Machine Learning-Informed Optimization for Modeling Flexible Energy Resources

by

Pietro Favaro

A thesis presented for the degree of
Philosophiae Doctor in Engineering Sciences & Technology

pursued in the Electrical Power Engineering Unit of the
University of Mons

Members of the Jury

Christophe Caucheteur, University of Mons	Chairman
Yi Wang, The University of Hong Kong	
Juan Miguel Morales, University of Málaga	
Erik Delarue, Katholieke Universiteit Leuven	
Zacharie De Grève, University of Mons	Secretary
François Vallée, University of Mons	Supervisor
Jean-François Toubeau, University of Mons	Co-supervisor

February 28, 2026

Strive for perfection in everything you do. Take
the best that exists and make it better.
When it does not exist, design it.

Sir Henry Royce

Acknowledgements

My uninterrupted 21-year journey as a student has come to an end. Reaching this milestone makes me reflect upon the chance I have had along the way. This PhD has been a long adventure made of excellence, discoveries, self-growth, and friendships.

Before individuals, I want to thank the Belgian educational system, the institutions I attended, and the many dedicated teachers I have encountered over the years. I grew up in a country where education is an affordable right. I want to reaffirm the importance of maintaining a high quality education accessible to all. At the university level, this goes hand in hand with cutting-edge research.

My very first personal thanks go to my supervisor, François Vallée. François, you saw something in me when I was still a first-year master student and you acted upon it. You placed your trust in me unconditionally from the beginning. Your unwavering support expressed through your kindness, care, and unlimited energy has no equal. By your side, I learned what leadership is all about: trust, empathy, modesty, and freedom. I will never thank you enough for everything you taught me, but I will carry it with me everyday.

The thanks to my co-supervizor, Jean-François Toubeau, would deserve their own chapter, if not thesis. What started as a one-month traineeship evolved in sharing an office for more than a year. This was arguably the most enriching time of my life. Never have I developed so fast in all directions (research, economy, writing, reading, sustainability) as during that time. With you, my research identity feels whole and the challenges we can address are limitless. Thank you for sharing with me your passion for research and academic excellence, you have been a true guide.

To Yury, thank you for welcoming me at Johns Hopkins University during a year. You have never ceased to challenge me, pushing me over my limits. This stay might have been a small step for Research, it was a giant leap for me. I am grateful for your guidance and your time.

I also wish to express my gratitude to all the amazing colleagues I have had over the years and places. Thank you to my PSMR colleagues for the awesome vibes in the lab: rare are the workplaces filled with so much joy. Zhirui, Tomas, Saroj, Boyu, and all the friends from CSEGA, you turned my stay in the U.S. into a wonderful time. To the group of Prof. Yi Wang, especially Xueyuan and Yangzhe, you are a true inspiration. Angel, thank you for transitioning from explaining the interconnection cost to attending a summer school and writing a PyEPO tutorial with me, to becoming a true friend.

I want to thank my parents for teaching me at an early age what excellence is. My father used to say: "When you do something, do it well." This seemingly innocuous maxim has shaped me and helped me reach this stage. Thank you to my mother, always available to lend me an attentive ear. To my mother's side, thank you for cultivating the art of debating, you laid the bases of what an analytical mindset should be. To my father's side, thank you for working hard providing at each generation a better start for the next one.

To Manon, Victor Hugo said "L'amour est une mer dont la femme est la rive." Thank you for being this welcoming land that brings me the calm and peace I so desperately need.

From the bottom of my heart,

Thank you.

Pietro.

Abstract

Decarbonization of power systems through large-scale renewable integration creates unprecedented grid management challenges. Variable wind and solar generation introduce significant uncertainty and volatility, requiring flexible resources to balance supply and demand in real time. Demand-side flexibility from building HVAC systems represent critical flexibility providers, but their optimal operation remains challenging. These assets exhibit complex nonlinear dynamics, such as heat pump efficiency curves and thermal dynamics, that conventional linear approximations inadequately represent. These misrepresentations results in suboptimal scheduling decisions.

A fundamental tension exists in energy system decision-making. Constrained optimization frameworks, particularly convex formulations, provide formal guarantees on solution quality and global optimality through well-established theory and efficient solvers. However, this mathematical rigor requires linearization of system dynamics, sacrificing modeling fidelity. Conversely, machine learning excels at capturing nonlinear system behavior from data but lacks the decision-making structure and solution quality guarantees that optimization provides. This thesis bridges this critical gap by developing a unified framework that harnesses the nonlinear expressiveness of machine learning while preserving the solution quality guarantees and tractability of constrained optimization.

We introduce neural network-constrained optimization, which embeds trained neural networks directly into mixed-integer optimization frameworks as exact piecewise linear reformulations. This approach enables precise modeling of unknown nonlinear thermal dynamics of buildings for day-ahead scheduling. Later, a multi-fidelity analysis demonstrates that incorporating nonlinearities through high-fidelity simulation significantly improves solution quality. However, fully nonlinear formulations render the optimization problem computationally intractable for practical implementation. To bridge this gap, we propose decision-focused learning: a framework that learns asset representations optimized for decision quality rather than statistical accuracy (e.g., mean

squared error minimization) on historical data. We first present a methodology for learning linear constraint parameters within convex optimization programs, directly aligning model learning with the downstream optimization objective. We subsequently extend this framework to non-differentiable programs (e.g., mixed-integer linear programs) using stochastic smoothing and score function gradient estimation (i.e., REINFORCE algorithm), enabling learning in inherently non-differentiable settings. This is particularly valuable for physical systems (or high-fidelity simulators of such systems) that are not differentiable. Crucially, this extended framework allows neural network parameters in neural network-constrained optimization to be learned end-to-end, accounting for true system nonlinearities while maintaining computational tractability.
Case studies on fictitious on multi-zone office buildings achieve sixfold reductions in cost prediction errors. These results validate that machine learning-informed optimization and decision-focused learning enable computationally tractable, and physically consistent decision-making strategies essential for grid flexibility in the renewable energy transition.

Contents

List of Figures

List of Tables

List of Acronyms

AC-OPF Alternative Current Optimal Power Flow.
aFRR automatic Frequency Restoration Reserve.
AHU Air Handling Unit.

CP Combinatorial Programs.

DCP Disciplined Convex Programming.
DFL Decision-Focused Learning.
DPP Disciplined Parametrized Programming.
DSO Distribution System Operator.

EV Electric Vehicle.

FB Fixed Binaries.
FCR Frequency Containment Reserve.

HVAC Heating, Ventilation, and Air-Conditioning.

ITO Identify-Then-Optimize.

KKT Karush-Kuhn-Tucker.

LP Linear Programs.
LSTM Long Short-Term Memory.

MAE Mean Absolute Error.
mFRR manual Frequency Restoration Reserve.
MIL Mixed-Integer Linear.
MILP Mixed-Integer Linear Program.

MIP	Mixed-Integer Program.
MIQP	Mixed-Integer Quadratic Program.
ML	Machine Learning.
MPC	Model Predictive Control.
MSE	Mean Squared Error.
NLL	Negative Log-Likelihood.
NN	Neural Network.
PHES	Pumped Hydro Energy Storage.
PV	Photovoltaic.
QP	Quadratic Program.
RC	Resistance-Capacitance.
ReLU	Rectified Linear Unit.
RES	Renewable Energy Source.
RL	Reinforcement Learning.
RMSE	Root Mean Squared Error.
SGD	Stochastic Gradient Descent.
SNR	Signal-to-Noise Ratio.
SPO+	Smart "Predict, then Optimize" Plus.
SS	Stochastic Smoothing.
SS-DFL	Stochastic Smoothing Decision-Focused Learning.
TSO	Transmission System Operator.

CHAPTER 1.

Introduction

1.1. Background and Motivation

The transition from fossil fuel-based to renewable energy-based electricity generation introduces fundamental challenges for power system operation and decision-making. On the supply side, the growing deployment of weather-dependent renewable energy sources, such as wind and solar, introduces substantial variability and uncertainty in electricity generation. On the demand side, sector electrification (heating, transportation, industry) and the adoption of flexible assets (heat pumps, electric vehicles, batteries) create new opportunities for load-side flexibility that remain largely untapped.

Instead of building new generation and transmission infrastructure to follow demand, modern power systems increasingly rely on demand-side flexibility (i.e., the ability to shift, shed, or modulate electricity consumption in response to grid conditions, market prices, and renewable availability). Power systems' shift from a supply-following to a demand-responsive paradigm has profound implications for building management, energy markets, and grid operations. From an application perspective, this thesis focuses on building Heating, Ventilation, and Air-Conditioning (HVAC)[1]. HVAC systems represent one of the largest untapped flexibility resources, accounting for 30% of the final electricity consumption in the USA [5], [6]. Yet this flexibility remains significantly underutilized because current control methods are fundamentally ill-suited to

[1]Note that Favaro and his co-authors also did major developments on Pumped Hydro Energy Storage (PHES) [1], [2], [3]. PHES is the largest utility-scale storage technology representing over 90% of the installed energy capacity worldwide in 2022 [4]. It stores energy by pumping water in altitude. Electricity is produced by releasing the water, which subsequently drives a turbine.

the task.

To maintain grid stability amid this transition, system operators and market participants must make operational decisions (e.g., HVAC scheduling in buildings, and energy arbitrage in PHES) based on imperfect models of system dynamics and uncertain market conditions. Historically, decision-making has followed a classical two-stage paradigm, named Identify-Then-Optimize (ITO): (i) *identify* a model of system dynamics from historical data, then (ii) *optimize* control decisions using that fixed model. The conventional modeling approach has severe limitations:

1. **Model-reality mismatch**: ITO treats system identification and control as decoupled tasks. A model trained to minimize prediction error (e.g., Mean Squared Error (MSE)) may be poor at the task it will ultimately support: decision-making. This is especially problematic when: (i) system behavior changes under control leading to a distribution shift (e.g., when HVAC scheduling modifies building thermal patterns from historical conditions); (ii) the relationship between the model error and the decision quality is asymmetric, non-linear, or non-monotonic [7].

2. **Incomplete physics representation**: Many systems have known constraints (thermal bounds, power limits, feasibility regions) that should be enforced structurally during learning of unknown system dynamics (building thermal dynamics, water-power conversion curve in hydropower plants). However, enforcing constraints strictly during learning is challenging and soft approaches such as penalty term in the loss function are preferred [8].

3. **Computational and structural barriers**: Real-world systems are often nonlinear or have discrete constraints (e.g., HVAC binary switching: heating or cooling, pumped hydro energy storage mode: pump or turbine). Embedding nonlinear and discrete models into optimization problems creates computational bottlenecks and renders the problem nonconvex [9].

This thesis addresses these interconnected challenges by developing Neural Network (NN)-constrained optimization and Decision-Focused Learning (DFL) frameworks that unify physics-based constraints with data-driven learning in a single end-to-end trainable pipeline. Rather than optimizing for prediction accuracy, these frameworks optimize directly for downstream decision quality, i.e., the decision-maker's ultimate goal.

1.2. Research Questions

The overarching research direction is motivated by four specific challenges in grid-interactive resource control and demand response:

Research Question I: Can we identify system dynamics and optimize control simultaneously, rather than sequentially?

In day-ahead building thermal scheduling (e.g., HVAC), the conventional ITO approach (i) learns a model of thermal dynamics by minimizing the prediction error on historical data, then (ii) solves an optimization problem using that fixed model to determine optimal setpoints [10]. This separation between model identification and control creates a fundamental mismatch between the training objective (prediction accuracy) and the decision objective (control cost minimization) [11].

The core problem is regime mismatch: when the HVAC system operates under historical passive control rules (e.g., simple setpoint temperature bands), it explores only a narrow region of the operating envelope. A thermal model trained on such data learns to be accurate in that narrow regime. However, when an optimization algorithm computes new setpoints to exploit flexibility (e.g., aggressive pre-cooling during low-price hours, or active load-shedding), the building operates in regimes not represented in historical data. Model parameters that were accurate for passive control may severely misestimate the actual cost consequences of control decisions in these new regimes, leading to suboptimal or even destabilizing control actions.

This raises the fundamental question: can we learn model parameters that are *task-aware*, i.e., optimized not to maximize prediction accuracy on passive historical data, but specifically to minimize the *downstream cost of control decisions*? Instead of separating identification and optimization, can we merge them into an end-to-end pipeline where model learning is driven directly by the quality of the resulting control decisions?

Research Question II: How to merge the expressiveness and versatility of data-driven neural networks with the optimality guarantees and constraint-satisfaction of mathematical optimization?

The thermal dynamics of buildings is highly nonlinear [12]. They depend on multiple coupled physical phenomena: heat transfer through building envelopes (temperature-dependent), HVAC system efficiency (load-dependent), internal gains from occupancy and equipment (stochastic), and solar irradiation (weather-dependent). Linear models, such as Resistance-Capacitance (RC) networks, provide interpretability and computational tractability but often fail to capture these nonlinearities accurately, especially across the wide range of operating conditions encountered in demand response scenarios [13], [14], [15].

Conversely, NNs offer exceptional representational power: they can learn complex input-output mappings from data without explicit physical equations. Moreover, NNs with a single hidden layer can arbitrarily well approximate any function given sufficient neurons[16]. However, embedding nonlinear NN models directly into optimization problems creates a severe computational barrier. The resulting optimization problem becomes non-convex and intractable, lacking the structure needed for efficient solution.

The core challenge is a mismatch of capabilities: NNs excel at learning from data but lack structure for optimization while mathematical optimization excels at finding good solutions but requires structured model classes (linear, convex, or mixed-integer linear). Therefore, this research question asks: can we leverage the piecewise-linear structure of Rectified Linear Unit (ReLU) NNs to reformulate them as exact mixed-integer linear constraints, enabling them to be embedded directly into optimization problems? If successful, this approach would combine the modeling versatility of NNs with the optimality and feasibility guarantees of mixed-integer programming.

Research Question III: How can gradients flow through non-differentiable assets/simulators and optimization problems to enable end-to-end learning?

In practice, evaluating the true performance of a building control decision requires more than solving an optimization problem. Building control decisions (temperature setpoints) are ultimately executed in the real building (or high-fidelity simulator), where the actual thermal response depends on true physics, occupancy patterns, and other factors not perfectly captured by any learned model. High-fidelity building simulation tools like EnergyPlus [17] or TRNSYS [18] are the gold standard for evaluating control performance, but they are non-differentiable: they consist of numerical solvers, iterative algorithms, and piecewise logic that produce no meaningful gradients.

More fundamentally, as thermal models become more sophisticated (e.g., embedding NNs in optimization, formulating stochastic problems with discrete decisions) the entire pipeline (model $\rightarrow$ optimization $\rightarrow$ simulator) becomes a composition of non-differentiable components. The optimization problem contains discrete variables (mode switches for heating/cooling), producing piecewise-constant solutions [19]. The simulator contains black-box physics engines.

At the same time, DFL explicitly aims to tune model parameters based on the quality of the final decisions, as measured by ex-post outcomes such as true energy cost and comfort violations. This creates a tension: DFL conceptually requires gradients to flow all the way from the realized outcome back to the model. Standard backpropagation assumes a smooth, differentiable objective landscape for DFL. However, the end of the pipeline (mixed-integer optimization plus simulator) is fundamentally non-differentiable. When that assumption of differentiability breaks down, gradient-based training becomes impossible. But we need gradient information to enable DFL.

This research question asks: can we extract meaningful gradient signals from the non-differentiable pipeline without explicitly computing derivatives? Can we estimate task-specific gradients that capture how changes in model parameters affect downstream decision cost, even when the loss landscape is piecewise constant or undefined? Specifically, can score-function gradient estimators (such as REINFORCE [20] from reinforcement learning) provide unbiased gradient estimates by leveraging the underlying probability distributions, enabling us to train through the full pipeline of non-differentiable simulators, discrete optimization, and complex NN constraints?

Research Question IV: How can we learn accurate thermal models that adapt to changing operating conditions while handling stochastic uncertainty and remaining tractable for optimization?

Building thermal dynamics are inherently nonlinear and time-varying. They depend on weather (ambient temperature, solar irradiation), occupancy (internal heat gains), and past decisions (thermal history). A single, static thermal model learned from data is fundamentally unable to remain accurate across this diversity [15]. Yet thermal models must be embedded in optimization problems, which requires tractability. The linear RC formulation provides computational tractability for optimization but may be too simplistic to capture the nonlinear dynamics accurately [21].

Consequently, we wonder how to develop hybrid models that are both expressive and tractable? Specifically, can we learn context-dependent, probabilistic RC models that adapt their parameters dynamically based on current conditions (weather, time of day, expected occupancy), thereby locally linearizing the nonlinear system behavior around the expected operating regime? By making RC parameters functions of context, we transform a global nonlinear problem into a sequence of local linear approximations, each tailored to the current circumstances.

Furthermore, how can we explicitly represent uncertainty in these models? Building thermal dynamics are fundamentally stochastic, e.g., we cannot predict occupancy or equipment use precisely. Can we develop probabilistic thermal models that captures both epistemic uncertainty (uncertainty due to incomplete knowledge or model simplifications, which can be reduced with more data) and aleatoric uncertainty (inherent randomness in the system due to unpredictable factors like occupant behavior)?

Finally, how can such probabilistic, context-dependent models be integrated into the downstream optimization? Rather than assuming a single deterministic HVAC power trajectory, can we use scenario-based robust optimization where control decisions are optimized to remain feasible and near-optimal across many possible realizations of thermal dynamics, sampled from the learned probability distributions? This would yield control strategies that are robust to uncertainty, maintaining occupant comfort even when actual thermal dynamics deviate from nominal forecasts.

1.3. Research Contributions

This thesis makes three major sets of contributions to DFL for grid-interactive building control, each corresponding to a distinct publication or major manuscript. These contributions systematically address the four research questions identified earlier, advancing the field from foundational decision-aware system identification to sophisticated stochastic modeling under uncertainty.

1.3.1. Contribution 1: Decision-Focused System Identification and Control

The first major contribution, presented in chapter 3, develops a foundational DFL framework for *simultaneous system identification and control optimization*. We identify building thermal dynamics while scheduling the HVAC system. This work directly addresses Research Question I by demonstrating that the classical ITO paradigm fundamentally fails when control decisions shift the building operating regime away from the conditions represented in historical training data.

Traditional approaches first identify a thermal model by minimizing prediction error (e.g., MSE) on passive operational data, where the HVAC system follows simple rule-based setpoints, then use this fixed model to solve an optimization problem for demand response. The critical insight of this contribution is that such models suffer from severe regime mismatch: they excel at predicting behavior under passive control but systematically fail when optimization computes aggressive demand-response maneuvers like pre-cooling during low-price periods or load shifting to align with renewable generation. The learned model parameters, optimized for statistical accuracy across historical regimes, misestimate the true cost resulting of these new control actions, leading to suboptimal or even infeasible schedules.

To overcome this, the contribution formalizes *decision-focused system identification*, where model parameters are learned to directly minimize ex-post control performance rather than prediction error. A linear lumped RC model represents the discretized thermal dynamics of the building. The framework integrates the RC model in a convex day-ahead scheduling of HVAC systems. Gradients can flow from the downstream decision quality (evaluated by a simulator) back through the optimization problem to the model parameters themselves, i.e., the Rs and Cs. This creates a true end-to-end learning pipeline where the model learns to be accurate specifically in the operating regimes that matter for control.

A key innovation is the *supervised decision-focused loss*, which evaluates model quality not through prediction metrics but through the actual performance of the resulting control decisions when executed in a high-fidelity simulator. The supervised DFL loss quantifies the sensitivity of the objective value to prediction errors by comparing the expected and ex-post HVAC power consumptions under the same control (i.e., indoor temperature setpoints) commands. The resulting loss is weighted by the electricity price at each time step to reflect the economic impact of forecasting inaccuracies. This learning setup requires no differentiability of the building or control simulator. The ex-post trajectories obtained from EnergyPlus provide the supervision signal. The approach is thus directly applicable to complex, black-box building physics.

The empirical validation on a realistic 15-zone office building in Denver, CO provided by the US Department of Energy [22] demonstrates that decision-focused identification reduces the absolute error between expected and ex-post power cost by a factor 6. Moreover, DFL achieves 1% lower operational costs compared to ITO baselines, with dramatically improved robustness to distribution shift (extreme weather conditions not seen during training). Critically, the learned models maintain statistical prediction accuracy comparable to traditional methods but concentrate their representational capacity on control-relevant dynamics, achieving superior generalization to new operating regimes.

1.3.2. Contribution 2: Neural Network-Constrained Optimization with Stochastic Smoothing

The second major contribution tackles Research Questions II and III by developing a NN-constrained mixed-integer optimization framework for day-ahead HVAC management that merges the expressive power of Machine Learning

(ML) with the optimality guarantees of mathematical programming. We detail the method and its results in chapter 4.

Building thermal dynamics exhibit profound nonlinearities that linear RC models cannot capture: temperature-dependent heat transfer coefficients, load-dependent HVAC efficiencies, coupled inter-zonal effects, and complex interactions between weather, occupancy, and control history. Even though NNs excel at learning these nonlinear mappings from data, embedding them directly into optimization creates intractable non-convex problems. A first approach approximates the network (sacrificing modeling fidelity) using convex hull relaxation [23]. A second approach looks at reformulating NNs as a set of constraints. In particular, NNs using piecewise linear activation functions such as ReLU can be exactly reformulated as a set of mixed-integer linear constraints [24], [25]. Such formulations were initially developed to verify trained NNs [26], for example against adversarial samples [24]. The first application to model unknown constraints in an optimization problem is in [27], where Murzakhanov et al. created a classifier embedded in an Alternative Current Optimal Power Flow (AC-OPF) to encode security constraints.

This contribution extends the use of NNs as regressors to model unknown constraints: it shows that feedforward ReLU NNs (whose piecewise-linear activation functions define convex polyhedral regions) can model building thermal dynamics within a larger HVAC scheduling optimization. Each ReLU neuron becomes a set of four mixed-integer constraints with a binary activation indicator, tight linear relaxations, and carefully bounded Big-M parameters. A critical practical innovation is adaptive Big-M tightening for the NN reformulation. Rather than using crude fixed bounds on neuron activations, the method dynamically recomputes tight bounds by exploiting known deterministic inputs (weather forecasts, time of day) and reducing their feasible intervals to singletons. This produces dramatically tighter Mixed-Integer Linear Program (MILP) relaxations (proven theoretically via interval analysis) without sacrificing exactness, enabling the framework to scale to deeper networks and realistic problem sizes. The resulting formulation embeds the full expressive power of the neural thermal model directly as optimization constraints while preserving global optimality guarantees through the MILP solver's optimality gap.

However, this exact reformulation creates a new challenge for DFL: Mixed-Integer Program (MIP) produces piecewise-constant solutions with respect to input parameters. Small changes in NN weights can cause discrete decisions to flip abruptly, yielding zero or undefined gradients that make standard back-propagation inapplicable. To overcome this, we leverage Stochastic Smoothing Decision-Focused Learning (SS-DFL), a novel training paradigm that treats

NN parameters as random variables drawn from a Gaussian smoothing distribution. The HVAC problem is solved repeatedly as a Mixed-Integer Quadratic Program (MIQP) under these random parameter realizations, and gradients are estimated using score-function (like the REINFORCE algorithm) estimators that relate parameter perturbations to downstream decision loss without ever differentiating through the solver itself.

Validation on a detailed 5-zone office building demonstrates that SS-DFL with NN-constrained optimization achieves major thermal comfort improvement (thermal comfort penalty divided by a factor 5) over convex DFL relaxations that sacrifice modeling fidelity while keeping HVAC cost roughly similar. The approach maintains computational tractability (solving 24-hour Mixed-Integer Quadratic Program in under 30 seconds) while delivering the solution quality guarantees of mixed-integer programming.

1.3.3. Contribution 3: Probabilistic Decision-Focused Learning for Multi-Zone Buildings Under Uncertainty

The third and culminating contribution, detailed in chapter 5, addresses Research Questions I, III, and IV through a *probabilistic DFL pipeline* for day-ahead thermal scheduling in multi-zone commercial buildings with fundamentally uncertain dynamics. This work, currently under preparation, represents the most complete realization of the thesis vision: end-to-end learning of context-dependent, probabilistic thermal models that drive robust optimization under real-world uncertainty.

Real buildings operate under profound stochasticity: fluctuating occupancy generates unpredictable internal heat gains, weather forecasts contain errors, occupants open windows and override controls unpredictably, and building material degrade over time. Static, deterministic thermal models, even sophisticated ones as NNs, cannot capture such evolving conditions. On the other hand, tractable optimization requires structure. The contribution resolves this through a *context-dependent probabilistic RC model* that parameterizes thermal capacitances, inter-zonal conductances, and HVAC efficiencies as time-varying functions of operating context (weather, calendar data, recent thermal history), learned by an LSTM encoder-decoder architecture. Rather than predicting point temperatures, the model outputs full predictive distributions.

This probabilistic model serves a dual purpose that elegantly unifies learning and optimization. First, the stochasticity enables score-function gradient estimation: by expressing decision loss as an expectation over sampled thermal trajectories, REINFORCE-style gradients propagate cleanly from ex-post outcomes back through the non-differentiable pipeline. Second, the sampled trajectories become scenarios in a scenario-based MIP, where a single shared temperature setpoint schedule is optimized to remain feasible and near-optimal across all plausible thermal realizations. This produces day-ahead schedules that robustly maintain comfort even when actual dynamics deviate from nominal forecasts.

The training procedure combines practicality with theoretical soundness. A physics-informed pre-training phase uses negative log-likelihood loss with ReLU positivity constraints and temporal regularization on RC parameters, providing a warm-start that encodes physical plausibility. The decision-focused fine-tuning phase then uses Monte Carlo scenario sampling and score-function gradients to optimize directly for realized EnergyPlus performance (true costs plus comfort violations), requiring no differentiability of either the MIP solver or building simulator.

The case study is performed on a detailed 3-floor, 15-zone medium office building in Denver (same as Contribution I). ITO warm-starts perform adequately across most models except for the best static RC model on historical forecast which completely collapses. All models underestimate the power cost of their decisions. In contrast, deterministic DFL tuning yields marginal gains (notably for dynamic models), while probabilistic DFL tuning with more scenarios progressively improves the loss at increased computational cost.

1.3.4. Synthesis Across Contributions

Together, these three contributions form a complete progression from decision-aware identification (Contribution 1) to neural-constrained optimization under non-differentiability (Contribution 2) to probabilistic, context-dependent modeling for real-world deployment (Contribution 3). They demonstrate that DFL can transform building HVAC systems from passive consumers to active grid resources, achieving dramatic cost savings while maintaining occupant comfort and computational tractability. The methods generalize far beyond buildings to any domain requiring learning-to-model under uncertainty, non-differentiability, and discrete decisions.

1.4. Thesis Organization

This thesis is organized as follows:

Chapter 2: Energy Transition and the Need for Flexible Energy Resources — provides essential context on the energy transition, the need for flexibility, and why buildings are critical. It establishes the motivation for developing better control methods for HVAC systems using machine learning.

Chapter 3: Decision-Focused Learning for Convex Control and Modeling of Building Thermal Dynamics — (Paper 1 – ACM E-Energy [28]) develops DFL for simultaneous system identification and control. It presents the general problem formulation, gradient computation via conic program differentiation, the supervised loss for non-differentiable systems, constraint relaxation for training robustness, and a case study on RC building model identification for a realistic 15-zone building (Denver, CO). The key contribution is demonstrating that end-to-end DFL outperforms ITO even with limited data, because it focuses learning on control-relevant parameters.

Chapter 4: Decision-Focused Learning for Non-Convex Control and Nonlinear Modeling of Building Thermal Dynamics — (Paper 2 – IEEE Transactions on Smart Grid [29]) develops DFL for simultaneous system identification and optimization, where NN thermal models are embedded as exact mixed-integer linear constraints. It presents the problem formulation, Stochastic Smoothing (SS) for gradient computation, the adaptive Big-M tightening technique, and a case study on a realistic 5-zone office building (Denver, CO). The key contribution is demonstrating that NN-constrained optimization, trained via DFL with SS, outperforms ITO and relaxed DFL methods while maintaining computational tractability.

Chapter 5: Probabilistic Decision-Focused Learning for Robust Control of Building HVAC Systems — (Paper 3 – Ongoing Work [30]) extends DFL to handle context-dependent, probabilistic thermal models and scenario-based robust optimization. It develops the architecture for context-aware forecasting, the integration of probabilistic models into scenario-based HVAC optimization, and the use of score-function gradient estimation to enable training through non-differentiable simulators and discrete optimization. A case study on a multi-zone office building demonstrates the approach and shows that stochastic DFL achieves 10–30% cost savings compared to ITO and is robust to weather extremes.

Chapter 6: Conclusion and Perspectives — synthesizes the contributions, discusses practical deployment considerations, identifies limitations of the current work, and outlines promising future research directions including: integration with electricity markets and demand response, multi-building aggregation and coordination, incorporation of occupant behavior models, and hybrid learning-optimization architectures combining DFL with reinforcement learning.

1.5. List of Publications

The following publications form the core of this thesis:

Paper 1 (Published): P. Favaro et al., "Decision-Focused Learning for Complex System Identification: HVAC Management System Application," in *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*, ser. E-Energy '25, New York, NY, USA: Association for Computing Machinery, Jun. 16, 2025, pp. 347–358, ISBN: 979-8-4007-1125-1. DOI: `10.1145/3679240.3734584`. Accessed: Jul. 15, 2025. [Online]. Available: `https://dl.acm.org/doi/10.1145/3679240.3734584` [12 pages, peer-reviewed]

Paper 2 (Published): P. Favaro et al., "Decision-Focused Learning for Neural Network-Constrained HVAC Scheduling," *IEEE Transactions on Smart Grid*, pp. 1–1, 2025, ISSN: 1949-3061. DOI: `10.1109/TSG.2025.3639887`. Accessed: Dec. 11, 2025. [Online]. Available: `https://ieeexplore.ieee.org/document/11275968` [13 pages, peer-reviewed]

Paper 3 (Ongoing): P. Favaro and J.-F. Toubeau, "Stochastic Decision-Focused Learning for Thermal Energy Scheduling in Multi-Zone Buildings," Feb. 2026. Accessed: Feb. 24, 2026. [Online]. Available: `https://hal.science/hal-05525850` [18 pages, working paper]

CHAPTER 2.

Energy Transition and the Need for Flexible Energy Resources

2.1. Introduction

The transition towards a renewable-based energy system with net-zero carbon emissions by 2050 is a central pillar of energy policy worldwide [31], [32]. This transition introduces a fundamental challenge: weather-dependent renewable energy sources such as wind and solar produce electricity intermittently and often unpredictably. Unlike conventional thermal power plants that can adjust output to follow demand, renewable generators are constrained by meteorological conditions. This weather-dependency creates periods of both scarcity (when demand exceeds supply) and abundance (when renewable generation exceeds demand). To maintain grid stability and reliability amid growing renewable penetration, a critical paradigm shift is underway. Modern power systems increasingly rely on demand-side flexibility.

Definition: Flexibility

Flexibility is the ability to change electricity generation, consumption, or storage in response to signals from the grid operator or market prices.

This shift from a supply-following to a demand-responsive paradigm has profound implications for building management, energy markets, and grid operations.

This chapter establishes the critical context for the thesis by exploring the role of demand-response in integrating renewable energy and maintaining grid stability, the primary flexible assets and demand response mechanisms in modern power systems, why building HVAC systems represent a critical untapped flexibility resource, how a Distribution System Operator (DSO) procures flexibility from buildings to manage local constraints, and the control challenges that DFL addresses. The chapter shows that buildings, particularly their HVAC systems, represent a significant flexibility potential that can contribute towards the decarbonization of the energy sector. However, current control methods are not suited to unlock this potential. The thesis contributions directly address this gap.

2.2. The Renewable Energy Transition

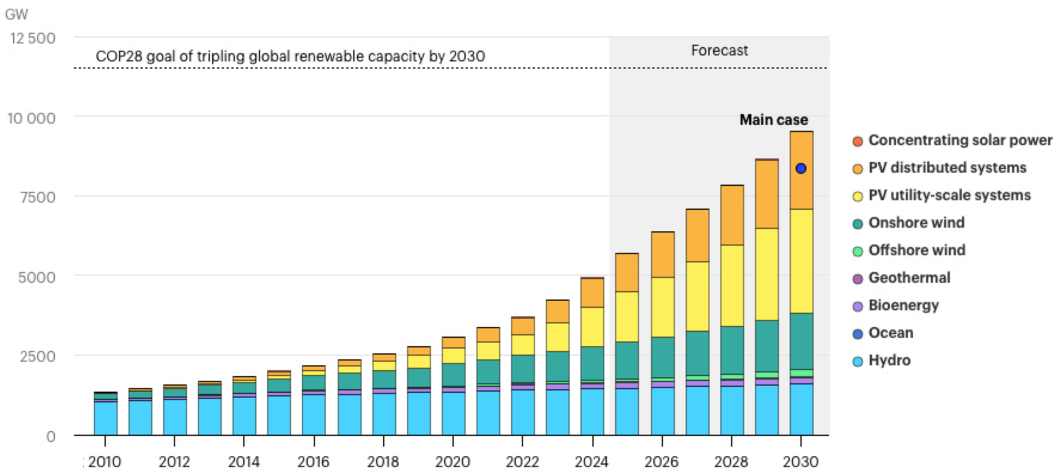


Figure 2.1.: Renewable electricity capacity worldwide from 2010 to 2024 and forecast of most probable scenario till 2030 [33].

Global renewable electricity capacity is expanding at an unprecedented pace, see Figure 2.1. A trend expected to continue following the agreement of COP28 (2023) to *"triple the world's installed renewable energy generation capacity to at least 11,000 GW by 2030"* [34]. The International Energy Agency reports that in 2023, Renewable Energy Source (RES) additions surpassed 50% year-on-year growth, with solar Photovoltaic (PV) and wind power accounting for over 90% of new generation capacity connected to grids globally [35]. By 2025, over 2.6 terawatts of new clean energy projects were awaiting grid connection in the United States alone, more than twice the current grid capacity [36]. However, this rapid expansion creates new operational challenges that were absent in conventional power systems dominated by dispatchable thermal generation.

2.2.1. Energy Transition Challenges

Three main challenges posed by the ramp-up of RES can be identified. First, the generation is variable and uncertain. Second, the integration of RES in the grid leads to increased curtailment (involuntary rejection of renewable generation due to insufficient demand or grid capacity).Third, RES hinders safe grid operation because they offer little voltage and frequency control.

Challenge 1: Variability and Uncertainty — The first challenge is variability and uncertainty. Solar and wind generation exhibit high temporal variability across multiple timescales. Solar generation follows predictable daily and seasonal cycles: generation peaks around midday and drops to zero at night, with seasonal variation driven by day length and cloud cover. Wind output, by contrast, is driven by weather patterns that are difficult to forecast more than a few hours ahead, with wind farms experiencing sudden changes when weather systems move through a region. This variability must be balanced by flexible supply or demand resources to maintain *the fundamental physics requirement of power systems: generation must equal consumption at all time.* This balance was previously maintained by quickly ramping thermal generators or demand-following hydro plants.

Challenge 2: Curtailment and Grid Integration — The second challenge is curtailment and grid integration costs. As renewable penetration increases, curtailment is rising globally [37]. This trend is likely to accelerate unless flexibility is procured or transmission infrastructure is expanded. Expanding transmission infrastructure takes many years of planning, environmental review, and permitting, whereas demand-side flexibility can be deployed at existing facilities within months. The economic case for flexibility is therefore compelling: avoiding curtailment through flexibility saves renewable generation that would otherwise be wasted, while avoiding CAPEX-intensive infrastructure expansions.

Challenge 3: Safe Grid Operation — The third challenge is frequency and voltage stability to ensure safe operation of the grid. Traditional synchronous generators provide inertia. Rotating generator shafts resist rapid frequency changes by converting kinetic energy stored into electrical energy [38]. This inertia is a critical grid service that has been provided automatically in the past. Renewable generators connected via power electronics (inverters) lack this natural inertia [39]. As renewable penetration increases and conventional generators retire, the grid's ability to withstand disturbances deteriorates. Modern grids therefore require fast-acting flexibility from energy storage assets and demand-side resources to restore frequency and voltage stability at millisecond to second timescales.

Together these challenges impact every aspect of power systems from the power generation–power consumption balance, to the markets, and to the network infrastructure and operation.

2.2.2. Impact on Distributions Systems

The energy transition is fundamentally transforming the distribution system's role. With distributed renewable generation (e.g., rooftop solar now represents about two third of installed PV capacity in Germany [40]), distributed storage (home and commercial batteries), and responsive loads (flexible HVAC, Electric Vehicle (EV) charging), the distribution network is becoming **active**. The DSO (i.e., the entity responsible for operating distribution networks) must now coordinate these resources to manage multiple new challenges that were absent in the centralized generation paradigm.

Congestion Management — DSOs must manage congestions. When many rooftop solar systems export power simultaneously during midday, or when many buildings charge EVs in the evening, local distribution lines can become overloaded, with power flows exceeding the line's thermal rating. This can damage equipment or, at minimum, violates grid code constraints. Rather than building new cables and substations (a capital-intensive solution taking years of planning), DSOs can use flexibility services from buildings and batteries to reduce simultaneous consumption or generation. For example, a DSO might contract with an aggregator to reduce consumption by 10 MW on specific distribution feeders when solar generation exceeds a threshold, thereby preventing line overload [41].

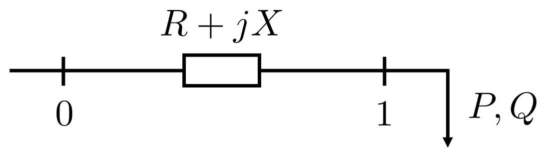


Figure 2.2.: Simple distribution system line with two buses, resistance R, and reactance X. The power consumption $P + jQ$ at node 1 is counted positive.

Voltage Stability — DSOs must maintain voltage stability. Electricity distribution systems operate with voltage targets of typically 400 V (nominal). High penetration of distributed generation can cause voltage to rise significantly above this target. When rooftop solar exports power to the distribution network, voltage rises especially at the end of long distribution branches where the solar systems are connected. Under mild assumptions[1], an estimate of the voltage drop magnitude between two busses $|\Delta V|$ of a distribution system aerial line (represented by Figure 2.2) can be obtained as:

$$|\Delta V_{01}| = |V_0| - |V_1| \simeq \frac{RP + XQ}{V_1} \tag{2.1}$$

where R and X are the total resistance and reactance of the line, P an Q are the active and reactive powers transiting on the line (counted positive from bus 0 to 1), and V is the voltage. Voltage outside acceptable ranges (typically 360–440 V, or $\pm 10\%$ according to European standard EN50160) causes equipment malfunction and can trigger protective devices that disconnect parts of the network. DSOs may use flexible loads and reactive power support (from battery inverters or other devices) to maintain voltage within acceptable ranges. For example, if voltage on a feeder is high due to solar export, the DSO can ask flexible loads (buildings with demand response capability) to increase consumption, thereby absorbing some of the solar export and reducing voltage [41].

Deferral of Infrastructure Investment — In locations with rapid demand growth (e.g., areas with many EV chargers being installed, or neighborhoods undergoing urban development), current substations and distribution lines may become inadequate. Building new infrastructure is expensive, time-consuming, and environmentally disruptive. DSOs can use flexibility services to defer or avoid these investments entirely. For example, if an urban neighborhood anticipates a 30% increase in electricity demand from EV charging over the next 5 years, rather than immediately upgrading the substation, the DSO can contract with buildings to provide peak-load reduction on days when demand is high, smoothing the peak and deferring the substation upgrade by several years. This is often called "flexibility-based infrastructure deferral" [41].

DSO/TSO Interaction — DSOs must coordinate with the Transmission System Operator (TSO). The DSO operates distribution networks (the last

[1] The capacitance of the line is assumed to be negligible and the phase difference between V_0 and V_1 to be very small.

mile to consumers), while the TSO operates transmission networks that connect generation to distribution systems. A critical challenge is that flexibility activated at the distribution level can have unintended consequences at the transmission level. For example, if a DSO activates flexibility to manage a local congestion problem by reducing demand on a specific distribution feeder, but this flexibility is activated simultaneously across many DSOs in the same region, the cumulative effect might be a sudden drop in demand at the transmission level, causing frequency to rise and requiring the TSO to activate different resources to maintain stability [42]. To address this, DSOs and TSOs must develop coordination frameworks where flexibility needs and availabilities are communicated bidirectionally, and algorithms ensure that flexibility activated at one level does not create problems at the other level [43].

2.3. The Need for Flexibility

Flexibility is recognized as one of the most cost-effective solutions for integrating high renewable penetration. A Brattle Group report estimated that demand-side flexibility in the United States could provide 200 gigawatts (GW) of capacity by 2030, which represents 20% of the estimated peak load in the same year, with a value exceeding $15 billion per year in avoided infrastructure investment [44]. Duke University researchers calculated that demand-side flexibility mechanisms could free up at least 76 GW (10% of current US peak demand) with an average load curtailment duration of only two hours per year, without requiring new infrastructure or technology breakthroughs [45]. At the global level, the International Energy Agency identifies flexibility (alongside storage and transmission investment) as one of three essential pillars for achieving the tripling of renewable capacity by 2030 without sacrificing grid stability [35].

Flexibility can be categorized along multiple dimensions. Although this section offers a very brief overview, there are extensive analyses such as [46]. Along the *temporal scale* dimension, energy-shifting flexibility operates over hours to days (e.g., pre-cooling a building during low-price periods to reduce consumption during peak hours), while frequency response flexibility operates at sub-second to minute timescales (e.g., a battery responding to grid frequency deviation within milliseconds). Along the *direction* dimension, downward flexibility increases consumption or reduces generation to absorb generation surplus, while upward flexibility reduces consumption or increases generation to meet demand during generation scarcity. Along the *source* dimension, supply-side flexibility comes from quickly dispatchable generation, demand-side flexibility comes from responsive loads, and storage-side flexibility comes from thermal or other energy storage. Along the activation mechanism dimension, price-based flexibility is implicit, as it reacts voluntarily to market price signals without prior contractual obligation. In contrast, incentive-based flexibility is explicit, since activation occurs under contractual arrangements or dedicated programs that compensate participants for providing predefined flexibility. Finally, constraint-based flexibility is triggered by system operators in response to physical limitations or contingencies in the grid, serving as a form of emergency response when market or contractual mechanisms are insufficient [47]. Table 2.1 summarizes this information.

Dimension	Category	Description / example
Temporal scale	Energy-shifting	Hours to days (e.g., pre-cooling a building during low-price periods to reduce consumption during peak hours).
	Frequency response	Sub-second to minutes (e.g., a battery responding to grid frequency deviation within milliseconds).
Direction	Upward	Increases consumption or reduces generation to absorb generation surplus.
	Downward	Reduces consumption or increases generation to meet demand during generation scarcity.
Source	Supply-side	Quickly dispatchable generation.
	Demand-side	Responsive loads.
	Storage	Thermal or electrical storage.
Activation mechanism	Price-based	Responds to electricity prices (market-based).
	Incentive-based	Responds to compensation mechanisms or grid signals (contract-based).
	Constraint-based	Activated when the grid faces physical limits (emergency response).

Table 2.1.: Summary of flexibility features.

The thesis focuses on demand-side flexibility from buildings, specifically, the ability of HVAC systems to shift electricity consumption over hours while maintaining occupants' comfort. This flexibility source is vast in magnitude. Buildings account for 18% of global electricity demand, marking an increase by 161% from 1990 to 2019 [48]. In the United States, the building sector accounts for 75% of the final electricity consumption, in which HVAC consumption contributes 40% [5], [6]. This flexibility source is widely distributed across electrical networks, ensuring that no single facility dominates supply, reducing concentration risk. It is increasingly connected to advanced controls via smart building management systems and the Internet of Things [49]. *Building flexibility remains significantly underutilized compared to its technical potential, largely because current control methods struggle to effectively and safely exploit thermal dynamics.* The thesis methods address this gap.

2.4. Flexible Assets and Demand Response Mechanisms

Modern power systems draw flexibility from diverse sources, each with different characteristics and suitable applications. Batteries, particularly lithium-ion systems, provide flexibility from milliseconds to hours, with response times under one second and typical duration of a few hours. Demand response from HVAC and water heating provides flexibility at minute-to-hour timescales with response times of 1–10 seconds. Pumped hydro facilities provide flexibility at hour-to-day timescales with response time in the order of the second and duration of several hours. Compressed air energy storage provides similar characteristics to pumped hydro. Fast-ramping gas turbines provide flexibility at minute-to-hour timescales with response times of 1–10 minutes and duration of 1 hour or longer [47], [50], [51].

Source	**Response Time**	**Timescale**	**Duration**
Lithium-ion Batteries	<1 s	ms to hours	Few hours
Demand Response (HVAC)	Seconds	Min to hours	Weather-dpdt 0.5–2 hours
Pumped Hydro	Seconds	Hours to days	Several hours
CAES	Seconds to min	Hours to days	Several hours
Fast-Ramp Gas Turb.	1–10 min	Min to hours	>1 hour

Table 2.2.: Flexibility Characteristics of Modern Power System Sources

As shown in Table 2.2, no single technology offers universal solution. Each technology has distinct characteristics tailored to specific operational context. Consequently, accommodating high renewable energy penetration necessitates a diverse technology mix. Figure 2.3 shows that storage is sufficient to cover

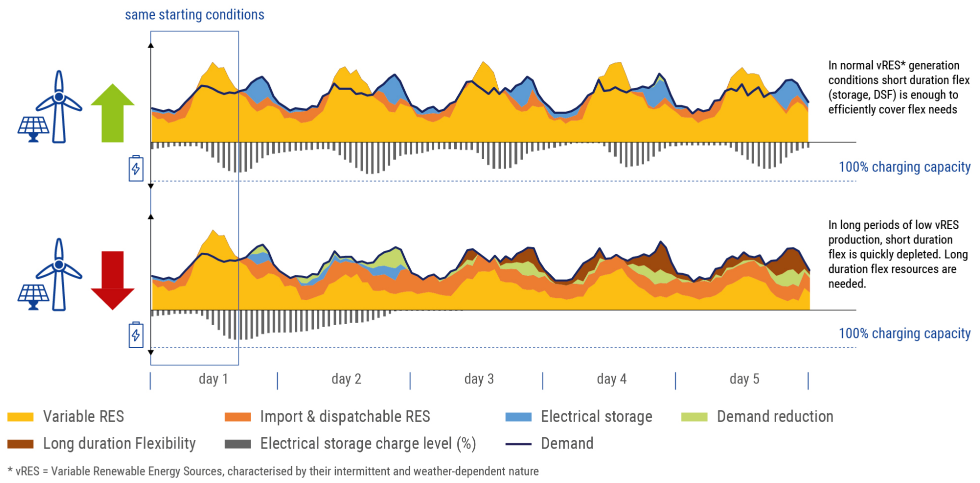


Figure 2.3.: Comparison of short- and long-duration flexibility needs for high renewable penetration: (top) normal weather with sufficient RES import/export; (bottom) low RES periods [52].

short duration flexibility needs of about one day. However, for longer period with low renewable energy generation, longer duration flexibility (e.g., PHES) and demand-side flexibility. In particular, residual net load (light green) highlights the role of diverse assets like HVAC beyond batteries.

2.4.1. Flexibility Procurements

Flexibility in Europe is procured by TSOs and DSOs through a set of partly coordinated markets and mechanisms that span different time scales. These arrangements have evolved substantially over the past decade, but they share the common aim of ensuring that sufficient upward and downward flexibility is available to maintain system balance and respect network constraints, while minimizing total system cost.

At the TSO level, flexibility for system balancing is mainly procured in advance via reserve capacity markets and then used in balancing energy markets. In the day-ahead time frame, TSOs procure reserve capacity products with defined activation times and durations. In European terminology, this includes Frequency Containment Reserve (FCR), automatic Frequency Restoration Reserve (aFRR), and manual Frequency Restoration Reserve (mFRR). Market participants submit offers to provide upward and/or downward reserve capacity, and the TSO clears the auction to meet exogenously defined reserve requirements at least cost, subject to technical constraints such as unit commitment, ramping limits and cross-zonal transmission capacity [53]. Flexibility can be provided by conventional generators, storage units, and demand-side resources, either directly or through aggregators.

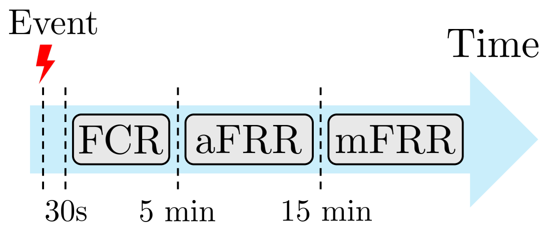


Figure 2.4.: Time line of the reserve activations in Europe.

In real time, balancing markets are used to activate this procured flexibility to correct deviations between scheduled and actual injections and withdrawals. Activation of aFRR and mFRR is typically organized through merit-order lists or real-time auctions, with resources remunerated both for capacity (availability) and for activated energy. FCR is usually contracted in advance with strict performance requirements in the sub-minute range, providing the first line of defense for frequency stability, while aFRR and mFRR correct imbalances on time scales from minutes to half an hour. Figure 2.4 summarizes the time line of the activation process. In most European systems, energy and reserve markets are cleared sequentially rather than jointly, so market design must carefully account for the technical interdependencies between energy dispatch and reserve availability [54].

2.4.2. Distribution System Operators and the Procurement of Flexibility

In recent years, *distribution-level flexibility services* have emerged as an important new market segment. DSOs increasingly procure flexibility to manage local constraints such as congestion (power flows exceeding line capacity), voltage violations (voltage outside acceptable ranges), and anticipated overloads (when new loads like EV chargers exceed current substation capacity). Rather than building new physical infrastructure (new substations, cables), DSOs can use flexibility services to manage these constraints more cost-effectively. This represents a fundamental business model shift for DSOs from infrastructure-centric to service-centric operation [41].
DSOs procure flexibility through several mechanisms adapted from practice at transmission level but tailored to distribution-specific needs.

Bilateral Contracts — Bilateral contracts are direct agreements between the DSO and a flexibility provider (e.g., a large building with demand response capability, or an aggregator representing many smaller buildings). The contract specifies the magnitude of flexibility (e.g., "reduce consumption by 5 MW"), the duration (e.g., "for up to 2 hours per day"), the response time (e.g., "within 10 minutes of activation"), and the compensation (e.g., "€50/MWh of consumption reduction") [41].

Tender processes — Tender processes are requests for proposals (RFP) where the DSO describes a flexibility need and aggregators submit bids specifying the price, quantity, and terms of flexibility they can provide. The DSO then selects the lowest-cost bids that satisfy the requirement. For example, a DSO might issue an RFP: "Procure 20 MW of downward flexibility (consumption reduction) on feeder X for the period May–September, activation duration up to 2 hours per day, response time within 15 minutes." Aggregators would submit bids specifying how much flexibility they can provide and at what price, and the DSO selects the least-expensive combination of bids. This type of solution is already implemented by UK Power Networks [55].

Market platforms — Market platforms are emerging in some countries, where flexibility providers continuously update their availability and pricing in a digital platform, and DSOs can procure flexibility from the platform in real-time. Examples include the Piclo Flex platform [56]. These platforms enable dynamic, auction-based procurement of flexibility, potentially reducing procurement costs compared to bilateral contracts [41].

DSO/TSO coordination — A critical emerging challenge in modern power systems is coordinating flexibility procurement between the DSO (managing distribution networks) and TSO (managing transmission networks). If a building's HVAC system is activated to provide congestion management service to the DSO (by reducing demand on a specific distribution feeder), but this activation simultaneously reduces demand at a moment when the TSO needs to increase demand to balance frequency, the two services work against each other, and the combined effect might be worse than doing nothing [41]. To address this, research institutions and grid operators worldwide have proposed hierarchical coordination schemes. The DSO identifies flexibility needs to manage local constraints and creates "flexibility bids"—quantities and timing of flexibility availability that could be used for local services. The DSO forwards any unused flexibility to the TSO, which can procure it for system-level services (frequency support, reserves). A coordination algorithm ensures that flexibility activated at the TSO level does not violate local DSO constraints (e.g., does not cause voltage violations or line overloads) [57]. Real-world implementations of this coordination are emerging in the European Union, with increasing emphasis on technical standards and market rules to ensure safe and efficient operation.

2.4.3. Demand Response

Demand response traditionally referred to short-term adjustments in electricity consumption by end-users in response to direct requests from utilities or grid operators, often with financial incentives[2]. The IEEE Power& Energy Society (PES) Task Force on "Demand Response in the DER Era" defines demand response as follows.

Definition: Demand Response

Demand response is the actions of customer-sited energy resources downstream of metering points to voluntarily, actively, and temporarily adjust their electricity production and/or consumption in response to signals (e.g., commands, prices, measurements) [58].

Modern demand response programs vary widely. Some are voluntary, where customers receive incentive payments for reducing consumption during high-demand periods. Others are mandatory, where grid code requirements compel certain customers (typically large industrial users) to maintain the ability to reduce consumption on short notice. The most advanced programs are real-time price-responsive, where electricity prices vary continuously (updated every 5–15 minutes), and consumers adjust consumption automatically in response to these dynamic retail prices, enabled by smart meters and control systems [59]. In contrast, most widespread pricing schemes are *time-of-use tariffs*. Under such tariffs, suppliers or network operators define time blocks (typically peak, shoulder, and off-peak periods) and assign higher prices to peak hours with high system demand and lower prices to off-peak hours when demand and generation costs are lower. Figure 2.5 displays the current time-of-use tariff offered by the main Walloon DSO, ORES [60].

Unlike conventional supply-side generation, demand response offers several unique advantages. First, *non-intrusive deployment*: flexibility can be implemented within existing buildings and devices without major new infrastructure

[2] The name *load flexibility* is sometimes preferred to demand response.

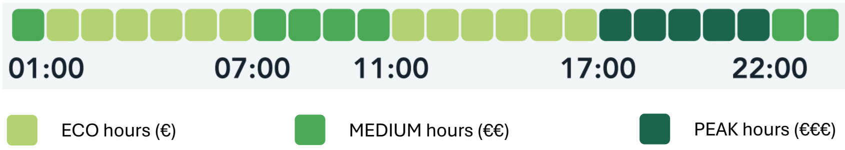


Figure 2.5.: Time-of-use tariff named *Impact* as released by ORES (Walloon main DSO) for 2026, adapted from [60].

investment. A building with a smart thermostat can participate in demand response without any capital expenditure beyond the controller itself. Second, *distributed availability*: unlike centralized power plants, flexible loads are distributed across thousands of buildings, providing natural geographic diversity and reducing concentration risk. If one flexibility provider defaults or fails, the remaining providers can still meet system needs. Third, *low environmental impact*: shifting demand rather than generating additional electricity reduces carbon emissions, air pollution, and thermal water pollution [61]. Fourth, *cost competitiveness*: demand response often costs less to procure than building new generation or storage capacity, making it economically attractive to both operators and customers [62].

However, demand response also introduces challenges that must be carefully managed. *Heterogeneity* is significant. For example, buildings and industrial processes vary widely. A small lightweight office building with large windows and poor insulation has very different flexibility characteristics than a heavy concrete hospital building. Similarly, occupancy patterns vary by building type and location, affecting thermal dynamics and comfort constraints [63]. *Observability* is another challenge: grid operators and aggregators typically cannot directly observe the state of individual assets, requiring forecasts and inference from electricity consumption data. This partial observability makes it difficult to predict and guarantee the flexibility available from each asset [63]. *Rebound effects* occur when, after a demand reduction event, consumption increases ("demand rebound"), for instance, to restore comfort in building and catch-up with delayed production in industry. If many assets experience rebound simultaneously, it can create a new peak in demand that coincides with other high-demand periods, reducing the net benefit of demand response [62]. *Consumer acceptance* limits the magnitude of flexibility that can be sustained: occupants experience discomfort if temperature setpoints are adjusted too aggressively, and sustained discomfort can reduce participation in demand response programs [64]. Managing these constraints while achieving grid benefits requires sophisticated control algorithms and careful coordination.

Scope Clarification

This thesis focuses on decision-making tools for operators of HVAC systems who wish to improve their participation in demand response schemes. The design and implementation of demand response schemes is out of the scope of this work. In subsequent chapters, demand response is always voluntary-based and price-based.

2.5. HVAC Systems as a Flexibility Resource

HVAC systems dominate building energy consumption worldwide and represent perhaps the largest, most accessible flexibility resource for the modern grid. In the United States for example, the HVAC consumption reaches 30% of the total electricity production [5], [6]. This enormous consumption level is only comparable to industrial processes and transportation. HVAC is also the single largest contributor to peak electricity demand on summer afternoons, when outdoor temperatures are highest and cooling demand is greatest [63].

From an economic perspective, HVAC flexibility is valuable because it can provide both energy and capacity services. Energy services arise from shifting consumption to hours when electricity prices are lower or carbon intensity is reduced. For example, during the middle of a sunny day when solar generation is high, electricity prices are typically low. By pre-cooling/heating a building during these low-price hours (and allowing temperature to drift higher during peak-price afternoon hours), the building can reduce the total electricity bill [65]. Aggregating this flexibility across thousands of buildings creates a demand response program that flattens the system load profile, reducing the need for expensive peak-generation resources [66]. Capacity services arise from reducing peak-hour consumption, which avoids costly infrastructure investments. Because peak-hour electricity demand determines the required capacity of generation, transmission, and distribution systems, a 5% reduction in peak demand can defer multi-million-euro investments in substation upgrades or new distribution lines [44]. This infrastructure deferral value is substantial, especially in fast-growing regions where new loads (such as EV charging) are rapidly increasing peak demand.

The physical basis of HVAC flexibility is rooted in a simple principle: *thermal inertia*. Buildings have mass (concrete structural elements, interior walls, furnishings) that stores thermal energy. This storage capacity allows buildings to tolerate temporary setpoint adjustments without violating comfort constraints. When an HVAC system is turned off on a summer afternoon, indoor temperature does not instantaneously jump to ambient temperature. Rather, heat flows slowly through the building envelope, and indoor temperature drifts upward gradually. This slow thermal response creates an opportunity for demand response: by pre-cooling/heating a building during off-peak hours (when electricity prices are low), storing that thermal energy in the building's mass, the HVAC system can be partially or fully disabled during peak hours when prices are high, allowing temperature to drift upward without exceeding comfort bounds.

The dynamics of indoor temperature in a building zone can be represented mathematically by a simple but effective *Reduced-order Resistance-Capacitance (RC) model* [13]:

$$C\frac{dT_{\text{in}}}{dt} = \frac{T_{\text{amb}} - T_{\text{in}}}{R} + Q_{\text{hvac}} + Q_{\text{internal}}, \tag{2.2}$$

where C is the thermal capacitance of the building zone (joules per kelvin), representing the thermal mass, T_{in} is the indoor temperature (kelvin), T_{amb} is the ambient temperature (kelvin), R is the thermal resistance of the building envelope (kelvin per watt), representing the insulation quality, Q_{hvac} is the heat transfer rate from the HVAC system (watts), and Q_{internal} is the internal heat gain from occupants, equipment, and solar radiation (watts). The RC model is a low-order approximation of building thermal dynamics, but it captures the essential behavior: the time constant $\tau_{RC} = RC$ quantifies the rate at which indoor temperature responds to external changes. In commercial buildings, τ_{RC} typically ranges from 1–8 hours. If the HVAC system is turned off and assuming no other heat gains, indoor temperature τ^{in} will drift slowly towards ambient temperature τ^{amb}. After a time period of τ_{RC}, 38.6% of the temperature difference $\tau^{\text{amb}} - \tau^{\text{in}}$ has reached inside:

$$\tau^{\text{in}}(\tau_{RC}) = \tau^{\text{in}}(t_0) + 0.386\big(\tau^{\text{amb}} - \tau^{\text{in}}(t_0)\big), \tag{2.3}$$

where t_0 is the time at which HVAC. After five times the constant τ_{RC}, the building indoor temperature is assumed to be τ^{amb}. This slow dynamics creates the opportunity: by pre-cooling or pre-heating a building during off-peak hours when electricity is cheap, the building can pass through peak hours with reduced HVAC operation, shifting electricity consumption from peak to off-peak periods.

The flexibility available from building HVAC systems can be quantified as the range of power consumption that the HVAC system can achieve while maintaining indoor temperature within comfort bounds. Formally, this feasible operating region is defined by:

$$P_{\text{hvac}}^{\min}(t) \leq P_{\text{hvac}}(t) \leq P_{\text{hvac}}^{\max}(t), \tag{2.4}$$

where $P_{\text{hvac}}^{\min}$ and $P_{\text{hvac}}^{\max}$ are the minimum and maximum feasible HVAC power at time t. These bounds are derived from multiple factors. Physical equipment limits mean that HVAC systems have maximum heating and cooling capacity (imposed by the equipment design) and minimum power required for ventilation (necessary for occupant health). Comfort bounds require that indoor temperature remains within acceptable limits, typically 20–24°C in winter and 22–26°C in summer, with stricter limits in hospitals and operating rooms. Thermal dynamics, governed by Equation (2.2), constrain how quickly temperature can be changed. Aggressive temperature adjustments that maximize flexibility in the short term may violate comfort bounds if sustained too long. Prediction uncertainty in ambient temperature, solar irradiation, and occupancy requires that flexibility plans account for forecast errors. A conservative flexibility strategy that guarantees comfort even under worst-case weather and occupancy scenarios provides less flexibility than an optimistic strategy that assumes perfect forecasts, but is more robust to forecast errors.

Research demonstrates that building HVAC flexibility is substantial and economically significant. A robust optimization study on unlocking thermal inertia for demand response found that buildings can reduce peak consumption by 15–30% during a two-hour demand response event while maintaining comfort constraints, simply by using pre-cooling strategies timed to the demand response period [65]. Across a portfolio of thousands of buildings, this translates to hundreds of megawatts of aggregated flexibility in a typical city. When this flexibility is valued at typical wholesale electricity market prices (e.g., €50–100/MWh during peak hours), the annual value of flexibility provision from a single commercial building with average cooling load of 500 kW could exceed €10,000 per year, providing attractive payback for investments in building controls and communication infrastructure.

2.6. Control Challenges and Opportunities for Machine Learning

The previous sections have established that HVAC flexibility is enormous in magnitude, widely distributed, economically valuable, and essential for integrating renewable energy and managing modern distribution networks. Despite buildings' physical capacity to provide flexibility, it remains significantly underutilized. The fundamental barrier is not technical inability but rather control and information challenges that are difficult to overcome with classical control methods. This section articulates these challenges and previews how the thesis methods address them.

Traditionally, most building HVAC control are rule-based thermostat. For instance, the temperature is kept constant at 21°C during the day but can fall down to 18°C during the night. Such simple control scheme do not allow for participation in demand response programs, which is the objective of this thesis.

Demand response in buildings hinges on the ability to (i) predict the thermal response to HVAC actions and exogenous conditions and (ii) encode this response in an optimization model that arbitrages dynamic electricity prices while preserving comfort and respecting equipment limits. From a modeling perspective, this requires a mapping

$$(x_t, u_t, w_t) \to x_{t+1}, \tag{2.5}$$

where x_t collects indoor temperatures and other relevant states at time t, u_t denotes the HVAC control inputs, and w_t represents exogenous disturbances (weather, occupancy, internal gains). Model identification enters precisely at the level of this mapping, either by parameterizing (simplified) physics-based models or by learning data-driven surrogates. Physics-based models require a complete knowledge of the building (layout, material, HVAC system) and the expertise to understand the data. Some building parameters, such as the materials resistance and capacitance, are intrinsically unknown due to aging and installation process which may cause large error. Therefore, this thesis focuses on data-driven models because of their versatility and their ability to model nonlinear relationship. A review of existing models for building thermal modeling is conducted in section 3.2.

Advanced data-driven building HVAC control follows a two-stage paradigm called ITO. In the first stage, called *identification*, the building manager collects months of historical building operation data. This data includes measurements of indoor temperature, ambient temperature, solar radiation, occupancy, and HVAC energy consumption. Using statistical methods, the building manager learns a data-driven model that best fit the historical data. The goal is to minimize prediction error on this historical data, often quantified as the MSE between predicted and measured temperatures. Once the model is identified, it is considered final and is used for all subsequent control decisions.

In the second stage, called *optimization*, the building manager uses the identified thermal model to solve an optimal control problem. For example, the optimization problem aims to find the HVAC setpoint trajectory over a day-ahead horizon that minimizes electricity cost, subject to comfort constraints (indoor temperature within acceptable bounds), physical constraints (HVAC equipment limits), and the thermal dynamics described by the identified model.

In practice, model identification and optimization are strongly coupled. If the model is inaccurate in operating regimes that are exploited for demand response (e.g., aggressive pre-heating before a peak price period), then even an optimal controller with respect to that model may perform poorly on the real building. *This motivates the use of ML not only for its versatile nonlinear architectures, but also to learn models that are tailored to the downstream control task.* This is the overarching goal of this thesis.

2.7. Summary and Thesis Positioning

Buildings, and particularly their HVAC systems, represent an enormous, underutilized flexibility resource that is essential for integrating renewable energy, managing modern distribution networks, and achieving net-zero carbon emissions. HVAC flexibility is physically large, economically valuable, widely distributed (thousands of buildings in every city), and technically mature (most HVAC systems can be retrofitted with smart controls). Despite this potential, current control methods (ITO) are not adapted because they separate model identification and control optimization. This separation results in model-reality mismatch and poor generalization to new operating regimes.

This thesis develops three interconnected contributions that enable better utilization of building flexibility through improved ML and control methods. The first contribution is *DFL for simultaneous identification and control*, ensuring that learned models are optimized for the quality of demand response actions rather than prediction accuracy. The second contribution is *integration of NN thermal models into optimization problems* via SS, enabling end-to-end training without requiring system differentiability. The building thermal modeling can capture nonlinear dynamics while focusing on decision quality. The third contribution is *a probabilistic DFL framework that trains a context-aware probabilistic RC model* so that the thermal dynamics model evolves with environmental information such as weather and occupancy forecasts. Moreover, this last contribution enables demand response actions robust to uncertainties affecting thermal comfort. Together, these methods enable the shift from passive buildings to grid-interactive, autonomous buildings that actively participate in electricity markets and flexibility services, contributing to a reliable, efficient, and decarbonized energy system.

CHAPTER 3.

Decision-Focused Learning for Convex Control and Modeling of Building Thermal Dynamics

This chapter studies DFL for building thermal dynamics modeling and control (convex policies only). The chapter starts with a quick summary of the context and motivations that are detailed in chapter 1 and chapter 2. Then, we review the literature about building thermal modeling and DFL (section 3.2). We present the contributions of the chapter in section 3.3. Section 3.4 describes the problem class, explains gradient computations necessary to learn the parameters, the loss function design for non-differentiable black-box systems, and training robustness enabled by constraint relaxation. Afterwards, section 3.5 presents the case study. In particular, sections 3.5.2 and 3.5.4 describe the specific optimization problem, and the loss function for the day-ahead HVAC scheduling, respectively. We report the results in section 3.5.5. Section 3.6 discusses the results and limitations of the case study. Section 3.7 draws the main conclusion and outlines directions for future improvements.

3.1. Motivation

Optimization is used to devise control policies such as for HVAC in buildings [67], the guidance of vehicles [68], and the planning of complex dynamic systems such as day-ahead energy scheduling of active buildings [69], or energy management systems for micro-grid users [70]. The optimization problem solution is the actions or decisions to apply to the system. However, an analytical model of the system dynamics is always assumed to be known and formulated within an optimization problem, usually as constraints. The necessity to have an analytical model hinders control tasks where physics-based models are unpractical; either because the system is a black box, or physical models are far too complex. Data-driven surrogates become impractical when system dynamics change drastically due to a newly applied control policy, rendering prior historical observations incomplete and minimally informative. Therefore, the conventional two-stage ITO approach is inefficient.

This chapter addresses the ITO approach inefficiencies using the day-ahead HVAC scheduling in buildings as a use case. HVAC scheduling becomes increasingly important because of the roll-out of smart meters and dynamic electricity tariffs. The underlying goal of such dynamic (e.g., time-of-use) tariffs is to prompt more flexibility at the consumption level to accommodate more renewable-based, and thus often weather-dependent and uncertain, generation (see subsection 2.4.3). Unlike lightning or essential equipment that needs to be available on demand, HVAC can be scheduled and benefit from the building thermal inertia to shift the consumption across the day. Consequently, building managers can save money and help network management by optimizing HVAC operations.

3.2. Literature Review

3.2.1. Building Thermal Modeling

The goal of day-ahead HVAC scheduling is to find the optimal temperature set point profiles for each building thermal zone to minimize the electricity cost for the upcoming day, alleviate the potential strain on the power grid, and ensure thermal comfort. Therefore, building management systems rely on a thermal model of the building that relates the HVAC operation to the indoor temperature.

Historically, the thermal model has always been assumed to be known. Most HVAC models are physics-based (i.e., white-box) [71]. Physics-based models rely on detailed information about a building's physical characteristics, including construction materials, insulation properties, and ventilation dynamics. This category includes high-fidelity simulators such as EnergyPlus[17]. While these models provide highly detailed representations of buildings [71], some parameters, such as the materials resistance and capacitance, are inherently uncertain due to factors like aging and variability in the installation process, which can lead to significant errors [72]. Moreover, it requires expertise in building thermal modeling, which may not be readily available at each building site. But, there also exist simpler models: aggregated (or lumped) RC models. These models are linear and based on circuit analogy between heat transfer and electricity. The parameters of these models can be computed

based on the construction materials [73].
On the other end of the spectrum lies fully data-driven (i.e., black-box) models that rely on statistical or ML techniques trained on historical weather and energy consumption data. These models range from simple linear regressions [74] to highly complex nonlinear architectures, such as NN models. NN were widely used for system identification [75], but cannot be easily integrated into optimization models; otherwise, the problem becomes non-linear and there is no guarantee on the solution quality [76], [77]. Furthermore, data-driven models perform poorly out of their training zone [71]. This hinders day-ahead HVAC scheduling applications since their goal is to explore the whole feasible domain to find the best course of action. Reinforcement Learning (RL) has proven to be effective for real-time set point control [78], emission minimization [79], and peak demand reduction [80]. However, inputs and assumptions underlying RL methods are often impractical. For instance, the active RL algorithm in [81] assumes real-time measurement of the occupants' clothing value, number, and thermal comfort. A review concludes that model-based deep RL must be favored to leverage some prior knowledge [78]. Similarly, the review by [82] states that the gray-box models are the most promising.
Gray-box models involve physics-based formulation in which parameters are obtained through data-driven techniques. Gray-box models require less data than black-box models [83]. Among such models, the RC model (also named lumped capacitance model or network model), a reduced order physics model, has exhibited good performances for buildings where there is no important indoor air convection [73]. Furthermore, the RC model is linear, thus lending convexity to the optimization problem. However, estimating the parameters of the model is challenging. A first option to estimate the parameters is to analyze the building materials and select default tabular data. In addition to requiring building blueprints, this method does not account for manufacturing dispersion (i.e., variations in product dimensions, properties, or performance due to inconsistencies in the production process), the on-site installation process, and the aging of materials. Therefore, data-driven parameter identification has led to significantly improved RC models [84]. In [75], physics-informed NNs were employed for control-oriented predictions. However, these NN approaches did not consider day-ahead scheduling, which introduces fundamentally different requirements. Table 3.1 summarizes the thermal modeling approaches[1].

[1]A comprehensive review of the literature is provided in [10], Section III.

Table 3.1.: Comparison of thermal dynamics modeling approaches.

	Model	**Complexity**	**Data Requirement**	**Ref.**
White box	High-fidelity simulator	Very high	Geometry, materials, weather	[17], [72], [85]
	Lumped RC (known R/C)	Low	R and C	[73]
Gray box	Lumped RC (parameter fitting)	Moderate	I/O measurements for R/C	[84]
	Physics-informed NN	High	Partial physics + data	[75]
Black box	NN	Moderate	Historical input–output data	[86], [87]
	Regression models	Low	Historical system data	[74]

3.2.2. Decision-Focused Learning

Aware of the limitations in estimating the parameters on historical data, we learn the parameters in a task-aware manner. In [88], the authors were the first to suggest training a ML model in a directed manner based on its impact on the downstream decisions. The weights of the ML model that is used to predict the uncertain parameters of the downstream optimization problem are learnt to minimize the decision error induced by the parameters misestimation. To that end, the authors established the gradient of the optimization problem solution with respect to the problem parameters. This enables back-propagation through the optimization problem. Only unconstrained quadratic problems were considered. By allowing loss functions that depend explicitly on downstream optimization decisions (and implicitly on the ML model output), the calculation of optimization problem gradient paves the way for DFL. The framework was then extended to stochastic quadratic problems [89]. Because the gradient of unconstrained problems is easier to compute, the constraints were relaxed in the objective function.

Bounded Linear Programs (LP) may lead to zero-valued or undefined gradient. Geometrically, the solution of LP belongs to the set of the vertices of the feasible space polytope [90]. Consequently, an infinitesimal change in the parameter values of the objective function has no impact on solution, i.e., the solution remains on the same vertex. However, as soon as objective parameter changes become large enough, the LP solution jumps to another vertex. This results in a piecewise constant objective function, leading to a gradient that is either zero or undefined with respect to the objective parameters, making it uninformative for gradient descent training. With respect to equality constraints (and binding inequality constraints) parameters, the gradient is defined almost everywhere, and undefined elsewhere. For Combinatorial Programs (CP), the discrete nature of the decision variables leads to zero-valued gradient almost everywhere, and undefined elsewhere. The first method to calculate the gradient of an LP is to calculate the gradient of its Karush-Kuhn-Tucker (KKT) conditions. The gradient of the KKT conditions is also null or undefined since the KKT conditions are an exact representation of the same LP. To address this issue, the objective function can be augmented by a L_2 regularization of the decisions variables. This turns the LP into a strong concave or convex quadratic program if it is a maximization or minimization, respectively [23]. The second approach is to design a surrogate loss function. Smart "Predict, then Optimize" Plus (SPO+) is such a convex function [91]. Interestingly, SPO+ extend beyond LP and can be applied to any optimization problem with a linear objective function and linear, convex, and integer constraints. However, the uncertain parameters cannot appear in the constraints. Moreover, SPO+ requires the optimal decisions to be known, which may not always be the case. The third option for LP and CP is stochastic smoothing. A random perturbation is applied to uncertain parameters to smooth the loss function transition and, thus, create informative gradients from the otherwise null and undefined derivative [19]. Unlike SPO+, the uncertain parameters can appear in constraints.

Leveraging the differentiability of conic programs [92], Agrawal et al. demonstrated how controller parameters within the objective function can be learned assuming full knowledge and differentiability of the system, and observation of the system dynamics [93]. It was applied to the forecasting of electricity prices through wind power prediction [94]. It was also employed to split the demand for flexibility from the transmission grid towards the flexibility of the distribution grid assets [95].

Few prior works have simultaneously addressed system identification (i.e., learning constraint parameters, specifically the parameters of the dynamics model) and optimization, nor have they considered scenarios where the controlled system is both unknown and non-differentiable. Tackling this challenge requires a profound understanding of optimization and decision-focused learning theory, combined with extensive domain-specific expertise. There are only three previous work that does not assume knowledge of the system dynamics and differentiability:

- In [96], a convex surrogate for both the optimization problem and the performance loss is constructed to train a ML forecaster. However, the construction of the surrogate model requires knowing the true value of the uncertain parameters.
- In [97], a differentiable Model Predictive Control (MPC) policy jointly learns the thermal dynamics and optimizes HVAC operations. Even though pioneering, their approach relies on linear thermal models that yield convex optimization problems, limiting its applicability to more realistic, non-convex building dynamics. Moreover, no other constraint than the unknown thermal dynamics is considered.
- In [98], Cui et al. extend the approach of [97] to include a forecaster upstream of the convex optimization problem. They jointly learn the linear thermal model and a forecaster predicting the thermal model residuals. The thermal model is kept unidimensional, even for the multi-zone buildings considered in the case study, by weighted averaging.

Other applications of DFL include forecasting electricity prices [94] and optimizing the allocation of flexibility between transmission and distribution grid assets [95].

3.3. Contributions

This chapter presents three major contributions.

First, we leverage advances in DFL to perform complex system identification and control simultaneously. Compared to [93] and [91], uncertain parameters to be learned are within the constraints of the optimization control policy. Due to end-to-end learning, the system model parameters are optimized for operating zones relevant to the control policy. This framework bypasses the need for extensive high-quality historical database reflecting control data distributions that are rarely available in practice.

Second, we formulate the performance loss that is necessary to compute the quality of the decisions and backpropagate the gradient with respect to the dynamics model parameters. Unlike in [93], we do not assume that the system is differentiable. The only requirement is for the system state variables used in the system dynamics model to be observable. In addition, we introduce a hierarchical loss to inform the learning about the HVAC system structure and its impact on the objective value.

Third, training is made robust to infeasibilities by relaxing constraints on the dynamics model output. Moreover, feasibility in the early stages of DFL training is prompted by performing a warm-start with a model pre-trained on historical data. A small noise is added on the parameters after pre-training to get the model out of the local minimum while retaining a maximum of prior knowledge.

The performance and relevance of the method is illustrated on the day-ahead HVAC scheduling of a realistic 15-zone building located in Denver, CO, US [99]. We model the building using a multi-dimensional RC model. Furthermore, the method's robustness is evaluated under a distribution shift in the input parameters. Specifically, we analyze the model's performance metrics on a dataset representing an exceptionally hot year, a scenario likely to become increasingly common due to global warming.

3.4. A Method for Simultaneous System Identification and Control

The proposed framework performs system identification and control simultaneously. To that end, we start by describing the addressed problem class, then come the explanation about the gradient computation needed for updating the parameters. Afterwards, we propose a DFL supervised loss that bypasses the need for a differentiable system. Lastly, we address the infeasibility that may arise during training. Figure 3.1 provides an overview of the proposed method.

3.4.1. Problem Class

Given a system with dynamics described by the unknown state transition function:

$$x_{t+1} = f(x_t, u_t, w_t),\ \forall t, \tag{3.1}$$

the goal is to learn the parameters θ of a proxy model $\hat{f}(\hat{x}_t, u_t; \theta)$ that represents the system dynamics, while keeping the formulation of the control policy ϕ convex (3.2). The known system state at a given time step $t \in \mathcal{T}$ is given by $x_t \in \mathbb{R}^n$, the expected state by $\hat{x}_t \in \mathbb{R}^n$, and the command or action by

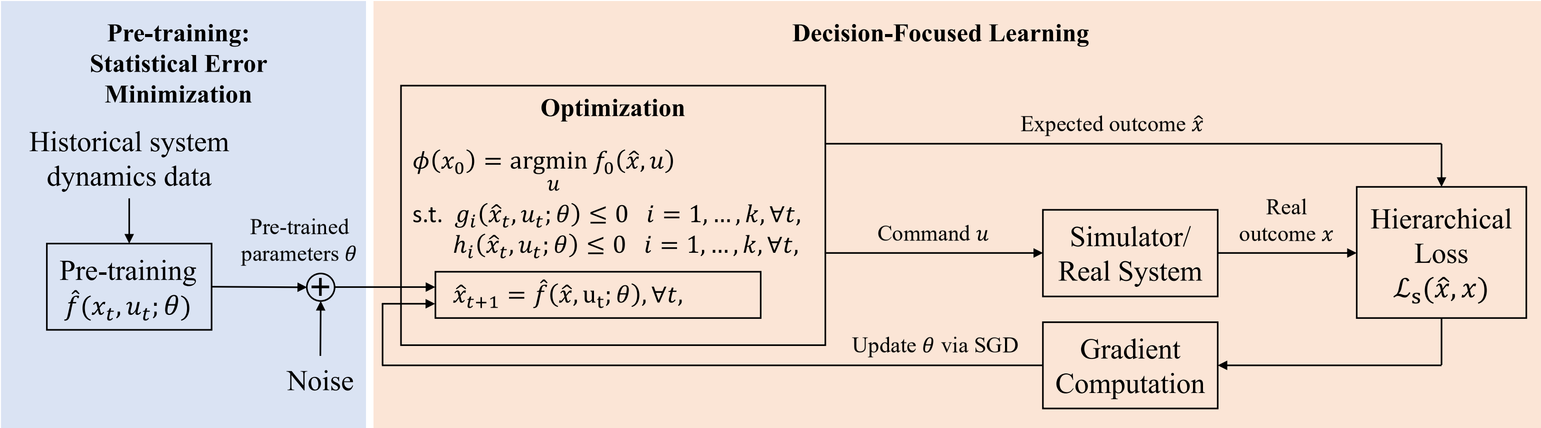


Figure 3.1.: The proposed framework starts by pre-training the system dynamics model on historical data. Then, the pre-trained model parameters are fed into the convex optimization control policy where they keep being updated on representative scenarios to minimize the loss metric evaluating the mismatch between the expected and observed system states.

$u_t \in \mathbb{R}^m$. The true state transition function can be affected by some noise w_t. We do not make any assumption on w_t.
A general formulation of the convex optimization control policy ϕ is given by:

$$\begin{aligned} \phi(x_0) = \underset{u}{\operatorname{argmin}}\, & f_o(\hat{x}, u) \\ \text{s.t. } & g_i(\hat{x}, u) \leq 0 \quad i = 1, ..., k,\ \forall t, \\ & h_i(\hat{x}, u) = 0 \quad i = 1, ..., l,\ \forall t, \\ & \hat{x}_{t+1} = \hat{f}(\hat{x}_t, u_t; \theta),\ \forall t. \end{aligned} \tag{3.2}$$

To keep the problem convex, the equality constraints h_i and $\hat{f}$ must be affine with respect to the optimization variables $\hat{x}$ and u whereas g_i must be convex. The convexity of the problem is crucial for the gradient computation as explained in section 3.4.2. In addition, convexity guarantees that the solution is the global optimum and offers tractable problems well handled by off-the-shelf solvers.

Even though we focus on control optimization policy problems and state transition functions, the proposed framework can be used to learn any parameters of a convex optimization problem.

3.4.2. Gradient Computation

We provide a brief explanation of CVXPYLayers in this section. The interested reader can find more information in Appendix B.

Let us assume we have a scalar function $\mathcal{L}$ that quantifies the loss of performance (i.e., the suboptimality) caused by the misestimated parameters θ. The parameters θ can be learned while controlling the system with the policy ϕ by minimizing $\mathcal{L}$. This requires computing the gradient $\frac{\partial \mathcal{L}}{\partial \theta}$. Then, θ can be updated iteratively by any gradient-based optimization algorithm. Misestimating θ leads to suboptimal commands u that are responsible for the performance loss. By applying the chain rule, the gradient becomes:

$$\frac{\partial \mathcal{L}}{\partial \theta} = \frac{\partial \mathcal{L}}{\partial u} \frac{\partial u}{\partial \theta}. \tag{3.3}$$

The first factor accounts for the performance loss caused by a variation of the commands u. The second factor creates the need for differentiating through the optimization problem. More specifically, the impact of a variation in the optimization parameters θ on the optimization output u must be determined.

An optimization problem can be considered as a mapping $O : \mathbb{R}^p \rightarrow \mathbb{R}^n$ between its parameters and the optimal solution. Differentiating such a problem

is challenging because of the complex and implicit mapping definition. Moreover, some classes of constrained programs, such as the linear programs, define piecewise-constant mappings in which the gradient is often null or nonexistent [11].
A general framework, named *CVXPYLayers*, is able to differentiate conic programs [92]. Since any convex program can be recast as a conic program [100], *CVXPYLayers* is very versatile. A conic program is of the form

$$\begin{aligned} &\min_x c^T x \\ &s.t.\ b - Ax \in \mathcal{K}, \end{aligned} \tag{3.4}$$

where $(A, b, c) \in (\mathbb{R}^{m\times n}, \mathbb{R}^m, \mathbb{R}^n)$ are the parameters of the problem, $x \in \mathbb{R}^n$ is the primal decision variable, $\mathcal{K} \subseteq \mathbb{R}^m$ is a nonempty, closed, convex cone. Under the assumption that (3.4) has a unique solution, its gradient can be computed. Solving a conic program can be done in three steps.

1. The data parameters (A, b, c) are used to build the corresponding skew-symmetric matrix Q [101], [102].

2. The self-homogeneous dual embedding problem, which reformulates the KKT conditions of the original conic program into a single system of equations and incorporates the matrix Q, is solved by finding a root (i.e., a zero) to its normalized residual map s [103]. The residual map is a function that quantifies the difference (or residual) between the current solutions of the primal and dual problems and the optimal solution. Therefore, it guides the search for the optimal solution, directing the optimization process.

3. The solution z is retrieved by R and mapped to a valid solution of the original problem.

Therefore, the mapping O can be seen as the composition of three submappings $R \circ s \circ Q$. By the chain rule, we have

$$\mathrm{D}O(A, b, c) = \mathrm{D}R(z)\ \mathrm{D}s(Q)\ \mathrm{D}Q(A, b, c) \tag{3.5}$$

with D, the derivative operator. The derivative of Q is straightforward since A, b, and c appear explicitly in its formulation. The derivative of the root-finding problem s is obtained via implicit differentiation. The derivative of the retriever R relies on the derivative of the euclidean projection of the self-homogeneous dual embedding solution z onto the feasible cone $\mathcal{K}^\star$ associated with the dual of the original conic problem (3.4) [92].

3.4.3. Performance Loss for Non-Differentiable System

As established by (3.3), the gradient of the performance loss with respect to the parameters θ must be defined and exist. The ideal loss prescribed in the literature is the regret $\mathcal{L}_r$ (3.6) [11], [91]. The regret assesses the objective value loss at optimality (marked by *) caused by the error between the actual θ and its estimator $\hat{\theta}$:

$$\mathcal{L}_r = f_o^*(\hat{\theta}) - f_o^*(\theta). \tag{3.6}$$

However, the definition of regret assumes the actual value of θ is known. Other loss functions have considered the transition state function f (3.1) as fully known and differentiable [93].

Many physical systems must be considered as black boxes, and thus non-differentiable, because of their inherent complexity and the unbearable cost to build an accurate system model. The gradient of black-box model can be estimated via simultaneous perturbation stochastic approximation [104]. However, it is unpractical if the system environment conditions are evolving (e.g. the weather) making the repetition of different commands in the same conditions impossible.
Consequently, we design a novel performance loss $\mathcal{L}_{\text{DFL}}$ that does not assume any knowledge about the system:

$$\mathcal{L}_{\text{DFL}} = \mathcal{L}_s(\hat{x}, x), \tag{3.7}$$

where $\mathcal{L}_s$ refers to any supervized loss such as the MSE loss or the Mean Absolute Error (MAE) loss. The loss $\mathcal{L}_{\text{DFL}}$ computes a statistical metric on the error between the expected system state $\hat{x}$, as predicted by the optimization model, and the actual observed realisation x. Similar to supervized learning, there is a target value, x, to which no gradient is attached. We refer to $\mathcal{L}_{\text{DFL}}$ as the supervized DFL loss. Unlike traditional methods that rely solely on historical data and remain agnostic on the downstream control policy, our approach benefits of more informed sampling of the input space u. This allows the state-transition model to allocate its modeling resources more effectively. Consequently, input regions that are critical to the control policy are modeled with greater precision, enabling even simple models to achieve high accuracy.

Interestingly, the regret $\mathcal{L}_r$ and the supervized DFL losses $\mathcal{L}_{\text{DFL}}$ do not pursue the same objective. On the one hand, the regret $\mathcal{L}_r$ trains the parameter $\hat{\theta}$ to bring the obtained objective value as close as possible to the objective value obtained with perfect knowledge of θ. On the other hand, the supervized DFL loss aims at improving the ability of the state transition proxy model $\hat{f}$ to match the true outcome. Notably, all performance losses would be the same if $\hat{f}$ is an error-free approximation of f.

3.4.4. Robustness to infeasibility

The updates of the constraint parameters by the SGD may lead the optimization program to become infeasible. A small number of infeasible optimization problems over a full training epoch may not be an issue if the remaining solvable samples update the parameters in a way which restores feasibility. However, such an approach lacks robustness. Therefore, we relax the constraints bounding the output of the system state model and formulate them as a quadratic regularization term in the objective function. To that end, the state model output is associated with a target value X^{g}. The target value is the desired state for the system and is chosen by the control policy designer. For example, the target value for the indoor temperature might be set to 21°C (70°F). A weight w defines the importance of meeting the target value. In addition to preventing infeasibility, the quadratic regularization term can transform linear programs into strongly convex quadratic programs, which, unlike linear programs, provide a non-null (and thus informative) gradient:

$$\begin{aligned} \phi(x) = \underset{u}{\operatorname{argmin}}\, & f_o(x,u) + w * (x - X^{\text{g}})^2 \\ \text{s.t. } & f_i(u) \leq 0 \quad i = 1, ..., k \ \forall t, \\ & h_i(u) = 0 \quad i = 1, ..., l \ \forall t, \\ & \hat{x}_{t+1} = \hat{f}(\hat{x}_t, u_t; \theta), \ \forall t. \end{aligned} \tag{3.8}$$

In addition, we propose a pre-training stage where the dynamics model is trained on historical data. It provides the policy with an initial estimate of the dynamics model parameters, which improves feasibility in the early stages of training. However, minimizing a task-agnostic loss on historical data may lead the pre-training to converge to a local minimum of the supervised DFL loss. Escaping this local minimum might require using a large step size in the gradient descent algorithm, which could harm training convergence. To address this, we apply calibrated Gaussian noise to the pre-trained parameters. The intensity of the noise must be carefully selected to retain as much pre-training information as possible while helping the model escape the local minimum.

3.5. Building Thermal Modeling and Day-ahead Control

We demonstrate the effectiveness of our method on the hourly day-ahead HVAC scheduling of a three-floor medium office building with 15 conditioned zones. We aim at learning the parameters of a multi-zone RC formulation modeling the building thermodynamics, while optimizing HVAC scheduling. In this section, we start by describing the building. The formulation of the optimization problem follows. Then, the data and their preprocessing are presented. Afterwards, we report and analyze the results of the proposed approach before comparing it to the conventional two-stage ITO performances. Finally, we assess the robustness of the model to distribution shift of the input data.

3.5.1. Building Description

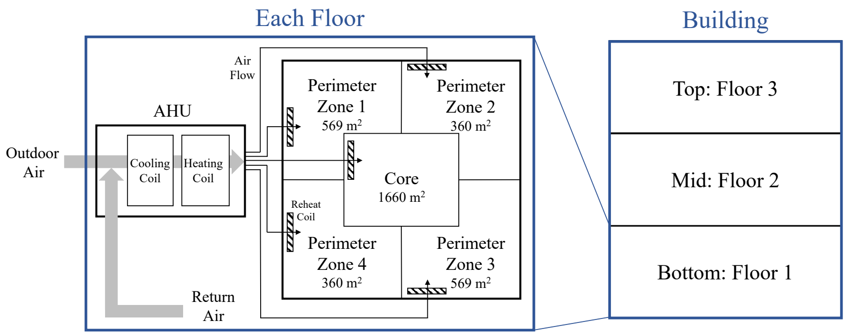


Figure 3.2.: The building is made of three similar floors. Each floor has its own variable volume Air Handling Unit (AHU) responsible for the thermal comfort of the five zones (four perimeter zones, and one core zone).

The building model is provided by the U.S. Department of Energy as part of EnergyPlus [17]. EnergyPlus is a state-of-the-art physics-based high-fidelity simulator for buildings. The building is an office building of 4928 m^2 (53,628 ft^2) assumed to be located in Denver, CO. It comprises three floors of six zones each, only five being thermally controlled. The sixth zone is a plenum that contains the ducts transporting the conditioned air. Each floor is equipped with an AHU. An AHU is an equipment that centrally conditions the air (either cooling or heating) before distribution to the zones via the duct network. The AHU is also responsible for the ventilation of the zones. The system in this building is a variable air volume system with reheat. The air flow rate reaching each zone can be modulated individually and, if necessary, the air can be heated just before entering the zone. Figure 3.2 provides a layout of the building floor and the AHU architecture. We make two assumptions with respect to the original template:

1. we replace the gas-fired heating coil of the AHU with an efficiency of 81 % by an electrical coil with an efficiency of 100 %;
2. we divide the AHU electricity consumption proportionally to the air mass flow rate of each zone served by the AHU, while the reheat coil electricity consumption is associated with the zone it supplies.

The building features a conventional occupancy schedule for offices; most of the activity takes place during weekdays from 7am to 6pm.

3.5.2. Control Formulation

The control problem aims to find the temperature profile for each zone $\tau^{\text{in}}_{t,z}$ that minimizes the electricity cost of HVAC for both heating and cooling while ensuring thermal comfort (3.9). The set of zones and time steps are $\mathcal{Z}$ (with cardinality Z, index z) and $\mathcal{T}$ (with cardinality T, index t), respectively. The electricity price consists of two terms. The first term (i) is the demand charge. The demand charge is proportional to the daily power consumption peak p^{d} and cost λ^{d}. The demand charge reflects the cost of network infrastructure to supply the requested power. The second term (ii) is the cost for the energy consumed over the time step p^{i}_t at price λ^{i}_t following the time-of-use tariff. Finally, the regularization term (iii) guarantees the problem feasibility, even during DFL. The regularization term penalizes the deviation of the indoor temperature $\tau^{\text{in}}_{t,z}$ from the target $T^{\text{tgt}}_{t,z}$ with a quadratic term. The weight $w_{t,z}$ adjusts the penalty incurred for thermal discomfort (i.e., the indoor temperature deviation from the target value) through the zones and time, for example, based on the occupancy.

$$\min_{\tau^{\text{in}}} \quad \underbrace{p^{\text{d}}\lambda^{\text{d}}}_{(i)} + \sum_{t=0}^{T} \left(\underbrace{p^{\text{i}}_t\lambda^{\text{i}}_t}_{(ii)} + \underbrace{\sum_{z=0}^{Z} w_{t,z}(\tau^{\text{in}}_{t,z} - T^{\text{tgt}}_{t,z})^2}_{(iii)} \right) \Delta t \tag{3.9}$$

The RC model in (3.10) accounts for multi-zonal building dynamics, where matrix $\alpha \in \mathbb{R}^{Z \times Z}$ represents inter-zonal heat transfer, vectors $R \in \mathbb{R}^{Z}$ and

$C \in \mathbb{R}^Z$ are zonal lumped resistances and capacitances, and vectors $\eta^{\mathrm{h}} \in \mathbb{R}^Z$ and $\eta^{\mathrm{c}} \in \mathbb{R}^Z$ are heating and cooling efficiencies, which include thermal losses of the duct system. For the 15-zone building, the RC model features 285 parameters. The parameters $\theta = (\alpha, \eta^{\mathrm{c}}, \eta^{\mathrm{h}}, R, C)$ will be learned while scheduling the HVAC system.

$$\tau^{\mathrm{in}}_{t+1} = \Delta t \left(\alpha \tau^{\mathrm{in}}_t + \frac{\left(\eta^{\mathrm{h}} p^{\mathrm{h}}_t - \eta^{\mathrm{c}} p^{\mathrm{c}}_t\right)}{C} + \frac{\left(\tau^{\mathrm{amb}}_t - \tau^{\mathrm{in}}_t\right)}{RC} \right) \quad \forall t \tag{3.10}$$

The initial conditions are given by:

$$\tau^{\mathrm{in}}_{0,z} = T^{\mathrm{in}}_{0,z} \quad \forall z. \tag{3.11}$$

For each zone, the HVAC capacities for cooling $\overline{P}^{\mathrm{c}}_{t,z}$ and heating $\overline{P}^{\mathrm{h}}_{t,z}$ are determined based on historical data. The constraints are imposed at the zonal (3.12) and floor (3.13) levels. The floor level constraint is essential to capture the maximum consumption of each AHU while the zonal constraint limits the amount of energy that can be dedicated to a specific zone. The set of floor is $\mathcal{F}$ with cardinality F, and index f. Each floor f is a set containing the zones z of that floor.

$$p^{\mathrm{c}}_{t,z} \leq \overline{P}^{\mathrm{c}}_{t,z},\ p^{\mathrm{h}}_{t,z} \leq \overline{P}^{\mathrm{h}}_{t,z} \quad \forall t, z \tag{3.12}$$

$$\sum_{z \in f} p^{\mathrm{c}}_{t,z} \leq \overline{P}^{\mathrm{c}}_{t,f},\ \sum_{z \in f} p^{\mathrm{h}}_{t,z} \leq \overline{P}^{\mathrm{h}}_{t,f} \quad \forall t, f \tag{3.13}$$

We define the HVAC electricity power as the sum of the cooling and heating electricity powers (3.14). Eq. (3.15) maintains the energy balance. In this reduced form, the HVAC electricity must be purchased from the grid.

$$p^{\mathrm{hvac}}_{t,z} = p^{\mathrm{h}}_{t,z} + p^{\mathrm{c}}_{t,z} \quad \forall t, z \tag{3.14}$$

$$\sum_{z=0}^{Z} p^{\mathrm{hvac}}_{t,z} = p^{\mathrm{i}}_t \quad \forall t \tag{3.15}$$

Constraint (3.16) defines the power peak demand as being greater than all power demand. Because the peak demand directly increases the electricity cost, this constraint is equivalent to $p^{\mathrm{d}} = \max \{p^{\mathrm{i}}_t : t \in T\}$, but it avoids making the problem bilevel.

$$p^{\mathrm{d}} \geq p^{\mathrm{i}}_t \quad \forall t \tag{3.16}$$

Finally, the power imported from the grid must be inferior to the line capacity $\overline{P}^{\mathrm{i}}$. This is equivalent to imposing the peak power demand to be inferior to $\overline{P}^{\mathrm{i}}$ (3.17). All power levels must be positive (3.18).

$$p^{\mathrm{d}} \leq \overline{P}^{\mathrm{i}} \tag{3.17}$$

$$p^{\mathrm{c}}_{t,z},\ p^{\mathrm{h}}_{t,z},\ p^{\mathrm{i}}_{t} \geq 0 \qquad \forall t, z \tag{3.18}$$

3.5.3. Datasets and Clustering

To simulate building heat transfers, two typical meteorological datasets are used. A typical meteorological dataset is a representative year of hourly weather data for a specific location, compiled from long-term observations. The first dataset is used as input to generate one year of historical data *without* any optimal scheduling. The second dataset provides weather scenarios for the scheduling optimization. Since optimizing and simulating the daily schedule for the 365 days of the year is burdensome, we perform a k-medoid clustering on the 365 days of weather data in the second dataset. The k-medoid algorithm is preferred to the k-mean algorithm to avoid the creation of fictitious average data. We focus on the ambient temperature because it is the only weather measurement necessary to the RC model (3.10). First, we identify three extreme days: the coldest, the hottest, and the day with the highest temperature variance. These three days are three fixed medoids of our medoid search. Then, we look for seven additional medoids to obtain the best partition of the space. The resulting temperature profiles of the k-medoid algorithm are given in Figure 3.3. The cluster means and standard deviations are reported in Figure 3.4.

3.5.4. Hierarchical Performance Loss

As seen in section 3.4.3, we can define any supervized DFL loss for learning the parameters $\theta = (\alpha, \eta^{\mathrm{c}}, \eta^{\mathrm{h}}, R, C)$ over the ten daily scenarios selected. In this specific case, the command is the indoor temperature profile for each zone $\tau^{\mathrm{in}}_{t,z}$ since it is the input of the thermostat. Therefore, the DFL loss focuses on the difference between the expected and the ex-post HVAC power consumptions. The underlying goal is to accurately capture the impact of HVAC power prediction errors on the objective value. Because the HVAC power p^{hvac} appears linearly in the objective function, we use a linear performance loss, the MAE. Furthermore, we weight the loss at each time step with the all-inclusive price coefficient $\lambda^{\mathrm{i,d}}_t$ that adds the peak demand price λ^{d}_t to the energy prices $\lambda^{\mathrm{i}}_t$for the time step with the highest ex-post power. Otherwise, $\lambda^{\mathrm{i,d}}_t$ and λ^{i}_t are equal.

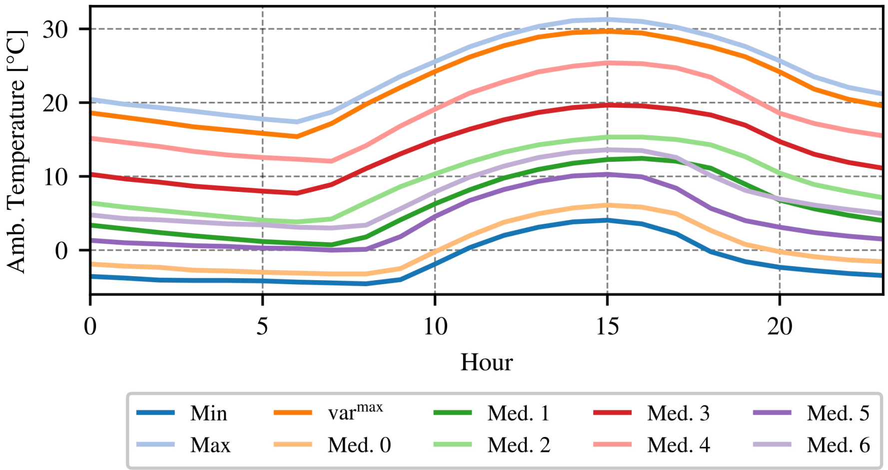


Figure 3.3.: Ambient temperature profiles of the ten medoids. The three first labels are extreme scenarios.

Finally, we leverage hierarchical loss. By order of importance, we minimize the error for (i) the whole building to obtain an accurate expectation of the electricity bill; (ii) each AHU which means for the sum of the five zones at each floor f; (iii) each zone z.
The final loss function is given by (3.19). The number of zones at each loss level is used to weight the importance. The building encompasses 15 zones ($w_b = 15$) while each floor features five zones ($w_f = 5\ \forall f$).

$$
\begin{aligned}
L_{\text{DFL}} =& \frac{1}{T}\sum_{t=0}^{T} \lambda_t^{\text{i,d}} \Bigg(w_b \cdot \text{MAE}\left(\hat{p}_{t,b}^{\text{hvac}}, p_{t,b}^{\text{hvac}}\right) \\
&+ w_f \sum_{f=0}^{F} \text{MAE}\left(\hat{p}_{t,f}^{\text{hvac}}, p_{t,f}^{\text{hvac}}\right) \\
&+ \sum_{z=0}^{Z} \text{MAE}\left(\hat{p}_{t,z}^{\text{hvac}}, p_{t,z}^{\text{hvac}}\right) \Bigg)
\end{aligned} \tag{3.19}
$$

where $\hat{p}_{t,f}^{\text{hvac}} = \sum_{z \in f} p_{t,z}^{\text{hvac}}$ is the HVAC power of floor f and $\hat{p}_{t,b}^{\text{hvac}} = \sum_{z=0}^{Z} p_{t,z}^{\text{hvac}}$ is the HVAC power of the whole building.

Note that the proposed supervized DFL loss introduces an outcome feedback loop in our bilevel calibration model for our linear thermal model [105]. The optimizer uses current parameters θ to set optimal temperatures, which EnergyPlus simulates into real power use $p_{t,z}^{hvac}$. The prediction error then re-

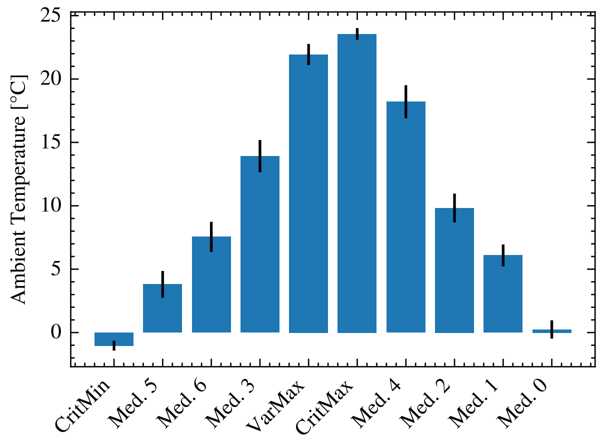


Figure 3.4.: Mean and standard deviation of the medoids. They are ordered to form a smooth cycle.

fines θ via L_{DFL}, looping back through outcomes. This risks measurement bias, as it tunes the model to its own optimized paths, potentially missing broader dynamics. Yet the feedback converges to an optimum outperforming model identification on historical data. This aligns with DFL, where model updates directly optimize decision quality rather than prediction accuracy alone. In practice, we approximate of the loss as

$$\nabla_\theta L_{\mathrm{DFL}}\big(\hat{p}^{\mathrm{hvac}}_{t,z}(\theta), p^{\mathrm{hvac}}_{t,z}(\tau^{\mathrm{in}}(\theta))\big) \approx \nabla_\theta L_{\mathrm{DFL}}\big(\hat{p}^{\mathrm{hvac}}_{t,z}(\theta), p^{\mathrm{hvac}}_{t,z}\big). \tag{3.20}$$

Nevertheless, a formal study of the approximation made by neglecting $\partial p^{hvac}_{t,z}/\partial\theta$ and the self-stabilization behavior under operational policies would strengthen the method.

3.5.5. Results

Our goal is to learn the parameters $\theta = (\alpha, \eta^{\mathrm{c}}, \eta^{\mathrm{h}}, R, C)$ that define the building heat transfers (3.10). We initialize the algorithm by the pre-training. During this warm-start period, the parameters are trained over the 365 days of the first typical weather file using an MSE loss. Then, an element-wise Gaussian noise $N \sim \mathcal{N}(0, \theta_i/25)$, which corresponds to a Signal-to-Noise Ratio (SNR) of 625 (i.e., 25^2), is applied independently on each element of the parameter vector. This operation aims at getting the RC model out of the pre-training local minimum while retaining a maximum of information.

Afterwards, we enter the DFL stage. The DFL training set is made of ten representative days obtained via clustering. In addition, we sample randomly

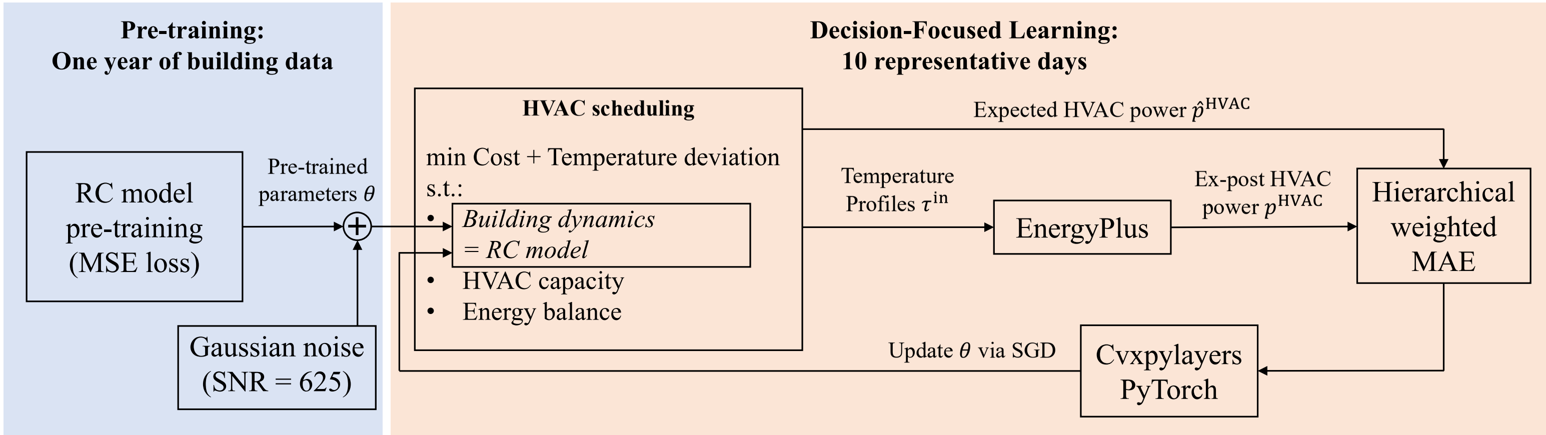


Figure 3.5.: The building RC model is pre-trained on one year of historical data before a small noise is applied on the parameters. They are then updated to perform the best day-ahead HVAC scheduling over 10 representative days with respect to the task-aware MAE loss.

in each cluster two days to form a validation set and a test set of ten days. We consider solving and simulating the solution once for each of the ten training problems as a training epoch of ten samples. Each problem is solved using the *ECOS* solver since it was reported to be the best performing solver for conic DFL [94]. *Cvxpylayer* computes the gradient of the optimization problem and *PyTorch* performs the backpropagation. As soon as a sample (i.e., a day) has been solved and simulated, the parameters are updated (i.e., pure Stochastic Gradient Descent (SGD)). To facilitate learning, we order the days to form a smooth cycle when looping over the epochs as depicted by Figure 3.4. We initialize *Adam*, a gradient-based optimization algorithm, with a learning rate at 0.001 and a polynomial decay of $\gamma = 0.9$ [106]. We set the maximum number of epochs to 50 with an early stopping featuring a 10-epoch patience. Figure 3.5 represents the whole process.

The time-of-use tariff is a rectangular function similar to the two-rate tariff proposed by Walloon main DSO ORES until 2025 [107]. From 7pm to 6am (excluded) the price is 0.3 €/kWh. It rises to 0.6 €/kWh from 6am to 7pm (excluded) as depicted by Figure 3.6. The increase highlights the higher demand during the day.

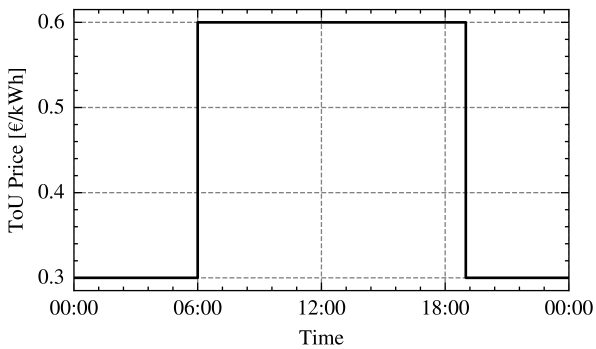


Figure 3.6.: Time-of-use tariff with an appreciation of electricity from 6am to 7pm.

The performances of the DFL training are presented in Figure 3.7. Overall, the training and validation metrics follow very similar trends. This suggests that ten medoids are sufficient for the model to generalize effectively. The first plot displays the evolution of the loss over the training epochs. The training is interrupted after 44 epochs by the early stopping because the minimum value for the validation loss is reached at the 34$^{\text{th}}$ epoch. Both the training and validation losses tend to decrease until the 34$^{\text{th}}$ epoch before leveling off. At epoch 34, the training and validation losses stand at 252 and 248, respectively.

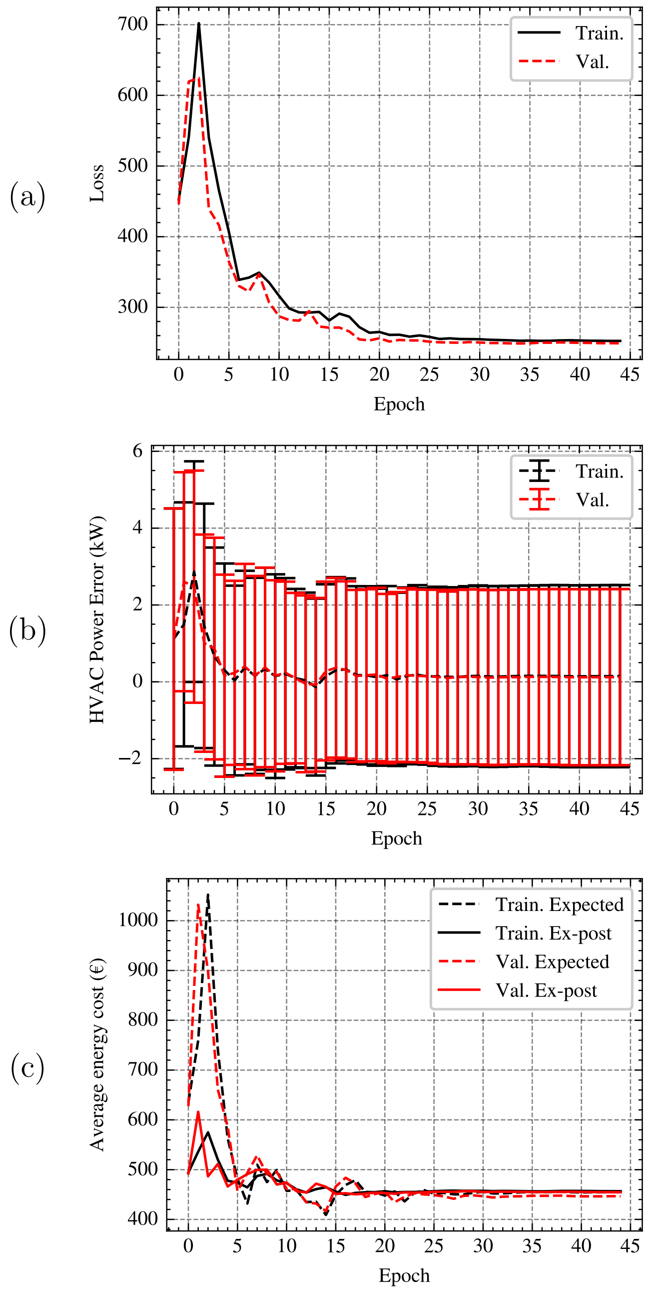
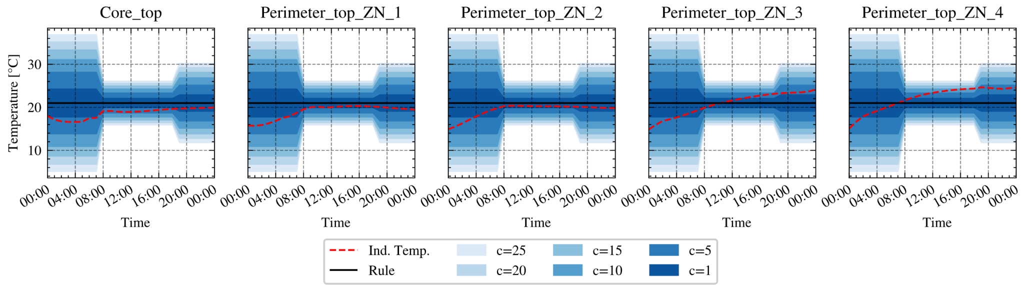


Figure 3.7.: (a) The DFL supervised loss (hierarchical weighted MAE) converges at much lower levels than initially; (b) the error mean converges towards zero and the standard deviation reduces; (c) the expected cost becomes a more accurate prediction of the ex-post cost while diminishing.

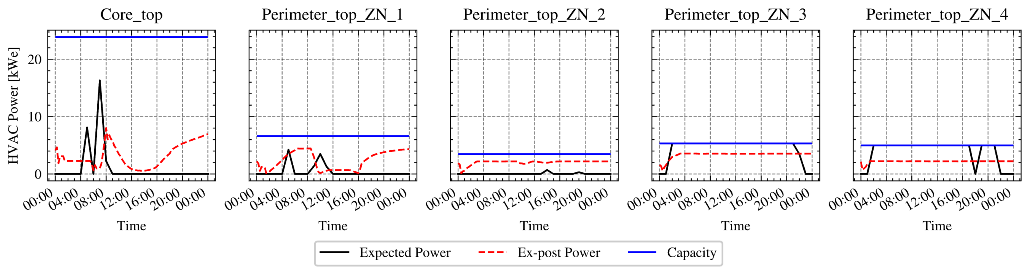
The center plot (figure 3.7.b) shows the evolution of the error, assuming a Gaussian distribution. After about five epochs, the error mean converges around 0 with a standard deviation at about 2.5 kW. Then, the error mean ripples slightly until the 25th epoch where it stabilizes at 0.15 with a standard deviation of 2.35. The training and validation sets exhibit very similar trends.
The last plot displays the evolution of the expected and ex-post electricity costs for the training and validation sets. The expected and ex-post costs are very volatile over the first half of the training. Over the second half, the expected and ex-post costs level off at about 450 € and 455 €, respectively. The bill prediction of the day-ahead planning is therefore accurate. It is worth noting that the training converges at the minimum ex-post cost over all epochs.
In conclusion, when surveying simultaneously the three metrics, the set of parameters obtained at the end of the 34th epoch offers the highest performance.

In comparison, the ITO model performs extremely poorly. The RC parameters of the ITO model were found by minimization of the MSE. Nevertheless, the MSE reported in Table 3.2 indicate clearly that the DFL model at the 34th epoch is better for each data set: training, validation, and test. The two models show very consistent metrics across the three data sets.
The hierarchical loss of the ITO test set stands at 652, more than two and a half times the value of the DFL model (253). However, the MAE is slightly lower for the ITO model, rising from 2.66 for the ITO model to 2.94 for the DFL model. This can be explained by the hierarchical loss that does not aim at minimizing the prediction error at the zonal level. The test set average error of the ITO model stands at -1.81 kW with a standard deviation of 2.39 kW. Since, on average, the expected power is much lower than the ex-post power, the electricity cost is underestimated. The expected cost is 85 € for an ex-post cost at 474 €. It represents a prediction error of 389 €. In contrast, the DFL model achieve a prediction error of 16 €. Very interestingly, the ex-post cost of the DFL model (468 €) is lower than for the ITO model (474 €). The foremost hindrance to DFL is the heavy computational burden associated with training. On a personal MacBook Pro, equipped with an Apple M1 Pro chip and 32 GB of memory, the DFL training took 7 hours compared to 16 minutes for the conventional supervized learning.

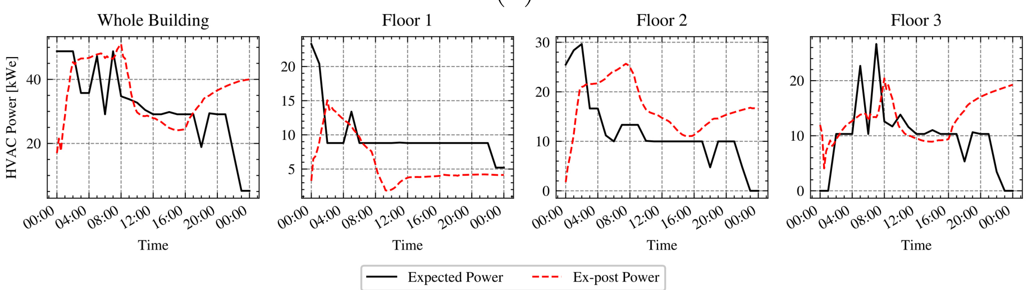
The scheduling of the building lower floor on the coldest day is shown in Figure 3.8. The upper part displays the temperature profiles within the five zones. The blue shades correspond to a given penalty per hour for deviating from the 21°C target. The discomfort penalty is the lowest during the night, from midnight to 7am, and the highest during the conventional working hours, reflecting the higher occupancy.
Overall, and aligned with the discomfort penalty, the temperature profiles stay



(a)



(b)



(c)

Figure 3.8.: Scheduling for the coldest day obtained by the DFL RC model at epoch 34. Subfigures are (a) the zonal indoor temperature of the top floor; (b) the zonal HVAC electric power of the top floor, and (c) the aggregated HVAC power for the whole building and each floor. The blue shades indicate the temperature deviations causing an objective value penalty lower than a given threshold. Temperature control is more loose during the night and evening because of the worker absence. The mismatch between the expected and ex-post HVAC powers in (b) can be explained by the intrinsic RC model limitations and the task-aware loss which gives little importance to accurate zonal predictions.

	ITO			DFL		
	Train.	Val.	Test	Train.	Val.	Test
Hierarchical loss	655	646	652	252	248	253
MAE (kW)	2.68	2.62	2.66	2.93	2.87	2.94
MSE (kW^2)	16.8	16.6	16.7	15.4	14.6	15.6
Error mean (kW)	-1.85	-1.86	-1.81	0.15	0.12	0.08
Error std (kW)	2.34	2.28	2.39	2.35	2.27	2.38
Expected cost (€)	82	75	85	454	446	452
Ex-post cost (€)	478	474	474	456	455	468
Cost error (€)	395	398	389	2	8	16
Training Time	16 min			7h02min		

Table 3.2.: Performances of the ITO model versus the DFL model at epoch 34.

close to the target of 21°C during the day, with greater deviations occurring in the evening and night. Still the deviations remain within admissible ranges. The electric power scheduling of the HVAC is shown in the lower two-thirds figure. The zonal electric power scheduling in subfigure 3.8.b is quite inaccurate. This can be explained by the hierarchical loss, which focuses primarily on having a precise day-ahead scheduling of the whole building. This is depicted by subfigure 3.8.c where the power trends are much better for the floor level and the entire building. At the building level, the attempt to shift the HVAC consumption towards cheap electricity stands out. Indeed, the HVAC consumption is high during the night, from 2am to 8am, and peaks again towards the end of the day, when the price of electricity goes down. The RC model is a linear approximation of the dynamics considering only the previous time step. Moreover, the outdoor temperature is the only weather input. Consequently, the limited modeling capability and input of the RC model, along with the complexity of the building's heat transfers, are responsible for the remaining inaccuracies. The real-time adjustment to the model inaccuracies and the changing conditions should be handled by a real-time controller.

3.5.6. Ambient Temperature Shift

We investigate the robustness of the proposed framework to the input distribution shift. In particular, we assess the model response to a hot year. We build the hot year by selecting the hottest sample from each cluster. For the sample associated with the hottest cluster, we add two Celsius degrees on top of the hottest day, assuming record-breaking temperatures. Figure 3.9 com-

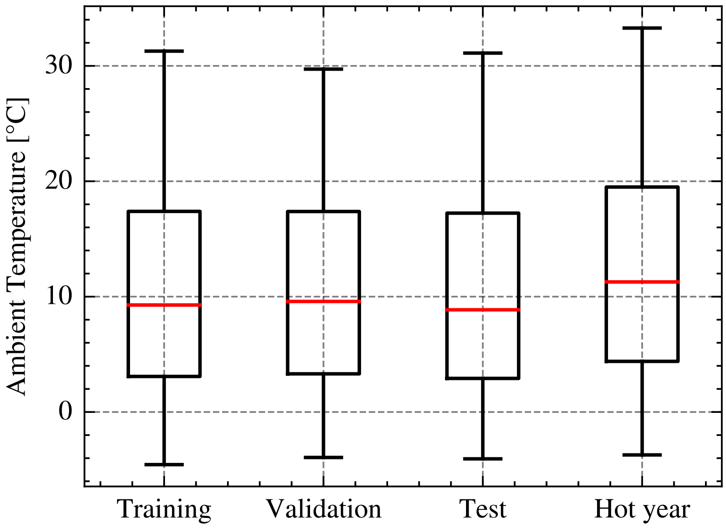


Figure 3.9.: Ambient temperature distribution of each data set. Whiskers represent minimum and maximum temperatures.

pares the ambient temperature distribution of the four data sets. The average and maximum ambient temperatures in the hot year set are higher than in any other set. The average is 2°C higher for the hot-year set than for the training set, thus reflecting a distribution shift.

Table 3.3 shows the metrics of the ITO and DFL models over the hot year data set. The results show a slight performance degradation for both models compared to the randomly sampled test set. Performance degradation is expected because the hot year data distribution is purposely different from the previous distribution.
The hierarchical loss stands at 685 (+5.1% compared to the test set) and 271 (+7.1%) for the ITO and DFL models, respectively. The error standard deviation remains virtually unchanged. The mean, however, worsens for the ITO model but improves for the DFL model. This results in a greater cost error for the ITO but not for the DFL.
Overall, DFL outperforms ITO and is less affected by input distribution shift than the ITO model.

3.6. Discussion

The proposed DFL strategy outperforms the ITO approach. The naive ITO framework that consists in MSE minimization over historical data leads to

	ITO	DFL
Hierarchical loss	685	271
MAE (kW)	2.76	2.95
MSE (kW^2)	18.3	15.8
Error mean (kW)	-1.96	0.1
Error std (kW)	2.39	2.39
Expected cost (€)	79	467
Ex-post cost (€)	500	482
Cost error (€)	420	15

Table 3.3.: Metrics of the ITO model and the DFL model over a dataset representing a distribution shift towards higher ambient temperature (i.e., hot year).

extremely poor dynamic model for day-ahead scheduling. In particular, the RC model is unable to appropriately fit the operating areas relevant for the control policy. This leads to potentially severe electric power scheduling underestimation resulting in unexpectedly high ex-post bills. In the case study, the ex-post cost was on average 6 times higher than the expected cost for the ITO model. This raises important questions about the use of RC model for the day-ahead HVAC scheduling, at least within an ITO framework.

To address potential concerns about baseline selection, we note that the ITO approach with RC modeling represents a standard supervised learning baseline for data-driven control. Yet, it systematically fails to capture tariff-aware operating regimes. Other benchmarks, such as the sophisticated rule-based control per ASHRAE Guideline 36, slightly underperform optimization-based control without even considering variable electricity prices [108]. It is therefore unlikely that such a benchmark will offer a better comparison.

In contrast, the proposed DFL approach significantly improves both the accuracy of the bill prediction and the cost reduction. All error metrics show improvement, except for the MAE. However, the DFL approach is much more computationally demanding. The DFL duration before triggering the early stopping is about 7 hours compared to 16 minutes for the ITO approach. This is due to the DFL need to solve, for each sample, the day-ahead HVAC scheduling optimization problem and simulating the results. Furthermore, software for DFL are not as mature as the ones for supervized learning. Future software development for DFL will likely harvest the power of parallel computing and new hardware leading to significant time reduction. In this specific case, due to the linear formulation of the optimization, the computational burden mainly lies with the simulation. We recommend considering using a reduced-order physics model of the simulator in the initial training steps. Since training is performed once and offline, longer time is not prohibitive.

A practical limitation of the proposed approach is the accessibility of high-quality training data in real-world applications. In this study, we relied on the high-fidelity EnergyPlus simulator, which provides detailed and consistent trajectories of building dynamics, control actions, and energy costs. Such data are extremely valuable because they enable systematic exploration of the operating space and generation of large datasets at low cost once the model is available. However, in operational settings, similarly rich data can be difficult to obtain. Even though our data usage remained sparse (ambient temperature, indoor temperature of each zone, cooling and heating consumption of each HVAC unit, and air volume injected in each zone), many sensors are necessary and hindrances are frequent (e.g., sensor outages, incomplete metering, limited historical records, or confidentiality constraints on building and tariff information). This suggests that, while the proposed framework works effectively for simulation-based environment and offline training, its deployment in practice will likely benefit from hybrid strategies combining measured data, calibrated digital twins, transfer learning, or reduced-order surrogate models to mitigate data scarcity.

Finally, we showed that the model is robust to an input data shift towards higher temperature, a proxy for climate change impacts on building loads. Regarding robustness to price signals, the current case study employs a fixed time-of-use tariff, which the DFL policy captures effectively. However, under fully dynamic pricing (e.g., day-ahead market prices or 5-min locational marginal prices), retraining or online adaptation may be required, as the policy is implicitly tuned to the training tariff structure. This could be tested by using dynamic price trajectories drawn from historical day-ahead/intraday markets, evaluating ex-post cost degradation relative to ITO and rule-based models.

3.7. Conclusion

This chapter presents a new framework that enables simultaneous identification and control (or planning) of a complex system. This is of particular interest for black-box systems in which dynamics change radically once the control policy is applied, thus making the historical observation little informative. The method exploits recent advances in decision-focused learning, and more specifically, in the calculation of the gradient of cone programs. The gradient of the convex optimization control policy output (i.e., the command) with respect to the policy parameters (e.g., the system dynamics model parameters) can be computed. Because a key constraint of the control policy is being learned, the feasibility domain evolves at each gradient descent step. We handle the potential infeasibility by relaxing the constraints on the system states and pre-learning the uncertain parameters on historical data. Furthermore, we propose a new type of loss function that bypasses the need for the system to be differentiable. Not only is this necessary for black-box systems, but also for actual physical systems that cannot be easily modeled.

We apply the proposed approach to the day-ahead HVAC scheduling of a 15-zone office located in Denver, CO. This case study showcases the efficient learning provided by our framework, and the ability to design complex task-aware loss functions to reflect the impact of the parameter misestimation on the objective value. The results show unambiguously the added value of simultaneous system identification and planning by displaying a lower ex-post cost along with a more accurate price forecast and a reduced error on the day-ahead HVAC scheduling.

However, to keep the scheduling problem convex and thus differentiable, the thermal model must be linear. The linear approximation limits the accuracy of the model and the applications. Therefore, in the next chapter, we leverage piecewise linear approximation to provide a more detailed model for building dynamics. Such piecewise models turn the optimization formulation into a MIP, which brings new challenges in the computation of the gradient.

CHAPTER 4.

Decision-Focused Learning for Non-Convex Control and Nonlinear Modeling of Building Thermal Dynamics

In chapter 3, we developed a DFL framework to simultaneously identify and control complex systems. The main shortcoming of the method is the necessity to have a convex control policy that enables differentiation of the optimization problem. Consequently, thermal models were limited to linear approximation since they appear as an equality constraint of the control problem. In this chapter, we extend the applicability of the idea to nonlinear thermal model which requires to obtain a gradient flow through a non-convex optimization program. We start with a motivation for the work in this chapter followed by a brief review of the literature and the contributions (section 4.1). In section 4.2, we describe the HVAC MS model, including the reformulation of NN as a set of mixed-integer linear constraints, and the derivation of the resulting MILP problem. In section 4.3, we introduce the case study and evaluate the performance of thermodynamic models of increasing complexity, from the RC linear model to the NN model. We compare the effectiveness of training these models in ITO and DFL fashions. In addition, we analyze the impact of the proposed tight formulation of ReLU, the standard deviation of the noise, and the number of samples on our DFL strategy. Finally, we summarize the key findings and outline the potential directions for future research in section 4.4.

4.1. Introduction

A major opportunity to leverage HVAC flexibility is the participation in the day-ahead electricity market, the main platform for trading electricity between grid actors. In Europe, this market is cleared daily, setting electricity volumes and prices for each quarter of hour of the following day since September 2025. This timeline aligns naturally with the slow thermal dynamics of buildings, where temperature adjustments from HVAC decisions and external conditions unfold gradually over time. Capturing these effects is crucial to maintain occupants' comfort [109].

Based on our willingness to account for nonlinear dynamics and the review conducted in section 3.2, we decide to focus on NNs. NNs can capture complex patterns with a limited manual modeling effort, and will thus be used in this chapter. To fully exploit the capabilities of NN models for HVAC control, it is essential to unify modeling and control tasks within a single framework. Such an integrated approach would allow HVAC systems to make energy-efficient planning decisions while maintaining occupants' comfort.

A first promising avenue is to model the control policy by a NN that outputs control decisions, allowing it to internalize the building dynamics. Traditionally, such NNs in real-time control are trained using RL [81], which directly optimizes decisions through interaction with the environment. Alternatively, self-supervised learning can be employed, but, similarly to RL, it often lacks guarantees on solution quality, especially in the presence of hard physical and operational constraints. For example, the method in [110] ensures feasibility but not optimality and cannot be applied to discrete problems, while extensions to mixed-integer nonlinear programming fail to guarantee either feasibility or optimality [111]. Self-supervised methods such as Primal-Dual Learning [112] and deep Lagrangian dual networks [113] improve feasibility by integrating dual information of the underlying optimization. However, Primal-Dual Learning relies on iterative dual updates and may struggle with nonconvex or stochastic problems, whereas the approach in [113], which predicts solutions to the AC-OPF problem, is designed around problem-specific structures, which prevents generalization to settings with uncertainty or combinatorial decision variables.

A second promising avenue is provided by NN-constrained optimization, where the learned thermal model, represented by a NN, is embedded directly into mathematical optimization as a set of constraints. This approach merges the expressiveness of NN with the stability and decision quality of constrained optimization. It has been shown that, by using a feedforward NN with ReLU, the constraints can be exactly encoded as a set of Mixed-Integer Linear (Mixed-Integer Linear (MIL)) equations [27].

In this chapter, we formulate the day-ahead HVAC Management System (MS) as an NN-constrained optimization problem. The NN, which models the building's thermal dynamics, is encoded precisely as a set of mixed-integer linear constraints within the optimization problem, enabling the optimization model to incorporate learned dynamics while ensuring decision quality and feasibility.

Scope Clarification

There are two main changes with respect to chapter 3:

- the thermal model must capture the nonlinear dynamics of the building and its HVAC system, no more linear approximation;
- the resulting scheduling problem is not convex which prevents differentiability by existing tools.

4.1.1. Literature Review

Although NNs offer strong modeling capabilities, their inherent approximation errors can propagate into suboptimal energy management decisions. Traditional ML models are trained to minimize statistical metrics such as MSE, thus overlooking how the learned model affects downstream control decisions. In contrast, DFL aligns ML training with optimization objectives to improve decision quality [114]. DFL typically relies on gradient-based training methods. The core challenge in this approach is two-fold: (i) computing the sensitivity of the optimization problem's solution with respect to its input parameters, and (ii) ensuring that the resulting gradients are meaningful and informative for learning [11].

In this chapter, to extend DFL to the scheduling of buildings with NN modeling of the thermal dynamics, we need to differentiate through a MIP, which is particularly difficult due to the presence of (discrete) integer decision variables. As a result, the gradient of the MIP solution with respect to the MIP parameters is typically either zero or undefined, making it incompatible with standard gradient-based learning methods. One possible approach is to replace the MIP with its continuous relaxation, allowing the use of the existing DFL methods such as implicit differentiation of the KKT conditions used in chapter 3 [93], [115]. While relaxing a MIP to its continuous counterpart reduces computational complexity, it alters the NN representation, i.e., from capturing a nonlinear hypersurface to approximating it as a polytope, compromising modeling fidelity. Another approach generates cuts to obtain a linear program that has the same solution as the original mixed-integer formulation [116]. However, in addition to the implementation difficulty, this method is extremely burdensome, since a new set of cuts must be generated for each training instance.

To tackle these challenges and thus enable DFL for the NN-constrained HVAC planning problem, we propose to leverage Stochastic Smoothing (SS). This technique introduces random perturbations to the uncertain parameters. Early approaches use this perturbation to smooth the "max" operator limiting its applicability to problems in which the unknown parameters are in the linear objective function [117]. We go further by extracting informative gradients from the task-specific loss using the score function of the REINFORCE algorithm [19]. Unlike the aforementioned methods, SS does not make any assumption about the optimization problem class, the position of the uncertain parameters (whether in the objective function and/or the constraints), and the specific form of the loss function. Therefore, SS obtains a gradient from ex-post signals directly such as the true realized cost or constraint deviation, without requiring differentiability. SS ensures high versatility of the proposed framework and is highly suitable for combinatorial optimization involving a non-differentiable dynamical system.

4.1.2. Contributions

The main contributions of this chapter can be summarized as:

1. We train a NN that models the building's thermal dynamics using DFL to optimize the decision quality rather than minimize a task-agnostic statistical metric. We formulate an HVAC day-ahead management as a MILP in which the building's thermal dynamics are modeled by a NN and embedded directly as optimization constraints. To enable effective and robust gradient-based training of the NN, we (*i*) reformulate the MILP problem as an MIQP to avoid infeasibility during learning, (*ii*) solve the MIQP with random perturbations applied to unknown NN parameters to encourage exploration, (*iii*) estimate gradients using a score-function method that directly relates learning with decision quality.

2. We embed a piecewise linear NN in the HVAC management system to obtain an expressive and tractable nonlinear model of the building's thermal dynamics. We reformulate each ReLU activation as a set of MIL constraints, providing solution quality guarantees via a MIP optimality gap. This approach yields a trade-off between modeling accuracy of the dynamics and computational efficiency.

3. Because reformulating NNs with ReLU activations in MILP requires Big-M constants, we improve the standard formulation by adaptively tightening them. Specifically, instead of relying on fixed Big-M bounds, we

dynamically adjust the feasible input intervals: when an input value is known in advance (i.e., it is a deterministic parameter), its interval is reduced to a single value. This adaptive strategy produces a tighter formulation than existing fixed Big-M methods [118], [119], leading to more efficient optimization without sacrificing correctness. The benefit of this approach is demonstrated through both theoretical analysis (section 4.2.4) and empirical validation (section 4.3.5).

Our framework bridges the gap between ML, constrained optimization, and real-world energy management, paving the way for more intelligent and efficient HVAC control strategies. We show the effectiveness of our approach on a realistic five-zone office building located in Denver, USA. Moreover, we compare the performance of multiple thermal models: various NN architectures and lumped RC model trained on historical data, or via traditional DFL methods.

4.2. Model and Methods

4.2.1. HVAC Management System

The objective of day-ahead HVAC MS is to minimize the following day's operational cost on behalf of the building manager. This scheduling task is challenging, as it must account for nonlinear building thermodynamics, efficient energy use, and sufficient thermal comfort for the occupants. In general, the easiest way to control the HVAC system is through thermostats, which are widely available in buildings. Therefore, the decision variable of the scheduling problem is the matrix of indoor temperature setpoints for each time step and thermal zone[1]. In the following model to optimize the day-ahead HVAC operation, we assume that the electricity cost λ_t^{i} is lower than the feed-in tariff λ_t^{e} for each time step.
The objective function (4.1) of the day-ahead HVAC MS is:

$$\min_{\tau^{\mathrm{in}}} \quad \underbrace{p^{\mathrm{d}}\lambda^{\mathrm{d}}}_{(i)} + \sum_{t=0}^{T} \left(\underbrace{p_t^{\mathrm{i}}\lambda_t^{\mathrm{i}} - p_t^{\mathrm{e}}\lambda_t^{\mathrm{e}}}_{(ii)} \right) \Delta t \tag{4.1}$$

where we minimize:

[1]A thermal zone in a building is a space or collection of spaces with similar space-conditioning requirements, typically sharing the same heating and cooling setpoint and controlled by a single thermostat or thermal control device.

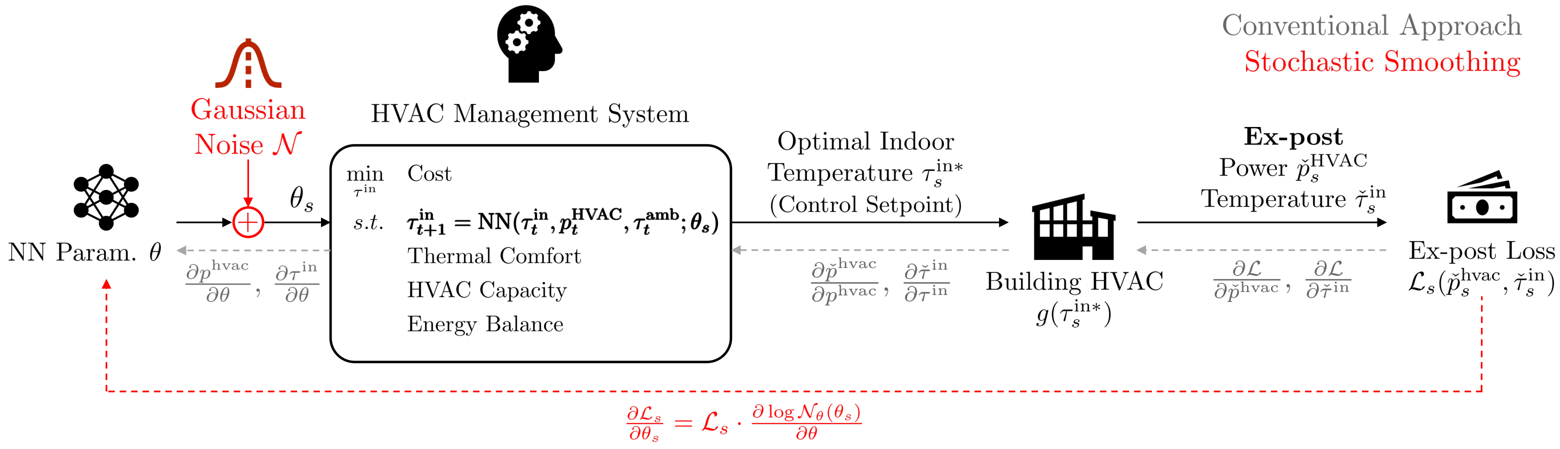


Figure 4.1.: Comparison of the proposed stochastic smoothing DFL pipeline (red) for HVAC management system with conventional DFL approaches (gray).

(*i*) the peak demand cost by applying a charge λ^{d} to the daily power consumption peak p^{d};

(*ii*) the electricity cost aggregated at the building level which is, at each time step, the difference between the cost of consuming grid electricity $\lambda_t^{\mathrm{i}} p_t^{\mathrm{i}}$ and the revenue from power injection into the grid $\lambda_t^{\mathrm{e}} p_t^{\mathrm{e}}$.

The optimization is constrained in (4.2) by the building thermal dynamics from time steps t to $t+1$. The relationship is modeled by a NN with parameters θ. The vector of zonal indoor temperatures τ_{t+1}^{in} is obtained based on the input from the previous time step t: the vector of zonal indoor temperatures $\tau_t^{\mathrm{in}} \in \mathbb{R}^Z$ where Z is the number of zones, the vector of electrical power for zonal heating $p_t^{\mathrm{h}} \in \mathbb{R}^Z$, the vector of electrical power for zonal cooling $p_t^{\mathrm{c}} \in \mathbb{R}^Z$, and ambient temperature $\tau_t^{\mathrm{amb}} \in \mathbb{R}$.

$$\tau_{t+1}^{\mathrm{in}} = \mathrm{NN}(\tau_t^{\mathrm{in}}, p_t^{\mathrm{h}}, p_t^{\mathrm{c}}, \tau_t^{\mathrm{amb}}; \theta) \quad \forall t \tag{4.2}$$

The initial state of the building, i.e., the set of initial zonal indoor temperatures, are given for all zones $z \in \mathcal{Z}$ by:

$$\tau_{0,z}^{\mathrm{in}} = T_{0,z}^{\mathrm{in}}. \tag{4.3}$$

In addition, zonal indoor temperature $\tau_{t,z}^{\mathrm{in}}$ is restricted to lie within comfort limits $\underline{T}_{t,z}^{\mathrm{in}}$ and $\overline{T}_{t,z}^{\mathrm{in}}$.

$$\underline{T}_{t,z}^{\mathrm{in}} \leq \tau_{t,z}^{\mathrm{in}} \leq \overline{T}_{t,z}^{\mathrm{in}} \tag{4.4}$$

Heating and cooling power consumptions are positive variables bounded by the HVAC capacities $\overline{P}_z^{\mathrm{c}}$ and $\overline{P}_z^{\mathrm{h}}$:

$$0 \leq p_{t,z}^{\mathrm{c}} \leq \overline{P}_z^{\mathrm{c}}, \quad \forall t, z; \tag{4.5}$$

$$0 \leq p_{t,z}^{\mathrm{h}} \leq \overline{P}_z^{\mathrm{h}}, \quad \forall t, z. \tag{4.6}$$

We define the zonal HVAC electrical power $p_{t,z}^{\mathrm{hvac}}$ as the sum of the powers for cooling and heating (4.7). The resulting energy balance is ensured by (4.8), in which the net power consumption of the building comprises the HVAC power consumption $p_{t,z}^{\mathrm{hvac}}$ and the non-dispatchable load P_t^{nd}, minus the building's on-site power generation P_t^{gen}, e.g., from rooftop photovoltaic panels.

$$p_{t,z}^{\mathrm{hvac}} = p_{t,z}^{\mathrm{h}} + p_{t,z}^{\mathrm{c}} \quad \forall t, z \tag{4.7}$$

$$p_t^{\mathrm{i}} - p_t^{\mathrm{e}} = \sum_{z=0}^{Z} p_{t,z}^{\mathrm{hvac}} + P_t^{\mathrm{nd}} - P_t^{\mathrm{gen}} \quad \forall t \tag{4.8}$$

The peak power demand p^{d} is the maximum power consumption (or injection) from the grid over the day. The peak demand is greater than or equal to all power exchanges $p_t^{\mathrm{i}} + p_t^{\mathrm{e}}$ across the day (4.9). While conceptually equivalent to defining $p^{\mathrm{d}} = \max\{p_t^{\mathrm{i}}, p_t^{\mathrm{e}} \mid t \in \mathcal{T}\}$, constraint (4.9) avoids introducing a bilevel structure into the optimization problem, preserving tractability.

$$p^{\mathrm{d}} \geq p_t^{\mathrm{i}} + p_t^{\mathrm{e}} \quad \forall t \tag{4.9}$$

At all times, power imported from or exported to the grid must comply with the line capacity $\overline{P}^{\mathrm{l}}$ of the connection between the building and the grid:

$$p_t^{\mathrm{i}} \leq \overline{P}^{\mathrm{l}} \quad \forall t, \tag{4.10}$$

$$p_t^{\mathrm{e}} \leq \overline{P}^{\mathrm{l}} \quad \forall t. \tag{4.11}$$

Finally, all power imports and exports must be positive:

$$p_t^{\mathrm{i}},\ p_t^{\mathrm{e}} \geq 0 \qquad \forall t. \tag{4.12}$$

4.2.2. Decision-Focused Learning

Problem Formulation

Our objective is to learn the NN parameters θ (i.e., weights and biases) of constraint (4.2). Instead of training θ in a task-agnostic manner on historical data, we aim to learn θ to directly improve HVAC management decisions as shown by Figure 4.1.

To formalize this DFL problem, we introduce r, a vector containing all stochastic parameters of the HVAC MS (e.g., weather conditions) excluding θ. We assume r follows the known training distribution $\mathcal{R}$. The goal is to minimize the task loss $\mathcal{L}$, which reflects the true operational goal of the HVAC MS. This naturally leads to a bilevel optimization problem, whose solution yields the optimal values of θ:

$$\min_{\theta} \mathbb{E}_{r\sim\mathcal{R}}[\mathcal{L}(\check{p}^{\text{hvac}}, \check{\tau}^{\text{in}})] \tag{4.13}$$

$$\text{s.t. } \check{p}^{\text{hvac}}, \check{\tau}^{\text{in}} = g(\tau^{\text{in}^*}) \tag{4.14}$$

$$\min f(p^{\text{hvac}}, \tau^{\text{in}}) \tag{4.15}$$

$$\text{s.t. } (4.2) - (4.12) \tag{4.16}$$

The inner problem (4.15)-(4.16) is the HVAC MS, including the NN formulation, which is thus a MILP. The outer problem aims to find θ by minimizing the expected task loss over the distribution of input parameters r (4.13). Ideally, the task loss is the ex-post value of the decisions. Thus, it depends on the realized ex-post measurements of the HVAC power $\check{p}^{\text{hvac}}$, and indoor temperature $\check{\tau}^{\text{in}}$. These ex-post measurements are obtained by controlling the HVAC system with optimal indoor temperature profiles τ^{in^*} (4.14). In (4.14), g is the mapping between τ^{in^*} and the ex-post measures $\check{p}^{\text{hvac}}$ and $\check{\tau}^{\text{in}}$. The function g models HVAC actuation, building response, and sensing.

Gradient Computation

Problem (4.13)-(4.16) is intractable and realistically large instances cannot be handled by off-the-shelf solvers. Therefore, we aim at learning θ by gradient descent. The backpropagation step of the conventional approach applies the chain rule, but for tractability this requires g to be differentiable. In practice, real systems, and even some building simulators, are not differentiable.

To address this issue, Stochastic Smoothing (SS) aims to learn directly from an ex-post loss signal, analogous to the reward in reinforcement learning (Figure 4.1). SS does not make any assumption about the form of the HVAC MS, the differentiability of g, and the structure of the ex-post loss. To that end, instead of producing a point estimate of θ, SS models it as a distribution, e.g., a Gaussian $\mathcal{N}(\theta, \sigma)$. Because of the stochasticity in θ, the loss $\mathcal{L}$ in (4.13) becomes an expectation:

$$\mathcal{L} = \mathbb{E}_{\theta_s\sim\mathcal{N}(\theta,\sigma)} \left[\mathcal{L}_s(\check{p}_s^{\text{hvac}}, \check{\tau}_s^{\text{in}})\right]. \tag{4.17}$$

By modeling probabilistic outputs, this approach avoids uninformative gradients that plague traditional deterministic methods. Even when the original gradient (e.g., from a LP) would be zero, taking the expectation over a smoothed distribution can yield informative, non-zero gradient signals. However, computing these gradients analytically is challenging.

To train the model, we thus employ score-function gradient estimation, a technique closely related to the REINFORCE algorithm from reinforcement learning. The approximation of the gradient becomes [19]:

$$\frac{\partial \mathcal{L}}{\partial \theta} = \mathbb{E}_{\theta_s \sim \mathcal{N}(\theta, \sigma)} \left[\mathcal{L}_s \frac{\partial \log \mathcal{N}_\theta(\theta_s)}{\partial \theta} \right]. \tag{4.18}$$

In practice, the expectation of the loss $\mathcal{L}$ is approximated using Monte Carlo with S samples [120]:

$$\frac{\partial \mathcal{L}}{\partial \theta} = \frac{1}{S} \sum_{s=1}^{S} \left[\mathcal{L}_s \frac{\partial \log \mathcal{N}_\theta(\theta_s)}{\partial \theta} \right]. \tag{4.19}$$

Although SS can accommodate any type of optimization problem, we reformulate the MILP as a MIQP by relaxing the indoor temperature constraint (4.4) and incorporating it as a quadratic penalty in the objective function (4.20).

Reformulating the problem as a MIQP is not strictly required to obtain meaningful gradients, since SS is agnostic to the optimization structure, but it helps stabilize the learning process. By penalizing constraint violations in the objective (in the MIQP) rather than enforcing them strictly (in the original MILP), the MIQP approach prevents infeasibility during training, thereby ensuring smoother and more robust model updates. Moreover, the hard constraints can be enforced at test time, if desired.

The quadratic term (iii) penalizes deviation of zonal indoor temperature $\tau^{\text{in}}_{t,z}$ from the target temperature $T^{\text{tgt}}_{t,z}$, reflecting the loss of thermal comfort of the occupants. The weight $o_{t,z}$ should be designed to reflect the occupancy of the zone, i.e., higher occupancy implies a greater importance of thermal comfort and thus a stronger penalty. In our case, $o_{t,z} \in [0, 1]$ is the ratio of zone occupancy (i.e., people count) to maximum occupancy of the entire building.

$$\min_{\tau^{\text{in}}} p^{\text{d}} \lambda^{\text{d}} + \sum_{t=0}^{T} \big(p^{\text{i}}_t \lambda^{\text{i}}_t - p^{\text{e}}_t \lambda^{\text{e}}_t + \underbrace{\sum_{z=0}^{Z} o_{t,z} (\tau^{\text{in}}_{t,z} - T^{\text{tgt}}_{t,z})^2}_{(iii)} \big) \Delta t \tag{4.20}$$

The resulting algorithm is detailed in Algorithm 1.

Algorithm 1 Algorithm for decision-focused learning of the HVAC management system.

```
Input:
Database of exogenous parameters R
Initial NN parameters θ_0
for each epoch do
    for each parameter sample r in R do
        for each sample s do
            Day-Ahead Stage:
            Sample θ_s ∼ N(θ, σ)
            Solve MIQP HVAC MS problem
            Simulator or Real Building:
            Thermostat control with τ_s^{in*} as setpoints
            Get ex-post power p̌_s^{hvac} and temperature τ̌_s^{in*}
            Ex-Post Analysis:
            Compute loss function L_s(p̌_s^{hvac*}, τ̌_s^{in*})
        end for
        Backward Pass:
        Compute gradient ∂L/∂θ ≈ (1/S) Σ_{s=1}^{S} L_s ∂log N_θ(θ_s)/∂θ
        Update NN parameters θ_{n+1} ← θ_n − α ∂L/∂θ
    end for
end for
```

Task Loss

The task loss $\mathcal{L}$ aims to evaluate the quality of the decisions made by the HVAC MS. The ideal task loss is the exact ex-post value of the decisions, which includes both the cost and the thermal comfort. Here, we align our loss $\mathcal{L}$ with the HVAC MS objective (4.20). We name the loss *Expost+* and define it as:

$$\mathcal{L} = \sum_{t=0}^{T} \underbrace{(\check{p}_t^{\text{hvac}} \check{\lambda}_t + \sum_{z=0}^{Z} o_{t,z} (\Delta \check{\tau}_{t,z}^{\text{in}})^2)}_{\text{Ex-post objective}} + \underbrace{\text{MSE}(\check{C} - C)}_{\text{MSE Power Cost}}, \tag{4.21}$$

where $\Delta\check{\tau}^{\text{in}}_{t,z} = \check{\tau}^{\text{in}}_{t,z} - T^{\text{tgt}}_{t,z}$ and $\text{MSE}(\check{C} - C) = \frac{1}{T}\sum_{t=0}^{T}(\check{p}^{\text{hvac}}\check{\lambda}_t - p^{\text{hvac}}\lambda_t)^2$. The ex-post price $\check{\lambda}_t$ depends on the ex-post import and export powers $\check{p}^{\text{i}}_t$ and $\check{p}^{\text{e}}_t$:

$$\check{\lambda}_t = \begin{cases} \lambda^{\text{i}}_t \text{ if } \check{p}^{\text{e}}_t = 0 \text{ and } \check{p}^{\text{i}}_t \neq \check{p}^{\text{d}}_t, \\ \lambda^{\text{e}}_t \text{ if } \check{p}^{\text{i}}_t = 0 \text{ and } \check{p}^{\text{e}}_t \neq \check{p}^{\text{d}}_t, \\ \lambda^{\text{i}}_t + \lambda^{\text{d}}_t \text{ if } \check{p}^{\text{i}}_t = \check{p}^{\text{d}}_t, \\ \lambda^{\text{e}}_t + \lambda^{\text{d}}_t \text{ if } \check{p}^{\text{e}}_t = \check{p}^{\text{d}}_t. \end{cases} \tag{4.22}$$

4.2.3. Neural Network MILP Formulation

Constraint (4.2) of the model presented in section 4.2.1 represents the building thermal dynamics over a time interval. We propose learning building thermal dynamics in a data-driven fashion by leveraging piecewise linear NNs, which are obtained by designing NNs with only piecewise linear activation functions. Specifically, we use ReLU activation functions, which are made up of two linear pieces. ReLU speeds up training and reaches greater accuracy than other activation functions for deep [121] and sparse NNs [122], but ReLU-based NNs may be nonconvex [123]. The ReLU function is defined as:

$$y = \max(0, \hat{y}). \tag{4.23}$$

However, embedding (4.23) directly into the larger HVAC MS optimization makes the overall problem intractable, as it requires solving a nested optimization at each ReLU. Consequently, each ReLU neuron $n \in \mathcal{N}$ is reformulated by a binary variable σ_n (4.24), two continuous variables y_n and $\hat{y}_n$ (4.25), and a set of constraints. The complexity of the resulting optimization problem scales with the number of neurons N.

$$\sigma_n \in \{0, 1\} \tag{4.24}$$

$$y_n, \hat{y}_n \in \mathbb{R} \tag{4.25}$$

The first step, as shown in Figure 4.2, is to compute the preactivation function (4.26) of the neuron n based on the input x, the weights w_n, and the bias b_n. For I inputs, $x \in \mathbb{R}^I$, $w_n \in \mathbb{R}^I$, and $b_n \in \mathbb{R}$. The preactivation value $\hat{y}_n$ is a scalar.

$$\hat{y}_n = \sum_{i=1}^{I} w_{n,i}x_{n,i} + b_n \quad \forall n \tag{4.26}$$

The second step is to choose a suitable formulation of the ReLU activation function, which is typically categorized into four main types: a disjunctive

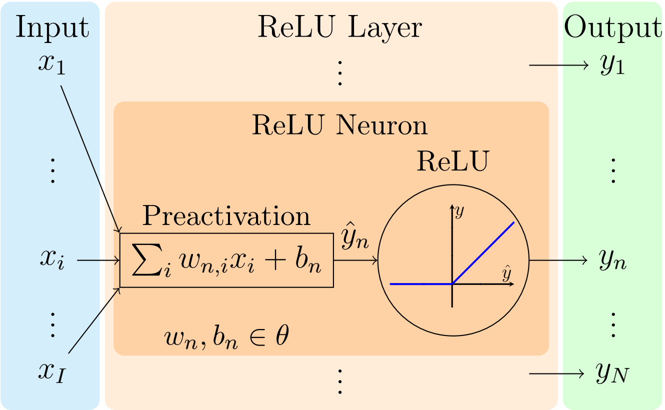


Figure 4.2.: Layer of neurons with Rectified Linear Unit (ReLU) activation function. The parameters are the weight w_i and the bias b.

formulation [124], a Big-M formulation [26], a strong formulation [25], and a partition-based formulation [125]. Despite offering a strong LP relaxation of the ReLU, the strong formulation requires an infinite number of constraints or a significant number of additional auxiliary variables, which hinders its performances. Here, we adopt the Big-M formulation for its simplicity and excellent empirical performance [1], [126], [127], [128].

First, the Big-M formulation of the ReLU defines the convex hull of the function as follows:

$$y_n \geq 0, \tag{4.27}$$

$$y_n \geq \hat{y}_n, \tag{4.28}$$

where $\hat{y}_n$, the preactivation value, is the input of the ReLU function and y_n is the output. The constraints (4.29) and (4.30) are alternatively binding depending on the value of the binary σ_n. The parameters $\hat{Y}_n^{\max}$ and $\hat{Y}_n^{\min}$ are the Big-M constants. When $\sigma_n = 0$, (4.29) is binding, imposing y_n to be null with (4.27). Variable $\hat{y}_n$ must be negative (4.28). In contrast, $\sigma_n = 1$ makes (4.30) binding, imposing $\hat{y}_n = y_n$ via (4.28). Variable $\hat{y}_n$ must be positive (4.27).

$$y_n \leq \hat{Y}_n^{\max} \cdot \sigma_n \tag{4.29}$$

$$y_n \leq \hat{y}_n - \hat{Y}_n^{\min} \cdot (1 - \sigma_n) \tag{4.30}$$

4.2.4. Improved tightness of the NN reformulation

The Big-M constants $\hat{Y}_n^{\min}$ and $\hat{Y}_n^{\max}$ need to be accurately determined to produce a tight formulation of the ReLU. Three approaches exist in the literature. A first naive approach is to record the minimum and maximum values of $\hat{y}_n$ during training. Although this method preserves the correlation between the NN inputs and remains simple, it is not well suited for NN-constrained optimization since physical constraints rather than the distribution of the NN training dataset should define the feasible domain [1]. More advanced methods formulate optimization problems to find the bounds $\hat{Y}_n^{\min}$ and $\hat{Y}_n^{\max}$ [24], [127]. Bound optimization is computationally cumbersome, making it impractical for online computations. The third approach is to calculate the bounds in all layers of the NN based solely on the input bounds using interval analysis [129]. Interval analysis enables efficient computation of bounds while ensuring physical consistency by appropriately setting the limits of the NN inputs. However, the correlation between inputs is lost.

In this work, the NN parameters θ are updated at each gradient descent during training via DFL (cf. section 4.2.2), such that the bounds $\hat{Y}_n^{\min}$ and $\hat{Y}_n^{\max}$ must be recomputed after each gradient descent step. To handle this efficiently, we use interval analysis, and further refine the method by setting the feasible interval of NN input to singleton (i.e., degenerate) interval if the input value is known prior to solving the optimization problem. This applies to all optimization parameters that are input to the NN.

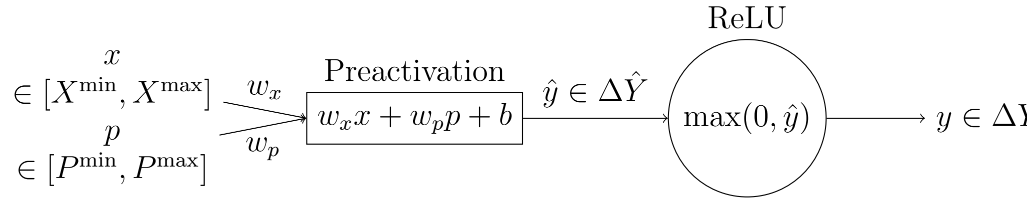


Figure 4.3.: Interval analysis of the bounds for a ReLU neuron with inputs p and x, where p is a parameter and x a variable in the overarching optimization.

Theorem 4.1 (Preactivation Range Reduction)**.** *Consider a neuron, reformulated as mixed-integer linear constraints using Big-M constants, with preactivation* $\hat{y} = w_x x + w_p p + b$*, where* $x \in \mathbb{R}^X$ *are unknown input variables with weights* $w^x \in \mathbb{R}^{1\times X}$*,* $p \in \mathbb{R}^P$ *are known parameters with weights* $w^p \in \mathbb{R}^{1\times P}$*, and* b *is the neuron bias. Assume that* x *and* p *are bounded such that* $x \in [X^{min}, X^{max}]^X$*, and* $p \in [P^{min}, P^{max}]^P$*. Let* $\Delta\hat{Y}$ *denote the feasible range of the preactivation* *under these bounds. When the feasible range of p is reduced to a degenerated interval $p \in [P_0, P_0]^P$, the feasible range of the modified preactivation $\hat{y}' = w_x x + w_p P_0 + b$ satisfies:*

$$\frac{|\Delta \hat{Y}'|}{|\Delta \hat{Y}|} \approx 1 - \frac{\sum_p^P w_p}{\sum_x^X w_x + \sum_p^P w_p} \tag{4.31}$$

where $|\cdot|$ denotes the Lebesgue measure (i.e., length) of the interval.

Proof. Assuming we have a ReLU neuron with two inputs $x \in \mathbb{R}$ and $p \in \mathbb{R}$ that are respectively variable and parameter of the optimization problem, as illustrated by Figure 4.3, we can demonstrate the tightening of the bounds by our approach. Let $\Delta(\cdot)$ denote the feasible interval of its argument. If

$$x \in \Delta X = [X^{\min}, X^{\max}], \tag{4.32}$$
$$p \in \Delta P = [P^{\min}, P^{\max}], \tag{4.33}$$

then, assuming without loss of generality that the associated weights $w_x \in \mathbb{R}$ and $w_p \in \mathbb{R}$ are positive[2], interval analysis defines the bounds of the interval of $\hat{y}$ as follows

$$\hat{Y}^{\min} = w_x X^{\min} + w_p P^{\min} + b, \tag{4.34}$$
$$\hat{Y}^{\max} = w_x X^{\max} + w_p P^{\max} + b. \tag{4.35}$$

Therefore, the interval length according to Lebesgue measure $|\cdot|$ on $\hat{y}$ is

$$|\Delta \hat{Y}| = \hat{Y}^{\max} - \hat{Y}^{\min} \tag{4.36}$$
$$= w_x |\Delta X| + w_p |\Delta P|. \tag{4.37}$$

By degenerating the parameter interval (4.33) as $p \in [P_0, P_0]$, where P_0 is the parameter value, equation (4.37) becomes

$$|\Delta \hat{Y}'| = w_x |\Delta X|. \tag{4.38}$$

[2] If a weight is negative, the lower bound of the input associated to that weight must be used in (4.34) and the upper bound in (4.35).

The ratio of the interval lengths is

$$\frac{|\Delta \hat{Y}'|}{|\Delta \hat{Y}|} = \frac{w_x|\Delta X|}{w_x|\Delta X| + w_p|\Delta P|}, \tag{4.39}$$

$$= 1 - \frac{w_p|\Delta P|}{w_x|\Delta X| + w_p|\Delta P|} \tag{4.40}$$

Assuming that the inputs have been normalized, which resulted in $|\Delta X| \approx |\Delta P|$, we have

$$\frac{|\Delta \hat{Y}'|}{|\Delta \hat{Y}|} \approx 1 - \frac{w_p}{w_x + w_p}. \tag{4.41}$$

Equation (4.41) can be extended to P parameter and X variable inputs as follows:

$$\frac{|\Delta \hat{Y}'|}{|\Delta \hat{Y}|} \approx 1 - \frac{\sum_p^P w_p}{\sum_x^X w_x + \sum_p^P w_p}, \tag{4.42}$$

which proves that the proposed method is at worst equivalent to the state-of-the-art since the weights have been assumed positive. The improvement in the bound tightness depends on the ratio of the sum of parameter input weights and the sum of all weights. ■

The bounds on y are

$$y \in [\max(0, \hat{Y}^{\min}), \max(0, \hat{Y}^{\max})]. \tag{4.43}$$

4.3. Case Study

We analyze the effectiveness of the proposed method on a realistic office building comprising five zones. We start in section 4.3.1 by describing the high-quality and publicly available building model, datasets, and tools used to design the case study, ensuring a fair comparison of results and reproducibility. Then, section 4.3.2 presents the different benchmarks to compare the performance of our DFL-based training procedure. In section 4.3.3, we report the performance of all the models learned in a DFL-fashion. Finally, in section 4.3.4, we further compare our DFL approach to the corresponding ITO baseline, where the building's thermal dynamics are learned using a standard statistical loss function independent of the optimization task. Our code is publicly available at https://github.com/PSMRB/dfl_hvac_management.

4.3.1. Data

Building Prototype

The US Department of Energy has developed EnergyPlus. EnergyPlus is a high-fidelity physics-based simulator for building energy modeling, continuously updated since 2001 [17]. EnergyPlus includes publicly available building models [22]. We selected an office building with five actively controlled thermal zones. The zones are located over one floor of 511 m^2. Figure 4.4 provides an illustration of the floor layout. Each zone is equipped with its own air-to-air

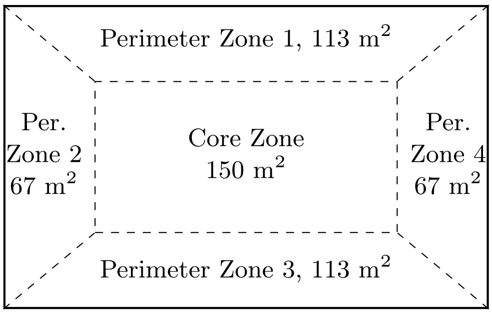


Figure 4.4.: Layout of the building floor.

heat pump, such that the HVAC system of each zone is fully independent. The only alteration to the building model is the replacement of the gas-fired heating coil by an electric coil of equal power.

Weather

Weather data and their processing is similar to section 3.5.3. Information is repeated here for the reader's convenience.

The building is assumed to be located in Denver, Colorado. To model building thermal dynamics, we utilize two typical meteorological datasets, each representing a full year of hourly weather data for Denver International Airport, derived from long-term observations. The first dataset is used to simulate one year of building operation using EnergyPlus's default heuristic control (i.e., without optimization), while the second provides weather scenarios used to evaluate optimized scheduling strategies.

Optimizing then simulating daily schedules for all 365 days is computationally demanding. Moreover, there exist days with similar weather that bring very little extra information to the learning process. Therefore, we apply k-medoids clustering to group similar weather conditions from the second dataset. The k-medoids algorithm is chosen over k-means to prevent the generation of synthetic average data. As ambient temperature is the only meteorological input required for the model (4.2), we use it as the primary clustering criterion. Initially, we select three fixed medoids corresponding to the most extreme conditions: the coldest day, the hottest day, and the day with the highest temperature variability. We then determine seven additional medoids to partition the rest of the dataset. The resulting temperature profiles from the k-medoids clustering are illustrated in Figure 4.5, with the cluster means and standard deviations summarized in Figure 4.6. In total, only ten carefully selected days are used to train the model that should generalize over the whole year.

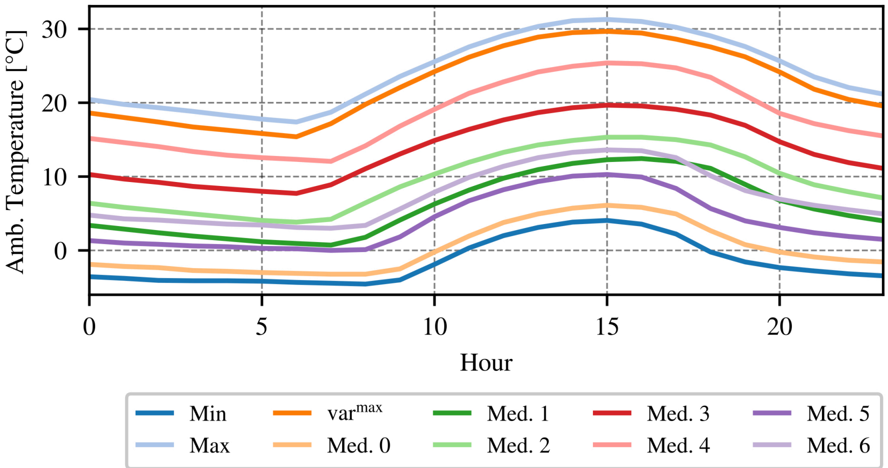


Figure 4.5.: Ambient temperature profiles of the ten medoids. The three first labels are extreme scenarios.

Time-of-use Tariff

Time-of-use tariff is similar to section 3.5.3. Information is repeated here for the reader's convenience.

To reflect typical load patterns and incentivize off-peak usage, we adopt a time-of-use electricity tariff: 0.6 \$/kWh from 6 a.m. to 7 p.m., the rate is double the base rate at 0.6 \$/kWh, and 0.3 \$/kWh during off-peak hours (see Figure 4.7).

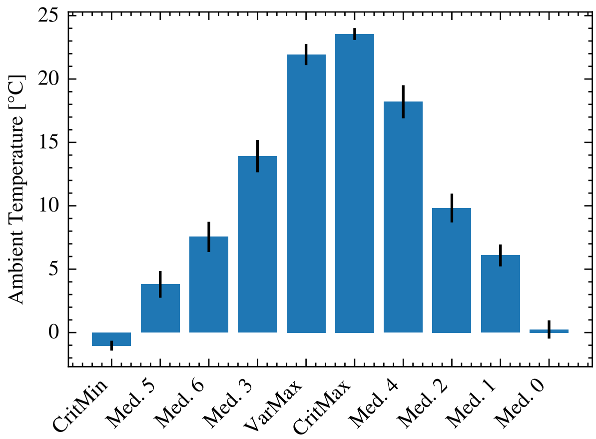


Figure 4.6.: Mean and standard deviation of the medoids. They are ordered to form a smooth cycle.

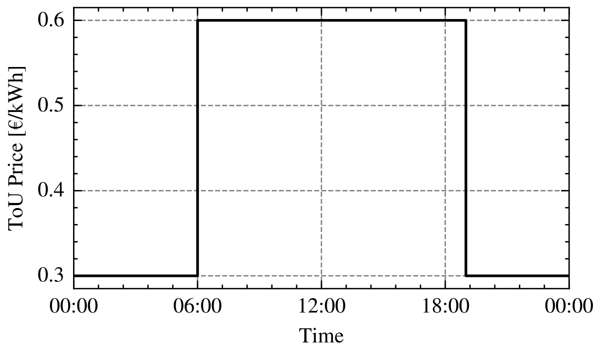


Figure 4.7.: Time-of-use tariff with an appreciation of electricity from 6am to 7pm.

4.3.2. Models and Benchmarks

Our goal is to learn the parameters θ of a piecewise linear NN, i.e., constraint (4.2), representing the thermal model of the building. We learn the parameters in a decision-focused fashion using gradient descent. To that end, we coded in *Python* and used *PyTorch* for computing the backpropagation. All computations were performed on an Intel Skylake 16-core Xeon 6142 processors and 16GB of RAM. To perform the gradient descent, we use *Adam* optimizer with an initial learning rate at 10^{-3} (except for the RC model where 10^{-2} was found to be better). The learning rate decays exponentially with $\gamma = 0.98$. We limit the number of epochs (i.e., the number of times the 10 scenarios are run) to 100 for the NNs, 150 for the RC model, and set an early stopping with a patience of 15.

We pre-train the NNs in a supervised way using the first dataset (i.e., one year of building operation with default heuristic control) to capture the building's main thermal characteristics. The resulting pre-trained NNs are then trained in a DFL fashion to learn additional and more complex relationships. This is achieved using SS.

SS learns the parameters of the NN-based thermal model embedded within an MIQP by introducing controlled random perturbations, which smooth the loss landscape and allow for gradient-based optimization. The resulting gradient is computed using the REINFORCE algorithm which enables bypassing the differentiation of the HVAC MS and EnergyPlus. We take a single sample $S = 1$ for the Monte Carlo approximation following conventional practice in the literature [19] and allowing practical implementation with a real-life system where only one outcome can be observed. We investigated five Gaussian noises of various intensity with standard deviations of 0.01, 0.05, 0.1, $\mu/10$, $\mu/2$ to perturb the parameters. The value of σ can be fixed or variable.

Our proposed DFL-based approach using SS is compared to other state-of-the-art DFL-based methods. These include (*i*) a Quadratic Program (QP) relaxation of the original optimization problem where the binaries are relaxed as continuous variables between 0 and 1 and (*ii*) a two-step approach that first fixes the integer variables, then solves the resulting continuous and differentiable QP problem. *Cvxpylayer* computes the gradient of QP problems that are solved with *ECOS* since it has the best performance for conic programs [94]. However, the MIQP is solved using *Gurobi*.

Because EnergyPlus is not differentiable, the QP relaxation and the two-step approach cannot directly optimize the ex-post value (4.21). Therefore, we introduced in the previous chapter (section 3.5.4) a *supervized DFL loss* that bypasses the need to differentiate EnergyPlus (or the real installation) because we assume that $\check{p}_t^{\text{hvac}}$ is independent on θ:

$$\mathcal{L}_{\text{DFL}} = \sum_{t=0}^{T} \frac{\check{\lambda}_t}{T} \left(|p_t^{\text{hvac}} - \check{p}_t^{\text{hvac}}| + \text{MAE}\left(p_{t,z}^{\text{hvac}}\right) \right), \tag{4.44}$$

where p_t^{hvac} is the total HVAC consumption of the building at time step t, and $\text{MAE}\left(p_{t,z}^{\text{hvac}}\right) = \frac{1}{Z}\sum_{z=0}^{Z} |p_{t,z}^{\text{hvac}} - \check{p}_{t,z}^{\text{hvac}}|$.

Regarding the NN architectures, we learn the parameters of three fully connected feedforward NNs with six inputs: the temperature in each zone (5 inputs) and the outdoor temperature. The inputs are normalized. The three architectures have one hidden layer with 2 (NN1), 5 (NN2), and 10 ReLU (NN3). The last layer of all architectures is a linear layer with five outputs:

the zonal temperatures at the next time step. Each combination of architecture and DFL method is run over five seeds, and the best outcome is reported.

Finally, we also compare the NNs to a lumped RC model, a linear model of the building based on the aggregated resistance and capacitance of each zone. For each zone, the RC model turns constraint (4.2) into:

$$\tau^{\mathrm{in}}_{z,t+1} = \tau^{\mathrm{in}}_{z,t} + \left(\frac{\eta^{\mathrm{h}}_z p^{\mathrm{h}}_{z,t} - \eta^{\mathrm{c}}_z p^{\mathrm{c}}_{z,t}}{C_z} + \frac{\tau^{\mathrm{amb}}_{z,t} - \tau^{\mathrm{in}}_{z,t}}{R_z C_z} \right) \Delta t. \tag{4.45}$$

Table 4.1 gathers the hyperparameters used in this case study. For more detail, you can visit the GitHub associated with the chapter.

Optimizer	Adam
Learning Rate	10^{-3} for NN
	0.02 for RC
Exponential Decay	0.98
Max. Nb. Epoch	100 (150 for RC)
Patience	15
SS Noise (σ)	0.01, 0.05, 0.1, $\mu/10$, $\mu/10$
S	1
NN1	1 layer, 2 ReLU
NN2	1 layer, 5 ReLU
NN3	1 layer, 10 ReLU

Table 4.1.: Hyperparameters.

4.3.3. DFL Results

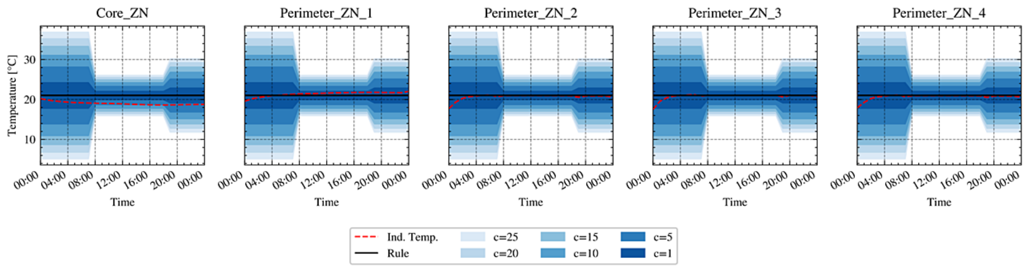
Before surveying the results of the DFL trainings in Table 4.2, we analyze in Figure 4.8 the results for the day with minimal outdoor temperature provided by the NN1 model trained with SS. Temperature deviations from the preferred temperature T^{tgt} occur mostly in the first four hours of the day. By 4 am, the four perimeter zones have reached T^{tgt} and roughly maintain it over the day. Very slight deviations appear after 6 pm. In contrast, the core zone is slightly under-heated with a temperature between 18°C and 20°C for the whole day. The zonal expected powers shown in Figure 4.8b are meaningless. This is because the thermal dynamics model focuses on delivering good decisions through minimization of the task loss. However, the overall trend of the expected aggregated power Figure 4.8c is right and arguably quite good given the high amount of perturbation that affects a building.

Table 4.2 reports the results of the various DFL-based models. For the two smaller NN architectures (NN1 & NN2), the three DFL methods—i.e., SS, QP-relaxation (QP), and QP subproblem with Fixed Binaries (FB)—converge successfully. For the most complex architecture, NN3, only QP training is feasible within 24 hours. Indeed, the complexity of the HVAC MS increases with the number of neurons. Consequently, since the SS and FB require to solve the MIQP for each sample, the training time quickly becomes computationally intensive.

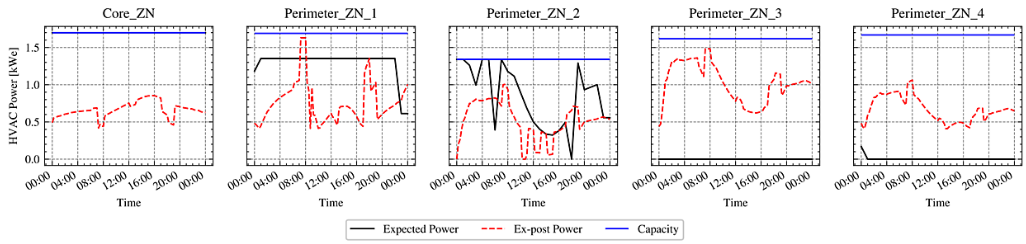
The first two rows, *Ex-post+* and Hierarchical Loss, are the metrics used for training. SS can directly minimize *Ex-post+* (i.e., ex-post power cost, thermal discomfort penalty, and power cost misestimation), whereas QP and FB methods minimize the error on the HVAC power through the hierarchical loss (4.44).

SS obtains the best ex-post value compared to QP, and FB. For NN1, SS ($\sigma = 0.01$) obtains the lowest *Ex-post+* at \$94. FB and QP follow at \$113 and \$130, respectively. All three NN1 models have an ex-post cost between \$41 and \$41.5. However, the model accuracy to predict the cost varies. SS is the most accurate model with a daily cost underestimation of \$3.7. QP and FB follow with an overestimation of \$4.69 and \$4.74, respectively. Regarding the temperature penalties, SS comes first with only \$31 of daily penalty compared to \$41 for FB, and \$57 for QP. The total computational time (training, validation, and test) ranges from 95 minutes for QP to 136 minutes for SS.
For NN2, the same ranking is observed as for NN1. SS ($\sigma = 0.01$) outperforms with an *Ex-post+* value at \$84. This is even better than \$94 for NN1. FB and QP follow with \$101 (\$113 for NN1) and \$182 (\$130 for NN1). The ex-post cost ranges from \$41.8 for SS to \$42.7 for FB. The cost estimation error is \$1.95 for SS, but rises to \$6.17 for QP. The temperature penalty goes from \$35 for SS to \$59 and \$77 for QP and FB, respectively. In conclusion, SS offers for both architectures the best thermal comfort and with excellent estimate of the power cost along with very competitive ex-post cost and computational burden.

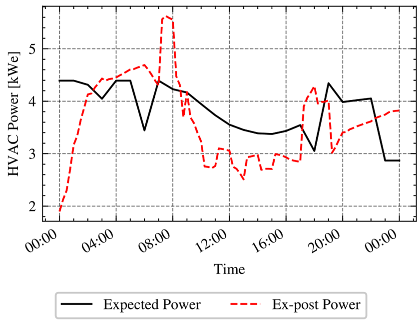
FB outperforms QP on almost all training and ex-post metrics. For NN1, FB has a better hierarchical loss (12.8) and *Ex-post+* (\$113) than QP (hierarchical loss at 14.3 and *Ex-post+* at \$130). Even though FB has a similar ex-post cost (\$41.0 versus \$41.1 for QP) and cost error (\$4.74 versus \$4.69 for QP), FB provides more thermal comfort (temperature penalty at \$41 versus \$57 for QP). FB and QP take advantage of NN2's stronger modeling abilities to improve their training loss (FB: 11.4 vs. 12.8; QP: 12.4 vs. 14.3)). However,



(a)



(b)



(c)

Figure 4.8.: Scheduling for the coldest day obtained by the SS NN1 model. Subfigures are (a) the zonal indoor temperature; (b) the zonal HVAC electric power, and (c) the aggregated HVAC power for the whole building. The blue shades indicate the temperature deviations causing an objective value penalty lower than a given threshold. Temperature control is more loose during the night and evening because of the worker absence. The mismatch between the expected and ex-post HVAC powers in (b) can be explained by the model limitations and the DFL loss which gives little importance to accurate zonal predictions.

	RC	NN1			NN2			NN3
		SS	QP	FB	SS	QP	FB	QP
Ex-post+ ($)	252	94	130	113	84	182	101	125
Hierarchical loss	24.7	20.4	14.3	12.8	22.2	12.4	11.4	13.2
MAE (kW)	0.6	0.52	0.55	0.55	0.5	0.54	0.54	0.54
MSE (kW2)	0.44	0.35	0.37	0.39	0.34	0.39	0.39	0.39
Error mean (kW)	-0.07	-0.08	0.07	0.04	-0.01	0.02	0.01	0.03
Error std (kW)	0.39	0.42	0.37	0.36	0.45	0.37	0.39	0.36
Expected cost ($)	32.3	37.9	45.8	45.8	39.9	47.4	43.1	45.4
Ex-post cost ($)	38.0	41.5	41.1	41.0	41.8	41.2	42.7	43.2
Cost error ($)	5.74	3.64	-4.69	-4.74	1.95	-6.17	-0.43	-2.22
Temp. Penalty($)	162	31	57	41	35	59	47	58
Nb. Epochs	150	71	36	54	43	25	94	100
Training time	03:30:24	01:08:34	00:48:59	00:56:49	04:17:05	00:49:33	07:19:17	04:25:18
Validation time	03:14:32	01:07:13	00:44:01	00:52:38	04:22:52	00:59:21	06:16:23	03:54:35 [3]
Test time	00:00:46	00:00:49	00:01:51	00:00:57	00:07:59	00:01:57	00:01:05	00:10:13

[4]Validation conducted on the continuous relaxation (i.e., QP-relaxation) of the problem.

Table 4.2.: Test metrics of the decision-focused learning models: NN with one layer of two ReLU (NN1), five ReLU (NN2), ten ReLU (NN3), and one multi-zone RC model (RC). Three methods make the MIQP problem differentiable: fixing the binaries to their optimal value (FB), relaxing the integer constraint (QP), or applying a stochatic smoothing (SS).

FB *Ex-post+* falls to \$101 reflecting better ex-post metrics whereas it increases to \$182 for QP. These results show (*i*) the inherent misalignment between the hierarchical loss and the ex-post value; (*ii*) the poor quality of the gradient obtained by relaxing the problem.

The poor gradient approximation generated by QP even undermines the performance of larger architectures such as NN3, causing them to underperform relative to smaller SS-trained architectures. Note that for NN3, the QP relaxation is also used for validation. Since the complexity of the HVAC MS grows with the number of neurons, SS and FB become too cumbersome. Indeed, FB and SS require to solve the MIQP at each sample. QP hierarchical loss is worse for NN3 than for NN2 going up from 12.4 to 13.2, while *Ex-post+* goes down from \$182 to \$125. The ex-post cost stands at \$43.2, slightly above the usual range. The cost overestimation (\$2.22) and the temperature penalty (\$58) become competitive. However, such a large NN with ten ReLU is long to train (almost four hours) and is still outperformed in terms of *Ex-post+* value by NN1 trained by SS. Moreover, solving the MIQP with NN3 at test time is about ten times longer than with NN1.

In contrast, since the RC model is linear, the resulting HVAC MS is inherently a continuous QP problem. Nevertheless, the RC model cannot capture the complexity of the thermal dynamics and exhibits the worst metrics among all models. The RC model is trained to minimize the hierarchical loss, which converges at 24.7 and *Ex-post+* at \$252. Regarding the ex-post metrics, the average daily ex-post cost of HVAC power is \$38.0, but it is largely underestimated at \$32.3. The average daily temperature penalty, reflecting the thermal discomfort, is \$150. Therefore, the ex-post cost is the lowest of all models because the RC model fails to provide comfortable temperatures. The total computational time for training, validation, and test is about seven hours.

In conclusion, SS outperforms DFL approaches that differentiate through the optimization problem (QP or FB), particularly when the underlying system or its simulator is non-differentiable. Unlike QP and FB, SS bypasses both the differentiation of the optimization layer and the system model, allowing the direct use of the true ex-post value (*Ex-post+*) as a training signal instead of relying on surrogate task losses. This results in superior gradient estimates and more informative training loss, enabling smaller models to outperform larger QP-trained counterparts while reducing computational overhead at inference.

4.3.4. ITO Results

Table 4.3 compares the results of the DFL-trained models with their ITO counterpart. ITO models are trained over one year of historical data by Mean

	RC		NN1		NN2		NN3	
	ITO	DFL	ITO	SS	ITO	SS	ITO	QP
Ex-post+ ($)	516	252	318	94	495	84	579	125
Hierarchical loss	54.2	24.7	35.2	20.4	43.7	22.2	39.4	13.2
MAE (kW)	0.59	0.6	0.5	0.52	0.51	0.5	0.53	0.54
MSE (kW2)	0.45	0.44	0.34	0.35	0.36	0.34	0.38	0.39
Error mean (kW)	-0.27	-0.07	-0.26	-0.08	-0.35	-0.01	-0.35	0.03
Error std (kW)	0.55	0.39	0.38	0.42	0.38	0.45	0.35	0.36
Expected cost ($)	21.3	32.3	26.8	37.9	21.7	39.9	20.5	45.4
Ex-post cost ($)	41.5	38.0	41.6	41.5	42.2	41.8	42.2	43.2
Cost error ($)	20.2	5.7	14.7	3.6	20.5	2.0	21.7	-2.2
Temp. Penalty($)	55	162	32	31	17	35	16	58
Nb. Epochs	0	150	0	71	0	43	0	100
Training time	-	03:30:24	-	01:08:34	-	04:17:05	-	04:25:18
Validation time	-	03:14:32	-	01:07:13	-	04:22:52	-	03:54:35
Test time	00:01:12	00:00:46	00:01:46	00:00:49	00:02:17	00:07:59	02:15:50	00:10:13

Table 4.3.: Comparison of the performance of the Identify-Then-Optimize approach to decision-focused learning via Stochastic Smoothing (SS) for each model—NN with one layer of two ReLU neurons (NN1), five ReLU neurons (NN2), ten ReLU neurons (NN3), and one multi-zone RC model (RC).

Square Error (MSE) minimization. The trained NN is then reformulated as a constraint of the HVAC MS. The ITO and DFL models are tested over the same ten days.

All models see a significant improvement with DFL compared to ITO. The three NN models see major improvements in terms of *Ex-post+* and hierarchical loss. *Ex-post+* falls from \$318 to \$94 for NN1, from \$495 to \$84 for NN2, and from \$579 to \$125 for NN3. The cost errors of the NN1 and NN2 ITO models are \$14.7 and \$20.5, compared to \$3.6 and \$2.0 for the SS models. The QP DFL with NN3 reduces the cost error from \$21.7 to \$2.2. Similarly, DFL of the RC parameters improves the hierarchical loss from 54.2 to 24.7 and *Ex-post+* from \$516 to \$252. It shows the possible improvement DFL can bring even for simple models as already noted in [130]. Because of DFL, the RC model can outperform more complex ITO models such as NNs. The *Ex-post+* value of the RC model is \$252, much lower than \$318 for NN1, \$495 for NN2, and \$579 for NN3.

When surveying only ITO models, NN1 is the best architecture, which shows that when ITO is used, a more complex architecture may not yield better decisions. NN1 reports an *Ex-post+* ITO value at \$318 far below the RC model (\$516), NN2 (\$495), and NN3 (\$579).

In conclusion, DFL significantly improves the quality of the ex-post metrics reflecting better and more realistic decisions compared to the conventional two-stage ITO approach. A simple DFL-trained model, such as the RC model, is to be preferred to ITO-trained NNs. Nevertheless, the improvement brought by DFL is limited by the modeling power of the architecture, as shown for RC model, and the quality of gradient approximation, as shown for NN3. Caution is warranted when resorting to ITO: increasing model complexity does not necessarily yield better decisions and may introduce additional computational burden.

4.3.5. Effect of Constraint Tightness on Solving Time

In subsection 4.2.4, we analytically showed how to tighten the MIL equations of a ReLU by dynamically adjusting the feasible interval of its inputs. In this case study, the NNs have 11 inputs, among which one is a parameter (i.e., the ambient temperature) and ten are decision variables (i.e., the five zonal indoor temperatures and the related HVAC power consumption). Applying (4.31) and assuming that the weights have the same magnitude, the expected theoretical improvement in tightness is approximately 9.1%.

As shown in Table 4.4, the tight formulation consistently reduces solving time across all test cases, with an average gain of 9.4%. This validates the the-

	NN1			NN2			NN3	Avg.
	SS	QP	FB	SS	QP	FB	QP	
SOTA	39.0	23.4	23.0	41.5	72.6	211	1443	264.4
Tight	32.0	23.2	19.8	38.2	66.0	180	1422	254.0
Gain (%)	17.9	0.9	13.9	8.0	9.1	14.7	1.5	9.4

Table 4.4.: Comparison of the solving time for the optimization during test (i.e., MIQP formulation) with and without the proposed tight formulation.

oretical prediction from (4.31) and demonstrates the practical advantage of our approach in terms of computational efficiency. Importantly, this improvement is achieved without sacrificing model accuracy or solution quality.

4.3.6. Noise intensity analysis

In this section, we analyze the impact of the noise intensity on the performance of SS-DFL. Since we use a Gaussian noise, we modify σ to investigate five level of noise: 0.01, 0.05, 0.1, $\mu/10$, and $\mu/2$. We further enable two modes for σ: it can either remain constant or be learned. Note that for $\mu/10$ and $\mu/2$, σ is either initialized as a fraction of μ and then kept constant, or evolves with μ. We focus on two key metrics[4]—the *Ex-post+* loss and the number of epochs necessary to converge—shown in Figure 4.9.

The analysis of noise impact on SS-DFL training performance reveals that the *Ex-post+* loss increases with the noise. This is true for both NN1 and NN2, whether σ is constant or variable. Interestingly, setting σ as a fraction of μ (i.e., σ is a fraction of the pre-training value found for each parameter) does not yield better results. Allowing σ to be trainable (variable configuration) does not improve the *Ex-post+* loss. Unlike μ, σ has no clear optimal value. It is a variable without physical reality introduced by SS. Therefore, making σ a parameter to learn complexifies training without a clear gain. Ultimately, it results in a loss of performance.

In both variable and constant configurations, the best performance is achieved with the smallest constant noise ($\sigma = 0.01$), and initializing all the σ's at the same value is to be preferred to fraction of μ.

There is no significant effect of σ on the training convergence speed, as measured by the number of epochs required. Additionally, whether σ is kept

[4]The complete tables are available in the repository.

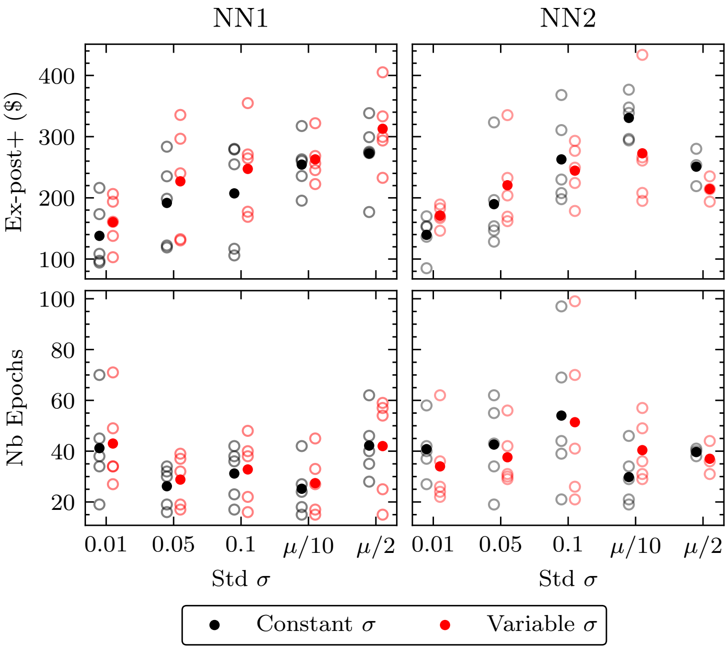


Figure 4.9.: Analysis of *Ex-post+* and the number of epochs for various definition of σ. The mean of the five trainings is in bold. The *Ex-post+* loss tends to increase with higher noises while no clear trend is observe on the number of epochs indicating little effect on the convergence rate.

constant or treated as a trainable parameter does not appear to influence convergence rates. This behavior is consistent across both NN1 and NN2 architectures.

4.3.7. Number of Samples Analysis

We complement our results by analyzing the impact of the sample size parameter S on performance, noting that all previous experiments were conducted with $S = 1$. The motivation for $S = 1$ was threefold. First, it reflects practical deployment conditions in real buildings, where only a single true realization of the building's response is observable at each time step due to the absence of simulators. Second, the score-function (REINFORCE) gradient estimator remains unbiased regardless of S. Increasing S decreases the variance of the gradient estimate but increases computational cost per training instance, leading to longer training times. Third, this choice aligns with the reinforcement learning literature (and the advice of the SS-DFL authors [19]), where policy gradient methods typically use single-sample estimates at each update to balance computational efficiency and estimator variance. Therefore, fixing $S = 1$ provides a practical and theoretically sound approach for our decision-focused learning framework.

In Figure 4.10, we can see the *Ex-post+* loss, the training time and the number of epochs for $S = 1, 2, 5, 10$[5]. All trainings were performed with σ set to its optimal value (i.e., constant $\sigma = 0.01$).The *Ex-post+* loss tends to increase with more samples for both NN1 and NN2. We believe this is because having more samples reduces the exploration during training. As expected, the number of epochs necessary for the training to converge decreases with S, reflecting the higher quality of the gradient estimation. This decrease is extremely steady for NN1, but less for NN2. Despite the number of epochs going down with S, the training time goes up as expected.

In conclusion, $S = 1$ leads to the best decision quality (i.e., minimal *Ex-post+* loss) and training time.

4.4. Conclusion

We presented an HVAC MS where the building thermal dynamics are modeled using NN. We started by improving the formulation of NN as constraints in optimization problem. Then, we learned the parameters of the NN using DFL. In order to ensure meaningful gradient and training robustness, the HVAC MS

[5]The complete tables are available in the repository.

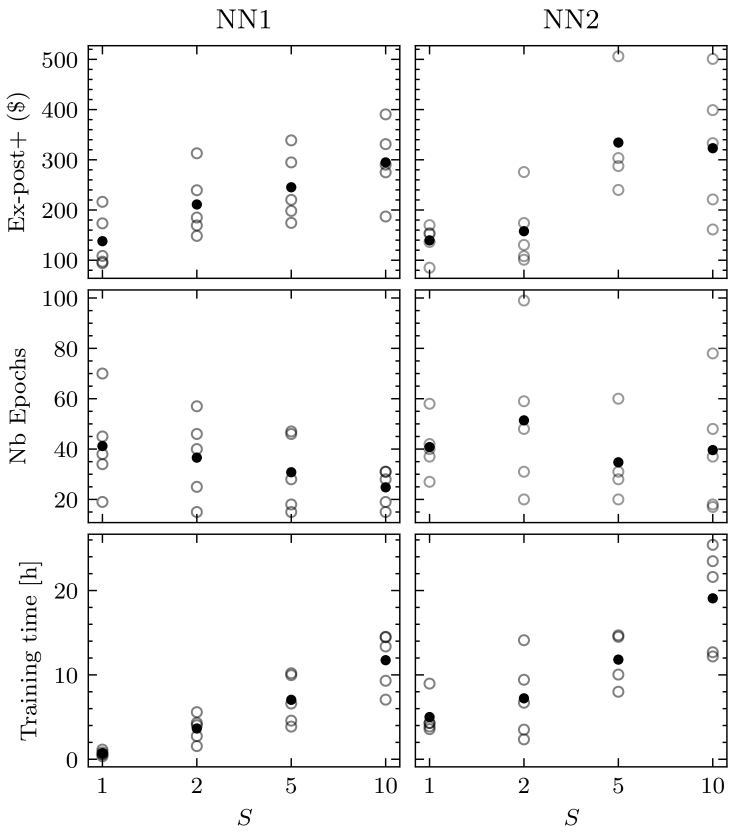


Figure 4.10.: Analysis of *Ex-post+*, the number of epochs, and the training time depending on the number of samples S. The mean of the five trainings is in bold. The *Ex-post+* loss tends to increase with more samples. Despite the number of epochs going down with S, the training time goes up.

is formulated as an MIQP where the thermal comfort is a quadratic penalty in the objective function rather than a hard constraint. Since MIQP are discrete by nature, we employed SS to produce an informative gradient without having to differentiate the MIQP and the building (or its simulator).

We tested our approach on a realistic five-zone building. Results show that SS outperforms the conventional two-stage approach where NN are trained on historical data and then embedded into the optimization. DFL with SS is to be preferred to more naive approaches that relax or fix the binaries of the mixed integer problem. A key insight is ITO models with better accuracy on historical data may lead to worse decisions. These results prove the misalignment of ITO metrics with decision quality.

Further research is required to strengthen scalability of NN-constrained optimization and robustness of the decisions. Developing optimization formulations for entire NN, rather than individual neurons, is an active field of study. We see a significant research opportunity in coupling SS and parameter distribution for stochastic optimization. Ideally, SS would integrate seamlessly with optimization under uncertainty that naturally has parameter distributions. We explore this idea in the next chapter.

CHAPTER 5.

Probabilistic Decision-Focused Learning for Robust Control of Building HVAC Systems

This chapter investigates the interactions between DFL performed with REINFORCE and decision-making under uncertainty. To that end, we make two additions to the conventional RC model. First, we make the RC parameters time-varying to adapt to changing conditions. Second, we expand the linear equation with the forecasted distribution of the temperature perturbation (i.e., the error of the RC model). We start by setting the scope of the chapter, motivating the design choices, and stating the contributions. Section 5.2 presents the global overview of the methodology before section 5.3 dives into the different parts of the decision pipeline. Section 5.4 details the training procedure, which begins with a statistically-driven pre-training phase to initialize the model parameters, followed by a DFL stage that fine-tunes the model based on control performance. Section 5.5 is dedicated to the case study where the impact of the different contributions is analyzed and discussed. Finally, the main conclusions and perspectives are summarized.

5.1. Introduction

5.1.1. Context and motivation

Buildings represent around 30% of the worldwide final energy demand, with approximately 50% devoted to space heating and cooling [131]. Consequently, with the increase of renewable generation, improving the timing and flexibility of thermal energy use is essential to align energy consumption with generation availability. This can be achieved through a reliable control of HVAC systems, which relies on the accurate modeling of building thermal dynamics

[132]. A variety of approaches have been developed to represent these thermal dynamics. Physics-based RC models are interpretable and generalizable, but the internal RC parameters are inherently unknown, and their linear structure prevent to capture nonlinear effects [73]. Then, fully black-box models based on machine learning [29] can capture complex dynamics but often require large datasets and may lack explainability. As a trade-off between both paradigms, grey-box formulations that combine physical insights (RC structure) with a data-driven parameter identification to balance physical model structure with data-driven consistency [84]. In this work, we build upon grey-box RC models to represent the thermal dynamics of multi-zone buildings to support the day-ahead energy optimization of HVAC systems.

5.1.2. Literature review on model identification

Existing data-driven methods for model identification, such as estimating the unknown RC parameters, face three major limitations that hinder their effectiveness for downstream decision-making (e.g., control) tasks.

First, most thermal models are static and context-agnostic, i.e., their parameters remain fixed regardless of changing external conditions [133] although the efficiency of heat pumps depends on the outdoor temperature and the properties of materials, such as their resistance and heat capacitance, depend on the humidity. Therefore, RC model with constant coefficient cannot capture the non-stationary, making the linear approximation accurate only in certain operating regimes while performing poorly in others [14]. To address this, we make the parameters of the RC model context-dependent. At each time step, the RC parameters adjust dynamically to adapt to exogenous conditions (e.g., weather, occupancy). As a result, the nonlinear thermal dynamics are approximated through local, context-driven linearizations, tailored to the expected operating regime.

Second, traditional RC models are deterministic and do not explicitly account for uncertainty [134], i.e., given an HVAC power input, the resulting indoor temperature is assumed to be perfectly known. When such models are embedded within control policies, this assumption can lead to poor performance due to real-world perturbations, such as fluctuating occupancy, window opening, equipment use and their associated heat gains, or other unpredictable disturbances [135]. In response, we model the thermal perturbations as a random variable. We predict the distributions of the thermal perturbations based on contextual inputs (e.g., weather, time of day, expected occupancy). Our formulation captures both epistemic uncertainty (which arises from incomplete knowledge of the building's physical characteristics or modeling simplifications), and aleatoric uncertainty (which reflects inherent variability in the system due to unpredictable factors such as occupant behavior, or window usage).

Third, the data-driven identification of model parameters typically focuses on minimizing a purely statistical loss between the model's predictions and observed thermal dynamics. This approach does not account for how model errors affect downstream decisions, i.e., it does not guarantee that the resulting model performs well in operating conditions that are most relevant for control effectiveness and occupant comfort [98]. To overcome this limitation, we adopt a DFL approach [11]. Instead of training the model to fit past data, DFL directly optimizes the parameters based on their impact on downstream control performance. DFL introduces a level of abstraction that goes beyond statistical capabilities, aligning model training with control performance rather than prediction accuracy alone [136]. We retain the SS-DFL technique of the previous chapter. SS-DFL is based on the score-function (REINFORCE) gradient estimator. This method enables gradient-based training by leveraging stochastic sampling when the loss can be written as an expectation over unknown parameters [19].

5.1.3. Contributions

This chapter proposes to train a forecaster that, for each time step, outputs a point estimate of the RC model parameters and the parameters of the distribution that describe the thermal perturbation. Using a DFL framework, the forecaster learns to provide an optimal estimate of the RC parameters and perturbation distribution with respect to the decision outcome (e.g., ex-post cost). The DFL method builds on the REINFORCE gradient estimator to bypass the non-differentiability of both the mixed-integer HVAC scheduling problem and the building simulator.

The stochasticity of the thermal perturbation plays a dual role in the approach: (*i*) it enables the use of REINFORCE by introducing randomness on the thermal perturbation (without which no gradient can be propagated), and (*ii*) this uncertainty feeds into scenario-based optimization, where sampled thermal perturbations are used to compute HVAC schedules that are robust to uncertainty in building thermal behavior.

The main contributions of the chapter can be summarized as:

1. We develop a multi-zone linear RC model with parameters that are adjusted at each time step and that considers the thermal perturbations. This augmented RC model produces a probabilistic output, explicitly accounting for both model limitations and real-world variability in building

operation. Moreover, the parameters of the RC model and the perturbation distribution are predicted by a forecaster that use contextual information (e.g., weather, expected occupancy) to adapt them to the changing operating conditions.

2. We integrate the augmented RC model into a scenario-based day-ahead HVAC scheduling framework to ensure robustness regarding uncertainty in thermal dynamics. The optimal temperature profile remains feasible over all scenarios of thermal perturbations.

3. The forecaster is trained end-to-end using DFL, where its parameters are learned to optimize control performance rather than prediction accuracy. We use the REINFORCE gradient estimator to enable training through the mixed-integer HVAC optimization and black-box simulator.

4. To improve training efficiency, we first apply standard supervised learning to pre-train the forecaster. This phase uses a physics-informed statistical loss that includes additional regularization terms to encourage physically plausible values for the parameters of the RC model (i.e., the output of the forecaster). The resulting warm start provides a strong initialization for the subsequent DFL phase, improving its convergence.

Scope Clarification

There are four main changes with respect to chapter 4:

- The thermal model is context-aware—it remains a linear RC model but its parameters are adjusted at each time step by a forecaster that adapts the coefficients of the RC model based on contextual information.
- The RC model includes a random variable modeling the thermal perturbation affecting each zone of the building.
- The scheduling problem becomes scenario-based to improve decision robustness.
- We use day-ahead prices to reflect dynamic pricing rather than time-of-use tariffs.

5.2. Overview of the method

Our overarching goal is to learn the parameters of a stochastic multi-zone model that captures a building's thermal dynamics for use in day-ahead HVAC optimization.

5.2.1. Stochastic RC model

To ensure tractability within the optimization, the building dynamics are modeled as linear with respect to HVAC power inputs, following the structure of a Resistance–Capacitance (RC) network, i.e.,

$$\boldsymbol{\tau}_{t+1}^{\text{in}} = \boldsymbol{\tau}_t^{\text{in}} + \boldsymbol{\Delta\tau}_t^{\text{in, RC}} + \epsilon_t, \tag{5.1a}$$

$$\begin{aligned}\text{with} \qquad \boldsymbol{\Delta\tau}_t^{\text{in, RC}} = \mathbf{C}_t^{-1} \odot \Big(&\boldsymbol{\eta}_t^{\text{heat}} \odot \mathbf{p}_t^{\text{heat}} - \boldsymbol{\eta}_t^{\text{cool}} \odot \mathbf{p}_t^{\text{cool}} \\ &+ \mathbf{G}_t^{\text{out}} \odot (\tau_t^{\text{out}}\mathbf{1} - \boldsymbol{\tau}_t^{\text{in}}) \\ &+ \textstyle\sum_{\text{cols}} \Big[\mathbf{G}_t^{\text{in}} \odot \Big(\boldsymbol{\tau}_t^{\text{in}}\mathbf{1}^\top - \mathbf{1}\,{\boldsymbol{\tau}_t^{\text{in}}}^\top\Big)\Big]\Big)\Delta t. \end{aligned} \tag{5.1b}$$

The thermal dynamics model in (5.1b) captures both external and inter-zonal heat exchanges. The state variable $\boldsymbol{\tau}_t^{\text{in}} \in \mathbb{R}^{|\mathcal{Z}|}$ denotes the indoor air temperatures for all zones $z \in \mathcal{Z}$ during time step $t \in \{t_1, \ldots, t_T\}$, while $\tau_t^{\text{out}} \in \mathbb{R}$ is the outdoor temperature. The vectors $\mathbf{p}_t^{\text{heat}}, \mathbf{p}_t^{\text{cool}} \in \mathbb{R}^{|\mathcal{Z}|}$ represent the electrical powers for heating and cooling delivered to each zone, scaled by their respective efficiency coefficients $\boldsymbol{\eta}_t^{\text{heat}}, \boldsymbol{\eta}_t^{\text{cool}} \in \mathbb{R}^{|\mathcal{Z}|}$. The vector $\mathbf{C}_t \in \mathbb{R}^{|\mathcal{Z}|}$ contains the thermal capacitances of each zone, while $\mathbf{G}_t^{\text{out}} \in \mathbb{R}^{|\mathcal{Z}|}, \mathbf{G}_t^{\text{in}} \in \mathbb{R}^{|\mathcal{Z}|\times|\mathcal{Z}|}$ are conductance matrices representing heat transfer with the outside and among zones, respectively. The first three terms in (5.1b) account for thermal inertia, HVAC input, and heat loss to the ambient. The final term models inter-zonal heat exchange by computing the pairwise temperature difference matrix $\boldsymbol{\tau}_t^{\text{in}}\mathbf{1}^\top - \mathbf{1}(\boldsymbol{\tau}_t^{\text{in}})^\top$, applying an element-wise product with $\mathbf{G}_t^{\text{in}}$, and aggregating column-wise to obtain the net thermal interaction vector. The operator $\odot$ denotes element-wise multiplication, and $\mathbf{1}$ is a vector of ones used for broadcasting quantities across zones.

The matrices $\mathbf{C}_t^{-1}$, $\mathbf{G}_t^{\text{out}}$, $\mathbf{G}_t^{\text{in}}$, $\boldsymbol{\eta}_t^{\text{heat}}$ and $\boldsymbol{\eta}_t^{\text{cool}}$, are indexed by time t to reflect their context-dependent nature, i.e., they are dynamically adjusted at each time step $t \in \{t_1, \ldots, t_T\}$ to locally linearize the building's thermal behavior around the expected operating regime. For notational compactness, we aggregate all time-varying parameters of the thermal model into a single

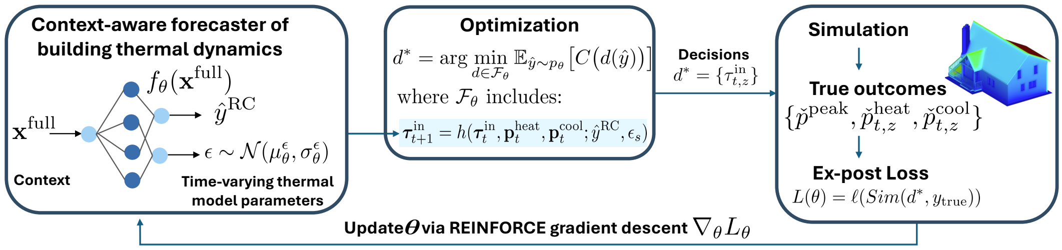


Figure 5.1.: Overview of the methodology.

representation $\hat{y}^{\text{RC}} := \left(\mathbf{C}_t^{-1}, \mathbf{G}_t^{\text{out}}, \mathbf{G}_t^{\text{in}}, \boldsymbol{\eta}_t^{\text{heat}}, \boldsymbol{\eta}_t^{\text{cool}}\right)_{t=t_1}^{t_T}$.

To account for uncertainty in the thermal response (5.1a), the RC model (5.1b) is further enriched with a stochastic perturbation term ϵ_t, representing uncertainty in the thermal response, here modeled as a Gaussian noise $\epsilon_t \sim \mathcal{N}(\mu_t^\epsilon, \sigma_t^\epsilon)$.

5.2.2. Architecture of the DFL training pipeline

We want to learn the weights θ of the forecaster predicting the RC parameters and distribution parameters of the thermal perturbation. We train the forecaster end-to-end with a task loss minimizing day-ahead HVAC energy costs while maintaining occupant comfort. We design a DFL pipeline composed of three stages: (i) a context-aware probabilistic model that forecasts the parameters of both the RC model (5.1b) and the distribution of thermal perturbation ϵ_t, (ii) a scenario-based HVAC scheduling that computes control actions based on sampled thermal perturbations, and (iii) a simulation-based evaluation that quantifies performance in terms of energy cost and comfort.

(i) Probabilistic forecaster — We define a forecaster f_θ that learns to adapt the RC model and the thermal perturbation to time-varying operating conditions based on the current context. The complete input context $\mathbf{x}^{\text{full}} \in \mathcal{X}$ available at the time of day-ahead decision-making, such as past indoor temperatures, HVAC power usage, and weather forecasts, are used as inputs to f_θ, which returns two complementary outputs: (i) the deterministic parameters $\hat{y}^{\text{RC}} = f_\theta^{\text{det}}(\mathbf{x}^{\text{full}})$, which define the nominal linear thermal model for the given context $\mathbf{x}^{\text{full}}$, and (ii) a probabilistic perturbation $\epsilon = \mathcal{N}(\mu^\epsilon(\mathbf{x}^{\text{full}}; \theta), \sigma^\epsilon(\mathbf{x}^{\text{full}}; \theta)) = f_\theta^{\text{prob}}(\mathbf{x}^{\text{full}})$ that represents the uncertainty in the thermal behavior of the building. This context-dependent model serves as a local approximation of the building's thermal response around the expected operating trajectory over the next-day horizon.

***(ii)* Day-ahead HVAC scheduling** — The stochastic RC model (5.1a) is then integrated into a scenario-based optimization framework. To that end, we draw S Monte Carlo samples $\{\hat{y}_s\}_{s=1}^{S}$, where each sample $\hat{y}_s$ combines the deterministic parameters $\hat{y}^{\mathrm{RC}}$ (identical in all samples) with a realization of the stochastic perturbation ϵ_s:

$$\hat{y}_s = \{\hat{y}^{\mathrm{RC}}, \epsilon_s\} \sim p_\theta = \left\{\hat{y}^{\mathrm{RC}}(\mathbf{x}^{\mathrm{full}}; \theta), \mathcal{N}\big(\mu^\epsilon(\mathbf{x}^{\mathrm{full}}; \theta),\, \sigma^\epsilon(\mathbf{x}^{\mathrm{full}}; \theta)\big)\right\}. \quad (5.2)$$

These sampled thermal responses $\hat{y}_s$ represent plausible realizations of the augmented RC parameters under uncertainty. They are used as input scenarios in the downstream HVAC optimization to approximate the expected cost in (5.3) over the distribution of thermal parameters $\hat{y}$ following the conditional predictive distribution $p_\theta(\cdot|\mathbf{x}^{\mathrm{full}})$:

$$d^* = \arg\min_{d\in\mathcal{F}_\theta} \mathbb{E}_{\hat{y}\sim p_\theta}\big[C\big(d(\hat{y})\big)\big] \approx \arg\min_{d\in\mathcal{F}_\theta(\{\hat{y}_s\})} \frac{1}{S}\sum_{s=1}^{S} C\big(d(\hat{y}_s)\big), \quad (5.3)$$

where the decision variable d^* represents the optimal control schedule, including discrete HVAC mode selections (e.g., heating or cooling activation) and temperature setpoints issued to each zone, pertaining to feasible set $\mathcal{F}_\theta(\{\hat{y}_s\})$, while $C(d)$ denotes the cost function of the HVAC scheduling problem (e.g., energy cost or discomfort).

***(iii)* Ex-post evaluation** — The optimal decisions d^* are passed into a non-differentiable simulator Sim, representing the true building response. Here, we use the EnergyPlus software, which yields the true HVAC power consumption resulting from executing the optimal temperature schedule d^* under realistic physical, operational, and occupancy conditions.

Finally, the pipeline is evaluated using the ex-post loss $L(\theta) \in \mathbb{R}$, defined as

$$L(\theta) = \ell\big(Sim(d^*, y_{\mathrm{true}})\big) \quad (5.4)$$

where y_{true} denotes the actual building parameters, and $\ell(\cdot)$ is the true ex-post performance metric (e.g., true costs and comfort deviations as further explained in section 5.3.3).

5.2.3. Challenges of the method

Training over this loss (5.4) presents two main challenges.

First, each sampled realization $\hat{y} \sim p_\theta$ is then passed to an MIQP, whose argmin map $d^*(\hat{y})$ is discrete and non-smooth, i.e., small changes in $\hat{y}$ (caused by updates in θ) can flip binaries and cause abrupt jumps in d^*, such that differentiation through the MIQP is ill-posed.

Second, the simulator *Sim* used to evaluate system performance is non-differentiable, further preventing end-to-end gradient propagation.

To address these difficulties, we estimate gradients using score-function estimators (also known as REINFORCE), as detailed in section 5.4. This enables end-to-end training of the entire DFL pipeline via gradient descent despite the non-differentiabilities in the optimization and simulator.

5.3. Methodology

5.3.1. Context-aware probabilistic forecaster of thermal dynamics

The forecaster is based on an encoder-decoder architecture built upon Long Short-Term Memory (LSTM) cells. LSTM encoder-decoder is a deep learning framework designed to capture temporal dependencies [137].

The multi-step probabilistic forecaster is a parametric model f_θ that maps the full input context $\mathbf{x}^{\text{full}}$, available at the day-ahead decision time t_0 (typically 12 a.m.), to the parameters of a stochastic RC model representing the building's thermal dynamics. The model parameters are time-varying over the prediction horizon $t_1{:}t_T$, allowing the linear RC model to adapt its approximation locally to the evolving operating conditions. Specifically, the model yields:

$$\left\{\hat{y}_t^{\text{RC}}, \mu_t^{\epsilon}, \sigma_t^{\epsilon}\right\}_{t=t_1}^{t_T} = f_\theta\left(\mathbf{x}^{\text{full}}\right), \tag{5.5}$$

The input features $\mathbf{x}^{\text{full}} = \left[\mathbf{x}^{\text{P}}_{t-H:t_0},\ \mathbf{y}_{t-H:t_0},\ \mathbf{x}^{\text{F}}_{t_1:t_T}\right]$ consists of past exogenous inputs $\mathbf{x}^{\text{P}}$ (past weather conditions), past indoor temperature values $\mathbf{y}$, and known future inputs $\mathbf{x}^{\text{F}}$ (weather predictions, calendar information) over the horizon $t_1{:}t_T$.

The forecaster output is the context-dependent parameters of the augmented RC model (5.1), which captures both the nominal thermal dynamics across all zones z of the building and the associated uncertainty:

- the deterministic parameters $\hat{y}_t^{\text{RC}}$ define the nominal linear thermal model for the expected operating regime at time t, and

- the probabilistic perturbation parameters $\mu_t^\epsilon, \sigma_t^\epsilon$ characterize uncertainty in the thermal response via a Gaussian noise $\epsilon_t \sim \mathcal{N}(\mu_t^\epsilon, \sigma_t^\epsilon)$.

Together, these outputs define a context-aware stochastic RC model that approximates the building's thermal dynamics under expected operating conditions. This augmented RC model with time-varying parameters is then passed to the downstream HVAC scheduling optimization, enabling robust control under uncertainty.

5.3.2. Formulation of the HVAC optimization

We consider the day-ahead HVAC scheduling problem for a multi-zone building under time-varying electricity prices, i.e., we shift heating and cooling loads to favorable times (e.g., low electricity prices) while maintaining thermal comfort. These prices are assumed to be deterministic, as they are provided by a building aggregator that coordinates building-level schedules to participate in wholesale electricity markets. We study a general setting in which a building is equipped with multiple Heat Pumps (HPs). Let $\mathcal{H}$ be the set of HPs, and let $\mathcal{Z}_h \subset \mathcal{Z}$ represent the set of zones served by HP $h \in \mathcal{H}$. We assume that each unit independently supplies a specific, non-overlapping subset of thermal zones, i.e., $\bigcup_{h\in\mathcal{H}} \mathcal{Z}_h = \mathcal{Z}$ and $\mathcal{Z}_h \cap \mathcal{Z}_{h'} = \emptyset$ for $h \neq h'$.

The problem is formulated as a scenario-based stochastic program, where $s \in \mathcal{S}$ are the thermal dynamics scenarios with associated probabilities π_s such that $\sum_{s\in\mathcal{S}} \pi_s = 1$. The decision variable in our formulation is a single temperature setpoint schedule $\tau_{t,z}^{\text{in}}$ for each zone z and time step t, which acts as the reference trajectory for the thermostat. This schedule is shared across all scenarios to ensure deployability, as only one control profile can be issued to the building automation system.

The objective function in (5.6) comprises three components: (i) the peak power consumption p_s^{peak} is priced at λ^{peak}; (ii) the power $p_{s,t}^{\text{tot}}$ at each time step t based on time-varying electricity prices λ_t, and (iii) a comfort penalty that penalizes deviations of the indoor temperature $\tau_{s,t,z}^{\text{in}}$ from the zone-specific target $T^{\text{target}}_{t,z}$, weighted by a coefficient $w_{t,z}$.

$$d^* \in \arg\min \underbrace{\sum_{s\in\mathcal{S}} \pi_s p^{\text{peak}}_s \lambda^{\text{peak}}}_{(i)} + \sum_{t=1}^{T} \Bigg(\underbrace{\sum_{s\in\mathcal{S}} \pi_s p^{\text{tot}}_{s,t} \lambda_t}_{(ii)} + \underbrace{\sum_{z\in\mathcal{Z}} w_{t,z} \left(\tau^{\text{in}}_{t,z} - T^{\text{target}}_{t,z} \right)^2}_{(iii)} \Bigg) \Delta t \tag{5.6}$$

The optimization is constrained by the building thermal dynamics modeled by (5.1). The set of initial zonal indoor temperatures $\tau^{\text{in}}_{s,t=0,z}$ are considered to be deterministically known for all zones z.

In centralized HP systems, all zones $z \in \mathcal{Z}_h$ served by the same HP unit $h \in \mathcal{H}$ must operate under a shared mode, either heating, cooling, or idle, at each time step. To model this constraint, we introduce binary variables $m_{t,h} \in \{0, 1\}$, which indicate the operating mode of h at time t, i.e., $m_{t,h} = 1$ corresponds to heating mode, and $m_{t,h} = 0$ to cooling mode. The idle mode is implicitly captured when both the heating and cooling powers are zero, i.e., $p^{\text{heat}}_{s,t,z} = p^{\text{cool}}_{s,t,z} = 0$ for all $z \in \mathcal{Z}_h$ (regardless of the value of $m_{t,h}$).

Given this shared operation, the total heating and cooling power delivered by HP h across its connected zones must respect the unit's aggregate capacity, conditioned on the active mode:

$$\sum_{z\in\mathcal{Z}_h} p^{\text{heat}}_{s,t,z} \leq m_{t,h} \cdot \bar{P}^{\text{heat}}_h, \quad \sum_{z\in\mathcal{Z}_h} p^{\text{cool}}_{s,t,z} \leq (1 - m_{t,h}) \cdot \bar{P}^{\text{cool}}_h \quad \forall s, t, h \tag{5.7}$$

where $\bar{P}^{\text{heat}}_h$ and $\bar{P}^{\text{cool}}_h$ are the maximum heating and cooling power capacities of h, respectively.

In addition, each individual zone z is also constrained by its own maximum heating and cooling power capacities:

$$0 \leq p^{\text{heat}}_{s,t,z} \leq \bar{P}^{\text{heat}}_z \quad \forall s, t, z \tag{5.8}$$

$$0 \leq p^{\text{cool}}_{s,t,z} \leq \bar{P}^{\text{cool}}_z \quad \forall s, t, z \tag{5.9}$$

We ensure power balance in (5.10) by stating that the total HP power consumption $p^{\text{tot}}_{s,t}$ equals the sum of the heating and cooling power delivered to all zones.

$$p_{s,t}^{\text{tot}} = \sum_{z \in \mathcal{Z}} \left(p_{s,t,z}^{\text{heat}} + p_{s,t,z}^{\text{cool}} \right) \quad \forall s, t \tag{5.10}$$

The building's peak power demand p_s^{peak} must not exceed the grid connection limit $\bar{P}^{\text{grid}}$, which defines the maximum power capacity available from the electric utility:

$$p_{s,t}^{\text{tot}} \leq p_s^{\text{peak}} \leq \bar{P}^{\text{grid}} \quad \forall s, t \tag{5.11}$$

5.3.3. Building simulator and Ex-post loss

The optimized temperature schedule is deployed in the EnergyPlus simulator to evaluate its real-world feasibility and performance [17]. EnergyPlus is widely regarded as the reference tool for high-fidelity building energy simulation. It is developed and maintained by the U.S. Department of Energy (DOE), which uses it for for developing and evaluating national building energy codes and efficiency standards.

EnergyPlus receives the zone-level temperature setpoints $\tau_{t,z}^{\text{in}}$ as control inputs and simulates the building's true thermal and operational response based on high-fidelity physical modeling and actual occupancy. The simulator attempts to follow the setpoints within physical limits, returning the actual heating and cooling power consumption $\check{p}_t^{\text{tot}}$ and any comfort violations due to unachievable targets. This enables the computation of the ex-post performance metric $\ell\big(Sim(d^*, y_{\text{true}})\big)$, as defined in (5.4), which captures both realized energy cost and comfort deviations.

The loss combines three components, i.e., (i) the realized energy cost, computed from the true electricity prices $\check{\lambda}_t$ and power consumption $\check{p}_t^{\text{tot}}$, (ii) the thermal comfort violations based on the deviation from the target indoor temperatures, and (iii) a disappointment penalty, which quantifies the mismatch (regret) between the expected and actual cost. Together, the first two components form the ex-post objective value of the control policy. We refer to this loss as *Ex-post+*. It is here defined as:

$$\ell(\cdot) = \sum_{t=t_0}^{t_T} \Bigg(\underbrace{\check{p}_t^{\text{tot}} \check{\lambda}_t}_{\text{Ex-post cost}} + \underbrace{\sum_{z=0}^{Z} w_{t,z} (\Delta \check{\tau}_{t,z}^{\text{in}})^2}_{\text{Comfort deviation}} \Bigg) + \underbrace{\text{RMSE}(\check{C} - C)}_{\text{RMSE Power Cost}}, \tag{5.12}$$

where $\Delta \check{\tau}_{t,z}^{\text{in}} = \check{\tau}_{t,z}^{\text{in}} - T_{t,z}^{\text{target}}$ and $\text{RMSE}(\check{C} - C) = \sqrt{\frac{1}{T} \sum_{t=t_0}^{t_T} (\check{p}_t^{\text{tot}} \check{\lambda}_t - p_t^{\text{tot}} \lambda_t)^2}$.

5.4. Training procedure

Before applying DFL in section 5.4.2, the forecaster f_θ is first pretrained using a traditional statistical learning procedure in section 5.4.1. This pretraining step provides a good initialization for the model parameters, improving both convergence stability and computational efficiency during DFL optimization.

5.4.1. Warm-start with physics-informed statistical loss

The proposed forecaster f_θ is trained to predict both deterministic parameters $\hat{y}^{\mathrm{RC}}$ and probabilistic perturbation terms $\mu^\epsilon, \sigma^\epsilon$.

To reflect the physical constraints of the system, the output layer enforces positivity on all physical quantities by applying a ReLU activation to $\hat{y}^{\mathrm{RC}}, \sigma^\epsilon$. This ensures that conductances, capacitances, efficiencies, and the standard deviation of the noise remain strictly non-negative. In addition, we force the matrix G^{int} to be null on its diagonal (no self-conductance). The mutual conductance must also be null if two zones do not share a wall or a ceiling. The resulting matrix must be symmetrical (heat transfer from zone z to z' has the same conductance as from z' to z). These three constraints (diagonal null, mutual conductance null for non-adjacent zones, symmetry) reduce the number of parameters to learn in G^{int} from 225 to 24.

The predicted model parameters are used to compute a probabilistic forecast of indoor air temperatures $\hat{\tau}^{\mathrm{in}}_{t,z}$ in all zones via the thermal dynamics model in (5.1). An intuitive approach would be to use the Negative Log-Likelihood (NLL) loss on the predicted temperature distributions. However, this loss disregards y^{RC} and learns only $\mu^\epsilon, \sigma^\epsilon$, as we proof in Appendix C. Therefore, the model specificity requires to craft a loss encompassing the multiple objectives:

1. $\hat{y}^{\mathrm{RC}}$ must provide a physically meaningful RC model,
2. $\hat{y}^{\mathrm{RC}}$ should evolve smoothly over time,
3. $\mu^\epsilon, \sigma^\epsilon$ must capture the probabilistic perturbation.

First, we use an MSE loss on the model prediction (without residual) $\hat{\tau}^{\mathrm{in}}_{\mathrm{RC}}$ with respect to the target τ^{in} to train the RC parameters. Second, we add a regularization term that penalizes temporal variability in the RC parameters, encouraging physically consistent and smoothly changing dynamics. Third, we assess how well ϵ captures both accuracy and uncertainty by employing a NLL loss between the predicted Gaussian distribution $\mu^\epsilon, \sigma^\epsilon$ and the indoor temperature perturbations that cannot be explained by the RC model alone $\epsilon^{\mathrm{true}} = \tau^{\mathrm{true}} - \hat{\tau}^{\mathrm{in}}_{\mathrm{RC}}$. The resulting loss is

$$\mathcal{L}_{\mathrm{WS}}(\theta) = \mathbb{E}_{t,z}\left[\mathrm{MSE}(\tau^{\mathrm{true}}_{t,z}, \hat{\tau}^{\mathrm{in}}_{\mathrm{RC,\ t,z}}) + \mathrm{CV}\big(\hat{y}^{\mathrm{RC}}\big) - \log \mathcal{N}\big(\epsilon^{\mathrm{true}}_{t,z} \,\big|\, \hat{\mu}^{\epsilon}_{t,z}, (\sigma^{\epsilon}_{t,z})^2\big)\right], \tag{5.13}$$

where $\tau^{\mathrm{true}}_{t,z}$ denotes the observed indoor temperature in zone z at time t, $\hat{\tau}^{\mathrm{in}}_{\mathrm{RC},t,z}$ denotes the predicted indoor temperature by the RC model without residual in zone z at time t, and $\mathcal{N}(\cdot \mid \mu, \sigma^2)$ represents the Gaussian likelihood parameterized by the predicted mean $\hat{\mu}^{\epsilon}_{t,z}$ and variance $(\sigma^{\epsilon}_{t,z})^2$. The term $\mathrm{CV}(\hat{y}^{\mathrm{RC}})$ is the sum of the temporal coefficient of variation defined as the ratio of the standard deviation and the absolute mean, $CV = \sigma/|\mu|$, of all the RC parameters (except for the residual), which enforces that they vary smoothly over time.

This formulation penalizes both systematic bias (through the mean prediction) and miscalibrated uncertainty (through the coefficient of variation), encouraging the model to accurately represent the stochastic variability of the thermal dynamics.

5.4.2. DFL training

We optimize the values of θ that parametrize the forecaster f_θ of the building thermal dynamics. Here, $x = \mathbf{x}^{\mathrm{full}}$ and y_{true} are two correlated random vectors characterized by their joint distribution $\mathcal{P}_{(x,y_{\mathrm{true}})}$. Practically, the true distribution $\mathcal{P}_{(x,y_{\mathrm{true}})}$ is unknown, and it is approximated via a training set $\mathcal{D} = \{x_i, y_{\mathrm{true}_i}\}_{i=1}^{N}$. The DFL training problem can thus be formulated as :

$$\begin{aligned}&\min_{\theta} \mathbb{E}_{(x,y_{\mathrm{true}})\in\mathcal{D}}\big[L(\theta)\big]\\ &\qquad = \min_{\theta} \mathbb{E}_{(x,y_{\mathrm{true}})\in\mathcal{D}}\big[\ell\big(Sim(z^*(\hat{y}), y_{\mathrm{true}})\big)\big].\end{aligned} \tag{5.14}$$

This problem is solved using gradient descent, i.e., an iterative algorithm that adjusts θ to minimize the expected return $L(\theta)$. The update rule for gradient descent is:

$$\theta \leftarrow \theta - \alpha \nabla_\theta \mathbb{E}_{(x,y_{true})\in\mathcal{D}}\big[L(\theta)\big]. \tag{5.15}$$

where α is the learning rate, controlling the step size of each update. The procedure requires to compute the gradient of the expected loss, which represents the direction in the parameter space Θ that increases $L(\theta)$ the most. Since the distribution of data $\mathcal{D}$ is fixed and independent of θ, the gradient of the expected loss reduces to an average of per-sample gradients over the entire training set:

$$\nabla_\theta \mathbb{E}_{(x,y_{true})\in\mathcal{D}}\big[L(\theta)\big] = \mathbb{E}_{(x,y_{true})\in\mathcal{D}}\big[\nabla_\theta L(\theta)\big] \tag{5.16}$$
$$= \frac{1}{|\mathcal{D}|}\sum_{(x,\,y_{\text{true}})\in\mathcal{D}}\big[\nabla_\theta L(\theta)\big]. \tag{5.17}$$

In practice, evaluating this full sum at every update is computationally expensive. Therefore, we approximate it using mini-batch stochastic gradient descent, where at each iteration the gradient is estimated from a small randomly sampled subset $\mathcal{B}$ of $\mathcal{D}$.

The non-differentiability of both the building simulator $\mathcal{S}$ and the mixed-integer optimization problem (5.3) prevents the use of backpropagation-based training. Instead, we leverage the score function (REINFORCE) gradient estimator, which provides unbiased gradient estimates in absence of differentiability, provided the loss function can be expressed as an expectation over a differentiable probability distribution parameterized by θ. Therefore, we build $p'_\theta(\cdot|\mathbf{x}^{\text{full}})$, defined as

$$p'_\theta(\cdot|\mathbf{x}^{\text{full}}) = \big\{\mathcal{N}\big(\hat{y}^{\text{RC}}(\mathbf{x}^{\text{full}};\theta), \sigma^{\text{RC}}\big), \mathcal{N}\big(\mu^{\epsilon}(\mathbf{x}^{\text{full}};\theta),\, \sigma^{\epsilon}(\mathbf{x}^{\text{full}};\theta)\big)\big\}\,, \tag{5.18}$$

where σ^{RC} is arbitrarily fixed to provide a distribution over $\hat{y}^{\text{RC}}$ that enables the use of REINFORCE to learn θ.

$$\nabla_\theta L(\theta) = \nabla_\theta \mathbb{E}_{\hat{y}\sim p'_\theta}\big[\ell(Sim(d^*(\hat{y}), y_{\text{true}}))\big]$$
$$= \mathbb{E}_{\hat{y}\sim p'_\theta}\big[\ell(Sim(d^*(\hat{y}), y_{\text{true}}))\cdot\nabla_\theta \log p'_\theta(\hat{y})\big] \tag{5.19}$$

where the expectation $\mathbb{E}_{\hat{y}\sim p'_\theta}$ is estimated using $i = \{1,\ldots,N\}$ Monte-Carlo samples from the predicted distribution p'_θ, i.e.,

$$\nabla_\theta L(\theta) \approx \frac{1}{N}\sum_{i=1}^{N}\Big[\ell\big(Sim\big(d^*(\hat{\mathbf{y}}^{(i)}),\, y_{\text{true}}\big)\big)\cdot\nabla_\theta \log p'_\theta\big(\hat{\mathbf{y}}^{(i)}\big)\Big]. \tag{5.20}$$

In practice, each REINFORCE sample $\hat{\mathbf{y}}^{(i)}$ is passed to the downstream optimization problem. To capture uncertainty in thermal dynamics, such as human occupancy or exogenous heat gains, we employ a scenario-based optimization approach that leverages the distributional information encoded in p'_θ. Hence, we treat each $\hat{\mathbf{y}}^{(i)}$ as an S-dimensional vector $\hat{\mathbf{y}}^{(i)} = (\hat{y}^{(i)}_1,\ldots,\hat{y}^{(i)}_S)$, where each component $\hat{y}^{(i)}_s \sim p_\theta(\cdot|\mathbf{x}^{\text{full}})$ is drawn independently to represent one possible scenario of building dynamics. Under this conditional independence assumption, the joint probability decomposes as

$$p'_\theta\big(\hat{\mathbf{y}}^{(i)}|\mathbf{x}^{\text{full}}\big) = \prod_{s=1}^{S} p_\theta\big(\hat{y}_s^{(i)}|\mathbf{x}^{\text{full}}\big), \tag{5.21}$$

and the corresponding score term in the REINFORCE gradient becomes a sum of individual log-probability gradients:

$$\nabla_\theta \log p'_\theta\big(\hat{\mathbf{y}}^{(i)}|\mathbf{x}^{\text{full}}\big) = \sum_{s=1}^{S} \nabla_\theta \log p_\theta\big(\hat{y}_s^{(i)}|\mathbf{x}^{\text{full}}\big). \tag{5.22}$$

Substituting this into Eq. (5.20) yields the Monte Carlo estimator

$$\begin{aligned}\nabla_\theta L(\theta) \approx \frac{1}{N} \sum_{i=1}^{N} &\ell\Big(Sim\big(d^*(\hat{\mathbf{y}}^{(i)}),\, y_{\text{true}}\big)\Big) \\ &\cdot \left(\sum_{s=1}^{S} \nabla_\theta \log p_\theta\big(\hat{y}_s^{(i)}|\mathbf{x}^{\text{full}}\big)\right),\end{aligned} \tag{5.23}$$

where $d^*(\hat{\mathbf{y}}^{(i)})$ is the optimal decision for the i-th Monte Carlo sample, obtained by solving the scenario-based optimization over the S sampled scenarios of thermal dynamics.

The resulting training is described in Algorithm 2.

Algorithm 2 DFL training with REINFORCE gradient estimator

Initialize parameters θ from warm-start
Initialize best validation loss $L_{\text{best}} \leftarrow +\infty$
for each epoch **do**
  for each training mini-batch $\mathcal{B}$ **do**
    Initialize gradient accumulator $g \leftarrow 0$
    for each $(x, y_{\text{true}}) \in \mathcal{B}$ **do**
      for $i = 1$ to N **do**
        Sample scenarios $\hat{\mathbf{y}}^{(i)}$ from $p_\theta(\cdot \,|\, \mathbf{x}^{\text{full}})$
        $d^*(\hat{\mathbf{y}}^{(i)}) \leftarrow$ solve scenario-based optimization
      end for
      $\nabla_\theta L(\theta) \leftarrow$ compute the gradient estimator using (5.23)
      $g \leftarrow g + \nabla_\theta L(\theta)$
    end for
    Update $\theta \leftarrow \theta + \alpha \cdot g$
  end for
  Compute validation loss
  Apply early stopping after 10 epochs without improving L_{best}.
end for

5.5. Case study

We evaluate the performance of the proposed DFL pipeline for identifying parameters of a building's thermal model on a medium office building (4,928m^2) located in Denver, CO (USA). The building is modeled using EnergyPlus [17], a high-fidelity, physics-based building simulator [22]. The building, already used in section 3.5.1, includes three floors, each of which being composed of 5 zones that are jointly thermally controlled by a single HVAC system. The simulations are conducted over a set of representative days, selected via clustering from a full-year dataset (see section 5.5.1). The building follows a typical office occupancy schedule, with most activity occurring on weekdays between 7a.m. and 6p.m.

All computations are carried out on a MacBook Pro with M1 Pro chip and 32GB of RAM.

5.5.1. Data & Clustering

Regarding weather, we use two hourly datasets representing a typical meteorological year for Denver. The first one is used for pre-training. The second one is used for DFL. Regarding price signals, we use the day-ahead market price of Belgium for 2024 available on the website of ENTSEO-e.

Since each forward pass in DFL is time-consuming, we want to maximize the information of each sample (i.e., a day). To that end, we cluster the data following a two-step procedure. First, we identify extreme scenarios. These extreme scenarios will be fixed centroids. Second, we run a k-medoid algorithm to partition the data between representative clusters and extreme scenario clusters.

We cluster with respect to the metric k. The metric k combines two components with high impact on our model accuracy and decisions: the weather and the electricity price. The weather component considers the outdoor temperature and the total horizontal solar radiation G. The distance between two days is computed as the Euclidean distance over the hourly vector $k \in \mathbb{R}^{24}$ defined as:

$$k = \underbrace{0.8\tilde{\tau}^{\text{out}} + 0.2\tilde{G}}_{Weather} + \underbrace{\tilde{\lambda}}_{Elec.\ Prices}, \tag{5.24}$$

where all variables $\tilde{\cdot}$ are normalized.

We set six extreme scenarios before starting the k-medoid algorithm: the days with the lowest and highest electricity prices, the days with the lowest and highest weather component, the days with the highest price variance, and weather component variance. It turns out that the day with the highest price variance is the same as the day with the highest price. Consequently, the number of extreme scenarios decreases to five.

Then, we run k-medoid algorithm to find ten clusters. Five medoids are fixed, they are the extreme days. Five extra medoids are randomly initialized. Figure 5.2 gives information about the medoids. The number of samples per cluster ranges from 2 to 68. Interestingly, the cluster with the maximum price is associated with very low outdoor temperature and sun. In contrast, the cluster associated with the minimum price is associated with mild temperature (average of about 20°C) and abundant sunshine.

During DFL training, the loss of each medoid is multiplied by the representativeness of this medoid (i.e., the probability that a day picked randomly belongs to the cluster associated with the medoid). This forces the model to focus on the representative days that are more likely to occur.

For validation, we randomly sample one day per cluster without replacement (i.e., the sampled day is not returned to the database). For testing, we again sample one day per cluster.

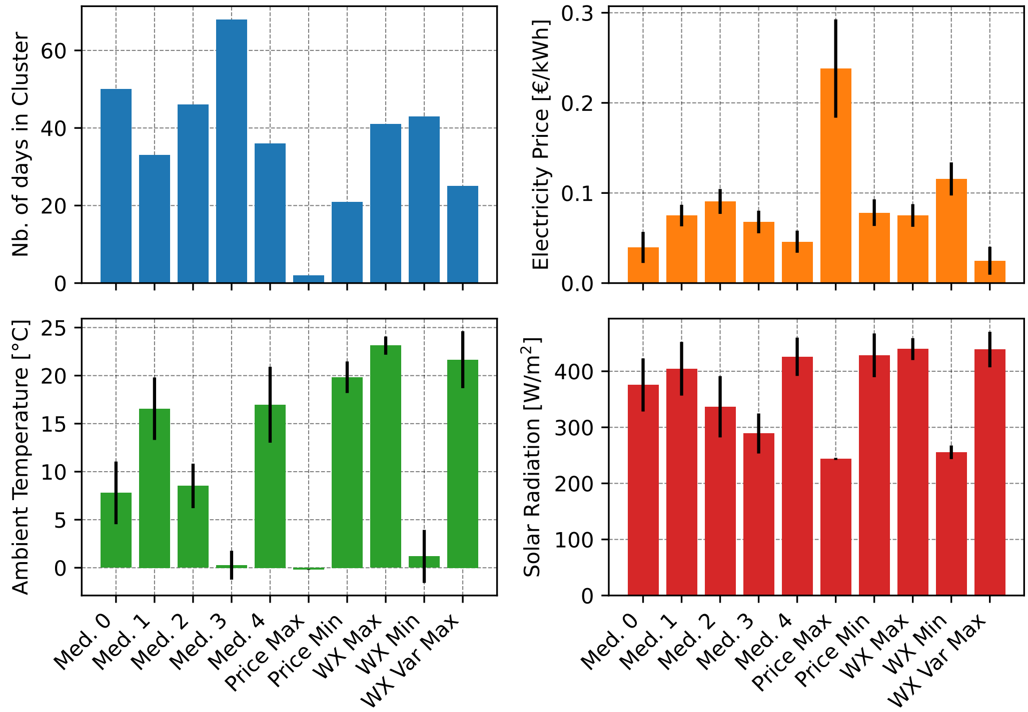


Figure 5.2.: Number of samples per cluster, as well as average and standard deviation for the electricity price, outdoor temperature, and horizontal solar radiations.

5.5.2. Benchmarks

We compare our reference stochastic dynamic spatial RC model (with positivity constraints) to four static RC model:

1. The first benchmark, RC, is a conventional RC model in its simplest form. The model has 60 unconstrained parameters: $\mathbf{C}^{-1}, \boldsymbol{\eta}^{\text{heat}}, \boldsymbol{\eta}^{\text{cool}}, \mathbf{G}^{\text{out}} \in \mathbb{R}^{|\mathcal{Z}|=15}$, and is defined as:

$$\begin{aligned}\boldsymbol{\tau}_{t+1}^{\text{in}} = \boldsymbol{\tau}_t^{\text{in}} + \mathbf{C}^{-1} \odot \Big(&\boldsymbol{\eta}^{\text{heat}} \odot \mathbf{p}_t^{\text{heat}} - \boldsymbol{\eta}^{\text{cool}} \odot \mathbf{p}_t^{\text{cool}} \\ &+ \mathbf{G}^{\text{out}} \odot (\tau_t^{\text{out}}\mathbf{1} - \boldsymbol{\tau}_t^{\text{in}})\Big)\Delta t.\end{aligned}$$

2. For the second benchmark, RC^+, we impose the positivity of the parameters to reflect their physical meaning.
3. The third benchmark, $\text{RC}_{G^{\text{int}},\epsilon}$, is the spatial version of RC with a residual. The model equation is similar to (5.1b) but ϵ is deterministic (no σ_ϵ). The number of parameters rises to 300, including 225 for the unconstrained indoor conductance matrix G^{int}.
4. The last benchmark, $\text{RC}^+_{G^{\text{int}},\epsilon}$, is the constrained version of $\text{RC}_{G^{\text{int}},\epsilon}$. All parameters must be positive and the number of non-zero elements in G^{int} is 24 by imposing symmetry, zeros on the diagonal and mutual conductance of non-adjacent zones. For more details, refer to subsection 5.4.1. The total number of parameters is 99.

First, we analyze the performance of the thermal models on historical data. Second, we evaluate the quality of the control policy based on these thermal models without further training (no DFL). Third, we assess the control policy after refining the models with DFL. For these three steps, no stochasticity is involved. Lastly, we evaluate the performance of DFL model when sampling 2, 5, and 10 scenarios for robust optimization.

5.5.3. Pre-training

In this section, we analyze the performance of the RC models trained, validated, and tested on the first dataset made of one year of historical data. The maximum number of epochs is set to 200 with an early stopping patience of 10 epochs if there is no improvement on the validation. Since the RC model finality is to be embedded in a day-ahead problem, we report the metrics over 24h-ahead forecasts in Table 5.1.

Model	RC	RC^+	$\text{RC}_{G^{\text{int}},\epsilon}$	$\text{RC}^+_{G^{\text{int}},\epsilon}$	$\text{DRC}^+_{G^{\text{int}},\epsilon}$
Loss	MSE	MSE	MSE	MSE	(5.13)
RMSE [°C]	1.78	1.77	1.27	1.63	0.27
MAE [°C]	1.30	1.33	0.85	1.20	0.15
MAPE [%]	5.90	6.02	3.91	5.42	0.70
R^2	0.32	0.34	0.66	0.44	0.98
Nb. Epochs	20	71	68	58	36
Time [s]	7	20	20	19	882

Table 5.1.: Test metrics for the pre-training on one year of historical data (i.e., rule-based control). The best model is by far the proposed context-aware dynamic RC (DRC) model despite parameters constrained to be positive to reflect physics and sparse G^{int} indicated by $.^+$. Indices G^{int} and ϵ indicate if the model has a spatial conductance matrix and a residual, respectively.

First, we compare the four static RC models; their parameters are constant over time. We start with the conventional RC model. The R^2 is 0.32 and the RMSE is 1.78°C. The second model RC^+ forces the parameters to be positive, which does not degrade performance. RC^+ is even slightly better. Albeit counterintuitive, the added information about the physics of the problem explains the improved performances in the day-ahead forecast. The third model $\text{RC}_{G^{\text{int}},\epsilon}$ has no positivity constraint but features a perturbation vector ϵ and inter-zonal admittance matrix G^{int}. Among the static models, it reaches the highest R^2 at 0.66 and lowest error metrics. Unlike for the non-spatial RC model, adding constraints on the RC parameters (not on ϵ) reduces the performances. This is because, in addition to positivity constraints, the number of non-zero elements in the indoor admittance matrix G^{int} falls from 225 to 24, which reduces the modeling power and thus the prediction accuracy.

The last column of Table 5.1 reports the performance of the dynamic RC (DRC) model. $\text{DRC}^+_{G^{\text{int}},\epsilon}$ exhibits excellent accuracy with an R^2 at 0.98 and an RMSE at 0.27°C. $\text{DRC}^+_{G^{\text{int}},\epsilon}$ outperforms all models significantly. This is due to the probabilistic forecaster ability to adapt the RC model parameters to the conditions along with the perturbation ϵ.

Regarding training time, the simple RC model takes 7 seconds while more evolved static versions need about 20 seconds. This can be explained by the lower number of epoch for the simple RC model to converge. $\text{DRC}^+_{G^{\text{int}},\epsilon}$ takes about 15 minutes despite a low number of epochs before convergence. The complexity of the model is responsible for this longer time. Because pre-training is performed once and off-line, training time is not critical.

5.5.4. ITO Results

Table 5.2 presents the optimization outcome when thermal models are plugged as a constraint directly after pre-training. This is the conventional two-stage ITO. To give an order of magnitude, the day-ahead scheduling problem has about 10,000 constraints, 10,000 variables, and 1300 parameters.

Metric	RC	RC^+	$RC_{G^{\text{int}},\epsilon}$	$RC^+_{G^{\text{int}},\epsilon}$	$DRC^+_{G^{\text{int}},\epsilon}$
Ex-post+	145.8	173.0	304.0	169.8	169.6
Ex-post Obj. Val.	95.8	89.0	237.8	96.1	103.1
Ex-post Power Cost	67.9	86.6	142.0	92.9	96.1
Ex-post Temp Penalty	27.9	2.4	95.8	3.2	7.0
Expected Obj. Val.	48.0	13.6	730.4	27.6	45.4
Expected Power Cost	22.0	12.2	97.0	25.5	39.0
Expected Temp Penalty	26.0	1.4	633.4	2.1	6.4
RMSE Cost	49.9	84.0	66.2	73.7	66.5

Table 5.2.: Optimization results of the ITO approach (no DFL).

Out of the five models, four have their weighted average ex-post objective value around €100 per day. Only the model $RC_{G^{\text{int}},\epsilon}$, which was the best static model on historical data, collapses entirely with an ex-post objective value of €237.8. The absence of physics-information led to a model with no robustness to regime chance and distribution shift. The four other models (RC, RC^+, $RC^+_{G^{\text{int}},\epsilon}$, $DRC^+_{G^{\text{int}},\epsilon}$) perform well. Each has upsides and downsides.

The RC model has the lowest ex-post power cost and most accurate power cost prediction (i.e., minimal Root Mean Squared Error (RMSE) on cost) but the highest thermal discomfort penalty.

The constrained RC model RC^+ has the lowest ex-post objective value at €89.0 per day. However, it has the greatest power cost misestimation with an RMSE on the cost at €84.

$RC^+_{G^{\text{int}},\epsilon}$ outperformed its unconstrained counterpart $RC_{G^{\text{int}},\epsilon}$. It has balanced performances with an ex-post objective value at €96.1, and a low thermal discomfort at €3.2.

The dynamic model $DRC^+_{G^{\text{int}},\epsilon}$ looses its advantage over the static models (except $RC_{G^{\text{int}},\epsilon}$). The complete change of dynamics due to demand response creates new links between contextual information (such as electricity price) and the operating points that the LSTM encoder-decoder did not learn. $DRC^+_{G^{\text{int}},\epsilon}$

has a high ex-post value at €103.1 driven by both high HVAC power cost at €96.1 and thermal discomfort at €7.0. The power cost RMSE is at €66.5.

Important conclusions can be drawn from these results. First, the forecasting performance of a model on historical data does not reflect once this model is part of a control problem. Second, all models underestimate the ex-post cost because optimization exploits the model inaccuracies that help improve the objective value.

5.5.5. Deterministic DFL Results

This section reports the results of the deterministic models after DFL training. For each thermal model, nine trainings were performed with different hyperparameters. The set of possible hyperparameters $\mathcal{H}$ is obtained by cartesian product of three standard deviation for SS on all deterministic parameters $\hat{y}^{\mathrm{RC}}$ and three learning rates:

$$\mathcal{H} = \{0.001, 0.1, \frac{\mu}{10}\}_\sigma \times \{0.0005, 0.001, 0.005\}_\alpha \tag{5.25}$$

Table 5.3 reports the models with the lowest validation loss.

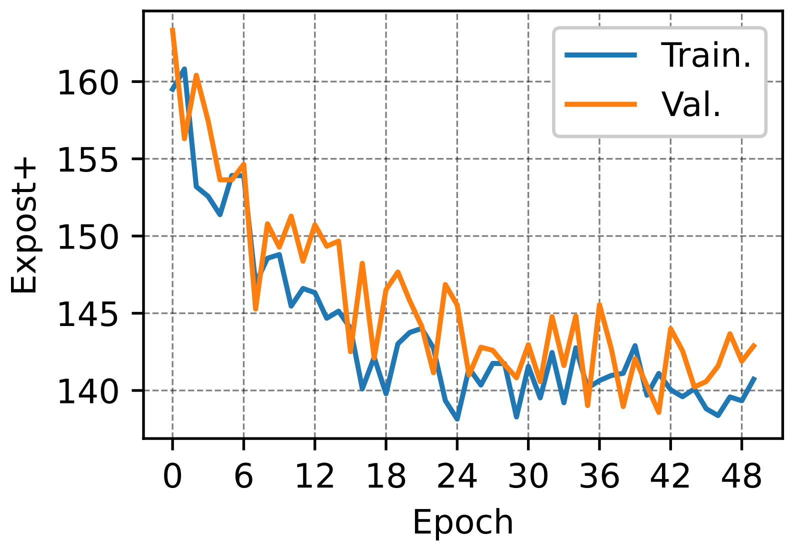


Figure 5.3.: Training and validation loss curves for the $\mathrm{DRC}^+_{G^{\mathrm{int}},\epsilon}$ model.

DFL optimizes for the *Ex-post+* loss (5.12). Overall, the results after DFL remain similar to those of the warm-start, especially for RC, RC^+, and $\mathrm{RC}^+_{G^{\mathrm{int}},\epsilon}$.

Metric	RC		RC^+		$RC_{G^{\text{int}},\epsilon}$		$RC^+_{G^{\text{int}},\epsilon}$		$DRC^+_{G^{\text{int}},\epsilon}$	
	ITO	DFL	ITO	DFL	ITO	DFL	ITO	DFL	ITO	DFL
Ex-post+	145.8	146.1	173	172	304	293.3	169.8	170.1	169.6	154.6
Ex-post Obj. Val.	95.8	97.7	89	88.6	237.8	212.2	96.1	96.8	103.1	99.6
Ex-post Power Cost	67.9	67.4	86.6	86.2	142	146.4	92.9	93.6	96.1	94.2
Ex-post Temp Penalty	27.9	30.3	2.4	2.4	95.8	65.8	3.2	3.2	7	5.4
Expected Obj. Val.	48	51.9	13.6	13.5	730.4	479.5	27.6	28.2	45.4	57.7
Expected Power Cost	22	23.3	12.2	12.1	97	108.1	25.5	26.1	39	53.3
Expected Temp Penalty	26	28.6	1.4	1.4	633.4	371.4	2.1	2.1	6.4	4.4
RMSE Cost	49.9	48.4	84	83.4	66.2	81.1	73.7	73.3	66.5	55
Nb. Epochs	-	29	-	13	-	15	-	22	-	49
Trng Time [s]	-	3258	-	1876	-	4424	-	4597	-	5029
Test Time [s]	-	244	-	237	-	140	-	207	-	105

Table 5.3.: Test results of the DFL models (and ITO models for ease of comparison) without stochasticity ($\epsilon = \mu_\epsilon$).

Only two models, $\mathrm{RC}_{G^{\mathrm{int}},\epsilon}$ and $\mathrm{DRC}^{+}_{G^{\mathrm{int}},\epsilon}$, see a significant improvement because of training. The change in $\mathrm{RC}_{G^{\mathrm{int}},\epsilon}$ is mostly due to the very poor warm-start making it simple to find better parameters. The improvement of $\mathrm{DRC}^{+}_{G^{\mathrm{int}},\epsilon}$ is more interesting. The *Ex-post+* loss falls from about 170 to 155, a reduction of roughly 9%. Both components of *Ex-Post+*, the ex-post objective value and the RMSE on the cost, decrease.

We show the training and validation curves of the model $\mathrm{DRC}^{+}_{G^{\mathrm{int}},\epsilon}$ in Figure 5.3. The training and validation losses follow very similar trends. The overall shape is the one of a negative exponential as expected. It is somewhat noisier than supervized learning curve which is normal for learning with REINFORCE. However, the results of the simplest model, RC, show that a better optimum in terms of *Ex-Post+* is possible. The models $\mathrm{RC}_{G^{\mathrm{int}},\epsilon}$ and $\mathrm{DRC}^{+}_{G^{\mathrm{int}},\epsilon}$ should be able to reach at least that optimum. We believe that the signal of the gradient is not strong enough to effectively guide the model, thereby making the warm-start the decisive factor. Further work should try to increase the batch size, and the number of samples N to improve the gradient quality. Better balancing exploration and exploitation could also be beneficial. For instance, the default standard deviation σ^{RC} for the RC parameters could decrease during training thereby forcing exploration in the early stages before exploitation.

5.5.6. Probabilistic DFL Results

In Table 5.4, very preliminary results of scenario-based robust optimization are presented. We evaluated the models for 2, 5, and 10 scenarios. Note that all models have the same warm-start presented in Table 5.2. Then, DFL trains each model considering the number of scenarios.

Overall, the loss *Ex-post+* decreases with the number of scenarios. From 2 to 5 scenarios, the decrease is driven by a diminution of the power cost. From 5 to 10 scenarios, the decrease is driven by a reduction of the comfort penalty. However, stochastic DFL optimization underperforms deterministic DFL optimization.

Adding scenarios increases the computation time. Test time went from about 100 seconds for optimization and simulation in the deterministic approach, to roughly 500 seconds for 2 scenarios, 800 for 5 scenarios, and 1000 for 10 scenarios. Since the simulation time remains the same, the increase is due to the time to solve the optimization problem. Training time increases to about four hours. Interestingly, training converges with a reducing number of epochs when the number of scenarios increases.

Nb. Scenarios S	2	5	10	Det.
Ex-post+	167.7	166.9	161	154.6
Ex-post Obj. Val.	103.7	103.3	99	99.6
Ex-post Power Cost	97.3	94.3	94.3	94.2
Ex-post Temp Penalty	6.4	9	4.7	5.4
Expected Obj. Val.	50.7	50.5	46.3	57.7
Expected Power Cost	44.6	41.5	42.8	53.3
Expected Temp Penalty	6.1	9	3.5	4.4
RMSE Cost	64	63.6	62	55
Nb. Epochs	34	13	11	49
Trng Time [s]	10783	16823	14335	5029
Test Time [s]	487	808	983	105

Table 5.4.: Test of the DFL $\mathrm{DRC}^{+}_{G^{\mathrm{int}},\epsilon}$ model with robust optimization built on 2, 5, and 10 scenarios. Deterministic (Det.) results are reproduced for comparison.

5.6. Conclusion

This chapter has extended the DFL framework to a stochastic, context-aware setting, where building thermal dynamics are modeled by a probabilistic, multi-zone RC model. The RC parameters evolve with exogenous information such as weather, hour of the day, and electricity price. By coupling this dynamic grey-box model with a scenario-based HVAC scheduling formulation and a REINFORCE-type gradient estimator, this chapter has shown how to propagate decision signals through a non-differentiable pipeline comprising a mixed-integer optimizer and a high-fidelity simulator.

The context-aware RC formulation significantly improves forecasting metrics on historical data. However, when embedded into day-ahead optimization, this dynamic RC model does not outperform static ones. *The numerical results indicate that improved forecasting accuracy does not automatically translate into better ex-post decision quality.* These observations confirm the misalignment between ITO metrics and decision performance already observed in previous chapters.

When performing DFL, the dynamic model has the most significant improvement, although the deterministic DFL experiments also underline the strong influence of the warm-start and the limited strength of the gradient signal in some configurations.

Methodologically, the chapter highlights both the promise and the difficulty of combining stochastic thermal models with DFL. On the positive side, the same probabilistic parameterization that enables REINFORCE-based gradients also naturally feeds scenario-based optimization, aligning model uncertainty with robust decision-making under uncertainty. On the other hand, the benefits of DFL are bounded by model expressiveness, gradient variance, the computational burden of the forward pass, and the number and representativeness of scenarios.

The current results are still preliminary, particularly for the DFL configurations, either deterministic or probabilistic, where computational constraints and gradient noise have so far prevented a complete evaluation. Nonetheless, the deterministic (single-scenario) experiments already suggest that context-aware, RC models trained with SS-DFL can improve ex-post objective value and cost prediction compared to their conventional ITO equivalent, while preserving tractability of the downstream optimization.

This opens several avenues for future work, including more expressive and physically structured stochastic models, tighter integration of the parameter distributions with chance-constrained or risk-aware formulations, improved variance reduction for REINFORCE, and a more systematic design of representative scenarios for DFL training.

CHAPTER 6.

Conclusion and Perspectives

This thesis develops a unified framework for model-based control of building Heating, Ventilation, and Air-Conditioning (HVAC) systems under dynamic prices, using optimization and Machine Learning (ML) to align thermal modeling with downstream decision quality. Across the thesis, we moved from convex, linear models and deterministic formulations to non-convex, neural and probabilistic models embedded in mixed-integer and scenario-based optimization. This progression has enabled increasingly realistic representations of building dynamics and uncertainty, while maintaining a clear focus on the decision-making task: scheduling HVAC in a way that is cost-effective, comfort-preserving, and robust to model error and exogenous variability.

6.1. Summary of main contributions

The thesis starts by motivating building HVAC as a major yet underused flexibility resource, and by identifying the core methodological gap of the conventional Identify-Then-Optimize (ITO) pipeline: models are trained to minimize statistical error, not to support good control actions in the regimes that matter for demand response. This leads to structurally misaligned training objectives and, in practice, to poor cost predictions and suboptimal schedules when these models are embedded in day-ahead scheduling problems.

To address this gap, three levels of Decision-Focused Learning (DFL) contributions were developed. At the convex level, we formulated a DFL framework for linear Resistance-Capacitance (RC) building models embedded in a quadratic program for day-ahead HVAC scheduling. By differentiating through the cone program, we showed how to update RC parameters to directly reduce an ex-post, task-aware loss defined on electricity bills and comfort, rather than

on temperature Mean Squared Error (MSE).
At the non-convex level, we embedded Rectified Linear Unit (ReLU) feedforward Neural Network (NN) as exact mixed-integer linear constraints in the HVAC management problem, obtaining a Mixed-Integer Quadratic Program (MIQP) that captures nonlinear thermal dynamics while preserving the optimality guarantees and interpretability of mathematical programming. We then introduced Stochastic Smoothing (SS) with a score-function gradient estimator (REINFORCE-type) to obtain informative gradients without differentiating the MIQP itself.
At the stochastic level, we proposed a probabilistic, context-aware RC model whose parameters evolve with exogenous forecasts (weather, occupancy proxies), and trained it within a scenario-based DFL scheme that explicitly targets robustness of decisions to both modeling and exogenous uncertainty. The same SS machinery provides gradients with respect to distributions over parameters, enabling decision-aware identification in a probabilistic setting.

These contributions are applied on realistic case studies (multi-zone office buildings driven by real weather and price data, with a high-fidelity simulator, EnergyPlus, in the loop), and consistently show that DFL-trained models can provide lower ex-post costs and more reliable bill predictions than their ITO counterparts, even when the latter achieve lower temperature prediction errors on historical data. The results highlight, in particular, that a simple DFL-trained RC model may outperform a more complex ITO-trained neural model when evaluated on decision metrics.

6.2. Impact on HVAC flexibility and demand response

Methodologically, the thesis clarifies how the misalignment between prediction metrics and control objectives undermines the economic and operational value of building flexibility in day-ahead markets. The ITO paradigm leads to models that are accurate in regions spanned by historical operation but unreliable in the regions explored by demand response (aggressive pre-cooling or pre-heating, new comfort envelopes, atypical thermal dynamics), which are exactly where value is created. The DFL approach redirects modeling capacity to these regions by using ex-post cost and comfort penalties as training signals.

From a power systems perspective, the main implication is improved valuation and exploitation of flexibility. By aligning model identification with market-driven objectives, building operators can operate more accurate and reliable demand response scheduling at the day-ahead horizon, reducing systematic underestimation of consumption and mitigating unexpected cost overruns.

A second implication is robustness to distribution shifts. The DFL-trained models show improved resilience to temperature distribution shifts (e.g., "hot years" in section 3.5.6) compared to ITO models calibrated on historical climates. This is critical under climate change, where future operating regimes systematically depart from past data.

A third implication is the power of hybrid models, which bridges physics-based and data-driven modeling. The thesis demonstrates that grey-box models (RC with data-driven parameter learning, possibly context-dependent and stochastic) strike a favorable compromise between interpretability, tractability, and control performance when trained with decision-aware losses. Fully black-box models trained in an ITO fashion are shown to be brittle in control tasks, even when they excel at pure forecasting.

Finally, by embedding neural models directly as optimization constraints instead of using end-to-end policy networks, the proposed approach preserves the structure of model predictive control and mathematical programming, which is important for explainability, safety constraints, and integration with existing industrial toolchains.

6.3. Methodological insights

The thesis offers several methodological insights that go beyond the specific HVAC application. First and foremost, this thesis repeatedly shows that decision metrics and prediction metrics may not align. The case studies consistently show that lower MSE on historical temperature data does not guarantee lower ex-post cost or better comfort when the model is used in a scheduling problem. This reinforces the need to evaluate and train models on decision-centric metrics.

Regarding gradient computation, we show that implicit differentiation of Karush-Kuhn-Tucker (KKT) conditions is effective and efficient for convex control policies. However, in the presence of integer variables and non-differentiable simulators, gradients through the optimization layer become either zero or undefined almost everywhere. SS with score-function estimators offers a principled alternative that does not require differentiability of the optimization problem or the physical system.

Because DFL is by nature computationally expensive, the trade-off between model complexity and trainability is stricter than for ITO. Increasing the complexity of the neural thermal model (more ReLUs, deeper networks) improves modeling capacity but significantly increases training and inference times due to the combinatorial explosion of the MIQP, with diminishing returns on ex-post cost when gradients become noisy. In several experiments, small neural models or even RC models with DFL training outperform larger architectures in terms of decision quality per unit of computational effort.

Lastly, the construction of the training dataset requires great care. Since each forward pass in DFL involves solving a large optimization problem (and often a simulation), the thesis introduces a clustering-based scenario selection procedure that maximizes information per sample: extreme days and representative medoids defined on joint weather-price features. This is essential to make DFL computationally tractable and points toward principled links between scenario reduction in stochastic programming and data selection in DFL.

6.4. Limitations

The thesis also makes explicit several limitations that bound the current applicability of the methods and motivate future research.

The first major limitation is computational burden. DFL training is orders of magnitude more expensive than standard supervised learning, primarily because each forward pass requires solving a day-ahead scheduling problem (quadratic program or MIQP) and simulating ex-post behavior. Even with modest model sizes and a limited number of scenarios, training times reach several hours on a single machine.

To maintain convex optimization or Mixed-Integer Program (MIP) structure, the thermal models are necessarily simplified. Linear RC models cannot capture important nonlinearities (e.g., radiation effects, complex HVAC equipment behavior), while piecewise-linear neural models, although more expressive, still approximate the true dynamics and can be sensitive to Big-M choices. This limitation comes directly from the type of optimization that can be handled by solvers.

Another challenge is gradient variance and warm-start dependence. The REINFORCE-based SS gradients have high variance, and the final DFL performance is often heavily influenced by the quality of the supervised warm-start. In some experiments, DFL offers limited improvement over ITO, indicating that the gradient signal is not always strong enough to systematically escape local minima.

Scalability to portfolio scale is an open problem. The case studies focus on relatively small buildings with 5 and 15 zones. Extending the same level of modeling and optimization detail to aggregations of hundreds or thousands of buildings, or to joint building–network problems, remains an open challenge due to both computational and data limitations.

Currently, there is a limited treatment of operational uncertainty. While the last chapter introduces probabilistic models and scenario-based optimization, the treatment of uncertainty is still preliminary. Uncertainty enters via parameter distributions and sampled scenarios, but formal chance constraints and risk measures (e.g., Conditional Value-at-Risk, CVaR) are not yet fully integrated into the DFL loop.

Finally, the proposed frameworks rely on the accessibility of high-quality training data in real-world deployments. The studies leveraged a high-fidelity EnergyPlus simulator, which generates rich and consistent trajectories of building dynamics, control decisions, and energy costs. In practice, data collection is often constrained by limited sensing infrastructure, incomplete metering, sensor outages, missing historical records, or confidentiality restrictions on building and tariff data. Although the proposed method relies only on a small set of variables (ambient and indoor temperatures, HVAC power consumption, and air flow per zone), the implementation of such sensing coverage in existing buildings remains challenging. Consequently, while the approach performs well under simulation-based and offline training conditions, practical deployment will likely require hybrid strategies, such as integrating measured data with calibrated digital twins, transfer learning, or reduced-order surrogate models, to alleviate data scarcity and improve robustness.

These limitations underline that DFL is not a drop-in replacement for ITO, but rather a powerful tool that must be used judiciously where the value of improved decisions justifies the extra complexity.

6.5. Perspectives for Future Research

Building on the insights and limitations above, several promising research directions emerge, which are listed below.

6.5.1. Richer and More Scalable Modeling

Surrogate models for simulation — The current DFL pipeline relies on high-fidelity simulators (e.g., EnergyPlus) to evaluate ex-post performance. A natural extension is to train fast surrogate models (e.g., reduced-order physics, emulator networks) that approximate simulator outputs sufficiently well for early DFL iterations, and progressively transition to the full simulator as training stabilizes.

Optimization formulations for entire neural networks — Embedding neural networks neuron-wise into Mixed-Integer Linear Program (MILP)/MIQP formulations via Big-M limits both tightness and scalability. Future work should explore formulations that operate at the layer or network level to obtain tighter and more scalable optimization models with explicit optimality and feasibility guarantees.

6.5.2. Advanced Decision-Focused Learning Techniques

Variance reduction — The REINFORCE estimator used here is unbiased but high-variance. Adopting variance reduction techniques (e.g., baselines) tailored to the HVAC scheduling context could yield more stable gradients and reduce the dependence on strong warm-starts.

Meta-learning and warm-start transfer — Given the cost of DFL for each building, methods to transfer knowledge between buildings become crucial. Meta-learning or transfer learning schemes could learn priors over thermal models or DFL parameters that are quickly adaptable to new buildings with limited additional DFL steps. This would support scale-up for aggregators managing heterogeneous portfolios.

Context-aware and adaptive models — The probabilistic, context-dependent RC models introduced in chapter 5 open the door to adaptive models that evolve over time with changing building usage and climate. A key research question is how to design online or periodic DFL updates that safely adapt model parameters while ensuring continuity of service and avoiding harmful forgetting.

Integration with real-time control. — The thesis focuses on day-ahead scheduling, with the implicit assumption that real-time controllers compensate for residual model errors. A natural extension is to co-design day-ahead and real-time policies: using DFL to train models and scheduling policies that are explicitly aware of the structure and limitations of the downstream real-time Model Predictive Control (MPC) or rule-based layer. This suggests a multi-level DFL framework spanning day-ahead, intraday, and real-time horizons.

6.5.3. Uncertainty, Robustness, and Risk

Chance-constrained and risk-aware DFL — Future work should combine SS with explicit chance constraints (e.g., probability of comfort violations) and risk measures on cost (e.g., CVaR of the bill) within the optimization layer. This requires generating feasible parameter distributions at scale, and designing gradient estimators that propagate through probabilistic constraints and risk functionals.

Formal robustness guarantees — While empirical robustness to temperature shifts is encouraging, there is a need for formal guarantees (e.g., bounds on worst-case cost or comfort violation under bounded model errors). Combining robust optimization, distributionally robust optimization, and DFL is an attractive direction that could bridge the gap between learning-based methods and the safety requirements of critical infrastructure.

6.5.4. System-Level Integration

Coupling with distribution networks and markets — The methods developed here operate at the building level with exogenous prices. Extending them to co-optimization with distribution networks (e.g., voltage constraints, transformer loading) and to endogenous prices (e.g., distribution-level market clearing with flexible buildings) is a natural extension. This would require integrating DFL within bilevel or hierarchical market-clearing formulations, possibly leveraging SS at multiple levels.

Interoperability and deployment — Finally, to impact real systems, the proposed methods must be integrated into existing building management systems and energy management platforms. This raises questions around interoperability (standardized interfaces between DFL models, simulators, and optimization engines), computational deployment (edge vs. cloud), and human-in-the-loop operation (e.g., presenting decision-focused models and their uncertainty to operators).

In conclusion, this thesis has shown that aligning model identification with control objectives through DFL fundamentally changes how we think about building thermal modeling and HVAC scheduling. By treating learning and optimization as a single, integrated problem, we can unlock more reliable and valuable building flexibility for power systems. At the same time, the work highlights substantial challenges in scalability, robustness, and deployment that define a rich research agenda at the intersection of power systems, optimization, and machine learning.

THE END

Bibliography

[1] P. Favaro, M. Dolányi, F. Vallée, and J.-F. Toubeau, "Neural network informed day-ahead scheduling of pumped hydro energy storage," *Energy*, vol. 289, p. 129 999, Feb. 15, 2024, ISSN: 0360-5442. DOI: 10.1016/j.energy.2023.129999. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544223033935.

[2] P. Favaro, M. Gobert, and J.-F. Toubeau, "Multi-fidelity optimization for the day-ahead scheduling of Pumped Hydro Energy Storage," *Journal of Energy Storage*, vol. 103, p. 114 096, Dec. 1, 2024, ISSN: 2352-152X. DOI: 10.1016/j.est.2024.114096. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352152X2403682X.

[3] H. Zheng, P. Favaro, Y. Dvorkin, and J. Drgoňa. "Accelerating Underground Pumped Hydro Energy Storage Scheduling with Decision-Focused Learning." arXiv: 2512.20880 [eess], Accessed: Dec. 27, 2025. [Online]. Available: http://arxiv.org/abs/2512.20880, prepublished.

[4] A. Bera, B. Gunawan, A. G. Benson, R. H. Byrne, and B. R. Chalamala, "CHAPTER 9: PUMPED HYDROELECTRIC STORAGE," en, in *U.S. DOE Energy Storage Handbook*, 2022.

[5] M. González-Torres, L. Pérez-Lombard, J. F. Coronel, I. R. Maestre, and D. Yan, "A review on buildings energy information: Trends, end-uses, fuels and drivers," *Energy Reports*, vol. 8, pp. 626–637, Nov. 2022, ISSN: 2352-4847. DOI: 10.1016/j.egyr.2021.11.280. Accessed: May 25, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S235248472101427X.

[6] S. Shoemaker, *NREL Researchers Reveal How Buildings Across United States Do—and Could—Use Energy*, en, https://www.nrel.gov/news/features/2023/nrel-researchers-reveal-how-buildings-across-the-united-states-do-and-could-use-energy.html, Ac-

cessed: 2024-05-25, Sep. 2023. Accessed: May 25, 2024. [Online]. Available: https://www.nrel.gov/news/features/2023/nrel-researchers-reveal-how-buildings-across-the-united-states-do-and-could-use-energy.html.

[7] H. Zhang et al., "Decision-Focused Learning for Power System Decision-Making Under Uncertainty," *IEEE Transactions on Power Systems*, vol. 41, no. 1, pp. 307–323, Jan. 2026, ISSN: 1558-0679. DOI: 10.1109/TPWRS.2025.3597806. Accessed: Feb. 4, 2026. [Online]. Available: https://ieeexplore.ieee.org/document/11122623.

[8] S. Cuomo, V. S. Di Cola, F. Giampaolo, G. Rozza, M. Raissi, and F. Piccialli, "Scientific Machine Learning Through Physics–Informed Neural Networks: Where we are and What's Next," en, *Journal of Scientific Computing*, vol. 92, no. 3, p. 88, Jul. 2022, ISSN: 1573-7691. DOI: 10.1007/s10915-022-01939-z. Accessed: Feb. 5, 2026. [Online]. Available: https://doi.org/10.1007/s10915-022-01939-z.

[9] S. Boyd and L. Vandenberghe, *Convex optimization [Livre]*. Cambridge university press, 2004, ISBN: 978-0-521-83378-3. [Online]. Available: https://search.ebscohost.com/login.aspx?direct=true&db=cat08949a&AN=udm.oai.umons.ac.be.SMARTU.2.257892&lang=fr&site=eds-live.

[10] J. Drgoňa et al., "All you need to know about model predictive control for buildings," *Annual Reviews in Control*, vol. 50, pp. 190–232, Jan. 2020, ISSN: 1367-5788. DOI: 10.1016/j.arcontrol.2020.09.001. Accessed: Jun. 24, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1367578820300584.

[11] J. Mandi et al., *Decision-Focused Learning: Foundations, State of the Art, Benchmark and Future Opportunities*, Aug. 2023. DOI: 10.48550/arXiv.2307.13565. arXiv: 2307.13565 [cs, math]. Accessed: Feb. 8, 2024. [Online]. Available: http://arxiv.org/abs/2307.13565.

[12] I. Jaffal and C. Inard, "A study of the nonlinearity of a building thermal behavior based on metamodeling," en, *E3S Web of Conferences*, vol. 111, S. Tanabe et al., Eds., p. 04 039, 2019, ISSN: 2267-1242. DOI: 10.1051/e3sconf/201911104039. Accessed: Feb. 5, 2026. [Online]. Available: https://www.e3s-conferences.org/10.1051/e3sconf/201911104039.

[13] Z. Wang, Y. Chen, and Y. Li, "Development of RC model for thermal dynamic analysis of buildings through model structure simplification," *Energy and Buildings*, vol. 195, pp. 51–67, Jul. 2019, ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2019.04.042. Accessed: Feb. 5, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378778818337654.

[14] J. Yang, H. Wang, L. Cheng, Z. Gao, and F. Xu, "A review of resistance–capacitance thermal network model in urban building energy simulations," *Energy and Buildings*, vol. 323, p. 114 765, Nov. 2024, ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2024.114765. Accessed: Feb. 5, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378778824008818.

[15] N. Gaikwad, K. Shankar, A. Dubey, A. Love, and O. Bergland, *Comparing Building Thermal Dynamics Models and Estimation Methods for Grid-Edge Applications*, arXiv:2508.09118 [eess], Aug. 2025. DOI: 10.48550/arXiv.2508.09118. Accessed: Feb. 5, 2026. [Online]. Available: http://arxiv.org/abs/2508.09118.

[16] G. Cybenko, "Approximation by superpositions of a sigmoidal function," en, *Mathematics of Control, Signals and Systems*, vol. 2, no. 4, pp. 303–314, Dec. 1989, ISSN: 1435-568X. DOI: 10.1007/BF02551274. Accessed: Feb. 25, 2026. [Online]. Available: https://doi.org/10.1007/BF02551274.

[17] D. Crawley et al., "EnergyPlus: Creating a New-Generation Building Energy Simulation Program," *Energy and Buildings*, vol. 33, pp. 319–331, Apr. 2001. DOI: 10.1016/S0378-7788(00)00114-6.

[18] University of Wisconsin–Madison. Solar Energy Laboratory, *TRNSYS, a transient simulation program*. Madison, Wis. : The Laboratory, 1975., 1975. [Online]. Available: https://search.library.wisc.edu/catalog/999800551102121.

[19] M. Silvestri et al., *Score Function Gradient Estimation to Widen the Applicability of Decision-Focused Learning*, en, arXiv:2307.05213 [cs], Jun. 2024. arXiv: 2307.05213 [cs]. Accessed: Aug. 30, 2024. [Online]. Available: http://arxiv.org/abs/2307.05213.

[20] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," en, *Machine Learning*, vol. 8, no. 3, pp. 229–256, May 1992, ISSN: 1573-0565. DOI: 10.1007/BF00992696. Accessed: Feb. 5, 2026. [Online]. Available: https://doi.org/10.1007/BF00992696.

[21] T. Morstyn et al., "OPEN: An open-source platform for developing smart local energy system applications," *Applied Energy*, vol. 275, p. 115 397, Oct. 2020, ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2020.115397. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261920309090.

[22] DOE and PNNL, *Commercial Prototype Building Models*, 2023. Accessed: Mar. 7, 2025. [Online]. Available: https://www.energycodes.gov/development/commercial/prototype_models.

[23] B. Wilder, B. Dilkina, and M. Tambe, "Melding the Data-Decisions Pipeline: Decision-Focused Learning for Combinatorial Optimization," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Section: AAAI Technical Track: Constraint Satisfaction and Optimization, vol. 33, Jul. 2019, pp. 1658–1665. DOI: 10.1609/aaai.v33i01.33011658. Accessed: Feb. 20, 2025. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/3982.

[24] V. Tjeng, K. Xiao, and R. Tedrake, *Evaluating Robustness of Neural Networks with Mixed Integer Programming*, Feb. 2019. DOI: 10.48550/arXiv.1711.07356. arXiv: 1711.07356 [cs]. Accessed: Apr. 15, 2024. [Online]. Available: http://arxiv.org/abs/1711.07356.

[25] R. Anderson, J. Huchette, W. Ma, C. Tjandraatmadja, and J. P. Vielma, "Strong mixed-integer programming formulations for trained neural networks," *Mathematical Programming*, vol. 183, no. 1, pp. 3–39, Sep. 2020, ISSN: 1436-4646. DOI: 10.1007/s10107-020-01474-5. Accessed: Feb. 15, 2024. [Online]. Available: https://doi.org/10.1007/s10107-020-01474-5.

[26] R. R. Bunel, I. Turkaslan, P. Torr, P. Kohli, and P. K. Mudigonda, "A Unified View of Piecewise Linear Neural Network Verification," in *Advances in Neural Information Processing Systems*, vol. 31, Curran Associates, Inc., 2018. Accessed: Feb. 14, 2025. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2018/hash/be53d253d6bc3258a8160556dda3e9b2-Abstract.html.

[27] I. Murzakhanov, A. Venzke, G. S. Misyris, and S. Chatzivasileiadis, "Neural Networks for Encoding Dynamic Security-Constrained Optimal Power Flow," *arXiv:2003.07939 [cs, eess, math]*, Oct. 2021. arXiv: 2003.07939 [cs, eess, math]. Accessed: Jan. 25, 2022. [Online]. Available: http://arxiv.org/abs/2003.07939.

[28] P. Favaro, J.-F. Toubeau, F. Vallée, and Y. Dvorkin, "Decision-Focused Learning for Complex System Identification: HVAC Management System Application," in *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*, ser. E-Energy '25, New York, NY, USA: Association for Computing Machinery, Jun. 16, 2025, pp. 347–358, ISBN: 979-8-4007-1125-1. DOI: 10.1145/3679240.3734584. Accessed: Jul. 15, 2025. [Online]. Available: https://dl.acm.org/doi/10.1145/3679240.3734584.

[29] P. Favaro, J.-F. Toubeau, F. Vallée, and Y. Dvorkin, "Decision-Focused Learning for Neural Network-Constrained HVAC Scheduling," *IEEE Transactions on Smart Grid*, pp. 1–1, 2025, ISSN: 1949-3061. DOI: 10.1109/TSG.2025.3639887. Accessed: Dec. 11, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/11275968.

[30] P. Favaro and J.-F. Toubeau, "Stochastic Decision-Focused Learning for Thermal Energy Scheduling in Multi-Zone Buildings," Feb. 2026. Accessed: Feb. 24, 2026. [Online]. Available: https://hal.science/hal-05525850.

[31] *European Climate Law - Climate Action - European Commission*, en. Accessed: Feb. 6, 2026. [Online]. Available: https://climate.ec.europa.eu/eu-action/european-climate-law_en.

[32] A. Fam and S. Fam, "Review of the US 2050 long term strategy to reach net zero carbon emissions," *Energy Reports*, vol. 12, pp. 845–860, Dec. 2024, ISSN: 2352-4847. DOI: 10.1016/j.egyr.2024.06.031. Accessed: Feb. 6, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352484724003883.

[33] *Renewable Energy Progress Tracker – Data Tools*, en-GB, Dec. 2025. Accessed: Feb. 6, 2026. [Online]. Available: https://www.iea.org/data-and-statistics/data-tools/renewable-energy-progress-tracker.

[34] *COP28: Global Renewables And Energy Efficiency Pledge*, en. Accessed: Feb. 6, 2026. [Online]. Available: https://www.cop28.com/en/global-renewables-and-energy-efficiency-pledge.

[35] "Renewables 2023 – Analysis," en-GB, Tech. Rep., Jan. 2024. Accessed: Feb. 6, 2026. [Online]. Available: https://www.iea.org/reports/renewables-2023.

[36] S. Directory, *US Grid Queue Backlog Holds Massive Clean Energy Capacity Hostage → Energy*, en-US. Accessed: Feb. 6, 2026. [Online]. Available: https://news.sustainability-directory.com/energy/us-grid-queue-backlog-holds-massive-clean-energy-capacity-hostage/.

[37] M. Laimon, "Renewable energy curtailment: A problem or an opportunity?" *Results in Engineering*, vol. 26, p. 104 925, Jun. 2025, ISSN: 2590-1230. DOI: 10.1016/j.rineng.2025.104925. Accessed: Feb. 6, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2590123025010011.

[38] T. U. Badrudeen, N. I. Nwulu, and S. L. Gbadamosi, "Low-inertia control of a large-scale renewable energy penetration in power grids: A systematic review with taxonomy and bibliometric analysis," *Energy Strategy Reviews*, vol. 52, p. 101 337, Mar. 2024, ISSN: 2211-467X. DOI: 10.1016/j.esr.2024.101337. Accessed: Feb. 6, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2211467X24000440.

[39] M. Z. Yameen, Z. Lu, F. F. M. El-Sousy, W. Younis, B. A. Zardari, and A. K. Junejo, "Improving frequency stability in grid-forming inverters with adaptive model predictive control and novel COA-jDE optimized reinforcement learning," en, *Scientific Reports*, vol. 15, no. 1, p. 16 540, May 2025, ISSN: 2045-2322. DOI: 10.1038/s41598-025-00896-5. Accessed: Feb. 6, 2026. [Online]. Available: https://www.nature.com/articles/s41598-025-00896-5.

[40] *Bundesnetzagentur - Press - Growth in renewable energy in 2024*, Jan. 2025. Accessed: Feb. 10, 2026. [Online]. Available: https://www.bundesnetzagentur.de/SharedDocs/Pressemitteilungen/EN/2025/20250108_EE.html.

[41] M. Utrilla and F. David, "Selecting mechanisms to acquire DSO flexibility services in the energy transition," en, Accepted: 2025-11-18T07:54:38Z, Ph.D. dissertation, 2025. Accessed: Feb. 10, 2026. [Online]. Available: https://repositorio.comillas.edu/xmlui/handle/11531/107248.

[42] K. Yoon, J. Leem, S. H. Lee, and J.-W. Park, "Cooperative Control of TSO and DSO: Management of Line Congestion and Frequency Response," *IEEE Access*, vol. 12, pp. 47 373–47 384, 2024, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2024.3382996. Accessed: Feb. 10, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10485417.

[43] *TSO/DSO Cooperation*, en-us, 2022. Accessed: Feb. 10, 2026. [Online]. Available: https://www.entsoe.eu/publications/tso-dso/.

[44] R. Hledik, J. Higham, A. Faruqui, and T. Lee, "The National Potential for Load Flexibility: Value and Market Potential Through 2030," en-US, Tech. Rep., Jun. 2021. Accessed: Feb. 6, 2026. [Online]. Available: https://www.brattle.com/insights-events/publications/brattle-study-cost-effective-load-flexibility-can-reduce-costs-by-more-than-15-billion-annually/.

[45] T. H. Norris, T. Profeta, D. Patino-Echeverri, and A. Cowie-Haskell, "Rethinking Load Growth: Assessing the Potential for Integration of Large Flexible Loads in US Power Systems," en, Nicholas Institute for Energy, Environment & Sustainability, Duke University, Text, Feb. 2025, Last Modified: 2025-11-07T16:15-05:00, p. 43. Accessed: Feb. 6, 2026. [Online]. Available: http://nicholasinstitute.duke.edu/publications/rethinking-load-growth.

[46] M. Z. Degefa, I. B. Sperstad, and H. Sæle, "Comprehensive classifications and characterizations of power system flexibility resources," en, *Electric Power Systems Research*, vol. 194, p. 107 022, May 2021, ISSN: 03787796. DOI: 10.1016/j.epsr.2021.107022. Accessed: Apr. 6, 2026. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S037877962100002X.

[47] B. Mohandes, M. S. E. Moursi, N. Hatziargyriou, and S. E. Khatib, "A Review of Power System Flexibility With High Penetration of Renewables," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 3140–3155, Jul. 2019, ISSN: 1558-0679. DOI: 10.1109/TPWRS.2019.2897727. Accessed: Feb. 6, 2026. [Online]. Available: https://ieeexplore.ieee.org/document/8635327/.

[48] P. R. Shukla, J. Skea, A. R. Reisinger, and IPCC, Eds., *Climate change 2022: mitigation of climate change*, en. Geneva: IPCC, 2022, ISBN: 978-92-9169-160-9.

[49] J. L. Hernández, I. de Miguel, F. Vélez, and A. Vasallo, "Challenges and opportunities in European smart buildings energy management: A critical review," *Renewable and Sustainable Energy Reviews*, vol. 199, p. 114 472, Jul. 2024, ISSN: 1364-0321. DOI: 10.1016/j.rser.2024.114472. Accessed: Feb. 6, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032124001953.

[50] P. D. Lund, J. Lindgren, J. Mikkola, and J. Salpakari, "Review of energy system flexibility measures to enable high levels of variable renewable electricity," *Renewable and Sustainable Energy Reviews*, vol. 45, pp. 785–807, May 2015, ISSN: 1364-0321. DOI: 10.1016/j.rser.2015.01.057. Accessed: Feb. 9, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032115000672.

[51] H. Khaloie, "Energy Storage Integration in Market-Based Energy Systems: Advanced Technologies and Decision-Making Models," Ph.D. dissertation, Jan. 2026. DOI: 10.13140/RG.2.2.26145.47205.

[52] *System Flexibility Needs for the Energy Transition*, en-us, Brussels, Dec. 2024. Accessed: Apr. 6, 2026. [Online]. Available: https://www.entsoe.eu/system-flexibility/.

[53] "The European power market: FCR, aFRR, mFRR & wholesale," Accessed: Feb. 9, 2026. [Online]. Available: https://www.enspired-trading.com/blog/the-european-power-market.

[54] K. Van Den Bergh and E. Delarue, "Energy and reserve markets: Interdependency in electricity systems with a high share of renewables," en, *Electric Power Systems Research*, vol. 189, p. 106 537, Dec. 2020, ISSN: 03787796. DOI: 10.1016/j.epsr.2020.106537. Accessed: Feb. 9, 2026. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0378779620303412.

[55] *Tender Hub*, en-GB. Accessed: Feb. 10, 2026. [Online]. Available: https://dso.ukpowernetworks.co.uk/flexibility/tender-hub.

[56] *The Marketplace for Energy Flexibility | Piclo*, en. Accessed: Feb. 10, 2026. [Online]. Available: https://www.piclo.com.

[57] M. Simões, A. G. Madureira, and J. A. P. Lopes, "A Comparison Between Decentralized Coordination Mechanisms for TSO-DSO Interaction: Hierarchical and Distributed Approaches," in *2025 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe)*, Oct. 2025, pp. 1–5. DOI: 10.1109/ISGTEurope64741.2025.11305627. Accessed: Feb. 10, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/11305627.

[58] J. L. Mathieu et al., "A New Definition and Research Agenda for Demand Response in the Distributed Energy Resource Era," *Policy and Regulation IEEE Transactions on Energy Markets*, vol. 3, no. 3, pp. 324–339, Sep. 2025, ISSN: 2771-9626. DOI: 10.1109/TEMPR.2025.3554734. Accessed: Feb. 6, 2026. [Online]. Available: https://ieeexplore.ieee.org/document/10938867.

[59] A. R. Jordehi, "Optimisation of demand response in electric power systems, a review," *Renewable and Sustainable Energy Reviews*, vol. 103, pp. 308–319, Apr. 2019, ISSN: 1364-0321. DOI: 10.1016/j.rser.2018.12.054. Accessed: Feb. 9, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032118308566.

[60] *Impact, un nouveau tarif pour les consommateurs flexibles*, fr, 2026. Accessed: Feb. 16, 2026. [Online]. Available: https://www.ores.be/comprendre-ma-facture/impact.

[61] J. Macknick, A. Miara, C. Tarr, R. Spellman, and C. Vörösmarty, "The power of efficiency: Optimizing environmental and social benefits through demand-side-management," American English, *Energy*, vol. 76, pp. 502–512, 2014, ISSN: 0360-5442. DOI: 10.1016/j.energy.2014.08.047. Accessed: Feb. 26, 2026. [Online]. Available: https://research-hub.nrel.gov/en/publications/power-of-efficiency-optimizing-environmental-and-social-benefits-/.

[62] P. Palensky and D. Dietrich, "Demand Side Management: Demand Response, Intelligent Energy Systems, and Smart Loads," en, *IEEE Transactions on Industrial Informatics*, vol. 7, no. 3, pp. 381–388, Aug. 2011, ISSN: 1551-3203, 1941-0050. DOI: 10.1109/TII.2011.2158841. Accessed: Feb. 9, 2026. [Online]. Available: http://ieeexplore.ieee.org/document/5930335/.

[63] R. C. Triolo, R. Rajagopal, F. A. Wolak, and J. A. de Chalendar, "Estimating cooling demand flexibility in a district energy system using temperature set point changes from selected buildings," *Applied Energy*, vol. 336, p. 120 816, Apr. 2023, ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2023.120816. Accessed: Feb. 9, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261923001800.

[64] C. D. Korkas, S. Baldi, I. Michailidis, and E. B. Kosmatopoulos, "Occupancy-based demand response and thermal comfort optimization in microgrids with renewable energy sources and energy storage," *Applied Energy*, vol. 163, pp. 93–104, Feb. 2016, ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2015.10.140. Accessed: Feb. 9, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261915013823.

[65] R. Yin, P. Xu, M. A. Piette, and S. Kiliccote, "Study on Auto-DR and pre-cooling of commercial buildings with thermal mass in California," *Energy and Buildings*, vol. 42, no. 7, pp. 967–975, Jul. 2010,

ISSN: 03787788. DOI: 10.1016/j.enbuild.2010.01.008. Accessed: Feb. 9, 2026. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0378778810000150.

[66] J. Zhang and Alejandro. D. Dominguez-Garcia, "Evaluation of Demand Response Resource Aggregation System Capacity Under Uncertainty," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4577–4586, Sep. 2018, ISSN: 1949-3053, 1949-3061. DOI: 10.1109/TSG.2017.2663780. Accessed: Feb. 9, 2026. [Online]. Available: https://ieeexplore.ieee.org/document/7839997/.

[67] E. Atam and L. Helsen, "A convex approach to a class of non-convex building HVAC control problems: Illustration by two case studies," *Energy and Buildings*, vol. 93, pp. 269–281, Apr. 2015, ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2015.02.026. Accessed: Jun. 24, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378778815001310.

[68] Z. Wang, "A survey on convex optimization for guidance and control of vehicular systems," *Annual Reviews in Control*, vol. 57, p. 100 957, Jan. 2024, ISSN: 1367-5788. DOI: 10.1016/j.arcontrol.2024.100957. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1367578824000269.

[69] S. M. Hosseini, R. Carli, and M. Dotoli, "Robust Day-Ahead Energy Scheduling of a Smart Residential User Under Uncertainty," in *2019 18th European Control Conference (ECC)*, Journal Abbreviation: 2019 18th European Control Conference (ECC), Naples, Italy: IEEE, Jun. 2019, pp. 935–940. DOI: 10.23919/ECC.2019.8796182.

[70] L. Phani Raghav, R. Seshu Kumar, D. Koteswara Raju, and A. R. Singh, "Optimal day ahead energy consumption management in grid-connected microgrids," *International Journal of Energy Research*, vol. 46, no. 2, pp. 1864–1881, Feb. 2022, Publisher: John Wiley & Sons, Ltd, ISSN: 0363-907X. DOI: 10.1002/er.7303. Accessed: Jun. 24, 2024. [Online]. Available: https://doi.org/10.1002/er.7303.

[71] Z. Afroz, G. Shafiullah, T. Urmee, and G. Higgins, "Modeling techniques used in building HVAC control systems: A review," *Renewable and Sustainable Energy Reviews*, vol. 83, pp. 64–84, Mar. 2018, ISSN: 1364-0321. DOI: 10.1016/j.rser.2017.10.044. Accessed: Jun. 24, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032117314193.

[72] D. D'Agostino, R. Landolfi, M. Nicolella, and F. Minichiello, "Experimental Study on the Performance Decay of Thermal Insulation and Related Influence on Heating Energy Consumption in Buildings," en, *Sustainability*, vol. 14, no. 5, p. 2947, Jan. 2022, Number: 5 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2071-1050. DOI: 10.3390/su14052947. Accessed: Jun. 24, 2024. [Online]. Available: https://www.mdpi.com/2071-1050/14/5/2947.

[73] K. J. Kircher and K. Max Zhang, "On the lumped capacitance approximation accuracy in RC network building models," *Energy and Buildings*, vol. 108, pp. 454–462, Dec. 2015, ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2015.09.053. Accessed: Jun. 24, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378778815302930.

[74] G. Qiang, T. Zhe, D. Yan, and Z. Neng, "An improved office building cooling load prediction model based on multivariable linear regression," *Energy and Buildings*, vol. 107, pp. 445–455, Nov. 2015, ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2015.08.041. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378778815302255.

[75] J. Drgoňa, A. R. Tuor, V. Chandan, and D. L. Vrabie, "Physics-constrained deep learning of multi-zone building thermal dynamics," *Energy and Buildings*, vol. 243, p. 110 992, Jul. 2021, ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2021.110992. Accessed: Aug. 8, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378778821002760.

[76] H. Huang, L. Chen, and E. Hu, "Model predictive control for energy-efficient buildings: An airport terminal building study," in *11th IEEE International Conference on Control & Automation (ICCA)*, ISSN: 1948-3457, Piscataway, NJ, USA: IEEE, Jun. 2014, pp. 1025–1030. DOI: 10.1109/ICCA.2014.6871061. Accessed: Jun. 24, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/6871061.

[77] Y. Zeng, Z. Zhang, and A. Kusiak, "Predictive modeling and optimization of a multi-zone HVAC system with data mining and firefly algorithms," *Energy*, vol. 86, pp. 393–402, Jun. 2015, ISSN: 0360-5442. DOI: 10.1016/j.energy.2015.04.045. Accessed: Jun. 24, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544215005009.

[78] L. Yu, S. Qin, M. Zhang, C. Shen, T. Jiang, and X. Guan, "A Review of Deep Reinforcement Learning for Smart Building Energy Management," *IEEE Internet of Things Journal*, vol. 8, no. 15, pp. 12 046–12 063, Aug. 2021, Conference Name: IEEE Internet of Things Journal, ISSN: 2327-4662. DOI: 10.1109/JIOT.2021.3078462. Accessed: Jun. 24, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/9426901.

[79] S. Jeen, A. Abate, and J. M. Cullen, "Low emission building control with zero-shot reinforcement learning," in *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, ser. AAAI'23/IAAI'23/EAAI'23, vol. 37, Washington, DC: AAAI Press, Feb. 2023, pp. 14 259–14 267, ISBN: 978-1-57735-880-0. DOI: 10.1609/aaai.v37i12.26668. Accessed: Jun. 24, 2024. [Online]. Available: https://doi.org/10.1609/aaai.v37i12.26668.

[80] J. Sun, Y. Zheng, J. Hao, Z. Meng, and Y. Liu, "Continuous Multiagent Control Using Collective Behavior Entropy for Large-Scale Home Energy Management," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, pp. 922–929, Apr. 2020, Section: AAAI Technical Track: Applications. DOI: 10.1609/aaai.v34i01.5439. Accessed: Jul. 1, 2024. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/5439.

[81] D. Jang, L. Yan, L. Spangher, and C. J. Spanos, "Active Reinforcement Learning for Robust Building Control," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 20, pp. 22 150–22 158, Mar. 2024, Section: AAAI Technical Track on AI for Social Impact Track. DOI: 10.1609/aaai.v38i20.30219. Accessed: Mar. 13, 2025. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/30219.

[82] R. Z. Homod, "Review on the HVAC System Modeling Types and the Shortcomings of Their Application," en, *Journal of Energy*, vol. 2013, no. 1, p. 768 632, 2013, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1155/2 ISSN: 2314-615X. DOI: 10.1155/2013/768632. Accessed: Jun. 24, 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1155/2013/768632.

[83] A. Afram and F. Janabi-Sharifi, "Review of modeling methods for HVAC systems," *Applied Thermal Engineering*, vol. 67, no. 1, pp. 507–519,

Jun. 2014, ISSN: 1359-4311. DOI: 10.1016/j.applthermaleng.2014.03.055. Accessed: Jun. 24, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1359431114002348.

[84] F. Belić, Ž. Hocenski, and D. Slišković, "Thermal modeling of buildings with RC method and parameter estimation," in *2016 International Conference on Smart Systems and Technologies (SST)*, Journal Abbreviation: 2016 International Conference on Smart Systems and Technologies (SST), Osijek, Croatia: IEEE, Oct. 2016, pp. 19–25. DOI: 10.1109/SST.2016.7765626.

[85] Z. Afroz, G. Shafiullah, T. Urmee, and G. Higgins, "Modeling techniques used in building HVAC control systems: A review," *Renewable & Sustainable Energy Reviews*, vol. 83, pp. 64–84, Dec. 2017. DOI: 10.1016/j.rser.2017.10.044.

[86] Y. -J. Kim, "A Supervised-Learning-Based Strategy for Optimal Demand Response of an HVAC System in a Multi-Zone Office Building," *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 4212–4226, Sep. 2020, ISSN: 1949-3061. DOI: 10.1109/TSG.2020.2986539.

[87] S. A. Nabavi, N. H. Motlagh, M. A. Zaidan, A. Aslani, and B. Zakeri, "Deep Learning in Energy Modeling: Application in Smart Buildings With Distributed Energy Generation," *IEEE Access*, vol. 9, pp. 125 439–125 461, 2021, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3110960.

[88] Y.-H. Kao, B. V. Roy, and X. Yan, "Directed Regression," in *Neural Information Processing Systems*, Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, Eds., vol. 22, Vancouver, B.C., Canada: Curran Associates, Inc., Dec. 2009. Accessed: Jun. 24, 2024. [Online]. Available: https://www.semanticscholar.org/paper/Directed-Regression-Kao-Roy/76efadb1274df410d7060ca0b184c4dba354e96c.

[89] P. L. Donti, B. Amos, and J. Z. Kolter, "Task-based end-to-end model learning in stochastic optimization," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17, Red Hook, NY, USA: Curran Associates Inc., Dec. 2017, pp. 5490–5500, ISBN: 978-1-5108-6096-4. Accessed: Jun. 24, 2024.

[90] M. S. Bazaraa, J. J. Jarvis, and H. D. Sherali, "The simplex method," in *Linear Programming and Network Flows*. Hoboken, NJ, US: John Wiley & Sons, Ltd, 2009, ch. 3, pp. 91–149, ISBN: 9780471703778. DOI: https://doi.org/10.1002/9780471703778.ch3. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780471703778.ch3.

[Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780471703778.ch3.

[91] Adam N. Elmachtoub, A. N. Elmachtoub, Paul Grigas, and P. Grigas, “Smart "Predict, then Optimize",” *arXiv: Optimization and Control*, vol. 68, no. 1, pp. 9–26, Oct. 2017, ARXIV_ID: 1710.08005 MAG ID: 2766820687 S2ID: 0eb3c110b1a42dc79a988f3a3e6f9a6c73a57cb0, ISSN: 0025-1909. DOI: 10.1287/mnsc.2020.3922. [Online]. Available: https://doi.org/10.1287/mnsc.2020.3922.

[92] A. Agrawal, S. Barratt, S. Boyd, E. Busseti, and W. M. Moursi, “Differentiating Through a Cone Program,” en, *Journal of Applied and Numerical Optimization*, vol. 1, pp. 107–115, 2019. Accessed: Jun. 21, 2024. [Online]. Available: https://jano.biemdas.com/archives/931.

[93] A. Agrawal, S. Barratt, S. Boyd, and B. Stellato, “Learning Convex Optimization Control Policies,” en, in *Proceedings of the 2nd Conference on Learning for Dynamics and Control*, ISSN: 2640-3498, Berkeley, CA: PMLR, Jul. 2020, pp. 361–373. Accessed: Jun. 24, 2024. [Online]. Available: https://proceedings.mlr.press/v120/agrawal20a.html.

[94] D. Wahdany, C. Schmitt, and J. L. Cremer, “More than accuracy: End-to-end wind power forecasting that optimises the energy system,” *Electric Power Systems Research*, vol. 221, p. 109 384, Aug. 2023, ISSN: 0378-7796. DOI: 10.1016/j.epsr.2023.109384. Accessed: Jun. 25, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378779623002730.

[95] L. Ortmann, F. Böhm, F. Klein-Helmkamp, A. Ulbig, S. Bolognani, and F. Dörfler, *Tuning and Testing an Online Feedback Optimization Controller to Provide Curative Distribution Grid Flexibility*, en, arXiv:2403.01782 [cs, eess], Mar. 2024. arXiv: 2403.01782 [cs, eess]. Accessed: Jun. 21, 2024. [Online]. Available: http://arxiv.org/abs/2403.01782.

[96] S. Shah, K. Wang, B. Wilder, A. Perrault, and M. Tambe, “Decision-focused learning without decision-making: Learning locally optimized decision losses,” in *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., vol. 35, New Orleans, LA: Curran Associates, Inc., 2022, pp. 1320–1332. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2022/file/0904c7edde20d7134a77fc7f9cd86ea2-Paper-Conference.pdf.

[97] B. Chen, Z. Cai, and M. Bergés, "Gnu-RL: A Precocial Reinforcement Learning Solution for Building HVAC Control Using a Differentiable MPC Policy," in *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, ser. BuildSys '19, New York, NY, USA: Association for Computing Machinery, Nov. 2019, pp. 316–325, ISBN: 978-1-4503-7005-9. DOI: 10.1145/3360322.3360849. Accessed: Aug. 21, 2025. [Online]. Available: https://dl.acm.org/doi/10.1145/3360322.3360849.

[98] X. Cui, J.-F. Toubeau, F. Vallée, and Y. Wang, "Decision-Oriented Modeling of Thermal Dynamics Within Buildings," *IEEE Transactions on Smart Grid*, vol. 16, no. 1, pp. 369–382, Jan. 2025, ISSN: 1949-3061. DOI: 10.1109/TSG.2024.3445574. Accessed: Aug. 8, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/10638763/.

[99] U.S. Department of Energy, en, https://www.energycodes.gov/sites/default/files/2023-10/ASHRAE901_OfficeMedium_STD2019.zip, Accessed: 2024-06-24, 2019.

[100] A. Nemirovski, "Advances in Convex Optimization: Conic Programming," en, in *International Congress of Mathematicians*, vol. 1, Zürich, Switzerland: EMS Press, 2007, pp. 413–444.

[101] Y. Ye, M. J. Todd, and S. Mizuno, "An O sqrt(nL)-Iteration Homogeneous and Self-Dual Linear Programming Algorithm," *Mathematics of Operations Research*, vol. 19, no. 1, pp. 53–67, Feb. 1994, Publisher: INFORMS, ISSN: 0364-765X. DOI: 10.1287/moor.19.1.53. Accessed: Jun. 24, 2024. [Online]. Available: https://pubsonline.informs.org/doi/abs/10.1287/moor.19.1.53.

[102] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, "Conic Optimization via Operator Splitting and Homogeneous Self-Dual Embedding," en, *Journal of Optimization Theory and Applications*, vol. 169, no. 3, pp. 1042–1068, Jun. 2016, ISSN: 0022-3239, 1573-2878. DOI: 10.1007/s10957-016-0892-3. Accessed: Jun. 21, 2024. [Online]. Available: http://link.springer.com/10.1007/s10957-016-0892-3.

[103] E. Busseti, W. M. Moursi, and S. Boyd, "Solution refinement at regular points of conic problems," en, *Computational Optimization and Applications*, vol. 74, no. 3, pp. 627–643, Dec. 2019, ISSN: 1573-2894. DOI: 10.1007/s10589-019-00122-9. Accessed: Jun. 24, 2024. [Online]. Available: https://doi.org/10.1007/s10589-019-00122-9.

[104] M. Chau and M. C. Fu, *An Overview of Stochastic Approximation*, en. New York, NY: Springer, 2015, pp. 149–178, ISBN: 978-1-4939-1384-8. DOI: 10.1007/978-1-4939-1384-8_6. Accessed: Jun. 24, 2024. [Online]. Available: https://doi.org/10.1007/978-1-4939-1384-8_6.

[105] N. Pagan, J. Baumann, E. Elokda, G. De Pasquale, S. Bolognani, and A. Hannák, "A Classification of Feedback Loops and Their Relation to Biases in Automated Decision-Making Systems," in *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, ser. EAAMO '23, New York, NY, USA: Association for Computing Machinery, Oct. 2023, pp. 1–14, ISBN: 979-8-4007-0381-2. DOI: 10.1145/3617694.3623227. Accessed: Apr. 7, 2026. [Online]. Available: https://doi.org/10.1145/3617694.3623227.

[106] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, arXiv:1412.6980 [cs], Jan. 2017. DOI: 10.48550/arXiv.1412.6980. Accessed: Jul. 1, 2024. [Online]. Available: http://arxiv.org/abs/1412.6980.

[107] *Le bihoraire change d'horaire*, fr. Accessed: Apr. 7, 2026. [Online]. Available: https://www.ores.be/comprendre-ma-facture/bihoraire.

[108] X. Lu, Y. Fu, and Z. O'Neill, "Benchmarking high performance HVAC Rule-Based controls with advanced intelligent Controllers: A case study in a Multi-Zone system in Modelica," *Energy and Buildings*, vol. 284, p. 112 854, Apr. 2023, ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2023.112854. Accessed: Apr. 7, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378778823000841.

[109] M. Yousefi, A. Hajizadeh, M. N. Soltani, and B. Hredzak, "Predictive Home Energy Management System With Photovoltaic Array, Heat Pump, and Plug-In Electric Vehicle," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 1, pp. 430–440, Jan. 2021, ISSN: 1941-0050. DOI: 10.1109/TII.2020.2971530.

[110] P. L. Donti, D. Rolnick, and J. Z. Kolter, *DC3: A learning method for optimization with hard constraints*, Apr. 2021. DOI: 10.48550/arXiv.2104.12225. arXiv: 2104.12225 [cs]. Accessed: Sep. 25, 2025. [Online]. Available: http://arxiv.org/abs/2104.12225.

[111] B. Tang, E. B. Khalil, and J. Drgoňa, *Learning to Optimize for Mixed-Integer Non-linear Programming with Feasibility Guarantees*, arXiv:2410.11061 [cs], May 2025. DOI: 10.48550/arXiv.2410.11061. Accessed: Aug. 22, 2025. [Online]. Available: http://arxiv.org/abs/2410.11061.

[112] S. Park and P. V. Hentenryck, *Self-Supervised Primal-Dual Learning for Constrained Optimization*, arXiv:2208.09046 [cs], Nov. 2022. DOI: 10.48550/arXiv.2208.09046. Accessed: Sep. 4, 2025. [Online]. Available: http://arxiv.org/abs/2208.09046.

[113] F. Fioretto, T. W. K. Mak, and P. V. Hentenryck, *Predicting AC Optimal Power Flows: Combining Deep Learning and Lagrangian Dual Methods*, arXiv:1909.10461 [eess], Dec. 2019. DOI: 10.48550/arXiv.1909.10461. Accessed: Sep. 4, 2025. [Online]. Available: http://arxiv.org/abs/1909.10461.

[114] R. Li et al., *Decision-Oriented Learning for Future Power System Decision-Making under Uncertainty*, Apr. 2024. DOI: 10.48550/arXiv.2401.03680. arXiv: 2401.03680 [eess]. Accessed: Mar. 13, 2025. [Online]. Available: http://arxiv.org/abs/2401.03680.

[115] B. Amos and J. Z. Kolter, "OptNet: Differentiable Optimization as a Layer in Neural Networks," en, in *Proceedings of the 34th International Conference on Machine Learning*, ISSN: 2640-3498, PMLR, Jul. 2017, pp. 136–145. Accessed: Jan. 7, 2025. [Online]. Available: https://proceedings.mlr.press/v70/amos17a.html.

[116] A. Ferber, B. Wilder, B. Dilkina, and M. Tambe, "MIPaaL: Mixed Integer Program as a Layer," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 02, pp. 1504–1511, Apr. 2020. DOI: 10.1609/aaai.v34i02.5509. Accessed: Feb. 20, 2025. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/5509.

[117] B. Tang and E. B. Khalil, *PyEPO: A PyTorch-based End-to-End Predict-then-Optimize Library for Linear and Integer Programming*, arXiv:2206.14234 [math], Apr. 2023. DOI: 10.48550/arXiv.2206.14234. Accessed: Aug. 21, 2025. [Online]. Available: http://arxiv.org/abs/2206.14234.

[118] F. Ceccon et al., *OMLT: Optimization & Machine Learning Toolkit*, Nov. 2022. DOI: 10.48550/arXiv.2202.02414. arXiv: 2202.02414 [stat]. Accessed: Mar. 17, 2025. [Online]. Available: http://arxiv.org/abs/2202.02414.

[119] S. Zhang, J. S. Campos, C. Feldmann, F. Sandfort, M. Mathea, and R. Misener, "Augmenting optimization-based molecular design with graph neural networks," *Computers & Chemical Engineering*, vol. 186, p. 108 684, Jul. 2024, ISSN: 0098-1354. DOI: 10.1016/j.compchemeng.2024.108684. Accessed: Mar. 17, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0098135424001029.

[120] S. Mohamed, M. Rosca, M. Figurnov, and A. Mnih, "Monte carlo gradient estimation in machine learning," *Journal of Machine Learning Research*, vol. 21, no. 132, pp. 1–62, 2020. [Online]. Available: http://jmlr.org/papers/v21/19-346.html.

[121] C. Baldassi, E. M. Malatesta, and R. Zecchina, "Properties of the Geometry of Solutions and Capacity of Multilayer Neural Networks with Rectified Linear Unit Activations," *Physical review letters*, vol. 123, no. 17, pp. 1–170 602, 2019, ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.123.170602.

[122] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," *Journal of machine learning research*, vol. 15, pp. 315–323, 2011, ISSN: 1532-4435.

[123] R. Balestriero and R. G. Baraniuk, "Mad Max: Affine Spline Insights Into Deep Learning," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 704–727, May 2021, ISSN: 1558-2256. DOI: 10.1109/JPROC.2020.3042100.

[124] E. Balas, "Disjunctive programming," *Annals of discrete mathematics*, vol. 5, pp. 3–51, 1979, ISSN: 0167-5060.

[125] C. Tsay, J. Kronqvist, A. Thebelt, and R. Misener, "Partition-Based Formulations for Mixed-Integer Optimization of Trained ReLU Neural Networks," in *Advances in Neural Information Processing Systems*, vol. 34, Curran Associates, Inc., 2021, pp. 3068–3080. Accessed: Feb. 14, 2025. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2021/hash/17f98ddf040204eda0af36a108cbdea4-Abstract.html.

[126] A. Alcántara and C. Ruiz, "A neural network-based distributional constraint learning methodology for mixed-integer stochastic optimization," *Expert Systems with Applications*, vol. 232, p. 120 895, Dec. 2023, ISSN: 0957-4174. DOI: 10.1016/j.eswa.2023.120895. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417423013970.

[127] D. Kenefake, R. Kakaodkar, M. Ali, and E. N. Pistikopoulos, "A novel neural network bounds-tightening procedure for multiparametric programming and control," *Computer Aided Chemical Engineering*, vol. 52, A. C. Kokossis, M. C. Georgiadis, and E. Pistikopoulos, Eds., pp. 1841–1846, Jan. 2023. DOI: 10.1016/B978-0-443-15274-0.50292-4. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780443152740502924.

[128] J. Zhou, Y. T. Chu, J. Ren, W. Shen, and C. He, "Integrating machine learning and mathematical programming for efficient optimization of operating conditions in organic Rankine cycle (ORC) based combined systems," *Energy*, vol. 281, p. 128 218, Oct. 2023, ISSN: 0360-5442. DOI: 10.1016/j.energy.2023.128218. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544223016122.

[129] R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*. Society for Industrial and Applied Mathematics, Jan. 2009, ISBN: 9780898716696. DOI: 10.1137/1.9780898717716. Accessed: Feb. 18, 2025. [Online]. Available: http://epubs.siam.org/doi/book/10.1137/1.9780898717716.

[130] P. Favaro, J.-F. Toubeau, F. Vallée, and Y. Dvorkin, "Decision-Focused Learning for Complex System Identification: HVAC Management System Application," in *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*, ser. E-Energy '25, New York, NY, USA: Association for Computing Machinery, Jun. 2025, pp. 347–358, ISBN: 979-8-4007-1125-1. DOI: 10.1145/3679240.3734584. Accessed: Jun. 24, 2025. [Online]. Available: https://dl.acm.org/doi/10.1145/3679240.3734584.

[131] L. F. Cabeza et al., "Buildings," in *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Intergovernmental Panel on Climate Change, Ed., Cambridge, UK and New York, NY, USA: Cambridge University Press, 2022.

[132] R. L. Kini, A. Vlachokostas, M. R. Brambley, and A. Rogers, "Occupant-centric demand response for thermostatically-controlled home loads," *IEEE Transactions on Smart Grid*, vol. 16, no. 3, pp. 2234–2245, 2025. DOI: 10.1109/TSG.2025.3540427.

[133] C. Ghiaus and N. Ahmad, "Thermal circuits assembling and state-space extraction for modelling heat transfer in buildings," *Energy*, vol. 195, p. 117 019, 2020, ISSN: 0360-5442.

[134] H. Shi, J. Liu, and Q. Chen, "An rc-network approach for hvac precooling optimization in buildings," *IEEE Transactions on Sustainable Computing*, vol. 7, no. 3, pp. 512–526, 2022. DOI: 10.1109/TSUSC.2019.2943491.

[135] O. B. Mulayim and M. Bergés, "On the Impact of Simulated Occupancy Behavior Assumptions on Reinforcement Learning for HVAC Controls," in *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*, ser. E-Energy '25, New York, NY, USA: Association for Computing Machinery, Jun. 2025, pp. 317–321, ISBN: 979-8-4007-1125-1. DOI: 10.1145/3679240.3734585. Accessed: Feb. 17, 2026. [Online]. Available: https://dl.acm.org/doi/10.1145/3679240.3734585.

[136] X. Cui, D. Qin, J.-F. Toubeau, F. Vallée, and Y. Wang, "Generalizable thermal dynamics modeling via personalized federated learning," *IEEE Transactions on Smart Grid*, vol. 16, no. 6, pp. 4698–4711, 2025. DOI: 10.1109/TSG.2025.3585942.

[137] J.-F. Toubeau, J. Bottieau, F. Vallée, and Z. De Grève, "Deep learning-based multivariate probabilistic forecasting for short-term scheduling in power markets," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1203–1215, 2019. DOI: 10.1109/TPWRS.2018.2870041.

[138] P. Favaro, C. Rasic, Y. Wang, and J.-F. Toubeau, "Safe Reinforcement Learning for Battery Energy Storage Participation in the Imbalance Settlement," *Policy and Regulation IEEE Transactions on Energy Markets*, pp. 1–13, 2025, ISSN: 2771-9626. DOI: 10.1109/TEMPR.2025.3639758. Accessed: Dec. 11, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/11277391.

[139] C. Dupont, P. Favaro, F. Vallée, B. Francois, and J. -F. Toubeau, "Decision-Focused Learning for Optimized Participation of an Industrial Consumer in Energy-Only and Reserve Markets," in *2024 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, Journal Abbreviation: 2024 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE), Oct. 2024, pp. 1–5. DOI: 10.1109/ISGTEUROPE62998.2024.10863804.

[140] A. Agrawal, B. Amos, S. Barratt, S. Boyd, S. Diamond, and J. Kolter, "Differentiable convex optimization layers," *Advances in Neural Information Processing Systems*, 2019.

APPENDIX A.

List of Publications

The publications of P. Favaro over the course of this PhD thesis are as follows.

Published

Peer-reviewed Journal papers

- P. Favaro et al., "Decision-Focused Learning for Neural Network-Constrained HVAC Scheduling," *IEEE Transactions on Smart Grid*, pp. 1–1, 2025, ISSN: 1949-3061. DOI: `10.1109/TSG.2025.3639887`. Accessed: Dec. 11, 2025. [Online]. Available: `https://ieeexplore.ieee.org/document/11275968`

- P. Favaro et al., "Safe Reinforcement Learning for Battery Energy Storage Participation in the Imbalance Settlement," *Policy and Regulation IEEE Transactions on Energy Markets*, pp. 1–13, 2025, ISSN: 2771-9626. DOI: `10.1109/TEMPR.2025.3639758`. Accessed: Dec. 11, 2025. [Online]. Available: `https://ieeexplore.ieee.org/document/11277391`

- P. Favaro et al., "Multi-fidelity optimization for the day-ahead scheduling of Pumped Hydro Energy Storage," *Journal of Energy Storage*, vol. 103, p. 114 096, Dec. 1, 2024, ISSN: 2352-152X. DOI: `10.1016/j.est.2024.114096`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S2352152X2403682X`

- P. Favaro et al., "Neural network informed day-ahead scheduling of pumped hydro energy storage," *Energy*, vol. 289, p. 129 999, Feb. 15, 2024, ISSN:

0360-5442. DOI: 10.1016/j.energy.2023.129999. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544223033935

Peer-reviewed Conference papers

- P. Favaro et al., "Decision-Focused Learning for Complex System Identification: HVAC Management System Application," in *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*, ser. E-Energy '25, New York, NY, USA: Association for Computing Machinery, Jun. 16, 2025, pp. 347–358, ISBN: 979-8-4007-1125-1. DOI: 10.1145/3679240.3734584. Accessed: Jul. 15, 2025. [Online]. Available: https://dl.acm.org/doi/10.1145/3679240.3734584

- C. Dupont et al., "Decision-Focused Learning for Optimized Participation of an Industrial Consumer in Energy-Only and Reserve Markets," in *2024 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, Journal Abbreviation: 2024 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE), Oct. 2024, pp. 1–5. DOI: 10.1109/ISGTEUROPE62998.2024.10863804

Under Review

- H. Zheng et al. "Accelerating Underground Pumped Hydro Energy Storage Scheduling with Decision-Focused Learning." arXiv: 2512.20880 [eess], Accessed: Dec. 27, 2025. [Online]. Available: http://arxiv.org/abs/2512.20880, pre-published

Working Paper

- P. Favaro and J.-F. Toubeau, "Stochastic Decision-Focused Learning for Thermal Energy Scheduling in Multi-Zone Buildings," Feb. 2026. Accessed: Feb. 24, 2026. [Online]. Available: https://hal.science/hal-05525850

APPENDIX B.

Differentiable Convex Optimization with CVXPYLayers

CVXPYLayers provide a mechanism to differentiate convex optimization problems. A CVXPYLayers represent the solution map of a convex optimization problem with respect to its parameters and exposes this map to automatic differentiation frameworks such as PyTorch, and TensorFlow. This section relies on material in [92], [102], [103], [140].

B.1. Problem Setup and Solution Map

Consider a parameterized convex optimization problem of the form

$$\begin{aligned} \text{minimize}_{x\in\mathbb{R}^n} \quad & f(x;\theta) \\ \text{subject to} \quad & g_i(x;\theta) \leq 0, \quad i = 1, \dots, m_i, \\ & h_j(x;\theta) = 0, \quad j = 1, \dots, m_e, \end{aligned} \tag{B.1}$$

where

- $x \in \mathbb{R}^n$ is the decision variable,
- $\theta \in \mathbb{R}^p$ is a parameter vector (the concatenation of all parameters in the problem),
- f is a convex objective function,
- g_i are convex inequality constraint functions, and
- h_j are affine equality constraint functions.

For each θ, assume that problem (B.1) has a unique optimal solution $x^\star(\theta)$. The CVXPYLayer represents the solution map

$$S : \mathbb{R}^p \to \mathbb{R}^n, \qquad S(\theta) = x^\star(\theta), \tag{B.2}$$

associated with problem (B.1).
In a learning setup, one typically defines a loss of the form

$$\mathcal{L}(\theta) = \ell\big(x^\star(\theta)\big) = \ell\big(S(\theta)\big), \tag{B.3}$$

for some differentiable function $\ell : \mathbb{R}^n \to \mathbb{R}$. Backpropagation then requires the gradient $\nabla_\theta \mathcal{L}(\theta)$. CVXPYLayers compute this gradient analytically by differentiating through the optimization problem.

B.2. Disciplined Parametrized Programming and Canonical Cone Program

CVXPY[1] uses the Disciplined Convex Programming (DCP) paradigm to guarantee convexity of optimization problems. For differentiable layers, the notion of Disciplined Parametrized Programming (DPP) is introduced to control how parameters may appear in the problem so that the final cone program data depend affinely on θ.

B.2.1. DCP and DPP

In DCP, parameters are conceptually treated as constants. In DPP, parameters are treated as affine expressions, and they must appear only in ways that preserve the convexity of the overall problem, and affine dependence of the canonical cone program data d (i.e., the variables and parameters in (B.4)) on the parameters θ. Non-allowed patterns include non-affine compositions of parameters, such as products of parameters $p_1 p_2$ or parameters used as exponents, since these would create nonlinear dependence of the cone data on θ.

[1]CVXPY is the open-source Python modeling language for optimization associated with CVXPYLayers.

B.2.2. Canonical cone program

Since any convex problem can be expressed as a conic problem [100], CVXPY transforms (canonicalizes) any DPP-compliant problem into a canonical cone primal-dual pair of the form

$$
\begin{array}{llll}
(P) & \min\limits_{z,s} & c^T z & \\
 & \text{s.t.} & Az + s = b, & \\
 & & (z,s) \in \mathbb{R}^n \times \mathcal{K}, &
\end{array}
\qquad
\begin{array}{lll}
(D) & \max\limits_{z,s} & -b^T \nu \\
 & \text{s.t.} & -A^T \nu + r = c, \\
 & & (r,\nu) \in \{0\}^n \times \mathcal{K}^*,
\end{array}
\tag{B.4}
$$

where

- $z \in \mathbb{R}^{n_z}$ is the canonical primal variable (including the original variable x and auxiliary variables),
- $s \in \mathbb{R}^{n_s}$ is a slack variable with $n_s \geq n_z$,
- $A \in \mathbb{R}^{m_e \times n_z}$ and $b \in \mathbb{R}^{m_e}$ specify constraints,
- $c \in \mathbb{R}^{n_z}$ is the cost vector,
- $\nu \in \mathcal{K}^*$ is the dual variable, with $\mathcal{K}^*$, the dual cone of $\mathcal{K}$,
- $r \in \{0\}^n$ is the dual residual, with $\{0\}^n$ the dual of $\mathbb{R}^n$
- $\mathcal{K}$ is a Cartesian product of standard convex cones: nonnegative orthant $\mathbb{R}^n_+$, second-order cones (a.k.a. ice-cream cones or Lorentz cones), semidefinite cones, etc.

Definition: Orthant

An orthant is one of the 2^n regions in $\mathbb{R}^n$ defined by fixing the signs (positive or negative) of each coordinate, analogous to a quadrant in 2D or octant in 3D.

B.3. Affine–Solver–Affine (ASA) Decomposition

The overall solution map S of a convex program can decomposed into three successive transformations: (i) the canonicalizer, (ii) the cone solver, and (iii) the retriever. Let

- $C : \mathbb{R}^p \to \mathcal{D}$ be the canonicalizer,
- $s : \mathcal{D} \to \mathcal{Y}$ be the cone solver's solution map,

- $R : \mathcal{Y} \to \mathbb{R}^n$ be the retriever,

where $\mathcal{D}$ is the space of cone program data $d := (A, b, c)$, and $\mathcal{Y}$ is the space of primal–dual variables $y := (z, s, \nu, r)$.

B.3.1. Canonicalizer

The tuple $d := (A, b, c)$ collects all the data from the cone program.
DPP guarantees the existence of an *affine* canonicalization map:

$$C : \mathbb{R}^p \to \mathcal{D}, \qquad C(\theta) = (A, b, c) := d. \tag{B.5}$$

C is the map reformulating the initial convex problem (B.1) into the canonic cone program (B.4).

To express the affine dependence between θ and d explicitly, consider the augmented parameter vector

$$\tilde{\theta} = \begin{bmatrix} \theta \\ 1 \end{bmatrix} \in \mathbb{R}^{p+1}. \tag{B.6}$$

There exist fixed sparse matrices Q_c, $R_{Ab}^{(i)}$, such that

$$\begin{aligned} c &= Q_c \tilde{\theta}, \\ \begin{bmatrix} A & b \end{bmatrix} &= \sum_{i=1}^{p+1} R_{Ab}^{(i)} \tilde{\theta}_i. \end{aligned} \tag{B.7}$$

Equivalently, by stretching our notation so that $d := \text{vec}(A, b, c)$, then

$$d = M\tilde{\theta}, \tag{B.8}$$

for a fixed sparse matrix M. CVXPYLayers compute and store this matrix (in an implicit sparse form) when constructing a `CvxpyLayer` object.

B.3.2. Solver

The solver map

$$s : (A, b, c) \mapsto y^\star = (z^\star, s^\star, \nu^\star, r^\star) \tag{B.9}$$

returns the optimal primal–dual point of the cone program (B.4). Assuming that strong duality holds, the KKT conditions are necessary and sufficient for optimality. Explicitly, $y^\star$ is primal-dual optimal when

$$Az^\star + s^\star = b,\ s^\star \in \mathcal{K}, \tag{B.10}$$
$$A^T\nu^\star + c = r^\star,\ \nu^\star \in \mathcal{K}^*, \tag{B.11}$$
$$(\nu^\star)^T s^\star = 0. \tag{B.12}$$

Equation (B.10) is primal feasibility, (B.11) is dual feasibility, and (B.12) is complementary slackness. Note that stationarity conditions, obtained by differentiating the Lagrangian with respect to x and ν, are identical to primal and dual feasibility. Complementary slackness can equivalently be replaced by the strong duality condition:

$$c^T z^\star + b^T \nu^\star = 0 \tag{B.13}$$

If we write these conditions under matrix-vector form, we get

$$\begin{bmatrix} r \\ s \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & A^T & c \\ -A & 0 & b \\ c^T & b^T & 0 \end{bmatrix} \begin{bmatrix} z \\ \nu \\ 1 \end{bmatrix}. \tag{B.14}$$

Definition: Homogeneous Function

A function is homogeneous of degree n if, when all the function's arguments are multiplied by the same scalar k, then the function's value is multiplied by k^n such that:

$$f(kx_1, ..., kx_n) = k^n f(x_1, ..., x_n),$$

for every $x_1, ..., x_n$ and $s \neq 0$.

Definition: Self-dual

A problem is self-dual if the primal and dual problem are identical.

Any y satisfying (B.14) is optimal for (B.4). However, if (B.4) is primal or dual infeasible, then (B.14) has no solution. Therefore, following the homogeneous self-dual embedding in [101], CVXPYLayers introduce two variables

$(\kappa, \tau) \in \mathbb{R}^n \times \mathbb{R}^n$ leading to:

$$\begin{bmatrix} r \\ s \\ \kappa \end{bmatrix} = \begin{bmatrix} 0 & A^T & c \\ -A & 0 & b \\ c^T & b^T & 0 \end{bmatrix} \begin{bmatrix} z \\ \nu \\ \tau \end{bmatrix} \tag{B.15}$$

$$\Leftrightarrow \qquad v = Qu. \tag{B.16}$$

Any solution $y = (z, s, r, \nu, \tau, \kappa)$ of the self-dual embedding belongs to one of these three categories:

1. If $\tau > 0$ and $\kappa = 0$, the point $(z/\tau, \nu/\tau, s/\tau)$ satisfies the KKT conditions of (B.4) and is a primal-dual solution.

2. If $\tau = 0$ and $\kappa > 0$, then the duality gap $c^T z + b^T \nu$ is negative indicating primal ($b^T \nu < 0$) and/or dual ($c^T z < 0$) infeasibility.

3. If $\tau = \kappa = 0$, nothing can be concluded about the problem except if $b^T \nu < 0$ or $c^T z < 0$ indicating primal and dual infeasibility, respectively.

To solve the homogeneous self-dual embedding, CVXPYLayers reformulate the problem as finding a zero to the normalized residual map of the embedding defined as:

$$\mathcal{N}(Z) = Q\Pi\left(\frac{Z}{|Z_{m+n+1}|}\right) + \Pi^*\left(\frac{Z}{|Z_{m+n+1}|}\right) \tag{B.17}$$

where

- $Z \in \mathbb{R}^{m+n+1}$ is the concatenation of primal z and slack s variables augmented for the mapping,
- Z_{m+n+1} is the last component of Z,
- Π is the Euclidean projection on $\mathcal{K}$,
- Π^* is the Euclidean projection on $-\mathcal{K}^*$.

For the interested reader, more about establishing the residual map is available at [103].

B.3.3. Retriever

The retriever R extracts the original decision variable $x^\star \in \mathbb{R}^n$ from the canonical solution. If the original variable x corresponds to a fixed subset of coordinates of z, then there exists a fixed selection matrix $P \in \mathbb{R}^{n \times n_z}$ such that

$$x^\star = R(y^\star) = Pz^\star. \tag{B.18}$$

B.3.4. Overall Mapping

The overall solution map represented by the CVXPYLayer is therefore

$$S(\theta) = R\big(s(C(\theta))\big). \tag{B.19}$$

This is the *Affine–Solver–Affine* (ASA) decomposition. The canonicalizer C is affine in θ. The retriever R is affine in y. The cone solver s is the only nonlinear component.

B.4. Differentiation via the Chain Rule

Let the training loss be

$$\mathcal{L}(\theta) = \ell\big(S(\theta)\big) = \ell\big(x^\star(\theta)\big), \tag{B.20}$$

and denote by

$$\bar{x} := \frac{\partial \mathcal{L}}{\partial x^\star} \in \mathbb{R}^n \tag{B.21}$$

the upstream gradient provided by the automatic differentiation framework. The derivative of S at θ can be written via the chain rule as

$$DS(\theta) = DR(y^\star)\, Ds(d)\, DC(\theta), \tag{B.22}$$

where $d = C(\theta)$ and $y^\star = s(d)$.
In reverse mode, one needs the adjoint of this derivative,

$$D^\top S(\theta) : \mathbb{R}^n \to \mathbb{R}^p, \qquad \bar{\theta} = D^\top S(\theta)\, \bar{x}. \tag{B.23}$$

Definition: Adjoint Matrix

The adjoint (or Hermitian) matrix $A^\top$ is the transpose of A where complex conjugation is applied to each element. For real matrices, the adjoint is identical to the transpose.

The chain rule gives

$$D^\top S(\theta) = D^\top C(\theta)\, D^\top s(d)\, D^\top R(y^\star). \tag{B.24}$$

Thus the backward pass through the CVXPYLayer consists of:

1. Applying $D^\top R$ to $\bar{x}$ to obtain an adjoint $\bar{y}$ with respect to the primal–dual variables y,
2. Applying $D^\top s$ to $\bar{y}$ to obtain an adjoint data $\bar{d}$ with respect to the cone data d,
3. Applying $D^\top C$ to $\bar{d}$ to obtain the gradient $\bar{\theta}$ with respect to the original parameters.

The main mathematical difficulty lies in computing the adjoint of the cone solver $D^\top s$ since the retriever R and the canonicalizer C are affine.

B.4.1. Differentiation of the Retriever

The retriever $R : \mathcal{Y} \to \mathbb{R}^n$ extracts $x^\star$ from the primal–dual variables $y^\star$. If $x^\star = Pz^\star$ for a fixed selection matrix $P \in \mathbb{R}^{n \times n_z}$, then the Jacobian of R with respect to y is a fixed linear map, and its adjoint acts as follows. Given $\bar{x} \in \mathbb{R}^n$, set

$$\bar{z} = P^\top \bar{x}, \qquad \bar{s} = 0, \qquad \bar{\nu} = 0, \qquad \bar{r} = 0, \tag{B.25}$$

so that

$$\bar{y} = (\bar{z}, \bar{s}, \bar{\nu}, \bar{r}) \tag{B.26}$$

is the adjoint fed into the cone-solver backward pass.

B.4.2. Implicit Differentiation of the Cone Solver

We summarize here the differentiation of the cone solver. For more information, please refer to [102].

Given $\bar{y}$, KKT/Newton matrix H (i.e., the linearization of the KKT system around the optimal primal-dual point $y^\star$) is the Jacobian of the normalized residual map $\mathcal{N}$:

$$H = D_y \mathcal{N}(y^\star, d), \tag{B.27}$$

and the data sensitivity:

$$B = D_d\mathcal{N}(y^\star, d) \tag{B.28}$$

Then, the adjoint linear system can be solved for the Lagrange multiplier of the linearized primal-dual constraints in the implicit differentiation formula w as follows:

$$H^T w = \bar{y} \tag{B.29}$$

The adjoint of the problem data $\bar{d}$ can be computed as

$$\bar{d} = -B^T w. \tag{B.30}$$

Finally, we have that

$$D^\top s(d) = -B^\top H^{-\top} \tag{B.31}$$

In practice, when H is singular or ill-conditioned (e.g., at non-differentiable points), a least-squares or regularized solution is used, which corresponds to a generalized derivative suitable for automatic differentiation.

B.4.3. Differentiation of the Canonicalizer

As established in Section B.2, when the data are stacked into a vector d, one has

$$d = M\tilde{\theta}, \tag{B.32}$$

for a fixed sparse matrix M. The Jacobian of C with respect to θ is therefore constant. Given an adjoint $\bar{d}$, the adjoint with respect to the parameters is

$$\bar{\tilde{\theta}} = M^\top \bar{d}. \tag{B.33}$$

The last component of $\bar{\tilde{\theta}}$, corresponding to the constant "1" in $\tilde{\theta}$, is discarded, and the remaining entries form the gradient $\bar{\theta} = \partial L/\partial\theta$. These are reshaped to match the shapes of the original CVXPY parameter tensors.

B.5. Summary

A CVXPYLayer realizes the solution map of a parameterized convex optimization problem and makes it differentiable by:

1. Representing the problem in a DPP-compliant form so that the cone program data (A, b, c) depend affinely on the parameters θ,
2. Decomposing the overall solution map as
   $$S(\theta) = R\big(s(C(\theta))\big), \tag{B.34}$$
   where C and R are affine maps and s is the cone solver solution map,
3. Applying the chain rule in reverse mode, with the nontrivial component $D^\top s$ obtained via implicit differentiation of the KKT system, which reduces to solving a linear system with the KKT matrix and applying the operator $B^\top$.

The result is an optimization layer whose forward pass solves a convex program and whose backward pass solves a linear system involving the KKT/Newton matrix, while parameter dependence is handled by precomputed sparse linear operators arising from the DPP-based canonicalization.

APPENDIX C.

Negative Log-Likelihood and Dynamic RC Model: Demonstration

In section 5.4, we explain that training the proposed dynamic RC model is challenging. Here, we demonstrate that

$$\mathcal{L}_{\text{NLL}}(\mu_\epsilon, \sigma_\epsilon, \epsilon^{\text{true}}; \theta) = \mathcal{L}_{\text{NLL}}(\mu_{\tau^{\text{in}}}, \sigma_{\tau^{\text{in}}}, \tau^{\text{true}}; \theta), \tag{C.1}$$

where $\epsilon^{\text{true}} := \tau^{\text{true}} - \hat{\tau}_{\text{RC}}^{\text{in}}$ is the perturbation and thus the target value of μ_ϵ, with $\hat{\tau}_{\text{RC}}^{\text{in}}$, the forecast of the RC model without residual.
By definition of log-likelihood, one gets

$$\frac{1}{2}\left[\log \sigma_\epsilon^2 + \frac{(\mu_\epsilon - \epsilon^{\text{true}})^2}{\sigma_\epsilon^2}\right] \stackrel{?}{=} \frac{1}{2}\left[\log \sigma_{\tau^{\text{in}}}^2 + \frac{(\mu_{\tau^{\text{in}}} - \tau^{\text{true}})^2}{\sigma_{\tau^{\text{in}}}^2}\right]. \tag{C.2}$$

Since ϵ is the only probabilistic parameter, we have $\sigma_\epsilon = \sigma_\tau$, which leads to

$$|\mu_\epsilon - \epsilon^{\text{true}}| \stackrel{?}{=} |\mu_{\tau^{\text{in}}} - \tau^{\text{true}}|. \tag{C.3}$$

Given that $\mu_\tau^{\text{in}} = \mu_\epsilon + \hat{\tau}_{\text{RC}}^{\text{in}}$ and the definition of ϵ^{true}, we have

$$|\mu_\epsilon - (\tau^{\text{true}} - \hat{\tau}_{\text{RC}}^{\text{in}})| \stackrel{?}{=} |\mu_\epsilon + \hat{\tau}_{\text{RC}}^{\text{in}} - \tau^{\text{true}}|. \tag{C.4}$$

And the equality is verified.