Australia's National
Science Agency

Role of price-based aggregated control of CER in market dispatch

Nariman Mahdavi, Xinlin Wang, Vahid Aryai, Ghulam Mohy-ud-din, Subbu Sethuvenkatraman

May 2026

Citation

Mahdavi N, Wang X, Aryai V, Mohy-ud-din G, Sethuvenkatraman S (2026) *Role of price-based aggregated control of CER in market dispatch*. CSIRO, Australia.

Important disclaimer

CSIRO advises that the information contained in this publication comprises general statements based on scientific research. The reader is advised and needs to be aware that such information may be incomplete or unable to be used in any specific situation. No reliance or actions must therefore be made on that information without seeking prior expert professional, scientific and technical advice. To the extent permitted by law, CSIRO (including its employees and consultants) excludes all liability to any person for any consequences, including but not limited to all losses, damages, costs, expenses and any other compensation, arising directly or indirectly from using this publication (in part or in whole) and any information or material contained in it.

CSIRO is committed to providing web accessible content wherever possible. If you are having difficulties with accessing this document please contact csiro.au/contact.

Acknowledgement of Country

CSIRO acknowledges the Traditional Owners of the lands, seas and waters of the are that we live and work on across Australia and pays its respects to Elders past and present. CSIRO recognises that Aboriginal and Torres Strait Islander peoples have made, and will continue to make, extraordinary contributions to Australian life including in cultural, economic, and scientific domains.

Contents

Figures

Tables

Acknowledgments

Australian Research for power system transformation (AR-PST)

Executive summary

In this project, we look at the whole electricity system from price responsive CERs located at the low-voltage (LV) level to their aggregations at the medium-voltage (MV) level, and a wholesale energy and FCAS market determining the wholesale spot price based on a synthetic National Electricity Market (NEM) model. In our proposed framework in Figure 1, we investigate the impact of spot price on the behaviour of an aggregation of price responsive resources (PRRs) through the identification of simple Flexibility Functions (FFs). FF is a map between spot price and aggregated demand of flexible loads in response to the price signal.

While the proposed framework in this project is applicable to any type of flexible loads, our focus is on the flexibility from HVAC systems in commercial buildings. The goal of a local optimisation in a building at the LV level is to minimise the chiller electricity demand or cost while satisfying thermal comfort and local grid requirements. Since the electricity demand of a single building is not at the required level to be connected to the synthetic NEM model, we introduce a total demand multiplying factor that magnifies the demand profile. The FF is then found based on an ensemble of such magnified buildings each of which using a different set of local optimisation goal and/or constraint. We will finally investigate how a change in dispatchable aggregation of flexibility can alter the spot price.

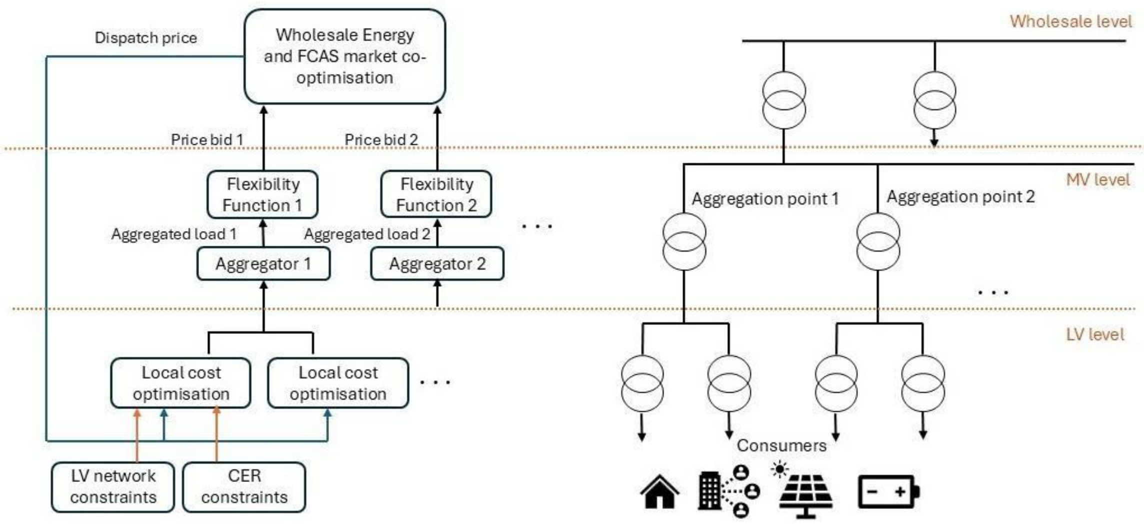

Figure 1 The proposed framework for studying the impact of price responsive resources on market dispatch

In summary, this research includes

- Implementation of the overall framework in Figure 1, where the Julia package NEMX.jl for energy and FCAS market co-optimisation is combined with Python Gym package for local RL-based chiller optimisation. The integration of these two environments includes running some standard testing scenario to make sure consistent results.
- Improving the existing NEMX package for timeseries analysis. While the original NEMX package runs snapshot optimisation scenarios, it cannot be run in timeseries mode required for flexibility analysis.

- Understanding the impact of variability and uncertainty of flexible loads on the dispatch price. Here, we investigated the impact of time-varying load profile and synchronous change in demand on the spot price.
- Investigating a potential spot price reduction by scheduling an aggregation of flexible loads and understanding the impact of their sizes, capacity-price bidding values, and locations.
- Understanding the impact of synchronous local optimisation (in an ensemble) on the dispatch price. This is simulated by training RL agents to receive AEMO's historical spot price signal and minimise the cost of energy at a building, while minimising the thermal discomfort and grid constraints.
  a) To diversify the experiments, a range of optimisation goals and constraints have been considered.
  b) Only a fixed operating envelop (OE) has been considered. We did not consider dynamic operating envelope.

1 Introduction

In an electricity market, a system operator schedules energy production and consumption, and deploy spinning or non-spinning reserve, e.g., to accommodate intermittent renewable energy sources. Australian electricity system is changing with high levels of Consumer Energy Resources (CER) ranging from batteries, PV and EV, where AEMO's ISP 2024 predicts coordinated CER storage will be nearly 50% of dispatchable capacity in the NEM by 2050. CERs, however, comes with uncertainty from environmental factors that may impact generation/consumption patterns to uncertainties on how consumers may respond to change in price, e.g., Time of Use tariffs. To mitigate the risks and utilise the opportunities, Australian government has developed CER roadmap and AEMC is currently carrying out rule changes for Integrating Price Responsive Resources (IPRR) enabling participation of CER in the National Electricity Market (NEM).

While PRR (price response resources) are capable of shaping demand in response to market signals, how their implicit flexible response will impact capacity allocation is not known. For instance, CER operation within the optimisation of the wholesale market and options for CER dispatching have identified by AEMO as highest priority focus areas. This includes incorporating Dynamic Operating Envelopes (DOEs) in the optimisation. Existing economic dispatch algorithms proposed in literature have mostly focused on the improvement of algorithm performance and consideration of more operational constraints such as spinning reserve and generator ramp rates. They did not dynamically engage an aggregation of CERs with time-varying price mechanisms such that the end-users are motivated to control their CERs to reduce electricity bill. Additionally, if these resources are effectively integrated into the wholesale market dispatch process, they can deliver essential power system services, and hence, i) reduce dispatch costs in the NEM, ii) reduce energy prices, and iii) support with reliability of grid.

To make that happen and with the increasing penetration of CERs, a coordination between Distribution Network Service Providers (DNSPs) and AEMO as the market operator is required. While the need for such coordination is imminent, there is not yet an interaction mechanism between AEMO and DNSPs to facilitate that. In this regard, there are two major considerations: i) CER need to be able to deliver these services while maintaining network reliability within network constraints, ii) Consumers who are the owner of these assets need to have simple ways to allow their resources to deliver grid services and be financially compensated in the easiest form.

Use of price signal to coordinate CER is considered an attractive approach due to its simplicity of implementation but has not been studied well. Although activities like Project Edith have attempted to evaluate the potential of price responsive CER on network services, price responsive CER coordination and its effect in delivering market services is not understood. This project aims to address this gap by evaluating ways that coordinated CER can participate in market dispatch while honouring network constraints and delivering energy bill savings to consumers.

This study will investigate the impact of flexible loads variability and uncertainty, CERs local site-level cost optimisation, and DOE constraints on the optimal scheduled/dispatch price provided by our NEM dispatch tool that co-optimises wholesale Energy and FCAS markets for a given FCAS product. In this study, we focus on the flexibility of HVAC loads in a commercial building with a local optimiser minimising the cost of electricity at the LV level given the thermal comfort and local network constraints provided by a DOE signal. We then consider an ensemble of such commercial buildings to reach to the aggregator level sitting at the MV scale. The flexibility function, which is a map between price signal and demand will then be calculated at the aggregator level. To ensure scalability, flexible loads will be modelled by data-driven machine learning (ML) tools and local cost optimisation will be based on Reinforcement Learning (RL) optimisation.

Note while our focus is on HVAC flexibility, our proposed framework is still valid for any other type of flexible CER if their control is happening at the site level. Direct control of the CERs by an aggregator is out of the scope of this project. Saying that and although we did not consider any load variation by an aggregator, the implicit flexible load variation through the alteration of the price signal can be considered by the aggregator. This future work will require a second optimiser at the MV level to adjust the price signal, using the price-demand flexibility function, such that the aggregated load profile is following a desired load profile.

Part I Market optimisation

2 NEMX.jl – A Physics-Based Co-Optimized Electricity and FCAS Market Modelling Framework

The Australian Energy Market Operator-operated National Electricity Market (NEM) is one of the largest interconnected electricity markets globally, spanning long transmission distances and integrating diverse generation technologies across multiple regions. Electricity dispatch in the NEM is centrally coordinated through a real-time market clearing process, where generation offers are optimized to meet demand at minimum cost while respecting system security constraints. A defining feature of the NEM is the co-optimization of energy and Frequency Control Ancillary Services (FCAS) within the dispatch engine (NEMDE). This co-optimization ensures that both energy production and frequency regulation requirements are simultaneously satisfied in each dispatch interval, enabling efficient allocation of generation resources across competing objectives. The FCAS markets in the NEM are essential for maintaining system frequency stability in response to disturbances such as generation outages or sudden load changes. These services operate across multiple timeframes from fast contingency response (seconds) to slower regulation services and are co-optimized with energy dispatch using shared capacity constraints and system-wide requirements.

The rapid transformation of the NEM towards a high-renewable, inverter-based power system introduces significant challenges for both market operation and system security. Increased penetration of variable renewable energy sources, such as wind and solar, reduces system inertia and increases reliance on fast frequency response services. As a result, the role of FCAS markets has become increasingly critical in ensuring secure operation under uncertainty. At the same time, the expansion of High Voltage Direct Current (HVDC) interconnectors is reshaping power flows across regions. HVDC links, such as Basslink and other interconnectors, provide controllable power transfer capabilities but introduce additional complexity in market clearing due to non-linear losses, controllability, and interaction with AC network constraints. Traditional market clearing models often approximate or neglect these characteristics, leading to inefficiencies and suboptimal dispatch outcomes. Furthermore, the increasing need for real-time, security-constrained and network-aware market clearing has highlighted limitations in simplified DC-based or zonal models. Accurate representation of AC power flow physics, voltage constraints, and converter behaviour is essential to ensure feasibility and reliability in operational decisions, particularly under stressed system conditions.

To address these challenges, NEMX.jl [1]has been developed as an open-source Julia-based package for co-optimized electricity and frequency control ancillary services (FCAS) market clearing. Built on the Julia-JuMP framework, the package enables physics-based modelling of both AC and HVDC

[1] Available [Online]: https://github.com/csiro-internal/ensys-NEMX.jl, https://data.csiro.au/collection/csiro%3A66378v2, https://doi.org/10.25919/v33w-8011

networks. NEMX.jl integrates detailed market mechanisms with network constraints, providing a comprehensive platform for studying operational and planning problems in the NEM.

2.1 Overview of NEMX.jl

NEMX.jl as shown in Figure 2 is designed to replicate key aspects of the NEM dispatch engine, incorporating both market-based and physics-based formulations. It utilizes the Synthetic-NEM-2000bus dataset[2], enabling large-scale, realistic simulations of the Australian power system. The key capabilities of this tool are listed as follows:

- Power and FCAS Bids in terms of price and quantity bands.
- FCAS trapezium constraints including enablement limits, bounds, and availability.
- Dispatchable loads participation in power and FCAS markets.
- Dispatchable bidirectional units such as storage systems and aggregators.
- Regional reference node spot pricing for power and FCAS.
- Power and FCAS unit power dispatch.
- Nodal locational marginal pricing (LMP).
- Marginal Loss Factor (MLF) modelling.
- Regional FCAS targets.
- Transmission system and AC as well as HVDC interconnector flows.

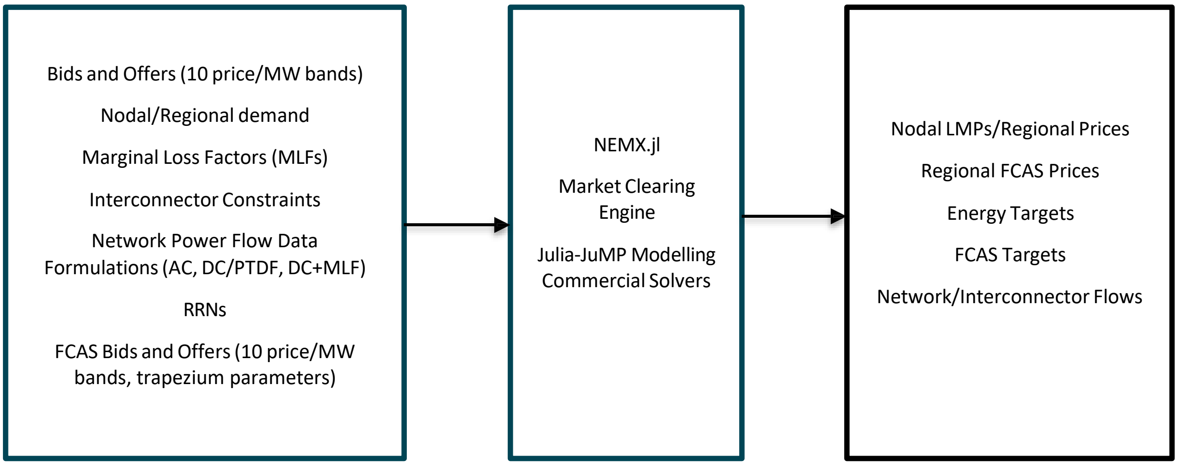

Figure 2 Conceptual architecture of NEMX.jl, illustrating the co-optimisation of the electricity and FCAS markets on a physics-based AC and HVDC network representation

[2] R. Heidari, M. Amos and F. Geth, "An Open Optimal Power Flow Model for the Australian National Electricity Market," 2023 IEEE PES Innovative Smart Grid Technologies - Asia (ISGT Asia), Auckland, New Zealand, 2023, pp. 1-5, doi: 10.1109/ISGTAsia54891.2023.10372618.
Available [Online]: https://github.com/csiro-energy-systems/Synthetic-NEM-2000bus-Data

The FCAS markets and spot price regions modelled in the NEMX.jl are listed in Table 1 and Table 2, respectively.

Table 1 10 FCAS markets modelled in NEMX

Service Name	Name Tag	Response Time Description
Regulation Raise	RReg	Minor frequency drop – primary frequency response
Regulation Lower	LReg	Minor frequency rise – primary frequency response
Very Fast Raise	R1S	Major frequency drop – 1s
Very Fast Lower	L1S	Major frequency rise – 1s
Fast Raise	R6S	Major frequency drop – 6s
Fast Lower	L6S	Major frequency rise – 6s
Slow Raise	R60S	Stabilise after major drop – 60s
Slow Lower	L60S	Stabilise after major rise – 60s
Delayed Raise	R5M	Return to normal operating band – 5m
Delayed Lower	L5M	Return to normal operating band – 5m

Table 2 Spot price regions in NEMX

Region Code	Region Name
area 1	New South Wales (NSW)
area 4	Queensland (QLD)
area 2	Victoria (VIC)
area 3	South Australia (SA)
area 5	Tasmania (TAS)

2.2 Synthetic NEM Dataset

The increasing need for high-fidelity modelling of large-scale power systems has driven the development of synthetic datasets that replicate the structural and operational characteristics of real-world networks while preserving confidentiality. Some examples from US and Europe are available in references [7], [8], and [9]. In the context of the Australian National Electricity Market (NEM), the Synthetic NEM dataset provides a comprehensive, publicly usable representation of the transmission system, enabling advanced research in system planning, operation, and market analysis.

Initially, a synthetic phasor-domain and EMT real-time digital simulation (Hypersim) dataset was developed based on a PSS®E network model [1], [2]. The dataset was subsequently extended for power network optimisation studies [3], with HVDC interconnectors later incorporated in [4]. The Synthetic NEM dataset is designed to emulate the topology, electrical characteristics, and geographic distribution of the real NEM, which spans multiple states and extensive transmission corridors. It comprises thousands of buses, transmission lines, transformers, generators, and loads, capturing the scale and operational complexity of the interconnected Australian power system. Commonly used variants, such as the S-NEM2000 dataset shown in Figure 3, include approximately 2,000 AC buses together with detailed representations of transmission infrastructure across all major NEM regions. The structural characteristics of S-NEM2000 are summarized in Table 3.

A key advantage of the Synthetic NEM dataset lies in its ability to support network-constrained and physics-based analyses. Unlike simplified test systems, it incorporates realistic impedance parameters, voltage levels, and inter-regional connections, enabling studies that require accurate representation of power flows, losses, and operational limits. This makes it particularly suitable for applications such as optimal power flow (OPF), security-constrained dispatch, and transmission expansion planning [5]. In addition to AC network representation, extended versions of the dataset include HVDC interconnectors and converter models [4], reflecting the growing importance of power electronics in the NEM. These extensions enable the study of hybrid AC–DC system behaviour, including controllable power flows, converter losses, and interactions between synchronous and inverter-based resources. Such capabilities are essential for analysing modern grid challenges associated with renewable integration and inter-regional energy exchange.

The dataset is also widely used in conjunction with time-series inputs, such as demand profiles, renewable generation traces, and market data, allowing for time-sequential and stochastic analyses. This facilitates the evaluation of system performance under varying operating conditions, including peak demand events, renewable variability, and contingency scenarios. As a result, the Synthetic NEM dataset provides a robust foundation for both operational studies and long-term planning assessments. Importantly, the Synthetic NEM dataset maintains a balance between data accessibility and realism. By abstracting sensitive infrastructure details while preserving key system characteristics, it enables collaboration between academia, industry, and policymakers. This has supported the development of open-source tools, benchmarking studies, and reproducible research in areas such as market design, system security, and resilience under extreme events [6], [5].

In summary, the Synthetic NEM dataset serves as a critical enabler for high-resolution modelling and analysis of the Australian power system. Its scalability, flexibility, and compatibility with modern optimization frameworks make it an essential resource for advancing research in power system engineering, particularly in the context of transitioning towards low-carbon and highly interconnected electricity networks.

Table 3 Structural characteristics of the S-NEM2000 dataset

Component	Quantity	Notes
AC buses	≈ 2,000	Obfuscated to preserve topology and confidentiality
AC transmission lines	≈ 3800	Includes 132 kV, 220 kV, 275 kV, 330 kV, and 500 kV circuits
HVDC interconnectors	3	Basslink (VIC to TAS), Murraylink (SA to VIC), Directlink (NSW to QLD)
Generators	≈ 261	Baseline 2024 set
Loads	≈ 912	Distributed across regional sub-area centroids
Regions	5	NSW, QLD, VIC, SA, TAS

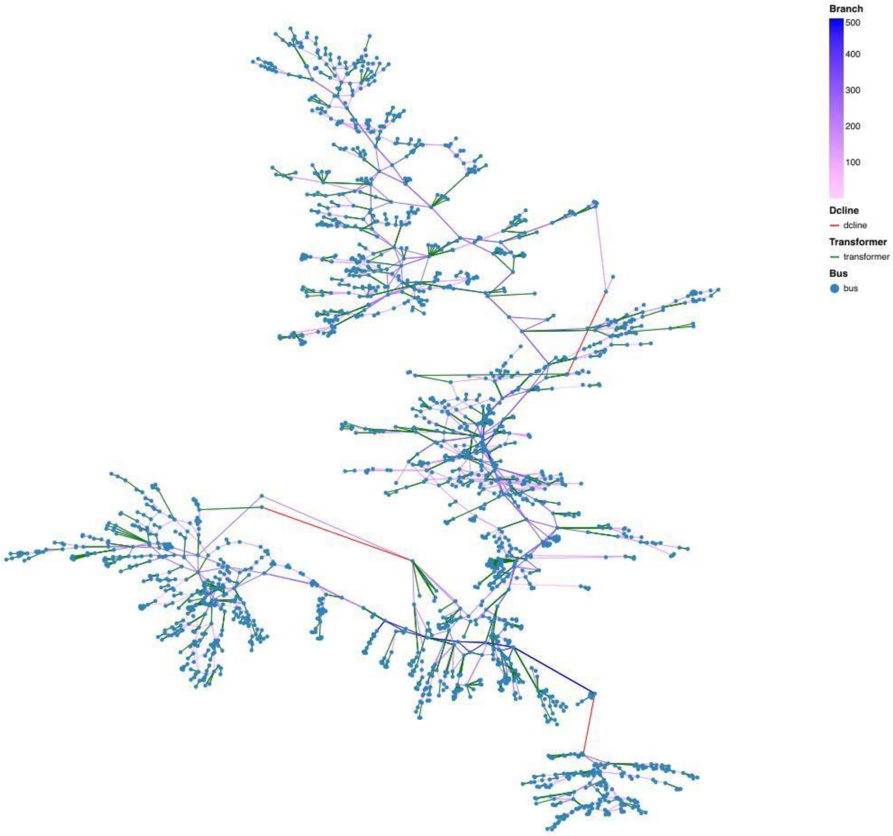

Figure 3 Synthetic NEM2000 dataset single line diagram

2.3 Example Regional Pricing Outcomes

This section presents representative regional pricing outcomes obtained from the co-optimized market clearing model for scenario 3 available in script `ex_opfcas_arpst.jl` in NEMX.jl. The results presented in Figures 4 and 5 illustrate spatial variations in electricity prices across the NEM regions, reflecting underlying network constraints, generation dispatch patterns, and inter-regional power flows. Differences in regional prices are driven by congestion, marginal loss factors, and the availability of low-cost generation, highlighting the importance of network-aware market modelling.

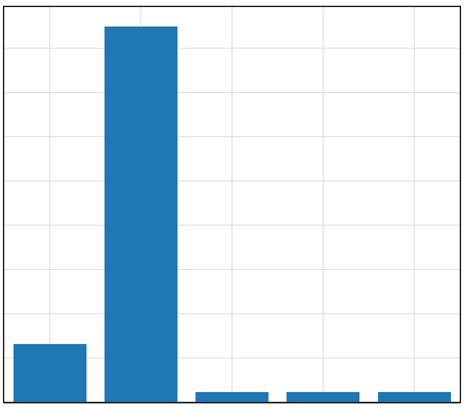

Figure 4 Energy prices at regional reference nodes

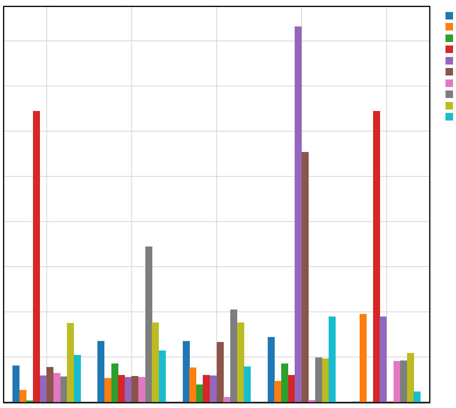

Figure 5 FCAS prices at regional reference nodes

This chapter has presented the NEMX.jl framework for physics-based co-optimised electricity and FCAS market modelling, documented its underlying synthetic-NEM dataset, and demonstrated its application to the analysis of regional energy and FCAS prices in market dispatch. The NEMX.jl framework provides the open-source, network-aware analytical foundation required to support the design and assessment of post-2027 market reforms in the NEM, and is positioned as the operational counterpart to the NEMDE in this project to study the price-responsive CER participation in the NEM as detailed in the next chapters.

Part II Local Optimisation

3 HVAC Gym model

Since in this project we are focusing on the price-based local flexibility of HVAC in a commercial building, we use CSIRO's HVAC Gym model. The Gym is a data-driven model of HVAC systems and their interactions with building environments at scale. The framework wraps a set of coupled machine-learning models each modelling a part of the HVAC-building interaction dynamics using Open AI's Gym. This way HVAC Gym models provide a standard framework for implementing data-driven control approaches, such as Reinforcement Learning (RL) agents, to control buildings' HVAC electricity demand. A full description of the methodology behind CSIRO's HVAC Gym can be found in the works of [13] with its successful real-world implementation [11]. An overview of the workflow for RK-based HVAC control using HVAC Gym given in Figure 6. In this project, we use the HVAC gym models for CSIRO Newcastle Energy Centre to implement the RL agents for controlling buildings HVAC as discussed in Section 4.

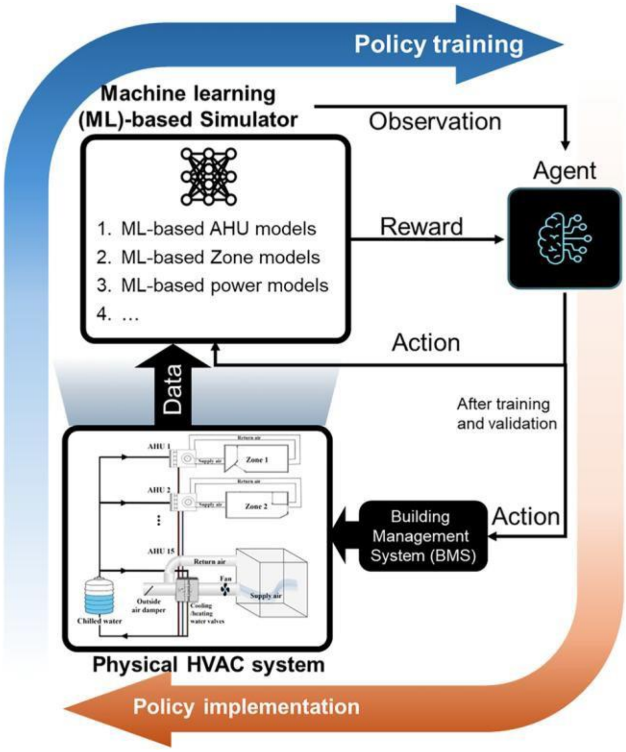

Figure 6 Overview of the workflow for RL-based HVAC control using HVAC Gym.

3.1 Simulation period

In this research, all the simulations are focused on a week of 2025 where the median temperature were the highest annually in Newcastle with the aim of considering a constrained gird period. Figure 7 shows the outside air temperature throughout that week and shows chiller power and total building load of the Newcastle site.

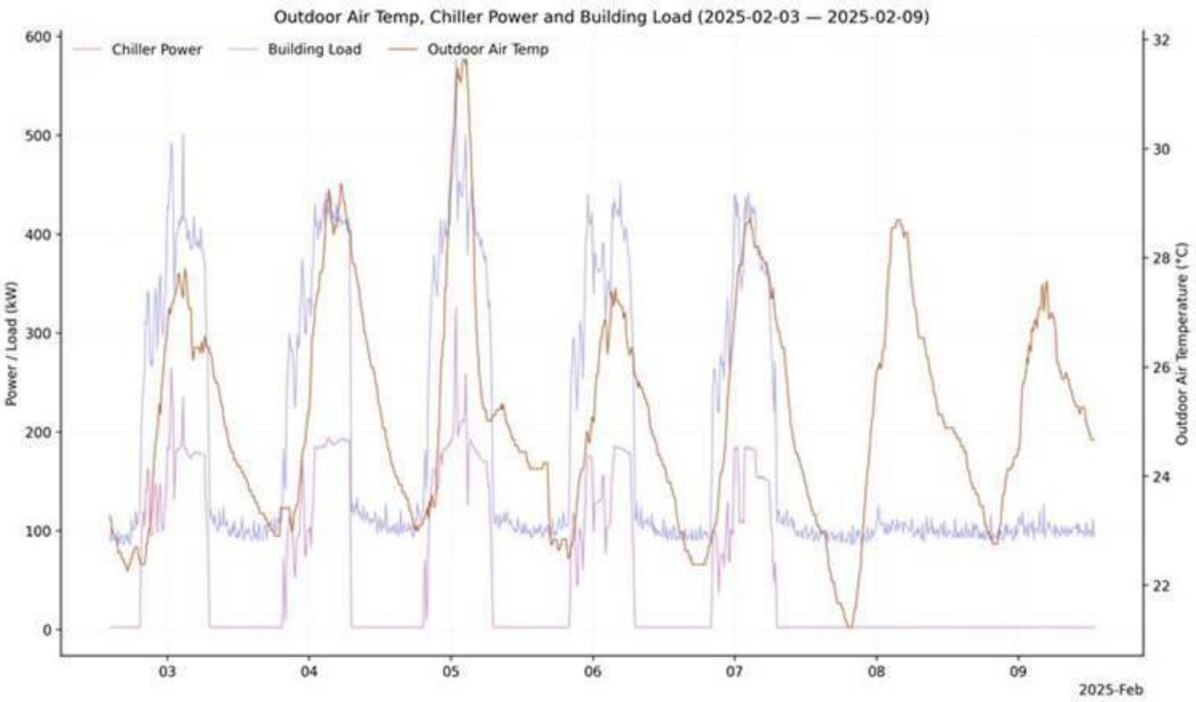

Figure 7 The warmest week of 2025 in Newcastle on median basis

3.2 Extrapolation of HVAC Gym models

Since simulation of the warmest week of the 2025 was intended in this report and given that the original HVAC gym models were developed and tested for 2023-2024, we examined the extrapolation capabilities of the models to make sure they are fit for such an outlier week simulation. To achieve this aim, we tried the following three evaluation approaches with their results listed in Table 4. As it can be seen both Newcastle and Clayton buildings Gym models provide more than 0.7 R-squared, in average, and hence, suitable for simulation.

1. Training Gym models on odd days and test on even days, and
2. Training Gym models using data from a year and testing them using data from a different time/year.
3. Training Gym models using 95th percentile of data and test them on the remaining extreme data. In other words, it checks if a model trained on typical target values (e.g., chiller power) can generalise to unusually low or high target values that were excluded from training.

Table 4 Extrapolation performance of HVAC Gym models

Approach	Building	Train period	Test period	R-Squared	RMSE
1	Newcastle	11/24-3/25 (odd days)	11/24-3/25 (even days)	0.71	0.04
1	Clayton	11/22-3/23 (odd days)	11/22-3/23 (even days)	0.87	1.58
2	Newcastle	8/24-3/25	11/23-3/24	0.73	0.048

2	Clayton	8/22-3/23	11/21-12/21	0.5	4.305
3	Newcastle	11/24-3/25 (95th percentile)	11/24-3/25 (the rest)	0.73	0.076
3	Clayton	11/22-3/23 (95th percentile)	11/22-3/23 (the rest)	0.83	5.17

3.3 Providing diversity in building models

To investigate the potential for improving model diversity when only two building models were available, an exploratory building-swapping strategy was examined. In this strategy, the Gym model developed for one site was tested using data from the other site, for example, applying the Newcastle Gym model to Clayton data, and vice versa. This approach was intended to assess if cross-site model use could expand the range of building responses available for training or evaluation. One problem to implement this building swapping, is that there are small differences in data availability between Clayton and Newcastle site. For example, Fan speed data that used in developing Newcastle Gym might not be available for Clayton site. Two different approaches used to address that data inconsistency:

1) Developing Gym models for each site only using a mutually available data between sites and ignore the data that exist for one site and not for the other site, and

2) Developing Gym models based on the original implementation but considering a constant value for the parameters those do not exist in a site. For example, assuming a constant 25% fan speed for Clayton site which does not have this data.

The figures below demonstrate the effectiveness of each of these two approaches. Figure 8 shows the underlying chiller power model of the Newcastle Gym performance if developed using the second approach and tested using Clayton site's data. Figure 9 demonstrates the results of the first approach described above. It can be concluded from metrics (for example a 0.7 vs 0.3 R-squared) that the second approach is better than the first one. This suggests that building swapping may be a useful direction for future work. However, this approach was not used in the final study. Therefore, the analysis is reported only as an indication of potential and not as part of the final research methodology.

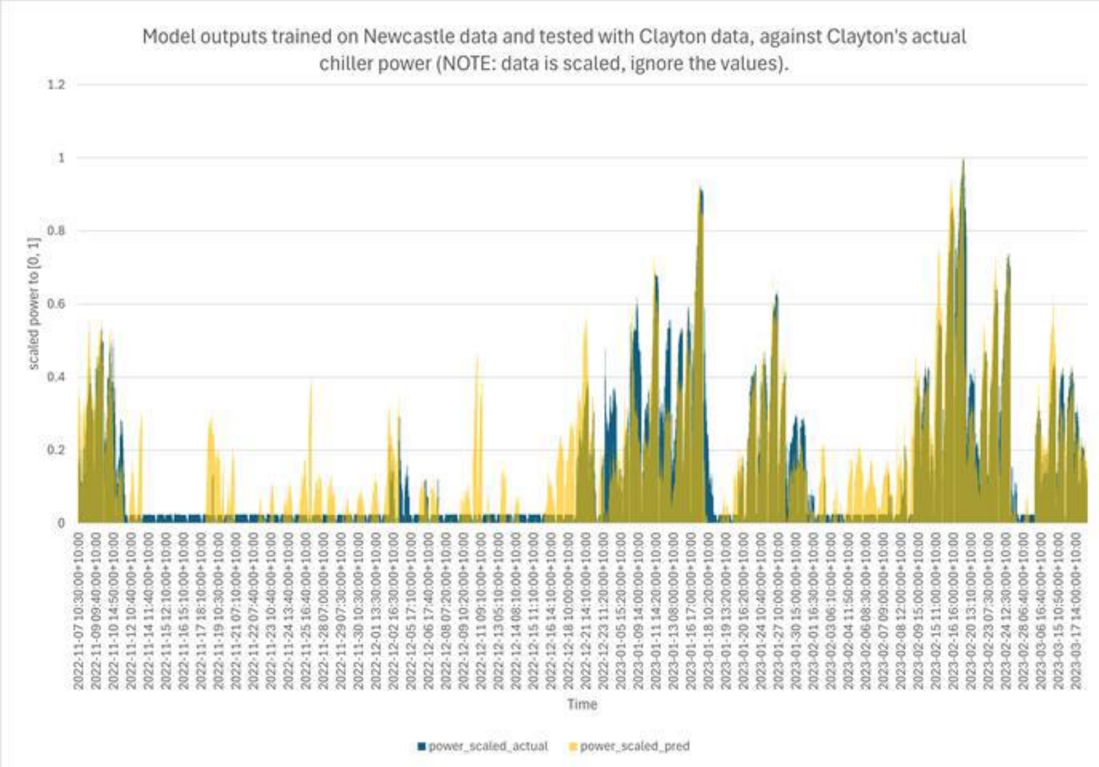

Figure 8 Train on Newcastle, Test on Clayton using the second approach

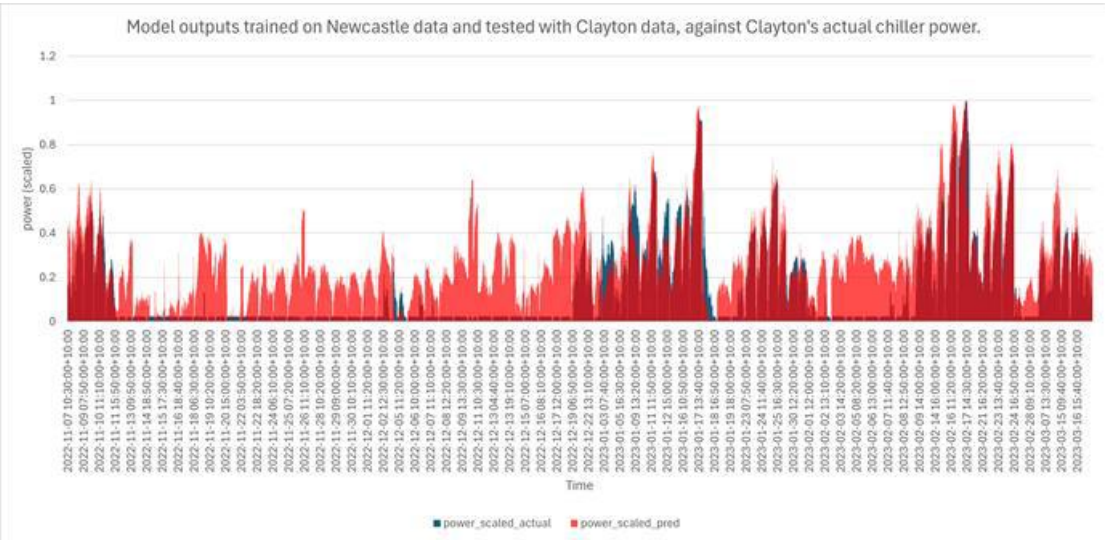

Figure 9 Train on Newcastle, Test on Clayton using the first approach

4 RL-Based Control Strategies for Smart Building Flexibility

The transition toward smart grid integration and sustainable building operations requires control systems capable of managing highly stochastic environments. Traditional heuristic and Rule-Based Control (RBC) strategies, while robust, are inherently static. They struggle to optimize the non-linear thermodynamics of building environments under time-varying disturbances, such as dynamic electricity pricing, fluctuating weather conditions, and unpredictable occupancy loads.

To address these limitations, Reinforcement Learning (RL) has emerged as a state-of-the-art, data-driven alternative. By formulating the HVAC control problem as a Markov Decision Process (MDP), RL algorithms can continuously learn and adapt optimal control policies through direct interaction with the environment, bypassing the need for computationally expensive, high-fidelity physics models.

4.1 The MDP Structure and the Critical Role of the Reward

To understand how different RL strategies manage building operations, it is necessary to pinpoint the underlying structure of the MDP, which is defined by three core interactions:

- **State Space (Observations):** The dynamic variables the agent observes at any given time step, such as indoor temperature, outdoor weather, and real-time electricity prices.
- **Action Space:** The control decisions available to the agent, such as adjusting temperature setpoints or modulating HVAC power.
- **Reward Signal:** The numerical feedback returned by the environment after an action is executed.

The reward mechanism is the fundamental driver of the RL framework. An agent possesses no inherent knowledge of "comfort" or "cost"; its sole mathematical objective is to discover a policy that maximizes cumulative rewards over time. Consequently, the formulation of the reward function dictates the entire behavioural outcome of the system. If the reward mechanism is poorly designed, the agent will exploit loopholes, leading to suboptimal or unsafe operations. Conversely, an effectively structured reward allows the system to dynamically balance complex, competing objectives.

By leveraging the dynamic nature of RL, it offers distinct operational advantages:

- **Model-Free Flexibility:** RL agents can learn optimal control policies directly from data and environmental interaction, bypassing the need to develop complex, computationally expensive mathematical models of a building's thermodynamics.
- **Exploration and exploitation:** RL thrives in stochastic environments. By continuously balancing the exploration of new actions with the exploitation of known rewarding strategies, the agent can dynamically adapt to shifting weather patterns, fluctuating energy

market signals, and changing thermal loads. This makes it uniquely suited for modern demand-response applications.

4.2 Control Architectures and Algorithm Selection

To navigate the competing objectives of minimizing operational costs while maximizing human comfort, control methodologies must be carefully structured. This involves both the overarching architectural framework and the specific algorithm deployed to execute the control policy.

4.2.1 The Evolution from General to Constrained Architectures

The management of multi-objective optimization in RL generally evolves through two foundational paradigms:

General Reinforcement Learning: In a General RL setup, competing objectives—such as energy savings versus thermal comfort—are managed by blending them into a single, unified scalar reward function [14, 15].

- **Operational Advantages:** The primary strength of General RL is its dynamic flexibility. By adjusting the weighting factors within the reward signal, the framework allows the agent to seamlessly shift its operational priorities in response to varying external market signals or environmental conditions.
- **Architectural Vulnerabilities:** The reliance on a unified scalar reward constitutes the framework's primary limitation. The tuning of penalty weights between competing objectives is highly sensitive and complex. A slight hyperparameter imbalance or an unforeseen environmental state can prompt the agent to aggressively prioritize economic savings at the severe expense of thermal comfort, or vice versa. Furthermore, General RL treats critical physical and operational limits merely as "soft" penalties rather than absolute rules, which introduces significant stability risks when deployed in physical building environments.

Constrained Reinforcement Learning: Because General RL often struggles to guarantee safe operation when faced with rigid real-world limitations, **Constrained RL** is utilized to address these critical shortcomings [16, 17]. Rather than relying solely on the sensitive tuning of penalty weights, Constrained RL reformulates the control problem as a Constrained Markov Decision Process (CMDP). In this architecture, critical operational boundaries—such as absolute indoor temperature limits or strict maximum power thresholds—are removed from the general reward pool and enforced as hard constraints. The agent is explicitly trained to optimize its primary economic objective strictly within a mathematically defined feasibility region. This transition from soft penalties to hard boundaries resolves the inherent vulnerabilities of general RL, ensuring robust, safe, and reliable real-world operation even during extreme grid events.

Note a constraint on the total power consumption can replicate fixed operating envelop signal coming from the DNSPs. During the training process, the RL agent will be penalised more if the power consumption go beyond the considered threshold. Therefore, the agent will ultimately learn to keep the consumption below the constraint value.

4.2.2 Suitability of the Soft Actor-Critic (SAC) Algorithm for Building Control

Regardless of whether a general or constrained architecture is employed, the choice of the underlying learning algorithm is critical to the system's success. For smart building applications, the Soft Actor-Critic (SAC) framework has emerged as a state-of-the-art solution [10]. SAC is an off-policy algorithm particularly well-suited for the continuous control tasks inherent to HVAC management (e.g., modulating variable air volume dampers or adjusting continuous temperature setpoints).

Its primary advantage lies in its entropy-augmented objective function; SAC operates by maximizing both the expected return and the entropy (randomness) of its policy. This inherent promotion of entropy encourages robust environmental exploration and actively prevents the agent from converging prematurely on suboptimal control strategies. The efficacy of this algorithm is well-supported by our recent research work, environment-adaptive SAC frameworks have been successfully deployed to manage complex, continuous action spaces and balance comfort with energy efficiency across distinct building typologies, including both single-zone residential and commercial office environments [10]. Consequently, SAC demonstrates superior sample efficiency and performance stability across varied runs. Its ability to reliably manage the highly stochastic and unpredictable conditions typical of physical thermodynamics makes it a critical enabler for transitioning RL control from theoretical simulation to safe, operational reality [11].

5 Experimental Framework and Reward Mechanisms

Building upon the theoretical distinctions between general and constrained architectures outlined in Section 4, this study constructs a comprehensive evaluation framework to assess their practical implications in operational environments. To systematically quantify how different optimization objectives dictate the balance between grid flexibility and occupant comfort, we formulated eight distinct control strategies. These strategies span both general and constrained RL paradigms and are driven by three primary reward functions (R). Each mechanism is designed to uniquely manage the trade-off between indoor thermal comfort (R_thermal), energy consumption (R_energy), and operational costs (R_cost) based on time-varying external signals. Note AEMO's historical spot price data have been used in this section for RL agents training.

- **Price-weighted Reward:** Balances thermal comfort and energy savings utilizing a price-derived weighting factor, w_price (ranging from 0 to 1). The agent dynamically adjusts its priorities based on real-time electricity market fluctuations.

  $R = (1 - w_{price}) \times R_{thermal} + (w_{price}) \times R_{energy}$

- **Outdoor-weighted Reward:** Adjusts the operational focus based on weather severity using an outdoor-derived weight, w_outdoor. This mechanism forces the agent to heavily prioritize thermal comfort during extreme external conditions, relegating cost to a secondary objective.

  $R = w_{outdoor} \times R_{thermal} + (1 - w_{outdoor}) \times R_{cost}$

- **Cost-Only Reward:** A specialized formulation implemented exclusively within the Constrained RL framework. The agent's strict objective is to minimize economic cost, treating thermal comfort boundaries not as a weighted reward, but as absolute, hard constraints.

  $R = -Cost$

To comprehensively account for varying occupant requirements and evaluate the impact of the "solution space" on agent performance, the strategies were tested against two distinct daytime thermal boundary sets: **T1** (Wide/Mild: 20°C - 26°C) and **T2** (Narrow/Strict: 22°C - 24°C). Nighttime requirements for both scenarios were maintained at a broader 15°C - 30°C. Table 5 summarizes the experimental scenarios.

Table 5 Summary of Control Strategies and Experimental Scenarios

Reward Mechanism	RL Framework	Constraint Focus	Scenarios
1. Price-Weighted	General RL	None	T1 & T2
	Constrained RL	Energy Usage	T1 & T2
2. Outdoor-Weighted	General RL	None	T1 & T2

3. Cost-Only	Constrained RL	Thermal Comfort	T1 & T2

5.1 Performance and Flexibility Analysis

The proposed strategies were evaluated using flexibility maps, which visually correlate external environmental severity (Temperature), market signals (Price), and the resulting system response (Power) and explained below.

5.1.1 General RL under price-weighted reward

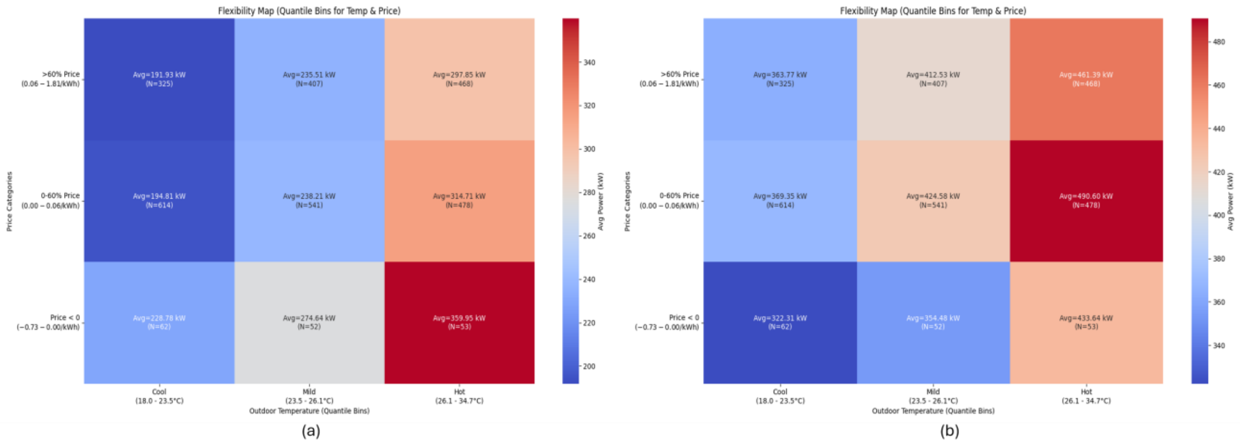

Figure 10 Performance of General RL under price-weighted reward: Comparison of (a) Wide Thermal Boundaries (T1) and (b) Narrow Thermal Boundaries (T2).

Figure 10 presents the flexibility maps among three key variables: External Environment (Temperature), Market Signal (Price), and System Response (Power). To quantify the flexibility of the proposed control strategies, we aggregated average power consumption into quantile-based bins for outdoor temperature (Cool, Mild, Hot) and real-time electricity price (Negative, Low, High).

Figure 10 (a) corresponds to the wide thermal boundary scenario (T1: 20-26C). This plot demonstrates a distinct price-sensitive response. As electricity prices increase (moving vertically from the bottom row to the top row), the agent successfully reduces power consumption, even under high thermal loads. For example, under 'Hot' outdoor conditions 26.1 - 34.7C, the average power consumption decreases from 359.95 kW during negative pricing events to 297.85 kW during peak pricing > 60% quantile. This indicates that the wide thermal tolerance allows the agent to effectively pre-cool the building or shed load in response to price penalties without violating comfort limits.

Although utilizing the same reward mechanism and RL framework, Figure 10 (b) illustrates a different control performance due to the narrow thermal boundary scenario (T2: 22-24C). This scenario exhibits behaviour dominated by thermal demand rather than price signals. The strict requirements of T2 significantly reduce the agent's feasible solution space, limiting its ability to modulate power usage for cost savings. Consequently, in the same 'Hot' temperature bin, power consumption increases alongside price, rising from 433.64 kW (negative price) to 461.39 kW (peak price). This suggests that under strict constraints, the agent is forced to prioritize thermal comfort over economic flexibility, as it lacks the "thermal slack" required to reduce demand during expensive periods.

5.1.2 Constrained RL under price-weighted reward

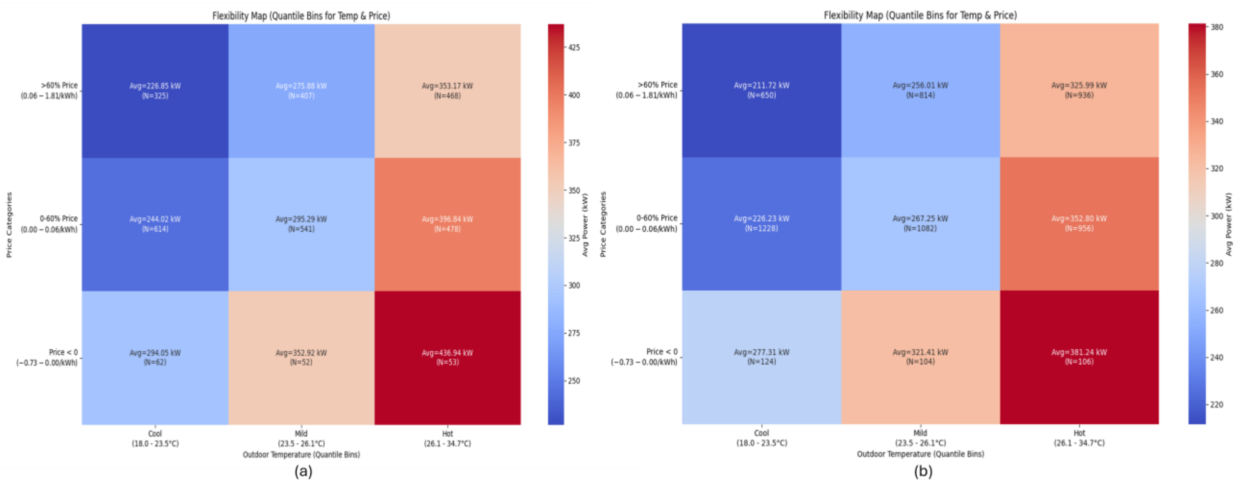

Figure 11 Performance of Constrained RL under price-weighted reward: Comparison of (a) Wide Thermal Boundaries (T1) and (b) Narrow Thermal Boundaries (T2).

Figure 11 displays the performance of the constrained RL framework. Unlike the General RL approach, which relies solely on weighted scalar rewards (thermal and energy usage rewards) to balance competing objectives, the constrained RL agent is explicitly trained to respect specific operational energy usage thresholds, specifically, the power demand is constrained to remain below 5 kW during operation. In this context, a penalty mechanism is triggered whenever energy consumption exceeds a predefined limit, effectively forcing the agent to prioritize energy reduction during critical periods regardless of the comfort-cost weighting balance.

Figure 11 (a) (Wide Boundaries T1) mirrors the trend observed in General RL, maintaining a strong price-sensitive response. Under 'Hot' conditions, the agent reduces consumption from 436.94 kW (Negative Price) to 353.17 kW (High Price). This confirms that when the solution space is large (wide thermal bounds), both General and Constrained RL can effectively optimize for cost.

However, a significant divergence in performance is observed in Figure 11 (b) (Narrow Boundaries T2). Recall that in the General RL scenario (Figure 10 (b)), the strict thermal constraints caused the agent to abandon cost optimization, leading to higher power usage during peak prices. In contrast, the Constrained RL agent in Figure 11 (b) maintains its load-shedding capability even under these strict thermal requirements. Under 'Hot' conditions, power consumption successfully drops from 381.24 kW (Negative Price) to 325.99 kW (High Price).

This result highlights the superior robustness of the Constrained RL framework. By explicitly penalizing threshold violations, the constrained agent is forced to discover more efficient control strategies—such as more aggressive pre-cooling or precise setpoint tracking—that allow it to respond to price signals even when the "thermal slack" is minimal.

5.1.3 General RL under outdoor-weighted reward

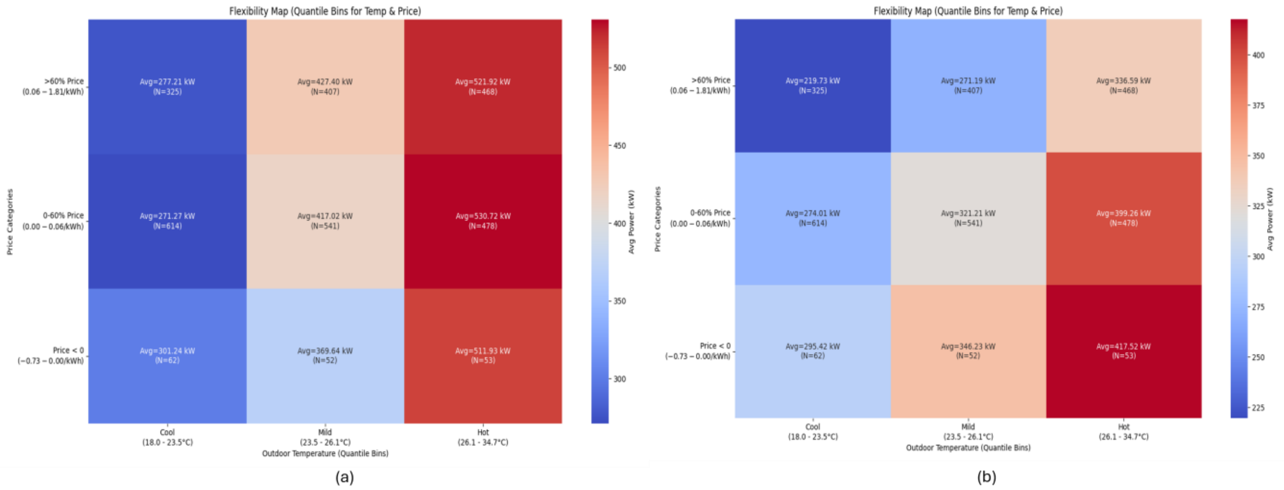

Figure 12 Performance of General RL under outdoor-weighted reward: Comparison of (a) Wide Thermal Boundaries (T1) and (b) Narrow Thermal Boundaries (T2).

Figure 12 illustrates the performance of the General RL agent under outdoor-weighted reward. In this mechanism, the weighting factor ($w_{outdoor}$) is determined by environmental severity rather than market signals. Consequently, the agent is designed to prioritize thermal comfort heavily during extreme outdoor conditions, relegating cost to a secondary objective.

This prioritization is clearly reflected in the flexibility maps, which show a distinct lack of price sensitivity. Unlike the vertical colour gradients seen in Figure 10 and Figure 11 (where power dropped as price rose), the columns in Figure 12 remain relatively uniform in colour regardless of the price level. In Figure 12 (a) (Wide Boundaries T1), the impact of the outdoor-driven weight is evident. Under 'Hot' outdoor conditions, the agent maintains extremely high-power consumption across all price bins to maximize comfort. During peak pricing events (>60% quantile), the average power consumption reaches 521.92 kW. This stands in sharp contrast to the Price-Driven Reward (Reward 1) shown in Figure 10 (a), which successfully reduced demand to 297.85 kW under identical conditions. This demonstrates that while Reward 2 leverages outdoor conditions to dynamically weight the trade-off between indoor thermal comfort and energy cost, economic flexibility is not prioritized, particularly when compared to other reward mechanisms within this test case.

A similar trend is observed in Figure 12 (b) (Narrow Boundaries T2), where the agent exhibits limited load-shedding capabilities in response to price increases. For instance, in the 'Hot' bin, consumption decreases only moderately from 417.52 kW (Negative Price) to 336.59 kW (High Price).

5.1.4 Constrained RL under cost-only reward

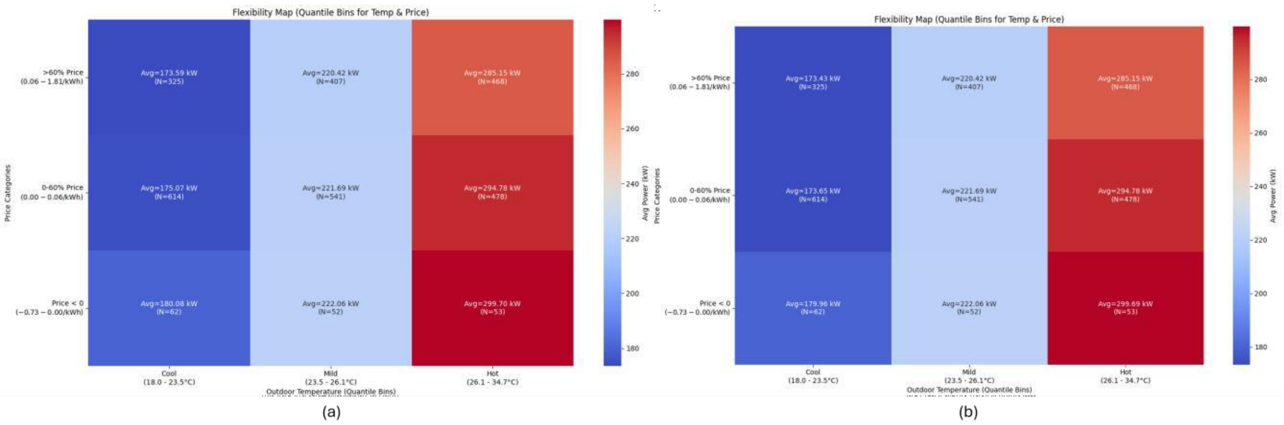

Figure 13 Performance of Constrained RL under cost-only reward: Comparison of (a) Wide Thermal Boundaries (T1) and (b) Narrow Thermal Boundaries (T2).

Figure 13 presents the performance of the Cost-Only reward mechanism implemented within the Constrained RL framework. In this formulation, the agent's primary objective is strictly to minimize operational costs, while thermal comfort is handled exclusively as a constraint. The most notable observation in this scenario is the overall reduction in energy intensity. Across all temperature and price bins, this strategy yields the lowest power consumption among all tested approaches. For example, under 'Hot' conditions and peak pricing, the average consumption is approximately 285.15 kW, significantly lower than the 325 kW observed in Reward 1 (Constrained) or the 336 kW in Reward 2. This indicates that a pure cost-minimization objective drives the agent to operate at the lower bound of the feasibility region, consuming only the bare minimum energy required to prevent constraint violations.

Furthermore, comparing Figure 13 (a) (Wide Boundaries T1) and Figure 13 (b) (Narrow Boundaries T2) reveals a remarkably consistent performance profile. This suggests that the "Cost-Only" signal is so dominant that the agent does not utilize the additional "thermal slack" provided by the wide T1 boundaries to improve comfort. Instead, regardless of whether the boundary is wide or narrow, the agent converges on the most efficient operational state allowed, demonstrating high stability but potentially lower comfort flexibility compared to the weighted reward mechanisms.

5.1.5 Aggregated Performance of All Strategies

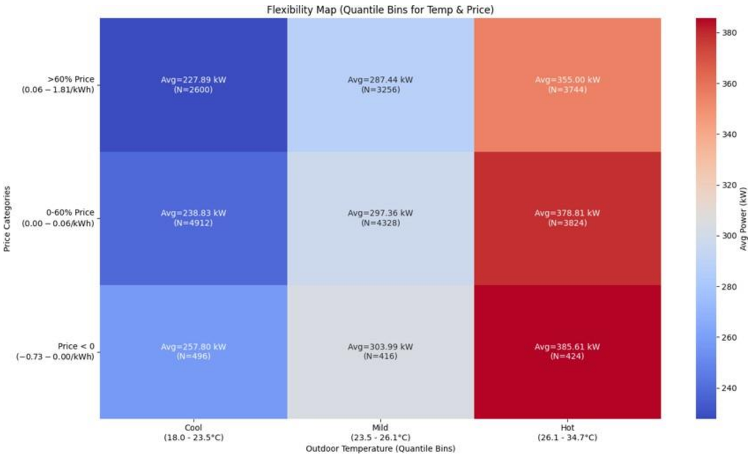

Figure 14 Aggregated Flexibility Map: Average performance across all 8 tested control strategies and thermal boundary conditions

Figure 14 presents the comprehensive flexibility map, aggregating the data from all 8 control strategies and both thermal boundary scenarios (T1 and T2). This visualization provides a holistic view of the "systemic flexibility potential" of the smart building environment when managed by the proposed portfolio of RL agents. The aggregated results reveal a robust, albeit smoothed, correlation between market signals and power consumption, where a consistent load-shifting trend remains visible. As electricity prices rise (moving from bottom to top), the average power consumption decreases across all temperature bins.

Most notably, in the 'Hot' outdoor condition, the aggregated system reduces its average power consumption from 385.61 kW during negative pricing events to 355.00 kW during peak pricing events. This represents a net reduction of approximately 30 kW. This reduction confirms that, on average, the proposed RL frameworks can deliver grid flexibility without compromising the overall operational stability of the building.

5.2 Quantitative Assessment via the Flexibility Index (FI)

To rigorously quantify the demand response capabilities of the proposed control architectures, we evaluated their performance using a formulated Flexibility Index (FI), adapted from the methodology proposed by Junker et al. for characterizing building energy flexibility [12].

The FI measures the capacity of a system to shift its operational behaviour in response to an external penalty signal—in this context, real-time electricity prices. The index compares the performance of a "penalty-aware" control strategy against a "penalty-ignorant" baseline. For this study, the baseline was established using a published general RL- based benchmark agent [10]. it trained exclusively on an outdoor-driven reward mechanism, meaning it lacked any visibility of market price signals [10]. On the other hand, all other approaches are trained with price signal.

The Flexibility Index based on real-time operational costs is defined as:

$$FI_{cost} = 1 - \frac{\sum Price(t) \times Energy_{w\ flexibility}(t)}{\sum Price(t) \times Energy_{wo\ flexibility}(t)}$$

To verify that the economic savings were achieved through actual load shedding rather than mere price arbitrage, the index was also evaluated purely on absolute energy consumption during peak periods:

$$FI_{energy} = 1 - \frac{\sum Energy_{w\ flexibility}(t)}{\sum Energy_{wo\ flexibility}(t)}$$

A positive FI indicates successful load shifting and flexibility in response to market signals, while a negative FI indicates that the penalty-aware strategy performed worse than the price-ignorant baseline, typically due to aggressive pre-cooling inefficiencies or constraint conflicts.

The calculated Flexibility Indices across the wide (T1: 20°C - 26°C) and narrow (T2: 22°C - 24°C) thermal boundaries reveal critical insights into the robustness of the evaluated architectures.

The Vulnerability of General RL (Reward No. 1)

Under the wide T1 boundary, the Price-Weighted General RL approach (R1) managed a marginal positive flexibility (FI_cost ∼ 0.37%), utilizing the broad "thermal slack" to navigate price signals. However, when subjected to the strict T2 boundary, the performance of this architecture collapsed drastically, plummeting to an FI_cost of -62.56% and an FI_energy of -68.99%. This extreme degradation quantitatively proves the primary vulnerability of General RL: when the physical solution space tightens, the agent abandons economic optimization entirely to avoid massive thermal discomfort penalties, resulting in severely rigid and inefficient grid behaviour.

The Superior Robustness of Constrained RL (Reward No. 3)

In stark contrast to the General RL models, the Cost-Only constrained RL approach (R3) demonstrated unparalleled robustness. By treating thermal limits as absolute, mathematically defined constraints rather than weighted penalties, the agent's sole objective was to minimize the cost penalty within the feasible region.

As a result, the R3 strategy was the only approach to yield consistently strong, positive Flexibility Indices regardless of the thermal boundary width. Under T1, it achieved an FI_cost of 6.24% and an FI_energy of 6.41%$. Remarkably, under the highly restrictive T2 scenario, it improved its relative performance, achieving the highest overall flexibility scores (FI_cost = 7.51%, FI_energy = 7.63%).

5.3 Price-demand mapping

Creating a function that links regional price to nodal demand is the intention of this section. To achieve this aim, we build a sequence-to-sequence regression model that learns the mapping from energy price, outdoor temperature, and hour-of-day to building electrical load (or cooling usage). The main idea is to predict not a single scalar output, but an entire window of daily load values given a corresponding window of input features. A one-to-one regression case is also investigated. In this setting, each sample contains one value of each input feature, namely energy price, outdoor temperature, and hour of day, and the model predicts the corresponding single value of building electrical load or cooling usage. This differs from the sequence-to-sequence setting, where a 24-timestep window of input features is mapped to a 24-timestep target sequence. The one-to-one model estimates the instantaneous relationship between the input variables and demand.

The data used for creating such a mapping function come from RL simulation outputs. Each input data file represents a different RL agent configuration, the description of which can be found in Table 5. To prepare the input data files, we slice out only the unseen period of each simulation to the RL (rows beyond a fixed index) and further filter to active occupancy hours by removing rows where the comfort boundary indicates unoccupied conditions. An aggregated version of all the cases is also assessed, in which the target column is the element-wise sum of load for all simulation files. Features in this case come from a single reference file, making the implicit assumption that exogenous conditions (price, temperature, time) are shared among scenarios.

For the regression task, we create sliding-window samples from the time series. Given a window length of 24 timesteps, each training example consists of the concatenation of all feature sequences over that window (so for 3 features × 24 steps = 72 input dimensions) mapped to the 24-step target sequence. This transforms the problem from a standard tabular regression into a multi-output regression in which the model must predict all 24 steps simultaneously.

The train/test split is done per file using an 80/20 chronological split (no shuffling to preserve temporal order), and the resulting windows from all files are then concatenated into global train and test arrays.

The model itself is a pipeline of standard scaling followed by a Ridge (linear) regression. Because the target is multi-dimensional (24 outputs), scikit-learn's Ridge natively fits a separate linear equation for each output step, but with shared regularisation. The scaler normalises the concatenated feature vectors to zero mean and unit variance before the linear model sees them.

Cross-validation uses a time-series aware splitting strategy that respects temporal ordering (each fold's validation set always comes strictly after its training set). The metrics computed both on the full flattened predictions (all steps pooled together) and per sequence step to check if prediction quality degrades at later horizons. The following tables 6-10 summarises the regression metrics for each of the 8 different settings in Table 5, as well as an aggregated ensemble of all of them. Additionally, Figure 15 shows the performance of the one to one and sequence to sequence models for predicting cooling usage and total load on test and train data. The developed models follow the following general mathematical formula:

$$M_c(t) = A_cP(t) + B_cT(t) + C_cH(t) + D_c$$

$$M_t(t) = A_tP(t) + B_tT(t) + C_tH(t) + D_t$$

Where $M_c(t)$ and $M_t(t)$ are cooling usage and total demand, respectively, $P(t)$ is the price $T(t)$ is the outside air temperature, $H(t)$ is time of day. The regression constants A, B, C are slopes of the learned linear model and D is its bias term. The slopes and bias terms of the derived models have been provided in Table 10 to Table 13.

The comparison between the one to one and sequence to sequence settings indicates that using a 24 timestep input window provides a small but consistent improvement. For cooling usage, the mean test R-squared increases only slightly from approximately 0.80 to 0.82. This suggests that much of the cooling response can be explained by the instantaneous values of price, outdoor air temperature, and hour of day. The sequence-to-sequence model still reduces the test RMSE in most cases, specifically for the aggregated case. This indicates that short-term temporal context contains useful information, although its added value is limited when the target variable is already strongly coupled with the current operating conditions.

The cooling demand is more directly governed by the selected explanatory variables, especially outdoor temperature and time of day. In contrast, total site load likely contains additional components that are not represented in the input feature set, such as plug loads, lighting, equipment schedules, stochastic occupant behaviour, and other non-HVAC demands. This is evidenced based on better metrics obtained for predicting cooling usage compared to that of total site load. The sequence-to-sequence setting improves total load prediction more clearly than cooling prediction, but the remaining error indicates that price, temperature, and hour of day alone might be insufficient to fully represent whole building demand dynamics. Therefore, the obtained mapping is more reliable for estimating price-responsive cooling behaviour than for explaining total site electricity use.

Table 6 Results for 1-1 regression between price, outside air temperature and hour of day and site cooling usage

Model	train_r2	test_r2	train_rmse	test_rmse	train_mae	test_mae
General_RL_R1_T1	0.89	0.90	18.72	18.36	9.60	10.85
General_RL_R1_T2	0.37	0.45	66.89	71.59	50.56	57.07
General_RL_R2_T1	0.87	0.88	41.41	38.92	34.93	32.00
General_RL_R2_T2	0.39	0.40	71.27	78.52	60.47	68.62
Constrained_RL_R1_T1	0.84	0.83	29.03	31.84	22.02	24.52
Constrained_RL_R1_T2	0.81	0.84	21.77	20.92	11.83	13.76
Constrained_RL_R3_T1	0.99	1.00	3.81	3.89	3.25	3.26
Constrained_RL_R3_T2	1.00	1.00	3.42	3.87	2.91	3.19
Summed_All_Files	0.95	0.94	111.90	124.81	85.24	97.66

Table 7 Results for 1-1 regression between price, outside air temperature and hour of day and site total load.

Model	train_r2	test_r2	train_rmse	test_rmse	train_mae	test_mae
General_RL_R1_T1	0.37	0.55	62.15	50.80	50.64	38.00
General_RL_R1_T2	0.17	0.37	97.25	86.40	80.94	70.43
General_RL_R2_T1	0.61	0.67	78.26	76.30	65.04	60.30
General_RL_R2_T2	0.21	0.35	91.88	88.55	73.09	72.90
Constrained_RL_R1_T1	0.45	0.65	64.43	53.91	52.34	41.42
Constrained_RL_R1_T2	0.28	0.51	63.87	48.17	51.84	35.73
Constrained_RL_R3_T1	0.37	0.57	61.89	51.24	51.32	39.27
Constrained_RL_R3_T2	0.37	0.57	61.62	51.09	51.03	39.25
Summed_All_Files	0.41	0.62	504.33	422.39	419.86	325.76

Table 8 Results for sequence to sequence (24 of each input to 24 of target) regression between price, outside air temperature and hour of day and site cooling usage.

Model	train_r2	test_r2	train_rmse	test_rmse	train_mae	test_mae
General_RL_R1_T1	0.90	0.90	17.95	18.10	9.94	11.12
General_RL_R1_T2	0.46	0.45	62.17	72.42	47.15	58.36
General_RL_R2_T1	0.88	0.89	39.72	36.98	33.48	30.09

General_RL_R2_T2	0.45	0.44	67.47	76.21	56.32	66.44
Constrained_RL_R1_T1	0.88	0.87	25.29	28.19	19.50	21.87
Constrained_RL_R1_T2	0.86	0.85	19.14	20.15	11.49	13.51
Constrained_RL_R3_T1	0.99	1.00	3.72	4.02	3.10	3.34
Constrained_RL_R3_T2	1.00	1.00	3.34	3.94	2.81	3.27
Summed_All_Files	0.96	0.95	100.00	116.86	77.73	93.70

Table 9 Results for sequence to sequence (24 of each input to 24 of target) regression between price, outside air temperature and hour of day and site total load

Model	train_r2	test_r2	train_rmse	test_rmse	train_mae	test_mae
General_RL_R1_T1	0.42	0.59	59.47	48.85	47.25	34.75
General_RL_R1_T2	0.28	0.41	90.71	83.20	75.00	67.24
General_RL_R2_T1	0.65	0.70	73.46	74.03	61.48	57.73
General_RL_R2_T2	0.28	0.38	87.72	87.19	68.65	71.11
Constrained_RL_R1_T1	0.51	0.68	61.33	51.99	49.00	38.51
Constrained_RL_R1_T2	0.35	0.55	60.59	46.17	48.23	32.67
Constrained_RL_R3_T1	0.44	0.61	58.40	48.88	47.64	35.16
Constrained_RL_R3_T2	0.44	0.61	58.16	48.71	47.38	35.18
Summed_All_Files	0.47	0.65	477.78	408.63	391.87	299.42

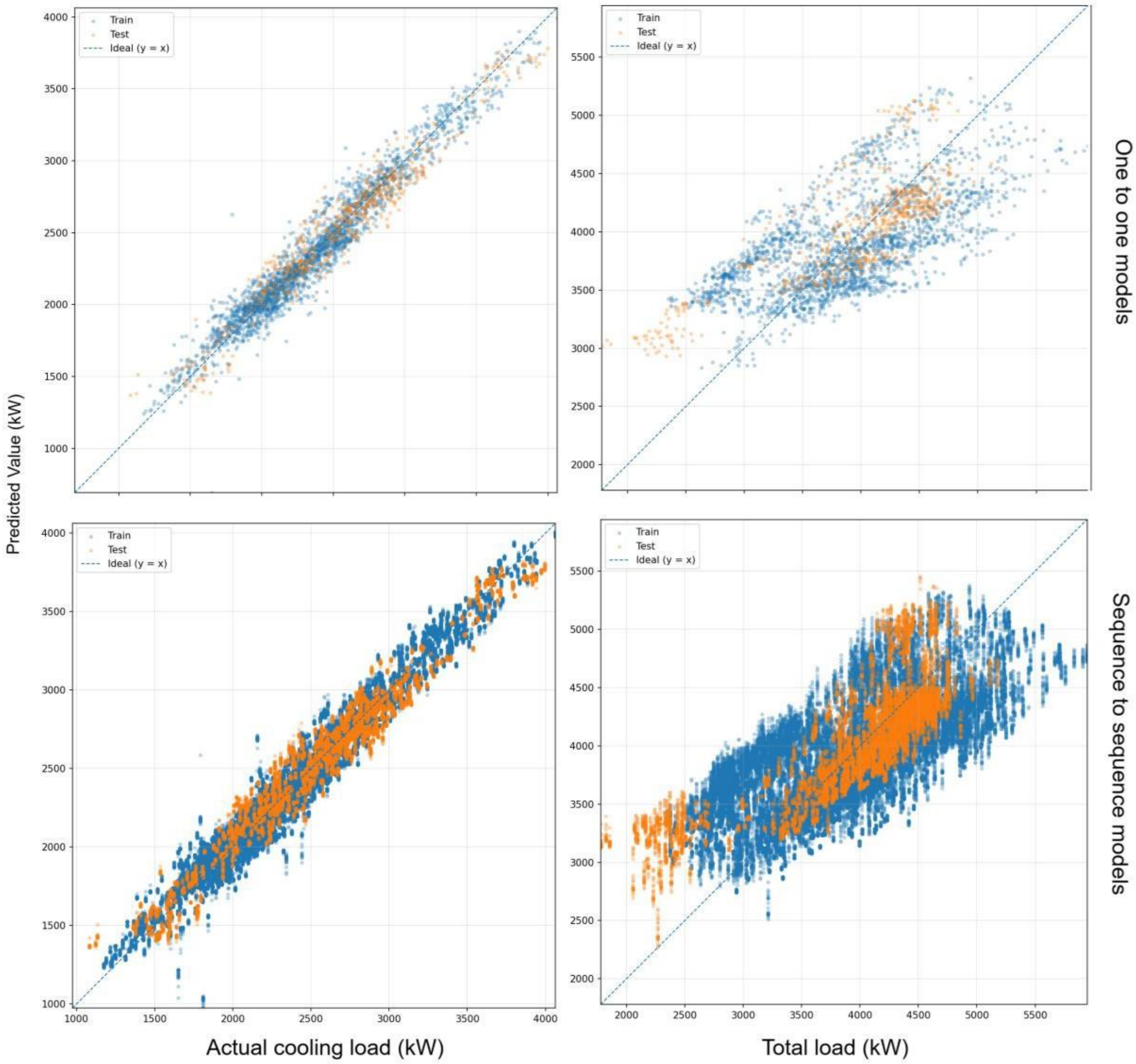

Figure 15 Performance of the developed sequence to sequence and one to one models for predicting total load and cooling usage. The top plots show the one-to-one models while the bottom plots show results for sequence to sequence models.

Table 10 Slopes and bias terms for the one to one cooling usage models.

Model	slope_price A_c	slope_outdoor_temp B_c	slope_hour C_c	Bias D_c
General_RL_R1_T1	-0.05518	18.05032	0.011514	-202.313
General_RL_R1_T2	-0.18938	17.19595	-0.49996	4.952389
General_RL_R2_T1	0.074839	36.15931	0.064647	-507.278
General_RL_R2_T2	-0.28182	18.75392	-2.02784	-119.62
Constrained_RL_R1_T1	-0.19025	22.36973	-0.48203	-241.944
Constrained_RL_R1_T2	-0.06372	15.52899	-0.80218	-125.398
Constrained_RL_R3_T1	-0.00124	17.91742	-0.27228	-217.654
Constrained_RL_R3_T2	0.000188	17.96409	-0.28541	-219.124
Summed_All_Files	-0.70637	163.8651	-4.28544	-1626.63

Table 11 Slopes and bias terms for the one to one total load models.

Model	slope_price A_t	slope_outdoor_temp B_t	slope_hour C_t	Bias D_t
General_RL_R1_T1	0.022076	16.55122	0.071135	16.29881
General_RL_R1_T2	-0.09028	15.74882	-1.10556	227.0936
General_RL_R2_T1	0.167743	33.41422	0.145567	-257.548
General_RL_R2_T2	-0.19289	16.48628	-1.52609	112.9148
Constrained_RL_R1_T1	-0.1144	20.72676	-0.34494	-20.2024
Constrained_RL_R1_T2	0.01486	13.92984	-0.60464	94.14533
Constrained_RL_R3_T1	0.083528	16.22024	-0.26085	5.730932
Constrained_RL_R3_T2	0.084527	16.37376	-0.28949	1.751004
Summed_All_Files	-0.02484	149.4511	-3.91487	180.1843

Table 12 Slopes and bias terms for the sequence to sequence cooling usage models.

Model	slope_price A_c	slope_outdoor_temp B_c	slope_hour C_c	Bias D_c
General_RL_R1_T1	-0.00571	0.76965	-0.01522	-203.145
General_RL_R1_T2	-0.01322	0.707249	-0.04334	24.81365
General_RL_R2_T1	0.009898	1.519185	0.005638	-525.418
General_RL_R2_T2	-0.02541	0.801452	-0.07034	-116.28
Constrained_RL_R1_T1	-0.0183	0.954865	-0.00717	-244.45
Constrained_RL_R1_T2	-0.00643	0.660432	-0.00113	-137.643
Constrained_RL_R3_T1	-3.13E-05	0.746515	-0.01874	-215.474
Constrained_RL_R3_T2	0.000121	0.752075	-0.02002	-218.994
Summed_All_Files	-0.05909	6.911424	-0.17032	-1636.59

Table 13 Slopes and bias terms for the sequence-to-sequence total load models.

Model	slope_price A_t	slope_outdoor_temp B_t	slope_hour C_t	Bias D_t
General_RL_R1_T1	0.007776	0.596479	-0.04367	76.96319
General_RL_R1_T2	0.002069	0.535361	-0.04351	291.556
General_RL_R2_T1	0.024734	1.27898	0.020102	-218.145
General_RL_R2_T2	-0.01083	0.600332	-0.06376	169.2696
Constrained_RL_R1_T1	-0.00476	0.777984	-0.0373	38.44794
Constrained_RL_R1_T2	0.007187	0.480775	-0.0286	146.1257
Constrained_RL_R3_T1	0.014045	0.564123	-0.03148	64.5475
Constrained_RL_R3_T2	0.014153	0.571462	-0.03369	60.23696
Summed_All_Files	0.054374	5.405495	-0.2619	629.0019

Part III Whole system simulation

6 Impact of price-responsive resources on dispatched price

For a whole system simulation from market-level optimisation (see part I) to local building level optimisation (see Part II), the SNEM 2000 grid model explained in section 2.2 should be coupled with HVAC Gyms described in section 3. This would create a closed loop for updating substation level demand using Gym and retrieving the resultant regional prices from the NEM model.

6.1 General network transformations

The optimisation framework in Part I, co-optimises the electricity and FCAS market at a given snapshot. For the closed-loop simulation of the whole system, however, a timeseries simulation is required. It basically means all the network parameters and inputs, including all the generator energy and FCAS bids, need to be updated at each iteration. Since such information are not available, we made the following simplifying assumptions:

- We take all the active power load values in the existing sNEM 2000 snapshot model and scaling them according to a given daily load profile such that the existing snapshot values occur at the time of daily peak demand.
    - We do the reactive power load scaling by preserving the original power factors of all loads across all time stamps.
    - In our experiments and to calculate the scaling factors, we use the NEM total demand across a typical week in February 2025. We first find the maximum demand across the entire week and then dividing all the other demand values by the maximum value to find the scaling factors over the entire week. The proposed scaling factors and the resulting dispatch price can be seen in Figure 16 with undesirable high pricing of up to $40,000 per MW.

Figure 16 Scaling factors for all the loads in NEMX.jl and the resulting dispatch price, when we run NEMX in timeseries mode.

- As it can be seen in Figure 16, the dispatch price is inversely related to demand, which is aligned with bidding curves of scheduled loads in Figure 17. In this illustrative example, only band 9 and 10 capacity will be dispatched as their price bids are above the spot price. In other words, the remaining bands won't be dispatched, since the spot price is not low enough to justify consumption on those bands.

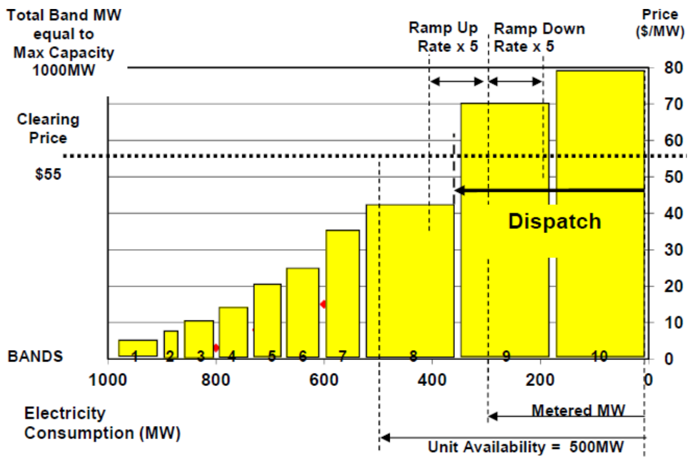

Figure 17 Price-demand bids of scheduled loads in NEM, taken from AEMO's guide to scheduled loads, July 2021

- The very high prices in Figure 16 in low demand hours is likely because of lack of tuning in the generators and scheduled demands bidding curves. Since updating ~300 of such curves is almost impossible, we propose a simple solution by using the same well-behaved arbitrary price-capacity bidding curve for all generators in the NEM. This selection ensures the flexible load demand across a day is the only deriving factor for changes in the price.
    - In our simulation, we used the following 10 quantity-price bidding values (the odd elements of this vector are the capacity, and the even elements are the associated price)

```
gen["cost"] = [0.0, -967.3, 1.1, -967.3, 4.375, 18.22, 8.75, 28.22, 8.75,
38.22, 7.75, 48.22, 7.75, 58.22, 7.75, 68.22, 7.75, 18.22, 0.0, 16057.2]
```

    - Note this well-behaved arbitrary price-capacity curve will remain the same across the entire window of a timeseries simulation.
    - Later in our experiments, we updated our quantity-price bidding values based on the fuel type of generation, but still the same across different time stamps.

```
if gen["fuel"] == "Coal"
            gen["cost"] = [0.0 0.0 2 0.98 4.0 9.85 6.0 35.58 6.5 54.02 7.0
83.74 7.5 137.9 8.0 216.69 9.0 315.14 10.0 1634.768]
        elseif gen["fuel"] == "Water"
            gen["cost"] = [0.0 0.0 1.5 0.0 4.0 10.07 6.0 31.83 6.5 49.06 7.0
72.63 7.5 90.77 9.0 117.94 10.0 281.2 11.0 1505.288]
        elseif gen["fuel"] == "Wind"
            gen["cost"] = [0.0 -965.5 0.5 -964.63 1.0 -182.5 1.5 -84.75 2.0 -
68.85 2.5 -45.05 3.0 -34.73 3.5 -2.99 6.0 19.51 7.0 160.96]
        elseif gen["fuel"] == "Solar" || gen["fuel"] == "Biomass" ||
gen["fuel"] == "CapBank/SVC/StatCom/SynCon" || gen["fuel"] == "Battery"
            gen["cost"] = [0.0 -984.7 0.1 -738.53 0.5 -172.32 1.0 -64.01 1.5
-58.1 3 -54.16 4.0 32 4.5 64 5.0 128 8.0 163.02]
        elseif gen["fuel"] == "Natural Gas" || gen["fuel"] == "Distillate"
            gen["cost"] = [0.0 0 1.0 0 2.0 124.22 2.5 198.95 3.0 298.11 3.5
377.6 4.0 557.49 4.5 1366.25 5.0 12841.65 8.0 16496.2]
end
```

- It can also be seen in Figure 16 than when the scaling factors are below 0.47, the market optimisation problem becomes infeasible and the price drops to zero. To avoid that, we consider a lower cap of 0.47, if the scaling factor becomes lower than that.

After the three above mentioned adjustments to the NEMX package, we can run a timeseries co-optimisation problem with reasonable price values as it can be seen in the following sections. It should also be noted that since we put our flexible demand in the NSW region, we only show the NSW regional reference price at bus 130.

6.2 Connecting a flexible demand to sNEM2000 model

We consider Newcastle Energy Centre chiller demand as an example of flexible demand in this study. However, since the total demand of the Newcastle site is around 1MW, see Figure 18, it is not yet at the required aggregated level to be connected to a (sub transmission) node in the sNEM 2000 model. Therefore, we have considered a constant total demand multiplying factor that simply multiplies the current demand of the Newcastle site by that factor before its connection to the sNEM 2000 model.

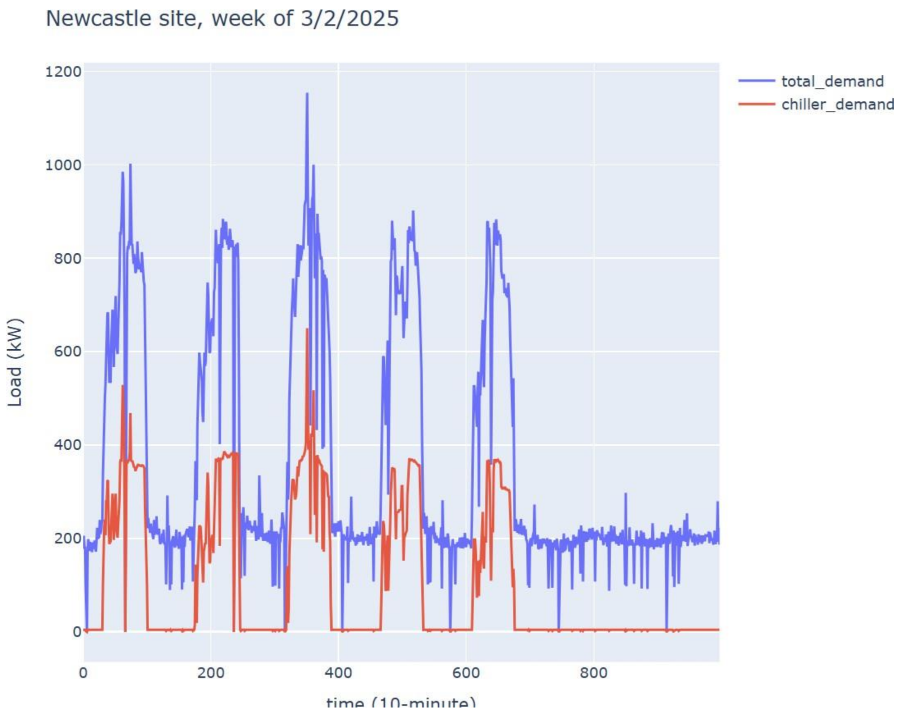

Figure 18 Flexible chiller demand and total demand of Newcastle Energy Centre in a typical week in February 2025

Based on the type of transmission lines and generator capacities/types, illustrated in Figure 19, we have identified the Node 690 of the sNEM 2000 model as the closest representation of the Newcastle site flexible demand.

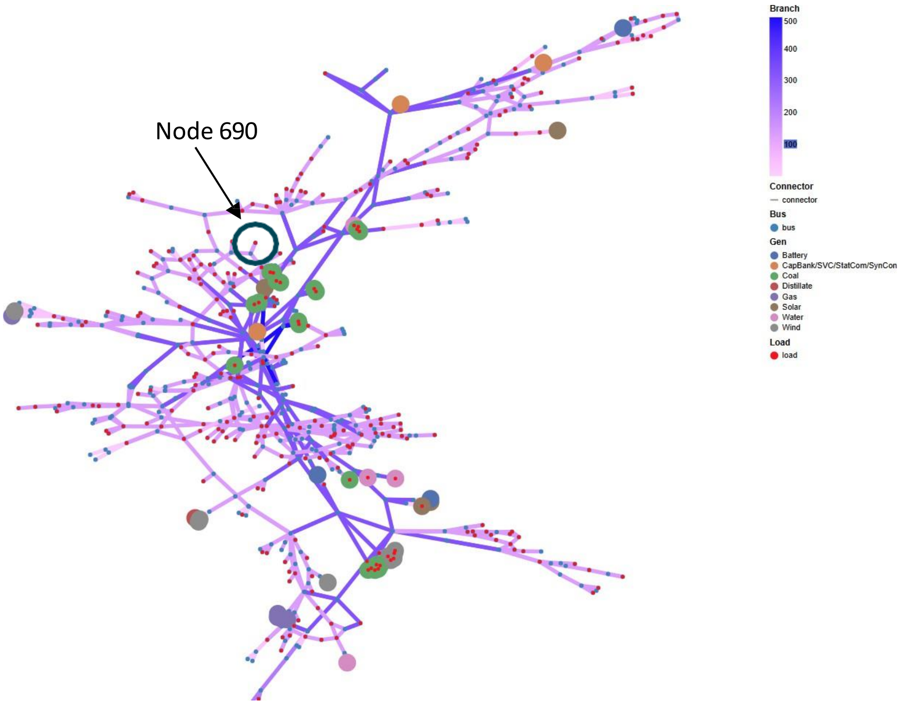

Figure 19 NSW section of the sNEM 2000 model and node 690 as Newcastle site representation

In the closed loop simulation, the specified node 690 active power is overwritten by the summation of

- underlying known non-chiller demand (fixed at each time step), and
- the dynamic flexible chiller power (HVAC Gym output varying at each step),

multiplied by the total demand multiplying factor.

6.3 Python-Julia wrapper function

For the closed loop simulation of the whole system, a wrapper function was developed in Python to set up and solve a Julia-based DC optimal power flow with co-optimised FCAS for one point in time (10-minute temporal resolution herein) during the desired simulation period. Each call to the function receives a new flexible demand value from the Newcastle Gym, a timestamp string, and a total demand multiplying factor.

The wrapper begins by looking up the corresponding timestamp in a static data frame containing Newcastle site's underlying (non-flexible) load and NEM total demand time-series. The underlying load will be added to the flexible demand value (output of the Gym after local RL price-based optimisation) to form the Newcastle site total demand. This total demand will then be multiplied by the constant multiplying factor before its attachment to the sNEM2000 model and running a market optimisation to find the price of the next step. The new price will ultimately change the flexible portion of the demand due to the local price-based optimisation, and the closed loop continues.

6.4 Simulation results

We first investigate the impact of non-scheduled flexible loads on dispatch price over the first day of a typical week shown in Figure 18. For simplicity, we exclude the RL price-based chiller demand optimisation and assume flexible demand is responding to the changes in price based on Time of Use (ToU) tariff. We consider the following two scenarios:

- Newcastle site total load with multiplying factor 50 is attached to the node 690, and
- Newcastle site total load with multiplying factor 50 is attached to the nodes 690, 670, 45, and 415 in Sydney/Newcastle area. The total load also has a constant 0.5 MW increase in demand between 5-7pm. This could represent electric vehicles or any other flexible demands that can be activated at the time of low ToU price. This means an overall increase of 25 MW aggregated demand (equivalent to 2500 EVs; 10 kW each) in each of the above 4 nodes in the sNEM 2000 model.

As it can be seen in Figure 20, and after addition of 25MW aggregated demand, the dispatch price has increased by 20% due to the synchronous jump in demand at 5pm. This phenomenon is due to the same flexible demand response to the same ToU price. However, this price jump can still happen with a dynamic pricing regime, if all flexible assets respond to the dynamic price in the same synchronous way. For example, if all EV charging is based on the same smart algorithm that minimises the cost according to the same spot price predictions, the same synchronous jump in demand will happen and a jump in price is expected.

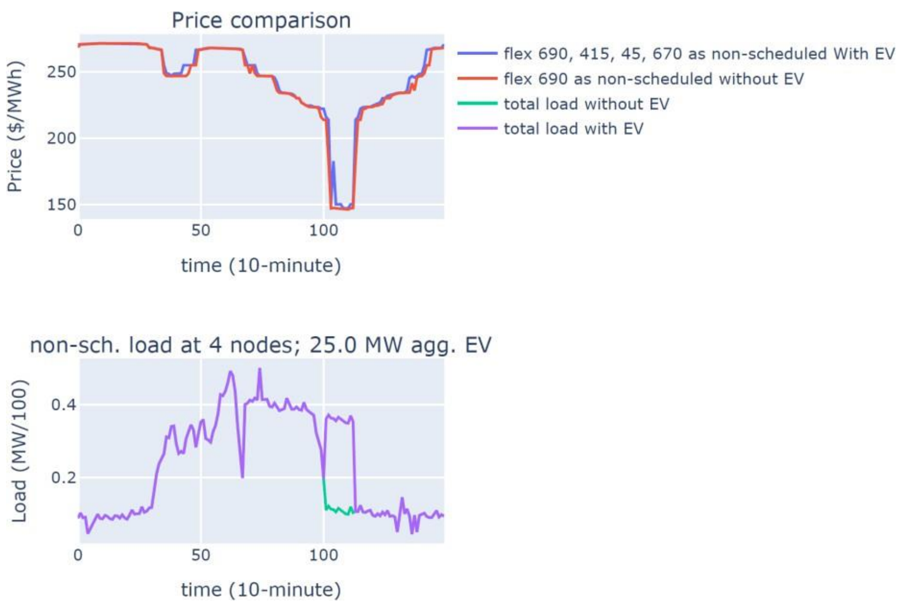

Figure 20 Impact of ToU-triggered non-scheduled flexible loads on dispatch price

Note the above dispatch price is based on the same generation cost curve for all generators regardless of their fuel types.

In the next experiment, we consider the impact of being a schedulable load on the dispatch price in the following scenarios:

- Newcastle site total load with multiplying factor 10 is attached to the node 690 as non-scheduled load, and

- Newcastle site total load with multiplying factor 10 is attached to the node 690 as scheduled load with band 1-5 as minimum demand at 0$/MW and band 6-10 as maximum demand at -10$/MW.

Note the minimum demand include all underlying loads with zero flexible demand, and the maximum demand is the total demand with full flexible load. Also note the negative bidding in band 6-10 ensures scheduling of the maximum demand if required.

As expected, and illustrated in Figure 21, the scheduled load is the same as the maximum available demand with a 10 MW peak. Therefore, the dispatch price is the same in both cases. However, if we keep increasing the multiplying factor, not all of such a cheap demand is required in the market, and hence, not all of it is dispatched. Figure 22 shows the results of the same scenario with multiplying factor 150. In this case, although the negative bidding encourages a move towards the maximum possible demand, only around 90 MW of maximum possible 150 MW is dispatched in the market. This further means the flexible cooling demand must be reduced accordingly.

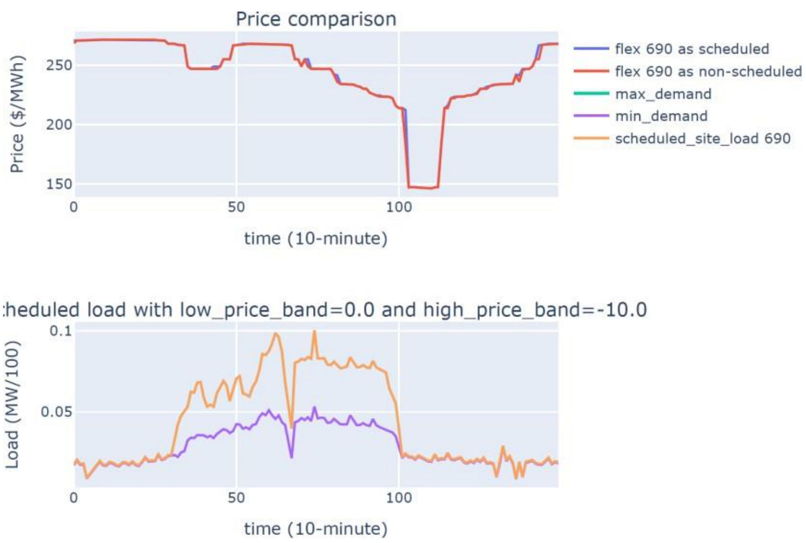

Figure 21 Impact of being schedulable load at node 690 on dispatch price with total demand multiplying factor 10

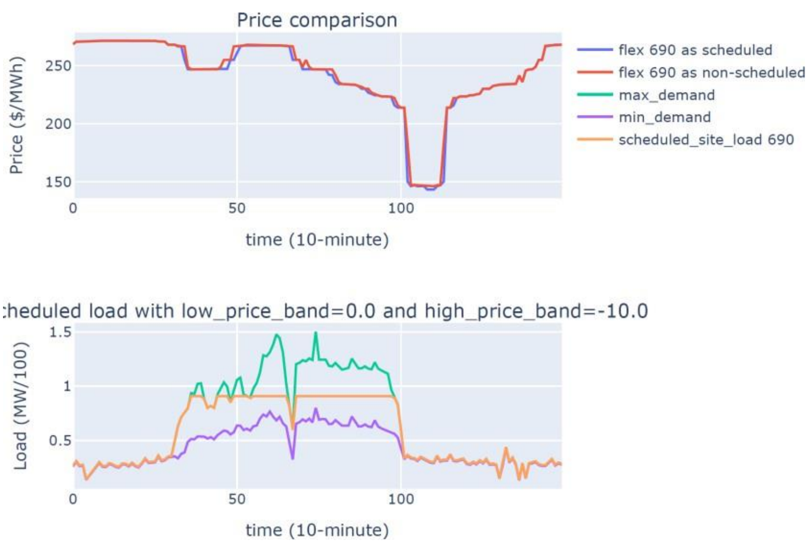

Figure 22 Impact of being schedulable load at node 690 on dispatch price with total demand multiplying factor 150

The reduction in flexible demand in Figure 22 (compared to the maximum value), causing a 5% reduction in dispatch price at around minute 500 (~8am). In other words, participation of a flexible load in the market and being schedulable can potentially reduce the spot price for everyone.

In the next experiment, we investigate the impact of scheduled load location in the following scenarios:

- Newcastle site total load with multiplying factor 150 is attached to the node 690 as non-scheduled load, and
- Newcastle site total load with multiplying factor 150 is attached to the node 605 as scheduled load with band 1-5 as minimum demand at 0$/MW and band 6-10 as maximum demand at -10$/MW.

Comparing the scheduled load prices in Figure 22 and Figure 23 reveals the importance of location of such scheduled load, where node 605 causes much lower dispatch prices than the same scheduled load at node 690 especially in the afternoon around minute 1000 (~5pm).

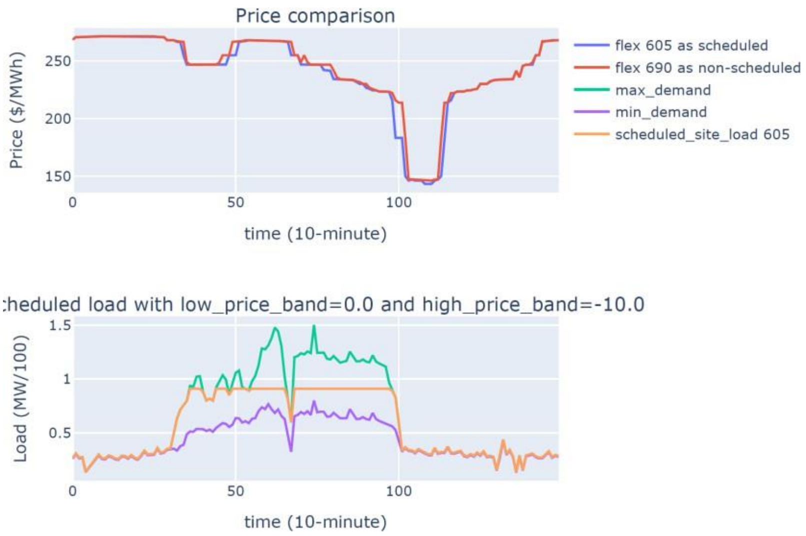

Figure 23 Impact of being schedulable load at node 605 on dispatch price with total demand multiplying factor 150

All of the above price plots in this section, so far, is based on the same generation cost curve for all generators regardless of the fuel type. If we consider the fuel-based generation bidding cost curves introduced in section 6.1, when scheduled load is at node 690 and multiplying factor is set at 150, we have the dispatch price in Figure 24 compared to the one in Figure 22.

Note, we still have the same conclusion with lower morning price in the case of scheduled load compared to the non-scheduled load at node 690. However, there are two main differences. First, the dispatch price is completely different with lower prices during the day because of having PV generation. Second, the price signal is more volatile due to the intermittent nature of renewable generation, and third, the dispatched load is at the minimum demand towards the afternoon hours.

It is worth mentioning that the NEM market model is very complex and with a little change, for example, in the bidding curve of a scheduled load, compared to the generation bidding curves, the dispatch price may change.

Figure 24 Same scenario as in Figure 22, but with different generation cost bidding curves based on fuel types

References

[1] F. Arraño-Vargas and G. Konstantinou, "Modular Design and Real-Time Simulators Toward Power System Digital Twins Implementation," in IEEE Transactions on Industrial Informatics, doi: 10.1109/TII.2022.3178713 (https://doi.org/10.1109/TII.2022.3178713

[2] F. Arraño-Vargas and G. Konstantinou, "Synthetic Grid Modeling for Real-Time Simulations," 2021 IEEE PES Innovative Smart Grid Technologies - Asia (ISGT Asia), 2021, pp. 1-5, doi: 10.1109/ISGTAsia49270.2021.9715654 (https://doi.org/10.1109/ISGTAsia49270.2021.9715654

[3] R. Heidari, M. Amos and F. Geth, "An Open Optimal Power Flow Model for the Australian National Electricity Market," 2023 IEEE PES Innovative Smart Grid Technologies - Asia (ISGT Asia), Auckland, New Zealand, 2023, pp. 1-5, doi: 10.1109/ISGTAsia54891.2023.10372618.

[4] G. Mohy-ud-din, R. Heidari, H. Ergun and F. Geth, "An AC-DC Power Flow Algorithm for Australian National Electricity Market," 2023 IEEE International Conference on Energy Technologies for Future Grids (ETFG), Wollongong, Australia, 2023, pp. 1-6, doi: 10.1109/ETFG55873.2023.10407659.

[5] Ghulam Mohy-ud-din, Rahmat Heidari, Hakan Ergun, Frederik Geth,
AC–DC security-constrained optimal power flow for the Australian National Electricity Market,
Electric Power Systems Research, Volume 234, 2024, 110784, ISSN 0378-7796,
https://doi.org/10.1016/j.epsr.2024.110784.

[6] NEMX.jl, https://doi.org/10.25919/v33w-8011

[7] Young S.J., Y.V. Makarov, R. Diao, R. Fan, R. Huang, J.G. O'Brien, and M. Halappanavar, et al. 2018.Synthetic Power Grids from Real World Models.In IEEE Power & Energy Society General Meeting (PESGM 2018), August 5-10, 2018, Portland, OR. Piscataway, New Jersey:IEEE.PNNL-SA-130491.doi:10.1109/PESGM.2018.8585792

[8] Birchfield, Adam, Xu, Ti, Shetye, Komal, Overbye, ThomasACTIVSg2000: A 2000-bus synthetic grid geolocated in Texas, http://item.bettergrids.org/handle/1001/437

[9] Gillioz, M., Dubuis, G. & Jacquod, P. A large synthetic dataset for machine learning applications in power transmission grids. *Sci Data* **12**, 168 (2025). https://doi.org/10.1038/s41597-025-04479-x

[10] Wang X, Mahdavi N, Sethuvenkatraman S, West S. An environment-adaptive SAC-based HVAC control of single-zone residential and office buildings. Data-Centric Engineering. 2025;6:e3. doi:10.1017/dce.2024.57

[11] Wang, X., Mahdavi, N., Aryai, V., Sethuvenkatraman, S., and West, S. R. (October 24, 2025). "From Simulation to Reality: A Study of Reinforcement Learning Control in Operational Building Environments." ASME. J. Eng. Sustain. Bldgs. Cities. November 2025; 6(4): 041006. https://doi.org/10.1115/1.4070007

[12] Junker, R. G., Azar, A. G., Lopes, R. A., Lindberg, K. B., Reynders, G., Relan, R., & Madsen, H. (2018). Characterizing the energy flexibility of buildings and districts. *Applied Energy*, 225, 175-182.

[13] Aryai, V., Mahdavi, N., West, S. and Henze, G., 2023, November. An automated data-driven platform for buildings simulation. In *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation* (pp. 61-68).

[14] Gao, G., Li, J. and Wen, Y., 2020. DeepComfort: Energy-efficient thermal comfort control in buildings via reinforcement learning. *IEEE Internet of Things Journal*, *7*(9), pp.8472-8484.

[15] Manjavacas, A., Campoy-Nieves, A., Jiménez-Raboso, J., Molina-Solana, M. and Gómez-Romero, J., 2024. An experimental evaluation of deep reinforcement learning algorithms for HVAC control. *arXiv preprint arXiv:2401.05737*.

[16] Zhang, B., Frison, L., Brox, T. and Bödecker, J., 2024. Constrained reinforcement learning for safe heat pump control. *arXiv preprint arXiv:2409.19716*.

[17] Paternain, S., Chamon, L., Calvo-Fullana, M. and Ribeiro, A., 2019. Constrained reinforcement learning has zero duality gap. *Advances in Neural Information Processing Systems*, *32*.

As Australia's national science agency, CSIRO is solving the greatest challenges through innovative science and technology.

CSIRO. Creating a better future for everyone.

Contact us

1300 363 400
+61 3 9545 2176
csiro.au/contact
csiro.au