to the model [244].

- $a_{i,i} > 0$ for $\forall i$, directly related to the thermal balance equations on each node. The coefficients represent combination of thermal resistance and capacitance values.

The calibration routine is executed by using a MatLab function, *fmincon*, a non-linear programming solver that finds the minimum of a specified function by considering a set of constraints. The algorithm is highly influenced by the initial condition; thus, the optimization process is combined with a constrained Latin Hypercube Sampling [245]. The selection of the building model and the identification of the parameters is also affected by the end use of the model. If the purpose is predictive control, the performance of the model over the prediction horizon must be included in the calibration function. In this paper, the Model Predictive Control Relevant Identification (MRI) is adopted to calibrate the model parameters [246].

The "Root Mean Square Error" (RMSE) represents a good evaluation criterion to define the fit of a model from measured data.

$$RMSE = \sqrt{\sum_{i=1}^{M} \frac{\left(y_i - \hat{y}_i\right)^2}{M}} \tag{3.12}$$

where $\hat{y}_i$ is the estimated output, y_i is the measured output and M is the number of observations. The "Fit function" (FIT) represents a good metric to define the prediction capabilities of a model [246]. This index is generated from the knowledge of the normalized RMSE.

$$NRMSE = \frac{RMSE}{\sqrt{\sum_{i=1}^{M} \frac{\left(y_i - \bar{y}_i\right)^2}{M}}} \tag{3.13}$$

where $\bar{y}_i$ is the mean value of the measured output. The FIT index is instead evaluated as follows:

$$FIT = \left(1 - NRMSE\right) \cdot 100\% \tag{3.14}$$

The best models are selected by considering the performance in terms of predictions using the FIT and RMSE indices.

Table 3.2: Different results taken from literature for calibrated thermal models.

	Guideline	*Hourly Calibration* **RMSE [°C]**	*Calibration with 6-hrs prediction horizon* **FIT [%]**
Based on indoor temperature	Baba et al. [247] Zhan et al. [248]	< 1	-
	Privara et al.[246]	-	> 40

3.3 Case study

The proposed methodology is applied to a house located in Québec (Canada), Figure 3.4, whose dataset for the calibration of the model is provided by Hydro-Québec. The information on the indoor temperatures for each thermal zone of the building are obtained from smart thermostats. The data include information from smart thermostats and measurements of actual temperature, record of setpoint temperature and heating energy consumption from each heating terminal. The data from the thermostats is sampled at a specific time interval, which is typically set to 15 minutes. This sampling time allows for consistent measurements of the indoor temperature dynamics within each thermal zone. The weather data is instead provided by the SIMEB tool [249]. It is important to note that there is no knowledge of the sketches and dimension of the building. Only information on the thermostats' attributes and location is available. Table 3.3 gives the general information related to the house: number of thermostats (corresponding to the initial number of thermal zones), type of heating system and number of floors. The heating is provided by baseboard heaters.

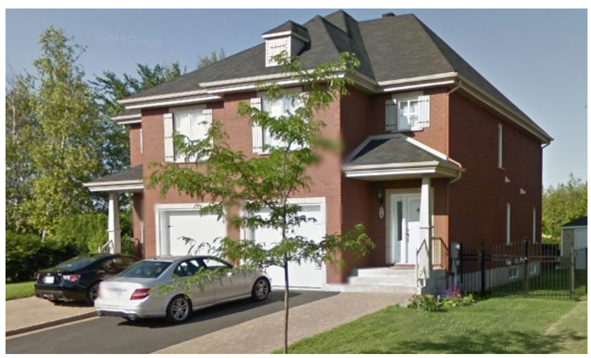


Figure 3.4: Picture of a typical two-storey Canadian house including a basement (Source: Google Maps).

Table 3.3: General information on the house data shown in the results.

Location	**Building type**	**Heating terminals**	**Number of Thermostats**
Bécancour, Québec	Split: 2 floors and semi-basement	Baseboard heaters	9

Here are also defined the hyperparameters fixed during the model reduction process. Table 3.4 provides: (i) the frequency range used for the spectral analysis, fixed between 0 and 48 cycles per day, and (ii) the number of poles, p, and number of zeros, z, which are related to the order of the transfer functions. It has been demonstrated that the best model is not always the one with the highest number of poles [69]. In this paper, the values are fixed to $z = 1$ and $p = 2$. based on the following considerations:

- During the model order reduction routine (Section 3.2.2), each thermal zone is modelled with transfer functions. In the Laplace Domain, the thermal dynamic of thermal zone nodes (dynamic of the air control volume) is assessed by using this transformation:

$C_{air} \cdot {dT_{air}}/{dt} \approx sC_{air} \cdot T_{air}$ [70]. Assuming the number of poles $p = 2$ means considering, on each thermal zone, the presence of two capacitances/poles (i.e. generally one light mass capacitance linked to the indoor air, and one heavy mass capacitance linked to the furniture and envelope). These two capacitances define a short and long thermal dynamic respectively, which is like considering two different time constants, τ [250].

- Furthermore, higher numbers of z and p can lead to gradient problems and overfitting when calibrating the transfer functions. The selected values of z and p represent a good compromise.

Moreover, as stated in Section 2.2.2, different values of $\Delta\tilde{T}_{set}$ can lead to models with different level of aggregation and so, different complexity. Hence, two values of $\Delta\tilde{T}_{set}$ are fixed to provide models with different detail on different thermal zones. For clarity, the concepts of *Low aggregation* and *High aggregation* will be associated to the $\Delta\tilde{T}_{set}$ values of 1 and 1.5 respectively. Several thermal models will be calibrated with prediction of 6, 12 and 24-hrs ahead.

Table 3.4: Hyperparameters to fix for the thermal modelling methodology.

Hyperparameters	***Values***
discrete frequency range in number of cycles, n	[0,48]
z, number of zeros	1
p, number of poles	2
$\Delta\tilde{T}_{set}$, setpoint threshold [K]	1, 1.5
ph, prediction horizon [h]	6, 12, 24

3.3.1 Retrofitting scenario: heat pump and photovoltaic panel

The developed model enables the consideration of convective heating effects. Modelling convective heating, whether from baseboard heaters or heat pumps, does not necessitate excessive complexity, as there is generally little delay between the heating input and the variation of the zone air temperature.

The presence of components with variable efficiency, according to several affecting variables, requires a performance curve. While for the baseboard heaters the conversion from electrical input to heating happens with a conversion factor around one, the coefficient of performance (COP) of an air source heat pump (ASHP) requires a proper evaluation. The capacity and COP of the ASHP are influenced by the ambient temperature and the inlet evaporator temperature. Thus, the formulas used in [251-253] are used to express the heat pump capacity and the COP performance curve. This will provide a model of the COP considering the effect of the outdoor temperature and the variable demand through a partial load factor (PLF).

A photovoltaic field is considered in this study. This technology is integrated in the building model by using the software PVLIB which is an open source photovoltaic performance modelling functions for MatLab and Python [254]. PVLIB provides a set of open-source modelling functions that enable users to accurately simulate PV system performance. These functions rely on input data such as location details (e.g., diffuse and beam solar radiation, wind speed, tilt angle, solar angles) and enable users to choose performance curves of PV panels from a database. The functions assess photovoltaic cell and module temperatures and, by factoring in incident solar radiation on the photovoltaic panel at a given tilt angle, compute the electricity output power of individual PV modules.

In Table 3.5 are provided the design variables used for the retrofitting scenario. The results will be compared with the reference baseline scenario. The retrofitting scenario will consider the utilization of a heat pump and photovoltaic panels which will substitute the actual heating terminals. A two-stage air-source heat pump [252] is considered to supply thermal energy to the building. The COP and PLR of the technology are evaluated according to references in literature and the catalogue. The features of the photovoltaic panels are chosen from a database accessible within the PVLIB library. The size of the PV array is set at 10 kW_p, which represents the maximum allowable size for residential applications in Québec. This size is typically feasible because there is generally around 50 m^2 of available building surface area suitable for PV installation with relatively optimal orientation.

Table 3.5: Properties of the system components.

	Parameters	***Values***
Heat pump: two-stage variable capacity air source heat pump [255]	1^{st} stage	6 kW_{th}
	2^{nd} stage	12 kW_{th}
	COP	Refs [251-253]
	T_{in} evaporator	T_{amb}
	T_{out} condenser	30 °C
	T_{limit} evaporator	-24 °C
Photovoltaic panels	Tilt angle, β	30°
	Orientation	South
	Type	Canadian Solar CS5P-220M
	$E_{el,p}$	10 kW_p

3.4 Results: thermal model development

This section shows the result of the automated thermal modelling routine. The methodology provides RC network models with prediction up to 24-hrs. The selection of these prediction horizons is a consequence of the periodic dynamic of buildings, which see the weather forecasts as their main affecting variable. The dataset used for calibration is selected and filtered, the model reduction methodology is then applied and provides different RC models for the chosen hyperparameters: distinguishing *Low aggregation* and *High aggregation* according to the values of $\Delta\tilde{T}_{set}$ equal to $1\,K$ and $1.5\,K$ respectively.

Figure 3.5 shows, for the considered case study, and for $\Delta\tilde{T}_{set} = 1\,K$ on the left and $\Delta\tilde{T}_{set} = 1.5\,K$ on the right: (a) the aggregation of thermal zones after every iteration, and (b) the number of parameters of each model which diminishes after every iteration. The legend provides information on the thermal zones lumped, independent and neglected after every iteration. In detail, the algorithm starts by neglecting thermal zones using the methodology described in

Section 3.2.3. Here, the thermal zones characterized by low temperature fluctuations and low energy request are neglected from the model creation. For this case study only one zone is neglected, corresponding to a basement zone (Basement Play-Room). After this procedure, the aggregation routine starts by following the process described in Table 3.1.

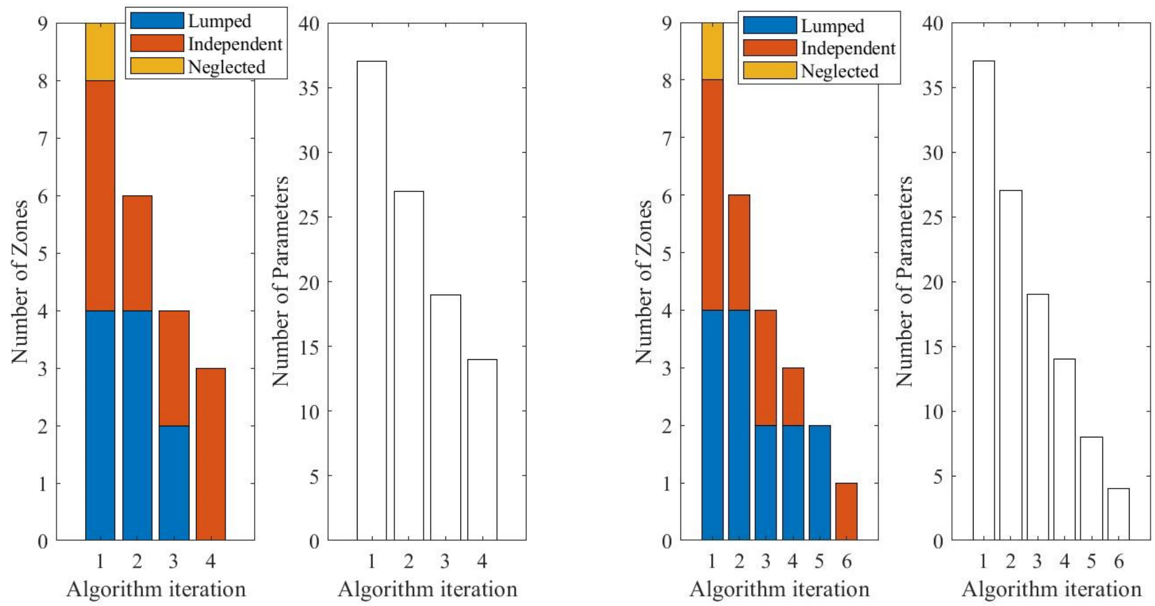


Figure 3.5: Results of the model reduction algorithm showing the number of thermal zones and parameters after every iteration for $\Delta\tilde{T}_{set} = 1\ K$ on the left and $\Delta\tilde{T}_{set} = 1.5\ K$ on the right.

The algorithm aggregates thermal zones only if three conditions are met (lines 13, 14 and 15 in Table 3.1) and provides, as a final step, the "dominant zones" of the considered building. As stated in the methodology, the aggregation is supervised by the setpoint temperature. So, once the algorithm finds a big correlation by comparing the transfer function models developed for each thermal zone, a possible aggregation is examined by looking at their setpoint profiles difference. During the procedure, the algorithm aggregates thermal zones with similar patterns and occupancy preferences (setpoints), and the algorithm stops when no more aggregation is possible. The algorithm generally stops when the value of ΔT_{set}, for each zone i and j, is higher than the threshold limit, $\Delta\tilde{T}_{set}$. Hence, when no more aggregation is possible, the algorithm stops and provides the final number of "dominant zones" with the required input variables: the RC thermal network structure. For the considered case study, the algorithm stops after 4 iterations for the case of $\Delta\tilde{T}_{set} = 1\ K$, and after 6 iterations for $\Delta\tilde{T}_{set} = 1.5\ K$. Table 3.6 and Figure 3.6 show the thermostat attributes and the building schematic for the aggregated thermal zones respectively. These highlight for $\Delta\tilde{T}_{set} = 1.5\ K$, one main zone, and for $\Delta\tilde{T}_{set} = 1\ K$ three main zones which corresponds to:

- 1st zone, which includes some basement zones,
- 2nd zone, which includes most of the basement and first floor thermal zones,
- 3rd zone, which includes all the second-floor thermal zones.

This aggregation can be identified as "floor aggregation" [100].

Table 3.6: Identified thermal zones with $\Delta\tilde{T}_{set} = 1\ K$ and $\Delta\tilde{T}_{set} = 1.5\ K$ showing the detailed attributes of the thermal zones of the case study.

<table>
<tr><th colspan="2">Identified "Dominant Zones"</th><th colspan="2">Building thermostat attributes</th></tr>
<tr><th>Low aggregation
$\Delta\tilde{T}_{set} = 1\ K$</th><th>High aggregation
$\Delta\tilde{T}_{set} = 1.5\ K$</th><th>Location</th><th>Thermal Zones</th></tr>
<tr><td>Neglected</td><td>Neglected</td><td>Basement</td><td>1. Basement Play-Room</td></tr>
<tr><td>Zone 1</td><td rowspan="8">Zone 1</td><td>Basement</td><td>2. Basement Cinema</td></tr>
<tr><td rowspan="4">Zone 2</td><td>Basement</td><td>3. Basement Living</td></tr>
<tr><td>1st floor</td><td>4. Main Living</td></tr>
<tr><td>1st floor</td><td>5. Main Kitchen</td></tr>
<tr><td>1st floor</td><td>6. Bathroom</td></tr>
<tr><td rowspan="3">Zone 3</td><td>2nd floor</td><td>7. Bedroom 1</td></tr>
<tr><td>2nd floor</td><td>8. Master Bedroom</td></tr>
<tr><td>2nd floor</td><td>9. Upstairs Bathroom</td></tr>
</table>

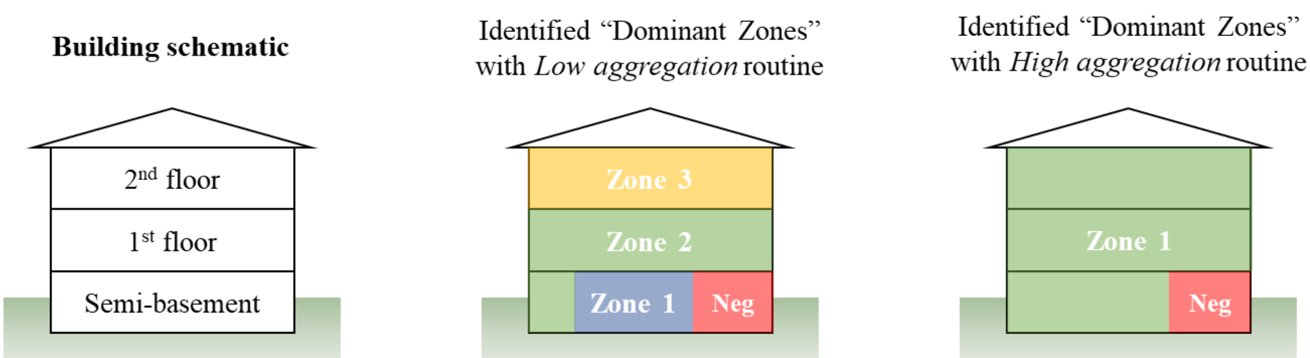


Figure 3.6: Building floor schematic representing the identified thermal zones after *Low aggregation* and *High aggregation* routine respectively.

The aggregation happens between adjacent zones or zones with same temperature fluctuations. Since the aggregation procedure is supervised by the setpoint difference, the algorithm may provide for the same building different RC thermal models according to the selected values of $\Delta\tilde{T}_{set}$. The comfort preferences are the main influencing variables of this procedure given that zones with same temperature setpoints are, as expected, more likely to be aggregated. For this reason, using different $\Delta\tilde{T}_{set}$ is important to guide the aggregation or lumping procedure. Hence, the algorithm is able to identify and lump adjacent thermal zones without any information on the building layout. Instead, it only receives as input the weather forecasts, indoor temperatures, heating input and set point profiles.

For this case study, the order of the model is reduced to few "dominant zones". This very low order may require the addition of hidden states: adding to the temperature node of a specific thermal zone another (capacitance) node which corresponds to the envelope mass. As described in the methodology, the next step of the RC model creation considers the addition of these hidden states on the model only if (differently from the up-to-date literature [256, 257]) they increase its accuracy significantly. This comparison is made by considering, as a reference, the RC model without any additional node. The proposed RC networks for the considered case study, which maximize the prediction accuracy, are reported in Figure 3.7 and Figure 3.8. A physical correlation between the identified RC networks and the building is provided in Figure 3.9 which shows the correlation between the thermal zone node and the air plus furniture, and between the zone envelope mass and the envelope of that specific zone. These figures show the identified "dominant zones", the additional "zone envelope mass" nodes, needed to further increase the accuracy of the model, and the input parameters. For this case study, the addition of the "zone envelope mass" node is required only for one "dominant zone" in case of both *Low aggregation* and *High aggregation* scenarios. The model reduction algorithm proposes, in case of *Low aggregation*, an RC network where the "dominant zones" 1 and 3 require the heating consumption and outdoor temperature as the only input, while the "dominant zone" 2 requires as inputs the outdoor temperature, solar radiation and the heating consumption. Hence, for the zones 1 and 3, the effect of the weather is mainly related to the influence of the outdoor temperature. Instead, for *High aggregation*, where only one "dominant zone" is identified, the values of outdoor temperature, solar radiation, and the heating consumption of the whole building are required as input.

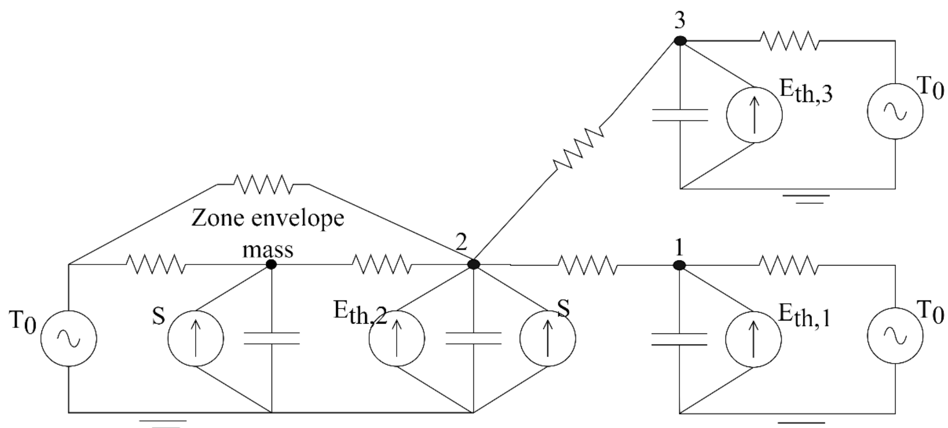


Figure 3.7: RC thermal network under *Low aggregation,* $\Delta\tilde{T}_{set} = 1\,K$, routine where 1, 2 and 3 are the thermal air zones (fast dynamics) and the "zone envelope mass" node is the zone envelope massive node related to zone 2 (slow dynamics).

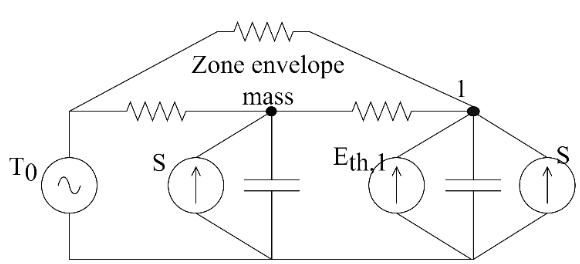


Figure 3.8: RC thermal network under *High aggregation,* $\Delta\tilde{T}_{set} = 1.5\,K$, routine where 1 is the thermal air zone (fast dynamic) and the "zone envelope mass" node is the zone envelope massive node of the zone 1 (slow dynamic).

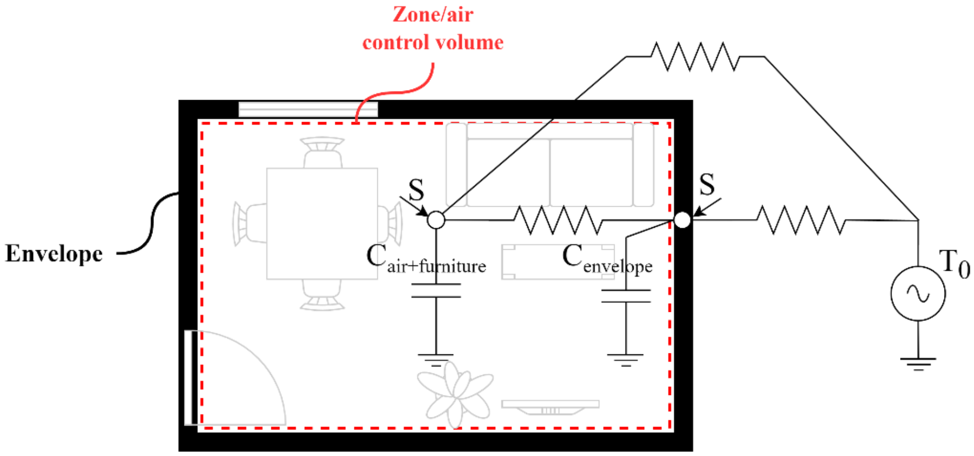


Figure 3.9: Physical meaning of the RC thermal network developed under *High aggregation*.

Table 3.7 shows the results of the calibration methodology and the performance of the developed models in terms of FIT index and RMSE, linked to the prediction capabilities of the models for 6, 12 and 24-hrs of prediction. According to the available literature, the calculated statistical indices are adequate for the prediction of performance 6, 12 and 24-hrs ahead.

Generally, the performance of the model decreases when increasing the prediction horizon during the calibration routine. Also, the results show a slight increase in the prediction ability of the model when increasing the level of aggregation and, so, reducing the order of the RC models. There are two reasons for this result: the aggregation is executed by weighing the different thermal zones with the maximum energy requested by each zone, thus the thermal zone with the higher $E_{th,max}$ has the highest impact on the temperature evaluation. Also, the value of the FIT index is evaluated by averaging the FIT index of each thermal zone, and these values of FIT for each zone may be different. Figure 3.10 and Figure 3.11 show the measured and predicted temperatures for each zone, showing the performance of the defined models for a prediction of 12-hrs during the validation period. The measured temperatures in the cited figures are evaluated during the aggregation routine using Equation (3.9).

Table 3.7: FIT index and order of the different models generated by varying $\Delta\tilde{T}_{set}$ and the prediction horizon.

Prediction horizon	*Low aggregation (3 dominant zones)* $\Delta\tilde{T}_{set} = 1\ K$			*High aggregation (1 dominant zone)* $\Delta\tilde{T}_{set} = 1.5\ K$	
	***FIT index* [%]**	***RMSE* [°C]**		***FIT index* [%]**	***RMSE* [°C]**
6-hrs	41.48	0.44		60.91	0.38
12-hrs	41.84	0.50	≠	57.41	0.44
24-hrs	41.83	0.53		51.68	0.50

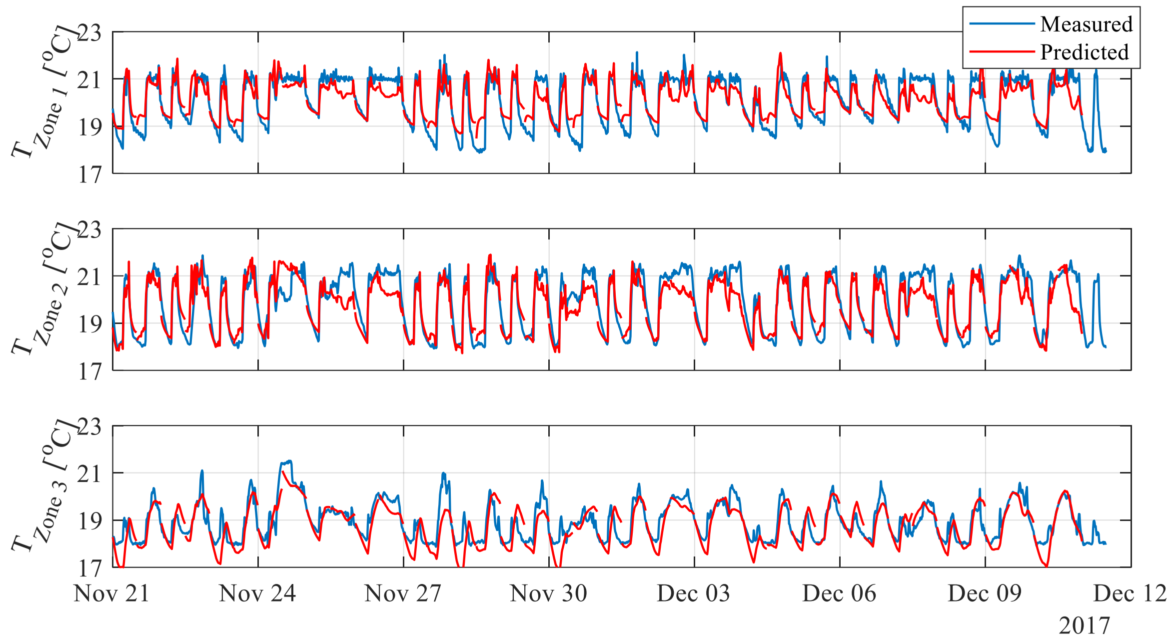


Figure 3.10: Measured and predicted zone temperatures for 12-hrs ahead in case of *Low aggregation* during the validation period: metrics of RMSE = 0.50 and FIT = 41.84.

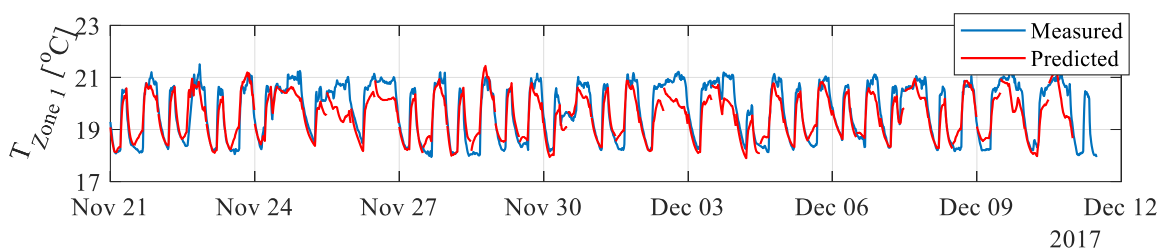


Figure 3.11: Measured and predicted zone temperatures for 12-hrs ahead in case of *High aggregation* during the validation period: metrics of RMSE = 0.44 and FIT = 57.41.

Finally, this modelling procedure showed the possibility to justify the choice of model structure and order based on the number of "dominant" zones within a specific building, with the aim of providing methodologies applicable at a larger scale. This paper introduces a methodology that bridges data acquisition from thermostats to model order determination, resulting in (i) a justification for the chosen model order in prior studies available in literature, and (ii) the development of a methodology applicable to any building equipped with smart thermostats. This approach aims to ensure the creation of a model structure that accurately represents the building's layout, even in cases where such layout details are unknown. Subsequently, through statistical analysis, important parameters capturing key aggregated physical aspects and other relevant characteristics are identified.

With this methodology, it is possible to create, automatically, different thermal models for the same building by varying the parameter $\Delta\tilde{T}_{set}$. The difference in terms of detail/complexity, related to the number of thermal zones (states or nodes), gives different advantages:

- Developing high order models, with higher number of thermal zones, allows for studying the effect of diverse control strategies (i.e., temperature setpoint variation) on different thermal zones of the building, seeing the effect on the energy consumption and energy flexibility of the building.
- Instead, a less detailed model gives an overall estimation of the performance of the building under a single strategy but reduces the uncertainty due to the thermal zone interaction which happens between each zone.

In the end, according to the obtained metrics, the models provided with this methodology are useful to predict and quantify the thermal performance of buildings, being useful for applications concerning retrofitting or demand response events. This allows the evaluation of the future demand of buildings under future price signals. The model complexity choice, as described, depends on the final end-use. For applications like retrofitting and design, where a general load understanding suffices, *High aggregation* is optimal. For more detailed control scenarios requiring deeper insights into thermal zone interactions, *Low aggregation* is preferred.

In terms of computational time, the model order reduction routine (Section 3.2.2) generates the RC thermal network of a specific building on a "minute timescale." This accelerates the model creation process, which now only requires parameter calibration. The calibration procedure, facilitated by the model predictive control relevant identification routine, depends on the prediction horizon and the number of "dominant thermal zones" in the building model (as discussed in 3.2.5). This procedure yields calibrated parameters on an "hour timescale." While the prediction horizon has less impact on result provision, the number of "dominant zones" significantly affects calibration time, increasing exponentially. This crucial aspect further supports the use of the *High aggregation* routine.

Furthermore, when considering a data-driven methodology for building thermal modelling and which is governed by a "zoning aggregation", it becomes challenging to establish a standardized "thermal model archetype" even for buildings of the same archetype (type of building sharing similar features). Since this methodology is governed by the concept of thermal zones and zoning, buildings of the same archetype may require different "thermal model archetypes" to represent their thermal response. The different utilization of the thermal zones inside each building highly affects the aggregation routine of the proposed algorithm.

It is important to note that the performance of the model is highly affected by the data input. Measurement noise or disturbances may affect respectively the actual measurement and the model accuracy. By combining the processes of data treatment with the whole reduction and calibration routine, it is possible to limit this issue. Therefore, in the future, novel smart thermostats may provide new and more accurate information that will affect positively the modelling procedures and the final accuracy.

3.5 Applications and analysis

To demonstrate the applicability of the developed energy models, two sections are introduced for describing:

- the indoor temperature trends and comfort constraints due to model predictive control, and

- the retrofit scenario which considers the change of the heating system, from electrical baseboard heaters to air source heat pumps in combination with PV. The size and properties of these technologies are reported in Table 3.5.

For simplicity, the building model resulted from the *High aggregation* routine for 24-hrs ahead of prediction is used in these scenarios. The models developed during the *Low aggregation* routine can be used to study the performance of diversified control action according to the different zone types. This will be assessed in a future study.

The results are presented for a one-month period. The control routine employed in the model predictive control scenario is based on a methodology developed in a previous study [55]. The control and prediction horizons are both fixed to 24 hours ahead. Here, the cost function prioritizes comfort and technology efficiency, with emphasis placed on minimizing the difference between the air temperature and the fixed setpoint profile set at 20°C between 7 am and 9 pm and 18°C during the night.

3.5.1 Temperature trends and comfort

The utilization of a model predictive control strategy gives the possibility to enhance the comfort condition in our system. This is shown in Figure 3.12, comparing the indoor temperature with the set point, and showing the temperature of the envelope node.

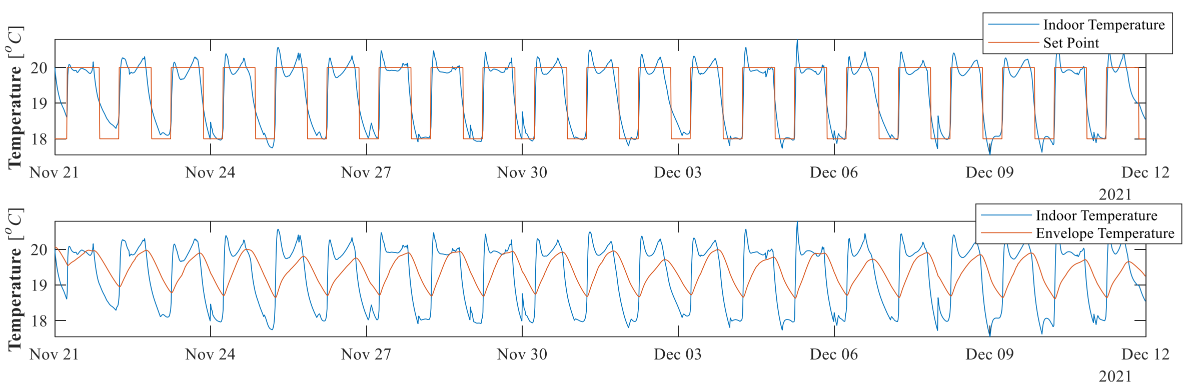


Figure 3.12: Comparison between the indoor, set point and envelope temperatures.

Therefore, the real advantages of predictive strategies can be seen in a control framework that, while considering system efficiency and indoor comfort conditions, also assesses the concept of energy flexibility through signals from the electrical grid. This is generally described as price responsiveness and will be addressed in the next Chapters.

3.5.2 Heat pump and photovoltaic refurbishment

In this sub-section are compared the energy consumption of the building equipped with baseboard heaters and the same building equipped with air source heat pumps and PV panels. The focus is mainly on the energy side, showing how the change in technology benefits the grid stress levels and consumption of the household. The simulation is done by fixing a prediction horizon equals to 24 hours ahead, corresponding to the control horizon.

Figure 3.13 shows the thermal and electrical demand of the building during the period of simulation while Figure 3.14 shows the weather forecasts for the considered period of study. Given the use of baseboard heaters, the thermal load corresponds to the electrical load required for the heating demand. The thermal demand exhibits high variation due to changes in the temperature set point of the indoor environment with peaks around the 7 am of each day. This variation directly impacts the electrical load of the building. The heating terminals used in this scenario are baseboard heaters; hence, the electrical quota required by these terminals corresponds to the thermal energy supplied to the indoor environment. This quota represents the *61.4%* of the whole electrical demand of the building.

The difference between the thermal load and the total electrical demand of the building is attributed to other loads (e.g., lighting, domestic hot water). This is evaluated using a specific daily trend generated from clustered historical data.

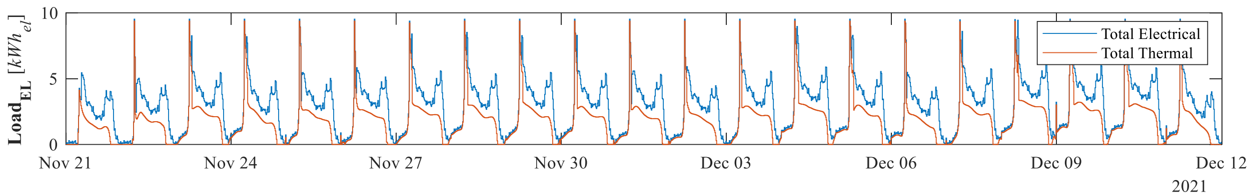


Figure 3.13: Thermal and electrical load of the reference scenario with baseboard heaters.

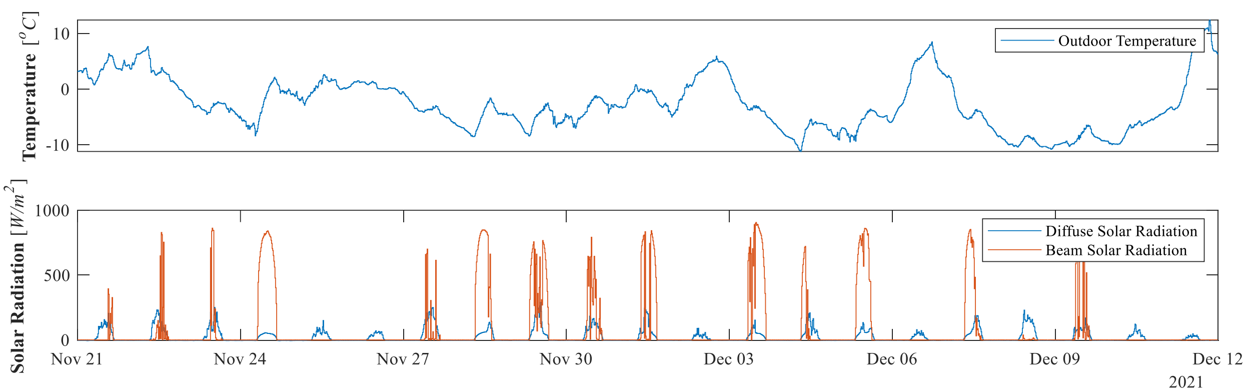


Figure 3.14: Weather forecasts for the considered period of study.

Figure 3.15 illustrates the thermal and electrical load, photovoltaic production, and net load of the refurbished building. Similar to the previous scenario, the thermal energy exhibits high peaks attributable to setpoint variations. However, the implementation of the air source heat pump mitigates the impact of the overall electrical load. In this case, the impact of other loads is comparable to the thermal energy demand of the building, which now represents *48%* of the entire electrical load. Compared to the previous scenario, there is a significant reduction in the electrical load. The adoption of only the air source heat pump results in a *25.8%* reduction in energy consumption compared to the reference scenario.

Furthermore, the production from PV panels contributes to an even greater reduction in the electrical load of the building. When comparing the net load of the refurbished scenario with the air source heat pump and PV panels, the reduction in energy consumption of the building is estimated to be *56.3%.* While *60%* of the photovoltaic electricity production is self-consumed by the building, the other *40%* can be potentially stored or exported to the grid.

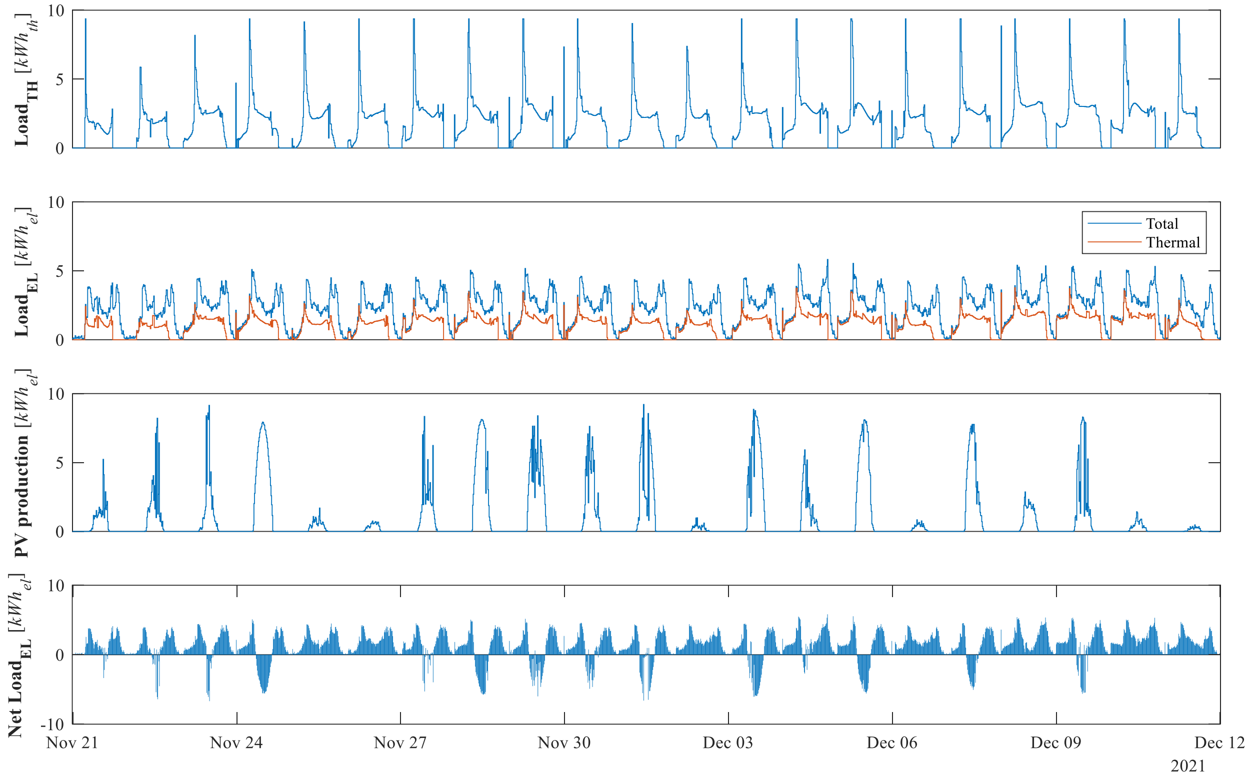


Figure 3.15: Thermal load, electrical load, photovoltaic production and net electrical load of the retrofitting scenario with air source heat pump and photovoltaic panels.

The potential to reduce electricity consumption and transition from passive consumer to producer not only offers economic benefits but also necessitates a re-evaluation of grid dynamics, prompting the introduction of bidirectionality to the grid. This shift underscores the importance of studying the interaction between buildings and the grid, particularly in terms of flexibility, by exploiting prediction of energy needs by mean of ROM. Surplus electricity generated by buildings equipped with photovoltaic systems can be either fed back into the grid or shared with neighboring buildings, effectively mitigating demand impact at the community level. However, the adoption of PV systems introduces new challenges, such as increased demand volatility, which can strain the grid during periods of low demand and high solar radiation—conditions often experienced during the summer months for the region of study. To address these challenges, integrating energy storage solutions such as batteries or utilizing electric vehicles becomes essential. These solutions not only enhance building self-consumption but also provide additional flexibility to the grid, helping to stabilize fluctuations in supply and demand.

Thus, the transition towards a more decentralized and renewable energy-based grid requires careful consideration of both technological solutions, through refurbishment studies, and grid management strategies to ensure reliability and efficiency.

3.6 Conclusion

This paper supports the automatic development of building thermal models by proposing a methodology that integrates data treatment with model reduction and creation. The need for model order reduction methodologies arises from the possibility of combining zones with similar setpoint profiles and occupant needs. This approach facilitates the creation of reduced order models for model predictive control using data from smart thermostats and building automation systems, hence, providing more compact thermal models with fewer parameters to calibrate.

The developed methodology allows for adjusting a hyperparameter to supervise the aggregation routine, enabling the development of multiple models with varying level of detail. These models can be used for different purposes, including control and design studies, with the selection of a specific model depending on its intended application. This work presents a novel methodology that combines model reduction and calibration techniques, incorporating both frequency and time domain approaches. One of the key considerations in developing this methodology is computational time, as many algorithms require extensive learning processes to create reliable models. The results provide, for the studied buildings, different RC building model archetypes characterized by a varying number of nodes/capacitances, influenced by the building type and occupancy preferences.

Accuracy of the models is achieved in accordance with other journal papers available in literature and with a minimum number of selected parameters. The developed models can accurately capture the dynamics of each building for prediction horizons of 6, 12 and 24-hrs.

To demonstrate the potential of this approach and offer insights into its specific application, one of the developed models is applied to a scenario involving model predictive control and the integration of other technologies. This showcases the consistency of the results and highlights possible applications of the developed models.

Future directions and limitations

The methodology employed in this study provides a methodology to capture the dominant building thermal dynamics by identifying the ‘dominant’ thermal zones. This identification of separate and compact thermal zones can enable more targeted and diversified control actions at the building level, enhancing the effectiveness and application of model-based predictive control algorithms in real buildings. Estimating the energy flexibility potential of each thermal zone can be challenging due to mutual interactions and uncertainties arising from indoor heat exchange. By adopting reduced-order models, these difficulties can be mitigated.

The results presented in this work pertain to a specific climate, specifically a Canadian climate, and are demonstrated using an in-depth example of a single house. Future studies will extend this methodology to a larger dataset, encompassing different types of buildings and climatic regions. This will provide comprehensive results on potential RC model archetypes and validate the developed methodology for climates other than winter-dominated ones.

Chapter: 4 Optimizing Energy Flexibility through Electricity Price-Responsiveness and Thermal Load Management in Buildings with Convective and Radiant Heating[4]

4.1 Introduction

Building energy modelling is essential for designing energy-efficient and flexible buildings that seamlessly integrate with the electrical grid. This chapter introduces a data-driven, control-oriented methodology using Resistance-Capacitance thermal network models to accurately forecast building thermal loads. It differentiates the impacts of fast and slow dynamics associated with different heating types—radiant and/or convective. A Model Predictive Control (MPC) framework optimizes coordination between the different building thermal dynamics, considering weather forecasts and price signals. The Varennes Library, a Net Zero Energy Institutional Building located in Québec (Canada), serves as a case study for performance assessment.

Validation of the developed model demonstrates its efficacy in enabling MPC to formulate effective control strategies. Findings reveal that high-mass radiant heating is strategically used before indoor setpoint variation or demand response events. Up to 70% of the building thermal load is delivered to the active envelope for off-peak heat storage and on-peak release. Conversely the ventilation heating is prioritized in proximity of the change in setpoint or grid tariff with percentages over 80%. Results show the adoption of weather clusters for generalizing the optimal control setting, highlighting their influence on thermal loads while maintaining robust ventilation and active envelope heating coordination. The comparison between the predictive control strategy and the existing rule-based control shows improvements in indoor temperature and energy flexibility. During the MPC routine, a constant price signal reduces grid stress, achieving Load Factor (LF) values up to 0.72 compared to 0.60 with rule-based control, while demand response, though critical peak pricing, optimally shifts up to 100% of the thermal load during peak price hours [258].

In detail, the analysis of the literature clearly underlines that while there has been considerable progress in optimizing ventilation and radiant heating systems independently, their integrated operation remains underexplored. The distinct thermal dynamics of these systems pose significant challenges that can be addressed through advanced predictive control algorithms. Current research has made strides in understanding and optimizing these systems in experimental

[4] This Chapter is based on the following journal paper: Maturo, Anthony, et al. «Optimizing energy flexibility through electricity price-responsiveness and thermal load management in buildings with convective and radiant heating systems». Energy and Buildings, January 2025, p. 115355. ScienceDirect, https://doi.org/10.1016/j.enbuild.2025.115355.

settings, but scalable and replicable methodologies to provide real-time or scenario-based control are lacking. This chapter aims to overcome these limits and gaps by addressing the following points:

i. Data-driven methodology to optimize the building thermal load in multi-zone buildings with radiant and/or convective heating under weather forecasts and dynamic tariffs.

ii. Clustering of weather forecasts to classify the building load profiles and generalize the optimal control settings.

iii. Provision of metrics to supervise the activation of heating terminals and to evaluate the total energy stored.

iv. Comparison of the most used energy flexibility indicators and demonstrate their advantages/limitations when applied to energy flexibility applications.

The Chapter is divided in several sections. Section 4.2 focuses on the methodology, by underlying the multiple steps used to produce building energy models, cluster weather data and optimally control the different building thermal dynamics. Section 4.3 describes the case study and Section 4.4 shows the results of the model development. Section 4.5 describes the reference and optimal thermal management. Sections 4.6 and 4.7 provide a final discussion and conclusions respectively.

4.2 Methodology

This section describes the workflow adopted for thermal energy modelling, control and performance evaluation. The modelling phase relies on Resistance-Capacitance (RC) model archetypes which enable the prediction of the building thermal load. Specifically, the selected model structure enables the study of the active building thermal mass and the coordination between the heating through ventilation and floor heating systems. The optimal control is evaluated through an MPC strategy, featuring the different time-lags of the technologies, and considering different weather forecasts and price signals. The performance of the proposed framework is evaluated through key performance indicators and the results are finally clustered with the weather data. This clustering/classification helps in generalizing the results over different weather scenarios.

A simplified schematic of the organization of this section is reported in Figure 4.1 with Section 4.2.1 describing the RC model archetype method, Section 4.2.4 the features of the MPC, Section 4.2.5 the hierarchical clustering technique applied to weather data, and Section 4.2.6 the selected key performance indicators.

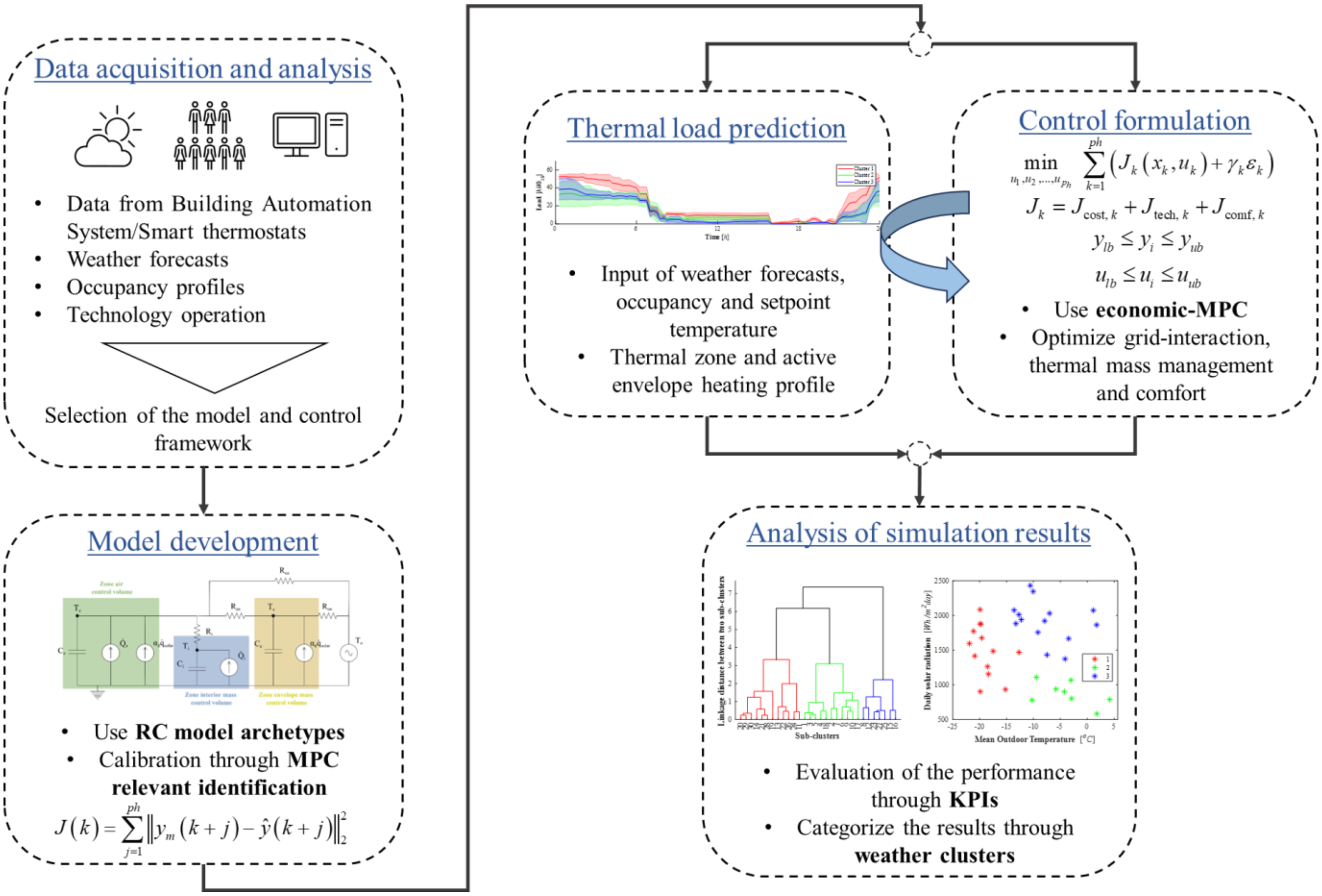


Figure 4.1: Simplified schematic of the proposed methodology.

4.2.1 Building modelling

The building thermal response is modelled using a lumped-parameter approach through RC thermal networks. The lumped-parameter approach consists of discretizing the temperature field of a thermodynamic system by identifying a certain number of representative nodes where an energy balance is computed. Each node is connected to the adjacent nodes by means of thermal resistances, and thermal capacitances are assigned to all elements that can store internal energy, such as walls and relevant volumes of air in case of buildings. These can be modelled by following a lumped approach with the indoor air mass lumped in a single uniform temperature node, also envelope elements and thermal masses (wall, roof, ceiling, floor, interior wall, window) lumped in multiple nodes. This approach is widely used by the research community involved in data-driven building energy and comfort assessment analyses, and it is more and more used for building control applications [113, 259-262]. During the modelling procedure, we distinguish three types of nodes related respectively to (i) thermal zone, identified with the subscript n, (ii) *active* envelope, identified with the subscripts n,i and connected to the thermal zone n, and (iii) *passive* envelope, identified with the subscript n,e and connected to the thermal zone n. The terms *active* and *passive* refer to whether heat can be delivered to or extracted from a specific node, respectively, with *active* always associated with an interior node and *passive* to an exterior node. The energy balance equations associated to each of these nodes are described by the following differential equations, accounting for controllable and uncontrollable inputs [39]:

$$C_n \cdot \frac{dT_n}{dt} = \sum_{n \neq m} U_{n,m}\left(T_m - T_n\right) + U_{n,i}\left(T_{n,i} - T_n\right) + U_{n,e}\left(T_{n,e} - T_n\right) + \\ + U_{n,o}\left(T_o - T_n\right) + \alpha_n \dot{q}_{\text{solar}} + \dot{Q}_{\text{gain},n} + \dot{Q}_n \tag{4.1}$$

$$C_{n,i} \cdot \frac{dT_{n,i}}{dt} = U_{n,i}\left(T_n - T_{n,i}\right) + \alpha_{n,i}\dot{q}_{\text{solar}} + \dot{Q}_{n,i} \tag{4.2}$$

$$C_{n,e} \cdot \frac{dT_{n,e}}{dt} = U_{n,e}\left(T_n - T_{n,e}\right) + U_{o,n,e}\left(T_o - T_{n,e}\right) + \alpha_{n,e}\dot{q}_{\text{solar}} \tag{4.3}$$

where T represents the temperature [K], C is the thermal capacitance [J/K], U is the thermal conductance [W/K], α is the effective solar factor, $\dot{q}_{\text{solar}}$ is the total horizontal irradiance [W/m^2], $\dot{Q}_{\text{gain},n}$ is the sensible heat gain to the indoor air node due to occupants, lights and equipment [W], $\dot{Q}_n$ is the heating/cooling input supplied to or removed from the air node i [W], $\dot{Q}_{n,i}$ is the heating/cooling input supplied to or removed from *active* envelope (or interior) node n,i [W]. The list of the major assumptions adopted for the developed lumped parameter models is reported below. Those assumptions have been proposed to simplify the complexity of the proposed models.

- The heat transfer between the different nodes is generally regarded as a one-dimensional process.
- The air within the studied building space is considered as perfectly mixed (air temperatures are uniformly and homogeneously distributed), meaning that contamination concentration correlates directly with the number of air changes occurring. This can be limiting for poorly mixed zones [263].
- The materials of the building are homogeneous with constant thermophysical properties.
- Constant and averaged heat transfer coefficients are assumed between the different building nodes. For the heat transfer between *active* envelope (i.e., floor heating) and air nodes, several papers available in literature show that with this approximation it is still possible to capture the dynamic with good and acceptable accuracy [143, 144, 264]. Therefore, by averaging the convective and radiative heat transfer, the application of the developed model is limited to a specific usage of the active envelope (i.e., heating or cooling).
- Ventilation heating, $\dot{Q}_n$, has an immediate effect on the thermal zone temperature and can be evaluated as a function of:

  $$\dot{Q}_n = \dot{m}_{vent} \cdot c_{p,air}\left(T_{supply} - T_n\right) \tag{4.4}$$

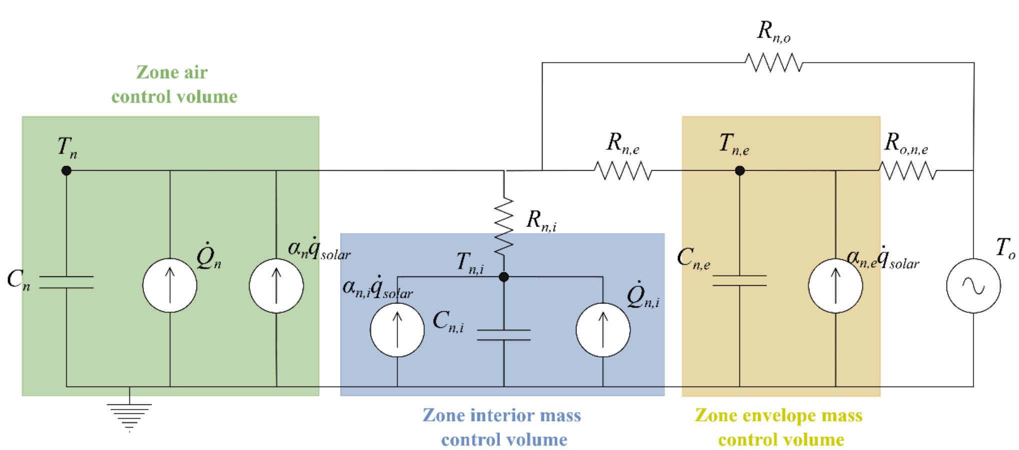
  where T_{supply} is the temperature at the delivery of the ventilation system, and $\dot{m}_{vent}$ is the air flowrate delivered to the thermal zone. This air flowrate may include a fresh-air component, $\dot{m}_{fresh}$, if the building is equipped with a dedicated outdoor Air Handling Unit (AHU). While $\dot{Q}_n$ is evaluated from Equation (4.1), the thermal quota related to the operation of outdoor air handling unit, $\dot{Q}_{AHU}$, can be evaluated as:

$$\dot{Q}_{AHU} = \dot{m}_{fresh} \cdot c_{p,air} \left(T_n - T_o \right) \cdot f \tag{4.5}$$

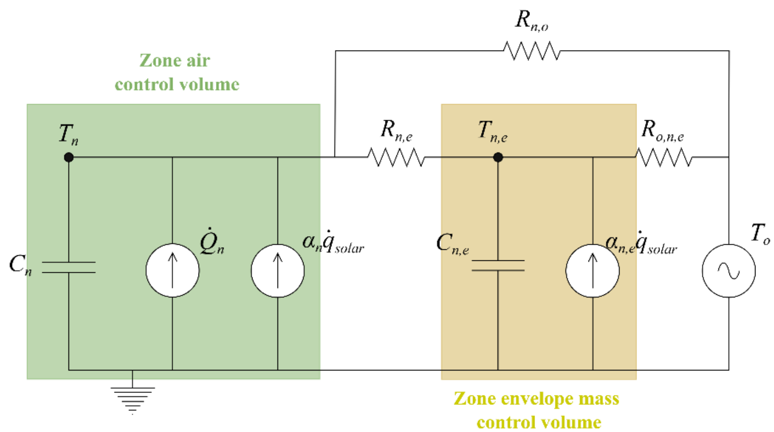
With f a factor that establishes how much energy is covered by heat recovery units or energy storage systems. The $\dot{m}_{fresh}$ component can be evaluated as a function of the building area or the variable number of occupants using the ASHRAE 62.1 guideline [265].

- Heat is not generated within the building construction elements despite for the *active* envelope node. Heat is extracted or supplied to the related node to provide cooling or heating respectively. This quota in input, which is a function of the supply water temperature and water flowrate, is lumped to one variable, $\dot{Q}_{n,i}$, received from the production unit.
- The distribution system of active envelope node is not modeled. If necessary, the energy use and energy loss associated to the distribution system could be considered in the efficiency of the production unit.

In this paper, the following RC thermal model archetypes have been adopted to capture the thermal dynamic of specific thermal zones (Figure 4.2). The dynamics of these RC thermal networks can be distinguished into (i) a fast-responding building element, which correspond to the thermal zone and includes the elements inside the air control volume (air plus furniture – node n), and (ii) slow-responding building elements of the specific thermal zone, distinguishing an interior mass node, associated in this study with an *active* envelope node (node n,i) and the exterior walls node or *passive* envelope node (node n,e). Each thermal zone of the building is modelled by using one of the proposed model archetypes, depending on the presence or absence of an *active* envelope node (i.e., floor heating/radiant slab) in the related zone. The final building RC model results from the connection of the RC model archetypes relative to each thermal zone of the building.



(a)



(b)

Figure 4.2: Adopted RC network archetypes for a single thermal zone with (a) active and passive nodes, and (b) passive node.

4.2.2 Internal gains

In buildings, the internal gains can be categorized into three major sources: miscellaneous electric loads, lighting, and occupants. The impact of these factors varies depending on the building type (e.g., residential, office, institutional). In residential buildings, internal heat gains from residents, particularly due to their use of appliances, can partially meet the heating demand [266]. In the case of office buildings, several studies have shown that the miscellaneous electric load serves as the most effective proxy variable for internal heat gains. This load constitutes the primary component and exhibits the highest correlation coefficient with internal heat gains [267]. Simultaneously, in office buildings, the lighting load might constitute 20–30% of the internal heat gains, depending on the implementation of energy-efficient lighting technologies such as dimmable lighting and LED. In institutional buildings, the primary factor influencing the internal gain quota is the occupancy [268]. The developed model incorporates the effects of occupants in each thermal zone, n, by integrating a convective heating allocation within the existing formulation, $\dot{Q}_{\mathrm{gain},n}$. This term is evaluated as follows:

$$\dot{Q}_{\mathrm{gain},n} = \dot{Q}_{\mathrm{MEL},n} + \dot{Q}_{\mathrm{lights},n} + \dot{Q}_{\mathrm{occupants},n} \tag{4.6}$$

where $\dot{Q}_{\mathrm{MEL},n}$ refers to the internal gain due to miscellaneous electric loads, $\dot{Q}_{\mathrm{lights},n}$ refers to the internal gain due to the lighting, $\dot{Q}_{\mathrm{occupants},n}$ is the internal gain due to the occupants.

4.2.3 Calibration procedure

Once the building RC structure is defined, the energy balances associated to all the nodes, generally described by Equation (4.1), are converted into a state-space formulation.

$$\dot{x} = Ax + Bu \tag{4.7}$$

$$y = Cx \tag{4.8}$$

where x are the state variables $x=\left[T_1 \ldots T_N \;\; T_{1,i} \ldots T_{N_1,i} \;\; T_{1,e} \ldots T_{N,e}\right]^T$, u are the input variables $u=\left[T_{out} \;\; \dot{q}_{solar} \;\; \dot{Q}_1 \ldots \dot{Q}_N \;\; \dot{Q}_{1,i} \ldots \dot{Q}_{N_1,i}\right]^T$ and y are the output variables $y=\left[T_1 \ldots T_N \;\; T_{1,i} \ldots T_{N_1,i}\right]^T$. The matrices A and B will be a combination of conductance, capacitance and solar factor values associated to each related node, while C is a null matrix with ones only on the first $N+N_1$ elements on the principal diagonal. N corresponds to the total number of thermal zones, while N_1 corresponds to the number of thermal zones characterized by indoor and exterior nodes. The difference $N-N_1$ corresponds to the thermal zones characterized by only exterior nodes.

During the calibration, the model parameters are bounded into a certain range in order to constraint values to have a physical meaning ($C_n \in \left(1,\ 10^9\right)$ [J/K], $C_{n,i} \in \left(10^3,\ 10^9\right)$ [J/K], $C_{n,e} \in \left(10^3,\ 10^9\right)$ [J/K], $U_{n,m} \in \left(1,\ 10^7\right)$ [W/K], $\alpha_n \in \left(10^{-5},\ 10^3\right)$) and they are subjected to the critical timestep constraint due to the use of the explicit discretization method. The explicit discretization method is a numerical technique used to solve differential equations, where the value of the state variables at the next timestep, $k+1$, is computed explicitly using known values from the current timestep, k. This method typically involves expressing the derivatives in terms of finite differences and solving directly for the future state.

The evaluation of the parameters is executed through an optimization routine using a MatLab function, *fmincon*. The algorithm is highly influenced by the initial model parameters guess; thus, the optimization process is combined with a *constrained* Latin Hypercube Sampling [245]. In this way, only the set of initial parameters that ensure RC model stability are selected as inputs for the calibration process. To provide a model with good prediction performance over a fixed prediction horizon, the MPC relevant identification (MRI) method is then used to calibrate the model parameters. Many studies have shown that MRI yields models with better accuracy in the prediction horizon, although at the cost of significantly increasing the computational resources needed [246]. This method is based on the minimization of the following cost function:

$$J(k)=\sum_{j=1}^{ph}\left\| y_m(k+j)-\hat{y}(k+j)\right\|_2^2 \tag{4.9}$$

where y_m is the measured value, $\hat{y}$ is the estimated value, and ph is the lengths of the prediction horizon. The cost function, J, is evaluated for each time step, k, over the considered calibration horizon. This formulation is a part of the Root Mean Square Error (RMSE), evaluated over the whole prediction period.

$$RMSE=\sqrt{\sum_{j=1}^{ph}\frac{\left(y_m(k+j)-\hat{y}(k+j)\right)^2}{ph}} \tag{4.10}$$

The most accurate model is selected according to the lowest RMSE and the highest value of the Fit function (FIT) [246]. The FIT is a common metric used to define the prediction capabilities of a model. This index is generated from the knowledge of the Normalized Root Mean Square Error (NRMSE):

$$NRMSE = \frac{RMSE}{\sqrt{\sum_{j=1}^{ph} \frac{\left(\bar{y}(k+j) - \hat{y}(k+j)\right)^2}{ph}}} \tag{4.11}$$

where $\bar{y}$ is mean value of the measured output. The FIT index is then evaluated as follows:

$$FIT = \left(1 - NRMSE\right) \cdot 100\% \tag{4.12}$$

The combination of these two metrics is crucial to assess the performance of the model. While the RMSE gives a general information on the discrepancy between the measured value and the estimated one, the FIT gives information on how the temperature trend follows the real dynamic. The calibration period used to capture the building thermal dynamic is equal to one week using a prediction horizon of 24 hours. The selection of the prediction horizon is a consequence of the periodic dynamic of buildings, which see the weather forecasts as their main affecting variable. The metrics RMSE and FIT for the calibrated model are evaluated over a period of one month, showing the accuracy and reliability of the developed RC model.

4.2.4 Model predictive control

Model predictive control is a high-performing algorithm based on a cost function evaluated over a prediction horizon, ph. The cost function gives the possibility to dynamically optimize the building-grid interaction while improving overall system efficiency and adhering to technological constraints. The control algorithm receives as input the disturbances and non-manipulates variables, which are the initial thermal zone temperature, setpoint temperatures, weather forecasts, occupancy profile, and grid signals. The inputs are provided to both the optimization problem and the building model that interact with each other. The developed RC model provides feedback on the indoor air temperature after receiving, as input from the optimization problem, the amount of thermal energy $\dot{Q}_n$ and $\dot{Q}_{n,i}$ provided to the thermal zones and to the active envelope nodes respectively. Once the optimization problem has reached the minimum possible cost function, the algorithm provides the optimal solution in terms of indoor air temperature condition and required heating input. These optimal set of values $\left[\dot{Q}_1 \; \ldots \; \dot{Q}_N \; \dot{Q}_{1,i} \; \ldots \; \dot{Q}_{N_1,i}\right]$ are then implemented over the selected control horizon, ch. The hyperparameters ch and ph affect the optimal control, and their values must be chosen according to the dynamics of the system of interest.

Within a general MPC framework, four key components exist: the cost function, constraints, system dynamics, and the current state. This was previously described in the general formulation of Equation (2.1). For this application, the selected cost function encapsulates the energy and cost dynamics of the system, including technological and comfort constraints. This work relies on the *economic*-MPC (e-MPC) to define the optimal set of control variables. This approach considers in the cost function the following terms:

$$J(t) = J_{\text{cost}}(t) + J_{\text{tech}}(t) + J_{\text{comf}}(t) \tag{4.13}$$

where t is the time step, J_{cost} represents the economic cost of the building operation, J_{tech} represents a penalty term for the operation of the technologies, and J_{comf} is a penalty for thermal zone temperature deviations from the preferred setpoint profile. The combination of these three penalties in the cost function enables a holistic approach to building energy management. The objective of the algorithm is to determine the optimal operation of the building energy systems based on fixed variables derived from grid signals, technology operation, and indoor setpoint temperatures. This approach ensures that the optimization process is driven by realistic and practical considerations, without undue influence from the developer.

The cost penalty addresses the economic aspect of energy consumption, factoring in time-of-use electricity prices, peak demand charges, and potential incentives. This penalty guides the algorithm to minimize operational costs, optimizing the use of energy resources in a financially efficient manner. By integrating cost considerations, the algorithm can dynamically respond to grid signals, shifting loads and adjusting operations to take advantage of lower energy prices and avoid high-cost periods. The economic cost term, J_{cost}, is defined as follows:

$$J_{\text{cost}}(t) = \frac{Q_{th}(t)}{COP(t)} \cdot cost_{en\ el}(t) + \frac{Q_{th}(t_p)}{COP(t_p)} \cdot cost_p \tag{4.14}$$

where t is the time step, and t_p denotes the time of the day at which peak energy consumption occurs. Moreover, Q_{th} is the thermal energy required, COP is the coefficient of performance of the heating system, $cost_{en\ el}$ is the electricity cost [*\$/kWh*$_{el}$], and $cost_p$ is the electricity cost for the peak of energy required for that day [*\$/kW*$_{el}$].

The technology penalty, J_{tech}, is introduced to consider the real operation of the technologies. This penalty bounds the variation of the heating demand from the different technologies, avoiding excessive variation using a specific ramp rate, $Q_{\text{ramp rate}}$.

$$J_{\text{tech}}(t) = \begin{cases} \lambda \cdot \Delta Q_{th}(t) = \lambda \cdot \left(\sum_{n=1}^{N} \left(Q_n(t) - Q_n(t-1) \right) + \sum_{n=1}^{N_1} \left(Q_{n,i}(t) - Q_{n,i}(t-1) \right) \right) & \text{if } \Delta Q_{th}(t) > Q_{\text{ramp rate}} \\ 0 & \text{else} \end{cases} \tag{4.15}$$

Introducing this penalty term in the cost function has a dual effect: improves operation of the heating system and limit the rebound effect [269] after periods of high price.

The comfort penalty, J_{comf}, depends on the difference between the indoor air temperature, T_n, and the setpoint temperature, T_{set} and quantifies deviations from the desired indoor setpoint temperatures, ensuring that occupant comfort is prioritized. The value of this cost function at time step t, is evaluated as follows:

$$J_{\text{comf}}(t) = \begin{cases} 0 & \text{if } T_n(t) > T_{set}(t) - \beta(t) \text{ and } T_n(t) < T_{set}(t) + \beta(t) \\ \varepsilon \cdot \sum_{n=1}^{N} \left(T_n(t) - T_{set}(t) \right)^2 & \text{else} \end{cases} \tag{4.16}$$

with β is the bound applied to the setpoint variation [°C], which provides flexibility to the indoor temperature variation through the hours of the day. The J_{comf} penalty is evaluated for all the building thermal zones, making the optimization *centralized*. The choice of the comfort cost function generally depends on the technologies involved [55]. The if-else structure introduces a non-linearity in the cost function but gives flexibility to the temperature fluctuations inside the building. This penalty is evaluated for all the thermal zones of the building. The value of the weight factor of the technology penalty, λ, and the weight factor of the comfort penalty, ε, are in the order of 1 [*$/kWh*] and 1 [*$/K*] respectively. These values give same weight to comfort and technology cost functions over the economic one.

The optimization routine of the management strategy is executed in MatLab using the function, *fmincon*. The algorithm uses the *interior point* method, particularly efficient for large-scale optimization problems with constraints commonly encountered in the engineering field. This method can handle both equality and inequality constraints effectively, making it versatile for modelling problems with nonlinear constraints. One of the key advantages of interior point method is the robustness since it is less prone to fall in local minima/maxima compared to gradient-based methods like gradient descent. This is essential for complex and non-convex optimization problems.

4.2.5 Hierarchical Clustering of Weather Data

Generalizing or classifying results and optimal control settings is essential for bridging the gap between simulation studies and real-world implementation. To ensure that control strategies are practical and robust across varying conditions, it is necessary to develop methods capable of adapting to diverse scenarios. This involves reducing the complexity of data and aggregating it into representative subsets while maintaining the accuracy needed for actionable insights. Clustering techniques are unsupervised learning algorithms useful for identifying patterns in large datasets. This allows for the aggregation of multiple results or scenarios into representative groups, called clusters. In the context of building energy management, clustering helps simplify the variability of weather conditions by categorizing similar days into distinct clusters. This enables simulations and analyses to focus on representative scenarios, reducing computational effort while maintaining a comprehensive understanding of the system. In this study, hierarchical clustering is employed to group weather data based on daily mean outdoor temperature and solar radiation levels. These variables are chosen for their critical influence on building energy performance.

Hierarchical clustering consists in building a binary merge tree, starting from the data elements stored at the leaves and proceed by merging two by two the "closest" sub-sets until it reaches the root of the tree that contains all the elements of the considered dataset. The clustering process is performed using the minimum variance algorithm (also known as Ward's method) with the Euclidean distance metric as *linkage distance* [270]. This approach minimizes the total within-cluster variance and ensures that clusters are compact and distinct. This technique is also called *agglomerative* hierarchical clustering since it starts from the leaves and merges iteratively subsets until it reaches the root. The graphical representation of this binary merge tree is called a dendrogram. This provides a visual representation of the clustering process, revealing relationships between different weather patterns [271].

One of the key advantages of hierarchical clustering is that it does not require pre-specification of the number of clusters, making it flexible for exploring datasets with unknown structure. Additionally, its hierarchical nature allows for analysis at different levels of granularity, offering insights into both broad and fine-grained patterns. However, it can be computationally intensive, particularly for large datasets, and may struggle with scalability compared to other methods such as k-means clustering. Despite these challenges, its ability to handle diverse weather data and avoid assumptions about cluster shapes makes it a suitable choice for this study.

The clustering results are used to identify representative weather scenarios, which serve as the basis for simulations. These simulations are conducted for each specific scenario to study the optimal control outputs. By aggregating the results of these simulations, the methodology ensures that the solutions are generalizable across different weather conditions. This approach facilitates a comprehensive analysis of building performance while maintaining scenario-specific applicability in real-world energy management systems.

4.2.6 Key performance indicators

The performance of the load management is evaluated through the key performance indicators synthetized in Sections 2.6.2 and 2.6.3.

In detail, Temperature State of Charge ($TSoC_{n,i}$), Radiative Ratio (RR) and Ramp Index ($Ramp$) provide information on the temperature and technology monitoring, while the Load Factor (LF), Flexibility Factor (FF) and Building Energy Flexibility Index ($BEFI$) are used to evaluate the building energy flexibility potential.

4.3 Case study

The proposed methodology is applied to the Varennes' Library (Figure 4.3), standing as the first institutional Net Zero Energy Building in Canada [161]. The building, inaugurated in 2014, is a municipal library situated in the town of Varennes, approximately 40 km away from Montreal (Québec).

The library achieves net-zero energy performance through on-site renewable systems, energy efficient technologies, and a solar design. It boasts a 110 kW$_p$ building integrated photovoltaic system, with 1/6 of it functioning as building integrated photovoltaic thermal collectors. This technology is linked to a heat recovery exchanger for fresh air preheating, connected with a dedicated outdoor air handling unit.

Figure 4.3: Picture of the Varennes' Library.

This unit processes the air and distributes it within the library through air channels controlled by dampers and fan coils. The terminals consist of ceiling/floor-displaced air diffusers. The building is equipped with four ground source heat pumps, connected to eight boreholes, with nominal heating capacity of 28.8, 26, 13 and 13 kW respectively. During the operation, the system level COP is calculated to be approximately 4.5 for heating and 4 for cooling since the source is a ground loop. This system caters to the thermal load of the library required by both the air handling unit, ventilation heating, and floor heating systems, which cover approximately 34% of the net surface area. These features emphasize the presence of convective terminals and active envelope components within the building. The building features demand-driven ventilation and the potential for natural cross-ventilation. The building spans a net surface area of 2000 m² distributed across two floors and a basement area, where the mechanical room is located. From the building plans, three main thermal zones can be identified: two on the first floor and one on the second floor. Heating and cooling in these thermal zones are supplied through ceiling-displaced air diffusers on the first floor and floor-displaced air diffusers on the second floor. Figure 4.4 shows that the building is characterized by three main thermal zones with the ones highlighted in green equipped with floor heating systems.

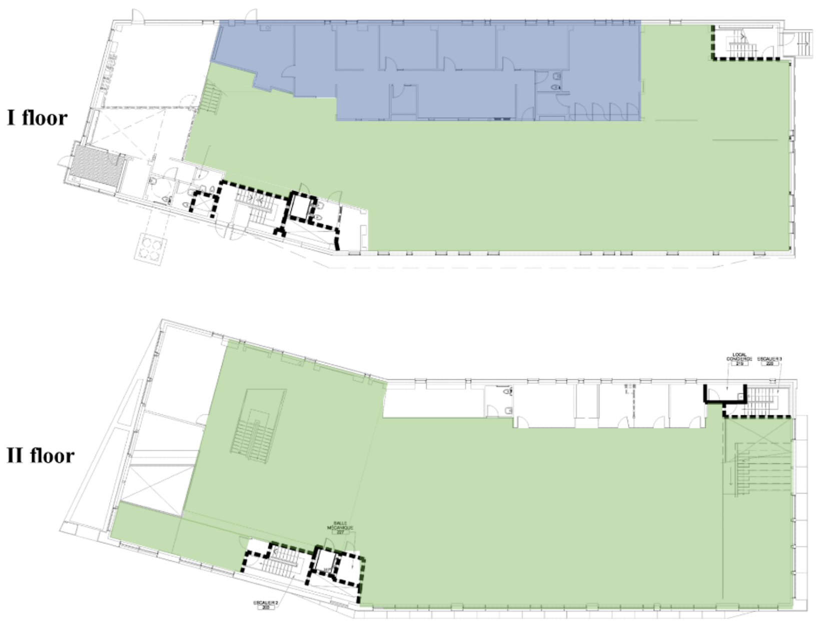


Figure 4.4: Layout of the main building thermal zones located at the first and second floor.

Operations in the building follow a planned schedule to minimize occupancy-related uncertainties and are supervised by building operators. The library is equipped with smart thermostats and sensors that measure indoor air temperature, slab temperature (embedded at the centre of the slab), humidity, CO_2 levels, flow rates, and the mode of operation of valves and pumps in the heating and cooling loops.

The Building Automation System (BAS) operates on a rule-based control approach, relying on setpoint temperature profiles, occupancy profiles, and demand control ventilation, which keeps CO_2 concentration below a defined threshold. On a demand for heating or cooling, or ventilation the geothermal loop and air handling unit respectively are activated. Those systems through glycol loops and air channels provide heating, cooling or fresh air to the different zones of the building. More details on the operation and integrated technologies of the library can be found in [161].

4.4 Results: Thermal model definition

This Section describes the developed RC thermal network for the Varennes' Library. The RC building model structure depends on the number of thermal zones and the presence of active envelopes (i.e., floor heating). As defined in Figure 4.4, this case study presents three main thermal zones, with two of them – represented in *green* in Figure 4.4 – characterized by a convective heating and floor heating system.

The presence of the floor heating is modelled through the adoption of *active* envelope nodes, using the RC model archetype of Figure *4.2* (a). While for the thermal zone equipped only with convective heating system – represented in *blue* in Figure 4.4 – is modelled using the RC model archetype of Figure *4.2* (b). The description of each node in the building RC network is reported in Table 4.1. With the adopted archetypes and the building zone schematic, an 8C14R is selected as a structure to capture the building thermal load (Figure 4.5).

Table 4.1: Description of the considered building RC-network thermal nodes.

Node	Description
1	Air node – ***I floor*** green area in Figure 4.4
2	Air node – ***I floor*** blue area in Figure 4.4
3	Air node – ***II floor*** green area in Figure 4.4
1i	Active envelope node – ***I floor*** floor heating
3i	Active envelope node – ***II floor*** floor heating
1e	Passive envelope node of Node 1
2e	Passive envelope node of Node 2
3e	Passive envelope node of Node 3

The parameters of the RC model are calibrated, considering a 24-hrs ahead prediction, with real data obtained from the BAS installed in the building. The acquisition system provides measurements of indoor air temperature, thermal energy and weather forecasts necessary for the calibration and validation periods (respectively one week and one month during the winter period). The calibration happens by adopting the methodology described in Section 4.2.3, employing the MPC relevant identification method using a prediction horizon of 24-hrs ahead. The parameters and performance metrics of the best set of model parameters are shown in the following tables and figures for training and testing periods. In detail, Table 4.2 provides information on model performance during the training/calibration period, while Table 4.3 and Figure 4.5 refer to the results during the validation/testing period. The results show acceptable accuracy with values of RMSE under 0.6 °C and FIT over 40% for prediction of one day/24-hours ahead. The performance indices of the individual nodes are still acceptable according to the values available in literature which suggest RMSE under 1°C and FIT values for prediction of 6 hrs ahead over 40% [246]. Finally, Table 4.4 shows the capacitance, conductance and effective solar factor values for the considered building model.

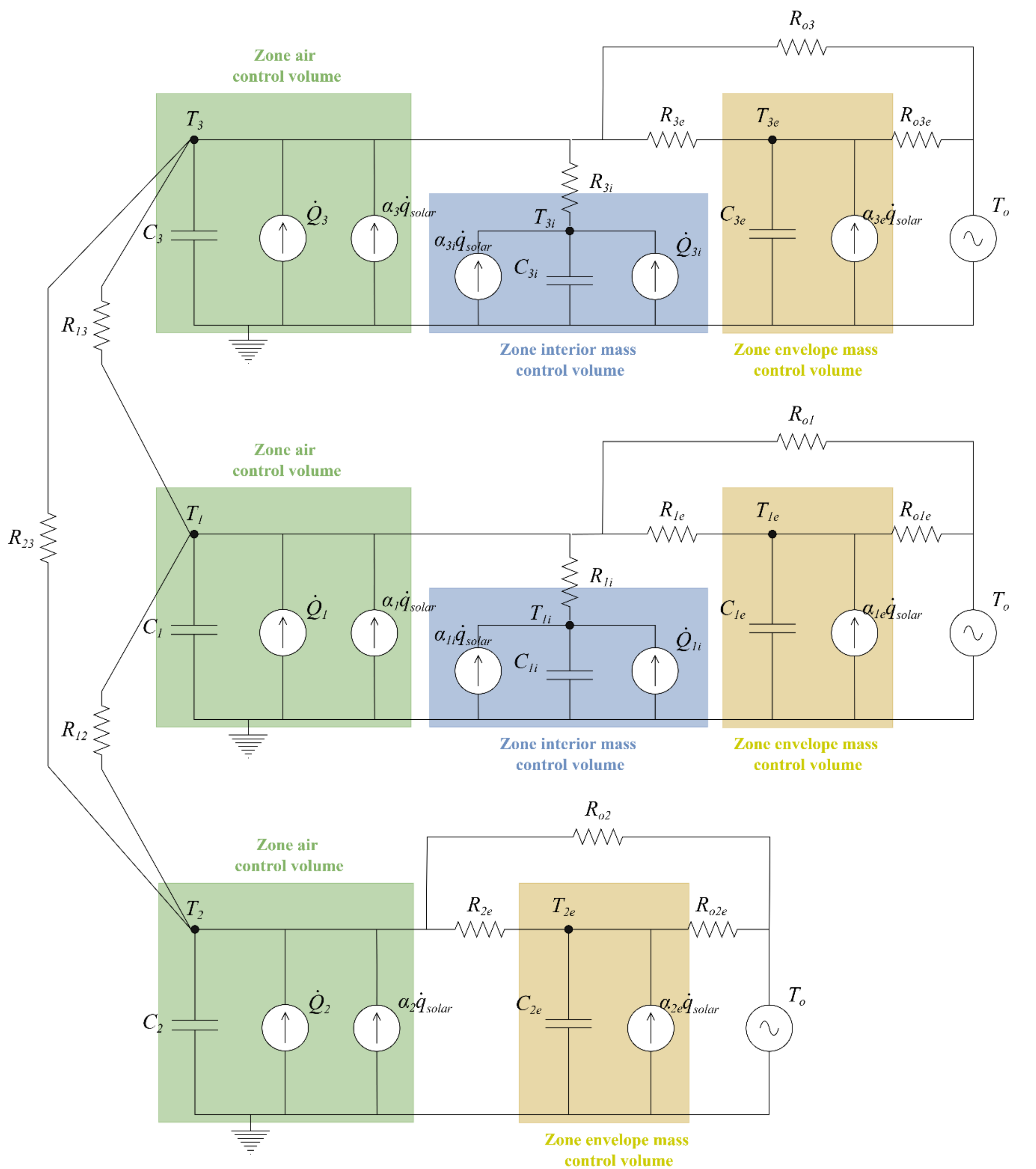


Figure 4.5: 8C14R network adopted for the Varennes' Library whose nodes are described in Table 4.1.

Table 4.2: Model performance for prediction of 24 hours ahead during the calibration period (21/12/2017 – 28/12/2017).

Node	1	2	3	1*i*	3*i*	Total
RMSE *[°C]*	0.192	0.244	0.309	0.122	0.237	0.221
FIT *[%]*	67.50	69.92	58.41	45.93	63.54	61.06

Table 4.3: Model performance for prediction of 24 hours ahead during the testing period (28/12/2017 – 01/02/2018).

Node	1	2	3	1*i*	3*i*	Total
RMSE *[°C]*	0.505	0.653	0.629	0.171	0.509	0.494
FIT *[%]*	46.15	43.00	37.66	56.65	30.20	42.73

It should be noted that the lumped nature of the parameters may make them lose their strict physical meaning and instead represent the physics of the system more loosely. Therefore, it is noted that the values of all estimated parameters are within logical physical bounds.

The developed model gives the possibility to study the thermal dynamic of the building distinguishing the different dynamics due to the thermal mass and the different heating terminals. The RMSE and FIT values for the developed model are within an acceptable range, making the choice of a liner model with constant parameters and the RC structure acceptable for the considered application.

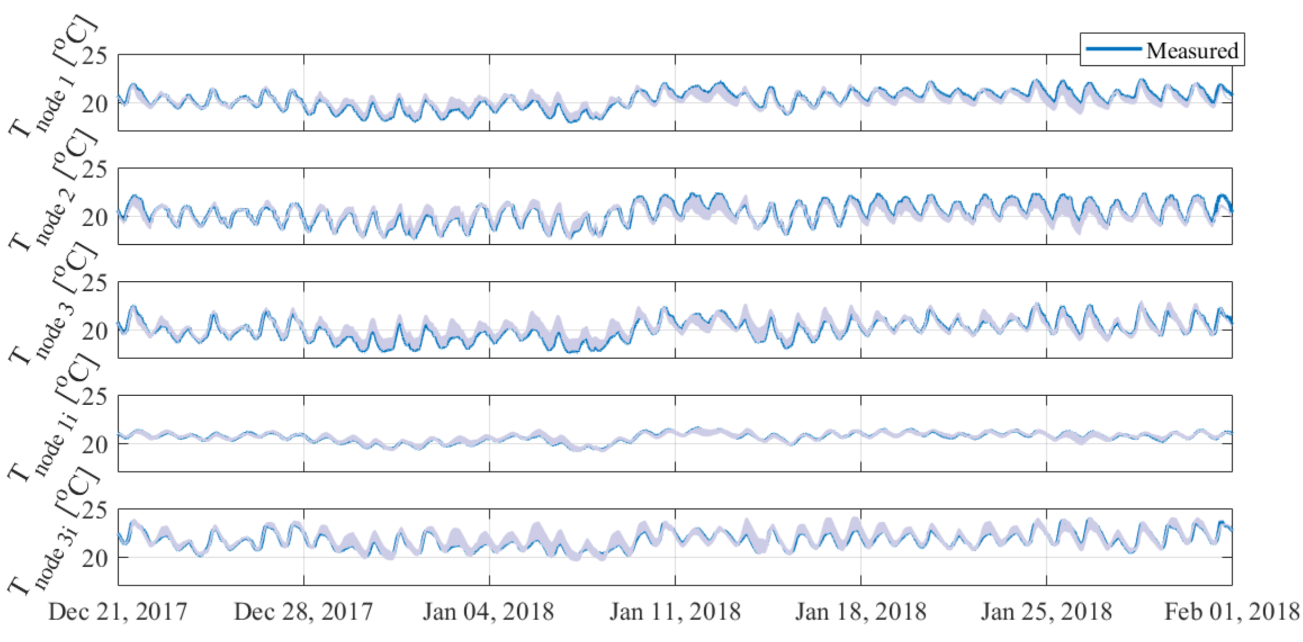


Figure 4.6: Performance of the model for 24 hours prediction during the validation and testing period (21/12/2017 – 01/02/2018) and for each thermal node (described in Table 4.1).

Table 4.4: Calibrated parameters of the 8C14R model.

Node n	**1**	2	3	$1i$	$3i$	$1e$	$2e$	$3e$
C_n [J/K]	$6.35 \cdot 10^4$	$2.47 \cdot 10^4$	$1.12 \cdot 10^5$	$2.46 \cdot 10^5$	$2.10 \cdot 10^5$	$8.26 \cdot 10^7$	$2.30 \cdot 10^6$	$8.18 \cdot 10^8$
α_n [-]	37.90	1.31	70.32	0.86	55.98	0.07	18.56	0.07
U_{on} [W/K]	$1.35 \cdot 10^2$	57.17	$1.01 \cdot 10^2$	-	-	$7.39 \cdot 10^3$	$1.12 \cdot 10^2$	$1.03 \cdot 10^5$
U_{1n} [W/K]	-	$1.78 \cdot 10^3$	~ 0	$5.38 \cdot 10^3$	-	$8.97 \cdot 10^2$	-	-
U_{2n} [W/K]	$1.78 \cdot 10^3$	-	~ 0	-	-	-	$5.70 \cdot 10^2$	-
U_{3n} [W/K]	~ 0	~ 0	-	-	$5.45 \cdot 10^3$	-	-	$4.62 \cdot 10^3$

4.5 Results: Thermal energy management

This Section describes the results of the thermal load management for the considered building. The methodology is based on the MPC routine described in Section 4.2.4, relying on the model provided in Section 4.2.6. The target is to optimize the coordination between the heating provided to the thermal zones through ventilation heating and the active envelope nodes. The results are compared with the measurements of the actual rule-based control inside the building. New key performance indicators, along with others currently available in literature, are used for the comparative assessment. This Section is divided into four parts: Section 4.5.1 describes the settings and inputs required for the optimization routine, Section 4.5.2 describes the effect of the optimal control on the temperature and thermal mass exploitation, Section 4.5.3 provides the results in terms of building thermal load and how the different pricing structures modify it, and Section 4.5.4 describes the building energy flexibility results.

4.5.1 Optimization settings and inputs

The selection of hyperparameters, ch and ph, in the control routine affects the performance of predictive algorithms (i.e., detection of dynamic changes in the input variables, such as weather forecasts and price signals). A shorter prediction horizon, ph, generally results in a slower response to long-term dynamics (i.e., activation of the active envelope). Conversely, a longer control horizon, ch, may offer slower adaptability to future and unforeseen changes. In building applications, good results are typically attained with lower control horizons (i.e., couple of hours) and higher prediction horizons (i.e., 6, 12 or 24 hours) [138]. Furthermore, it is essential to select a compromise solution that balances between optimal performance and computational time. In this study, a prediction horizon of 12 hours is chosen, while the control horizon is fixed at 2 hours. This selection will enable the detection of price variation, solar effect and the different building thermal dynamics. The methodology is applied to different weather forecasts and price signals. Here are described the input variables:

- **Simulation period**: the simulation length is one day, and it is executed for 30 days with different winter weather conditions. The explicit method is the numerical simulation method adopted.
- **Weather forecasts and clusters** [272]: as described in Section 4.2.5, agglomerative hierarchical clustering is used to cluster the weather data contained in the selected simulation period. The daily values of mean outdoor temperature [*°C*] and solar radiation [Wh/m^2day], reported in Figure 4.7, are the parameters used to create groups/clusters which share similar features. According to the methods used, the algorithm identifies three main weather clusters. The three clusters can be categorized into: "Very Cold" days (red colour), "Cold Cloudy" days (green colour) and "Cold Sunny" days (blue colour). This classification is acceptable given the presence, in the selected simulation period, of few cloudy days during the "Very Cold" period, with daily solar radiation values mainly concentrated between 1100 to 1800 [Wh/m^2day]. Figure 4.7 shows the dendrogram of the three identified weather clusters (left) and the values of temperature and solar radiation (right). This cluster identification for weather data helps in simplifying the representation of the results and in generalizing the optimal control setting for a larger range of daily weather conditions which belongs to a specific cluster.

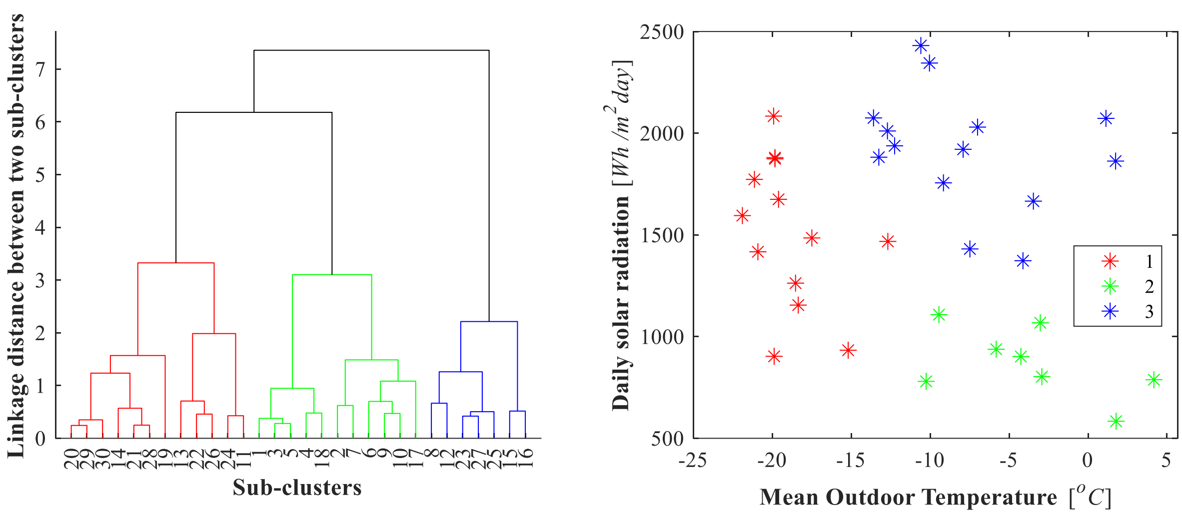


Figure 4.7: Dendrogram of the three identified weather clusters (left) and values of daily mean outdoor temperature and daily solar radiation (right).

- **Occupancy schedule**: the occupancy is fixed as constant for every day of simulation, assuming a gaussian distribution with peak period between 12 am to 1 pm and by fixing a maximum number of people – equal to 50 people – during that time. The maximum value is estimated with an occupancy meter located in the library. To account for the occupants' impact during operation, it is essential to derive an occupant profile distribution throughout the day and subsequently multiply it by the [*W/person*] component. A moderately active office worker averagely generate sensible heat of 73.2 [*W/person*], and latent heat of 58.6 [*W/person*], which is equivalent to 131.8 [*W/person*] [267, 273].
- **Fresh air ventilation**: fresh air is provided in the building through infiltration (connected to the term $U_{n,o}$ of Equation (4.1)), and by a component, $\dot{m}_{fresh}$, corresponding to the fresh airflow delivered from a dedicated outdoor-air system. Since the building is equipped

with a dedicated outdoor air handling unit, $\dot{m}_{fresh}$ is non-null and is evaluated as a function of the building area and the opening hours of the building. According to the ASHRAE 62.1 guideline [265], for a library, this value should be at least 0.6 [$L/s{\cdot}m^2$]. By multiplying this term by the total building area, the thermal energy required can be evaluated with Equation (4.5). The selected $\dot{m}_{fresh}$ trend corresponds to the actual operation of the building, and it is kept constant during the occupied hours. It is important to underline that the fresh air ventilation component is not considered as controllable in this study. For this reason, and given the application over a winter period, the thermal energy results showed in the next sections refer only to the heating supplied to the thermal zones, $\dot{Q}_n$ and $\dot{Q}_{n,i}$ respectively, evaluated from the RC model Equations (4.1) and (4.2).

- **Heating system performance**: the system level COP is calculated to be approximately 4.5 for heating and 4 for cooling since the source is a ground loop. Therefore, a more detailed evaluation can be integrated in Equation (4.14) as done in [262].
- **Setpoint temperature**: this value varies from 18 to 22 °C from unoccupied to occupied periods and it is applied to all the building thermal zones. The occupied periods are from 7 am to 9 pm.
- **Price signals**: two price signals are considered to study the performance of the energy system. These do not consider the utility distribution charges and includes:
    - Constant price obtained from Rate M of the utility Hydro-Québec [274]. This rate applies to a medium-power contract with a maximum power demand over 50 kW.
    - Critical-peak pricing structure (or Demand Response) obtained from Rate flex M of the utility Hydro-Québec [274]. The peak period is generally generated by considering the predicted values of the outdoor temperature, given the winter dominated climate.

Table 4.5: Main features of the price signals.

Price signal	*Constant price:* Rate M	*Demand response:* Rate Flex M
Minimum price [*$/kWh*]	0.055	0.035
Maximum price [*$/kWh*]	0.055	0.553
Maximum price variation [*$/kWh*]	0	0.518
Peak price [*$/kW*]	16.139	16.139
Period of high price	-	6 am to 9 am
Dependencies	-	Outdoor temperature
Additional expenses	Distribution rate	Distribution rate

4.5.2 *Effect of optimal control on thermal zones temperature and thermal mass exploitation*

This section compares the results in terms of indoor air temperature and active envelope temperature trends. The results show the different temperature fluctuations obtained with the adopted strategies and the exploitation of the building thermal mass during the operation of the building. For simplicity the following figures refer to a period of one week, and according to the outdoor temperature and irradiance values, reported in Figure *4.8*, it is possible to associate the considered days to specific clusters (the 2nd, 5th, 6th and 7th day ∈ Cluster 1).

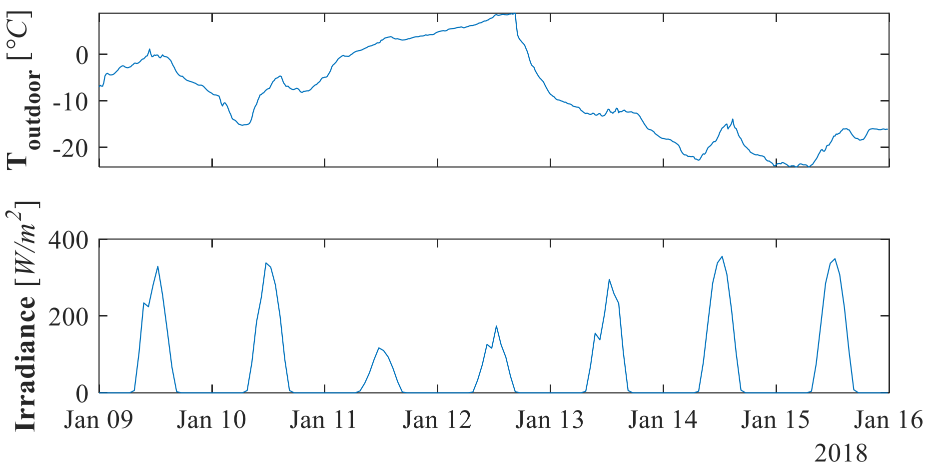


Figure 4.8: Weather forecasts for the considered period.

The *Reference Scenario with a rule-based control* corresponds to the actual operation of the building. *Figure 4.9* shows the second-floor thermal zone, T_3, active envelope, $T_{3,i}$ and setpoint temperature with the active envelope state of charge and the radiant heating mode for this considered scenario. Due to the adopted rule-based control, the heating is provided simultaneously and equally to the second-floor thermal zone, through convective terminals, and to the active envelope. As shown in *Figure 4.9*, in the first hours of the day, the thermal zone temperature is not able to reach the desired setpoint. Furthermore, according to the trend of $TSoC$, the thermal mass of the second-floor thermal zone is not exploited to the fullest, especially in periods of high demand (Very Cold days ∈ Cluster 1).

During this scenario, the radiant heating system is ON for the 71% of the simulation period, and the heating delivered to the indoor environment through ventilation heating and radiant heating appears to be equally divided. During the operation, thermal energy is stored in the active envelope and released during occupied periods simultaneously. While convective terminals directly affect the zone temperature, the activation of the heating through the thermal mass cannot rely on a setpoint air temperature trend based on the occupancy (library open from 7 am to 9 pm). Instead, the activation of the active envelope (floor heating) must account for its slow thermal dynamics which can also vary according to the structure of the heating system (i.e., placing of the pipes, slab thickness, material, etc.).

Thanks to the developed building thermal model, the model-based control approach can consider, during the optimization routine, different building thermal dynamics, different operation of the heating terminals, weather forecast and price signals from the electrical grid. This allows a diversification of the contributions coming from the convective heating (i.e., ventilation) which affects directly the thermal zone temperature and the active envelope (floor heating). The heating provided to each thermal zone, whether through convective or radiant heating (depending on the specific zone), is uniformly distributed and adheres to the hypotheses described in Section 2.1.

Each thermal zone is served by distinct heating inputs: zone 1 by $\dot{Q}_1$ and $\dot{Q}_{1,i}$, zone 2 by $\dot{Q}_2$, and zone 3 by $\dot{Q}_3$ and $\dot{Q}_{3,i}$. Although the optimization problem is *centralized*—considering the comfort of all thermal zones and overall energy consumption within a single cost function—the distinct heating inputs ensure that the specific demands of each zone are met. These demands are addressed in accordance with the setpoint temperature, weather forecasts, and the unique thermal dynamics of each zone.

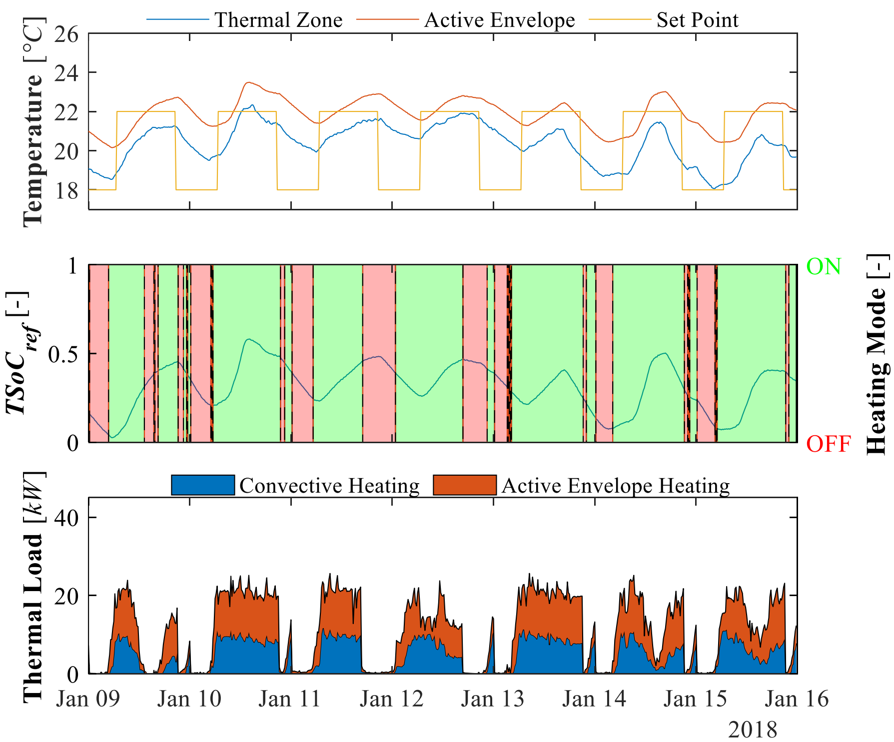


Figure 4.9: *Reference Scenario – rule-based control*: second floor thermal zone, active envelope and setpoint temperature (top), active envelope state of charge and its heating mode (middle), convective and active envelope heating differentiation for the second-floor zone (bottom).

By employing a model predictive control routine (with constant price signal), *Table 4.6* shows, for a period of one week, the comparison in terms of thermal energy consumption of the whole building and mean temperature values of the two thermal zones equipped with a radiant heating system (Zone 1 and Zone 3) and the one characterized by only convective heating (Zone 2). These temperature values represent important components for the evaluation of each thermal zone operative temperature [275]. The scenarios involve *full-convective*, *full-radiant*, and *mixed convective-radiant heating*, and are executed by varying the maximum thermal energy that each specific terminal can provide. The results show that the MPC strategy, compared to the reference, provides better indoor comfort conditions and lower energy consumption. The optimal control of the building thermal mass contributed to further reduce the energy consumption values as defined by the *full-radiant* scenario. Therefore, since in this scenario, the heating is supplied only by the radiant slab, the slab temperatures may reach very high values. These temperatures may negatively affect the occupants comfort. For this reason, the combination with the ventilation heating can reduce the possible thermal discomfort in the indoor environment, finally demonstrating that the coordination of ventilation and radiant heating is worth studying.

Table 4.6: Comparison between different heating terminals utilization for different scenarios.

	Thermal demand $[MWh]$	First floor			Second floor	
		$T_{1,mean}$	$T_{1,i,mean}$	$T_{2,mean}$	$T_{3,mean}$	$T_{3,i,mean}$
Reference	3.03	20.6	20.8	20.6	20.3	21.8
MPC full-convective	2.97	22.4	22.5	21.9	21.8	22.1
MPC full-radiant	2.53	22.3	23.1	21.6	21.7	23.3
MPC mixed convective-radiant	2.60	22.1	22.2	21.9	22.1	22.8

By focusing on the coordination between ventilation heating and radiant heating, *Figure 4.10* and *Figure 4.11* provides more detail on two scenarios respectively: (i) *Model predictive control with constant price signal* (Rate M) and (ii) *Model predictive control with demand response event* (Rate Flex M). These figures show the second-floor thermal zone, T_3, active envelope, $T_{3,i}$ and setpoint temperature with the active envelope state of charge and the radiant heating mode for each specific scenario. From the results of the simulation, with the predictive control, the active envelope is heated before the variation in the indoor setpoint temperature. In both scenarios, the second-floor thermal zone temperature can follow the change in setpoint. This condition improves the exploitation of the building thermal mass, diversifying the heating input to different terminals (i.e., fan coils and floor heating), and improves indoor thermal conditions.

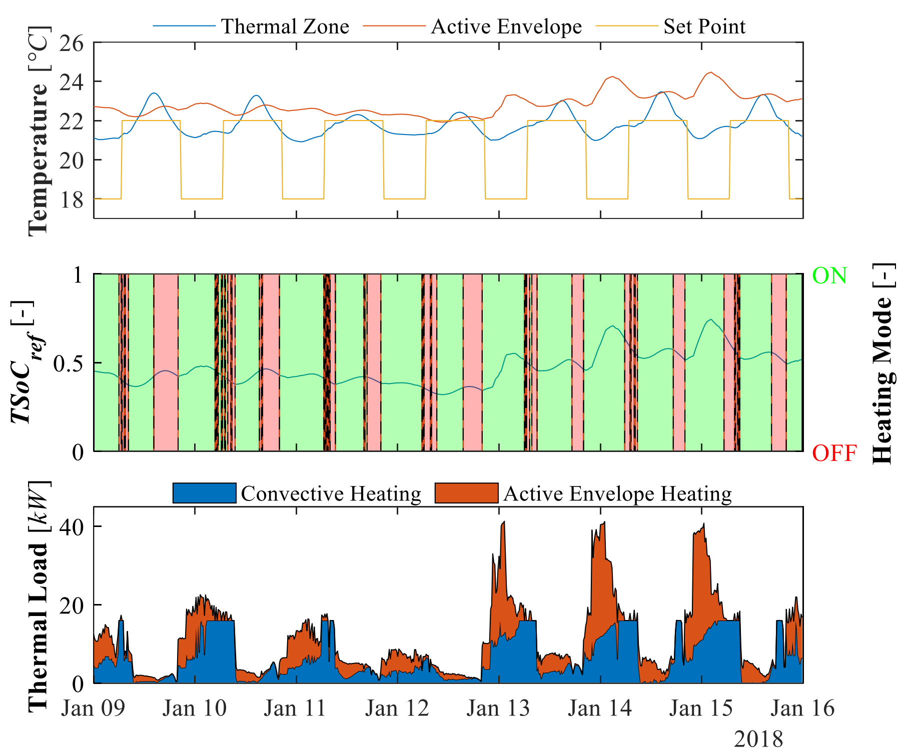


Figure 4.10: *Model predictive control with Rate M*: second-floor thermal zone, active envelope and setpoint temperature (top), active envelope state of charge and its heating mode (middle), differentiation of the convective and active envelope heating for the second-floor zone (bottom).

In these figures, it is noteworthy that the intensity of the preheating is influenced by the outdoor temperature. Examining the 2nd, 5th, 6th and 7th days, which correspond to "Very Cold" days, reveals a clear differentiation between the convective heating and the active envelope heating, as shown in Figure 4.10 and Figure 4.11. This differentiation indicates that active envelope heating plays a crucial role in maintaining the indoor setpoint temperature and, in case of Figure 4.11, avoiding energy consumption during periods of high prices. Furthermore, comparing the trends of *TSoC* in Figure *4.10* and Figure *4.11* reveals that slab utilization is higher during the DR scenario. This is consequence of the high electricity price from 6 am to 9 am, which force the heating system to activate before the price increase. These observations also enhance the importance of creating weather clusters to better diversify control actions during the building operation. Therefore, further discussion on this topic is continued in the next sections.

During *Model predictive control with constant price signal* and *Model predictive control with demand response event*, the radiant heating system is ON for 70% and 62.5% of the simulation period, respectively, showing a significant reduction compared to the reference scenario.

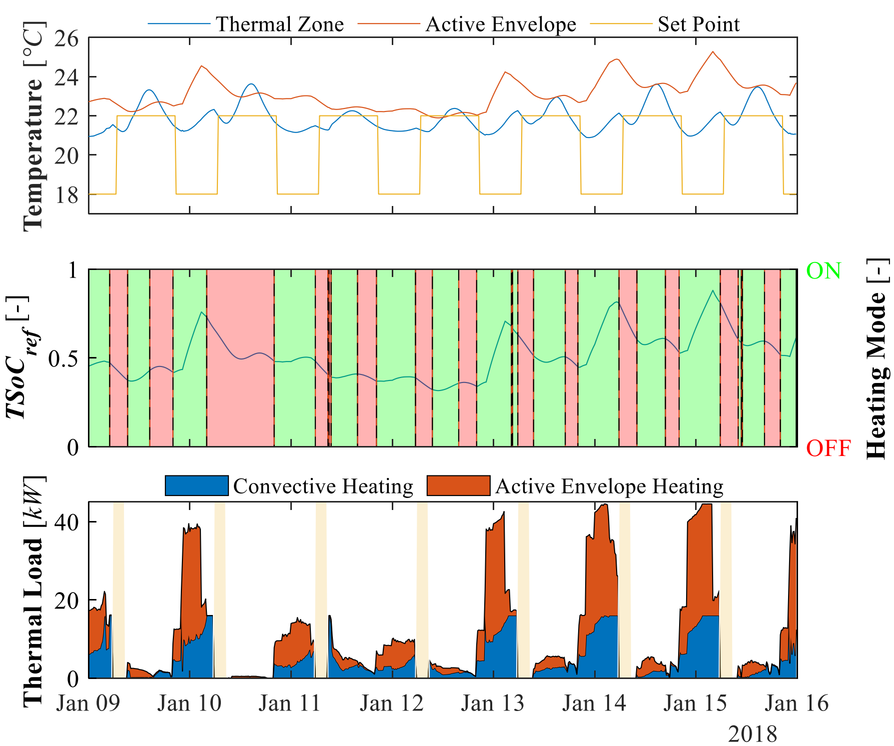


Figure 4.11: *Model predictive control with Rate Flex M*: second-floor thermal zone, active envelope and setpoint temperature (top), active envelope state of charge and its heating mode (middle), differentiation of the convective and active envelope heating for the second-floor zone with high price in orange area (bottom)

4.5.3 Effect of optimal control on overall building thermal load profile

The difference in the thermal management, considering the effect of the different tariff structures, is further described in this section by comparing the results with a reference scenario. The results related to the different pricing structures (Rate M and Rate Flex M) are obtained by using the optimal control described in the methodology, based on the RC thermal model and the MPC routine.

Reference scenario – Rule-based control

The reference scenario, corresponding to the actual rule-based control applied in the library, provides real measurements gathered during the operation of the building. According to the established rule-based control, the heating system is activated only during the occupied periods (setpoint change – from 7 am to 9 pm). The ventilation and the floor heating systems are activated at the same time and there is no diversification in their control. As described before, this condition may cause inefficiency during the building operation since the building thermal mass is not exploited to the fullest.

Figure 4.12 gives information on the daily thermal energy demand, the peak and the minimum and maximum variation of the thermal demand through the day (*Ramp Down* and *Ramp Up* indices, respectively). This figure is based on a box chart representation which is combined with a red trend showing the average values of each label for each weather cluster. As expected, the higher thermal consumption happens during Cluster 1, corresponding to "Very Cold" days. The peak of demand is not really affected by the different weather conditions and its values are mostly between 35 to 42 kW_{th}. The values of the *Ramp Up* index are not affected by the different weather clusters, unlike the *Ramp Down* index which sees an increase for very cold days reaching $-34\ kW_{th}$.

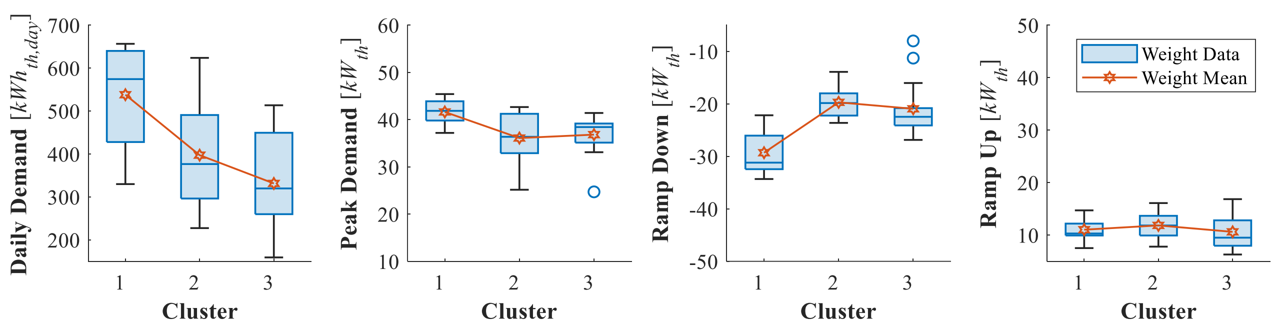


Figure 4.12: *Reference Scenario – rule-based control*: daily energy demand, peak of demand, ramp down and ramp up for different weather clusters.

More insights on the actual operation of the building are reported in Figure *4.13* which shows, through the 24 hours, the Radiative Ratio (RR) index, representing the percentage of thermal power provided to the active envelope nodes versus the total thermal demand, and the trend of the thermal load. The trends of RR and building thermal load are reported by considering the mean value of the trend for each weather cluster, identified in Figure 4.7, and the standard deviation (combining a solid line and a shaded area). The figure shows that the energy demand is concentrated mainly during the occupied hours, from 7 am to 9 pm. During this period, the radiant heating supplies around the 50% of the required thermal energy for the whole building. This means that, with a basic rule-based control, systems with different time-lags are controlled in the same way, hence, the actual BAS does not distinguish the different dynamics. The thermal load shows that no preheating is in effect, and the coordination between the heating provided to the thermal zones and the active envelopes is not optimal.

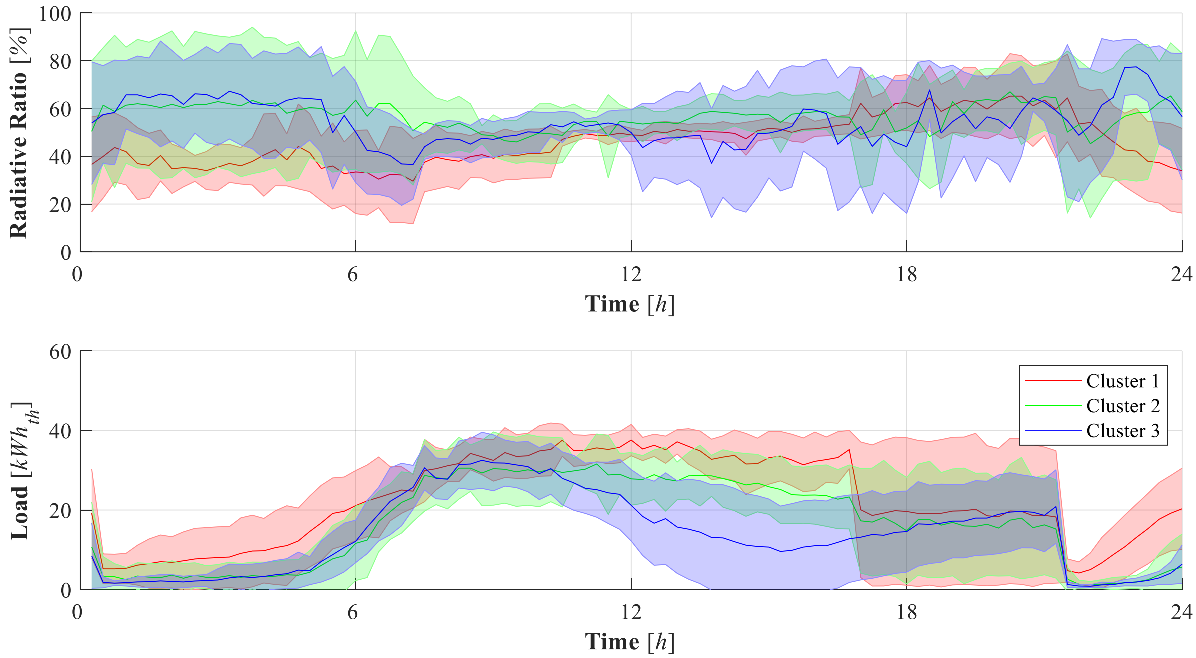


Figure 4.13: *Reference Scenario – rule-based control*: radiative ratio and thermal demand profiles for different weather clusters.

Optimal scenario – Model Predictive Control with Rate M and Rate flex M

The use of predictive control strategies combined with pricing schemes introduces a "grid effect" in the optimization routine. This creates a dual target: (i) reduce building operation costs and (ii) improve building-grid interaction. Two pricing signals, constant (Rate M) and demand response/critical peak pricing (Rate Flex M) are used to study the optimal building thermal load management.

Figure 4.14 and *Figure 4.15* show, for Rate M and Rate Flex M respectively, and for different weather clusters, the daily thermal energy demand during the period of study, the peak demand and the minimum and maximum variation of the thermal load through the day (related to the *Ramp Up* and *RampDown* indices respectively). In terms of daily energy demand, a higher difference between the weather clusters can be found, and, unlike the reference scenario, a lower difference between the energy consumption for Cluster 2 and 3 can be spotted. By comparing their average and median values, the difference in terms of energy consumption between Rate M and Rate Flex M is very low.

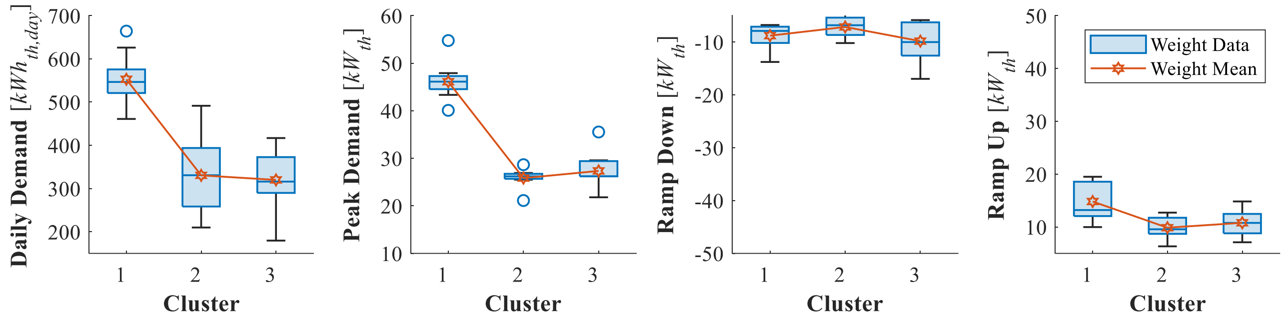


Figure 4.14: *Model predictive control with Rate M*: daily energy demand, peak of demand, ramp down and ramp up for different weather clusters.

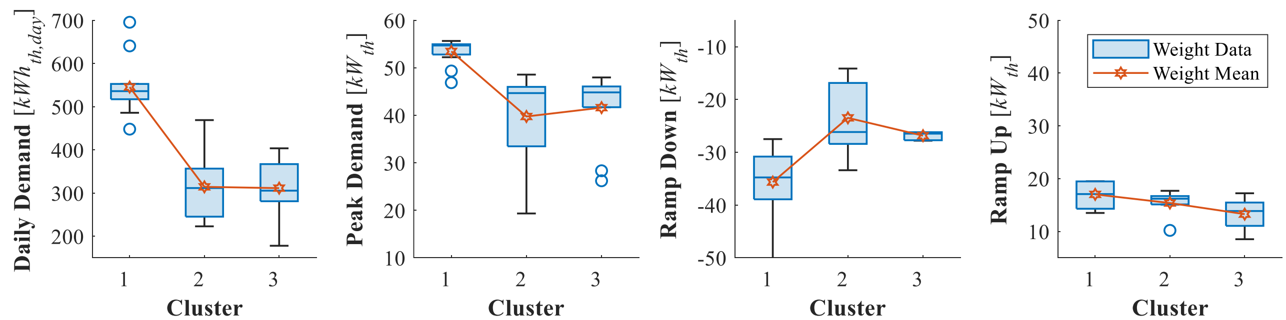


Figure 4.15: *Model predictive control with Rate Flex M*: daily energy demand, peak of demand, ramp down and ramp up for different weather clusters.

The effect of the different pricing signals can be noticed in the values of peak and energy variation through the day. As shown in Figure 4.12, the use of a rule-based control, which activates the heating system only when the setpoint varies, affects negatively the peak of the demand. This peak is always located at the time of setpoint change (7 am – from 18 to 22 °C) and it is not affected by the different weather clusters. Instead, the use of a predictive control strategy allows a better distribution of the energy demand through the hours of the day. This leads to a reduced peak and diversification in control for the different weather clusters.

For Rate M, Figure 4.14 shows that the severe weather conditions of Cluster 1 force the system to activate with peaks around $42 \text{ to } 46\ kW_{th}$. The less harsh weather conditions and the improved solar gain prediction allows a better distribution of the building thermal load through the hours of the day, leading to peak values around $26 \text{ to } 30\ kW_{th}$. Furthermore, the values of $RampUp$ and $RampDown$ indices demonstrate the better distribution of the building thermal load. The minimum/maximum fluctuation of the building thermal load is always higher/fewer than $-15/+20$ respectively.

Conversely, Figure 4.15 shows the effect of a critical peak pricing structure (Rate Flex M) on the building consumption. The peak of thermal demand becomes very high due to the price variation, reaching values over $50\ kW_{th}$ for the Cluster 1. The $RampDown$ index is also highly affected by this variation since the thermal load needs to instantaneously approach to zero when the price changes. It is important to note that the $RampUp$ index is still bounded under $20\ kW_{th}$, consequence of the technology penalty used in the cost function, which penalizes the high variation of the demand. Figure 4.16 and Figure 4.17 show the Radiative Ratio (RR) and daily thermal load profiles for different weather clusters and for the different pricing structures. The RR index provides a better understanding of how the heating supply is distributed through the hours of the day. In case of constant price signal (Figure 4.16), preheating is prioritized by the active envelope, which stores heat and release it during the day. The values of RR range between 40 and 70% depending on the cluster (lower outdoor temperature leads to higher preheating), and show minimal influence of the weather cluster in the hours during the setpoint temperature change (6 am to 9 am) and the late hours of the day (8 pm to 12 am). During this period the hourly deviation of the RR index is approximately $\pm 12\%$.

In case of Rate Flex M (Figure 4.17), the RR index maintains values between 55 to 65% during preheating, indicating an even reduced influence of the weather cluster during the early hours of the day. By disregarding the behaviour near the cost variation during extreme weather conditions, the variation of the Radiative Ratio can be assumed to be within a $\pm 5\%$ range, which is even smaller compared to the case with a constant tariff. From this analysis, it is evident that while the weather cluster significantly affects the building thermal load, its impact on the optimal daily RR and its average value is very limited. The trend of the RR index indicates that, for both Rate M and Rate Flex M, ventilation heating is preferred during the change in indoor setpoint temperature (6 am to 8 am) where RR of 20% means 80% of energy provided by the ventilation heating. By optimizing the hourly delivery of thermal energy to both the ventilation terminals and the active envelope (floor heating), better utilization of the building thermal mass is achieved.

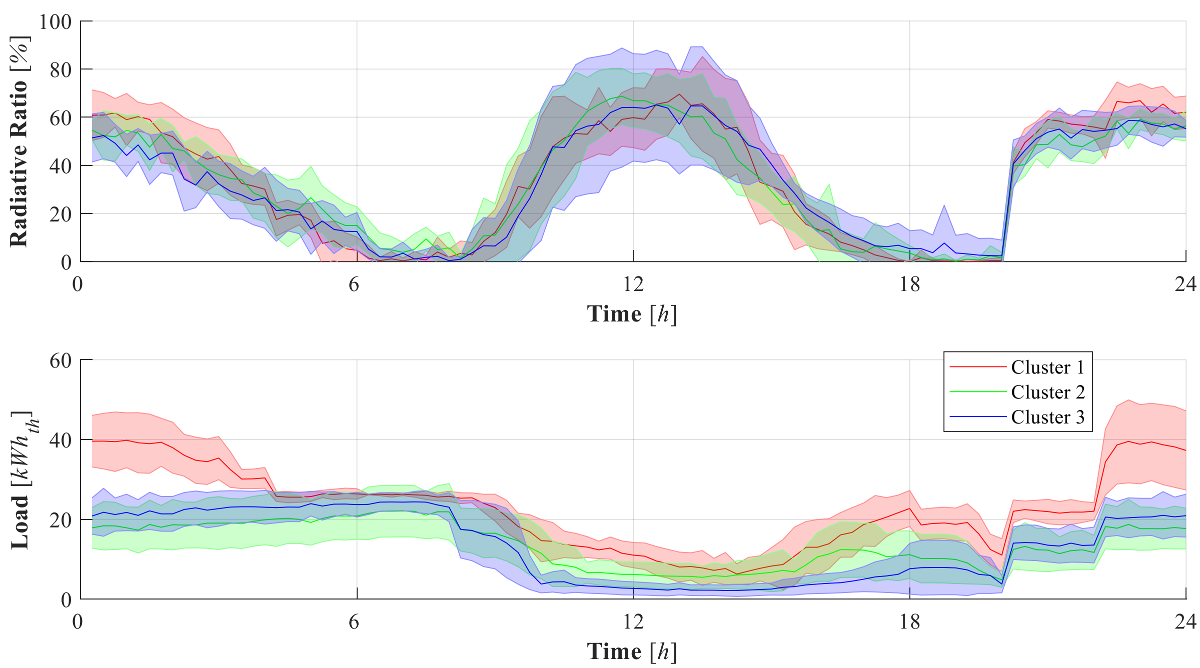


Figure 4.16: *Model predictive control with Rate M*: radiative ratio and thermal demand profiles for different weather clusters.

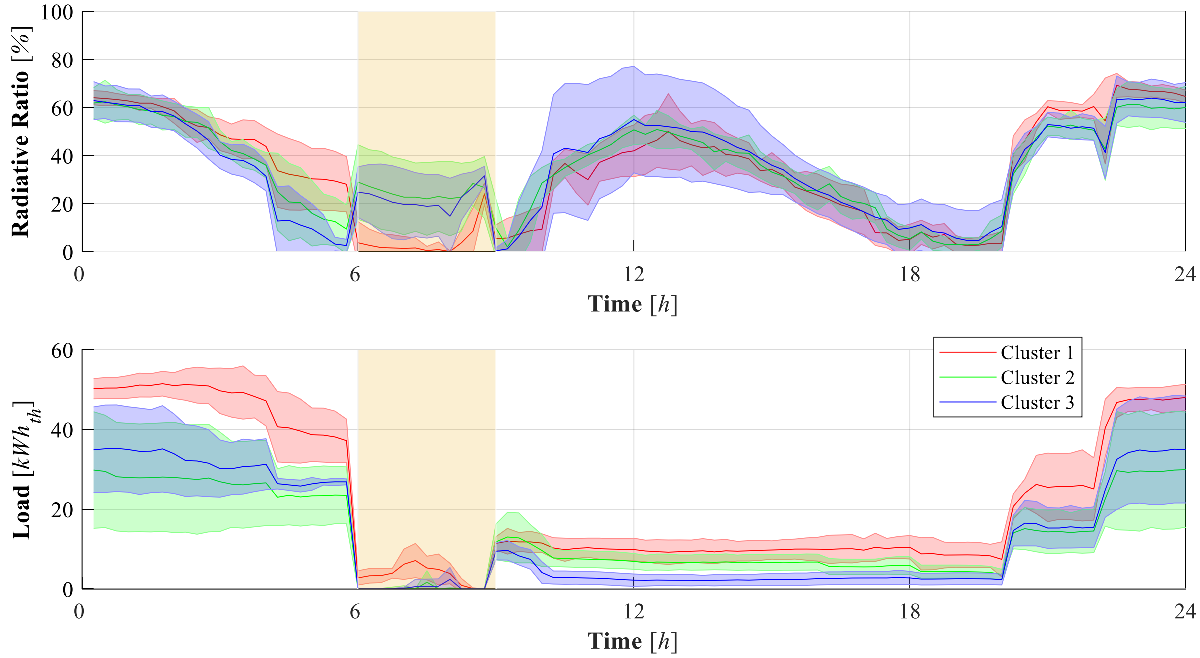


Figure 4.17: *Model predictive control with Rate Flex M*: radiative ratio and thermal demand profiles for different weather clusters (DR in orange area).

4.5.4 Effect of optimal control on the building energy flexibility

This Section quantifies the results obtained from the different pricing signals in terms of energy flexibility. The reference scenario is also reported to show the advantages of the predictive control and the effect of the different pricing signals. The following conclusions are assessed:

- *Load Factor* (LF): this represents one of the most important metrics to evaluate the stress on the grid due to the building load variability. It ranges between 0 and 1, where values closer to 1 means less stress on the grid due to lower variation of the load profile through the day. Numerical results show that the reference scenario provided LF values ranging from 0.20 to 0.60, the constant price signal provided LF values ranging from 0.34 to 0.72, and the demand response scenario provided LF values ranging from approximately 0.22 to 0.5. This metric is not ideal for characterizing the efficiency of profiles with rule-based control, which shows that the LF is optimal for very cold days for the considered building. While this index provides a general idea of the stress on the grid, it does not offer sufficient insights into the actual energy flexibility that the building can provide. In the case of predictive control strategies, the trend of LF shows benefits with a constant price signal like Rate M, but it significantly decreases for Rate Flex M. Ultimately, this metric gives an overall view of the daily load management of the building. If more detailed information (i.e., every time-step, hour) is needed by grid operators, the LF alone is inadequate and needs to be combined with other indicators.
- *Flexibility Factor* (FF): this index ranges between 1 and -1. Values closer to 1 indicates high flexibility, while values closer to -1 indicate low flexibility. Since this indicator evaluates the energy flexibility provided only during periods of high price, there is no need to make a distinction in terms of weather clusters. Numerical results show that the

FF ranges from 0.72 to 0.79 in the case of constant price signals, while it ranges from 0.94 to 1 in the case of Rate Flex M. The results demonstrates that the DR event outperforms the other pricing structure in terms of flexibility from on-peak (selected as ranging from 6 am to 9 am) to off-peak hours. It is important to note that the constant price signal also provides more flexibility than the reference scenario due to the predictive management strategy, which preheats the floor heating during the first hours of the day. This is consequence of the comfort penalty, $J_{\text{comf},k}$, used during the optimization control routine. In fact, the peak hour corresponds to the beginning of the occupied period of the building, which is when the indoor setpoint temperature changes. The *FF* metric is important for focusing on specific parts of the day, distinguishing "low price" from "high price" periods.

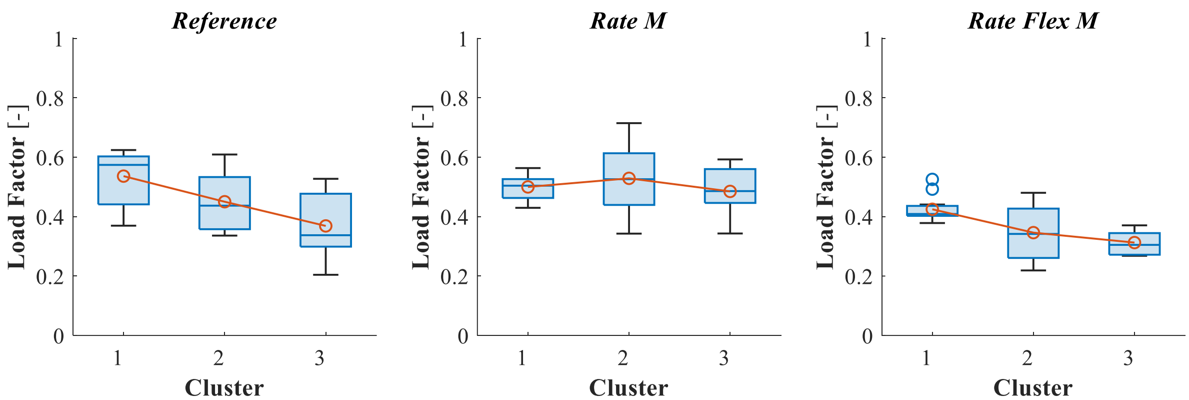


Figure 4.18: Load factor for different weather cluster and control scenarios.

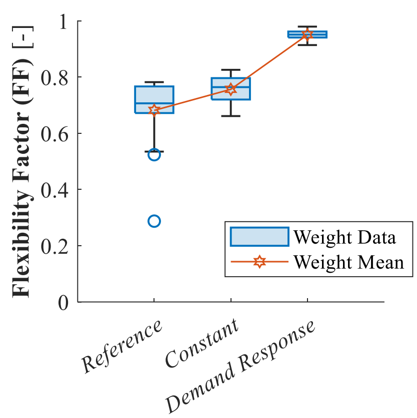


Figure 4.19: Flexibility factor for the different price signals.

Finally, the Building Energy Flexibility Index (*BEFI*) is used to evaluate the difference between the energy requested during the flexible (optimal) scenario and the reference scenario, corresponding to the rule-based control. This metric allows for varying the length of the considered flexibility period and shows, hour by hour, the energy flexibility provided to the grid. With the chosen strategy, the system effectively achieves flexibility during peak demand hours: (i) positive values of *BEFI* indicates the possibility to shift that amount of energy to other hours of the day, (ii) negative values means lower flexibility in those specific hours of the day corresponding to energy demand shifted during those hours. In both constant and DR events the control strategy successfully shifts most of the building energy consumption to off peak hours, thereby reducing the demand during peak hours. According to Figure *4.20*, the proposed methodology allows for a significant portion of the energy consumption to be shifted from the peak to off-peak hours, ultimately contributing to load shifting and, in some cases, to a more balanced load profile. Figure *4.20* illustrates the values of thermal energy flexibility that the building can provide for each hour of the day under different pricing structures. The results indicate that the DR strategy outperforms the other pricing structure during the period 6 am to 9 am, due to the concentrated price variation during that period.

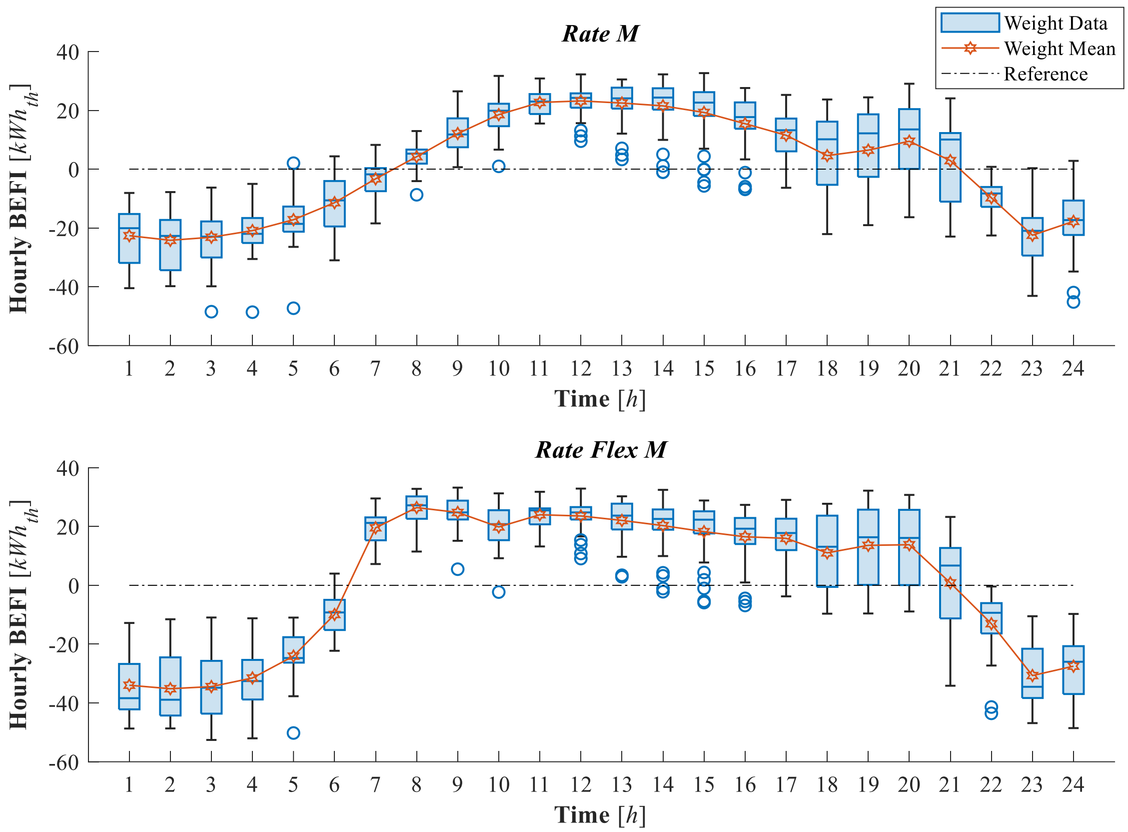


Figure 4.20: Building Energy Flexibility Index for Rate M and Rate Flex M.

4.6 Discussion

The methodology employed for the model structure selection allows the creation of RC thermal networks to study the thermal dynamics in multi-zone buildings characterized by different heating terminals. Applied to an institutional building characterized by three main thermal zones, the selected model structure and the multi-step ahead calibration process captures the building thermal response. The model is calibrated using real data and the physical interpretation of its parameters allows the model to be implemented under different weather and control scenarios.

A methodology based on control-oriented archetypes, combined with a calibration based on real data, can be applied to other case studies, characterized by different thermal zones, heating terminals with different features (i.e., slab thickness, maximum heating supply) and system configuration. The developed methodology only requires data on indoor air temperature, slab temperature and amount of energy provided by each one of the terminals. Therefore, buildings characterized by an increased number of thermal zones may require model order reduction methodologies to further reduce the total number of thermal zones to "dominant thermal zones," as done in [276]. As soon as the "dominant thermal zones" are identified, the use of control-oriented archetypes still represents the final step to create the building thermal model.

The adoption of an economic-Model Predictive Control (e-MPC) framework has proven to be a key strategy for reducing building energy consumption while enhancing the interaction with the grid [138]. In the adopted framework, soft constraints are integrated within the cost function, and includes technology operation, occupant comfort, and economic costs. This integration allows for a balanced approach to address competing objectives, ensuring that the resulting solutions are not only optimal but also feasible. Notably, the use of soft constraints mitigates the risk of convergence issues that are often encountered in optimization problems involving hard constraints.

A significant strength of e-MPC lies in its capability to directly incorporate grid signals so the system can respond dynamically to external signals, improving building-grid interactions and contributing to grid stability. This study demonstrated the effectiveness of the e-MPC framework by analyzing two price signal strategies: constant pricing and demand response or critical peak pricing.

- Constant Pricing: This pricing strategy highlighted the model ability to leverage the building active envelope to exploit its thermal mass. By doing so, the system optimizes the energy performance under steady price conditions while maintaining occupant comfort and minimizing energy costs.
- Critical Peak Pricing: Under this dynamic pricing signal, the model demonstrated its capability to preheat the building during off-peak hours. This proactive strategy reduces energy consumption during critical peak hours, alleviating stress on the grid and enhancing the economic performance of the system.

In the end, a demand response or critical peak pricing structure generally increase peak and load variation. This variation can be mitigated by adopting smoother price signals or by combining thermal load optimization, which exploits the building thermal mass, with the adoption of other storage systems (i.e., batteries). While the building thermal mass may have limitations in peak shaving applications due to its finite capacity and thermal inertia, it always provides potential in load shifting applications.

By accounting for dynamic grid signals and effectively managing building thermal mass, e-MPC not only reduces energy consumption but also promotes greater energy flexibility. This is particularly crucial in modern smart grid environments, where the interaction between buildings and the electrical grid plays a pivotal role in achieving sustainable energy systems.

Results on temperature and heating load profiles show that coordinating different heating terminals leads to reduced energy waste and improved thermal mass utilization, positively affecting the indoor comfort conditions. The use of indicators, such as *RampUp* and *RampDown*

, is important for analyzing the building load profile. These metrics are closely related to the concept of the *rebound effect* [277-279], which needs to be limited to avoid increased energy demand after periods of high electricity price. For this case study, these indices demonstrates that high price variations are generally linked to significant fluctuations in demand throughout the day. The presence of periods with high prices forces the system to supply thermal energy during low-price hours. Applying this same strategy on a large scale may only shift the problem of high peak during other hours of the day, potentially maintaining or worsening the issue of high stress on the grid.

The use of the Radiative Ratio (RR) metric provides information on how the thermal energy required by the building thermal zones is distributed to the different heating terminals. The results show that while the weather clusters significantly affect the building thermal load, their impact on the optimal daily RR and its average value is very limited. This may be a consequence of the high insulation levels of the considered building. Therefore, it is interesting to see how weather cluster can help in generalizing the control scenario, moving towards a more sophisticated scenario-based control. This optimal control may happen by selecting specific trends of setpoint temperature (different for each heating terminal) or by modulating the flow rate delivered to the different heating terminals in case of a centralized system.

Finally, the three energy flexibility metrics, including the *Load Factor, Flexibility Factor* and *Building Energy Flexibility Index*, provide sufficient information for selecting the appropriate control scenario during specific periods. The grid can use the information obtained by adopting this methodology to different buildings and predict the aggregated energy consumption. The selection of a specific tariff needs to be supported by metrics that quantify the effect of the chosen tariff on actual building consumption. Rate Flex M provides the best load profile if the grid target is to shift energy consumption to off-peak hours. However, if applied to many buildings, this profile may only shift peak demand to different times. Studies focusing on the effect of different dynamic tariffs at a larger scale are necessary to evaluate the advantages and disadvantages of selecting a combination of specific tariffs at the community level. Therefore, studies like the proposed one are essential to support the selection of price signals and to understand the contribution that each building can provide during balancing services and load leveling scenarios.

4.7 Conclusions

This study focuses on optimizing the coordination between different building thermal dynamics, specifically ventilation heating and active envelope systems (floor heating). The findings demonstrates that improved coordination enhances indoor thermal conditions and provides greater energy flexibility to the electrical grid.

The proposed methodology utilizes RC model archetypes to build and optimize the building thermal load through model predictive control. By leveraging archetypes, the methodology simplifies model creation and selection, which is particularly valuable for multi-zone buildings with various heating terminals. The methodology studies the building participation under different pricing signals, focusing on shifting load from peak to off-peak hours and providing accurate thermal load forecasting for the utility grid.

The study evaluated various thermal management strategies, comparing rule-based control with MPC strategies under different pricing signals. Results showed that MPC significantly improves thermal comfort and energy efficiency by making better use of the building's thermal mass. In contrast, rule-based control may fail to fully utilize the thermal mass, leading to inefficiencies, especially during high-demand periods.

A notable finding from this study is the minimal variation in the Radiative Ratio (RR) across different weather clusters. Under a constant price signal, the average RR variation was $\pm$ 12%. However, during DR events, this variation decreased to around $\pm$ 5%. The hierarchical clustering method used in this study was crucial in identifying this result, as it allowed for a detailed analysis of weather clusters and their impact on the RR index. This minimal variation underscores the effectiveness of using clusters to simplify the application and coordination of heating terminals in real-world scenarios, making it easier to optimize control strategies and enhance operational efficiency.

The analysis of different pricing structures (Rate M and Rate Flex M) revealed distinct impacts on energy consumption and load management. Under a constant price signal, preheating by the active envelope resulted in a more even distribution of thermal energy. However, DR events caused significant fluctuations in demand due to price variations. During "Very Cold" days (Cluster 1), daily thermal energy demand peaked between 42 to 46 kW_{th}, compared to 26 to 30 kW_{th} in milder conditions. These findings show the importance of strategic energy management to mitigate stress on the electrical grid during high price periods.

Key performance indicators, such as the $RampDown$ and $RampUp$ indices, were critical for analysing building load profiles and ensuring compliance with technological constraints. High price variations led to notable demand fluctuations, with $RampDown$ values reaching -34 kW_{th} during very cold days. This indicates that without careful control, high price variations can shift peak demand to other times, potentially increasing grid stress. Therefore, the building thermal mass, while limited in peak shaving applications, remains essential for load shifting applications.

In summary, the selected KPIs provide valuable insights into the effects of different pricing signals and control strategies on building energy management. The use of predictive control strategies, particularly under constant price signals, led to significant improvements in energy flexibility metrics. The *Flexibility Factor* increased notably during DR events, reflecting enhanced flexibility. The *Load Factor* improved with constant price signals but decreased during DR events due to concentrated price variations. The *Building Energy Flexibility Index* demonstrated the system capability to shift energy consumption effectively during peak hours. These results emphasize the efficiency gains from advanced control strategies and dynamic tariff structures, supporting sustainable energy practices and informing future policy decisions to enhance grid stability and operational efficiency.

Future directions and limitations

This study has several limitations that should be acknowledged. Firstly, the model does not consider the fresh air supply due to occupants as controllable. Instead, it relies on a predefined schedule. Future work could integrate an occupancy estimator to dynamically adjust the fresh air supply based on real-time occupancy data.

Secondly, the current model provides a basic representation of radiative heating. It could be further enhanced by incorporating more detailed aspects of radiative heat transfer and system complexity. Adding detailed modelling of radiative heating elements and their interaction with the building's thermal mass could improve the precision of the thermal dynamics simulation.

Additionally, the clustering approach relies on mean outdoor temperature and total solar radiation as the primary weather variables. This simplification may overlook other important weather factors (i.e., wind speed), which can influence thermal dynamics. Future studies could incorporate a broader range of meteorological variables to enhance the robustness and accuracy of the weather clusters.

Chapter: 5 Clustering-driven Design and Predictive Control of Hybrid PV-Battery Storage Systems for Demand Response in Energy Communities

5.1 Introduction

This chapter presents a comprehensive methodology for selecting typical days and evaluating how controllable building loads influence the design and operation of grid-supportive technologies, specifically photovoltaic (PV) and battery storage systems. Typical days are identified through dynamic time warping (DTW) and hierarchical clustering approaches, supported by six internal validation metrics. Grey-box and regression models are employed to predict building energy consumption, while PV and battery models assess system performance. A two-level Model Predictive Control (MPC) framework is employed to optimize the buildings demand and coordinate the operation of grid-supportive technologies. At the first level, a distributed MPC algorithm manages thermal loads in individual buildings to enable demand response. At the second level, a supervisory MPC optimizes the operation of the hybrid PV-battery storage system to achieve targeted grid flexibility. The case study considers a virtual community in Varennes, Québec, consisting of institutional and residential buildings. Through efficient thermal load management, the methodology shows that community peak demand can be reduced by over 40% compared to current operational practices, and the required capacity of grid-supportive systems can be reduced by up to 26% in a *worst-case* scenario analysis.

In detail, with the introduction of advanced algorithms and improvements in system configurations, the design optimization problem becomes more complex. This complexity arises from large system scales, multi-aspect objectives, dynamic tariff influences, and the co-planning of system size and operation. The adoption of typical days to characterize the building load profiles can simplify these issues, strengthening the link between design and control. Furthermore, many studies focus on conventional optimization methods, simplified models, and static load profiles, overlooking critical aspects such as battery aging, dynamic load estimation, and the interaction between design and control. In this framework, this chapter proposes a data-driven methodology to address these challenges, with the following contributions:

- Dynamic Load Models: Development of data-driven models to predict the building load, studying how utility signals and weather forecasts affect this profile.
- Integrated Design and Control Framework: Investigation of the influence of adjusted load profiles on the optimal sizing and operation of PV and battery systems, accounting for objectives such as flexibility, cost reduction, and resilience.
- Community-Scale Optimization: Analysis of the aggregated load of building community to explore how single-building control strategies impact community-scale operations.

- Connecting Design and Control: Representative weather periods are extracted using dynamic time warping (DTW) and hierarchical clustering, offering insights into reducing technology size while enhancing energy flexibility and building community resilience.

The study demonstrates how coordinated predictive management strategies can effectively balance the dual objectives of economic benefits for end-users and operational optimization for the electrical grid.

The Chapter is divided in several sections. Section 5.2 focuses on the methodology, Section 5.3 describes the case study and Section 5.4 shows the results from model development, to reference and optimal control at single building and micro-grid levels. Sections 5.5 and 5.6 provide a final discussion and conclusions respectively.

5.2 Methodology

This section describes the methods used in this paper. Dynamic time warping (DTW) and hierarchical clustering are employed to select the most representative instance for drawing conclusions from the simulation. The participation level of buildings in demand response programs is assessed by combining data-driven grey-box models and distributed predictive control. The operation of the microgrid, which is characterized by a hybrid photovoltaic-battery system and supports the building community, is modelled using available libraries and controlled through a supervisory control. Section 5.2 is divided into five subsections with main approaches summarized in Figure 5.1.

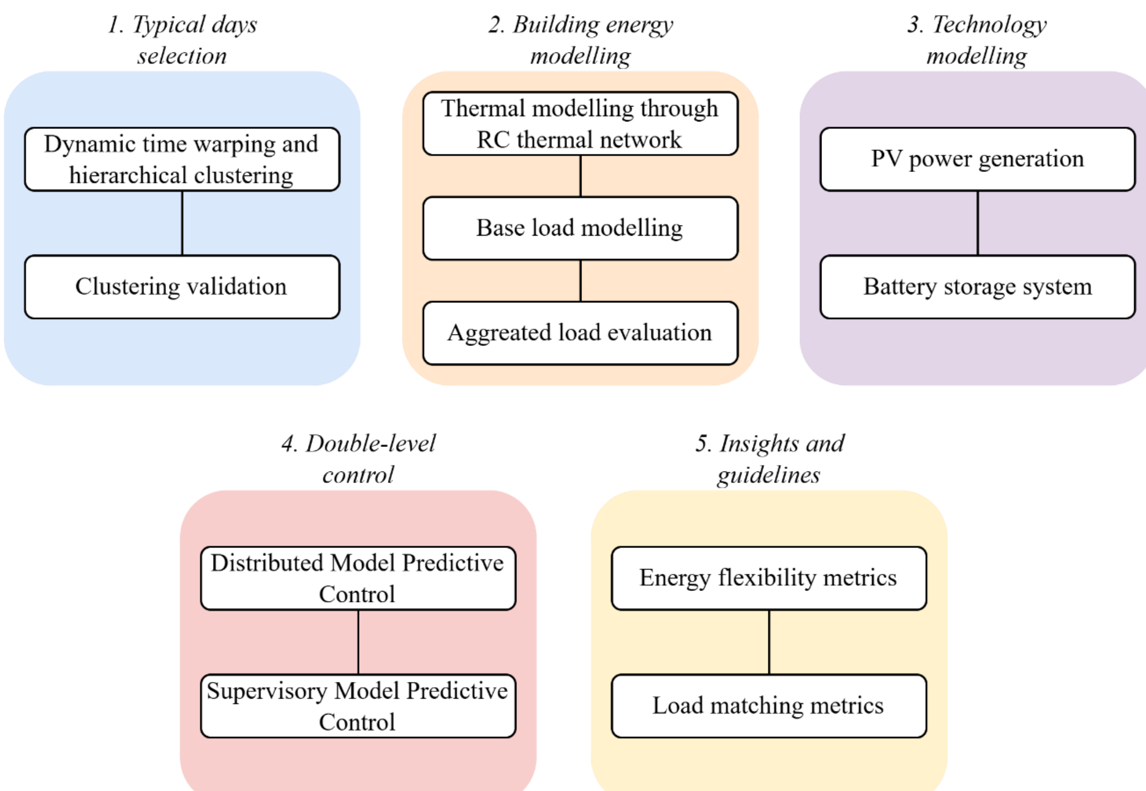


Figure 5.1: Schematic of the adopted methodology.

5.2.1 Typical days selection

Data-driven and control-oriented models are crucial for studying load modifications in grid-interactive applications. These models are typically suited for short-term simulations, and their applicability across different scenarios often requires frequent calibration [280]. Furthermore, the adoption of advanced control strategies, such as model predictive control, increases the computational burden and questions their feasibility for design applications.

To address these challenges, this paper employs an unsupervised learning method (hierarchical clustering) to identify the most representative instance for drawing conclusions from the simulation [211, 281]. As an unsupervised learning task, it is essential to validate the quality of the clustering results. Clustering validation has long been recognized as a key factor for the success of clustering applications. Clustering validation can be broadly categorized into external and internal validation. The primary distinction between these two approaches lies in the use (or lack thereof) of external information. External validation measures rely on knowledge of the "true" number of clusters in advance and are typically used to select the optimal clustering algorithm for a specific dataset. In contrast, internal validation measures do not require additional information and can be used both to select the best clustering algorithm and determine the optimal number of clusters.

This paper uses dynamic time warping (DTW) to evaluate the distances between each observation. These distances are then clustered using hierarchical cluster tree with *ward* method – which is an inner squared distance metric to minimize the cluster variance. The DTW metric is ideal for application with not-aligned time series, while studies employing data with equal length or shape comparison applications can rely on different metrics, such as Euclidean distance and correlation distance respectively. In this paper, the quality of each cluster is evaluated using cluster validity indices (CVIs). The final clustering solution and its corresponding number of clusters are determined by selecting the partition that yields the best CVIs values. This is made possible through the use of the CVIK toolbox, a MatLab-based tool for automatic data clustering [282]. This tool is implemented using the representation (b) in Figure *5.2*, following an internal validation process. However, the wide variety of existing CVIs presents a challenge in selecting one that adequately reflects the underlying structure of the input data. The selection of the CVIs used in this study was guided by the most relevant metrics in literature [283]. Seven internal validation indices—Silhouette [284], Davies-Bouldin [285], S-Dbw, Calinski-Harabasz, Xie-Beni [286], and Dunn's index [287]—were used to determine the optimal number of clusters.

Key factors influencing building energy demand include occupancy, outdoor temperature, and solar radiation, with this study primarily focusing on the latter two. To this concern, outdoor temperature and solar radiation data are clustered separately to capture their distinct trends and impact on building performance and solar technologies operation. This separation ensures that the trends in each variable are clustered by considering its temporal variation. Based on the identified clusters from combinations of these two variables, typical days are selected to extract representative periods from the overall period of study. These will represent the most characteristic instance for drawing conclusions from the simulation.

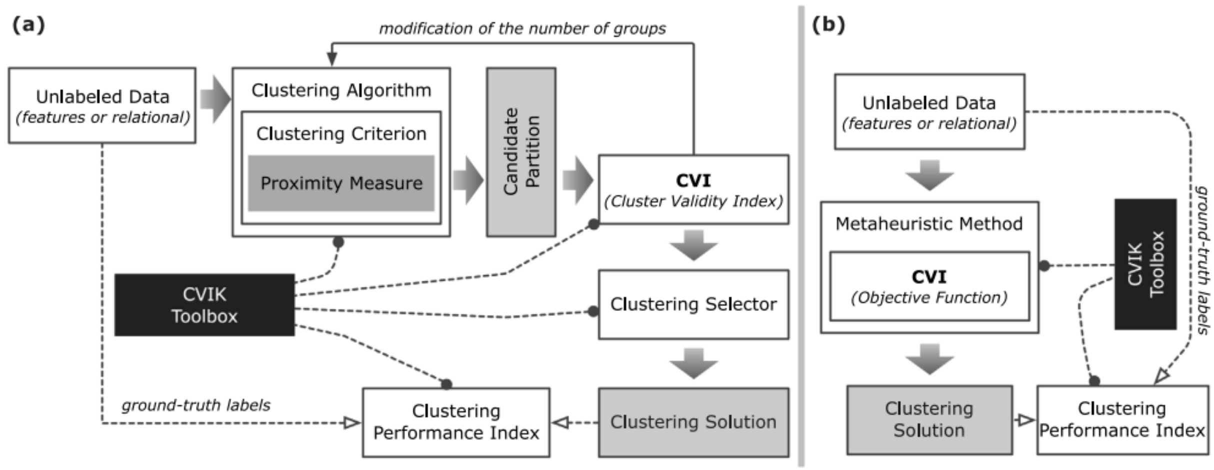


Figure 5.2: Integration of the CVIK Toolbox into the clustering task pipeline to determine the number of clusters automatically. In (a), CVIK's indices are used as external validation functions. In (b), CVIK's indices are internal clustering criteria (Image from [282]).

5.2.2 Building energy modelling

The building load comprises various components, including thermal, miscellaneous equipment, and domestic hot water loads. This study distinguishes between controllable and uncontrollable loads, evaluating each type using distinct methods. Controllable loads can be managed or varied based on signals from utilities or users.

Thermal loads are considered controllable in this study by varying the heating input and leveraging buildings thermal mass. The buildings thermal dynamics is modelled using Resistance Capacitance (RC) thermal networks, which represent the building as a simplified circuit of resistances and capacitances that describe heat transfer and storage characteristics [81]. These models allow the study of indoor temperature dynamics considering exterior disturbances (e.g., solar radiation, ambient temperature), occupancy patterns, and internal gains (e.g., equipment or lighting). Additionally, they allow the inclusion of heating inputs as control variables, enabling the prediction and optimization of energy use while maintaining occupant comfort. The RC thermal model of the i-th building can be represented with this state-space formulation:

$$T_i(t) = A_i \cdot T_i(t-1) + B_i \cdot u_i(t) \tag{5.1}$$

where T_i represents the indoor air temperature of the i-th building [°C], A_i is its state matrix, B_i is the input to state matrix, and u are the inputs. The input variables include the outdoor temperature, solar radiation and the thermal power, Q_i [W]. More detail on the creation of these models is available in previously published studies on automated building modelling from smart thermostat data [276, 288] and RC model archetypes [100].

Uncontrollable loads, by contrast, follow predefined schedules and include components like mechanical systems, lighting, and ancillary loads. These loads consist of both variable and fixed components. The variable component is estimated with regression-based models, with the controllable load as an input. This approach draws on measurements and catalogue data. In contrast, the fixed component is estimated by considering an average daily profile obtained from the available dataset.

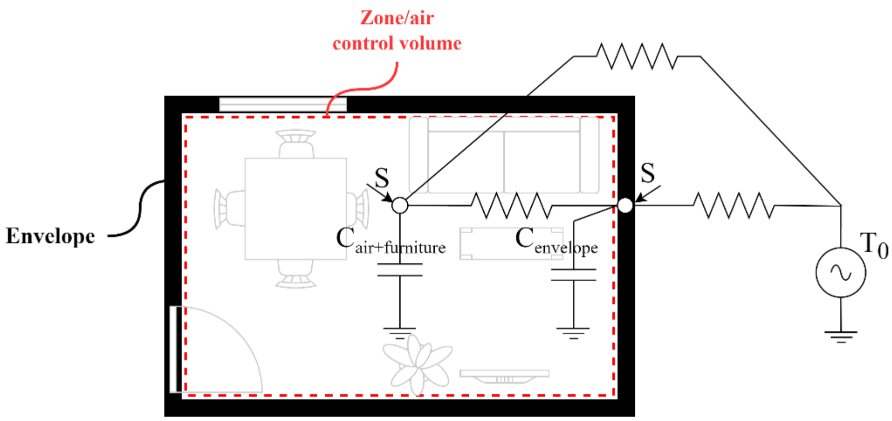


Figure 5.3: Typical 2C3R representation of the building thermal dynamic (Image from [276]).

Overall, the energy demand of the i-th building can be calculated with the following equations, comprising of the controllable and uncontrollable component.

$$P_{c,i} = \eta_{th,i} \cdot Q_i \tag{5.2}$$

$$P_{u,i} = P_{u,f,i} + P_{u,v,i} = P_{u,f,i} + f_{reg,i}\left(Q_i\right) \tag{5.3}$$

where $P_{c,i}$ is the controllable load $[W]$ (or $[Sm^3]$ according to the application), $\eta_{th,i}$ is the conversion factor and Q_i is the heating input $[W]$. $P_{u,i}$ is the uncontrollable load $[W]$ with $P_{u,f,i}$ and $P_{u,v,i}$ the fixed and variable components respectively $[W]$, and $f_{reg,i}$ represent the regression model used to calculate the uncontrollable component function of the building thermal demand.

By evaluating the fixed load profile, and the RC and regression models it is possible to characterize the total load of each building. In case of N full-electric buildings, creating a community, the aggregated load can be calculated as $P_{agg} = \sum_{i=1}^{N}\left(P_{c,i} + P_{u,i}\right)$ $[W]$.

5.2.3 Technology modelling: Hybrid PV-battery storage system

This study considers a hybrid PV-battery storage system. The PV technology is modelled using PVLIB, an open-source software providing photovoltaic performance modelling for MatLab and Python [254]. PVLIB enables accurate simulation of PV system performance by incorporating input data such as diffuse and beam solar radiation, wind speed, tilt angle, and solar angles. Additionally, PVLIB allows users to select performance curves for PV panels from an extensive database. The key formulations are presented below:

$$P_{PV} = A_{PV} \cdot I_{tot} \cdot \eta_{PV} \cdot \eta_{other} \tag{5.4}$$

$$I_{tot} = f\left(\text{tilt angle, solar angles}, I_b, I_d\right) \tag{5.5}$$

$$\eta_{PV} = f\left(P_{MPP,STC}, C_{t,P}, T_{cell}\right) \tag{5.6}$$

With P_{PV} representing the PV power generation [*W*], A_{PV} the PV area [*m*^{2}], η_{PV} corresponding to the module efficiency and η_{other} corresponding to all the other losses of the plant (i.e., transmission system, inverter). I_{tot} is the total irradiation [*W/ m*^{2}] which is function of the PV tilt angle, solar angles and its beam [*W/ m*^{2}] and diffuse [*W/ m*^{2}] components, I_b and I_d. Finally, $P_{MPP,STC}$ stands for the maximum power production [*W*] in standard test conditions, $C_{t,P}$ represents the coefficient for power variation and T_{cell} the cell temperature [°*C*]. The functions assess photovoltaic cell efficiency by evaluating cell temperature. By accounting for incident solar radiation on the photovoltaic panel at a given tilt angle, these functions compute the electricity output power of individual PV modules. Finally, multiplying this output by the area of the PV field yields the total electricity production.

The battery storage serves as a back-up, enhancing system reliability and significantly reducing the burden on the grid. The maximum capacity of the battery and the size of the photovoltaic field are critical variables to determine during the design phase. During operation, battery charging occurs when there is excess power produced by the PV system or in anticipation of high price or demand response (DR) events. Conversely, discharging typically occurs during periods of high prices or peak consumption. The battery can supply electrical energy to the buildings or store energy coming from the PV system and the electrical grid. These operations affects the battery state of charge, SoC, which is bounded by a minimum and maximum value, $SoC_{\min}$ and $SoC_{\max}$, respectively, based on its depth of discharge, DoD. The battery charging and discharging processes are dynamically modelled using Equation (5.7):

$$SoC(t) = SoC(t-1)\cdot\left(1-\delta\right) + \frac{\eta_{CH}}{C_{NOM}}\cdot P_{CH}(t)\cdot\Delta t - \frac{1}{\eta_{DIS}C_{NOM}}\cdot P_{DIS}(t)\cdot\Delta t \tag{5.7}$$

where t is the time step, δ is the self-discharge rate of the battery, η_{CH} and η_{DIS} are the conversion efficiency of the battery system during the charging and discharging phases, respectively, C_{NOM} is the battery nominal capacity [*Wh*], P_{CH} is the power in input to the battery [*W*], and P_{DIS} is the power in output from the battery [*W*].

5.2.4 Control formulation

To optimize energy management within the building community supported by the hybrid PV-battery microgrid, a two level control strategy is implemented, consisting of a distributed control at the first level and a supervisory control at the second level (Figure 5.4). Both levels use a Model Predictive Control (MPC) routine. The objective of the MPC controller is to minimize a certain cost function over a fixed prediction horizon, ph, while respecting the specified constraints. In building applications, the strength of MPC lies in utilizing a mathematical model to predict the future response of the building. These predictions allow MPC to select control actions that align with the set objectives, systematically balancing comfort requirements, technological constraints, and weather forecasts [138]. The optimization routine thus accounts for the future impact of disturbance variables on system dynamics, applying control variable values within a fixed control horizon, ch.

The first level control strategy, represented in Figure 5.4, employs an economic Model Predictive Control (e-MPC) which is applied individually in each building of the community. In this formulation, the building thermal dynamic with the variables in input, corresponding to the weather forecasts and the heating input, are used to evaluate the optimal controllable load profile of each building over the considered prediction horizon.

The cost function of this local MPC penalizes energy consumption during periods of high prices or demand response events, while still ensuring indoor temperatures in specific bounds and accounting for building thermal dynamic and occupant preferences. The optimization routine can be described by the following equations:

$$\min_{Q_i(\kappa),\dots,Q_i(\kappa+ph)} \sum_{t=\kappa}^{\kappa+ph} \left(J_{economic,i}(t) + J_{comfort,i}(t) \right)$$
$$\text{Subject to} \begin{cases} T_i(t) = A_i \cdot T_i(t-1) + B_i \cdot u_i(t) \\ u_i(t) = \left[T_{out}(t),\ S(t),\ Q_i(t) \right] \\ 0 \le Q_i(t) \le Q_{i,\max} \end{cases} \tag{5.8}$$

where:

$$J_{economic,i}(t) = c_{energy}(t) \cdot \eta_i \cdot Q_i(t) \cdot \Delta t \tag{5.9}$$

$$J_{comfort,i}(t) = \begin{cases} 0 & \textit{if} \quad T_i(t) > T_{set,i}(t) - \beta(t) \ \text{ and } \ T_i(t) < T_{set,i}(t) + \beta(t) \\ \lambda_i \cdot \left(T_i(t) - T_{set,i}(t) \right)^2 & \textit{else} \end{cases} \tag{5.10}$$

In this formulation, c_{energy} is the energy cost *[\$/kWh$_{el}$]* (or [*\$/Sm^3*]), T_i is the indoor air temperature of the i-th building [*°C*], $T_{set,i}$ is the setpoint temperature of the i-th building [*°C*], and β is the bound applied to the setpoint variation [*°C*], which provides flexibility to the indoor temperature variation through the hours of the day.

The dual penalties for energy consumption and comfort create an optimization framework that balances these competing objectives, guiding the controller to find an effective compromise between "comfort" and cost reduction. The comfort cost function allows flexibility in indoor temperature fluctuations during periods of low setpoint values, enabling the indoor temperature to approach or exceed the setpoint as conditions permit. The weight factor, λ, equal to 1 [*\$/°C*], influences the prioritization of comfort relative to the economic term in the cost function [111]. The resulting load profile, $P_{c,i}$, as input and control variable of the optimization related to the i-th building, maintains occupant comfort while contributing to the DR program.

The second-level control strategy, on the other hand, coordinates microgrid resources (PV production and battery storage) to alleviate grid stress during peak periods. Using an MPC approach, the cost function of this level aims to minimize consumption at specific times of the day by factoring in PV production and the aggregated demand from all buildings. This supervisory control penalizes electricity use during demand response periods and peak creation from battery charging, while the battery system is managed to meet predefined SoC_{target} through the day based on PV availability or grid signals. In the absence of grid signals, the cost function discourages excessive battery usage by applying penalties for frequent SoC fluctuations, promoting efficient battery use and mitigating degradation. The cost formulation is described in the following equations:

$$\min_{P_{CH/DIS}(\kappa),\ldots,P_{CH/DIS}(\kappa+ph)} \left(\sum_{t=\kappa}^{\kappa+ph} \left(J_{grid}(t) \right) + J_{peak} + J_{SoC} \right) \tag{5.11}$$

where:

$$\begin{aligned} J_{grid}(t) &= c_{en,el}(t) \cdot P_{grid}(t) \cdot \Delta t \\ &= c_{en,el}(t) \cdot \left(P_{agg}(t) + P_{CH}(t) - P_{DIS}(t) - P_{PV}(t) \right) \cdot \Delta t \end{aligned} \tag{5.12}$$

$$J_{peak} = c_{peak} \cdot \left(\max_{t \in [\kappa, \kappa+ph]} P_{grid}(t) \right) \tag{5.13}$$

$$J_{SoC} = \psi \cdot \left| SoC_{target}\left(t^{*}\right) - SoC\left(t^{*}\right) \right| \tag{5.14}$$

During the optimization, the variables P_{CH} and P_{DIS}, associated with charging and discharging conditions respectively, are controlled to optimize battery operation, represented by the state variable, SoC, which is maintained between minimum, SoC_{min}, and maximum, SoC_{max}, values. The charging condition is defined as the difference between the target SoC_{target} that the battery should achieve at the time step of interest, t^{*}, and the SoC at that time step, multiplied by a weight factor, ψ. Furthermore, during the operation, the peak of energy consumption is penalized by c_{peak} [*$/kW*].

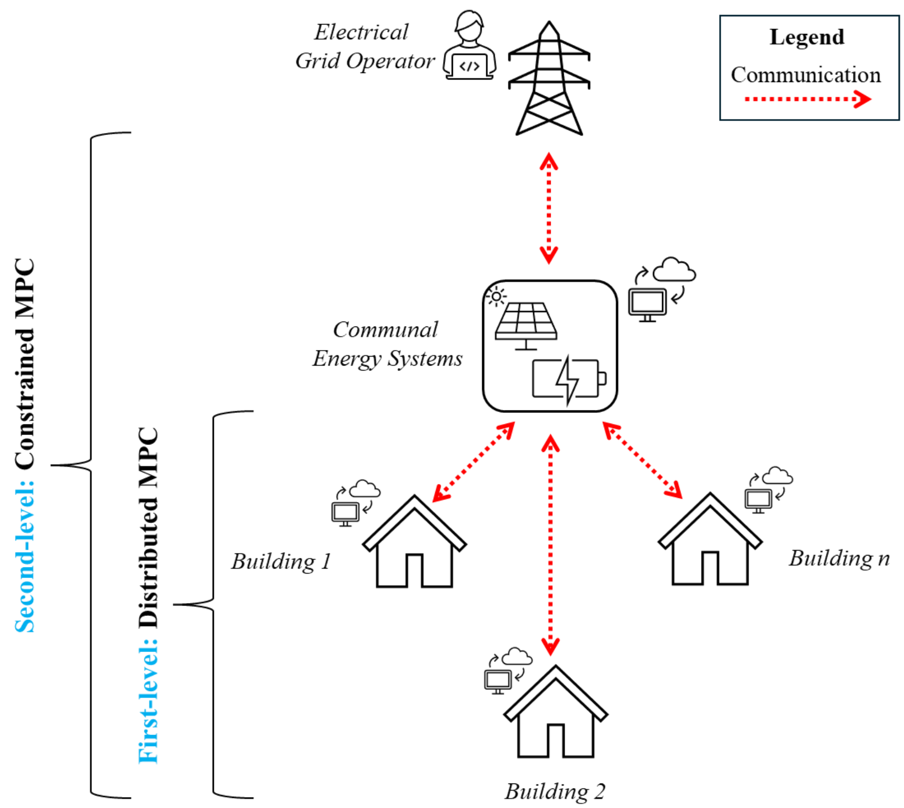


Figure 5.4: Schematic of the first and second level control.

Figure *5.5* shows that four primary SoC_{target} are identified through the day and are evaluated using the following equation: one for the end day, one before the demand response period, one in the evening during the solar radiation hours, and another for all other periods.

$$SoC_{target}\left(t^{*}\right)=\begin{cases} SoC_{end-day} & \mathit{if}\ \text{end of the day} \\ \dfrac{\sum_{t^{*}}^{\kappa+ph} P_{agg}(t)}{C_{NOM}}+SoC_{end-day} & \mathit{elseif}\ \text{start of a demand response event} \\ \dfrac{\sum_{\kappa}^{t^{*}} P_{PV}(t)}{C_{NOM}}+SoC_{end-day} & \mathit{elseif}\ \text{excessive photovoltaic generation} \\ SoC(t-1) & \mathit{else} \end{cases} \tag{5.15}$$

The SoC_{target} is adjusted based on what the optimizer anticipates during the optimization process and may vary according to projected consumption and PV generation levels within the prediction horizon. By coordinating the energy profiles of buildings participating in the demand response program and optimizing the combined load through available energy resources, the supervisory control provides a more efficient management of community-wide demand. This strategic approach leverages both local optimization and community-level load control, enhancing grid stability and reliability during demand response periods.

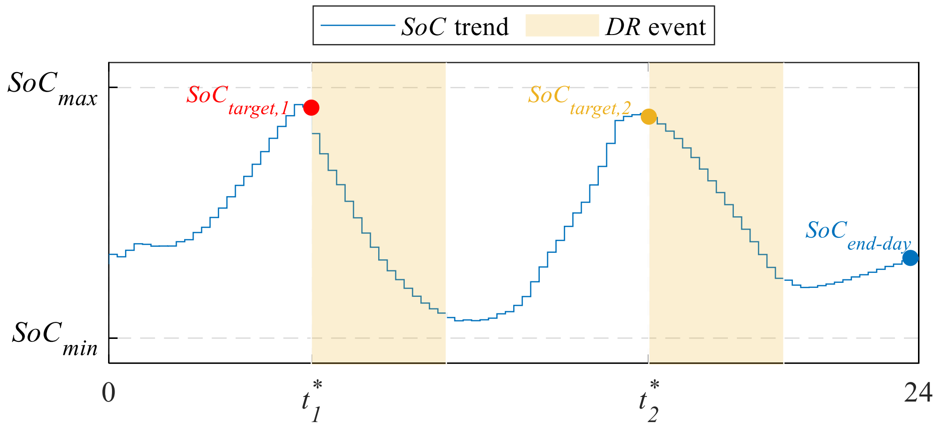


Figure 5.5: Example of optimal SoC trend according to specific targets, SoC_{target}.

5.2.5 Key performance indicators

The performance of the load management and the impact of varying design parameters are evaluated through the following key performance indicators (KPIs) that primarily focus on energy flexibility and grid interaction. The Building Energy Flexibility Index ($BEFI$), recognized as a baseline-required metric [55], quantifies the difference between energy consumption during a well-defined reference (baseline) scenario and a flexible scenario over a specific period, $\Delta\kappa$ [231]. Two primary variants of this metric are considered:

$$BEFI_{kWh} = \sum_{\kappa}^{\kappa+\Delta\kappa} \left(P_{agg}\right)_{ref} - \sum_{\kappa}^{\kappa+\Delta\kappa} \left(P_{agg}\right)_{flex} \tag{5.16}$$

$$BEFI_{\%} = \frac{BEFI_{kWh}}{\sum_{\kappa}^{\kappa+\Delta\kappa} \left(P_{agg}\right)_{ref}} \cdot 100 \tag{5.17}$$

This metric offers a comprehensive evaluation of the energy shifted or reduced during a given period and identifies the specific hours of the day when these changes occur.

The Loss of load probability (LLP) is considered as a load matching metric and represents the frequency or magnitude of a system inability to meet load demand. This metric can also indicate the mean percentage of the load unmet by the installed system. LLP is calculated as the ratio of the total energy deficit – which the grid should supply – to the total load demand over a defined period [225, 226]. It can be expressed as:

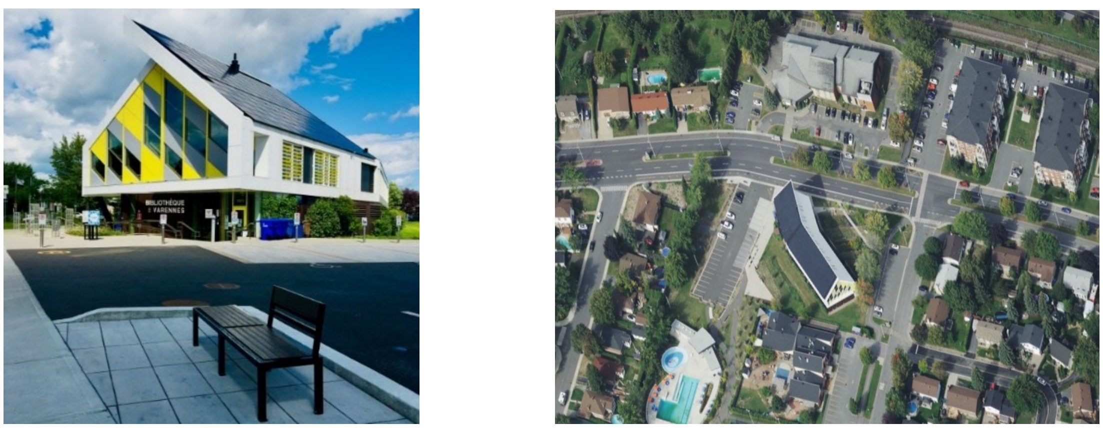
$$LLP = \frac{\sum_{\kappa}^{\kappa+\Delta\kappa} P_{grid}}{\sum_{\kappa}^{\kappa+\Delta\kappa} P_{agg}} \tag{5.18}$$

Both $BEFI$ and LLP provide valuable insights into the operational and design implications of control strategies and technology sizing, particularly in the context of building-grid interactions.

5.3 Case study

The case study features a virtual community incorporating the Varennes' Library, recognised as Canada's first institutional Net Zero Energy Building, along with residential buildings whose data has been provided by Hydro-Quebec – the public utility in Québec, Canada. Located in Varennes, QC, the library is a sustainable municipal facility developed through a collaboration between the municipality, CanmetENERGY (Natural Resources Canada), and Concordia University. Spanning 2000 m² across two floors and a basement, the library features a 110-kWe building-integrated photovoltaic system, with one-sixth functioning as building-integrated photovoltaic thermal collectors. Its energy systems include four ground-source heat pumps (system seasonal COP=4.5 during winter [221]) connected to eight boreholes, which provide heat/cool through the air handling unit and the floor heating/cooling system covering 34% of the net surface area. A building automation system facilitates monitoring and dataset extraction for analysis and testing. Further details about the building are available in [161].

The residential house dataset contains detailed information on four residential buildings. This dataset includes data on actual temperature measurements, setpoint temperature records, heating energy consumption from baseboard heaters ($\eta_{th} \approx 1$), and electricity usage from other loads such as domestic hot water systems, lighting, and other appliances.



(a) (b)

Figure 5.6: (a) the Varennes' library, a sustainable Net Zero Institutional Building, (b) a map of a virtual community in Varennes imported from Google Maps.

For all the buildings, the thermal load and associated technologies are modelled to account for the effects of weather forecasts and desired set point profiles. This thermal load represents the only controllable load at the individual building level in this study. For the reference operation, the indoor temperature set point, T_{set}, is fixed to 20°C from 6:00 to 20:00 for residential buildings [103] and to 22°C from 7:00 to 21:00 for institutional buildings. In all the other hours of the day, the set point is 18°C. The adopted set point values reflect a common trend in residential and institutional buildings in the region of Québec. In contrast, reference profiles based on fixed trends or regression models will be applied for all other loads (base loads) of the buildings. The microgrid, equipped with a hybrid PV-battery storage system, plays a critical role in managing the energy flow within the community during demand response events. Key properties of the system, including the photovoltaic module specifications, battery capacity, efficiency, and operational constraints, are summarized in Table 5.1.

Table 5.1: Technology properties of the hybrid PV-battery system.

Technology	**Variable**	
Photovoltaic field	*Type*	Canadian Solar CS5P-220M
	Installed kW_p	[10:120]
	β	30°
	η_{loss}	0.98
Battery storage	*Type*	Lithium-ion
	C_{NOM}	[10:300]
	η_{CH} / η_{DIS}	0.98
	δ	0.1% a day
	DoD	0.70
	SoC_{min}	0.20
	SoC_{max}	0.90
	$SoC_{end-day}$	0.60

The economic parameters used in this study, presented in Table 5.2, play an important role in the operation of the distributed and supervisory MPC routines. These include electricity costs and incentive structures provided by Hydro-Québec, with high rewards during peak hours (demand response) [289]. Demand response periods are defined to encourage energy reduction or shifting during times of high grid stress. The start and duration of the demand response event affects the thermal load demand, and the SoC_{target} values with the associated time step of interest, t^*. According to the high price periods, for this application, two values of t^* are identified and are 6:00 and 16:00 respectively. Finally, to further support grid reliability, incentives are offered to motivate buildings participating in the program to actively adjust their energy usage patterns. Table 5.2 also provides the values of the cost function peak penalty related to the supervisory control that optimizes the battery operation.

Table 5.2: Economic features used during the optimization routine.

Tariffs [289]	**Electricity price** *[$/kWh]*	**Demand response/High price periods**	**Energy shifted reward** *[$/kWh]*	***Cost function* peak penalty** *[$/kW]*
Institutional Buildings: Rate M	0.058	6:00 to 9:00 16:00 to 20:00	0.58 for severe and very cold days 0.02 for others	16.139
Residential Buildings: Rate D	0.067	6:00 to 9:00 16:00 to 20:00	0.55 for severe and very cold days 0.03 for others	

5.4 Results

This section presents the results of a case study applied to the period December 01, 2019, to March 21, 2020, which covers the winter season. The study shows the importance of thermal load management in enabling buildings to balance energy consumption and mitigate typical peak demand patterns. The findings are organized into five subsections to guide the reader through the distinct aspects of the analysis. In detail, Section 5.4.1 outlines the selection of typical days and the identified clusters, Section 5.4.2 details the development of the building energy models, Section 5.4.3 presents the results of the distributed control at the local level, Section 5.4.4 discusses the outcomes of the supervisory control, highlighting the optimal operation of the technologies, and Section 5.4.5 discusses how the optimal coordination between local and supervisory control affect the optimal size of the technologies during the *worst-case* scenario.

5.4.1 Typical days selection

The typical days selection process involved clustering weather data (from December 01, 2019, to March 21, 2020) to define the most representative periods in which to run the analysis. For outdoor temperature, the CVIs are suggesting two clusters as a viable solution. Therefore, given the similarity between the values of the indicators for two and three clusters (Table B.*1* in Appendix B), the latter are identified as viable option, balancing representativeness and computational efficiency. The S-Dbw index, while effective at detecting subclusters [283], often suggested a higher number of clusters, which was not feasible for this application due to the computational cost outweighing the benefits of increased detail in the clustering process. In this paper, *Severe Cold*, *Very Cold* and *Cold* outdoor temperatures are considered. For solar radiation, the results were suggesting the selection of two clusters (mid-high and mid-low radiation). This hypothesis may be acceptable but given the presence of solar technologies the selected number of clusters was improved to three. This provides a better representation of the different radiation trends. In this paper, *Low*, *Mid* and *High* solar radiation levels are considered. More information about the values of the internal validation metrics is provided in Table B.*2* in Appendix B.

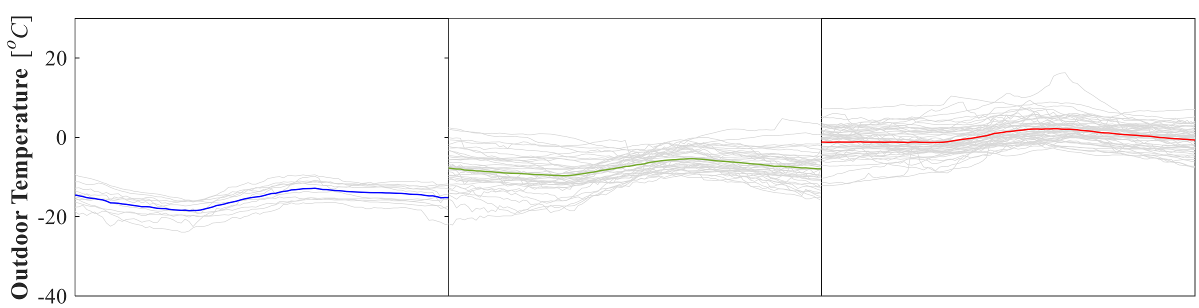


Figure 5.7: Selected number of clusters for outdoor temperature predictions.

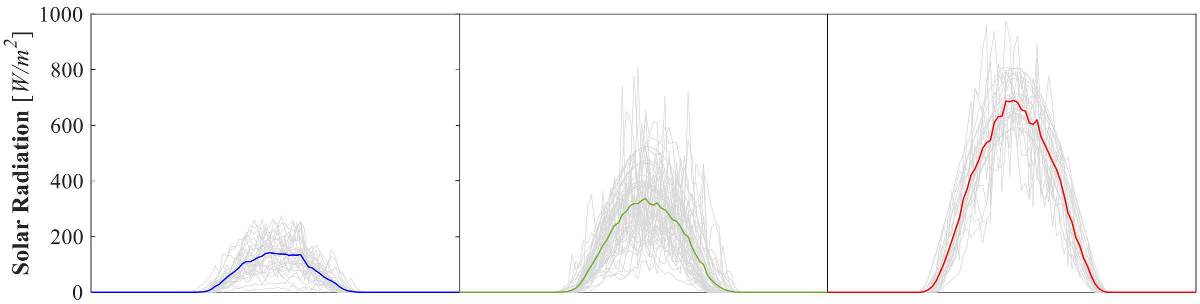


Figure 5.8: Selected number of cluster for solar radiation predictions.

Combining three clusters for outdoor temperature with three clusters for solar radiation resulted in nine typical days. These days are used to define nine representative periods in which to assess the analysis. This approach provides a robust foundation, not only to simplify monthly simulations to a limited number of days, but also to study thermal load management and microgrid performance under diverse yet computationally manageable weather conditions.

Table 5.3 presents the selected typical days, identified based on outdoor temperature and solar radiation clustering. The table includes the occurrences and weights of each typical day over the analysed period. The representative periods in which to run the analysis are determined using the medoids of the identified outdoor temperature and solar radiation clusters while considering a *warm-up* period, which accounts for outdoor temperature and solar radiation in the previous day, limiting the influence of initial conditions in the optimization routine [290].

Table 5.3: Features of the most representative periods used in this paper.

		Cluster of Outdoor Temperature	**Cluster of Solar Radiation**	**Days in selected period**	**Weight**
#1	*Severe Cold-Low*	Severe Cold	Low	2	0.018
#2	*Severe Cold-Mid*	Severe Cold	Mid	6	0.055
#3	*Severe Cold-High*	Severe Cold	High	4	0.037
#4	*Very Cold-Low*	Very Cold	Low	9	0.082
#5	*Very Cold-Mid*	Very Cold	Mid	24	0.220
#6	*Very Cold-High*	Very Cold	High	9	0.082
#7	*Cold-Low*	Cold	Low	20	0.183
#8	*Cold-Mid*	Cold	Mid	22	0.202
#9	*Cold-High*	Cold	High	13	0.119

5.4.2 Building energy models

This section presents the results of thermal modelling for each building, focusing on performance metrics. As shown in Table 5.4, predictions with an RMSE for 24-hrs ahead for both calibration and validation of the models, over different periods, are below 1°C and are considered acceptable. Figure 5.9 illustrates the predicted indoor temperature of House #01, showing the overall capability of the model to follow the building thermal dynamics. The RC models estimate indoor air temperature based on weather forecasts and thermal energy in input with enough accuracy. Subsequent sections examine how thermal energy distribution throughout the day impacts preheating and indoor temperature conditions. Leveraging building thermal mass, with indoor air environments typically having time constants of 1–3 hours [106], allows for thermal energy shifting while maintaining indoor temperatures into acceptable bounds.

Table 5.4: RC thermal model structure and performance for the different buildings.

Building	**Structure**	**RMSE** *[°C]*	
		Calibration	**Validation**
Varennes' Library	8C16R	0.221	0.494
House #01	2C3R	0.503	0.656
House #02	4C7R	0.405	0.633
House #03	2C3R	0.585	0.727
House #04	2C3R	0.642	0.745

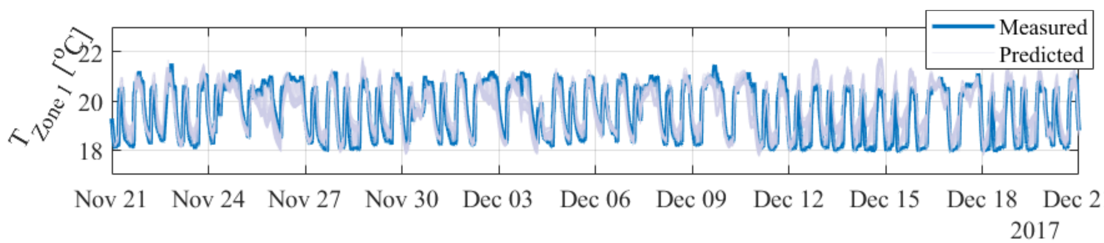


Figure 5.9: House #01 measured and predicted indoor temperature trend for each time step with predictions of 24-hours ahead during the validation period.

In terms of base load modelling, for the institutional building, a regression model is employed to estimate fan coils electrical consumption. The model is generated by considering as inputs the fresh air flow rate and the heating input required by the system. Figure *5.10* shows the comparison between the predicted and measured values, considering a ±20% bound around the ideal model. For the fans located in the dedicated outdoor-air handling unit, catalogue curves are employed. For circulating pumps in the mechanical system, using a linear regression model can lead to significant errors due to the cascade operation of the four ground-source heat pumps. The pumps electricity consumption will depend on specific ranges on building's thermal demand. Finally, other loads, such as lighting and miscellaneous equipment, are represented using a constant profile.

Instead, for residential buildings, energy consumption measurements enabled the generation of two daily profiles accounting for uncontrollable loads—one for weekdays and another for weekends. In this study, the weekday profile is used to evaluate, in combination with the thermal energy demand, the total electrical load of each residential building.

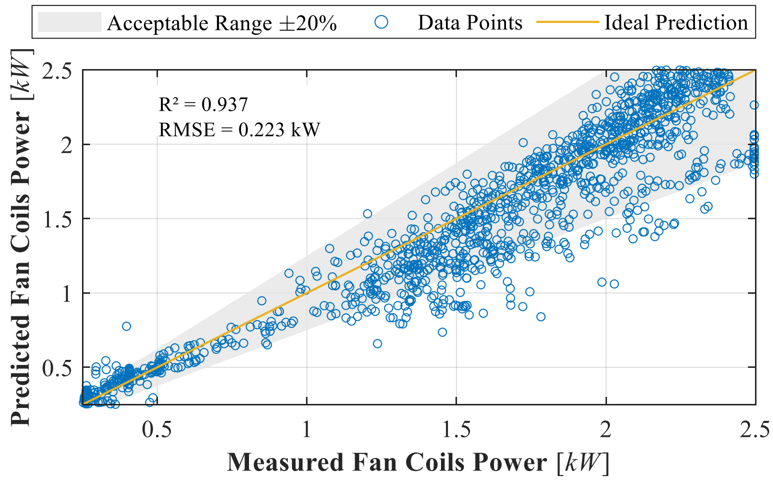


Figure 5.10: Comparison between measured and predicted electricity consumption of the fan coils.

5.4.3 Reference and Flexible Scenario: Distributed control for thermal load management

This section describes the thermal load management in each building of the community under a reference and a flexible scenario. The reference scenario relies on a proportional controller, and it is implemented in residential buildings, while an MPC with a constant price signal and no demand response is employed for the institutional building (to optimize the different heating terminals in the environment – radiant slab and convective heating). In all buildings and scenarios no unacceptable zone temperature variations due to overheating are observed, with indoor temperatures maintained within 17–22°C for residential buildings (as shown for one of the residential buildings in Figure 5.11) and within 18–25°C for the institutional building, where the set points vary between 18–22°C compared to the 18-20°C in residential buildings.

The flexible scenario relies on the methodology described in Section 5.4.3, relying on a distributed economic MPC applied individually at each building in the community. The dual penalties for energy consumption and comfort create an optimization framework that balances these competing objectives, guiding the controller to find a thermal energy trend for each building which compromises between "comfort" and cost reduction.

To better understand what happens inside each building, in terms of thermal energy management, Figure 5.11 illustrates, for a *Severe Cold-Low* typical day, the comparison between the usual operation of the heating terminals in the building versus the MPC. The figure shows, with a 15-minute time step, the indoor temperature profiles and thermal energy of a representative thermal zone. The MPC effectively preheats the building during lower-cost hours, shifting thermal energy consumption to off-peak hours and ensuring temperature in acceptable bounds. It is interesting to note that part of the thermal energy consumption still happens during the DR event. This is consequence of the adopted weight, λ_i, during the economic MPC routine, which limits excessive indoor temperature reduction.

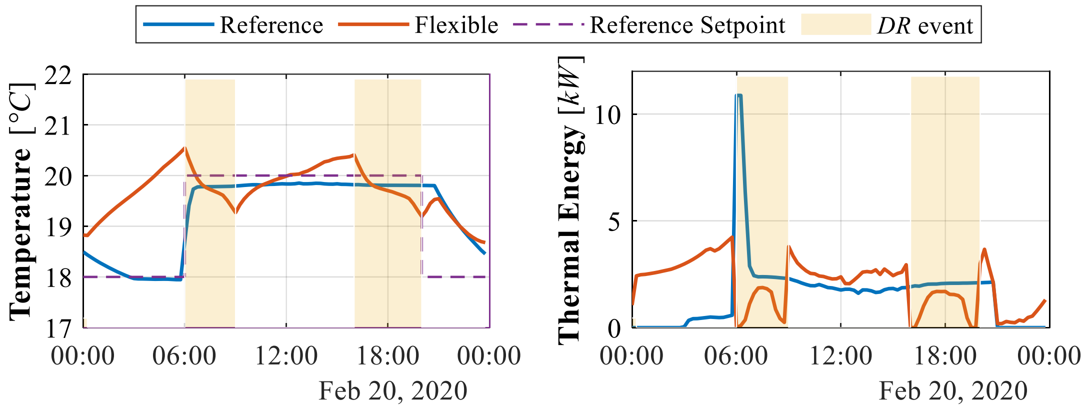


Figure 5.11: Reference and flexible scenarios in terms of indoor temperature and thermal energy for a thermal zone of a specific building during a *Severe Cold-Low* typical day.

The comparison between the two management strategies is showed in Table 5.5. The adoption of predictive control slightly increases the energy consumption of buildings, especially during *Cold-High* days (overall ranging between +0.6% to +8.6%). Conversely, the predictive control significantly reduces peak demand, with reductions ranging from 40.0 to 51.7 kW compared to the 63.5 to 71.7 kW range with mainly proportional control. By leveraging the building thermal mass, the system provides thermal energy during off-peak hours and limits it during peak periods, effectively flattening the load profile.

Table 5.5 shows that peak demand is closely tied to the timing of residential buildings' set-point temperature adjustments. Under flexible control, the peak shifts to just before or after demand response events, sometimes creating an energy *rebound* right after the DR period. The BEFI demonstrates considerable potential of the thermal energy management strategy, reaching values up to 41.5% during colder days. This corresponds to up to 41.5% of energy consumption shifted from on-peak hours to off-peak hours. Economic analysis reveals substantial rewards for buildings participating in demand response programs, especially during "Severe Cold" and "Very Cold" days, with potential savings outweighing the electrical energy cost of \$0.058–0.067/kWh (currently in Québec).

A broader community perspective reveals the impact of varying participation levels in demand response programs. By focusing on the demand response from 6 am to 9 am, Figure *5.12* demonstrates that as more buildings participate, the aggregated electrical energy consumption becomes more responsive and optimized. The BEFI is an important metric to show the effects of the varying participation intensity for each typical day scenario, considering *Min, Mean* and *Max* values related to which building is participating to the program. The BEFI values directly influence the sizing and operation of grid-supportive energy systems.

Table 5.5: Summary of aggregated results obtained from the local control for each typical day.

		Severe Cold			*Very Cold*			*Cold*		
		Low	*Mid*	*High*	*Low*	*Mid*	*High*	*Low*	*Mid*	*High*
Energy consumption [kWh]	*Reference*	1017	952	862	919	854	762	808	745	685
	Flexible	1036	987	899	925	880	810	837	786	744
Peak demand [kW]	*Reference*	71.7	71.6	69.0	69.8	70.4	63.5	68.7	64.8	63.5
	Flexible	48.2	51.7	47.4	43.2	44.9	43.1	40.4	40.0	41.0
Hour of peak demand [h]	*Reference*	6:00	6:00	6:00	6:00	6:00	6:00	6:00	6:00	6:00
	Flexible	20:15	20:15	20:15	9:00	9:00	9:00	9:00	9:00	20:15
BEFI [%] (6 am to 9 am)	*Reference*	-	-	-	-	-	-	-	-	-
	Flexible	41.5	38.8	35.2	41.1	38.0	31.8	37.7	31.53	26.7
BEFI [%] (4 pm to 8 pm)	*Reference*	-	-	-	-	-	-	-	-	-
	Flexible	21.5	10.6	9.2	19.8	9.0	6.9	15.7	7.2	4.7
Total Reward *[$]*	*Reference*	-	-	-	-	-	-	-	-	-
	Flexible	60.8	43.8	36.6	55.6	38.7	27.8	1.6	1.5	1.2

Furthermore, the *Mean* trends in Figure *5.12* show interesting results for the *Severe Cold-Low* and *Very Cold-Low* days where the BEFI trends have very similar patterns. At the same time, also *Very Cold-Mid* and *Cold-Low* day have very similar patterns. These underlines a common behaviour and operation for some combinations of outdoor temperature and solar radiation clusters. An interesting difference is provided by the *Max* BEFI trend which shows that for a lower number of buildings participating in demand response, the community BEFI provided during the demand response is higher for *Very Cold-Low* typical days compared to the *Severe Cold-Low* ones. This happens to avoid excessive temperature reduction during those hours, as also showed in Figure *5.11*, where the building can still request thermal energy during the demand response. By changing the weight, λ_i, during the distributed economic MPC routine, it is possible to obtain different results and BEFI values, according to the "importance" that the comfort has for each of the customers. These findings highlight the importance of flexible participation scenarios, where the methodology adapts to different numbers of buildings engaging in demand response programs. The results provide valuable insights for the application of local control strategies that maximize flexibility and minimize operational costs, aligning with sustainable energy management goals. The next link to provide is on how this optimal operation is interconnected when designing grid-supportive energy systems.

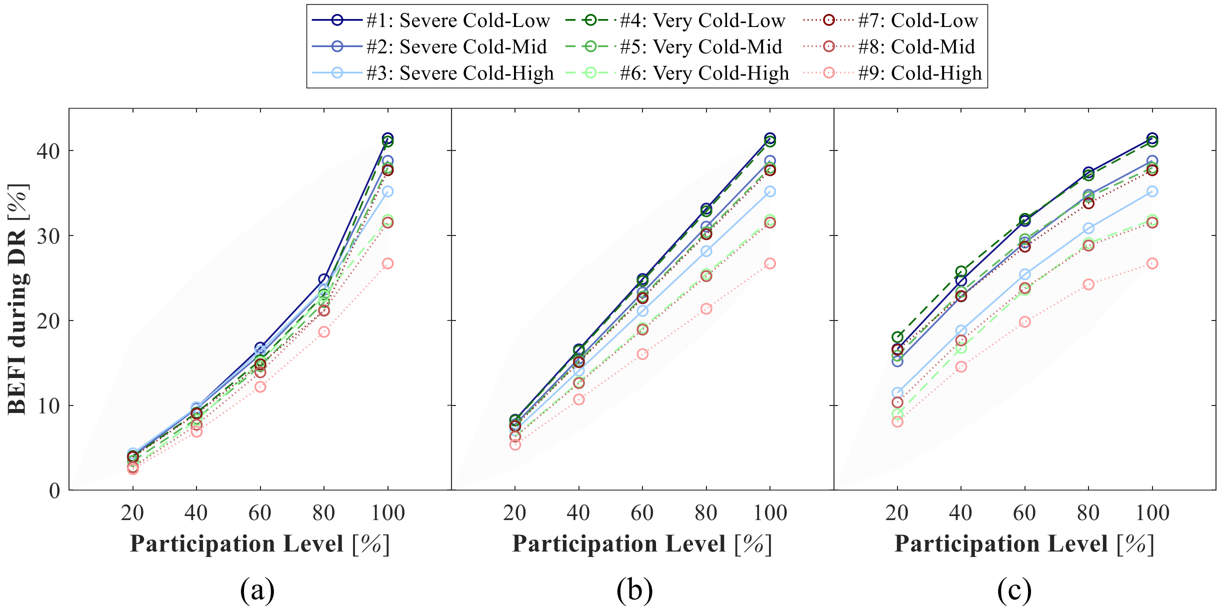


Figure 5.12: (a) Minimum/Min, (b) Average/Mean and (c) Maximum/Max building energy flexibility index during demand response event (6 am to 9 am) evaluated by varying the buildings participation level.

5.4.4 Supervisory control applied to the microgrid: PV and battery storage operation

The integration of photovoltaic (PV) and battery storage systems is crucial for supporting grid balancing, particularly during periods of high renewable energy penetration and demand response (DR) events. While thermal load management in buildings offers a degree of energy flexibility, it often falls short of addressing grid stability requirements. By combining PV and battery systems with thermal load control, flexibility can be significantly improved, allowing for a more effective response to grid demands during critical periods.

For instance, an analysis of a *Severe Cold-Low* day, with all buildings participating in demand response alongside a PV system and battery capacity of 80 kW and 200 kWh respectively, highlights the influence of these factors. Electricity demand scenarios—including the reference case, local distributed MPC, and double-level MPC control—demonstrate that the combined PV and battery system under double-level MPC control performs best during DR events. This double-level control enables the community load to align closely with the local MPC control scenario while ensuring that the selected battery size sufficiently meets most energy consumption during two DR periods.

As shown by the electricity demand trend in Figure *5.13*, the PV-battery combination also mitigates energy increase (*rebound*) after the first DR event (from 6:00 to 9:00). With higher solar radiation levels, this effect is expected to be even more pronounced. In contrast, without PV-battery support, the increase in building demand after the event is more significant, as seen in the reference scenario (red line). Following the second DR event, the battery transitions into a charging phase to replenish its state of charge $\left(SoC_{end-day}\right)$, temporarily increasing consumption as it cannot supply energy to the community. This causes the peak demand to shift from early hours of the day to later hours, after the second DR event. The analysis also highlights the battery ability to maintain its SoC within predefined bounds, as depicted in Figure *5.13*. However, the low solar radiation characteristic of the day limits PV electricity production, underscoring the importance of effective system integration. In this scenario, the 80 kW PV cannot fully meet the battery's charging needs, which instead forces a 10.4% increase in electricity demand from 9:00 to 16:00.

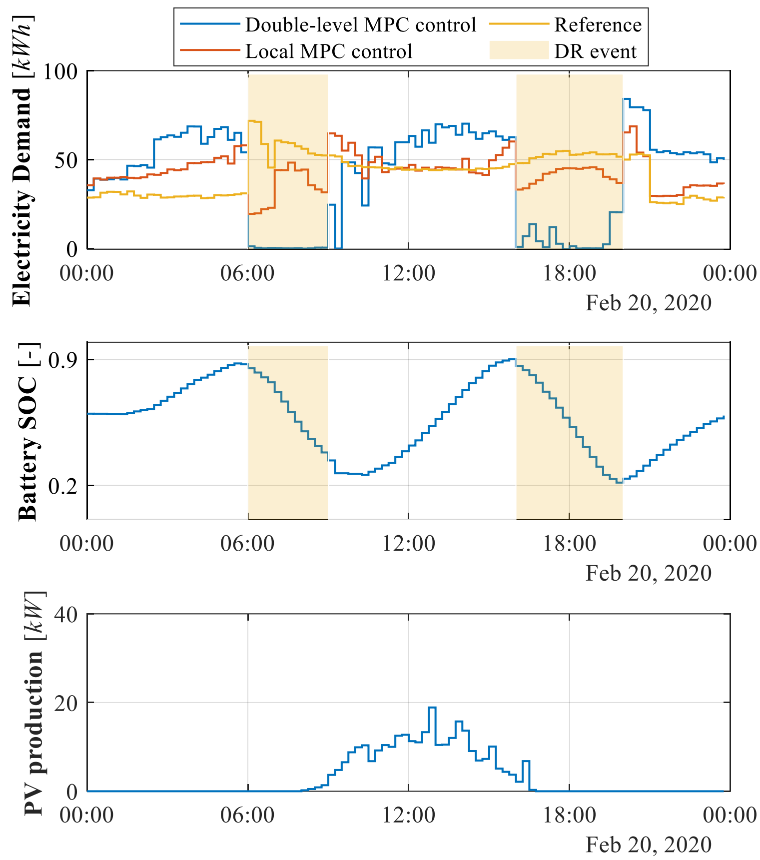


Figure 5.13: Electricity demand, battery SOC and PV production for a *Severe Cold-Low* day scenario considering a battery of 200 kWh and PV size of 80 kW.

5.4.5 Supervisory control applied to the microgrid: design and control in the Worst-Case scenario

Achieving grid flexibility requires meticulous design and operational strategies that dynamically coordinate these systems. The operation of the battery system is closely tied to the design principles of the proposed methodology and the support provided by PV generation. The optimal interactions between PV and battery storage technologies depend on three key factors: (i) the level of building participation in DR events, quantified using the Building Energy Flexibility Index (BEFI), (ii) the extent to which the electrical grid is leveraged to cover building loads during these periods, and (iii) the selection of the typical day on which to assess the analysis.

The electrical grid may opt for a system configuration that covers only a portion of the community's energy demand while leveraging building energy flexibility. The interplay between flexibility and the optimal sizing of photovoltaic and battery systems highlights the importance of the loss of load probability (LLP) indicator. LLP serves as a key metric, providing insights into the system's ability to maintain supply during grid-supporting operations. It effectively evaluates the trade-offs between flexibility, system sizing, and overall performance.

Given the presence of several typical days, it becomes necessary to choose a specific day to focus on for detailed analysis and system sizing. The days can be classified based on their energy consumption values, as provided in Table 5.5, with the *Severe Cold-Low* day identified as the *worst-case* scenario. Sizing the system for this specific case ensures that all other conditions are covered. Alternatively, by using the weight factors provided in Table 5.3 and classifying the days from the highest to the lowest energy consumption values, decisions can be made regarding the final analysis. For instance, the system could be sized with 100% probability of covering the virtual community's needs by sizing it under the *worst-case* scenario corresponding to the *Severe Cold-Low* day, with 95% probability by sizing the system under the *Very Cold-Low* day, or by considering the most recurrent scenario corresponding to the *Very Cold-Mid* day.

In this section, detailed results are provided for the *worst-case* scenario. However, the final decision on system sizing can be made based on these three factors, balancing the desired level of coverage probability with system cost and performance considerations.

Figure 5.14 shows, for a *Severe Cold*-Low typical day, the connection between different levels of LLP and the varying size of PV and battery storage systems. The figure is a contour plot generated by fitting cubic curves through data points which have similar battery sizes. While this method may introduce some nonlinearities in certain regions, the primary focus of the plot is to highlight overall trends and key patterns. The results indicate that as the LLP value increases, the influence of PV size on the battery storage system diminishes. This occurs because the battery system primarily ensures achieving the lowest possible LLP during the first demand response event, which lacks any solar radiation. In contrast, the PV system contributes partially during the second demand response event (16:00 to 20:00), mainly by supporting battery charging and reducing electricity demand during solar radiation hours. Consequently, for some LLP values and for a specific BEFI value, the required storage size becomes uniquely defined.

The analysis demonstrates how battery sizing adjusts dynamically based on the extent of aggregated load coverage by the electrical grid during specific DR events (LLP indicator). For instance, battery capacity decreases from 236 kWh to 184 kWh and eventually to 125 kWh as the grid assumes greater responsibility for the load. Additionally, fixing the LLP value reveals significant battery size reductions with a decrease of 22.0% for LLP = 0.1, 15.4% for LLP = 0.2, 22.8% for LLP = 0.3, and 13.8% for LLP=0.4.

For days with low or moderate solar radiation and higher thermal energy demand, a similar trend is observed. For certain LLP and BEFI values, there is a single optimal battery size required to meet the community load. This highlights that increasing PV capacity alone is insufficient to reduce building participation in DR events to achieve specific LLP targets. In these scenarios, building flexibility and the appropriate battery size remain essential to meet LLP objectives.

This trend changes for days with high solar radiation, where the impact of PV size becomes significant. PV capacity not only reduces the required battery size but also decreases electricity demand from 9:00 to 16:00, enabling the PV-battery storage system to sustain itself during demand peaks later in the day. Detailed results for each typical day, including PV-battery sizing trends for various LLP and BEFI values, are presented in Figure B.*1* in Appendix B. Additionally, Figure B.*2* in Appendix B provides further insights into the electricity demand reduction from 9:00 to 16:00 due to PV production. By fitting cubic curves to specific parameter values while varying other variables influenced by DR participation—evaluated through the Building Energy Flexibility Index (BEFI)—this analysis establishes a solid basis for sizing and operational planning of PV and battery systems. These curves identify the optimal battery capacity or just the energy (kWh) required by grid-supportive systems under different LLP values and weather cluster scenarios, emphasizing the importance of system adaptability to grid requirements and typical day conditions.

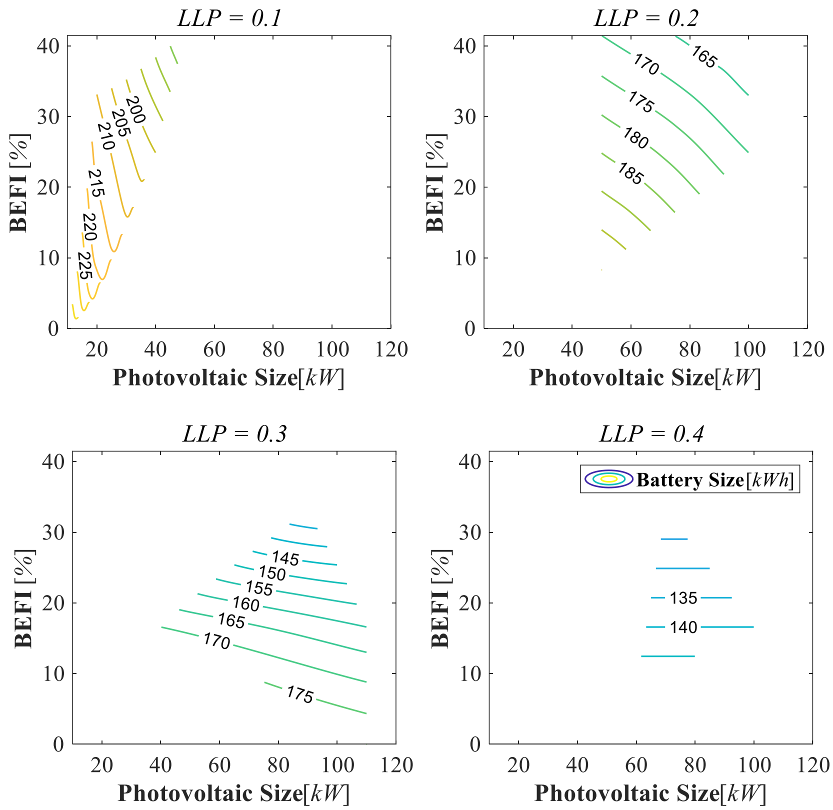


Figure 5.14: Effect of community BEFI on the battery and photovoltaic size for different Loss of Load Probability values and for a *Severe Cold – Low* Radiation Day.

5.5 Discussion

The results of this study underscore the importance of identifying typical weather clusters using dynamic time warping (DTW) and hierarchical clustering methods tailored for time series data. This approach enables the selection of representative days, capturing the variability in weather patterns and ensuring that energy management strategies are robust and adaptable to diverse scenarios.

Local control, achieved through optimal thermal load management, has proven to be a critical factor in this analysis. The localized impact of thermal load optimization must be considered when designing and operating grid-supportive technologies. Ignoring these localized effects could lead to suboptimal system sizing and reduced operational efficiency.

The findings also highlight the essential role of battery storage in addressing the first demand response (DR) event. PV systems play a supportive role by covering part or all of the battery charging needs before the second peak, and by reducing overall electricity consumption during hours of solar radiation. System sizing decisions can be guided by analysing the *worst-case* scenario. In this study, results for this specific scenario were emphasized, as they provide insights into the system resilience during critical conditions, even though this scenario is not the most frequent. Alternatively, sizing could be based on different probabilities of covering the virtual community's needs or the most recurrent conditions.

This study illustrates how LLP and the Building Energy Flexibility Index (BEFI) influence the sizing and operation of the microgrid components supporting the utility during demand response events. These events can be likened to resilience scenarios, such as grid outages or periods of heightened blackout risk. In this study, the system was sized with a focus on critical DR periods and peak hours. However, the same methodology can be adapted for other objectives, such as load flattening, which may require a different operational strategy for the battery. Additionally, the framework is flexible enough to incorporate alternative pricing signals or longer-duration DR events, expanding its applicability to various energy management contexts.

5.6 Conclusions

This study presents a robust methodology that combines dynamic thermal load models, integrated design and control frameworks, and community-scale optimization to address the dual challenges of energy flexibility and operational efficiency. Dynamic load models were developed to generate buildings load profiles based on real data, enabling detailed analyses of how grid pricing signals influence the optimal control. The integrated design and control framework revealed the impact of adjusted load profiles on the optimal sizing and operation of PV and battery systems, demonstrating the potential to achieve objectives such as enhanced flexibility, cost reduction, and resilience. Furthermore, by examining aggregated building loads at the community scale, the methodology highlights how single-building control strategies influence broader operational outcomes, emphasizing the importance of coordinated energy management.

Using dynamic time warping (DTW) and hierarchical clustering to extract representative weather periods, the methodology delivers valuable insights into reducing distributed energy resource/technology sizes and costs while simultaneously improving energy flexibility and resilience within building communities. For instance, during *Severe Cold-Low* solar radiation days, scenarios with PV and battery systems employing double-level control proved effective, with specific battery-PV combinations significantly mitigating peak demand and providing the required flexibility to the electrical grid.

In the context of demand response events, the electrical grid may aim to balance energy flexibility with optimal system sizing by configuring technologies to cover only a portion of the community's energy demand. This approach underscores the importance of leveraging building energy flexibility while ensuring system reliability. In this study, BEFI and LLP serve as robust metrics for evaluating the trade-offs between flexibility, photovoltaic and battery system sizing, and overall performance, guiding the design of efficient and resilient energy systems.

When analysing the loss of load probability for low radiation days, the results showed that PV size had minimal influence under low solar radiation. In contrast, battery size variations dictated system performance, with reductions in battery size (or battery usage) of up to 26%, in the *worst*-case scenario, for specific LLP and BEFI scenarios. Days with high solar radiation exhibited a stronger influence of PV size on flexibility, further emphasizing the need for scenario-specific optimization. Importantly, the methodology also showcased how PV systems mitigate rebound effects and sustain the battery operation for a big part of the day.

This methodology offers a powerful tool for optimizing the sizing and operation of PV-battery systems at both building and community scales. By integrating predictive modelling, clustering techniques, and dynamic load management, it provides a comprehensive framework for supporting the electrical grid in achieving economic and operational goals. As building communities become central to future energy ecosystems, tools like this are essential for enabling the grid to adapt to predicted energy consumption patterns while enhancing resilience and efficiency.

Chapter: 6 Conclusions and Future Works

This thesis explores methods to harness the energy flexibility potential of buildings through thermal load management and coordination with various energy systems. Building energy modelling has played a crucial role in this effort, advancing the concept of smart buildings, where monitoring infrastructure enables a detailed representation of energy consumption. Resistance-capacitance (RC) thermal models were pivotal in studies involving advanced control strategies, such as model predictive control (MPC).

A key contribution is the development of a model order reduction methodology that directly evaluates the structure and parameters of accurate thermal models, effectively capturing building thermal dynamics. This methodology enhances data preprocessing by selecting the most informative data instances, cleaning data, and neglecting or lumping thermal zones. This approach enables the creation of compact models that limit uncertainty from thermal zone interactions while preserving sufficient detail to account for varying occupancy preferences across zones. The methodology bridges RC model generation with automated development and thermal zoning. Furthermore, it serves as a fundamental step toward automating the aggregation of building models for district-level energy management and optimization. By offering a systematic approach to generate accurate, scalable models, this methodology lays the groundwork for integrating multiple buildings into a cohesive, automated system that can be expanded to larger-scale district-level applications, facilitating energy flexibility and coordinated control across entire neighborhoods or urban areas.

The thesis also examines the use of model archetypes to optimize thermal performance in buildings operating on schedules, where the structure of mechanical components (e.g., HVAC systems) significantly influences model design. While not fully automated, this approach is especially valuable for buildings with complex mechanical systems and architecture.

The adoption of economic Model Predictive Control (e-MPC) was critical for incorporating building thermal dynamics while accounting for occupancy, weather forecasts, and grid signals. Tariff structures, such as time-of-use pricing, critical peak pricing, and demand response scenarios, were key in achieving energy flexibility by reducing peak demand and shifting energy consumption to off-peak hours. This optimization framework effectively leverages building thermal mass, avoiding excessive peaks and variations in energy consumption. For example, studies on a residential buildings' dataset and the Varennes Library, using an MPC routine with a constant or dynamic price signal, demonstrated optimal thermal load management across full-convective, full-radiant, and mixed convective-radiant terminals.

The thesis introduced multiple cost functions to balance indoor comfort, technology operation, and building-grid interaction through dynamic tariffs, incentives, and load-shaping strategies. Various control architectures—centralized, distributed, double-level, and hierarchical—were analyzed to coordinate technologies and objectives. Centralized control proved most effective for single-building applications, integrating all influencing factors into a single cost function. At the community scale, where increased objectives can lead to convergence challenges, distributed architecture emerged as a viable approach. These enable individual buildings to optimize energy consumption while responding to grid signals, with supervisory control at the community level ensuring overall system efficiency.

Clustering techniques were employed to extract typical days based on solar radiation and outdoor temperature. This approach enables scenario-based simulations that reduce computational time and facilitates the creation of multiple control scenarios tailored to specific conditions. The resulting optimal control scenarios can be easily implemented and integrated through rule-based architectures in building automation systems. The development of a methodology to identify typical days strengthens the connection between design and control studies. Each typical day allows for the evaluation of specific scenarios, enabling the optimization of technology sizing based on key parameters such as energy consumption and operational requirements. The optimal design variable can be selected by assigning a weight to the solution provided by each typical day—for example by reflecting its frequency of occurrence over the study period and considering its importance, such as higher energy consumption in worst-case scenarios. This approach parallels traditional design methodologies, which balance worst-case and more flexible scenarios, and integrates them into control analyses. Ultimately, typical days provide a structured framework for linking design decisions with control strategies, ensuring both robust performance and energy efficiency.

The thesis also highlights the importance of key performance indicators (KPIs) and demonstrates how metrics related to operation—such as energy flexibility, load matching, and grid interaction—can inform technology design.

By integrating automated and innovative modelling techniques, establishing robust control frameworks for buildings with multiple energy systems, selecting appropriate KPIs for monitoring and evaluation, and employing clustering techniques to link design and control, this thesis advances control management strategies at the building level. It also provides guidance for designing grid-supportive technologies and optimizing control strategies at larger scales, quantifying their impact on the coordination of energy systems.

6.1 Contributions

The major contributions resulting from this thesis stem directly from the thesis objectives presented in Section 1.2:

- Automated Model Order Reduction for Thermal Load Prediction: Developed automated, data-driven modelling techniques using smart thermostat data to predict building energy demand, balancing accuracy and computational efficiency for real-time predictive control.

- Control-Oriented Model Archetypes for Buildings Operating on a Schedule: Designed control-oriented model archetypes to optimize building operations based on schedules, enhancing energy flexibility and minimizing waste through dynamic grid responsiveness.

- Optimization of Energy Flexibility through Thermal Load Management: Introduced strategies leveraging building thermal mass and price-based demand response to optimize energy flexibility, reduce peak demand, and ensure measurable performance through KPIs.

- Design and Control for Energy Flexibility at Community Scale: Developed methodologies for integrating and optimizing grid-supportive technologies within microgrids, enabling efficient system sizing and enhanced energy flexibility for sustainable energy systems.

6.2 Publications

6.2.1 Journal publications

[1] *A. Maturo*, A. Petrucci, C. Forzano, G.F. Giuzio, A. Buonomano, A. Athienitis, «Design and Environmental Sustainability Assessment of Energy-Independent Communities: The Case Study of a Livestock Farm in the North of Italy». Energy Reports, vol. 7, November 2021, page. 8091–107. ScienceDirect, https://doi.org/10.1016/j.egyr.2021.05.080.

[2] *A. Maturo*, A. Buonomano, A. Athienitis, «Design for Energy Flexibility in Smart Buildings through Solar Based and Thermal Storage Systems: Modelling, Simulation, and Control for the System Optimization». Energy, vol. 260, December 2022, page. 125024. ScienceDirect, https://doi.org/10.1016/j.energy.2022.125024.

[3] H. Li, H. Johra, F. de Andrade Pereira, T. Hong, J. Le Dreau, *A. Maturo*, M. Wei, Y. Liu, A. Saberi-Derakhtenjani, Z. Nagy, A. Marszal-Pomianowska, D. Finn, S. Miyata, K. Kaspar, K. Nweye, Z. O Neill, F. Pallonetto, B. Dong, «Data-Driven Key Performance Indicators and Datasets for Building Energy Flexibility: A Review and Perspectives». Applied Energy, vol. 343, August 2023, page. 121217. ScienceDirect, https://doi.org/10.1016/j.apenergy.2023.121217.

[4] *A. Maturo*, C. Vallianos, A. Buonomano, A. Athienitis, «A Novel Multi-Level Predictive Management Strategy to Optimize Phase-Change Energy Storage and Building-Integrated Renewable Technologies Operation under Dynamic Tariffs». Energy Conversion and Management, vol. 291, September 2023, page. 117220. ScienceDirect, https://doi.org/10.1016/j.enconman.2023.117220.

[5] *A. Maturo*, C. Vallianos, B. Delcroix, A. Buonomano, A. Athienitis, «Automated Model Order Reduction for Building Thermal Load Prediction Using Smart Thermostats Data». Journal of Building Engineering, vol. 96, November 2024, p. 110492. ScienceDirect, https://doi.org/10.1016/j.jobe.2024.110492.

[6] *A. Maturo,* A. Buonomano, A. Athienitis, «Optimizing energy flexibility through electricity price-responsiveness and thermal load management in buildings with convective and radiant heating systems». Energy and Buildings, January 2025, p. 115355. ScienceDirect, https://doi.org/10.1016/j.enbuild.2025.115355.

[7] *A. Maturo*, C. Vallianos, B. Delcroix, A. Buonomano, A. Athienitis, «Clustering-Driven Design and Predictive Control of Hybrid PV-Battery Storage Systems for Demand Response in Energy Communities». Renewable Energy. *Under review*

6.2.2 Conference proceedings

[1] G. Barone, A. Buonomano, C. Forzano, G.F. Giuzio, *A. Maturo*, A. Palombo and A. Petrucci. Dynamic simulation for highly energy-independent communities design: the case study of a livestock farm in the North of Italy. Published at SDEWES 2020, 15th Conference on Sustainable Development of Energy, Water and Environment Systems, Cologne, Germany – September 2020.

[2] *A. Maturo*, A. Buonomano, A. Athienitis. Simulation and control for energy management: an energy management strategy applied to a multi-zone building coupled with solar-based and energy storage technologies. Published at SDEWES 2021, 16th Conference on Sustainable Development of Energy, Water and Environment Systems, Dubrovnik, Croatia – October 2021.

[3] *A. Maturo*, A. Petrucci, A. Buonomano, A. Athienitis. Thermal and electrical modelling of a double-skin façades integrating bifacial photovoltaics: energy and economic performance assessment. Published at SDEWES 2021, 16th Conference on Sustainable Development of Energy, Water and Environment Systems, Dubrovnik, Croatia – October 2021.

[4] *A. Maturo*, A. Athienitis, B. Delcroix. A data-driven frequency domain system identification approach to define house archetypes and flexibility. Published at COBEE 2022, 5th International Conference on Building Energy and Environment, Montreal, Québec, Canada – July 2022.

[5] *A. Maturo*, C. Vallianos, A. Buonomano, A. Athienitis. Model predictive control for energy flexibility of a building coupled with advanced solar and energy storage technologies. Published at SDEWES 2022, 17th Conference on Sustainable Development of Energy, Water and Environment Systems, Paphos, Cyprus – October 2022.

[6] *A. Maturo*, A. Athienitis, B. Delcroix. A novel model reduction and calibration methodology to define lumped parameters of building thermal models. Published at IBPSA 2023, 18th IBPSA International Conference and Exhibition Building Simulation 2023, Shanghai, China – September 2023.

[7] *A. Maturo*, A. Buonomano, A. Athienitis. Modelling and control for energy flexibility distinguishing fast and slow response in a Net Zero Institutional Building: the Varennes Library. Published at SDEWES 2023, 18th Conference on Sustainable Development of Energy, Water and Environment Systems, Dubrovnik, Croatia – September 2023.

[8] *A. Maturo*, A. Buonomano, A. Athienitis. The role of Net Zero Energy Buildings in building cluster as energy community. Published at SDEWES 2024, 19th Conference on Sustainable Development of Energy, Water and Environment Systems, Rome, Italy – September 2024.

[9] *A. Maturo*, B. Delcroix, A. Buonomano, A. Athienitis. The nexus between design and control: a data-driven approach for leveraging flexibility potential of micro-grids. Published at SyNERGY MED 2024, 3rd International Conference on Energy Transition in the Mediterranean Area, Limassol, Cyprus – October 2024.

[10] *A. Maturo*, B. Delcroix, A. Buonomano, A. Athienitis. Thermal and electrical load optimization in building clusters for energy flexibility in grid interaction. Published at ASHRAE Winter Conference 2025, Orlando, Florida – February 2025.

6.3 Recommendations for future works

This research provides a solid foundation for improving building energy flexibility and optimizing thermal load management, yet several areas remain open for further exploration.

- Occupancy Estimation and Indoor Air Quality Analysis: Future studies should focus on developing models to estimate building occupancy and monitor CO_2 concentration. These models would provide deeper insights into occupant behavior and air quality, enhancing the evaluation of indoor comfort conditions. Integrating these parameters into energy management systems can lead to more comprehensive control strategies that balance energy efficiency with occupant well-being.

- Peer-to-Peer Energy Markets: The development of non-hierarchical distributed control routines, such as peer-to-peer frameworks, represents a promising direction for optimizing energy systems at the community level. These routines would allow buildings to exchange energy directly, fostering an energy market where individual buildings collaboratively manage energy resources. Such an approach could provide significant benefits to the electrical grid by enhancing flexibility, reducing peak demand, and improving resilience.

- Advanced Multi-sensor Integration: Incorporating new multi-sensor technologies that measure variables beyond standard parameters could greatly enhance energy models. By integrating data from sensors that track additional environmental and operational factors, these systems could provide better model validation, improve the generalizability of control strategies, and generate novel control parameters. This advancement would strengthen the connection between monitoring systems and energy models, enabling more precise and adaptive control frameworks.

- Exploration of Diverse Renewable Energy Resources: Expanding the focus beyond photovoltaic (PV) systems to include other renewable energy sources, such as wind or geothermal, offers opportunities to diversify energy generation and enhance grid interaction. Coupling these technologies with battery storage systems could create more robust and flexible energy solutions. Studies on the integration and coordination of multiple renewable sources would be valuable for developing holistic strategies to decarbonize the building sector.

By addressing these recommendations, future research can further advance the integration of energy flexibility into building systems, support the transition to smart grids, and contribute to a sustainable and resilient energy future.

Bibliography

1. Cao, X., X. Dai, and J. Liu, *Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade.* Energy and Buildings, 2016. **128**: p. 198-213.
2. Asif, M. and T. Muneer, *Energy supply, its demand and security issues for developed and emerging economies.* Renewable and Sustainable Energy Reviews, 2007. **11**(7): p. 1388-1413.
3. Lawrence, T.M., et al., *Ten questions concerning integrating smart buildings into the smart grid.* 2016. **108**: p. 273-283.
4. Sinsel, S.R., R.L. Riemke, and V.H. Hoffmann, *Challenges and solution technologies for the integration of variable renewable energy sources—a review.* Renewable Energy, 2020. **145**: p. 2271-2285.
5. Al-Shetwi, A.Q., et al., *Grid-connected renewable energy sources: Review of the recent integration requirements and control methods.* Journal of Cleaner Production, 2020. **253**: p. 119831.
6. Liang, X., *Emerging power quality challenges due to integration of renewable energy sources.* IEEE Transactions on Industry Applications, 2016. **53**(2): p. 855-866.
7. Impram, S., S.V. Nese, and B. Oral, *Challenges of renewable energy penetration on power system flexibility: A survey.* Energy Strategy Reviews, 2020. **31**: p. 100539.
8. Bird, L., M. Milligan, and D. Lew, *Integrating variable renewable energy: Challenges and solutions*. 2013, National Renewable Energy Lab.(NREL), Golden, CO (United States).
9. Obi, M., R.J.R. Bass, and S.E. Reviews, *Trends and challenges of grid-connected photovoltaic systems–A review.* 2016. **58**: p. 1082-1094.
10. Gonzalez-Salazar, M.A., T. Kirsten, and L. Prchlik, *Review of the operational flexibility and emissions of gas-and coal-fired power plants in a future with growing renewables.* Renewable and Sustainable Energy Reviews, 2018. **82**: p. 1497-1513.
11. Evans, A., V. Strezov, and T.J. Evans, *Assessment of utility energy storage options for increased renewable energy penetration.* Renewable and Sustainable Energy Reviews, 2012. **16**(6): p. 4141-4147.
12. Rodrigues, E., et al., *Energy storage systems supporting increased penetration of renewables in islanded systems.* Energy, 2014. **75**: p. 265-280.
13. Mwasilu, F., et al., *Electric vehicles and smart grid interaction: A review on vehicle to grid and renewable energy sources integration.* Renewable and Sustainable Energy Reviews, 2014. **34**: p. 501-516.
14. Tuballa, M.L. and M.L. Abundo, *A review of the development of Smart Grid technologies.* Renewable and Sustainable Energy Reviews, 2016. **59**: p. 710-725.
15. Kanakadhurga, D. and N. Prabaharan, *Demand side management in microgrid: A critical review of key issues and recent trends.* Renewable and Sustainable Energy Reviews, 2022. **156**: p. 111915.
16. Golmohamadi, H., *Demand-side management in industrial sector: A review of heavy industries.* Renewable and Sustainable Energy Reviews, 2022. **156**: p. 111963.
17. Behrangrad, M., *A review of demand side management business models in the electricity market.* Renewable and Sustainable Energy Reviews, 2015. **47**: p. 270-283.

18. Pérez-Lombard, L., J. Ortiz, and C. Pout, *A review on buildings energy consumption information.* Energy and Buildings, 2008. **40**(3): p. 394-398.
19. Swan, L.G. and V.I. Ugursal, *Modeling of end-use energy consumption in the residential sector: A review of modeling techniques.* Renewable and Sustainable Energy Reviews, 2009. **13**(8): p. 1819-1835.
20. Amasyali, K. and N.M. El-Gohary, *A review of data-driven building energy consumption prediction studies.* Renewable and Sustainable Energy Reviews, 2018. **81**: p. 1192-1205.
21. Uddin, M., et al., *A review on peak load shaving strategies.* Renewable and Sustainable Energy Reviews, 2018. **82**: p. 3323-3332.
22. Li, H., et al., *Energy flexibility of residential buildings: A systematic review of characterization and quantification methods and applications.* Advances in Applied Energy, 2021. **3**: p. 100054.
23. Tang, H., S. Wang, and H. Li, *Flexibility categorization, sources, capabilities and technologies for energy-flexible and grid-responsive buildings: State-of-the-art and future perspective.* Energy, 2021. **219**: p. 119598.
24. Chen, Y., et al., *Measures to improve energy demand flexibility in buildings for demand response (DR): A review.* Energy and Buildings, 2018. **177**: p. 125-139.
25. Jensen, S.Ø., et al., *IEA EBC annex 67 energy flexible buildings.* Energy and Buildings, 2017. **155**: p. 25-34.
26. IEA-EBC. *Annex 81 - Data-Driven Smart Buildings*. Available from: https://annex81.iea-ebc.org/.
27. IEA-EBC. *Annex 82 - Energy Flexible Buildings Towards Resilient Low Carbon Energy Systems*. Available from: https://annex82.iea-ebc.org/.
28. Drgoňa, J., et al., *All you need to know about model predictive control for buildings.* 2020. **50**: p. 190-232.
29. Le Dréau, J. and P. Heiselberg, *Energy flexibility of residential buildings using short term heat storage in the thermal mass.* Energy, 2016. **111**: p. 991-1002.
30. Zhao, P., et al., *Evaluation of commercial building HVAC systems as frequency regulation providers.* Energy and Buildings, 2013. **67**: p. 225-235.
31. Wu, X., et al., *Hierarchical control of residential HVAC units for primary frequency regulation.* IEEE Transactions on Smart Grid, 2017. **9**(4): p. 3844-3856.
32. Zou, W., et al., *A review on integration of surging plug-in electric vehicles charging in energy-flexible buildings: Impacts analysis, collaborative management technologies, and future perspective.* Applied Energy, 2023. **331**: p. 120393.
33. Venegas, F.G., M. Petit, and Y. Perez, *Active integration of electric vehicles into distribution grids: Barriers and frameworks for flexibility services.* Renewable and Sustainable Energy Reviews, 2021. **145**: p. 111060.
34. Stinner, S., K. Huchtemann, and D. Müller, *Quantifying the operational flexibility of building energy systems with thermal energy storages.* Applied Energy, 2016. **181**: p. 140-154.
35. Niu, J., et al., *Flexible dispatch of a building energy system using building thermal storage and battery energy storage.* Applied Energy, 2019. **243**: p. 274-287.
36. Lefebure, N., et al., *Distributed model predictive control of buildings and energy hubs.* Energy and Buildings, 2022. **259**: p. 111806.
37. Le Dréau, J., et al., *Developing energy flexibility in clusters of buildings: A critical analysis of barriers from planning to operation.* Energy and Buildings, 2023: p. 113608.

38. Vigna, I., et al., *New domain for promoting energy efficiency: Energy Flexible Building Cluster.* Sustainable Cities and Society, 2018. **38**: p. 526-533.
39. Maturo, A., A. Buonomano, and A. Athienitis, *Design for energy flexibility in smart buildings through solar based and thermal storage systems: Modelling, simulation and control for the system optimization.* Energy, 2022. **260**: p. 125024.
40. Rullo, P., et al., *Integration of sizing and energy management based on economic predictive control for standalone hybrid renewable energy systems.* Renewable energy, 2019. **140**: p. 436-451.
41. Baniasadi, A., et al., *Optimal sizing design and operation of electrical and thermal energy storage systems in smart buildings.* Journal of Energy Storage, 2020. **28**: p. 101186.
42. Mlecnik, E., et al., *Policy challenges for the development of energy flexibility services.* Energy Policy, 2020. **137**: p. 111147.
43. Al Dakheel, J., et al., *Smart buildings features and key performance indicators: A review.* Sustainable Cities and Society, 2020. **61**: p. 102328.
44. IEA-EBC. *Annex 84 - Demand Management of Buildings in Thermal Networks*. Available from: https://annex84.iea-ebc.org/.
45. Satchwell, A., Piette, Mary Ann, Khandekar, Aditya, Granderson, Jessica, Frick, Natalie Mims, Hledik, Ryan, Faruqui, Ahmad, Lam, Long, Ross, Stephanie, Cohen, Jesse, Wang, Kitty, Urigwe, Daniela, Delurey, Dan, Neukomm, Monica, and Nemtzow, David, *A National Roadmap for Grid-Interactive Efficient Buildings.* 2021.
46. Di Silvestre, M.L., et al., *How Decarbonization, Digitalization and Decentralization are changing key power infrastructures.* Renewable and Sustainable Energy Reviews, 2018. **93**: p. 483-498 %@ 1364-0321.
47. IEA, *Digitalisation and Energy*. 2017.
48. Candanedo, J., et al., *Data-Driven Smart Buildings: State-of-the-Art Review.* 2023.
49. Haider, H.T., O.H. See, and W. Elmenreich, *A review of residential demand response of smart grid.* Renewable and Sustainable Energy Reviews, 2016. **59**: p. 166-178.
50. Halvgaard, R., et al. *Economic model predictive control for building climate control in a smart grid*. in *2012 IEEE PES innovative smart grid technologies (ISGT)*. 2012. IEEE.
51. Paterakis, N.G., O. Erdinç, and J.P. Catalão, *An overview of Demand Response: Key-elements and international experience.* Renewable and Sustainable Energy Reviews, 2017. **69**: p. 871-891.
52. Samad, T., E. Koch, and P. Stluka, *Automated demand response for smart buildings and microgrids: The state of the practice and research challenges.* Proceedings of the IEEE, 2016. **104**(4): p. 726-744.
53. Madina, C., et al., *Technologies and protocols: the experience of the three smartNet pilots.* TSO-DSO Interactions Ancillary Services in Electricity Transmission Distribution Networks: Modeling, Analysis Case-Studies, 2020: p. 141-183.
54. Kazmi, H., et al., *Multi-agent reinforcement learning for modeling and control of thermostatically controlled loads.* Applied Energy, 2019. **238**: p. 1022-1035.
55. Maturo, A., et al., *A novel multi-level predictive management strategy to optimize phase-change energy storage and building-integrated renewable technologies operation under dynamic tariffs.* Energy Conversion and Management, 2023. **291**: p. 117220.
56. Aste, N., M. Manfren, and G. Marenzi, *Building Automation and Control Systems and performance optimization: A framework for analysis.* Renewable Sustainable Energy Reviews, 2017. **75**: p. 313-330.

57. Deb, C. and A. Schlueter, *Review of data-driven energy modelling techniques for building retrofit.* Renewable Sustainable Energy Reviews, 2021. **144**: p. 110990.
58. Ravelo, B., L. Rajaoarisoa, and O. Maurice, *Thermal modelling of multilayer walls for building retrofitting applications.* Journal of Building Engineering, 2020. **29**: p. 101126.
59. Cecconi, F.R., A. Khodabakhshian, and L. Rampini, *Data-driven decision support system for building stocks energy retrofit policy.* Journal of Building Engineering, 2022. **54**: p. 104633.
60. Lirola, J.M., et al., *A review on experimental research using scale models for buildings: Application and methodologies.* Energy and Buildings, 2017. **142**: p. 72-110.
61. Lv, C., et al., *Model predictive control based robust scheduling of community integrated energy system with operational flexibility.* Applied energy, 2019. **243**: p. 250-265.
62. Buonomano, A., et al., *Dynamic building energy performance analysis: A new adaptive control strategy for stringent thermohygrometric indoor air requirements.* Applied Energy, 2016. **163**: p. 361-386.
63. Buonomano, A. and A. Palombo, *Building energy performance analysis by an in-house developed dynamic simulation code: An investigation for different case studies.* Applied Energy, 2014. **113**(0): p. 788-807.
64. Loyola-Gonzalez, O., *Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view.* IEEE Access, 2019. **7**: p. 154096-154113.
65. Harish, V. and A. Kumar, *A review on modeling and simulation of building energy systems.* Renewable and Sustainable Energy Reviews, 2016. **56**: p. 1272-1292.
66. Guidotti, R., et al., *A survey of methods for explaining black box models.* ACM computing surveys, 2018. **51**(5): p. 1-42.
67. Vivian, J., et al., *A comparison between grey-box models and neural networks for indoor air temperature prediction in buildings.* Journal of Building Engineering, 2024: p. 108583.
68. Yin, H., Z. Tang, and C. Yang, *Predicting hourly electricity consumption of chillers in subway stations: A comparison of support vector machine and different artificial neural networks.* Journal of Building Engineering, 2023: p. 107179.
69. Beccali, G., et al., *Is the transfer function method reliable in a European building context? A theoretical analysis and a case study in the south of Italy.* Applied thermal engineering, 2005. **25**(2-3): p. 341-357.
70. Derakhtenjani, A.S. and A.K. Athienitis, *A frequency domain transfer function methodology for thermal characterization and design for energy flexibility of zones with radiant systems.* Renewable Energy, 2021. **163**: p. 1033-1045.
71. Athienitis, A., M. Stylianou, and J. Shou, *A methodology for building thermal dynamics studies and control applications.* ASHRAE Transactions, 1990. **96**(CONF-9006117-).
72. Candanedo, J.A. and A.K. Athienitis, *Simplified linear models for predictive control of advanced solar homes with passive and active thermal storage.* 2010.
73. Candanedo, J.A., A. Allard, and A.K. Athienitis, *Predictive Control of Radiant Floor Heating and Transmitted Irradiance in a Room with High Solar Gains.* ASHRAE Transactions, 2011. **117**(2).
74. Chen, Y., A.K. Athienitis, and K.E. Galal, *A charging control strategy for active building-integrated thermal energy storage systems using frequency domain modeling.* Energy and Buildings, 2014. **84**: p. 651-661.

75. Zhuang, J., Y. Chen, and X. Chen, *A new simplified modeling method for model predictive control in a medium-sized commercial building: A case study.* Building and Environment, 2018. **127**: p. 1-12.
76. Athienitis, A., H. Sullivan, and K. Hollands, *Discrete Fourier series models for building auxiliary energy loads based on network formulation techniques.* Solar energy, 1987. **39**(3): p. 203-210.
77. Athienitis, A., *Building thermal analysis.* Electronic Book Mathcad, MathSoft, Boston, USA, 1994.
78. Wanasundara, S.N., et al., *Detecting thermal anomalies in buildings using frequency and temporal domains analysis.* Journal of Building Engineering, 2023: p. 106923.
79. Li, H., et al., *Characterizing patterns and variability of building electric load profiles in time and frequency domains.* Applied Energy, 2021. **291**: p. 116721.
80. Viot, H., et al., *Fast on-Site Measurement Campaigns and Simple Building Models Identification for Heating Control.* Energy Procedia, 2015. **78**: p. 812-817.
81. Cattarin, G., et al., *Empirical validation and local sensitivity analysis of a lumped-parameter thermal model of an outdoor test cell.* Building and Environment, 2018. **130**: p. 151-161.
82. Gouda, M., S. Danaher, and C. Underwood, *Building thermal model reduction using nonlinear constrained optimization.* Building and Environment, 2002. **37**(12): p. 1255-1265.
83. Antoulas, A.C., D.C. Sorensen, and S. Gugercin, *A survey of model reduction methods for large-scale systems*. 2000.
84. Kim, D., et al., *A methodology for generating reduced-order models for large-scale buildings using the Krylov subspace method.* Journal of Building Performance Simulation, 2020. **13**(4): p. 419-429.
85. Goyal, S. and P. Barooah, *A method for model-reduction of non-linear thermal dynamics of multi-zone buildings.* Energy and Buildings, 2012. **47**: p. 332-340.
86. Athienitis, A., M. Chandrashekar, and H. Sullivan, *Modelling and analysis of thermal networks through subnetworks for multizone passive solar buildings.* Applied mathematical modelling, 1985. **9**(2): p. 109-116.
87. Boglietti, A., et al., *Stator-winding thermal models for short-time thermal transients: Definition and validation.* IEEE Transactions on Industrial Electronics, 2015. **63**(5): p. 2713-2721.
88. Boodi, A., et al., *Building thermal-network models: a comparative analysis, recommendations, and perspectives.* Energies, 2022. **15**(4): p. 1328.
89. Deng, K., et al. *Building thermal model reduction via aggregation of states*. in *Proceedings of the 2010 American Control Conference*. 2010. IEEE.
90. Mossolly, M., K. Ghali, and N. Ghaddar, *Optimal control strategy for a multi-zone air conditioning system using a genetic algorithm.* Energy, 2009. **34**(1): p. 58-66.
91. Shin, M. and J.S. Haberl, *Thermal zoning for building HVAC design and energy simulation: A literature review.* Energy and Buildings, 2019. **203**: p. 109429.
92. Banihashemi, F., M. Weber, and W. Lang, *Model order reduction of building energy simulation models using a convolutional neural network autoencoder.* Building Environment, 2022. **207**: p. 108498.
93. Shi, Z. and W. O'Brien, *Building energy model reduction using model-cluster-reduce pipeline.* Journal of Building Performance Simulation, 2018. **11**(5): p. 553-567.

94. Shin, M. and J.S.J.J.o.B.E. Haberl, *A procedure for automating thermal zoning for building energy simulation.* 2022. **46**: p. 103780.
95. Vallianos, C., A. Athienitis, and B. Delcroix, *Automatic generation of multi-zone RC models using smart thermostat data from homes.* Energy and Buildings, 2022. **277**: p. 112571.
96. Wang, J., et al., *Development and validation of a second-order thermal network model for residential buildings.* Applied Energy, 2022. **306**: p. 118124.
97. Wang, J., et al., *Predicting home thermal dynamics using a reduced-order model and automated real-time parameter estimation.* Energy and Buildings, 2019. **198**: p. 305-317.
98. Xiao, Z., et al., *Impacts of data preprocessing and selection on energy consumption prediction model of HVAC systems based on deep learning.* Energy and Buildings, 2022. **258**: p. 111832.
99. Shen, P. and H. Wang, *Archetype building energy modeling approaches and applications: A review.* Renewable and Sustainable Energy Reviews, 2024. **199**: p. 114478.
100. Candanedo, J.A., et al., *Control-oriented archetypes: a pathway for the systematic application of advanced controls in buildings.* Journal of Building Performance Simulation, 2022: p. 1-12.
101. Privara, S., et al., *Building modeling: Selection of the most appropriate model for predictive control.* Energy and Buildings, 2012. **55**: p. 341-350.
102. Arroyo, J., F. Spiessens, and L. Helsen, *Identification of multi-zone grey-box building models for use in model predictive control.* Journal of Building Performance Simulation, 2020. **13**(4): p. 472-486.
103. Vallianos, C., et al., *Online model-based predictive control with smart thermostats: application to an experimental house in Québec.* Journal of Building Performance Simulation, 2024. **17**(1): p. 94-110.
104. Doma, A., et al., *Investigating the Thermal Performance of Canadian Houses Using Smart Thermostat Data.* ASHRAE Transactions, 2021. **127**(1).
105. Huchuk, B., S. Sanner, and W. O'Brien, *Evaluation of data-driven thermal models for multi-hour predictions using residential smart thermostat data.* Journal of Building Performance Simulation, 2022. **15**(4): p. 445-464.
106. Vallianos, C., J. Candanedo, and A. Athienitis, *Thermal modeling for control applications of 60,000 homes in North America using smart thermostat data.* Energy and Buildings, 2024. **303**: p. 113811.
107. *EN ISO 13790:2008 - Energy performance of buildings - Calculation of energy use for space heating and cooling*. 2008.
108. *EN ISO 52016-1:2017 - Energy performance of buildings - Energy needs for heating and cooling, internal temperatures and sensible and latent heat loads - Part 1: Calculation procedures*. 2017.
109. Shen, P., W. Braham, and Y. Yi, *Development of a lightweight building simulation tool using simplified zone thermal coupling for fast parametric study.* Applied Energy, 2018. **223**: p. 188-214.
110. Andrade-Cabrera, C., et al., *Ensemble Calibration of lumped parameter retrofit building models using Particle Swarm Optimization.* Energy and Buildings, 2017. **155**: p. 513-532.
111. Sigounis, A.-M., C. Vallianos, and A. Athienitis, *Model predictive control of air-based building integrated PV/T systems for optimal HVAC integration.* Renewable Energy, 2023. **212**: p. 655-668.

112. Ioannidis, Z., et al., *Modeling of double skin façades integrating photovoltaic panels and automated roller shades: Analysis of the thermal and electrical performance.* Energy and Buildings, 2017. **154**: p. 618-632.
113. Athienitis, A.K., et al., *Assessing active and passive effects of façade building integrated photovoltaics/thermal systems: Dynamic modelling and simulation.* Applied Energy, 2018. **209**: p. 355-382.
114. Guo, J., et al., *A dynamic state-space model for predicting the thermal performance of ventilated electric heating mortar blocks integrated with phase change material.* Energy and Buildings, 2021. **244**: p. 111010.
115. Petrucci, A., et al., *Development of energy aggregators for virtual communities: The energy efficiency-flexibility nexus for demand response.* Renewable Energy, 2023. **215**: p. 118975.
116. Berouine, A., et al., *A predictive control approach for thermal energy management in buildings.* Energy Reports, 2022. **8**: p. 9127-9141.
117. De Coninck, R. and L. Helsen, *Practical implementation and evaluation of model predictive control for an office building in Brussels.* Energy and Buildings, 2016. **111**: p. 290-298.
118. Joe, J., et al., *Virtual storage capability of residential buildings for sustainable smart city via model-based predictive control.* Sustainable Cities and Society, 2021. **64**: p. 102491.
119. Barone, G., et al., *A new thermal comfort model based on physiological parameters for the smart design and control of energy-efficient HVAC systems.* Renewable and Sustainable Energy Reviews, 2023. **173**.
120. Buonomano, A., et al., *Temperature and humidity adaptive control in multi-enclosed thermal zones under unexpected external disturbances.* Energy and Buildings, 2017. **135**: p. 263-285.
121. Aguilar, J., et al., *A systematic literature review on the use of artificial intelligence in energy self-management in smart buildings.* Renewable and Sustainable Energy Reviews, 2021. **151**: p. 111530.
122. Panchalingam, R. and K.C. Chan, *A state-of-the-art review on artificial intelligence for Smart Buildings.* Intelligent Buildings International, 2021. **13**(4): p. 203-226.
123. Shamsi, M.H., et al., *Feature assessment frameworks to evaluate reduced-order grey-box building energy models.* Applied Energy, 2021. **298**: p. 117174.
124. Greenacre, M., et al., *Principal component analysis.* Nature Reviews Methods Primers, 2022. **2**(1): p. 100.
125. Meng, Q., et al. *Relational autoencoder for feature extraction*. in *2017 International joint conference on neural networks (IJCNN)*. 2017. IEEE.
126. Chicco, G., R. Napoli, and F. Piglione, *Comparisons among clustering techniques for electricity customer classification.* IEEE Transactions on power systems, 2006. **21**(2): p. 933-940.
127. Gulbinas, R., A. Khosrowpour, and J. Taylor, *Segmentation and classification of commercial building occupants by energy-use efficiency and predictability.* IEEE Transactions on Smart Grid, 2015. **6**(3): p. 1414-1424.
128. Park, D.C., et al., *Electric load forecasting using an artificial neural network.* IEEE transactions on Power Systems, 1991. **6**(2): p. 442-449.
129. Chen, B.-J. and M.-W. Chang, *Load forecasting using support vector machines: A study on EUNITE competition 2001.* IEEE transactions on power systems, 2004. **19**(4): p. 1821-1830.

130. Yang, S., et al., *Experimental study of a model predictive control system for active chilled beam (ACB) air-conditioning system.* Energy and Buildings, 2019. **203**: p. 109451.
131. Taheri, S., P. Hosseini, and A. Razban, *Model predictive control of heating, ventilation, and air conditioning (HVAC) systems: A state-of-the-art review.* Journal of Building Engineering, 2022. **60**: p. 105067.
132. Zhang, C., et al., *A review of integrated radiant heating/cooling with ventilation systems-Thermal comfort and indoor air quality.* Energy and Buildings, 2020. **223**: p. 110094.
133. Lydon, G.P. and A. Schlueter, *Small-scale experiments on the operational performance of a lightweight thermally active building system.* Journal of Building Engineering, 2023. **78**: p. 107372.
134. Wang, Z., et al., *A model to compare convective and radiant heating systems for intermittent space heating.* Applied Energy, 2018. **215**: p. 211-226.
135. Ran, F., et al., *A virtual sensor based self-adjusting control for HVAC fast demand response in commercial buildings towards smart grid applications.* Applied energy, 2020. **269**: p. 115103.
136. Romaní, J., A. de Gracia, and L.F. Cabeza, *Simulation and control of thermally activated building systems (TABS).* Energy and Buildings, 2016. **127**: p. 22-42.
137. Andriamamonjy, A., D. Saelens, and R. Klein, *An automated IFC-based workflow for building energy performance simulation with Modelica.* Automation in Construction, 2018. **91**: p. 166-181.
138. Drgoňa, J., et al., *All you need to know about model predictive control for buildings.* Annual Reviews in Control, 2020. **50**: p. 190-232.
139. Chen, Q., N. Li, and W. Feng, *Model predictive control optimization for rapid response and energy efficiency based on the state-space model of a radiant floor heating system.* Energy and Buildings, 2021. **238**: p. 110832.
140. Zhang, D., et al., *Experimental study on control performance comparison between model predictive control and proportion-integral-derivative control for radiant ceiling cooling integrated with underfloor ventilation system.* Applied Thermal Engineering, 2018. **143**: p. 130-136.
141. Zhang, D., et al., *Experimental investigation on model predictive control of radiant floor cooling combined with underfloor ventilation system.* Energy, 2019. **176**: p. 23-33.
142. Zhang, D., X. Xia, and N. Cai, *A dynamic simplified model of radiant ceiling cooling integrated with underfloor ventilation system.* Applied Thermal Engineering, 2016. **106**: p. 415-422.
143. Joe, J. and P. Karava, *A model predictive control strategy to optimize the performance of radiant floor heating and cooling systems in office buildings.* Applied Energy, 2019. **245**: p. 65-77.
144. Viot, H., et al., *Model predictive control of a thermally activated building system to improve energy management of an experimental building: Part I—Modeling and measurements.* Energy and Buildings, 2018. **172**: p. 94-103.
145. Bradley, P., M. Leach, and J. Torriti, *A review of the costs and benefits of demand response for electricity in the UK.* Energy Policy, 2013. **52**: p. 312-327.
146. Hu, M., et al., *Price-responsive model predictive control of floor heating systems for demand response using building thermal mass.* Applied Thermal Engineering, 2019. **153**: p. 316-329.

147. Li, H., et al., *Integrated building envelope performance evaluation method towards nearly zero energy buildings based on operation data.* Energy and Buildings, 2022. **268**: p. 112219.
148. Ibrahim, H., et al., *Energy storage systems—Characteristics and comparisons.* 2008. **12**(5): p. 1221-1250.
149. Mariaud, A., et al., *Integrated optimisation of photovoltaic and battery storage systems for UK commercial buildings.* Applied energy, 2017. **199**: p. 466-478.
150. Solangi, N.H., et al., *MXene-based phase change materials for solar thermal energy storage.* Energy Conversion and Management, 2022. **273**: p. 116432.
151. Shoeibi, S., et al., *A comprehensive review of nano-enhanced phase change materials on solar energy applications.* Journal of Energy Storage, 2022. **50**: p. 104262.
152. Unander, F. *Energy indicators and sustainable development: the International Energy Agency approach*. in *Natural Resources Forum*. 2005. Wiley Online Library.
153. Athienitis, A., et al., *Investigation of the thermal performance of a passive solar test-room with wall latent heat storage.* Building environment, 1997. **32**(5): p. 405-410.
154. Wu, D., et al., *Multilayer assembly of phase change material and bio-based concrete: A passive envelope to improve the energy and hygrothermal performance of buildings.* Energy Conversion and Management, 2022. **257**: p. 115454.
155. Forzano, C., et al., *Building integrating phase change materials: A dynamic hygrothermal simulation model for system analysis.* Journal of sustainable Development of Energy, Water Environment Systems, 2019. **7**(2): p. 325-342.
156. Buonomano, A. and F. Guarino, *The impact of thermophysical properties and hysteresis effects on the energy performance simulation of PCM wallboards: Experimental studies, modelling, and validation.* Renewable and Sustainable Energy Reviews, 2020. **126**: p. 109807.
157. Zhou, G. and M. Pang, *Experimental investigations on the performance of a collector–storage wall system using phase change materials.* Energy Conversion and Management, 2015. **105**: p. 178-188.
158. Morovat, N., et al., *Simulation and performance analysis of an active PCM-heat exchanger intended for building operation optimization.* Energy and Buildings, 2019. **199**: p. 47-61.
159. Bullich-Massagué, E., et al., *A review of energy storage technologies for large scale photovoltaic power plants.* Applied Energy, 2020. **274**: p. 115213.
160. Vassiliades, C., et al., *Assessment of an innovative plug and play PV/T system integrated in a prefabricated house unit: Active and passive behaviour and life cycle cost analysis.* Renewable Energy, 2022. **186**: p. 845-863.
161. Dermardiros, V. and P. Scott Bucking PhD, *Energy performance, comfort, and lessons learned from an institutional building designed for net zero energy.* ASHRAE Transactions, 2019. **125**: p. 682-695.
162. Li, Z., et al., *Multi-objective energy and exergy optimization of different configurations of hybrid earth-air heat exchanger and building integrated photovoltaic/thermal system.* 2019. **195**: p. 1098-1110.
163. Tomar, V., et al., *Performance analysis of a prototype solar photovoltaic/wickless heat pipe embedded aluminum curtain wall-heat pump water heating system.* 2022. **258**: p. 115559.
164. Lidula, N. and A. Rajapakse, *Microgrids research: A review of experimental microgrids and test systems.* Renewable and Sustainable Energy Reviews, 2011. **15**(1): p. 186-202.

165. Akram, U., et al., *A review on rapid responsive energy storage technologies for frequency regulation in modern power systems.* Renewable and Sustainable Energy Reviews, 2020. **120**: p. 109626.
166. Delfanti, M., D. Falabretti, and M. Merlo, *Energy storage for PV power plant dispatching.* Renewable Energy, 2015. **80**: p. 61-72.
167. Parra, D., et al., *An interdisciplinary review of energy storage for communities: Challenges and perspectives.* Renewable and Sustainable Energy Reviews, 2017. **79**: p. 730-749.
168. Cabrera-Tobar, A., et al., *Review of advanced grid requirements for the integration of large scale photovoltaic power plants in the transmission system.* Renewable and Sustainable Energy Reviews, 2016. **62**: p. 971-987.
169. Code, T., *Network and system rules of the German transmission system operators.* VDN-ev beim VDEW, 2007.
170. Energinet, D., *Technical regulation 3.2. 2 for PV power plants with a power output above 11 kW.* Energinet, Fredericia, Denmark, Tech. Rep, 2015.
171. Brisebois, J. and N. Aubut. *Wind farm inertia emulation to fulfill Hydro-Québec's specific need*. in *2011 IEEE Power and Energy Society General Meeting*. 2011. IEEE.
172. Li, B., et al., *Review on photovoltaic with battery energy storage system for power supply to buildings: Challenges and opportunities.* Journal of Energy Storage, 2023. **61**: p. 106763.
173. Dai, R., R. Esmaeilbeigi, and H. Charkhgard, *The utilization of shared energy storage in energy systems: A comprehensive review.* IEEE Transactions on Smart Grid, 2021. **12**(4): p. 3163-3174.
174. Nair, U.R., et al., *Grid congestion mitigation and battery degradation minimisation using model predictive control in PV-based microgrid.* IEEE Transactions on Energy Conversion, 2020. **36**(2): p. 1500-1509.
175. Talent, O. and H. Du, *Optimal sizing and energy scheduling of photovoltaic-battery systems under different tariff structures.* Renewable energy, 2018. **129**: p. 513-526.
176. Parra, D., et al., *Optimum community energy storage system for PV energy time-shift.* Applied Energy, 2015. **137**: p. 576-587.
177. Parra, D., et al., *Optimum community energy storage system for demand load shifting.* Applied Energy, 2016. **174**: p. 130-143.
178. Parra, D., et al., *Optimum community energy storage for renewable energy and demand load management.* Applied Energy, 2017. **200**: p. 358-369.
179. Zhou, J., et al., *Economic and resilience benefit analysis of incorporating battery storage to photovoltaic array generation.* Renewable Energy, 2019. **135**: p. 652-662.
180. Sepúlveda-Mora, S.B. and S. Hegedus, *Resilience analysis of renewable microgrids for commercial buildings with different usage patterns and weather conditions.* Renewable Energy, 2022. **192**: p. 731-744.
181. Kok, K. and S. Widergren, *A society of devices: Integrating intelligent distributed resources with transactive energy.* IEEE Power Energy Magazine, 2016. **14**(3): p. 34-45.
182. Ahmethodzic, L. and M. Music, *Comprehensive review of trends in microgrid control.* Renewable energy focus, 2021. **38**: p. 84-96.
183. Li, R., et al., *Ten questions concerning energy flexibility in buildings.* Building and Environment, 2022. **223**: p. 109461.
184. Council, T., *Gridwise transactive energy framework version 1.0.* The GridWise Architecture Council, Tech. Rep, 2015.

185. Tushar, W., et al., *Peer-to-peer energy systems for connected communities: A review of recent advances and emerging challenges.* Applied Energy, 2021. **282**: p. 116131.
186. Onumanyi, A.J., et al., *Transactive energy: State-of-the-art in control strategies, architectures, and simulators.* 2021. **9**: p. 131552-131573.
187. Albadi, M.H. and E.F. El-Saadany. *Demand response in electricity markets: An overview*. in *2007 IEEE power engineering society general meeting*. 2007. IEEE.
188. Faruqui, A. and S. Sergici, *Household response to dynamic pricing of electricity: a survey of 15 experiments.* Journal of regulatory Economics, 2010. **38**(2): p. 193-225.
189. Shirani, F., et al., *'I'm the smart meter': Perceptions of smart technology amongst vulnerable consumers.* Energy Policy, 2020. **144**: p. 111637.
190. Chamaret, C., V. Steyer, and J.C. Mayer, *"Hands off my meter!" when municipalities resist smart meters: Linking arguments and degrees of resistance.* Energy Policy, 2020. **144**: p. 111556.
191. Nilsson, A., et al., *Household responsiveness to residential demand response strategies: Results and policy implications from a Swedish field study.* Energy policy, 2018. **122**: p. 273-286.
192. Nojavan, S., K. Zare, and B. Mohammadi-Ivatloo, *Optimal stochastic energy management of retailer based on selling price determination under smart grid environment in the presence of demand response program.* Applied Energy, 2017. **187**: p. 449-464.
193. Balijepalli, V.M., et al. *Review of demand response under smart grid paradigm*. in *ISGT2011-India*. 2011. IEEE.
194. Pina, A., C. Silva, and P. Ferrão, *The impact of demand side management strategies in the penetration of renewable electricity.* Energy, 2012. **41**(1): p. 128-137.
195. Boisvert, R.N., et al., *Customer response to RTP in competitive markets: A study of Niagara Mohawk's standard offer tariff.* The Energy Journal, 2007: p. 53-74.
196. Doostizadeh, M. and H. Ghasemi, *A day-ahead electricity pricing model based on smart metering and demand-side management.* Energy, 2012. **46**(1): p. 221-230.
197. Dagoumas, A.S. and M.L. Polemis, *An integrated model for assessing electricity retailer's profitability with demand response.* Applied Energy, 2017. **198**: p. 49-64.
198. Petrucci, A., et al., *Modelling of a multi-stage energy management control routine for energy demand forecasting, flexibility, and optimization of smart communities using a Recurrent Neural Network.* Energy Conversion and Management, 2022. **268**: p. 115995.
199. Program, E.; Available from: https://www.monecowatt.fr/.
200. Mariano-Hernández, D., et al., *A review of strategies for building energy management system: Model predictive control, demand side management, optimization, and fault detect & diagnosis.* Journal of Building Engineering, 2021. **33**: p. 101692.
201. Candanedo, J.A., et al., *Control-oriented archetypes: a pathway for the systematic application of advanced controls in buildings.* 2022: p. 1-12.
202. Wu, L., et al., *Economic model predictive control of integrated energy systems: A multi-time-scale framework.* Applied Energy, 2022. **328**: p. 120187.
203. Lefebure, N., et al., *Distributed model predictive control of buildings and energy hubs.* Energy Buildings, 2022. **259**: p. 111806.
204. Niveditha, N. and M.R. Singaravel, *Optimal sizing of hybrid PV–Wind–Battery storage system for Net Zero Energy Buildings to reduce grid burden.* Applied Energy, 2022. **324**: p. 119713.

205. Barbour, E. and M.C. González, *Projecting battery adoption in the prosumer era.* Applied Energy, 2018. **215**: p. 356-370.
206. Hemmati, R. and H. Saboori, *Stochastic optimal battery storage sizing and scheduling in home energy management systems equipped with solar photovoltaic panels.* Energy and Buildings, 2017. **152**: p. 290-300.
207. O'Shaughnessy, E., et al., *Solar plus: Optimization of distributed solar PV through battery storage and dispatchable load in residential buildings.* Applied Energy, 2018. **213**: p. 11-21.
208. Ashtiani, M.N., et al., *Techno-economic analysis of a grid-connected PV/battery system using the teaching-learning-based optimization algorithm.* Solar Energy, 2020. **203**: p. 69-82.
209. Koskela, J., A. Rautiainen, and P. Järventausta, *Using electrical energy storage in residential buildings–Sizing of battery and photovoltaic panels based on electricity cost optimization.* Applied Energy, 2019. **239**: p. 1175-1189.
210. Zhang, Y., T. Ma, and H. Yang, *A review on capacity sizing and operation strategy of grid-connected photovoltaic battery systems.* Energy and Built Environment, 2024. **5**(4): p. 500-516.
211. Schütz, T., et al., *Comparison of clustering algorithms for the selection of typical demand days for energy system synthesis.* Renewable Energy, 2018. **129**: p. 570-582.
212. Wakui, T., H. Kawayoshi, and R. Yokoyama, *Optimal structural design of residential power and heat supply devices in consideration of operational and capital recovery constraints.* Applied Energy, 2016. **163**: p. 118-133.
213. Wakui, T. and R. Yokoyama, *Optimal structural design of residential cogeneration systems in consideration of their operating restrictions.* Energy, 2014. **64**: p. 719-733.
214. Wakui, T. and R. Yokoyama, *Optimal structural design of residential cogeneration systems with battery based on improved solution method for mixed-integer linear programming.* Energy, 2015. **84**: p. 106-120.
215. Moradi, S., et al., *Optimal integrated sizing and planning of hubs with midsize/large CHP units considering reliability of supply.* Energy Conversion and Management, 2017. **148**: p. 974-992.
216. Schütz, T., et al., *Optimal design of energy conversion units for residential buildings considering German market conditions.* Energy, 2017. **139**: p. 895-915.
217. Schütz, T., et al., *Optimal design of energy conversion units and envelopes for residential building retrofits using a comprehensive MILP model.* Applied Energy, 2017. **185**: p. 1-15.
218. Tostado-Véliz, M., D. Icaza-Alvarez, and F. Jurado, *A novel methodology for optimal sizing photovoltaic-battery systems in smart homes considering grid outages and demand response.* Renewable Energy, 2021. **170**: p. 884-896.
219. Merkel, E., R. McKenna, and W. Fichtner, *Optimisation of the capacity and the dispatch of decentralised micro-CHP systems: A case study for the UK.* Applied Energy, 2015. **140**: p. 120-134.
220. Buoro, D., et al., *Optimal synthesis and operation of advanced energy supply systems for standard and domotic home.* Energy Conversion and Management, 2012. **60**: p. 96-105.
221. Dermardiros, V., *Data-Driven Optimized Operation of Buildings with Intermittent Renewables and Application to a Net-Zero Energy Library*. 2020, Concordia University.
222. Morovat, N., et al., *Heuristic model predictive control implementation to activate energy flexibility in a fully electric school building.* Energy, 2024. **296**: p. 131126.

223. Li, R., et al., *Ten questions concerning energy flexibility in buildings.* 2022. **223**: p. 109461.
224. Salom, J., et al., *Analysis of load match and grid interaction indicators in net zero energy buildings with simulated and monitored data.* Applied Energy, 2014. **136**: p. 119-131.
225. Kazem, H.A., T. Khatib, and K. Sopian, *Sizing of a standalone photovoltaic/battery system at minimum cost for remote housing electrification in Sohar, Oman.* Energy and Buildings, 2013. **61**: p. 108-115.
226. Khatib, T., I.A. Ibrahim, and A. Mohamed, *A review on sizing methodologies of photovoltaic array and storage battery in a standalone photovoltaic system.* Energy Conversion and Management, 2016. **120**: p. 430-448.
227. Ruan, J., et al., *Time-varying price elasticity of demand estimation for demand-side smart dynamic pricing.* Applied Energy, 2022. **322**: p. 119520.
228. Razmara, M., et al., *Building-to-grid predictive power flow control for demand response and demand flexibility programs.* Applied Energy, 2017. **203**: p. 128-141.
229. Li, H., et al., *Data-driven key performance indicators and datasets for building energy flexibility: A review and perspectives.* Applied Energy, 2023. **343**: p. 121217.
230. Liu, M. and P. Heiselberg, *Energy flexibility of a nearly zero-energy building with weather predictive control on a convective building energy system and evaluated with different metrics.* Applied Energy, 2019. **233**: p. 764-775.
231. Gaucher-Loksts, E., A. Athienitis, and M. Ouf, *Design and energy flexibility analysis for building integrated photovoltaics-heat pump combinations in a house.* Renewable Energy, 2022. **195**: p. 872-884.
232. Kotsiantis, S.B., D. Kanellopoulos, and P.E. Pintelas, *Data preprocessing for supervised leaning.* International journal of computer science, 2006. **1**(2): p. 111-117.
233. Lukasik, M., et al. *Does label smoothing mitigate label noise?* in *International Conference on Machine Learning*. 2020. PMLR.
234. Schafer, R.W., *What is a Savitzky-Golay filter?[lecture notes].* IEEE Signal processing magazine, 2011. **28**(4): p. 111-117.
235. Simpkins, A., *System Identification: Theory for the User, 2nd Edition (Ljung, L.; 1999) [On the Shelf]*. IEEE Robotics & Automation Magazine, 2012. **19**(2): p. 95-96.
236. Melikov, A., et al., *Air temperature fluctuations in rooms.* 1997. **32**(2): p. 101-114.
237. Maturo, A., A. Athienitis, and B. Delcroix, *A data-driven frequency domain system identification approach to define house archetypes and flexibility.*
238. Beccali, G., et al., *Single thermal zone balance solved by transfer function method.* Energy and Buildings, 2005. **37**(12): p. 1268-1277.
239. Ljung, L., *System identification*, in *Signal analysis and prediction*. 1998, Springer. p. 163-173.
240. Garnier, H., M. Mensler, and A. Richard, *Continuous-time model identification from sampled data: implementation issues and performance evaluation.* International journal of Control, 2003. **76**(13): p. 1337-1357.
241. Newey, W.K., *Efficient instrumental variables estimation of nonlinear models.* Econometrica: Journal of the Econometric Society, 1990: p. 809-837.
242. Fornasini, E. and G. Marchesini, *Doubly-indexed dynamical systems: State-space models and structural properties.* Mathematical systems theory, 1978. **12**(1): p. 59-72.
243. Merikoski, J.K., et al., *A best upper bound for the 2-norm condition number of a matrix.* Linear algebra and its applications, 1997. **254**(1-3): p. 355-365.

244. Smith, G.D., *Numerical solution of partial differential equations: finite difference methods*. 1985: Oxford university press.
245. Petelet, M., et al., *Latin hypercube sampling with inequality constraints.* AStA Advances in Statistical Analysis, 2010. **94**(4): p. 325-339.
246. Privara, S., et al., *Building modeling as a crucial part for building predictive control.* Energy and Buildings, 2013. **56**: p. 8-22.
247. Baba, F.M., et al., *Calibration of building model based on indoor temperature for overheating assessment using genetic algorithm: Methodology, evaluation criteria, and case study.* Building Environment, 2022. **207**: p. 108518.
248. Zhan, S. and A. Chong, *Data requirements and performance evaluation of model predictive control in buildings: A modeling perspective.* Renewable and Sustainable Energy Reviews, 2021. **142**: p. 110835.
249. Millette, J., S. Sansregret, and A. Daoud. *SIMEB: Simplified interface to DOE2 and EnergyPlus-A user's perspective–Case study of an existing building*. in *Proceeding of the 12th International Building Performance Simulation Asssociation (IBPSA) Conference, Sydney, Australia*. 2011.
250. Vivian, J., et al., *An evaluation of the suitability of lumped-capacitance models in calculating energy needs and thermal behaviour of buildings.* Energy and Buildings, 2017. **150**: p. 447-465.
251. Long, J., et al., *Study on energy-saving operation of a combined heating system of solar hot water and air source heat pump.* 2021. **229**: p. 113624.
252. Safa, A.A., et al., *Performance of two-stage variable capacity air source heat pump: Field performance results and TRNSYS simulation.* 2015. **94**: p. 80-90.
253. Kamel, R. and A. Fung, *Theoretical Estimation of the Performance of a Photovoltaic-Thermal Collector (PV/T) System Coupled with a Heat Pump in a Sustainable House in Toronto.* ASHRAE Transactions, 2014. **120**(1).
254. Stein, J.S., et al. *PVLIB: Open source photovoltaic performance modeling functions for Matlab and Python*. in *2016 ieee 43rd photovoltaic specialists conference (PVSC)*. 2016. IEEE.
255. Safa, A.A., A.S. Fung, and R. Kumar, *Performance of two-stage variable capacity air source heat pump: Field performance results and TRNSYS simulation.* Energy and Buildings, 2015. **94**: p. 80-90.
256. Li, Y., et al., *Grey-box modeling and application for building energy simulations - A critical review.* Renewable and Sustainable Energy Reviews, 2021. **146**: p. 111174.
257. Wang, Z. and Y. Chen, *Data-driven modeling of building thermal dynamics: Methodology and state of the art.* Energy and Buildings, 2019. **203**: p. 109405.
258. Maturo, A., A. Buonomano, and A. Athienitis, *Optimizing energy flexibility through electricity price-responsiveness and thermal load management in buildings with convective and radiant heating systems.* Energy and Buildings, 2025: p. 115355.
259. Serasinghe, R., N. Long, and J.D. Clark, *Parameter identification methods for low-order gray box building energy models: A critical review.* Energy and Buildings, 2024: p. 114123.
260. Li, X. and J. Wen, *Review of building energy modeling for control and operation.* Renewable and Sustainable Energy Reviews, 2014. **37**: p. 517-537.
261. Buonomano, A. and A. Palombo, *Building energy performance analysis by an in-house developed dynamic simulation code: An investigation for different case studies.* Applied Energy, 2014. **113**: p. 788-807.

262. Maturo, A., et al., *A novel multi-level predictive management strategy to optimize phase-change energy storage and building-integrated renewable technologies operation under dynamic tariffs.* Energy Conversion Management, 2023. **291**: p. 117220.
263. Wang, L. and Q. Chen, *Validation of a coupled multizone-CFD program for building airflow and contaminant transport simulations.* HVAC R Research, 2007. **13**(2): p. 267-281.
264. Joe, J. and P. Karava, *Agent-based system identification for control-oriented building models.* Journal of Building Performance Simulation, 2017. **10**(2): p. 183-204.
265. American Society of Heating, R.a.A.-C.E.A., *ANSI/ASHRAE Standard 62.1-2022: Ventilation for Acceptable Indoor Air Quality*. 2022, ASHRAE.
266. Fransson, V., H. Bagge, and D. Johansson, *Impact of variations in residential use of household electricity on the energy and power demand for space heating–Variations from measurements in 1000 apartments.* Applied Energy, 2019. **254**: p. 113599.
267. Wang, Z., T. Hong, and M.A. Piette, *Data fusion in predicting internal heat gains for office buildings through a deep learning approach.* Applied Energy, 2019. **240**: p. 386-398.
268. Yang, J., et al., *Energy performance model development and occupancy number identification of institutional buildings.* Energy and Buildings, 2016. **123**: p. 192-204.
269. Sorrell, S., J. Dimitropoulos, and M. Sommerville, *Empirical estimates of the direct rebound effect: A review.* Energy policy, 2009. **37**(4): p. 1356-1371.
270. Szekely, G.J. and M.L. Rizzo, *Hierarchical clustering via joint between-within distances: Extending Ward's minimum variance method.* Journal of classification, 2005. **22**(2): p. 151-184.
271. Jaeger, A. and D. Banks, *Cluster analysis: A modern statistical review.* Wiley Interdisciplinary Reviews: Computational Statistics, 2023. **15**(3): p. e1597.
272. Canada, N.R., *CanmetENERGY in Varennes' Rooftop Weather Station Dataset [Available upon request]*. 2023.
273. Ashrae, A.H.F. and G. Atlanta, *American society of Heating.* Refrigerating Air-Conditioning Engineers, 2009. **1**.
274. Hydro-Québec. *Electricity Rates*. 2023; Available from: https://www.hydroquebec.com/data/documents-donnees/pdf/electricity-rates.pdf.
275. Soleimani-Mohseni, M., B. Thomas, and P. Fahlen, *Estimation of operative temperature in buildings using artificial neural networks.* Energy and Buildings, 2006. **38**(6): p. 635-640.
276. Maturo, A., et al., *Automated model order reduction for building thermal load prediction using smart thermostats data.* Journal of Building Engineering, 2024: p. 110492.
277. Georges, E., et al., *Residential heat pump as flexible load for direct control service with parametrized duration and rebound effect.* Applied Energy, 2017. **187**: p. 140-153.
278. Meng, Q., et al., *Load rebound suppression strategy and demand response potential of thermal storage HVAC systems: An experimental and simulation study.* Journal of Energy Storage, 2023. **73**: p. 108872.
279. Du, Q., et al., *The energy rebound effect of residential buildings: Evidence from urban and rural areas in China.* Energy Policy, 2021. **153**: p. 112235.
280. Klanatsky, P., F. Veynandt, and C. Heschl, *Grey-box model for model predictive control of buildings.* Energy and Buildings, 2023. **300**: p. 113624.

281. Teichgraeber, H. and A.R. Brandt, *Clustering methods to find representative periods for the optimization of energy systems: An initial framework and comparison.* 2019. **239**: p. 1283-1293.
282. José-García, A. and W. Gómez-Flores, *CVIK: A Matlab-based cluster validity index toolbox for automatic data clustering.* SoftwareX, 2023. **22**: p. 101359.
283. Liu, Y., et al. *Understanding of internal clustering validation measures*. in *2010 IEEE international conference on data mining*. 2010. Ieee.
284. Rousseeuw, P.J., *Silhouettes: a graphical aid to the interpretation and validation of cluster analysis.* Journal of computational applied mathematics, 1987. **20**: p. 53-65.
285. Davies, D.L. and D.W. Bouldin, *A cluster separation measure.* IEEE transactions on pattern analysis machine intelligence, 1979(2): p. 224-227.
286. Xie, X.L. and G. Beni, *A validity measure for fuzzy clustering.* IEEE Transactions on Pattern Analysis Machine Intelligence, 1991. **13**(08): p. 841-847.
287. Dunn, J.C., *A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters.* 1973.
288. Maturo, A., A. Athienitis, and B. Delcroix, *A novel model reduction and calibration methodology to define lumped parameters of building thermal models.*
289. Hydro-Québec. *Electricity Rates Effective April 1, 2024.* 2024; Available from: https://www.hydroquebec.com/data/documents-donnees/pdf/electricity-rates.pdf.
290. dos Santos, G.H. and N. Mendes, *Analysis of numerical methods and simulation time step effects on the prediction of building thermal performance.* Applied thermal engineering, 2004. **24**(8-9): p. 1129-1142.

Appendix A: Results of the MOR methodology applied to other buildings

This section demonstrates the practical application of the proposed methodology for providing reduced-order models for thermal load prediction in buildings equipped with smart thermostats. It presents the model order reduction results for other two different buildings, detailing the aggregation procedure, temperature trends for each building, and the calibrated model for a prediction horizon of 12-hrs.

This section serves as a proof of concept for the structure identification of the selected buildings, also showing that same model archetypes/structures can be applied to different buildings.

Building #1

Table A.1: Building #1 – General information.

Location	**Building type**	**Heating terminals**	**Number of Thermostats**
Trois-Rivières, Québec	Bungalow: 1 floor and basement	Baseboard heaters and fan coils	10

Table A.2: Building #1 – Identified thermal zones with $\widetilde{\Delta T_{set}} = 1\ K$ and $\widetilde{\Delta T_{set}} = 1.5\ K$ showing the detailed attributes of the thermal zones.

<table>
<tr><th colspan="2">Identified “Dominant Zones”</th><th colspan="2">Building thermostat attributes</th></tr>
<tr><th>Low aggregation
$\widetilde{\Delta T_{set}} = 1\ K$</th><th>High aggregation
$\widetilde{\Delta T_{set}} = 1.5\ K$</th><th>Location</th><th>Thermal Zones</th></tr>
<tr><td colspan="2" rowspan="2">Neglected</td><td>Basement</td><td>1. Basement Laundry</td></tr>
<tr><td>Basement</td><td>2. Basement Workspace</td></tr>
<tr><td colspan="2">Zone 1</td><td>Basement</td><td>3. Basement Living</td></tr>
<tr><td colspan="2" rowspan="4">Zone 2</td><td>Basement</td><td>4. Basement Office</td></tr>
<tr><td>Basement</td><td>5. Basement Washroom</td></tr>
<tr><td>1st floor</td><td>6. Main Living</td></tr>
<tr><td>1st floor</td><td>7. Main Dining</td></tr>
<tr><td colspan="2" rowspan="3">Zone 3</td><td>1st floor</td><td>8. Master Bedroom</td></tr>
<tr><td>1st floor</td><td>9. Bedroom 1</td></tr>
<tr><td>1st floor</td><td>10. Bedroom 2</td></tr>
</table>

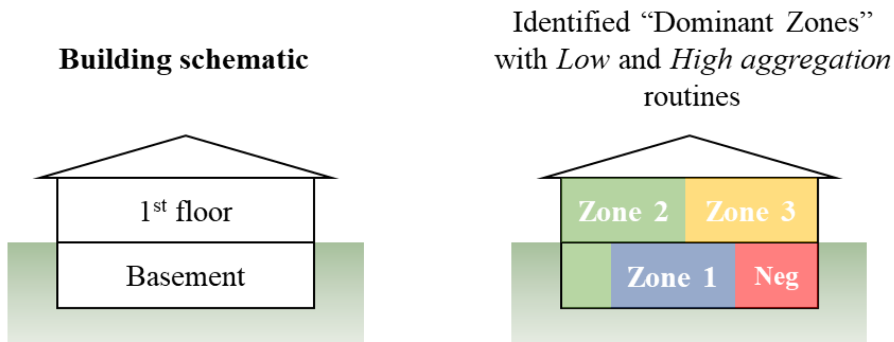


Figure A.1: Building #1 –Floor schematic representing the identified thermal zones after Low aggregation and High aggregation routine respectively.

Table A.3: Building #1 – FIT index and order of the different models generated by varying $\widetilde{\Delta T_{set}}$ and the prediction horizon.

Prediction horizon	*Low aggregation (3 dominant zones)* $\widetilde{\Delta T_{set}} = 1\ K$			*High aggregation (1 dominant zone)* $\widetilde{\Delta T_{set}} = 1.5\ K$	
	***FIT index* [%]**	***RMSE* [°C]**		***FIT index* [%]**	***RMSE* [°C]**
12-hrs	25.78	0.47	=	25.78	0.47

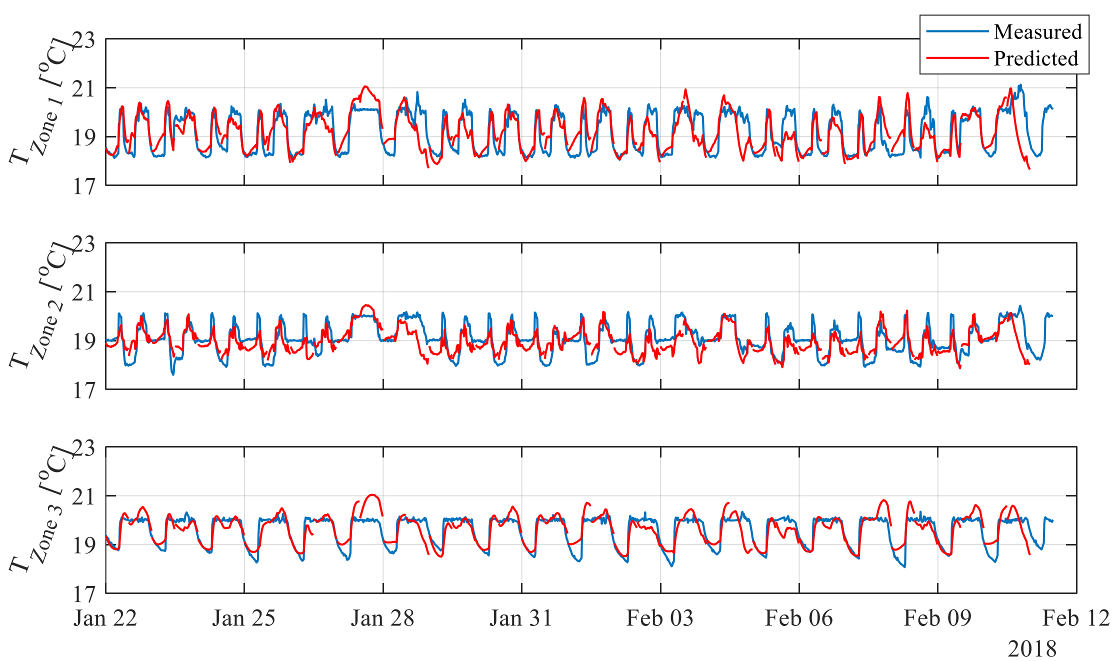


Figure A.2: Building #1 – Measured and predicted zone temperatures for 12-hrs ahead in case of *Low* and *High aggregation* during the validation period: metrics of $RMSE$ = 0.47 and FIT = 25.8.

Building #1 model matrices calibrated with the optimal structure determined during the *Low* and *High aggregation* routines with prediction horizon set at 12-hrs.

$$A = \begin{bmatrix} 0.7402 & 0.2520 & 0.0011 & 0 \\ 0.3536 & 0.3111 & 0.0014 & 0.3287 \\ 0.0104 & 0.0898 & 0.8998 & 0 \\ 0 & 0.0032 & 0 & 0.9967 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.0067 & 7.2376 \cdot 10^{-4} & 1.4007 \cdot 10^{-4} & 0 & 0 \\ 0.0053 & 0 & 0 & 3.8961 \cdot 10^{-4} & 0 \\ 0 & 0 & 0 & 0 & 3.3347 \cdot 10^{-4} \\ 1.9544 \cdot 10^{-4} & 3.4961 \cdot 10^{-8} & 0 & 0 & 0 \end{bmatrix}$$

Building #2

Table A.4: Building #2 – General information.

Location	**Building type**	**Heating terminals**	**Number of Thermostats**
Trois-Rivières, Québec	Cottage: 2 floors and basement	Baseboard heaters	11

Table A.5: Building #2 – Identified thermal zones with $\widetilde{\Delta T_{set}} = 1\ K$ and $\widetilde{\Delta T_{set}} = 1.5\ K$ showing the detailed attributes of the thermal zones.

Identified "Dominant Zones"		***Building thermostat attributes***	
Low aggregation $\widetilde{\Delta T_{set}} = 1\ K$	*High aggregation* $\widetilde{\Delta T_{set}} = 1.5\ K$	*Location*	*Thermal Zones*
Neglected	Neglected	Basement	*1.* Basement Bathroom
		1st floor	*2.* Entrance
		2nd floor	*3.* Upstairs Office
Zone 1	Zone 1	Basement	*4.* Basement Bedroom
Zone 2		Basement	*5.* Basement Living
		Basement	*6.* Basement Electric
Zone 3		1st floor	*7.* Main Living
		1st floor	*8.* Main Dining
		1st floor	*9.* Kitchen
Zone 4		2nd floor	*10.* Master Bedroom
Zone 5		2nd floor	*11.* Bedroom 1

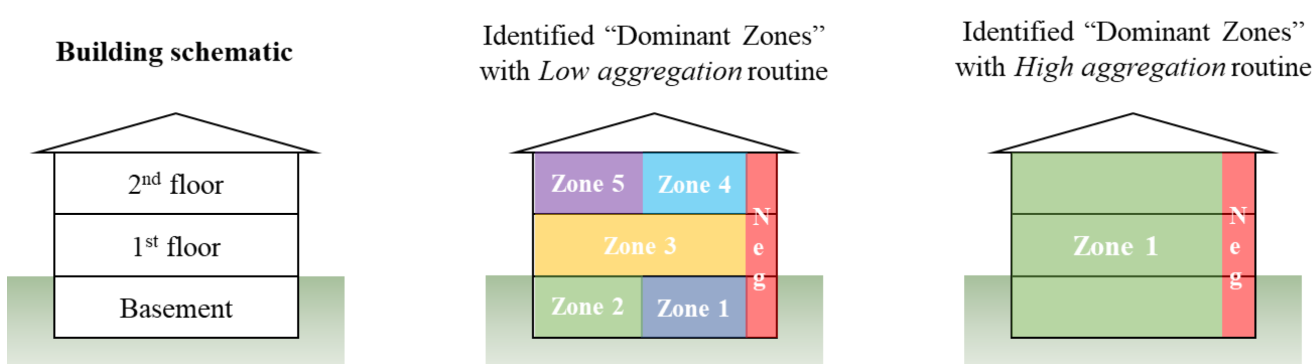


Figure A.3: Building #2 – Floor schematic representing the identified thermal zones after *Low aggregation* and *High aggregation* routine respectively.

Table A.6: Building #2 – FIT index and order of the different models generated by varying $\widetilde{\Delta T_{set}}$ and the prediction horizon.

Prediction horizon	*Low aggregation (3 dominant zones)* $\widetilde{\Delta T_{set}} = 1\ K$			*High aggregation (1 dominant zone)* $\widetilde{\Delta T_{set}} = 1.5\ K$	
	***FIT index* [%]**	***RMSE* [ºC]**		***FIT index* [%]**	***RMSE* [ºC]**
12-hrs	47.89	0.55	≠	48.74	0.48

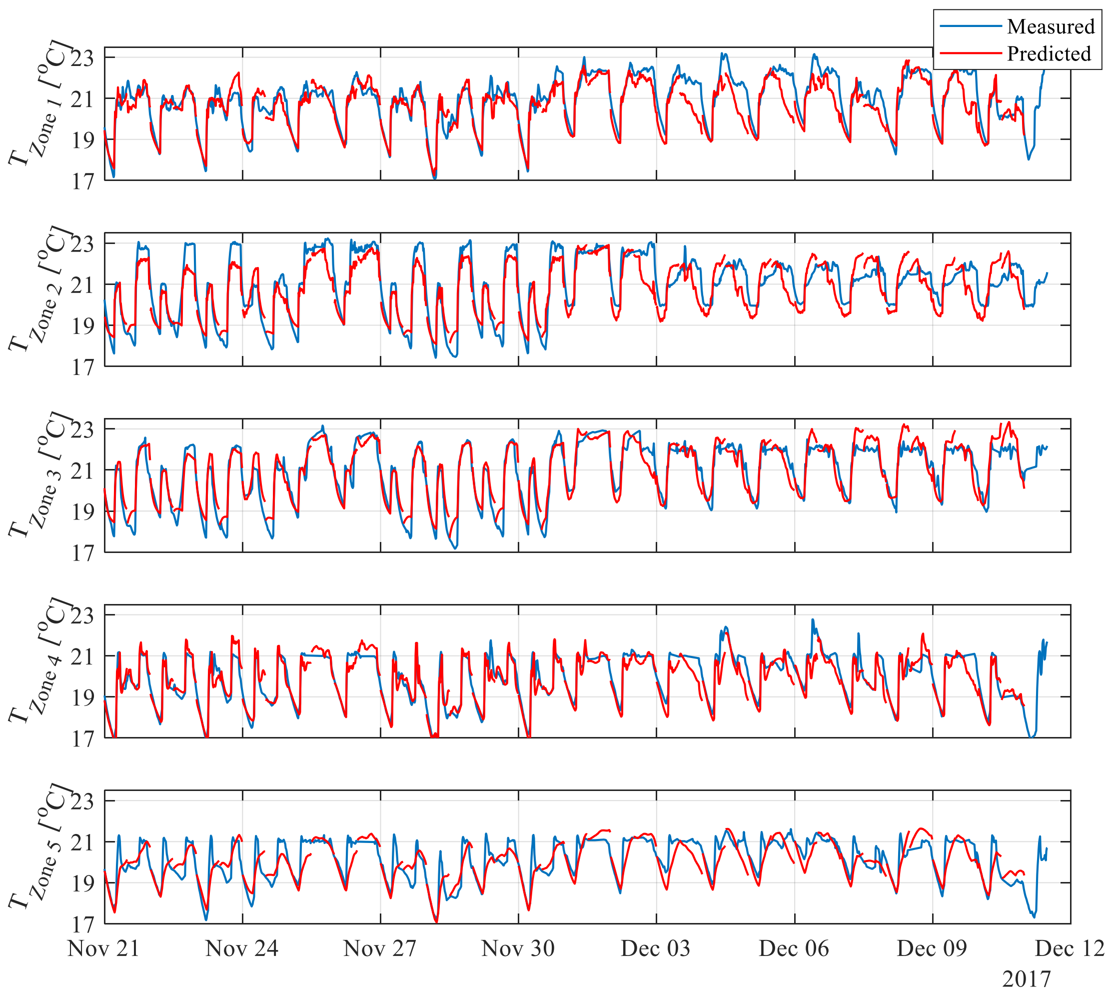


Figure A.4: Building #2 – Measured and predicted zone temperatures for 12-hrs ahead in case of *Low aggregation* during the validation period: metrics of $RMSE$ = 0.55 and FIT = 47.89.

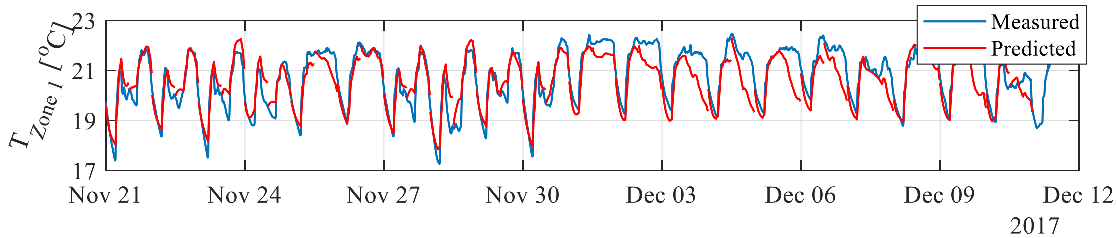


Figure A.5: Building #2 – Measured and predicted zone temperatures for 12-hrs ahead in case of *High aggregation* during the validation period: metrics of $RMSE$ = 0.48 and FIT = 48.74.

Building #2 model matrices calibrated with the optimal structure determined during the *Low* and *High aggregation* routines with prediction horizon set at 12-hrs.

$$A = \begin{bmatrix} 0.2162 & 0.2761 & 0 & 0 & 0.4967 & 0 \\ 5.0295 \cdot 10^{-7} & 1.7545 \cdot 10^{-5} & 0.5685 & 0 & 0 & 0.4072 \\ 0 & 0.3935 & 0.6065 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.4472 & 0.5357 & 0 \\ 0.0903 & 0 & 0 & 1.2911 \cdot 10^{-7} & 0.9072 & 0 \\ 0 & 0.0135 & 0 & 0 & 0 & 0.9865 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.0110 & 0.0014 & 5.9379 \cdot 10^{-4} & 0 & 0 & 0 & 0 \\ 0.0243 & 0 & 0 & 6.8127 \cdot 10^{-4} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 6.7228 \cdot 10^{-4} & 0 & 0 \\ 0.0171 & 0.0015 & 0 & 0 & 0 & 0.0021 & 0 \\ 0.0025 & 0 & 0 & 0 & 0 & 0 & 7.8961 \cdot 10^{-5} \\ 2.2535 \cdot 10^{-8} & 7.2197 \cdot 10^{-5} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$A = \begin{bmatrix} 0.8975 & 0.0955 \\ 0.0171 & 0.9829 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.0070 & 2.2518 \cdot 10^{-4} & 8.7999 \cdot 10^{-5} \\ 6.3388 \cdot 10^{-9} & 7.1722 \cdot 10^{-5} & 0 \end{bmatrix}$$

Furthermore, in this section are provided the state space A and B matrices of the developed models, for the building described in Section 3.3, in case of *Low aggregation* and *High aggregation* and for prediction horizon of 6, 12 and 24-hrs.

- *Low aggregation* with prediction horizon set at 6-hrs.

$$A = \begin{bmatrix} 0.9867 & 0.0124 & 0 & 0 \\ 0.2997 & 0.0184 & 0.6790 & 7.7421 \cdot 10^{-4} \\ 0 & 0.1814 & 0.8103 & 0 \\ 0 & 0.0666 & 0 & 0.3355 \end{bmatrix}$$

$$B = \begin{bmatrix} 8.6237 \cdot 10^{-4} & 0 & 4.6701 \cdot 10^{-5} & 0 & 0 \\ 0.0022 & 7.5193 \cdot 10^{-5} & 0 & 4.4414 \cdot 10^{-4} & 0 \\ 0.0083 & 0 & 0 & 0 & 2.5345 \cdot 10^{-4} \\ 0.5979 & 0.2566 & 0 & 0 & 0 \end{bmatrix}$$

- *Low aggregation* with prediction horizon set at 12-hrs.

$$A = \begin{bmatrix} 0.1741 & 0.8259 & 0 & 0 \\ 0.3746 & 9.7686 \cdot 10^{-6} & 0.3307 & 0.2885 \\ 0 & 0.2613 & 0.7278 & 0 \\ 0 & 0.0269 & 0 & 0.9731 \end{bmatrix}$$

$$B = \begin{bmatrix} 8.780 \cdot 10^{-8} & 0 & 4.9794 \cdot 10^{-4} & 0 & 0 \\ 0.0063 & 3.3024 \cdot 10^{-7} & 0 & 2.0656 \cdot 10^{-4} & 0 \\ 0.0109 & 0 & 0 & 0 & 1.6953 \cdot 10^{-9} \\ 4.6344 \cdot 10^{-9} & 1.1052 \cdot 10^{-4} & 0 & 0 & 0 \end{bmatrix}$$

- *Low aggregation* with prediction horizon set at 24-hrs.

$$A = \begin{bmatrix} 0.0651 & 0.9349 & 0 & 0 \\ 0.3312 & 1.3331 \cdot 10^{-6} & 0.4185 & 0.2473 \\ 0 & 0.2582 & 0.7315 & 0 \\ 0 & 0.0155 & 0 & 0.9845 \end{bmatrix}$$

$$B = \begin{bmatrix} 2.4925 \cdot 10^{-8} & 0 & 5.2339 \cdot 10^{-4} & 0 & 0 \\ 0.0030 & 2.0226 \cdot 10^{-4} & 0 & 2.2948 \cdot 10^{-4} & 0 \\ 0.0103 & 0 & 0 & 0 & 7.322 \cdot 10^{-11} \\ 3.333 \cdot 10^{-10} & 1.1052 \cdot 10^{-4} & 0 & 0 & 0 \end{bmatrix}$$

- *High aggregation* and prediction horizon set at 6-hrs.

$$A = \begin{bmatrix} 0.8923 & 0.0987 \\ 0.0316 & 0.9684 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.0090 & 1.0975 \cdot 10^{-11} & 8.9531 \cdot 10^{-5} \\ 1.3859 \cdot 10^{-5} & 1.2566 \cdot 10^{-4} & 0 \end{bmatrix}$$

- *High aggregation* and prediction horizon set at 12-hrs.

$$A = \begin{bmatrix} 0.8730 & 0.1176 \\ 0.0319 & 0.9681 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.0094 & 1.6718 \cdot 10^{-11} & 9.9987 \cdot 10^{-5} \\ 2.6170 \cdot 10^{-10} & 9.6913 \cdot 10^{-5} & 0 \end{bmatrix}$$

- *High aggregation* and prediction horizon set at 24-hrs.

$$A = \begin{bmatrix} 0.8609 & 0.1292 \\ 0.0289 & 0.9711 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.0099 & 8.2288 \cdot 10^{-9} & 1.1022 \cdot 10^{-4} \\ 1.3632 \cdot 10^{-8} & 6.6536 \cdot 10^{-5} & 0 \end{bmatrix}$$

Appendix B: Detailed results of Chapter 5

Table B.1: Internal validation metrics for Outdoor Temperature cluster.

Cluster Number	*Silhouette*	*Davies-Bouldin*	*Davies-Bouldin 2*	*S_Dbw*	*Calinksi-Harabasz*	*Xie-Beni*	*Dunn's*
2	**0,670**	**0,750**	**0,750**	0,438	**146,952**	**0,182**	0,555
3	0,563	0,767	0,767	0,280	127,268	0,185	**0,579**
4	0,515	0,870	0,997	0,224	116,797	0,419	0,341
5	0,536	1,029	1,150	0,204	109,004	0,399	0,319
6	0,515	1,001	1,224	0,174	91,260	0,575	0,260
7	0,521	1,120	1,520	0,205	81,184	1,115	0,181
8	0,503	1,051	1,224	**0,137**	96,329	0,839	0,277
9	0,447	1,144	1,434	0,150	89,153	0,933	0,255
10	0,489	1,284	1,584	0,148	83,758	0,880	0,266

Table B.2: Internal validation metrics for Solar Radiation cluster.

Cluster Number	*Silhouette*	*Davies-Bouldin*	*Davies-Bouldin 2*	*S_Dbw*	*Calinksi-Harabasz*	*Xie-Beni*	*Dunn's*
2	**0,759**	**0,604**	**0,604**	0,373	**175,655**	**0,111**	**0,877**
3	0,586	0,913	1,047	0,259	142,119	0,453	0,453
4	0,469	1,029	1,150	0,267	108,322	0,407	0,480
5	0,427	1,041	1,495	0,214	83,796	1,591	0,239
6	0,457	1,380	1,843	0,201	80,461	2,197	0,191
7	0,405	1,450	1,791	**0,190**	78,152	1,926	0,209
8	0,422	1,745	2,228	0,212	70,067	1,841	0,209
9	0,389	2,008	2,765	0,194	61,721	4,233	0,137
10	0,402	2,050	3,195	0,225	60,011	3,893	0,137

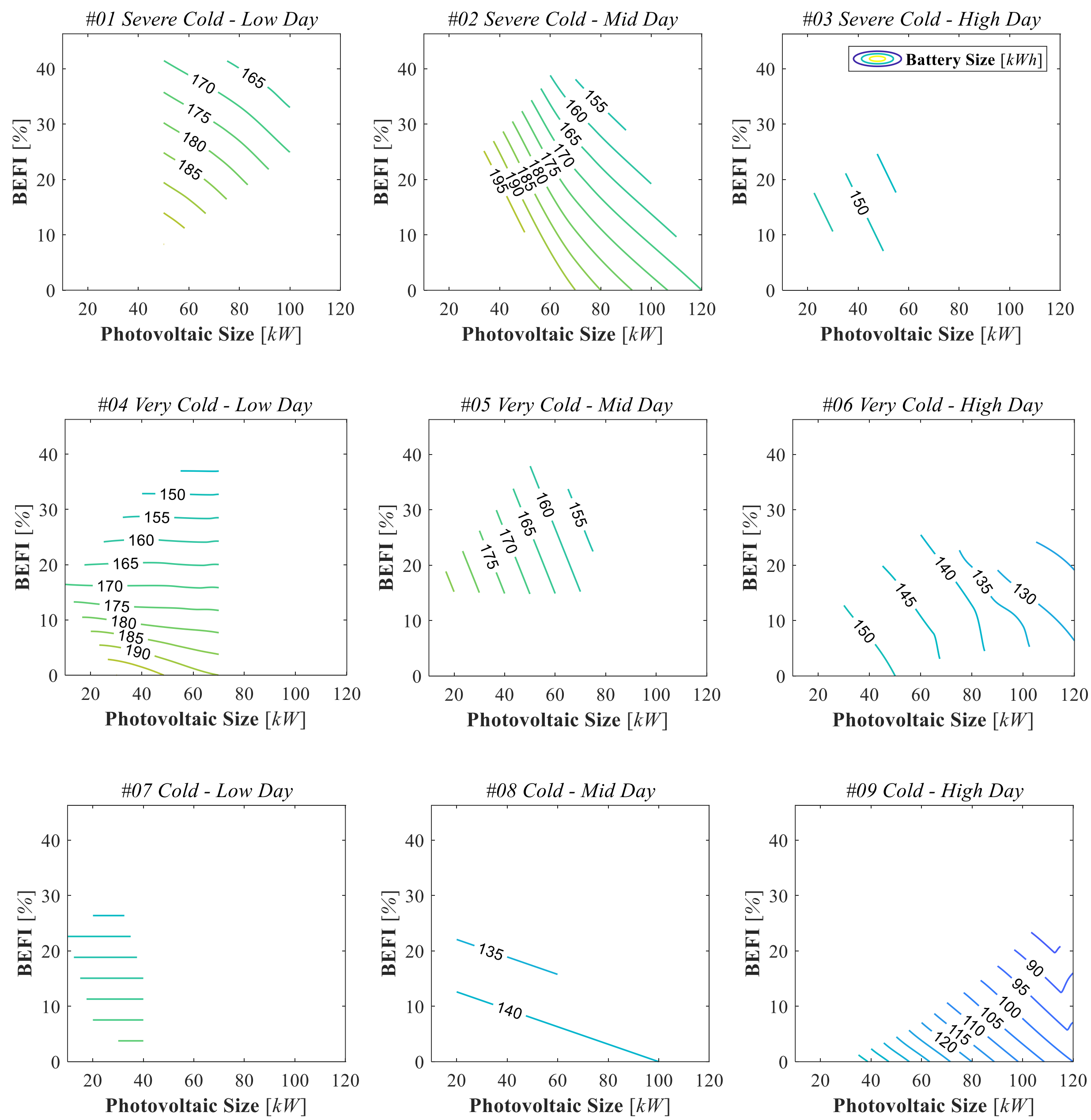


Figure B.1: Optimal PV-battery size combination for each typical day with loss of load probability equal to 0.2.

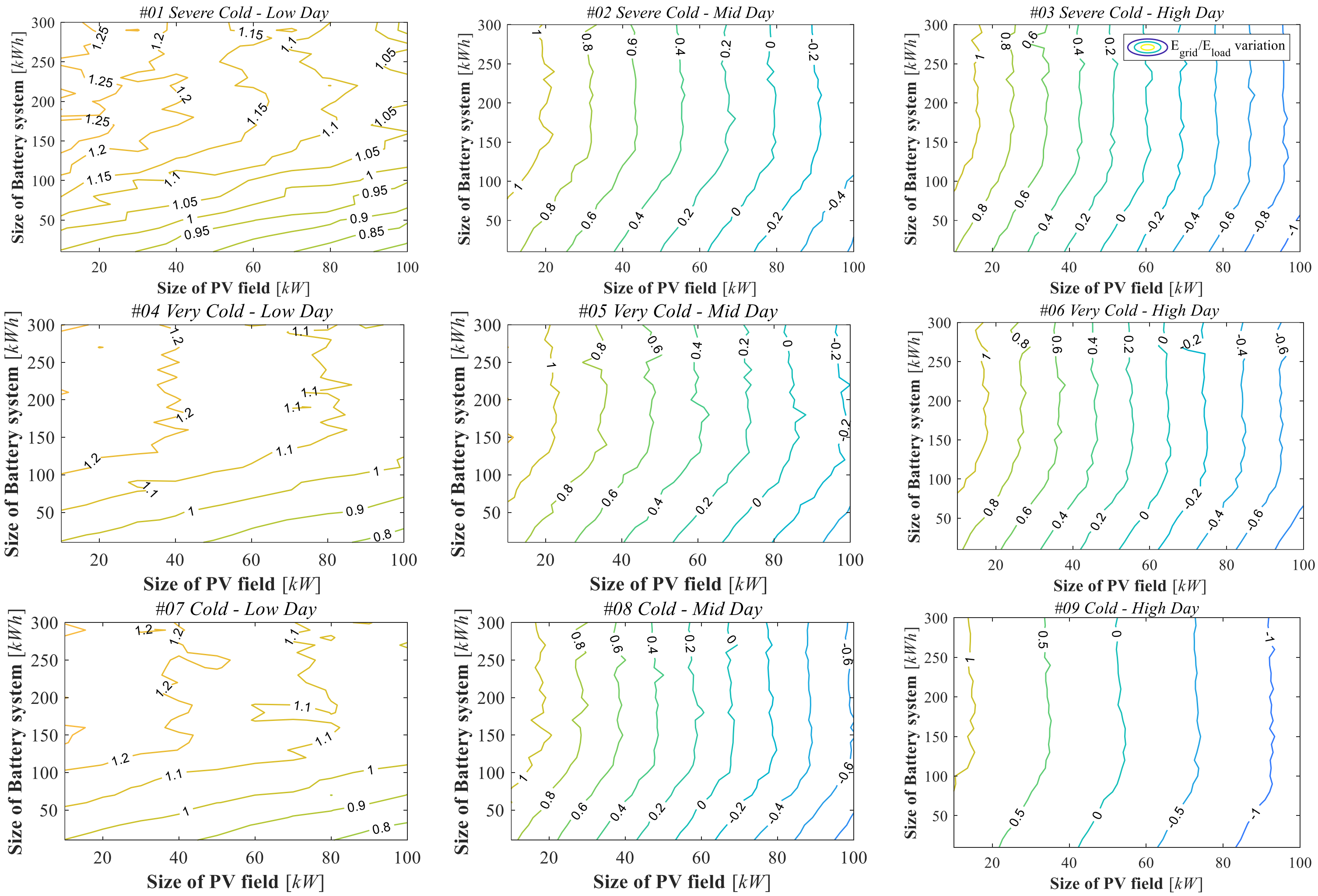


Figure B.2: Comparison between energy demand to the grid and community demand from 9:00 to 16:00 for different values of PV and battery size and with all the buildings participating in demand response.

Appendix C: Supplementary information on case studies dataset

Hydro-Québec pilot project: 30 Houses

The dataset originates from a pilot project conducted by Hydro-Québec, which aimed to evaluate remote thermostat control as a means of managing demand response events. The dataset comprises measurements collected from 30 single-family homes in Québec, all of which use electric baseboard heaters as their primary heating source. These homes were instrumented with advanced monitoring equipment to capture detailed energy usage and thermal behaviour data.

Instrumentation and Data Collection

Smart Thermostats: All conventional thermostats in the homes were replaced with communicating smart thermostats. These devices recorded key variables, including:

- Setpoint Temperature: The desired indoor temperature set by the occupants.
- Measured Temperature: The actual temperature recorded by the thermostat.
- Hourly Energy Consumption: Energy consumed by the heating system on an hourly basis.

Electrical Panel Monitoring: In 10 of the 30 homes equipped with wall-mounted heat pumps, additional monitoring systems (*Egauge*) were installed at the electrical panel. These systems recorded minute-level energy consumption of major electrical circuits, including:

- Heating circuits
- Domestic hot water systems
- Appliances (e.g., stoves, refrigerators)
- Other loads (e.g., ventilation systems, pool pumps)

Whole-House Consumption: For all 30 homes, the total electrical consumption was recorded at 15-minute intervals using data from Hydro-Québec's smart meters.

Dataset Timeframe

The smart thermostats were installed in late 2016, and the shared dataset covers the period from May 1, 2017, to April 30, 2018. It is noteworthy that no demand response events were conducted during this timeframe, ensuring that the data reflects natural occupant behaviour and baseline energy usage patterns.

Data Structure

The dataset includes several files with detailed measurements:

House Description File: Contains metadata about the homes, such as house ID, location (linked to historical weather data), occupant demographics, building characteristics (e.g., number of floors, total area, construction year), installed equipment (e.g., heat pumps, ventilation systems, pools)

Thermostat Data: Provides details on each thermostat's location, power rating, and associated heating circuit.

Smart Meter Data: Logs whole-house energy consumption at 15-minute intervals, with average power consumption recorded in kilowatts (kW).

Egauge Data: Logs minute-level energy consumption for specific circuits in homes with heat pumps.

Thermostat Measurements: Includes variables such as measured and setpoint temperatures (°C) and hourly energy consumption (Wh).

Limitations and Data Gaps

- Data completeness varies, with potential communication losses leading to gaps in the dataset.
- For thermostat data, the recording frequency for temperature variables can be irregular, depending on occupant interactions or system conditions.

Regulvar monitoring: Varennes library

The Varennes Library, located in Varennes, Québec, near Montreal, stands as Canada's first institutional net-zero energy building. Completed in 2015, this two-storey, 24,000-square-foot facility exemplifies sustainable design and energy efficiency.

Situated on the site of the former library, the new structure is one of only about 10 buildings across Canada that produce as much energy as they consume, and the City of Varennes is the first in Québec with a net-zero institutional building on its territory.

The library employs a comprehensive building management system (BMS) – accessible online through the enteliWEB platform of DeltaControls – to monitor and control its mechanical and electrical systems, ensuring optimal performance and energy efficiency. This system allows for real-time data collection and analysis, facilitating proactive maintenance and operational adjustments.

Mechanical Systems

The library's heating and cooling are managed by four ground-source heat pumps connected to eight boreholes. These heat pumps operate in parallel to meet the building's thermal demands, including those of the air handling unit (AHU), fan coils in various areas, and radiant slabs.

The system comprises two single-stage heat pumps prioritized for operation, while two two-stage heat pumps are utilized as needed. Heat pumps and associated pumps remain off when there is no demand for heating or cooling. For instance, during unoccupied summer periods, the system can shut down entirely if conditions permit.

The chilled water setpoint adjusts between 6.7°C (44°F) and 10°C (50°F) based on valve positions, while the heating setpoint ranges from 26.6°C (80°F) to 34°C (93°F). Heat pumps are sequenced to maintain these setpoints according to demand. Upon activation of a heat pump, both its hot and cold side pumps start. In scenarios where the borehole water temperature is below 7.2°C, free cooling is utilized by circulating water without engaging the heat pumps.

During cooling demand with excess heat (network temperature 4°C above setpoint), valves to the boreholes open to dissipate heat, with pumps P-12 and P-13 starting in sequence to provide the required flow. Conversely, during heating demand with insufficient energy on the cold side (network temperature 4°C below setpoint), the same valves and pumps operate to extract energy from the boreholes.

Setpoints adjust gradually, not exceeding a 1°C change per 5 minutes, especially during mode transitions between heating and cooling. If heat pumps cannot meet the demand, auxiliary heating is provided by starting pump P-09 and activating the boiler to maintain the network setpoint.

A differential pressure sensor modulates the bypass valve to maintain the required flow to the boreholes. Heating and cooling pumps are controlled by differential pressure sensors to meet zone demands, shutting down when there is no call for heating or cooling. A three-way valve on the chilled water line modulates to supply water at a minimum temperature of 6.7°C (44°F).

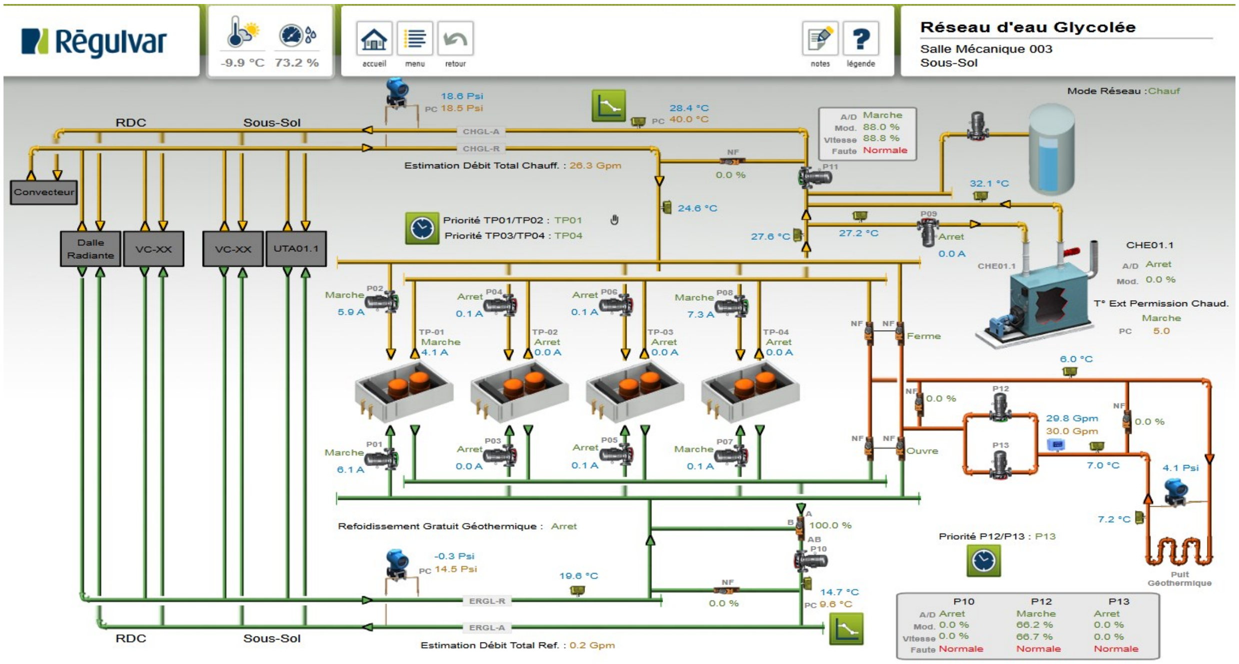


Figure C.1: Glycol loop control graphic interface with the four ground source heat pumps, electric heater, circulating pumps and boreholes.

Air Handling Unit (AHU)

The AHU supplies fresh air to all fan coil units in the building. This air can be preheated using heat recovered from building-integrated photovoltaic thermal (BIPV/T) collectors and a thermal wheel. In winter, a temperature sensor modulates the thermal wheel, BIPV/T air dampers (VE-04), and the heating valve to maintain the supply air setpoint, which adjusts between 13°C and 18°C based on room demand. In summer, the thermal wheel is stopped, VE-04 dampers expel air outside, heating is disabled, and the cooling valve modulates to maintain the setpoint, set 1°C higher than in winter.

The AHU operates according to a programmed occupancy schedule. Upon startup, supply and exhaust fans activate, dampers open, and heating, cooling, and humidification become available. Supply fan speed is modulated to ensure adequate airflow to all terminal units, aiming for the most open terminal box to be at least 90% open. Exhaust airflow is balanced by adjusting the flow rates of exhaust boxes BT-1.07 and BT-2.01 based on the operation of local exhaust fans in areas like restrooms.

Restroom exhaust fans operate based on occupancy detection, continuing for 15 minutes after the space is vacated. Room humidity is maintained at 30% in winter by modulating the humidifier, with a high-limit sensor in the supply air stream set at 85% to prevent condensation. The system shuts down upon freeze detection.

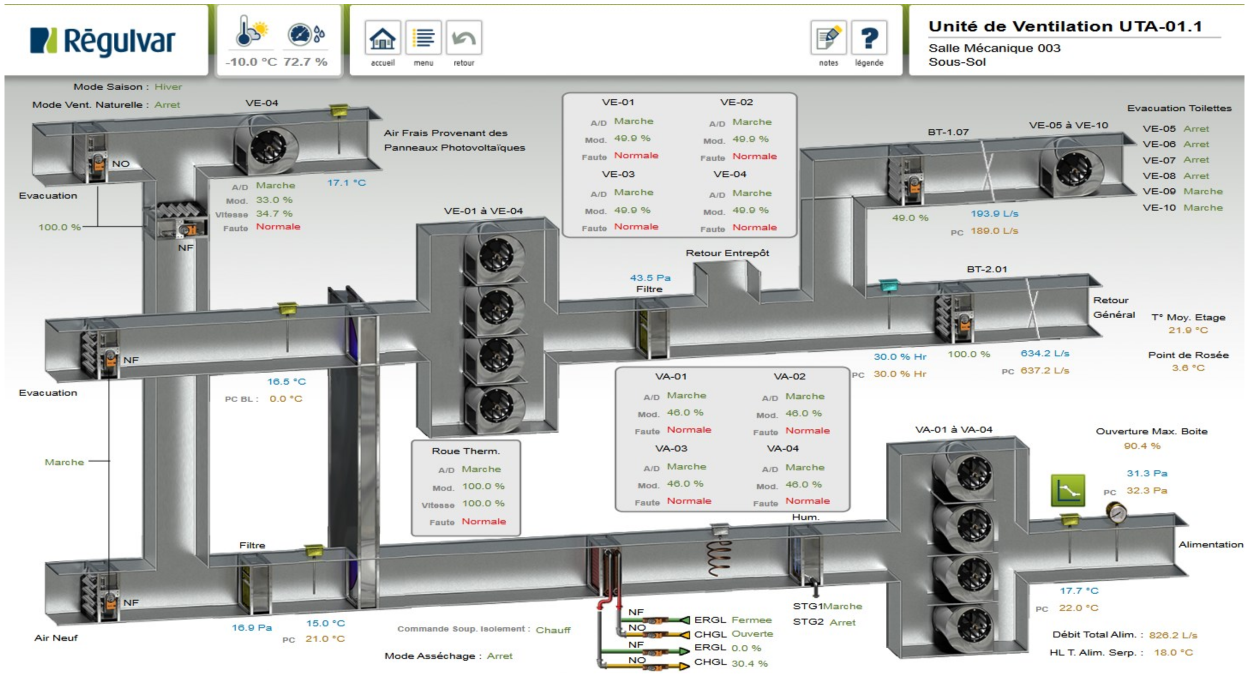


Figure C.2: Air handling unit control graphic interface with thermal wheel, fans and heating/cooling supplier.

Ventilation and Terminal Units

The building utilizes both natural and mechanical ventilation strategies. In cooler seasons, motorized windows can open to facilitate natural ventilation. During this time, the AHU supply fans are stopped, and exhaust is limited to restrooms and storage areas. The AHU supply fans can restart based on room CO_2 levels. Windows open automatically when the outdoor temperature is between 13°C and 22°C and is lower than the average indoor temperature. In unoccupied periods, only upper-level windows may open, and windows close automatically upon detecting rain.

Spaces are conditioned using fan coil units and radiant floor slabs. The radiant slabs are integrated with the building's heating and cooling systems, using glycol circulated through embedded pipes to manage thermal loads efficiently. These slabs provide the primary stage of heating and cooling in occupied spaces.

During occupied periods, fan coil units operate continuously, with room sensors controlling the six-way valves of the radiant slabs to maintain desired temperature setpoints (22°C in winter and 25°C in summer). In unoccupied periods, fan coil units are typically off, but the radiant slabs remain active to maintain setpoints. Fresh air supplied to fan coil units is controlled between minimum and maximum flow rates based on room CO_2 levels, targeting a maximum of 850 ppm. When FCUs are off or during unoccupied periods, fresh air dampers close. Fan speed adjusts between 10% and maximum based on demand.

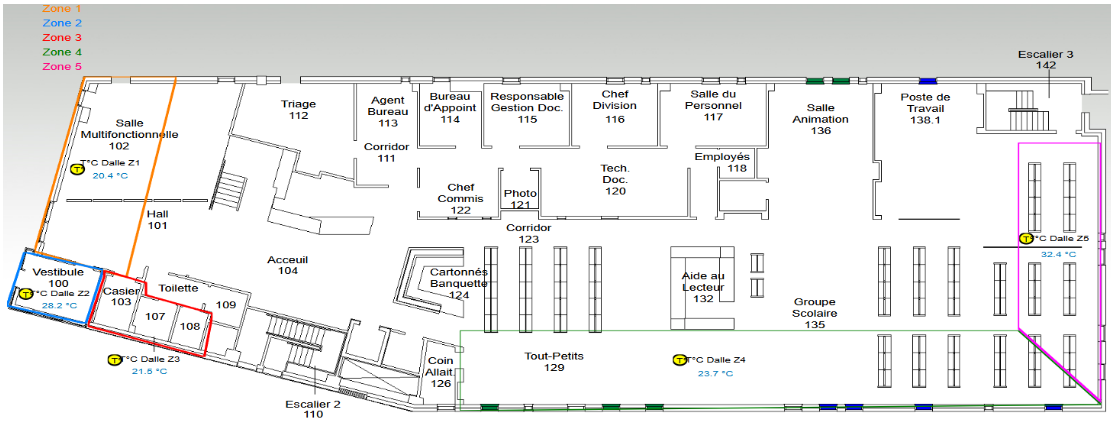


Figure C.3: First floor plan with focus on the area covered by the five radiant slabs.

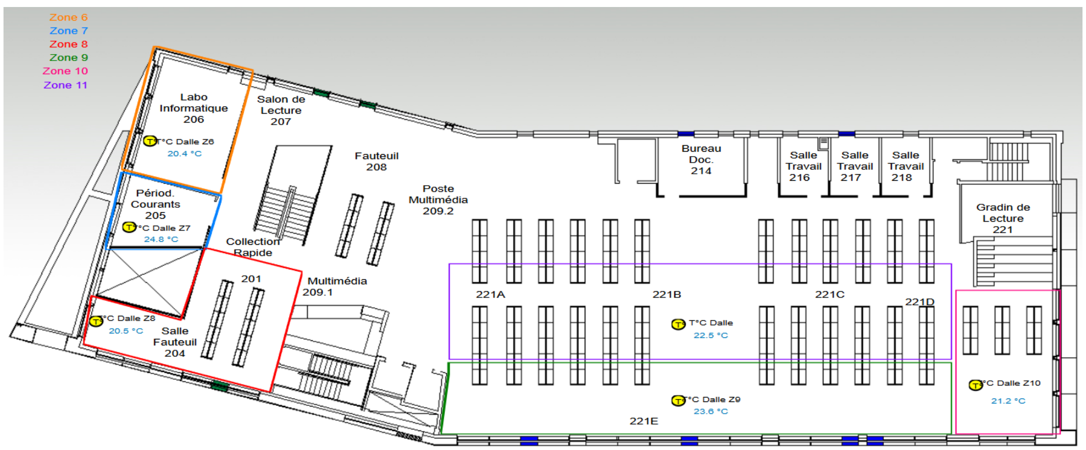


Figure C.4: Second floor plan with focus on the area covered by the six radiant slabs.

Appendix D: Sample of MatLab codes for RC Model Structure Identification, Clustering and MPC strategies

This section provides a selection of source code snippets used in the development and implementation of the studies outlined in this thesis. These examples are meant to illustrate key methodologies and algorithms applied throughout the research, particularly those related to building energy modelling, clustering, and optimal energy management. The full code will be available at the following GitHub repository: https://github.com/anthonymaturo.

RC model structure identification

```
%% Automatic approach to define the RC model structure for the provided data
% The approach is based on a cascade process with two main steps:
% - Spectral analysis: to neglect certain thermal zones from the model.
% - Transfer function estimate: to aggregate thermal zones based on certain
conditions.

% Information on the required inputs:
% - Data of both thermal zones and weather for a period longer than one week with a
%   time step of 15 minutes.
% - Thermal zones data should include three columns: Tset (set point temperature),
%   Tin (indoor temperature)and Eth (heating input).

close all
clear all
clc

global info_tfest

%% DATA DEFINITION
% Load thermostat and weather data for each house, saved with their own name “H68” to
% “H98”
load DATA_HOUSES
par.numberofhouses = 30;
par.START          = 1;
par.STOP           = 672;  % 7 days


%%%% MODEL ORDER REDUCTION LAYER: SINGLE BUILDING
contatore = waitbar(0, 'Simulation single Houses');

for i = 0:par.numberofhouses
    houseDataVar = ['H' num2str(68+i) '_cal'];  % dataset used for order reduction
    if eval(['size(' houseDataVar ', 1) == 0;'])
        houseStruct = struct('data', [], 'Ethmax', [], 'neglectedzones_spa', [],
        'info_tfest', []);
    else
```

```
% Weather data
Tout = eval(['H' num2str(68+i) '_cal{par.START:par.STOP,1};']);
G = eval(['H' num2str(68+i) '_cal{par.START:par.STOP,2};']);

info_tfest = []; Tset = []; Tzone = []; Q = []; Q_SPA = [];

DATA = eval(['H' num2str(68+i) '_cal{par.START:par.STOP,3:end};']);
for j = 1:(size(DATA,2)/3)
    Tset(:,j)  = DATA(:,3*(j-1)+1);
    Tzone(:,j) = DATA(:,3*(j-1)+2);
    Q(:,j)     = DATA(:,3*(j-1)+3);
    eval(['Q_SPA(:,j) = QSPA_H' num2str(68+i) '{par.START:par.STOP,j};']);
end
Ethmax = eval(['Ethmax_H' num2str(68+i) ';']);


%% FIRST STEP: Spectral analysis
% This step receives as input the dataset on the indoor air temperature,
% setpoint and heating input for each thermal zone and provides as output the
% reduced dataset.
par.Ts = 900;                           % Time step (seconds)
par.winSize = 96;                       % Window size
par.N = par.winSize / 2;                % Number of frequencies
par.freq = (0:1:par.N) * 2 * pi / 86400; % Frequency vector

[Tzone, Tset, Q_SPA, Ethmax, neglectedzones_spa] = spa_evolution(Tzone, Tset,
   Q_SPA, Tout, G, Ethmax, par);

sizespa = size(neglectedzones_spa, 1);
if sizespa > 0
    for kkkk = 0:sizespa-1
        Q(:, neglectedzones_spa(end-kkkk)) = [];
    end
end


%% SECOND STEP: Transfer function estimate
% This layer receives as input the reduced dataset and run the aggregation
% routine for the specific building. The output will be the structure and
% dataset required to calibrate the building thermal model
par.nz = 1;   % Number of zeros
par.np = 2;   % Number of poles
par.thresholdtfest = 0;
par.thresholdsetpoint = 1.5; % 1.5 for "Low Aggregation" routine,
                             % 1.0 for "High Aggregation" routine.

par.opt = tfestOptions('InitializeMethod', 'all');
par.opt.EnforceStability = 1;

for iteration = 1:size(Tzone, 2)
    [Tzone, Tset, Q, Ethmax, info] = tfest_evolution_singlebuilding(Tzone,
       Tset, Q, Tout, G, Ethmax, par);
    info_tfest = [info_tfest; info];
    stopcriteria(iteration) = size(Tzone, 2);
    if iteration > 1 && stopcriteria(iteration) == stopcriteria(iteration-1)
```

```
                break;
            end
        end

        % Aggregate results
        DATA = [];
        for k = 1:size(Tzone, 2)
            DATA = [DATA, Tset(:,k), Tzone(:,k), Q(:,k)];
        end

        % Results of each single house
        houseStruct.data = [Tout, G, DATA];
        houseStruct.data_val = createdataval(eval(['H' num2str(68+i) '_val']),
           info_tfest, neglectedzones_spa);
        houseStruct.Ethmax = Ethmax;
        houseStruct.neglectedzones_spa = neglectedzones_spa;
        houseStruct.info_tfest = info_tfest;
    end
    waitbar(i / par.numberofhouses);
end
close(contatore);

% Save results
save('Results_H68to98_lowaggregation.mat');
```

Clustering

```
clear
close all
clc

%%%%%%% First Layer
%% Settings
settings.Sf = timerange('12/01/2019 00:00:00','03/21/2020 00:00:00');

cd inputs
% Weather data loading
load('weather2019_2023.mat');
settings.timef = weather(settings.Sf,:).Date;
settings.Toutf = weather(settings.Sf,:).Tout;
settings.GloSf = weather(settings.Sf,:).GloS; settings.GloSf(settings.GloSf < 0) = 0;
settings.DirSf = weather(settings.Sf,:).DirS; settings.DirSf(settings.GloSf < 0) = 0;
settings.DifSf = weather(settings.Sf,:).DiffS;settings.DifSf(settings.GloSf < 0) = 0;
settings.windf = weather(settings.Sf,:).wind;
clear weather
cd ..

%% Typical Days
% Perform clustering and select the simulation period based on the result
lunghezza1 = 96;        % used for reshaping
lunghezza2 = 7;         % used for reshaping
Tout_24h = reshape(settings.Toutf, lunghezza1, size(settings.Toutf,1)/lunghezza1);
```

```
GloS_24h = reshape(settings.GloSf, lunghezza1, size(settings.GloSf,1)/lunghezza1);
%% Outdoor Temperature: Hierarchical Clustering with Dynamic Time Warping (DTW)
% Normalize the data row-wise (by days)
Tout_24h_normalized = normalize(Tout_24h, 2);

% Calculate pairwise DTW distances
nDays = size(Tout_24h_normalized, 2);
distances = zeros(nDays);
for i = 1:nDays
    for j = 1:nDays
        if i < j
            distances(i, j) = dtw(Tout_24h_normalized(:, i)', Tout_24h_normalized(:,
              j)');
        elseif i > j
            distances(i, j) = distances(j, i);  % Ensure symmetry
        end
    end
end
% Convert the distances matrix to a condensed form
condensedDistances = squareform(distances);

% Perform hierarchical clustering
NumberofClusters = 1:10;
Z = linkage(distances, 'ward');
for i = NumberofClusters
    clust_tout(:, i) = cluster(Z, 'MaxClust', i);
end

% Evaluate the optimal number of clusters based on various metrics
CVIcbi = {'sil','db','db2','sdbw','ch','xb','gd41'};  % Clustering validation indices
addpath([pwd '/adanjoga-cvik-toolbox-2920beb/proximity']);
addpath([pwd '/adanjoga-cvik-toolbox-2920beb/cvi']);
for kk = 1:size(CVIcbi, 2)
    eva_tout = evalcvi(clust_tout, CVIcbi(kk), Tout_24h');
    Table_bestToutK = [Table_bestToutK, table(eva_tout.OptimalK, 'VariableNames',
       CVIcbi(kk))];
    Table_evaTout = [Table_evaTout, table(eva_tout.FitnessValues',
      'VariableNames', CVIcbi(kk))];
end

% Select the optimal number of clusters
% The value selected is the most recurrent if its probability is over 80%
% according to the indicators selected in this study.
T = tabulate(Table_bestToutK.Variables);
sorted_T = sortrows(T, -2);  % Sort by frequency
most_frequent_values = sorted_T(1:2, 1);  % First two rows contain the most frequent
                                          % values
frequencies = sorted_T(1:2, 3);  % First two rows contain the frequencies
if frequencies(1) > 80
    cluster_DTW.Nclusters_Tout = most_frequent_values(1);
else
    cluster_DTW.Nclusters_Tout = max(most_frequent_values);
end

cluster_DTW.idx_Tout(:, 1) = clust_tout(:, cluster_DTW.Nclusters_Tout);
```

```
% Medoids evaluation for Tout and each cluster
for targetCluster = 1:cluster_DTW.Nclusters_Tout
    clusterPoints = Tout_24h(:, cluster_DTW.idx_Tout == targetCluster);
    distances = pdist(clusterPoints');
    distanceMatrix = squareform(distances);
    distanceSums = sum(distanceMatrix, 2);
    [~, medoidIdx] = min(distanceSums);
    medoid_Tout(:, targetCluster) = clusterPoints(:, medoidIdx);
end

% Mean evaluation for Tout and each cluster
for targetCluster = 1:cluster_DTW.Nclusters_Tout
    clusterPoints = Tout_24h(:, cluster_DTW.idx_Tout == targetCluster);
    mean_Tout(:, targetCluster) = mean(clusterPoints, 2);
end
```

MPC strategy

```
function [Housescons] = LauncherHouses(settings)

% Initialize the weather and solar radiation data
for i = 1:settings.numberTypicalDays
    Tout(:,i) = settings.("TypicalDay0" + num2str(i)).Tout;
    GHI(:,i) = settings.("TypicalDay0" + num2str(i)).GloS;
end

% Define house parameters
N          = settings.N;
Ndr        = settings.Ndr;   % Residential buildings in demand response
Nnotdr     = N - Ndr;
price_Dr   = settings.price_DR;
price_notDr  = settings.price_notDR;
price      = [repmat(price_Dr, 1, Ndr), repmat(price_notDr, 1, Nnotdr)];
peakprice = settings.peakprice;

% Load house models and attributes
info1      = load('Houses_RC_Models');
MODELS     = info1.MODELfinal;
attributes = info1.houses;
info2 = load('Houses_DeterministicValues_weekdays');
reference.otherloadWD = info2.reference.otherloadWD;

% Generate the base load profiles
for i = 1:size(reference.otherloadWD, 1)
    reference.profile(:,i) = repmat(reference.otherloadWD{i,1}, size(Tout,1)/96, 1);
end

% Fixed variables
Duration = size(Tout, 1) - 96;
days     = (Duration + 96) / 96;
T_set    = OccupancyHouse([6, 21], 12, 4, 10);       % Set point creation
T_set    = repmat(T_set, days, 1);                   % Celsius
```

```
% Model Predictive Control settings
ph_v = ones(1, N) * 1; ph_v(1, 1:Ndr) = 4 * 12;     % Prediction horizon vector
ch_v = ones(1, N) * 1; ch_v(1, 1:Ndr) = 4 * 2;      % Control horizon vector

% Optimization options for fmincon
optionsfmincon = optimoptions('fmincon', 'UseParallel', true);

% Run simulation for each typical day
for kkk = 1:settings.numberTypicalDays
    TempHistory = [];
    Wheat = [];
    statevariables = [];
    controlvariables = [];

    for i = 1:N
        % Define zones and parameters
        nozones  = attributes(i).nozones;
        ph = ph_v(i);
        ch = ch_v(i);

        % Boundary conditions
        statevariables(i) = size(MODELS{i, 1}.posA, 1);
        controlvariables(i) = nozones;
        A = -eye(controlvariables(i) * ph, controlvariables(i) * ph);
        b = zeros(controlvariables(i) * ph, 1);
        Aeq = [];
        beq = [];
        lbfmincon = -0.001 * ones(controlvariables(i), ph)';
        ubfmincon = (attributes(i).Ethmax' .* ones(controlvariables(i), ph))';

        if i == 1
            [TempHistory(:,1:statevariables(i)), Wheat(:,1:controlvariables(i))] =
                Main_MPC_parfor_Houses(ch, ph, Duration, MODELS{i, 1},
                controlvariables(i), statevariables(i), nozones, A, b, Aeq, beq,
                lbfmincon, ubfmincon, optionsfmincon, peakprice, Tout(:, kkk),
                GHI(:, kkk), price(:, i), T_set);
        else
            [TempHistory(:, 1 + sum(statevariables(1:i-1)):sum(statevariables)), ...
                Wheat(:, 1 + sum(controlvariables(1:i-1)):sum(controlvariables))] =
                Main_MPC_parfor_Houses(ch, ph, Duration, MODELS{i, 1},
                controlvariables(i), statevariables(i), nozones, A, b, Aeq, beq,
                lbfmincon, ubfmincon, optionsfmincon, peakprice, Tout(:, kkk),
                GHI(:, kkk), price(:, i), T_set);
        end
    end
end
end
```