HEC MONTRÉAL
École affiliée à l'Université de Montréal

Prosumer-Distributor Dynamics in Distributed Energy Resources Integration: Strategies for Load Management and Energy Optimization

par
Seyyedreza Madani

Thèse présentée en vue de l'obtention du grade de Ph. D. en administration
(spécialisation Sciences de la décision)

juillet 2024

HEC MONTRÉAL
École affiliée à l'Université de Montréal

Cette thèse intitulée :

Prosumer-Distributor Dynamics in Distributed Energy Resources Integration: Strategies for Load Management and Energy Optimization

Présentée par :

Seyyedreza Madani

a été évaluée par un jury composé des personnes suivantes :

Georges Zaccour
HEC Montréal
Président-rapporteur

Pierre-Olivier Pineau
HEC Montréal
Directeur de recherche

Laurent Charlin
HEC Montréal
Codirecteur de recherche

Sanjay Dominik Jena
ESG – UQAM
Membre du jury

Fuzhan Nasiri
Concordia University
Examinateur externe

Marie-Ève Rancourt
HEC Montréal
Représentante du directeur de HEC Montréal

Résumé

Cette thèse doctorale présente des travaux sur l'application de techniques d'optimisation avancées et des algorithmes d'apprentissage par renforcement (RL) pour améliorer les stratégies de réponse à la demande (DR) dans trois domaines clés : la technologie Véhicule-réseau ("vehicle-to-grid" ou V2G), la tarification de pointe critique ("critical peak pricing" ou CPP) et la gestion de l'énergie dans les bâtiments commerciaux. L'objectif principal est de développer des solutions innovantes qui optimisent la consommation d'énergie, réduisent les coûts et atténuent les impacts environnementaux tout en considérant la dynamique complexe des systèmes électriques modernes et les objectifs diversifiés des différents intervenants.

Le premier article explore les stratégies optimales d'investissement et de contrôle pour les ressources énergétiques distribuées ("distributed energy resources" ou DER) et les systèmes de gestion de l'énergie ("energy mangement system" ou EMS) du point de vue du distributeur et du prosommateur. Un modèle d'analyse de scénario mathématique est développé pour simuler le fonctionnement des DER et des EMS, en utilisant des données de consommation et de production réelles, ainsi que des structures de coûts du réseau électrique du Vermont. L'étude évalue différents scénarios d'investissement et structures tarifaires, en soulignant le rôle crucial de la technologie V2G dans l'amélioration de la rentabilité des investissements DER. Les résultats fournissent des aperçus précieux sur la dynamique complexe du déploiement des maisons intelligentes et soulignent l'importance de considérer les objectifs des différents intervenants lors de la conception des stratégies d'investissement et de contrôle.

Le deuxième article se concentre sur l'application des algorithmes RL pour identifier les stratégies CPP optimales en présence de profils de prosommateurs diversifiés. L'étude intro-

duit le concept de tarification dynamique ciblée comme solution pour optimiser la réponse à la demande tout en assurant l'équité entre les participants. En intégrant différents profils de prosommateurs, reflétant la pénétration croissante des DER tels que les panneaux photovoltaïques, les batteries et les véhicules électriques, l'analyse identifie la participation optimale des clients au CPP. Des simulations complètes et l'application des algorithmes RL démontrent que les stratégies CPP ciblées améliorent considérablement les performances et prolongent la viabilité des programmes CPP par rapport aux scénarios d'offre de masse. Les résultats soulignent également le rôle influent des batteries et des véhicules électriques dans la réduction de la charge de pointe, suggérant la nécessité de politiques ciblées et de structures incitatives pour encourager l'adoption de ces technologies.

Le troisième article présente un nouveau cadre DR pour optimiser la consommation d'électricité dans les bâtiments commerciaux de petite et moyenne taille. L'étude développe un modèle mathématique qui catégorise les charges des bâtiments en non-contrôlables, contrôlables, et consommation HVAC, visant à minimiser les coûts, les émissions de CO_2, et l'insatisfaction des occupants tout en maximisant la réduction de la charge de pointe. Trois algorithmes RL sont mis en œuvre et comparés à des approches heuristiques traditionnelles en utilisant huit semaines de données de consommation hivernale d'un bâtiment commercial. Les résultats montrent que les algorithmes RL, en particulier les combinaisons PPOD & TD3, atteignent des taux de réduction de charge de pointe dépassant 25 %, accompagnés de réductions significatives de coûts et d'avantages environnementaux. L'analyse intègre également l'impact des variations de température extérieure et des évaluations des risques utilisant les métriques Value at Risk (VaR) et Conditional Value at Risk (CVaR), fournissant une évaluation complète des stratégies proposées.

La thèse contribue au développement de stratégies de gestion de l'énergie efficaces et respectueuses de l'environnement, avec des implications pour la politique et les pratiques industrielles dans les transitions énergétiques durables. En tirant parti des techniques d'optimisation avancées et des algorithmes RL, les études fournissent des aperçus précieux sur la dynamique complexe du déploiement des maisons intelligentes, le comportement des prosommateurs et la gestion de l'énergie des bâtiments commerciaux. Les résultats soulignent l'importance de considérer les objectifs diversifiés des différents intervenants, le rôle

croissant des DER et le potentiel de la tarification dynamique ciblée dans l'optimisation des stratégies de réponse à la demande.

De plus, la thèse met en lumière la nécessité de politiques ciblées et de structures incitatives pour encourager l'adoption de la technologie V2G, des batteries et des véhicules électriques, car ces technologies jouent un rôle crucial dans la réduction de la charge de pointe et l'amélioration de la rentabilité des investissements DER. Les cadres et méthodologies présentés peuvent servir de fondement pour la recherche future et le développement d'outils pratiques pour soutenir la prise de décision dans le secteur de l'énergie, contribuant finalement à la transition vers un système électrique plus durable et efficace.

En outre, la thèse sert de preuve de concept du rôle important que les techniques d'optimisation avancées et les algorithmes d'apprentissage automatique, en particulier l'apprentissage par renforcement, peuvent jouer dans les applications énergétiques futures. En explorant l'application de ces méthodes de pointe dans trois domaines critiques - la technologie V2G, la tarification dynamique et la gestion de l'énergie dans les bâtiments commerciaux - la recherche démontre le potentiel de ces techniques pour optimiser la consommation d'énergie, réduire les coûts et atténuer les impacts environnementaux. La mise en œuvre réussie des algorithmes RL pour identifier les stratégies CPP optimales et obtenir un écrêtement substantiel des charges de pointe, des réductions de coûts et des avantages environnementaux dans les bâtiments commerciaux met en évidence le potentiel de transformation de ces méthodes dans le secteur de l'énergie. En tant que telle, cette thèse pose les bases de recherches et de développements ultérieurs sur les techniques avancées d'optimisation et d'apprentissage par renforcement, mettant en valeur leur promesse de façonner l'avenir de la gestion de l'énergie et de contribuer à la transition vers un système électrique plus durable et plus efficace.

Mots-clés

Ressources énergétiques distribuées; Véhicule-à-réseau; Réponse à la demande; Apprentissage par renforcement; Tarification de l'électricité; Contrôle des bâtiments; Optimisation.

Méthodes de recherche

Recherche quantitative; Modélisation mathématique; Simulation.

Abstract

This doctoral thesis presents an investigation into the application of advanced optimization techniques and reinforcement learning (RL) algorithms to improve demand response (DR) strategies in three key areas: vehicle-to-grid (V2G) technology, critical peak pricing (CPP), and energy management in commercial buildings. The overarching goal is to develop innovative solutions that optimize energy consumption—meaning to adjust and balance energy use to align with grid demand, reduce peak loads, and improve efficiency—while also reducing costs and mitigating environmental impacts, all within the complex dynamics of modern power systems and the diverse objectives of various stakeholders.

The first chapter delves into the optimal investment and control strategies for distributed energy resources (DER) and energy management systems (EMS) from the perspectives of both the distributor and the prosumer. A simulation model is developed for scenario analysis of the operation of DER and EMS, utilizing real-world consumption, generation data, and cost structures from the Vermont electricity grid. The study evaluates different investment scenarios and tariff structures, emphasizing the crucial role of V2G technology in enhancing the profitability of DER investments. The results provide valuable insights into the complex dynamics of smart-home deployment and highlight the importance of considering the objectives of different stakeholders when designing investment and control strategies.

The second chapter focuses on the application of RL algorithms to identify optimal CPP strategies in the presence of diverse prosumer profiles. The study introduces the concept of targeted dynamic pricing as a solution to optimize demand response while ensuring fairness among participants. By integrating different prosumer profiles, reflecting the increasing penetration of DERs such as photovoltaic panels, batteries, and electric vehicles, the analysis

identifies the optimal level of customer participation in CPP—referring to how customers adjust their energy usage in response to price signals. Comprehensive simulations of various prosumer behaviors and system conditions demonstrate that targeted CPP strategies significantly improve performance and extend the viability of CPP programs compared to mass offering scenarios. The results also highlight the influential role of batteries and electric vehicles in peak load reduction, suggesting that focused policy and incentive structures could be essential to offset any potential changes in system costs and to encourage the broader adoption of these technologies.

The third chapter presents a novel DR framework for optimizing electricity consumption in small and medium-sized commercial buildings. The study develops a mathematical model that categorizes building loads into non-controllable, controllable, and HVAC consumption, aiming to minimize costs, CO_2 emissions, and occupant dissatisfaction while maximizing peak load shaving. Three RL algorithms are implemented and compared against traditional heuristic approaches using eight weeks of winter consumption data from a commercial building. The results show that RL algorithms, particularly PPOD & TD3 combinations, achieve peak load shaving ratios exceeding 25%, along with significant cost reductions and environmental benefits. The analysis also incorporates the impact of outdoor temperature variations and risk assessments using value at risk (VaR) and conditional value at risk (CVaR) metrics, providing a comprehensive evaluation of the proposed strategies.

The thesis contributes to the development of efficient and environmentally conscious energy management strategies, with implications for both macro-level policy and micro-level industry practices in sustainable energy transitions. By leveraging advanced optimization techniques and RL algorithms, the studies provide valuable insights into the complex dynamics of smart-home deployment, prosumer behavior, and commercial building energy management. The findings highlight the importance of aligning the diverse objectives of various stakeholders—acknowledging that these objectives may not always be considered in parallel but may require sequential or coordinated decision-making. Additionally, the research emphasizes the increasing role of DERs and the potential of targeted dynamic pricing in optimizing demand response strategies.

Moreover, the thesis highlights the need for focused policy and incentive structures to

encourage the adoption of V2G technology, batteries, and electric vehicles, as these technologies play a crucial role in peak load reduction and enhancing the profitability of DER investments. The presented frameworks and methodologies can serve as a foundation for future research and the development of practical tools to support decision-making in the energy sector, ultimately contributing to the transition towards a more sustainable and efficient power system.

Additionally, the thesis serves as a proof of concept for the significant role that advanced optimization techniques and Machine Learning algorithms, particularly RL, can play in future energy applications. By exploring the application of these cutting-edge methods in three critical areas—V2G technology, dynamic pricing, and energy management in commercial buildings—the research demonstrates the potential for these techniques to optimize energy consumption, reduce costs, and mitigate environmental impacts. The successful implementation of RL algorithms in identifying optimal CPP strategies and achieving substantial peak load shaving, cost reductions, and environmental benefits in commercial buildings underscores the transformative potential of these methods in the energy sector. While the developed models are tailored to specific case studies, they offer valuable insights and can be adapted to a broader range of applications in future research. As such, this thesis lays the groundwork for further research and development of advanced optimization and ML techniques, showcasing their promise in contributing to the transition towards a more sustainable and efficient power system.

Keywords

Distributed energy resources; Vehicle-to-grid; Demand response; Reinforcement learning; Electricity pricing; Building control; Optimization.

Research methods

Quantitative research; Mathematical modeling; Simulation.

Contents

List of Tables

List of Figures

List of acronyms

BAS Building Automation Systems

BAT Batteries

CPE Critical Peak Event

CPP Critical Peak Pricing

CVaR Conditional Value at Risk

D3QN Double Dueling Deep Q Network

DDPG Deep Deterministic Policy Gradient

DER Distributed Energy Resources

DR Demand Response

EMS Energy Management Systems

ESS Energy Storage Systems

EV Electric Vehicle

FXD Flex D rate

GEB Grid-interactive Efficient Buildings

GHG GreenHouse Gas

HEC Hautes études commerciales

HVAC Heating, Ventilation, and Air Conditioning

IoT Internet of Things

IRR Internal Rate of Return

MDP Markov Decision Process

MPC Model Predictive Control

NPV Net Present Value

NYISO New York Independent System Operator

PhD Doctorat

PPP Prosumer Presence Percentage

PPOD Proximal Policy Optimization with Discrete actions

PPO Proximal Policy Optimization

PV Photovoltaic

PVGIS Photovoltaic Geographical Information System

RL Reinforcement Learning

RNS Regional Network Service

RTLMP Real-Time Locational Marginal Price

RTP Real-Time Pricing

SACD Soft Actor Critic with Discrete actions

SDGs Sustainable Development Goals

SOC State Of Charge

TD3 Twin Delayed Deep Deterministic Policy Gradient

TOU Time-Of-Use

VaR Value at Risk

V2G Vehicle to Grid

WCO Winter Credit Option

Acknowledgements

First and foremost, I would like to express my sincere and heartfelt gratitude to my supervisor, Dr. Pierre-Olivier Pineau. His guidance and encouragement throughout my PhD journey were indispensable. From allowing me the freedom to explore my own research interests to offering unwavering support during both the highs and lows of this process, Dr. Pineau has been an outstanding mentor. His trust in my abilities and insightful feedback have played a significant role in shaping this work, and I am truly fortunate to have had the opportunity to work under his supervision.

I would also like to thank my co-supervisor, Dr. Laurent Charlin, for his valuable advice and dedication. His thoughtful contributions, along with his patience and expertise, have enriched this thesis. I deeply appreciate his support and involvement in my research.

I am also grateful to my committee members, Dr. Sanjay Dominik Jena, Dr. Fuzhan Nasiri, and Dr. Georges Zaccour, for their constructive feedback and their willingness to share their time and expertise. Their insights have contributed greatly to improving the quality of this research.

Special thanks are due to the Fonds de recherche du Québec – Nature et technologies (FRQNT) for the funding that made this work possible. I would also like to acknowledge the support and collaboration of Mitacs, dcbel, Brainbox AI, and Schneider Electric, whose partnerships were crucial in advancing this research.

Finally, I would like to extend my deepest thanks to my parents, whose unwavering love, encouragement, and belief in me have been my greatest sources of strength. Without their support, none of this would have been possible.

I also want to dedicate this work to all those who have fought for freedom in Iran and

Canada. Their struggles and sacrifices have paved the way for opportunities like mine, allowing me the chance to pursue my education and my dreams.

Preface

This thesis consists of three articles listed as follows:

1. Madani, S., & Pineau, P. O. (2024). Investment in vehicle-to-grid and distributed energy resources: Distributor versus prosumer perspectives and the impact of rate structures. Utilities Policy, 88, 101736.

2. Madani, S., & Pineau, P. O. & Charlin, L. "Smart Grids, Smart Pricing: Employing Reinforcement Learning for Prosumer-Responsive Critical Peak Pricing" To be submitted.

3. Madani, S., & Pineau, P. O. & Charlin, L. & Desage, Y. "Towards Sustainable Energy Use: Reinforcement Learning for Demand Response in Commercial Buildings" To be submitted.

General Introduction

The global energy sector is undergoing a transformative shift, driven by the urgent need to combat climate change, ensure energy security, and transition towards a more sustainable future (Adelekan et al., 2024; Gielen et al., 2019). This transition is characterized by the rapid adoption of renewable energy sources, the decentralization of energy production through the proliferation of Distributed Energy Resources (DER), and the development of smart grid technologies that enable more efficient and flexible management of electricity supply and demand (Kabeyi and Olanrewaju, 2022). These changes are fundamentally reshaping the traditional power system paradigm, presenting both challenges and opportunities for stakeholders across the energy value chain.

One of the key challenges in this context is the effective integration of DER, such as PhotoVoltaic (PV) panels, battery storage systems, and Electric Vehicles (EVs), into the power grid (Khalid, 2024). DER have the potential to provide numerous benefits, including reduced greenhouse gas emissions, improved grid resilience, and increased energy efficiency (Zitelman, 2024). However, their intermittent and distributed nature also introduces new complexities in terms of grid operation, stability, and economic viability (Lopes et al., 2007). As a result, there is a growing need for innovative strategies and technologies that can optimize the deployment and management of DER, while ensuring the reliability and affordability of electricity supply.

Another critical aspect of the energy transition is the emergence of prosumers – consumers who actively participate in the energy market by producing, storing, and trading electricity (Oprea and Bâra, 2024). The rise of prosumers is enabled by the declining costs of DER technologies, as well as the development of smart metering and communication

infrastructures that allow for real-time monitoring and control of energy flows. Prosumers have the potential to play a significant role in the transition towards a more sustainable and decentralized energy system, by contributing to peak load reduction, providing flexibility services, and fostering the adoption of clean energy technologies (Kotilainen, 2019). However, the integration of prosumers into the power grid also presents new challenges, such as the need for appropriate market designs, tariff structures, and incentive mechanisms that can align the interests of prosumers with those of other stakeholders (Botelho et al., 2021).

In this context, demand Response (DR) has emerged as a key strategy for managing the increasing complexity and variability of the power system. DR refers to the ability of consumers to adjust their electricity consumption in response to market signals or grid conditions, such as changes in electricity prices or the availability of renewable energy. By providing flexibility on the demand side, DR can help to balance supply and demand, reduce peak loads, and improve the overall efficiency and reliability of the power system (Silva et al., 2020). However, the effective implementation of DR requires the development of advanced control and optimization algorithms that can accurately predict and influence consumer behavior, while taking into account the diverse characteristics and preferences of different consumer segments.

This thesis aims to contribute to the development of such algorithms and strategies, by exploring three key aspects of the energy transition: the investment in Vehicle-to-Grid (V2G) and DER technologies, the application of reinforcement learning (RL) for critical peak pricing (CPP) in the presence of prosumers, and the use of RL for DR in commercial buildings. Through a combination of mathematical modeling, simulation, and real-world data analysis, the thesis seeks to provide new insights and tools for optimizing the deployment and operation of DER, designing effective DR programs, and fostering the transition towards a more sustainable and efficient energy system.

The thesis also highlights the critical role that advanced optimization techniques and machine learning, particularly RL, can play in enabling the energy transition. By leveraging the power of these computational methods, the studies presented in this work demonstrate how complex challenges, such as the optimal deployment and management of DER, the design of effective demand response programs, and the improvement of energy efficiency in

commercial buildings, can be addressed in innovative and effective ways. The successful application of RL algorithms across different domains serves as a proof of concept for their potential to revolutionize energy applications in the future, by providing adaptive, robust, and data-driven solutions to the multifaceted problems posed by the evolving energy landscape. As such, this thesis contributes to the growing recognition of the transformative potential of advanced optimization and machine learning techniques in the energy sector, and underscores the importance of continued research and development in this area to support the transition towards a more sustainable and resilient energy future.

The first chapter, titled "Investment in Vehicle-to-Grid and Distributed Energy Resources: Distributor versus Prosumer Perspectives and the Impact of Rate structures," addresses the question of who should invest in and manage DER, and what tariff structures should be used to incentivize their adoption. The increasing penetration of PV panels, EVs, and V2G technologies presents significant opportunities for improving grid performance and supporting the energy transition. However, the optimal allocation of investment responsibilities between distribution companies and prosumers, as well as the design of appropriate rate structures, remains a complex and largely unresolved issue.

To shed light on this issue, the chapter develops a mathematical scenario analysis model that simulates the operation of DER and energy management systems, using real-world consumption and generation data from the Vermont electricity grid. The model considers different investment scenarios, in which either the distribution company or the prosumer invests in and controls the DER, and evaluates their impact on the profitability and system-wide costs of DER deployment. The results highlight the importance of incorporating V2G technology to enhance the economic viability of DER investments, and provide insights into the trade-offs between different investor objectives and rate structures. The economic viability of DER investments is enhanced by incorporating V2G technology because it enables prosumers to capitalize on price differentials in the electricity market, storing energy when prices are low and either using or selling it when prices are high. This ability to leverage price fluctuations makes DER investments more financially attractive, particularly under rate structures that incentivize such behavior.

The second chapter, "Smart Grids, Smart Pricing: Employing Reinforcement Learning

for Prosumer-Responsive Critical Peak Pricing," focuses on the design of CPP programs in the presence of prosumers. CPP is a dynamic pricing scheme that aims to incentivize consumers to reduce their electricity consumption during critical peak events, by offering higher prices or rebates for load reduction. While CPP has been shown to be effective in reducing peak loads and improving grid efficiency, its implementation becomes more challenging in the context of prosumers, whose consumption and generation patterns can be more variable and unpredictable.

To address this challenge, the chapter proposes the use of RL algorithms for optimizing CPP strategies in the presence of prosumers. RL is a subfield of machine learning that enables agents to learn optimal policies through trial-and-error interactions with an environment, without requiring explicit modeling of the underlying system dynamics. By applying RL to the problem of CPP design, the chapter aims to develop more adaptive and robust pricing strategies that can effectively incentivize prosumers to contribute to peak load reduction, while avoiding unintended consequences such as the emergence of new peaks.

The study employs simulations and comparative analysis of different RL algorithms, considering both mass and targeted CPP offering scenarios. The results demonstrate the potential of RL-based CPP strategies to significantly improve the performance and extend the viability of CPP programs, particularly in the presence of high prosumer penetration. The chapter also highlights the influential role of batteries and EVs in enabling more effective peak load management, and suggests the need for targeted policy and incentive structures to encourage their adoption.

The third chapter, "Towards Sustainable Energy Use: Reinforcement Learning for Demand Response in Commercial Buildings," shifts the focus to the application of RL for DR in the commercial building sector. Commercial buildings account for a significant share of global electricity consumption, and their energy use patterns are often characterized by high variability and inefficiency. As a result, there is a growing interest in developing advanced control and optimization strategies that can enable more sustainable and cost-effective energy management in commercial buildings.

The chapter presents a novel DR framework that integrates RL algorithms with a mathematical model of building energy dynamics, considering non-controllable loads, controllable

loads, and Heating, Ventilation, and Air Conditioning (HVAC) systems. The objective of the framework is to minimize energy costs, CO_2 emissions, and occupant dissatisfaction, while maximizing peak load shaving potential. The study implements and compares three state-of-the-art RL algorithms for the HVAC load, namely PPOD, D3QN, and SACD, against traditional heuristic approaches, using real-world consumption data from a commercial building over an eight-week winter period.

The results demonstrate the superiority of RL-based DR strategies, particularly the combination of PPOD and TD3 algorithms, in achieving significant peak load shaving (exceeding 25%), cost reductions, and environmental benefits. The chapter also incorporates the impact of outdoor temperature variations and risk assessment using Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) metrics, providing a more comprehensive and robust evaluation of the proposed DR framework. The findings contribute to the growing body of knowledge on the application of RL for building energy management, and have important implications for the development of more sustainable and efficient energy practices in the commercial sector.

In summary, this thesis addresses critical challenges and opportunities in the ongoing energy transition, by exploring the application of reinforcement learning algorithms, mathematical optimization techniques, and scenario analysis models, along with emerging technologies for the optimal deployment and management of DER, the design of effective DR programs, and the improvement of energy efficiency in commercial buildings. The three chapters that comprise the thesis make novel contributions to the fields of energy systems modeling, machine learning, and sustainability, and provide valuable insights and tools for policy makers, industry practitioners, and researchers working towards a more sustainable and resilient energy future.

References

Adelekan, O. A., Ilugbusi, B. S., Adisa, O., Obi, O. C., Awonuga, K. F., Asuzu, O. F., and Ndubuisi, N. L. (2024). Energy transition policies: a global review of shifts towards renewable sources. *Engineering Science & Technology Journal*, 5(2):272–287.

Botelho, D. F., Dias, B. H., de Oliveira, L. W., Soares, T. A., Rezende, I., and Sousa, T. (2021). Innovative business models as drivers for prosumers integration-enablers and barriers. *Renewable and Sustainable Energy Reviews*, 144:111057.

Gielen, D., Gorini, R., Wagner, N., Leme, R., Gutierrez, L., Prakash, G., Asmelash, E., Janeiro, L., Gallina, G., Vale, G., et al. (2019). Global energy transformation: a roadmap to 2050.

Kabeyi, M. J. B. and Olanrewaju, O. A. (2022). Sustainable energy transition for renewable and low carbon grid electricity generation and supply. *Frontiers in Energy research*, 9:743114.

Khalid, M. (2024). Smart grids and renewable energy systems: Perspectives and grid integration challenges. *Energy Strategy Reviews*, 51:101299.

Kotilainen, K. (2019). Energy prosumers' role in the sustainable energy system. In *Affordable and clean energy*, pages 1–14. Springer.

Lopes, J. P., Hatziargyriou, N., Mutale, J., Djapic, P., and Jenkins, N. (2007). Integrating distributed generation into electric power systems: A review of drivers, challenges and opportunities. *Electric power systems research*, 77(9):1189–1203.

Oprea, S.-V. and Bâra, A. (2024). Generative literature analysis on the rise of prosumers and their influence on the sustainable energy transition. *Utilities Policy*, 90:101799.

Silva, B. N., Khan, M., and Han, K. (2020). Futuristic sustainable energy management in smart environments: A review of peak load shaving and demand response strategies, challenges, and opportunities. *Sustainability*, 12(14):5561.

Zitelman, K. (2024). Advancing electric system resilience with distributed energy resources: a review of state policies.

Chapter 1

Investment in Vehicle-to-Grid and Distributed Energy Resources: Distributor versus Prosumer Perspectives and the Impact of Rate structures

Abstract

Photovoltaic panels, electric vehicles, and vehicle-to-grid technologies are becoming more common and hold significant promises to improve the grid and foster the energy transition. However, significant questions remain unanswered with respect to who should invest in this equipment and what tariff should be used. This paper examines whether the distribution company or prosumer should invest in and manage Distributed Energy Resources (DER), the ideal combination of DER to utilize, and the appropriate tariff to implement. Central to this analysis is the assessment of different stakeholder objectives, particularly from the investor's perspective, where net present value is used as the primary criterion for evaluating the different investment scenarios. Additionally, the impact of these scenarios on the annual

system cost is calculated. A mathematical scenario analysis model is developed to simulate the operation of DER and energy management systems. This model utilizes the Vermont electricity grid's real-world consumption, generation data, and cost structures. The results underscore the significance of incorporating vehicle-to-grid technology to enhance the profitability of DER investments. This inclusion of specific data sources and stakeholder criteria aims to provide insight into the complex dynamics of smart-home deployment.

1.1 Introduction

The residential subsector of electricity consumers is the second largest worldwide after the industrial subsector (U.S. Energy Information Administration (EIA), 2021c) and the largest in many countries, such as the US (with a share of 39% (U.S. Energy Information Administration (EIA), 2021a)). Equipping homes with Distributed Energy Resources (DER) and Energy Management Systems (EMS) can help the grid owner address many challenging problems the electricity grid faces. First, since the residential peak demand and Photo-Voltaic (PV) solar panel generation usually do not coincide (see for instance Figure 1.1)), self-generation alone (without storage) cannot cover peak loads. Peak load pressures the grid operator to meet high demand through expensive peaking power plants and network capacity. These resources are more expensive than usual generation resources and impose massive ramp-up and ramp-down operational costs upon the grid operator (Imcharoenkul and Chaitusaney, 2020). Moreover, during off-peak hours, the network capacity is under-utilized, while capacity limits during peak hours endanger grid reliability (Salisbury and Toor, 2016).

Second, to increase the generation capacity during peak hours, the grid operator tends to use dispatchable resources that often run on fossil fuels, increasing the grid's carbon footprint (Singh, 2021). Reducing environmental damage is another advantage of shaving the peak loads in the grid. Finally, due to the sharp decrease in battery cost in recent years as well as incentives and regulations, a substantial increase in Electric Vehicle (EV) penetration in the market is expected in the coming years (International Energy Agency (IEA), 2020; Wu et al., 2015). Introducing EVs to the prosumer's home increases power consumption and

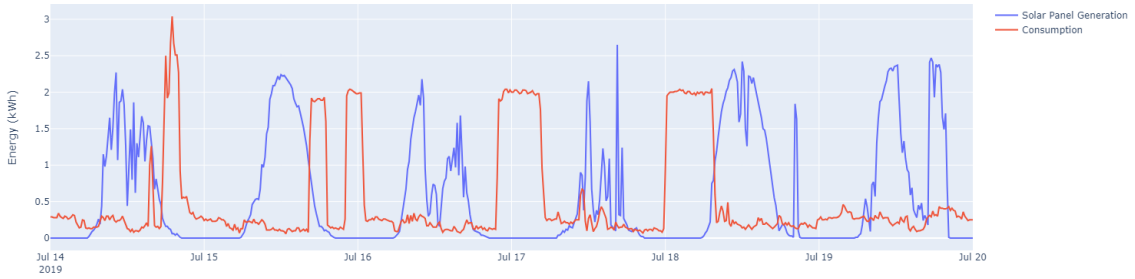

Figure 1.1: Generation and load (with 15-minute time steps) in a sample week in July

changes the load profile, especially if the vehicle is connected to the grid ("vehicle-to-grid" or V2G) so that it can feed electricity into it. V2G has the potential to shave peak loads since the batteries are often available; about 90% of the cars are parked during peak hours in California (Thomas et al., 2018). Deploying V2G can thus reduce generation costs during peak hours (Mozafar et al., 2018). On the contrary, unmanaged EV penetration may create new peaks during the formerly off-peak hours (Sioshansi, 2012).

There is a body of research studying the advantages and disadvantages of V2G. Drivers and barriers influencing the deployment of V2G are identified in (Aasbøe, 2021); among other factors, the authors conclude that increasing the battery size can boost the profitability of V2G. Also, the environmental and economic benefits of embedding V2G into EMS are quantified in (Wei et al., 2022) by formulating a multi-objective optimization problem and using a Non-dominated Sorting Genetic Algorithm; the results highlight the importance of electricity and gas prices on the system design and profitability of V2G. Bentley et al. consider the physical characteristics of EV batteries and the impact of battery degradation on the profitability of V2G (Bentley et al., 2021), and Ali et al. show that even in developing regions, using V2G returns economic benefits and reduces CO_2 emissions (Ali et al., 2020). Considering reliability, cost, and emissions as the measures of profitability, Bibak and Tekiner-Mogulkoc investigate the impact of V2G on the grid (Bibak and Tekiner-Mogulkoc, 2021) and show that V2G can increase grid efficiency. How V2G benefits EV owners trading electricity in the Netherlands is studied in (Raj, 2019). The findings of this study highlight the monetary value of V2G and show that a greater penetration of PV can increase the

profitability of V2G. Li et al. assess the profitability of V2G for EV owners, power plants, and power grid companies (Li et al., 2020). They conclude that V2G yields positive income for all parties under suitable parameters. For instance, the EV owner would gain a positive income if the peak load price is more than three times the off-peak price. Also, electric trucks have more battery capacity and can earn even more revenue by participating in V2G (Zhao et al., 2016).

However, some researchers believe that investment in V2G would not pay off due to battery degradation costs (Gough et al., 2017) and high investment costs (Peterson et al., 2010). Mullan et al. suggest that too much additional infrastructure expansion investment is needed, which V2G is not able to compensate (Mullan et al., 2012), and Curtin et al. state that the lack of savings is the critical barrier against V2G for the consumers (Curtin et al., 2019). Accordingly, the profitability of V2G is contingent on various factors, and its overall profitability remains dependent on the context.

Furthermore, investment in the consumer's side by equipping its homes with DER and EMS and enabling them to feed back electricity to the grid are well-studied solutions to address the previously mentioned problems (Krozer, 2013) which have attracted significant attention in the literature (Segreto et al., 2020). However, turning former passive consumers' homes into smart ones requires wise investment planning, the right governmental incentives, well-designed tariff structures, and making the prosumer capable of making online decisions (Avau et al., 2021). These topics have not yet been well addressed in the literature.

Several studies have investigated the impact of implementing different electricity tariff designs and pricing schemes in combination with vehicle-to-grid (V2G) charging. Richardson calculated premium tariff rates for V2G peak power in Ontario, Canada similar to existing feed-in-tariff programs for renewables that could encourage more participation in V2G (Richardson, 2013). Huang and Wu proposed a dynamic tariff-subsidy method for congestion management in distribution networks with high penetration of solar PV, heat pumps, and V2G-enabled electric vehicles (Huang and Wu, 2019). They showed the efficacy of using both positive (tariff) and negative (subsidy) regulation prices to solve congestion issues.

Other work has looked specifically at time-of-use (TOU) tariffs for V2G. Ma et al. developed an optimal dispatch strategy for a wind-PV-battery microgrid with V2G under TOU

tariffs to minimize operation costs (Ma et al., 2019). Aguilar-Dominguez et al. analyzed the potential impact of V2G operating under different TOU tariffs on a household's electricity demand and bill savings (Aguilar-Dominguez et al., 2020). They found savings of 30-85% could be achieved with V2G and rooftop PV compared to no storage. Jeon et al. compared smart/managed EV charging strategies under varying renewable energy penetration levels, including a TOU tariff approach (Jeon et al., 2020). Their results highlighted the larger benefits of smart V2G control, especially at higher renewable shares.

In terms of assessing end-user preferences, Baumgartner et al. used a survey to explore EV owners' motivations and willingness-to-pay for a V2G charging tariff design (BAUMGARTNER et al., 2022). They found more experienced EV users had a higher acceptance of lower minimum range requirements and lower expected monetary savings from V2G participation. This suggests the importance of designing business models for V2G that align with specific user requirements.

In summary, recent work has evaluated different electricity pricing schemes and tariff structures in combination with V2G to demonstrate their ability to encourage participation, reduce grid congestion, minimize costs and emissions, and deliver bill savings for prosumers. However, further analysis is still needed to optimize these tariff designs and quantify trade-offs between various stakeholders. Therefore, this study concentrates on two of these less investigated aspects:

1. **Who should invest?** DER and EMS are often discussed as tools for consumers, but there is a limited discussion on who should actually invest and control their usage. Given the decreasing cost of DER and EV, more and more prosumers are adopting them, and the share of emerging investors in DER has increased in recent years (Bergek et al., 2013). As studied in (Bergek and Mignon, 2017), different motivations, including cost minimization, grid stability maximization, and environmental concerns, encourage different types of investors to come to this market. Among the main types of investors, this study considers the distributor and the prosumer as two candidates to invest in DER and EMS and control them. Since each has different objectives, they tend to control DER and EMS in different ways, which has different consequences on the grid and returns different profitability for each party. In particular, the prosumer

seeks annual bill minimization. On the other side, as shown in (Asensio et al., 2016), by utilizing demand response and Energy Storage Systems (ESS), the distributor can avoid network and supply expansion costs in the long run. Then, with different objectives, the distributor tries to shave the peak load and manage the local grid. The different objectives of the two parties do not necessarily converge and lead to different operating strategies. Therefore, this study formulates a model for controlling DER and EMS, defines different objective functions for each party, and carries out a cost-and-benefit analysis of each investment scenario for each party under different tariff structures are subject to.

2. **What rate should be used?** Carbon tax and market-based regulations are hard to implement due to political concerns, while tariff-based strategies are easier to apply (Smith and Urpelainen, 2014). However, a wise tariff design is needed. Under an energy-based pricing structure, the higher penetration of DER could increase the distributor's cost and reduce its income (Haapaniemi et al., 2017). Moreover, imposing different tariff structures affects the profitability of investment scenarios on DER (Kök et al., 2018). Accordingly, this study investigates the effect of different rate structures (as shown in Table 1.3) on the profitability of each investment scenario. Tariffs can be an option (decision) for both agents; the distributor can make a list of rate schemes available to the prosumer, and the prosumer chooses the best one proposed.

The remainder of this chapter is structured as follows: Section 2 introduces the model's elements, data, and two performance evaluation measures for the problem and formulates two models for the distributor and the prosumer, respectively. Section 3 discusses the results, and section 4 presents the main conclusions, indicates the limitations of this study, and suggests some future research lines.

1.2 Model Description and Methodology

This section offers a comprehensive overview of the research methodology and explicitly presents certain assumptions to guide interpretation. We assume all parameters are deterministic and are specified at the outset. The study focuses on investment decisions made by

a single household without considering shared or coalition investments. Additionally, this analysis does not account for the mass adoption effect, where many prosumers might shift their peak load and create new peak loads. Moreover, our approach centers on the load pattern of one household, which may differ from aggregated load patterns. This framework outlines the data sources, parameters, and essential components to assess investments' profitability in DER, V2G, and EMS under realistic conditions. The subsections detail specific aspects of the model, including data and parameters, cost components, and the mathematical model. See the appendix for a more detailed version of the mathematical model with the numeric values of parameters.

1.2.1 Data and Parameters

Investment in DER, V2G technology and EMS is gaining popularity due to environmental and managerial incentives. However, the profitability of such investments has not been extensively examined, particularly in realistic conditions and when subsidies are no longer provided (Bertsch and Di Cosmo, 2020). This study evaluates the profitability of these investments and determines the most suitable conditions by utilizing real-world data obtained from four single detached houses in Vermont, USA, during the year 2019. Table 1.1 presents the consumption patterns of these houses, highlighting their variations and means. The following section selects two houses to calculate and present results when each agent invests in and controls DER and EMS. The models are also applied to data from the other houses to ensure the consistency of the results. Figure 1.2 illustrates the electricity demand of these houses during the first week of 2019, including consumption and solar PV generation data recorded at 15-minute intervals. Vermont is a leading state in DER utilization, attributed to policies implemented approximately 20 years ago that encouraged solar PV investments and established a robust net-metering program (Allen, 2019). Vermont's solar PV generation has consistently increased over the years (see Figure 1.3). The DER market has a wide range of available technologies with varying capacities and costs. This variety presents prosumers (consumers who also produce energy) and distributors with numerous investment options (Kosmadakis et al., 2013). For this study, five DER home investment scenarios are examined, along with their corresponding purchasing and installation costs. These scenar-

Table 1.1: Consumption characteristics for an entire year by 15-min intervals (values in kWh)

House number	mean	std	min	25%	50%	75%	max	sum
1	0.99	0.89	0.00	0.44	0.64	1.20	9.16	8,709
2	0.78	1.06	0.00	0.24	0.40	0.72	8.08	6,862
3	1.61	1.60	0.00	0.60	1.04	2.04	13.12	14,148
4	1.06	1.21	0.00	0.32	0.60	1.20	9.72	9,345

ios, described in detail in Table 1.2, are compared against a Status Quo scenario, which serves as a reference point for analysis.

The total costs of the scenarios are estimated by the company dcbel (selling bidirectional EV chargers and inverters). Consumers are assumed to purchase the EV independently of their smart-home operations, as in the market (see Figure 1.3). The only additional cost is the V2G equipment. This assumption explains why the cost of the EV option is relatively lower. We note that the EV battery capacity is 60 kWh, but only 60% of this capacity is assumed to be available for the V2G system. The growing trend in EV purchases further justifies this assumption. Consumers are increasingly inclined to buy EVs due to various incentives such as decreasing battery costs, environmental benefits, and efforts to address climate change. However, it is crucial to understand that the primary motivation for purchasing an EV is not as a tool for home electricity management. While the potential to use an EV to reduce electricity bills might add an extra layer of motivation, it is far from being the primary reason for their purchase. Therefore, our assumption rests on the premise that the EV is purchased independently of its potential utility in home energy management, and any additional cost savings or benefits derived from its integration into home energy systems are incidental to this primary purpose. During working days from 7:45 am to 5:15 pm, we assume that EVs are not accessible, as they are at work and consume 10 kWh of stored energy from their batteries.

Additionally, the distributor procures electricity from the market based on the Real-Time Locational Marginal Price (RTLMP), which changes every five minutes. The distributor offers different electricity rates to consumers for their electricity purchases. Prosumers choose a specific rate based on their demand patterns and available DER. Table 1.3 outlines the five rates available to prosumers, each with a fixed daily fee and a variable charge per kWh.

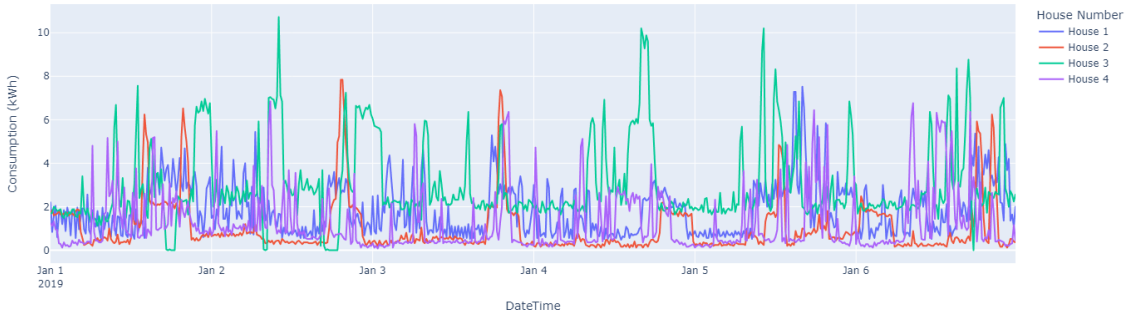

Figure 1.2: Consumption pattern in different houses in the first week of 2019

Table 1.2: Investment scenarios

Scenario	Capacity of the standalone battery (kWh)	Charge/discharge capacity (kWh) (per 15 min.)	Available capacity of the EV's battery (kWh)	Solar PV capacity (kWh)	Total cost (purchase and installation, USD)
0 – Status Quo	0	0	0	0	0
1 – Battery	40.5	3.75	0	0	27,300
2 – Battery + PV	27	2.5	0	10	35,125
3 – EV	0	1.9	36	0	4,000
4 – EV + PV	0	1.9	36	10	17,145
5 – Battery + EV + PV	10	1.9	36	10	23,724

The simplest rate, called the uniform rate, is volumetric, with a fixed daily fee of $0.492 and a per kWh charge of $0.169.

To incentivize prosumers to shift their consumption from peak to off-peak hours, the distributor offers a 'Time-of-Use' (TOU) rate, with higher prices during peak hours and lower prices during off-peak hours. Also, under the 'TOU EV & Uniform' rate, all consumption is charged at the 'Uniform' rate except for the EV, which follows the 'TOU' rate. These three tariffs presented in Table 1.3 are the main tariffs offered in Vermont and are typical options considered by utilities. The authors introduce the subsequent two tariffs to explore the implications of consumers engaging in the wholesale market. Under the 'RTLMP' rate, the consumer buys and sells back electricity at the wholesale market price, and under the 'TOU & RTLMP' rate, prosumers purchase electricity at the 'TOU' rate and sell excess electricity back to the grid at the 'RTLMP' rate. The electricity rates for a sample week in March are depicted in Figure 1.4.

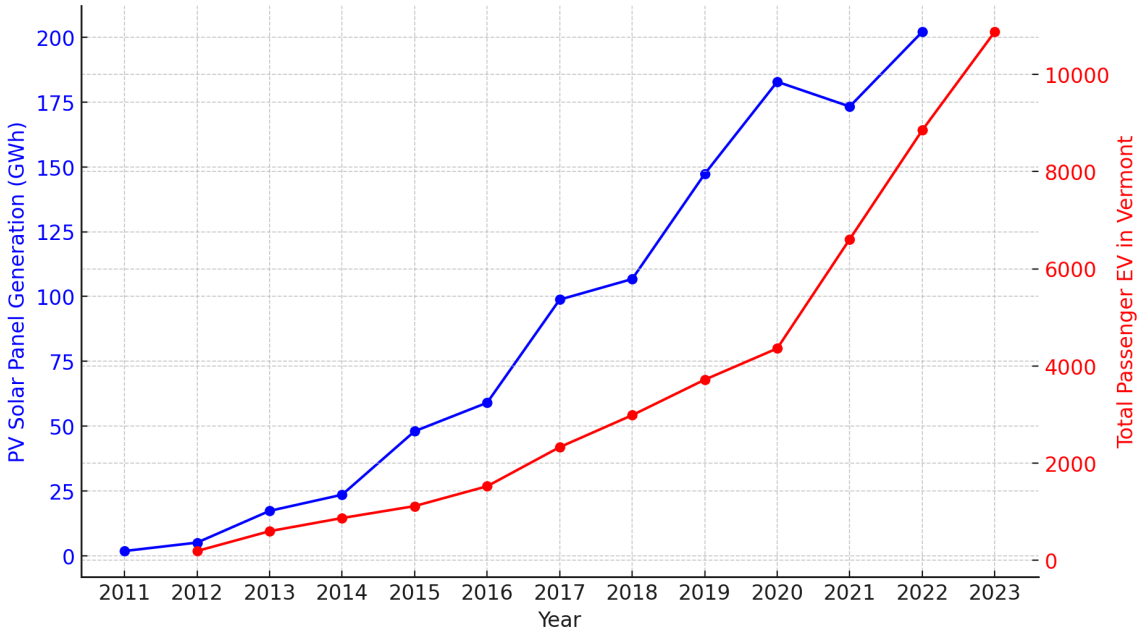

Figure 1.3: Growth of Passenger EVs and solar panel generation in Vermont (Vermont Department of Environmental Conservation, 2023; U.S. Energy Information Administration (EIA), 2021b)

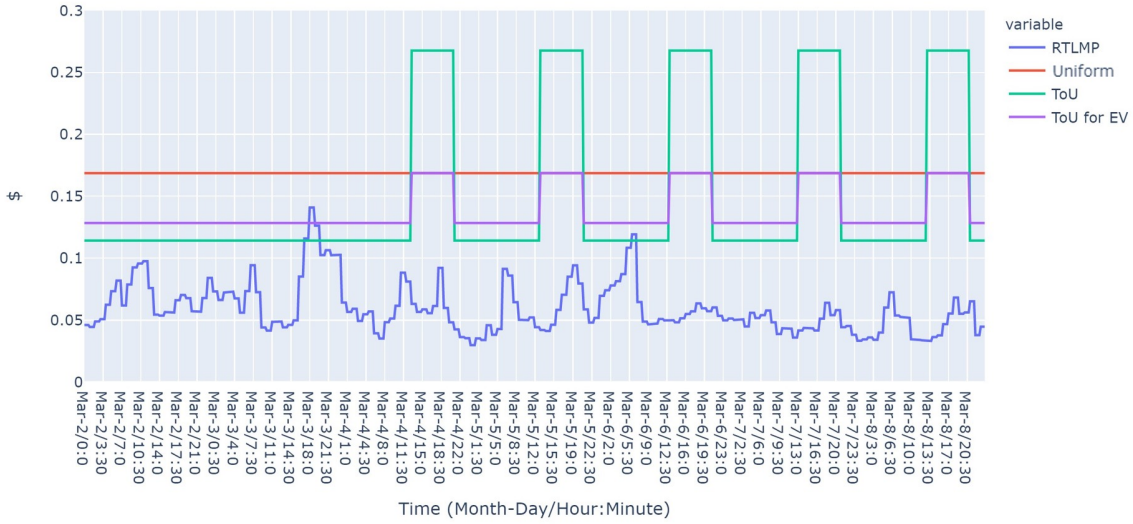

Figure 1.4: Electricity rates over a week in March 2019

Table 1.3: Five rates proposed to prosumer

Rate Name	Fixed charge ($/day)		buys at ($/kWh)		sells at (if allowed) ($/kWh)
Uniform	0.492		0.16859		0.16859
TOU	0.651	Peak Hours (1pm-9pm):	0.26771	Peak Hours:	0.26771
		Off-Peak Hours:	0.11411	Off-Peak Hours:	0.11411
TOU EV & Uniform	0.651	General Usage (any time):	0.16859		
		Peak Hours EV Charge:	0.16859		0.16859
		Off-Peak Hours EV Charge:	0.12831		
RTLMP	0		Market Price		Market Price
TOU & RTLMP	0.651	Peak Hours (1pm-9pm):	0.26771		Market Price
		Off-Peak Hours:	0.11411		

1.2.2 Cost Components

As previously discussed, each party involved in the context has distinct objectives, as their respective objective functions encompass various cost elements. The prosumer's tariffs and costs are explicitly defined. In accordance with the applicable tariff structure, which may incorporate volumetric and time-dependent factors, the prosumer aims to minimize the overall expenditure associated with energy consumption from the power grid. Conversely, the distributor's costs encompass two significant categories[1]:

- **Wholesale Load Costs**: These costs represent the substantial portion of expenses related to the procurement of wholesale electricity and consist of three primary cost classes:

  1. **Energy**: This pertains to the provision of demand and is determined by the volume of consumed energy.
  2. **Capacity**: This refers to the payment made to ensure the secure electricity supply during peak hours and is determined based on the monthly peak capacity.
  3. **Additional costs**: This category incorporates supplementary charges such as ancillary market charges, administrative fees, and other expenses that typically do not rely on volume and peak consumption. However, these costs are not considered in the present model.

- **Regional Network Service (RNS)**: These costs cover the transmission expenses incurred by the distributor accessing the New England grid and delivering electricity

[1] https://www.iso-ne.com/markets-operations/market-performance/load-costs/

to the end-user. These costs are calculated based on the monthly peak capacity, as defined by the New England Independent System Operator.

The distributor's supply cost is formulated using the two primary cost categories. It includes the first category's energy component and monthly peak capacity component. A peak capacity component from the second category is added to the initial peak capacity cost. For more detailed information regarding the distributor's costs in New England, please refer to (Independent System Operator – New England" (ISO-NE), 2021b,a)

1.2.3 Modeling

In contrast to traditional homes where electricity consumption is passive, electricity flow in a smart home requires continuous management at each time step. By following price signals, the electricity flow within the home is optimized to ensure demand fulfillment while minimizing the overall cost. However, as the number of decisions to be made increases, the complexity of the problem escalates. Halman et al. demonstrated that even a simplified version of this problem, where a home is equipped with an ESS and a PV and engages in electricity trading with the grid, is NP-Hard (Halman et al., 2018). Therefore, the complexity of this problem necessitates the development of a mathematical model to control the electricity flow and the inevitable utilization of an EMS.

This study presents a mathematical scenario analysis model that simulates the operation of DER and EMS within a smart home. The model's objective must be adjusted based on the party responsible for investing in and controlling DER and EMS. The prosumer aims to optimize the system to minimize the total electricity consumption costs (or, equivalently, maximize total revenue) according to their applicable tariff. On the other hand, the distributor strives to minimize the overall supply costs while also considering network costs, which are directly proportional to the monthly peak load. Peak loads necessitate an increase in service capacity, and the network cost is determined by multiplying the coefficients of RNS and forward Capacity (Cap) costs by the monthly peak load. Figure 1.5 illustrates the DER components in a smart home, consisting of a solar PV system, a stationary ESS, and an EV alongside the grid and the load. The load is deterministic, and demand response measures to shift demand are not employed by the prosumer in this model. The distributor is allowed

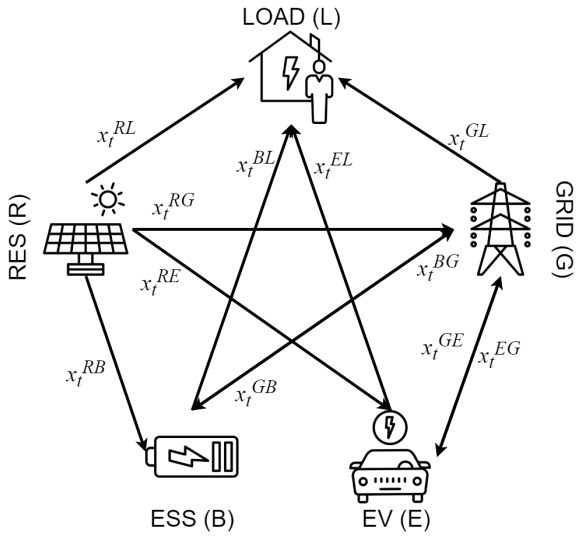

Figure 1.5: Electricity flows in the smart home

to inject electricity generated by residential DER into the grid. An identifier parameter (Itrade) is introduced to account for the impact of decentralized trade allowances on the profitability of investment scenarios. The ESS and EV batteries possess identical charging/discharging capacities and loss rates. Additionally, a minimum charging level is defined for the EV to ensure optimal functionality throughout its lifespan.

All parameters in the model are considered deterministic and provided at the outset. Figure 1.5 depicts the continuous variables representing the flows between the nodes in the system graph. Furthermore, a dummy variable (x_k^{peak}) represents the monthly peak load used in the computation of capacity and network costs, while two state variables track the ESS and EV charging levels. The final model is a linear model comprising 15 constraints that govern the electricity flow within the home, satisfy minimum requirements, and compute the state variables for the subsequent time step. An additional constraint is included in the prosumer's model to adhere to the trade allowance condition. The model, under various parameter configurations and with two different objective functions, is solved using Gurobi in Python. It is important to note that the complete model is described in the appendix, and the accompanying codes are provided in the attachment.

1.2.4 Net Benefits

The internal rate of return (IRR) and net present value (NPV) are financial metrics commonly used in investment analysis. IRR measures the anticipated rate of return that an investment is expected to generate throughout its duration, whereas NPV assesses the difference between the present value of cash inflows and outflows using a discount rate. NPV is preferable due to its consideration of a specific discount rate, which facilitates more effective comparisons between various projects and provides a more precise gauge of an investment's value and profitability by reflecting the genuine opportunity cost of capital, unlike IRR which presupposes reinvestment at the same rate. Furthermore, NPV fosters superior comparability between different projects by considering their absolute value, thereby enabling more well-informed decision-making. When comparing five investment scenarios with identical durations, NPV is particularly suitable. In scenarios where durations are equal, the timing becomes irrelevant, and NPV becomes focused on the magnitude of cash flows. This approach accurately captures the opportunity cost of capital and allows for direct comparison. By evaluating NPV values, one can effectively identify the most financially advantageous option, thus establishing NPV as the preferred metric for informed decision-making. The formulas to calculate these indices are as follows:

$$NPV = \sum_{n=1}^{\mathrm{N}} \{\frac{\text{Yearly Saving}}{(1+i)^{\mathrm{N}}}\} - (\text{Product price} + \text{Installation and Labor cost})$$

$$0 = \sum_{n=1}^{\mathrm{N}} \{\frac{\text{Yearly Saving}}{(1+IRR)^{\mathrm{N}}}\} - (\text{Product price} + \text{Installation and Labor cost})$$

$$\text{Yearly Saving} = \text{Total Cost - Base Cost} + \text{EV Usage Cost}$$

$$i \doteq \text{Discount rate} = 0.0619$$

$$N \doteq \text{Life time of the product(s)} = 15 \text{ years}$$

1.3 Results

This section presents the outcomes of the distributor and the prosumer investing in and controlling DER and EMS. Additionally, a sensitivity analysis is conducted in the third

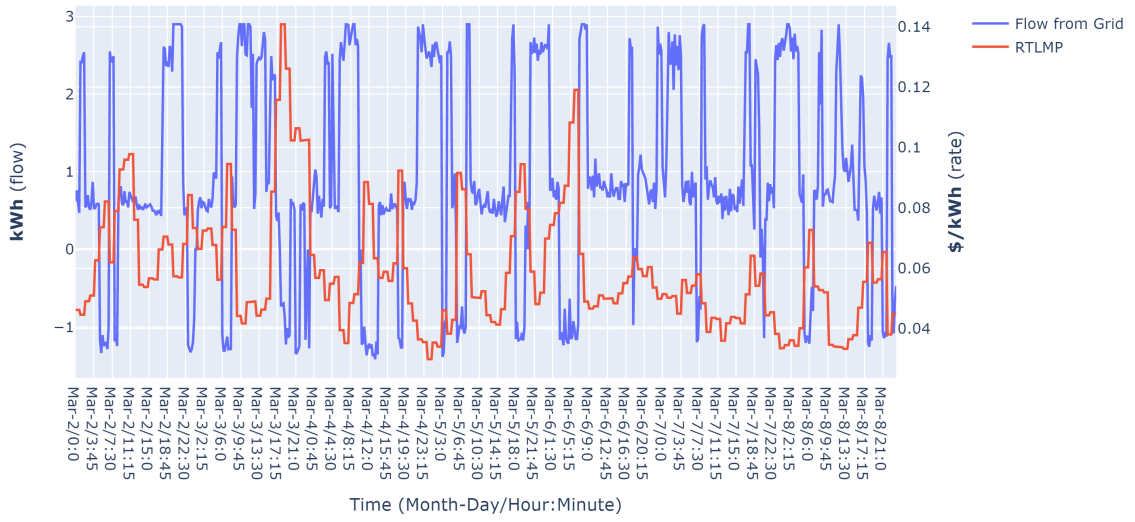

Figure 1.6: Electricity flow from the grid under s3 (EV) and RTLMP where distributor controls

subsection, focusing on variations in the load pattern, battery capacity, and solar panel system size.

1.3.1 Distributor in Control

This section examines the profitability and consequences of the proposed investment scenarios in which the distributor controls the system. The distributor leverages the available DER within the home and engages in decentralized trading activities with the grid to minimize costs related to load covering and network utilization. As an example, Figure 1.6 illustrates the electricity obtained from the grid (positive) and injected into the grid (negative) under-investment scenario 3 (EV) with Real-Time Locational Marginal Price (RTLMP) rates in a selected week of March. The figure demonstrates how the distributor strategically acquires electricity from the grid and engages in decentralized trading based on the prevailing price signals. When the price is low and the electric vehicle (EV) is available, the distributor charges the EV battery. Conversely, when the price increases, the stored energy is utilized to cover the load or can be sold back to the grid to generate a profit.

The Sankey diagrams presented in Figure 1.7 visually represent the overall energy flows between various nodes depicted in Figure 1.5. In scenario 0 (Status Quo), the grid fulfills

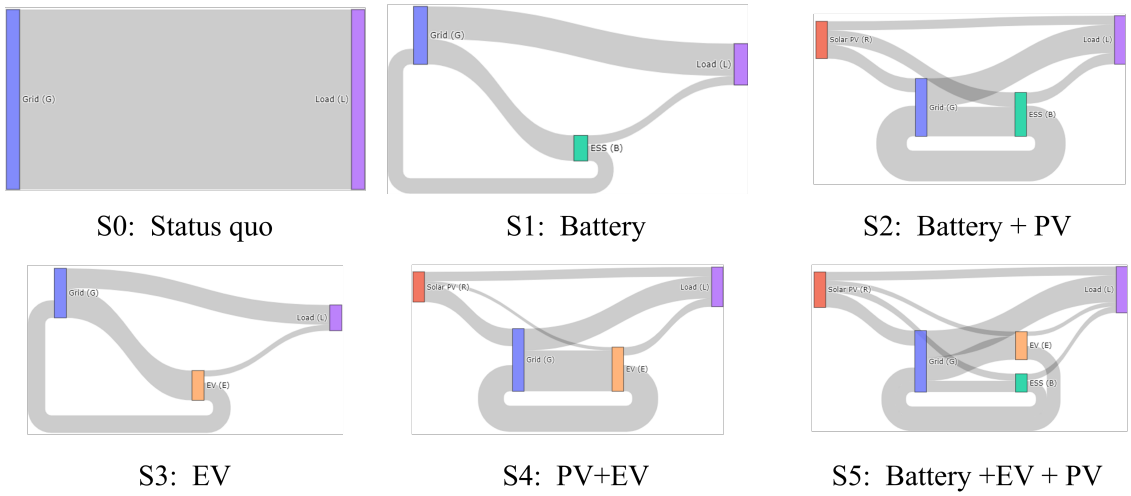

Figure 1.7: Yearly electricity flows under different investment scenarios where distributor controls

the entire demand. In alternative scenarios where a battery or an electric vehicle (EV) is introduced to the home, their capacities are primarily utilized to leverage electricity price differentials, while the remaining capacity is employed to meet the demand when prices are high. Furthermore, in scenario 4, the unavailability of the EV prevents the distributor from storing the electricity generated by the photovoltaic (PV) system. Conversely, scenario 2 exhibits a substantial energy flow from the PV system to the battery.

The distributor's objective function is used to solve the model and evaluate the profitability of various investment scenarios. The results in Table 1.4 demonstrate the potential annual savings for the distributor investing in and controlling the electricity load. These savings range from \$813 in scenario 3 (EV) to \$2,163 in scenario 5 (EV+PV+Battery). However, when taking into account investment costs, discount rate, and product lifespan, scenario 3 (EV) emerges as the only profitable option, with a net present value (NPV) of \$3,803. Consequently, connecting an electric vehicle (EV) to the home is the most favorable investment scenario among the proposed options.

Table 1.4 also provides an overview of the key annual outcomes derived from the distributor's model. The "Net trade" column indicates the net funds exchanged with the grid for optimal electricity supply, denoted by positive (+) or negative (-) values. The "System cost" column reflects the total costs associated with RNS and Cap costs resulting from the chosen

Table 1.4: Key results of the distributor's model for the representative House #3

Scenario	Net trade (A) ($)	System cost (B) ($)	Total cost (A+B) ($)	Saving compared to S0 ($)	NPV ($)	IRR (%)	Maximum peak load (kWh)	Total from grid (kWh)	Total to grid (kWh)
S0: Status Quo	505	2,050	2,555	0	0	NaN	13.12	14,148	0
S1: Battery	422	486	908	1,647	-11,498	-1.22	3.76	19,745	4,769
S2: Battery+PV	61	647	708	1,847	-17,403	-2.83	5.08	17,637	13,860
S3: EV	-79	1,821	1,742	813	3,801	18.80	11.64	26,057	7,352
S4: EV+PV	-399	1,605	1,206	1,350	-4,196	2.15	11.60	22,596	15,459
S5: EV+PV+Battery	-380	772	392	2,163	-2,970	4.20	6.32	20,069	12,954

Figure 1.8: NPV and IRR of scenarios where distributor controls

investment scenario. The "Total cost" represents the sum of net trade and system costs the distributor aims to minimize. The "Saving" column indicates the yearly cost savings, irrespective of initial costs. The NPV and IRR columns incorporate these costs and measure the profitability of each investment scenario using different evaluation methods.

Additionally, Table 1.4 includes information on the maximum monthly peak load ("Maximum peak load") observed during the year, as well as the total energy flow taken from and sent to the grid ("Total flow taken from the grid" and "Total flow sent to grid" columns). Furthermore, Figure 1.8 illustrates that scenario 3 (EV) remains the most viable and profitable option across four homes with various load patterns.

1.3.2 Prosumer in Control

The prosumer is responsible for paying only the energy costs, as computed from the selected tariff. Therefore, this section aims to evaluate the profitability of various investment scenarios for the prosumer while also analyzing the costs imposed on the grid. Additionally, the prosumer will examine the impacts of the distributor's imposed rates and the restriction on injecting electricity back into the grid. An illustration depicting the prosumer's control is presented in Figure 1.9. This figure showcases the electricity the selected house takes from the grid at the 'TOU' rate, considering both the Status Quo (S0) and Battery+PV+EV (S5)

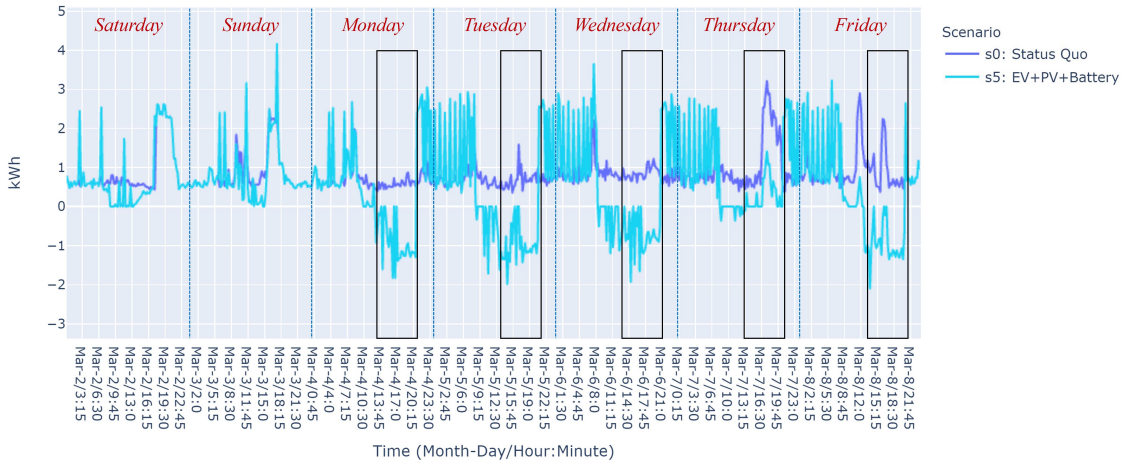

Figure 1.9: Flow from the grid at 'TOU' and under scenarios 0 and 5 where prosumer controls

scenarios during a sample week in March. The black boxes represent peak hours throughout the week when electricity prices are high during business days. The prosumer adjusts its energy consumption based on price signals to minimize its overall energy costs.

The key findings resulting from the analysis of the prosumer's model applied on House 4 and under various conditions are presented in Table 1.5. The table includes several columns: 'Total cost' and 'System cost' represent the annual expenses incurred by the prosumer and the costs imposed on the grid, respectively. The 'Saving' column quantifies the yearly savings achievable by the prosumer through each investment scenario compared to the Status Quo. For each scenario, the 'NPV' and 'IRR' columns provide the Net Present Value and Internal Rate of Return measurements, respectively. Moreover, the 'Maximum flow taken from the grid' indicates the highest monthly peak load observed throughout the year. The 'Total flow taken from the grid' and 'Total flow sent to grid' columns display the overall energy obtained from and sold back to the grid by the prosumer over the year. The last two columns, 'NPV' and 'System Cost,' allow us to explore the potential outcomes if the prosumer could not sell electricity to the grid. These columns demonstrate the NPV and system cost associated with the respective scenario where such trade is not permitted.

The results presented in Table 1.5 and Figure 1.10 illustrate that in situations where the prosumer is unable to sell surplus electricity to the grid, except for scenario 3 (EV), where the investment yields a net present value of $888 at the 'RTLMP' rate, all other investment

Table 1.5: Key results of the prosumer's model

Selling back to grid is →		Allowed								NOT Allowed	
Rate	Scenario	System cost ($)	Total cost ($)	Saving compared to S0 ($)	NPV ($)	IRR (%)	Highest monthly peak (kWh)	Total from grid (kWh)	Total to grid (kWh)	NPV ($)	System cost ($)
Uniform	S0: Status Quo	1,631	1,756	0	0	NAN	10	9,345	0	0	1,631
	S1: Battery	1,631	1,756	0	-27,300	NAN	10	9,345	0	-27,300	1,631
	S2: Battery+PV	1,555	-189	1,945	-16,468	-2	10	6,735	8,935	-25,144	1,130
	S3: EV	2,662	1,813	-57	-4,549	NAN	17	12,843	0	-4,552	2,672
	S4: EV+PV	2,457	-132	1,888	965	7	17	9,063	7,765	-11,528	2,445
	S5: EV+PV+Battery	2,543	-132	1,888	-5,614	2	17	9,063	7,765	-14,573	2,464
TOU	S0: Status Quo	1,631	1,819	0	0	NAN	10	9,345	0	0	1,631
	S1: Battery	4,102	309	1,510	-12,818	-2	23	17,753	7,327	-23,420	3,972
	S2: Battery+PV	2,727	-1,189	3,007	-6,277	3	17	8,440	9,920	-23,742	1,658
	S3: EV	2,845	706	1,112	6,672	27	17	18,697	5,133	-754	2,720
	S4: EV+PV	2,770	-1,294	2,957	11,222	15	17	15,634	13,615	-8,445	2,529
	S5: EV+PV+Battery	2,726	-1,667	3,486	9,714	12	16	16,130	13,845	-11,594	2,478
TOU EV & Uniform	S0: Status Quo	1,631	1,756	0	0	NAN	10	9,345	0	0	1,631
	S1: Battery	1,631	1,756	0	-27,300	NAN	10	9,345	0	-27,300	1,631
	S2: Battery+PV	1,555	-191	1,946	-16,455	-2	10	6,735	8,935	-25,144	1,130
	S3: EV	1,620	670	1,086	6,413	26	10	39,438	23,418	-1,073	1,592
	S4: EV+PV	1,571	-1,276	3,032	11,938	16	10	37,504	33,028	-8,798	1,520
	S5: EV+PV+Battery	1,571	-1,276	3,032	5,359	10	10	37,504	33,028	-12,329	1,490
RTLMP	S0: Status Quo	1,631	320	0	0	NAN	10	9,345	0	0	1,631
	S1: Battery	4,009	28	293	-24,492	-17	22	31,756	19,975	-26,446	3,807
	S2: Battery+PV	2,957	-200	520	-30,137	-15	17	20,104	20,678	-33,030	2,079
	S3: EV	2,631	-266	586	1,623	12	16	21,672	7,754	888	2,509
	S4: EV+PV	2,604	-591	911	-8,406	-3	16	19,265	16,892	-11,269	2,373
	S5: EV+PV+Battery	2,584	-640	960	-14,516	-6	16	22,938	20,066	-17,283	2,426
TOU&RTLMP	S0: Status Quo	1,631	1,819	0	0	NAN	10	9,345	0	-	-
	S1: Battery	3,988	1,412	406	-23,404	-15	23	9,784	135	-	-
	S2: Battery+PV	1,619	478	1,341	-22,266	-6	17	3,348	4,938	-	-
	S3: EV	2,775	1,480	339	-748	3	17	13,079	63	-	-
	S4: EV+PV	2,656	710	1,109	-6,507	-0	17	9,270	7,823	-	-
	S5: EV+PV+Battery	2,511	410	1,409	-10,210	-1	17	7,042	5,398	-	-

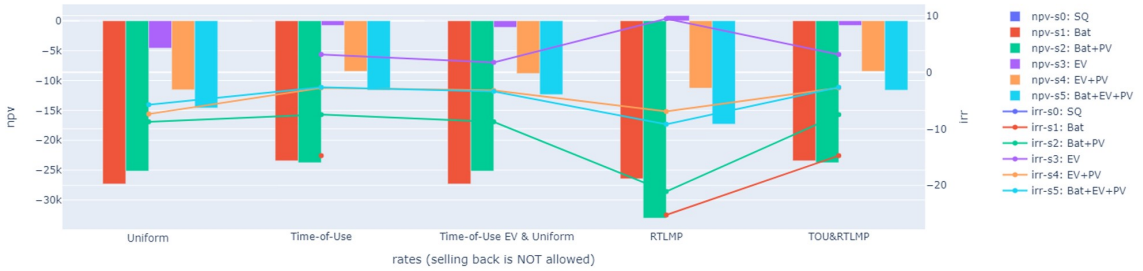

Figure 1.10: NPV & IRR of scenarios where prosumer controls and selling back is NOT allowed

scenarios across different rate structures are unprofitable. In general, if the prosumer cannot sell electricity to the grid, they cannot take advantage of price differentials, resulting in an investment that does not generate returns.

Figure 1.11 displays the NPV and IRR of the five investment scenarios, considering various rates when the prosumer has control over DER and EMS, with the ability to sell

Figure 1.11: NPV & IRR of scenarios where prosumer controls and selling back is allowed

electricity back to the grid. When the prosumer buys energy at 'TOU' rates and sells at 'RTLMP' rates, all investments yield negative NPVs. This finding is attributed to the prosumer purchasing energy at a higher rate and selling it to the wholesale market at a lower rate. However, if the prosumer buys and sells energy at 'RTLMP' rates, scenario 3 (EV) generates a positive NPV of +$1,623. Consequently, a standalone prosumer who has connected their EV to the grid can generate benefits under the wholesale market rate. Regarding the remaining rates, the NPVs of the first and second scenarios are negative. Hence, connecting an EV through a V2G device becomes crucial to making investments beneficial. Among the remaining rates, scenario 4 (EV+PV) yields the highest NPV, reaching +$11,938. The suitable rates for the prosumer are 'TOU EV & Uniform','TOU', and 'Uniform,' respectively. Another notable finding is that a significant price gap during peak and off-peak hours substantially impacts the profitability of the investment scenarios. This gap allows the prosumer to store electricity when the price is low and sell or use it when the price is high.

Figure 1.12 illustrates the costs incurred by the distributor when the prosumer controls DER and EMS and is allowed to sell electricity back to the grid. The results indicate that the 'TOU EV & Uniform' rate corresponds to the lowest annual cost for the distributor. This finding is further supported by Figure 1.13, which can explain the superiority of the 'TOU EV & Uniform' rate for the distributor. Since the objective function of the prosumer does not include network costs, implementing smart technologies and making homes smarter can lead to increased peak loads. As peak loads constitute a significant portion of network

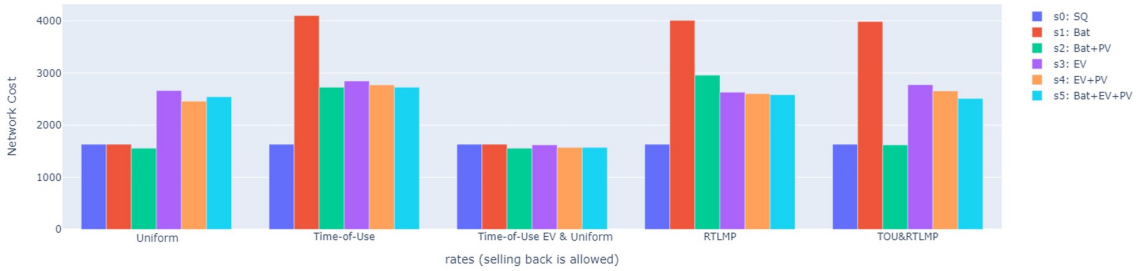

Figure 1.12: System cost where prosumer controls and selling back is allowed

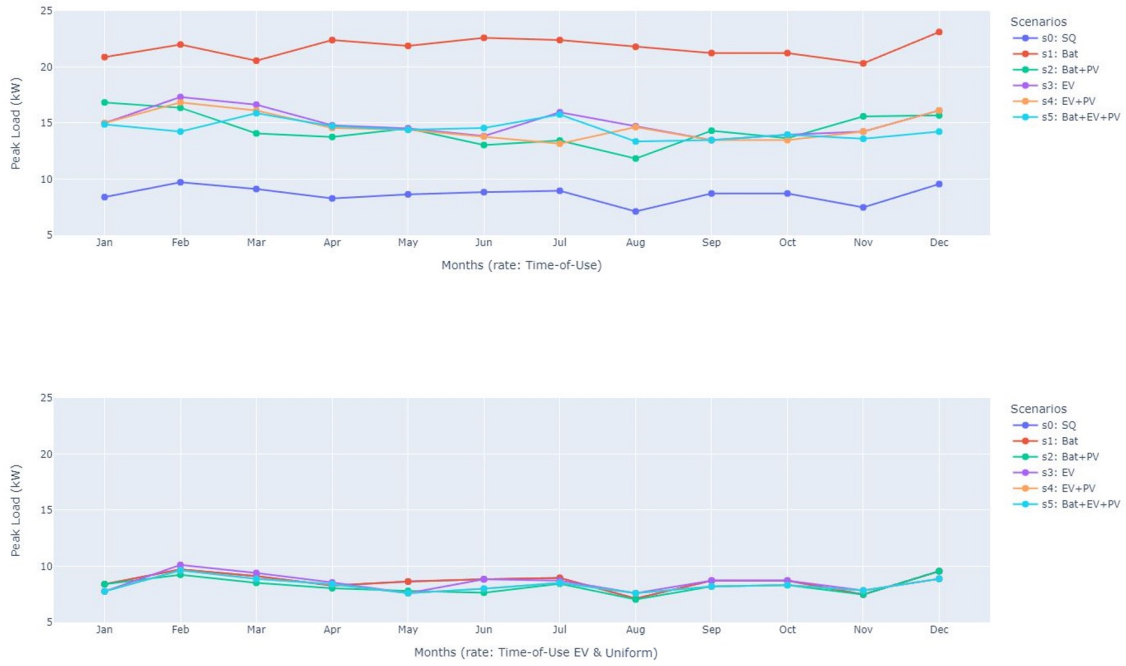

Figure 1.13: Peak loads at 'TOU' and 'TOU EV & Uniform' where prosumer controls

costs, the 'TOU EV & Uniform' rate has minimal impact on monthly peak loads, making it the most favorable option for the distributor to recommend to prosumers. On the other hand, the 'TOU' rate increases peak loads across all tariff structures.

Table 1.5 demonstrates that when the prosumer cannot engage in decentralized trade, it results in a decrease in the NPV or, at best, no change in NPV across all investment scenarios and rate structures. Moreover, except for scenario 3 (EV) at the 'Uniform' rate, preventing the prosumer from selling back electricity reduces system costs for all other rates and

scenarios. However, if the distributor modifies the trade allowance condition, the prosumer can adjust its chosen rate and investment scenario accordingly. This strategic interaction between the prosumer and the distributor is analyzed to determine the equilibrium in this one-shot game, where the decisions of one agent affect the outcome of the other. Table 1.6 identifies the optimal decisions for the prosumer, who controls DER and EMS, and the distributor. The game unfolds with the distributor making the initial move as the leader, deciding whether to permit the prosumer to sell back electricity to the grid and which tariff is offered. Subsequently, the prosumer, acting as the follower and considering the leader's decision, selects an investment scenario.

Hence, Table 1.6 provides insights into the best response of the prosumer for each combination of rate scheme and trade allowance. It also presents the corresponding costs for the prosumer and the system under the selected investment scenario. The table indicates that when selling back electricity to the grid is prohibited, prosumers would not change from the status quo unless they can select the 'RTLMP' rate and invest in connecting their EV to the grid. This combination yields an NPV of $888 for the prosumer while increasing the annual system cost by $878. It is important to note that when selling back is not allowed, the 'TOU & RTLMP' rate is essentially the same as being a passive consumer under the 'TOU' rate. A significant takeaway from this scenario is that restricting the prosumer from selling back to the distributor would push them to expose themselves to the wholesale market price if they can. Conversely, when the prosumer is allowed to inject electricity into the grid, the optimal choice for the prosumer is to select the 'TOU EV & Uniform' rate and invest in the combination of electric vehicle (EV) and photovoltaic (PV) systems (S4). This particular choice results in an NPV of $11,938 and reduces the annual system cost by $60. Comparing the two distributors' decisions reveals that allowing the prosumer to sell back electricity benefits both parties. The prosumer's NPV increases by $11,050, and the annual system cost decreases by $938. Consequently, the optimal decision for the distributor is to permit the prosumer to sell back electricity. Consequently, the prosumer chooses the 'TOU EV & Uniform' rate and invests in EV and PV systems.

Table 1.6: Prosumer's best response to distributors decisions

selling back ↓	rate →	Uniform	TOU	TOU EV & Uniform	RTLMP	TOU & RTLMP
allowed	prosumer's choice	S4	S4	**S4**	S3	S0
	prosumer's NPV ($)	965	11,222	**11,938**	1,623	0
	system annual cost ($)	2,457	2,770	**1,571**	2,631	1,631
not allowed	prosumer's choice	S0	S0	S0	S3	-
	prosumer's NPV ($)	0	0	0	888	-
	system annual cost ($)	1,631	1,631	1,631	2,509	-

1.3.3 Sensitivity Analysis

A series of sensitivity analyses are conducted on the load patterns and the size and capacity of the PV system and stationary battery in this subsection to ensure the consistency of the results obtained. It can be concluded that if the results are robust, the findings can be generalized to gain valuable insights. The distributor and prosumer models are run using the consumption records of all four households (Table 1.7 in the Appendix presents the sensitivity analysis results). Despite fluctuations in NPVs across different scenarios and houses, the relative profitability order remains consistent. In the case of the distributor model, scenario S3 (EV) emerges again as the only profitable option for all households, and the findings for the prosumer model remain unchanged.

Subsequently, in scenarios 1 (Bat), 2 (Bat+PV), and 5 (Bat+PV+EV) that include a stationary battery, the new NPV, and IRR are computed for the cases where battery sizes range from 0.33 to 2 times the original battery capacity. The resulting NPVs and IRRs are graphically presented in Figure 1.14. The findings reveal two key observations. First, the NPV of the scenarios decreases as the battery capacity increases. Second, changing battery capacity does not change the optimal investment option. The optimal investment scenario would change if the new NPV obtained using new battery capacity and, under a specific rate, is higher than all other investment scenarios and rates. The results show that for scenario 1 (Bat), all combinations return negative NPVs. In scenario 2 (Bat+PV), the 1/3 capacity demonstrates the highest NPVs, yet they remain lower than the maximum NPV achieved by other scenarios for each rate. This trend persists across all rates in scenario 5 (Bat+PV+EV), except for the 'Uniform' rate. Table 1.5 shows that scenario 4 (EV+PV)

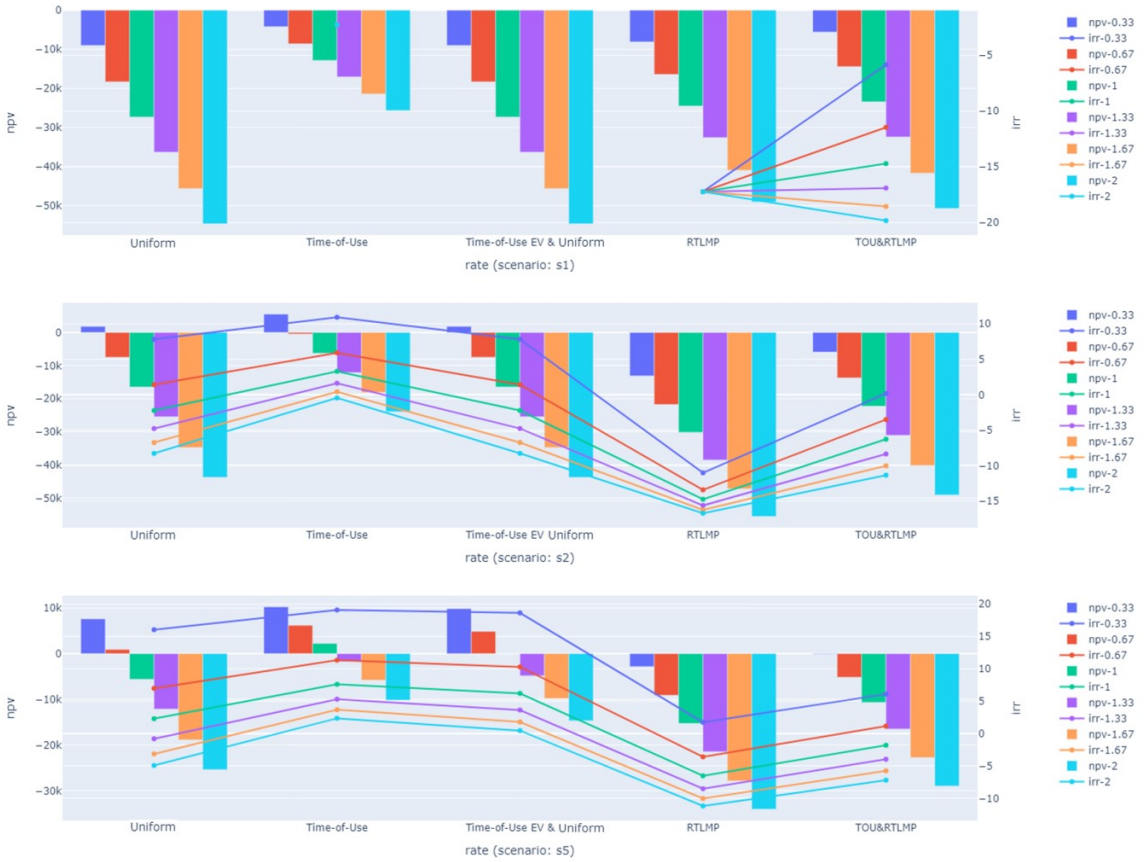

Figure 1.14: NPV & IRR for scenarios 1, 2, and 5 under different battery capacities

yields a positive NPV of +$965, while the 1/3 battery capacity in scenario 5 (Bat+PV+EV) under the 'Uniform' rate attains an NPV of +$7,607. Nevertheless, this value still falls short of the highest NPV obtained under the 'TOU EV & Uniform' rate in scenario 4 (PV+EV), which amounts to +$11,938.

The investment scenarios were analyzed in the preceding subsection using generation data from a 10 kW photovoltaic (PV) system. However, the significant capacity of this PV system led to higher investment costs for the scenarios, with an increase of +$13,145. The distributor and prosumer models are solved to assess the impact of PV size on the profitability of the scenarios and validate the previous subsection's findings, considering PV sizes of 2 kW, 4 kW, 6 kW, and 8 kW in addition to the default 10 kW. Figure 1.15 illustrates

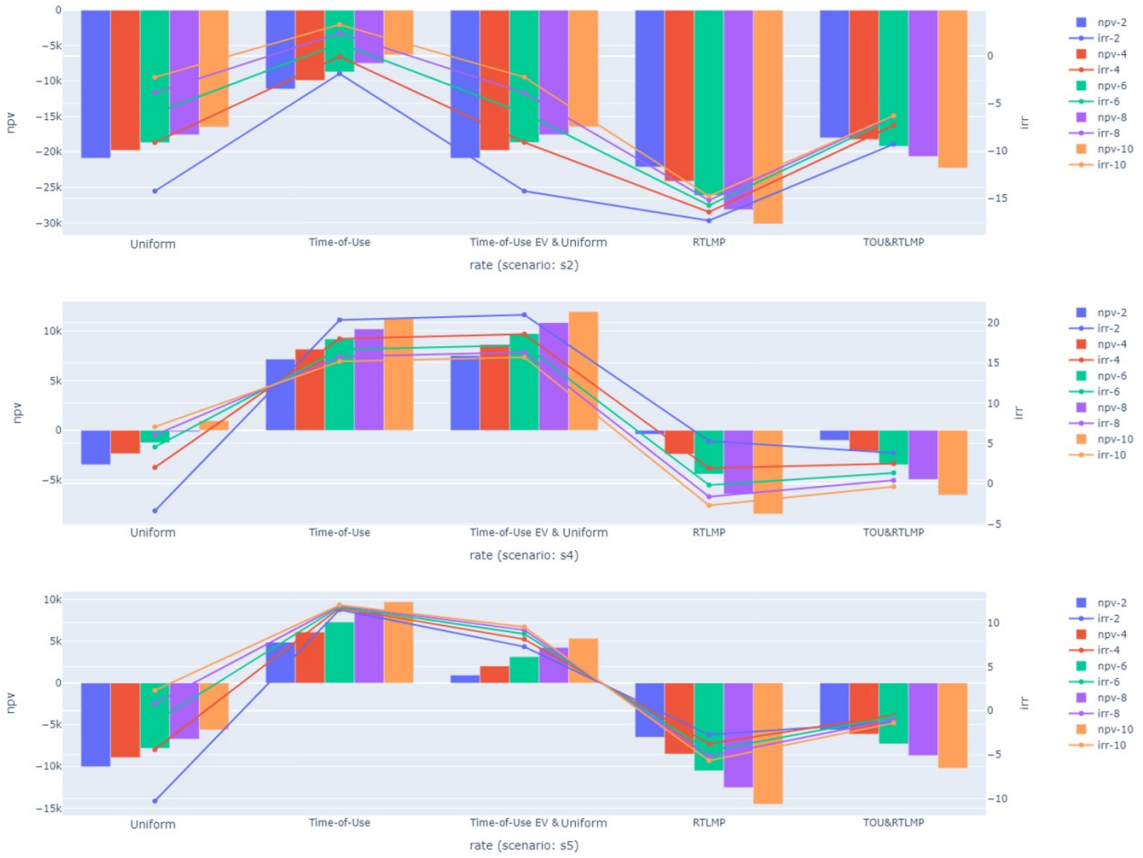

Figure 1.15: NPV & IRR for scenarios 2, 4, and 5 under PV capacities

that as the PV size decreases, the NPVs of the scenarios also decline. Consequently, the recommended PV size utilized in the previous subsection remains the most favorable among the different sizes depicted in Figure 1.15, and the results are still valid for the rates that return positive NPV for the prosumer's investment. However, it should be noted that this trend holds for the rates 'Uniform,' 'TOU,' and 'TOU EV & Uniform.' In contrast, when considering 'RTLMP' and 'TOU & RTLMP,' larger PV installations lead to lower NPVs, particularly when selling rates are based on the market rates. This finding suggests that larger PV installations are advantageous only if prosumers can sell their excess energy at their buying rate.

1.4 Conclusion & Outlook

This analysis considers the profitability and consequences of investment scenarios wherein a distributor or prosumer controls a residential energy system. The potential benefits and tradeoffs associated with various investment options were explored through modeling and analysis, considering factors such as DER, electricity trading, and load management strategies. The objective was to determine the most favorable investment scenario and evaluate its benefits in the residential electricity market context.

Who should invest? The results demonstrate that the investment can benefit both agents, and under suitable conditions, the prosumer's investment can also benefit the distributor and shave the peak loads.

What combination of DER and tariff should be used? The analysis provides valuable insights into the optimal investment choices for distributors and prosumers. For distributors, significant cost savings and profitability can be achieved by connecting an EV to the residential energy system and engaging in electricity trading with the grid based on price signals. Among the proposed investment scenarios, scenario 3 (EV) emerged as the most profitable option, with a positive NPV of $3,803. By strategically leveraging the available DER and participating in decentralized trading activities, distributors can minimize costs related to load covering and network utilization, leading to improved profitability.

For prosumers, the profitability of investment scenarios depends on the ability to sell surplus electricity back to the grid. The analysis revealed that, except for scenario 3 (EV) under some rate structures, positive NPVs could not be generated by prosumers when the sale of electricity back to the grid is forbidden. Therefore, allowing prosumers to engage in electricity decentralized trading and take advantage of price differentials is crucial for making investment scenarios profitable. Additionally, connecting an EV to the residential energy system was identified as a crucial element in ensuring the profitability of investment scenarios for prosumers. Moreover, bundling EV and PV (scenario 4) and choosing the 'TOU & EV' rate was identified as the best option for the prosumers, with a positive NPV of $11,938.

A sensitivity analysis was also conducted to examine the robustness of the findings. The results demonstrated that the relative profitability order remained consistent despite varia-

tions in load patterns and the size/capacity of the PV system and stationary battery. The conclusions drawn from the analysis can thus be generalized and provide valuable insights for stakeholders in the residential energy market.

In conclusion, this study highlights the potential cost savings and profitability that can be achieved through investments in residential energy systems, with strategic control over DER, electricity trading, and load management. Connecting EVs to the system and engaging in decentralized trading activities benefited distributors, while prosumers benefited from selling surplus electricity back to the grid and capitalizing on price differentials. However, policymakers and stakeholders should consider the specific rate structures and regulatory conditions to ensure the viability and profitability of investment scenarios. Based on the findings, it is recommended that policymakers facilitate the decentralized electricity trading mechanisms, promote the adoption of EVs in residential energy systems, and create supportive regulatory frameworks that enable prosumers to sell back electricity to the grid. These measures would enhance the profitability of the investment scenario and thus contribute to the sustainable development of the residential energy consumption subsector.

However, this study is limited in several aspects. First, this study takes all parameters as given at the beginning and is deterministic. Nevertheless, in the actual market, some parameters are revealed a few minutes before the agent decides. For instance, although one can use the forecasts of real-time electricity price, consumption, and renewable generation, the deterministic model presented in this chapter cannot handle the variance and risk involved in their predictions. Therefore, developing a stochastic dynamic model would make the model more realistic, and the risks involved could be assessed and managed. Second, making a shared investment in DER by a group of neighbors or sharing the existing DER devices can benefit a coalition of prosumers. Also, the coalition members may have local electricity trade through an aggregator and be more independent in the grid; investigating such coalitions can introduce a new agent with more power and capabilities to this context and will identify new opportunities and challenges of investing in smart homes. Furthermore, since this study considers only one prosumer, it ignores the effect of mass adoption by the prosumers on the total load and grid prices. Indeed, if a large enough group of prosumers change their consumption pattern and follow the same price incentive, they may create new

peak loads, imposing tremendous costs on the grid. Studying the consequences of the mass adoption of smart-home technologies and demand response is another direction that can be considered for exploration.

Moreover, in this study, the distributor tries to shave the monthly peak loads of a single house, which is based on that house's personal load. However, different houses may have different peaks, and the profitability of an investment scenario depends on how the price signals and demand patterns match. Hence, more reliable results would be obtained when aggregated consumer loads are considered. Another future research direction could be investigating the coordination and settlement challenges that arise when simultaneously controlling EV charging and their interaction with the power distribution network. Specifically, the study could explore methods to effectively coordinate control between the EV owner and the distributor while also addressing the complexities introduced by the mobility of EVs and the need for additional settlement systems. The research aims to maximize the overall welfare of both parties involved in the EV charging process.

1.5 Appendices

1.5.1 Mathematical Model

In order to optimally manage the electricity flow in the smart home under different tariff schemes and investment scenarios, two linear mathematical models are developed in this section. As explained before, the investment and management of the smart home can be made by the distributor or the prosumer; hence, the elements of the problem will be explained, and then two models will be formulated.

1.5.1.1 Indexes and Parameters

This study uses the generation and consumption records of four single detached houses in 2019, and the smart-home management problem will be solved for the whole year of 2019 (8,760 hours) with 15-minute time steps (t). When 1 kWh is used during this time interval, 4 kW of power/capacity is used. Moreover, the network capacity/transmission costs are

calculated based on the monthly peak loads. Therefore, two indexes are used in this study as below:

$$t \in \{\mathcal{T}\} \qquad \text{episode } t \text{ in horizon } [1, 2, 3, ..., T]$$
$$k \in \{\mathcal{K}\} \qquad \text{month } k \text{ in horizon } [1, 2, 3, ..., K]$$

Below is the list of parameters:

$B^{\max} \doteq$ capacity (kWh) of battery $= \{0, 10, 27, 40.5\}$

$B^{\min} \doteq$ minimum allowed electricity (kWh) in battery $= 0$

$E^{\max} \doteq$ capacity (kWh) of EV $= 60$

$E^{\min} \doteq$ minimum allowed electricity (kWh) in EV $= 24$

$U^{\mathrm{c}} \doteq$ charging capacity (kWh) of battery/EV during the interval $= \{0, 1.9, 2.5, 3.75\}$

$U^{\mathrm{d}} \doteq$ discharging capacity (kWh) of battery/EV during the interval $= \{0, 1.9, 2.5, 3.75\}$

$1 - \eta^{\mathrm{c}} \doteq$ charging loss rate $= 0.05$

$1 - \eta^{\mathrm{d}} \doteq$ discharging loss rate $= 0.05$

$P^{\mathrm{rns}} \doteq$ cost coefficient (\$/kW) of RNS $= 9$

$P^{\mathrm{cap}} \doteq$ cost coefficient (\$/kW) of capacity $= 5$

$P_t^{\mathrm{buy}} \doteq$ electricity price (\$/kWh) to buy from grid at time t

$P_t^{\mathrm{sell}} \doteq$ electricity price (\$/kWh) to sell to grid at time t

$L_t \doteq$ load (kWh) at time t

$A_t \doteq$ availability of EV at time t under the chosen scenario

$V_t \doteq$ EV usage (kWh) (for vehicle riding) at time t under the chosen scenario

$R_t \doteq$ electricity generation (kWh) from solar panel at time t

$I^{\mathrm{trade}} \doteq$ indicates if it is allowed to sell electricity to grid

1.5.1.2 Decision and State Variables

As depicted in Figure 1.5, the decision variable set indicates the electricity flows and the monthly peak loads in the problem; the list of the decision variables is as below:

$x_t^{\mathrm{GL}} \doteq$ electricity (kWh) from grid to load at time t

$x_t^{\mathrm{GB}} \doteq$ electricity (kWh) from grid to battery at time t

$x_t^{\mathrm{GE}} \doteq$ electricity (kWh) from grid to EV at time t

$x_t^{\mathrm{RL}} \doteq$ electricity (kWh) from solar panel to load at time t

$x_t^{\mathrm{RB}} \doteq$ electricity (kWh) from solar panel to the battery at time t

$x_t^{\mathrm{RE}} \doteq$ electricity (kWh) from solar panel to EV at time t

$x_t^{\mathrm{RG}} \doteq$ electricity (kWh) from solar panel to the grid at time t

$x_t^{\mathrm{BL}} \doteq$ electricity (kWh) from battery to load at time t

$x_t^{\mathrm{EL}} \doteq$ electricity (kWh) from EV to load at time t

$x_t^{\mathrm{BG}} \doteq$ electricity (kWh) from battery to grid at time t

$x_t^{\mathrm{EG}} \doteq$ electricity (kWh) from EV to grid at time t

$x_k^{\mathrm{peak}} \doteq$ maximum electricity (kWh) taken from the grid in an episode in month k

Moreover, the energy stored in the storage devices is defined as the state variables of the problem:

$b_t^{\mathrm{B}} \doteq$ state of the battery (available electricity (kWh) in battery) at time t

$b_t^{\mathrm{E}} \doteq$ state of the EV (available electricity (kWh) in battery) at time t

1.5.1.3 Distributor's Model

The model below tries to minimize the annual casts of the distributor that invests and controls the system:

$$\min \sum_{t=1}^{T} \Big(P_t^{buy}(x_t^{GL} + x_t^{GB} + x_t^{GE}) - P_t^{sell}(x_t^{RG} + \eta^d x_t^{BG} + \eta^d x_t^{EG}) \Big) + 4 \sum_{k=1}^{12} (P^{rns} + 1.35 P^{cap}) x_k^{peak} \tag{1}$$

subject to:

$$x_t^{\mathrm{GL}} + \eta^{\mathrm{d}}(x_t^{\mathrm{BL}} + x_t^{\mathrm{EL}}) + x_t^{\mathrm{RL}} \geq L_t \qquad \forall t \tag{2}$$

$$x_t^{\mathrm{BL}} + x_t^{\mathrm{BG}} \leq b_t^{\mathrm{B}} \qquad \forall t \tag{3}$$

$$x_t^{\mathrm{EL}} + x_t^{\mathrm{EG}} \leq b_t^{\mathrm{E}} A_t \qquad \forall t \tag{4}$$

$$x_t^{\mathrm{GE}} + x_t^{\mathrm{RE}} \leq E^{\max} A_t \qquad \forall t \tag{5}$$

$$x_t^{\mathrm{BL}} + x_t^{\mathrm{EL}} + x_t^{\mathrm{BG}} + x_t^{\mathrm{EG}} \leq U^{\mathrm{d}} \qquad \forall t \tag{6}$$

$$x_t^{\mathrm{GB}} + x_t^{\mathrm{GE}} + x_t^{\mathrm{RB}} + x_t^{\mathrm{RE}} \leq U^{\mathrm{c}} \qquad \forall t \tag{7}$$

$$B^{\min} \leq b_t^{\mathrm{B}} \leq B^{\max} \qquad \forall t \tag{8}$$

$$E^{\min} \leq b_t^{\mathrm{E}} \leq E^{\max} \qquad \forall t \tag{9}$$

$$x_t^{\mathrm{RL}} + x_t^{\mathrm{RB}} + x_t^{\mathrm{RE}} + x_t^{\mathrm{RG}} \leq R_t \qquad \forall t \tag{10}$$

$$b_t^{\mathrm{B}} + \eta^{\mathrm{c}}(x_t^{\mathrm{GB}} + x_t^{\mathrm{RB}}) - (x_t^{\mathrm{BL}} + x_t^{\mathrm{BG}}) = b_{t+1}^{\mathrm{B}} \qquad \forall t \tag{11}$$

$$b_t^{\mathrm{E}} + \eta^{\mathrm{c}}(x_t^{\mathrm{GE}} + x_t^{\mathrm{RE}}) - (x_t^{\mathrm{EL}} + x_t^{\mathrm{EG}}) - \frac{V_t}{\eta^{\mathrm{d}}} = b_{t+1}^{\mathrm{E}} \qquad \forall t \tag{12}$$

$$x_t^{\mathrm{GL}} + x_t^{\mathrm{GB}} + x_t^{\mathrm{GE}} \leq x_k^{\mathrm{peak}} \qquad \forall t \in k, k \tag{13}$$

$$b_0^{\mathrm{B}} = B^{\min}, b_0^{\mathrm{E}} = E^{\min} \tag{14}$$

$$x_t^{RL}, x_t^{RB}, x_t^{RE}, x_t^{RG}, x_t^{BG}, x_t^{EG}, x_t^{BL}, x_t^{EL}, x_t^{GL}, x_t^{GB}, x_t^{GE}, x_k^{peak}, b_t^B, b_t^E \geq 0 \qquad \forall t \tag{15}$$

Where objective function 1 consists of three terms (load covering cost + revenue of selling back electricity to the grid + network charges (transmission+capacity)) and minimizes electricity consumption cost. Constraint 2 makes sure the load is covered. Constraints 3 to 5 state that electricity taken out of the battery or EV must be firstly less than the available energy stored in them and secondly, electricity can be taken from or sent to EV, when it is available. Constraints 6 and 7 indicate the charging and discharging limits. Constraints 8 and 9 make sure the state of the battery and EV are in the allowed range of stored energy

in them. Constraint 10 checks electricity sent from the solar panel would be less than or equal to the generated electricity and constraints 11 and 12 calculate the state of battery and EV in the next episode. Constraint 13 calculates the maximum electricity taken from the grid in month k. Finally, constraint 14 shows that the initial value of the states of the battery and EV are equal to their minimum values, and, constraint 15 specifies the domain of the decision variables.

1.5.1.4 Prosumer's Model

When the prosumer invests and controls the system, s/he does not consider the network charges; then, the new objective function would be different.

$$\min \sum_{t=1}^{\mathrm{T}} \left(P_t^{\mathrm{buy}} (x_t^{\mathrm{GL}} + x_t^{\mathrm{GB}} + x_t^{\mathrm{GE}}) - P_t^{\mathrm{sell}} (x_t^{\mathrm{RG}} + \eta^{\mathrm{d}} x_t^{\mathrm{BG}} + \eta^{\mathrm{d}} x_t^{\mathrm{EG}}) \right) \tag{16}$$

Moreover, along with the constraints 2 to 15, to evaluate the effects of allowance of selling back the energy to the grid, a new parameter I^{trade} with a constraint as below is added to the model:

$$x_t^{\mathrm{RG}} + x_t^{\mathrm{BG}} + x_t^{\mathrm{EG}} \leq (U^{\mathrm{d}} + R_t) I^{\mathrm{trade}} \qquad \forall t \tag{17}$$

where objective function 16 minimizes electricity consumption cost and constraint 17 makes sure the prosumer can sell electricity to the grid if the decentralized trade is allowed.

The HTML report and coding files are shared in the online appendix.

1.5.2 Detailed NPVs in Sensitivity Analysis

Table 1.7: Scenarios NPVs for all houses when prosumer controls and selling back is allowed

rate	**S0**	**S1**	**S2**	**S3**	**S4**	**S5**	**House**
RTLMP	0	-24,493	-30,137	1,623	-8,406	-14,516	
TOU&RTLMP	0	-24,109	-23,946	-1,258	-7,490	-11,331	
TOU	0	-12,819	-6,279	6,671	11,221	9,712	House 1
TOU EV & Uniform	0	-27,300	-16,455	6,413	11,938	5,359	
Uniform	0	-27,300	-16,468	-4,549	964	-5,615	
Distributor controls	0	-15,953	-19,769	1,745	-6,458	-5,134	
RTLMP	0	-24,493	-30,137	1,623	-8,407	-14,517	
TOU&RTLMP	0	-24,305	-24,016	-764	-7,645	-11,808	
TOU	0	-12,820	-6,279	6,672	11,221	9,711	House 2
TOU EV & Uniform	0	-27,300	-16,455	6,413	11,938	5,359	
Uniform	0	-27,300	-16,469	-4,549	963	-5,616	
Distributor controls	0	-14,903	-18,702	2,528	-5,484	-4,189	
RTLMP	0	-24,492	-30,136	1,623	-8,406	-14,516	
TOU&RTLMP	0	-22,528	-20,952	-437	-5,420	-8,964	
TOU	0	-12,816	-6,275	6,673	11,224	9,716	House 3
TOU EV & Uniform	0	-27,300	-16,455	6,414	11,939	5,360	
Uniform	0	-27,300	-16,467	-4,549	966	-5,613	
Distributor controls	0	-11,498	-17,403	3,803	-4,196	-2,970	
RTLMP	0	-24,492	-30,137	1,623	-8,406	-14,516	
TOU&RTLMP	0	-23,404	-22,266	-748	-6,507	-10,210	
TOU	0	-12,818	-6,277	6,672	11,222	9,714	House 4
TOU EV & Uniform	0	-27,300	-16,455	6,413	11,938	5,359	
Uniform	0	-27,300	-16,468	-4,549	965	-5,614	
Distributor controls	0	-13,750	-17,467	3,242	-5,332	-2,908	

References

Aasbøe, T. N. (2021). Importance of drivers and barriers for v2g? Master's thesis, Norwegian University of Life Sciences, Ås.

Aguilar-Dominguez, D., Dunbar, A., and Brown, S. (2020). The electricity demand of an ev providing power via vehicle-to-home and its potential impact on the grid with different electricity price tariffs. *Energy Reports*, 6:132–141. 4th Annual CDT Conference in Energy Storage &Its Applications.

Ali, H., Hussain, S., Khan, H. A., Arshad, N., and Khan, I. A. (2020). Economic and environmental impact of vehicle-to-grid (v2g) integration in an intermittent utility grid. In *2020 2nd International Conference on Smart Power & Internet Energy Systems (SPIES)*, pages 345–349. IEEE.

Allen, B. (2019). Adopting sustainable innovations: A case on renewables' integration into the grid.

Asensio, M., de Quevedo, P. M., Muñoz-Delgado, G., and Contreras, J. (2016). Joint distribution network and renewable energy expansion planning considering demand response and energy storage—part i: Stochastic programming model. *IEEE Transactions on Smart Grid*, 9(2):655–666.

Avau, M., Govaerts, N., and Delarue, E. (2021). Impact of distribution tariffs on prosumer demand response. *Energy Policy*, 151:112116.

BAUMGARTNER, N., KELLERER, F., RUPPERT, M., HIRSCH, S., MANG, S., and FICHTNER, W. (2022). Does experience matter? assessing user motivations to accept a vehicle-to-grid charging tariff. *Transportation Research Part D: Transport and Environment*, 113:103528.

Bentley, E., Putrus, G., Lacey, G., Kotter, R., Wang, Y., Das, R., Ali, Z., and Warmerdam, J. (2021). On beneficial vehicle-to-grid (v2g) services. In *2021 9th International Conference on Modern Power Systems (MPS)*, pages 1–6. IEEE.

Bergek, A. and Mignon, I. (2017). Motives to adopt renewable electricity technologies: Evidence from sweden. *Energy Policy*, 106:547–559.

Bergek, A., Mignon, I., and Sundberg, G. (2013). Who invests in renewable electricity production? empirical evidence and suggestions for further research. *Energy Policy*, 56:568–581.

Bertsch, V. and Di Cosmo, V. (2020). Are renewables profitable in 2030 and do they reduce carbon emissions effectively? a comparison across europe.

Bibak, B. and Tekiner-Mogulkoc, H. (2021). Influences of vehicle to grid (v2g) on power grid: An analysis by considering associated stochastic parameters explicitly. *Sustainable Energy, Grids and Networks*, 26:100429.

Curtin, J., McInerney, C., Gallachóir, B. Ó., and Salm, S. (2019). Energizing local communities—what motivates irish citizens to invest in distributed renewables? *Energy Research & Social Science*, 48:177–188.

Gough, R., Dickerson, C., Rowley, P., and Walsh, C. (2017). Vehicle-to-grid feasibility: A techno-economic analysis of ev-based energy storage. *Applied energy*, 192:12–23.

Haapaniemi, J., Narayanan, A., Tikka, V., Haakana, J., Honkapuro, S., Lassila, J., Kaipia, T., and Partanen, J. (2017). Effects of major tariff changes by distribution system operators on profitability of photovoltaic systems. In *2017 14th International Conference on the European Energy Market (EEM)*, pages 1–6.

Halman, N., Nannicini, G., and Orlin, J. (2018). On the complexity of energy storage problems. *Discrete Optimization*, 28:31–53.

Huang, S. and Wu, Q. (2019). Dynamic tariff-subsidy method for pv and v2g congestion management in distribution networks. *IEEE Transactions on Smart Grid*, 10(5):5851–5860.

Imcharoenkul, V. and Chaitusaney, S. (2020). Optimal renewable energy integration considering minimum dispatchable generation operating costs. pages 494–497.

Independent System Operator – New England" (ISO-NE) (2021a). Monthly regional, network load cost report, october 2021.

Independent System Operator – New England" (ISO-NE) (2021b). Wholesale load cost report, november 2021.

International Energy Agency (IEA) (2020). Global ev outlook 2020. *URL: https://www. iea. org/reports/global-ev-outlook-2020*.

Jeon, W., Cho, S., and Lee, S. (2020). Estimating the impact of electric vehicle demand response programs in a grid with varying levels of renewable energy sources: Time-of-use tariff versus smart charging. *Energies*, 13(17).

Kök, A. G., Shang, K., and Yücel, Ş. (2018). Impact of electricity pricing policies on renewable energy investments and carbon emissions. *Management Science*, 64(1):131–148.

Kosmadakis, G., Karellas, S., and Kakaras, E. (2013). *Renewable and Conventional Electricity Generation Systems: Technologies and Diversity of Energy Systems*, pages 9–30. Springer London, London.

Krozer, Y. (2013). Cost and benefit of renewable energy in the european union. *Renewable Energy*, 50:68–73.

Li, X., Tan, Y., Liu, X., Liao, Q., Sun, B., Cao, G., Li, C., Yang, X., and Wang, Z. (2020). A cost-benefit analysis of v2g electric vehicles supporting peak shaving in shanghai. *Electric Power Systems Research*, 179:106058.

Ma, Y., Chen, Y., Chen, X., Deng, F., and Song, X. (2019). Optimal dispatch of hybrid energy islanded microgrid considering v2g under tou tariffs. 107.

Mozafar, M. R., Amini, M. H., and Moradi, M. H. (2018). Innovative appraisement of smart grid operation considering large-scale integration of electric vehicles enabling v2g and g2v systems. *Electric Power Systems Research*, 154:245–256.

Mullan, J., Harries, D., Bräunl, T., and Whitely, S. (2012). The technical, economic and commercial viability of the vehicle-to-grid concept. *Energy Policy*, 48:394–406.

Peterson, S. B., Whitacre, J., and Apt, J. (2010). The economics of using plug-in hybrid electric vehicle battery packs for grid storage. *Journal of Power Sources*, 195(8):2377–2384.

Raj, P. (2019). Assessing the monetary value of vehicle-to-grid considering battery degradation: Agent-based approach.

Richardson, D. B. (2013). Encouraging vehicle-to-grid (v2g) participation through premium tariff rates. *Journal of Power Sources*, 243:219–224.

Salisbury, M. and Toor, W. (2016). How and why leading utilities are embracing electric vehicles. *The Electricity Journal*, 29(6):22–27.

Segreto, M., Principe, L., Desormeaux, A., Torre, M., Tomassetti, L., Tratzi, P., Paolini, V., and Petracchini, F. (2020). Trends in social acceptance of renewable energy across europe—a literature review. *International Journal of Environmental Research and Public Health*, 17(24):9161.

Singh, R. (2021). Solar-city plans with large-scale energy storage: Metrics to assess the ability to replace fossil-fuel based power. *Sustainable Energy Technologies and Assessments*, 44:101065.

Sioshansi, R. (2012). Or forum—modeling the impacts of electricity tariffs on plug-in hybrid electric vehicle charging, costs, and emissions. *Operations Research*, 60:506–516.

Smith, M. G. and Urpelainen, J. (2014). The effect of feed-in tariffs on renewable electricity generation: An instrumental variables approach. *Environmental and resource economics*, 57(3):367–392.

Thomas, D., Deblecker, O., and Ioakimidis, C. S. (2018). Optimal operation of an energy management system for a grid-connected smart building considering photovoltaics' uncertainty and stochastic electric vehicles' driving schedule. *Applied Energy*, 210:1188–1206.

U.S. Energy Information Administration (EIA) (2021a). Electric power annual.

U.S. Energy Information Administration (EIA) (2021b). Vermont electricity profile 2020.

U.S. Energy Information Administration (EIA) (2021c). World electricity final consumption by sector, 1974-2019.

Vermont Department of Environmental Conservation (2023). Zero emission vehicles (zev).

Wei, H., Zhang, Y., Wang, Y., Hua, W., Jing, R., and Zhou, Y. (2022). Planning integrated energy systems coupling v2g as a flexible storage. *Energy*, 239:122215.

Wu, X., Freese, D., Cabrera, A., and Kitch, W. A. (2015). Electric vehicles' energy consumption measurement and estimation. *Transportation Research Part D: Transport and Environment*, 34:52–67.

Zhao, Y., Noori, M., and Tatari, O. (2016). Vehicle to grid regulation services of electric delivery trucks: Economic and environmental benefit analysis. *Applied energy*, 170:161–175.

Chapter 2

Smart Grids, Smart Pricing: Employing Reinforcement Learning for Prosumer-Responsive Critical Peak Pricing

Abstract

This chapter focuses on Critical Peak Pricing (CPP), a rising smart pricing option, as a demand response strategy for peak load shaving in electricity grids. By integrating different prosumer profiles into the analysis, reflecting the increasing penetration of distributed energy resources such as photovoltaic panels, batteries, and electric vehicles, we can identify the optimal customer participation in CPP. Through comprehensive simulations and the application of reinforcement learning algorithms, we analyze the effectiveness of CPP programs, both in mass and targeted offering scenarios. The results reveal that while CPP is effective in incentivizing load shifting, its efficacy diminishes with increasing prosumer participation, leading to new peaks. To counteract this, we propose targeted dynamic pricing strategies demonstrating significantly improved performance and extended viability. The study also highlights the influential role of batteries and electric vehicles in peak load reduction,

suggesting a need for focused policy and incentive structures.

2.1 Introduction

As the world grapples with the escalating demand for clean electricity, the pursuit of sustainable energy consumption patterns has brought peak load shaving to the forefront of energy management strategies (International Energy Agency (IEA), 2024). Peak load shaving, a crucial aspect of Demand Response (DR) efforts, mitigates the disparity between high and low loads, reducing stress on energy infrastructure and lowering electricity costs and carbon emissions from reliance on inefficient, fossil-fuel-based peak power plants. Amidst this global challenge, dynamic pricing emerges as a pivotal tool for addressing peak load challenges, with Critical Peak Pricing (CPP) holding particular promise.

This chapter explores the application of CPP to incentivize consumers to reduce energy consumption during critical peak events through financial incentives or penalties. The rationale behind focusing on CPP, rather than other dynamic pricing strategies such as Time-of-Use (TOU) or Real-Time Pricing (RTP), stems from its comparative growth and potential for impact. While the United States has observed a significant increase in the adoption of TOU and RTP programs for residential customers (48% and 154%, respectively, from 2013 to 2022), CPP has only seen modest growth of 17% in the same sector. However, the adoption of CPP programs for commercial and industrial customers has surged by 80% and 90%, respectively (U.S. Energy Information Administration, 2023). This divergent trend highlights the untapped potential of CPP in the residential sector, motivating this study to investigate CPP's viability and efficacy for residential customers.

Implementing CPP, however, presents challenges, the foremost being accurately forecasting peak load periods. Fluctuating weather conditions, dynamic market situations, and diverse consumer behaviors contribute to the complexity of predicting these events (Chan et al., 2012). In addition, determining the optimal number of Critical Peak Events (CPEs) is crucial for balancing financial incentives and avoiding undue strain on utility companies. CPEs are typically 3 or 4 hour time windows of anticipated high demand during which distributors offer rate incentives to lower consumption.

The emergence of Distributed Energy Resources (DERs) such as PhotoVoltaic (PV) solar panels, standalone Batteries (BATs), and Electric Vehicles (EVs) has introduced a new dynamic to energy consumption patterns, giving rise to prosumers — consumers who not only consume but also produce and store energy (Parag and Sovacool, 2016). Prosumers introduce additional variability and unpredictability in load profiles, further complicating forecasting peak load events and determining appropriate CPE announcements. Their ability to adjust energy consumption and generation in response to price signals, with objectives that may diverge from peak load reduction goals, poses additional challenges for CPP implementation.

To address these challenges, this study introduces Reinforcement Learning (RL) optimization algorithms to identify CPEs in scenarios where multiple prosumer profiles coexist. The complexity of this problem arises from diverse prosumer behaviors and objectives, as well as the dynamic environment where parameters are realized in real-time. Deep RL techniques are well-suited to handle such complexities by learning optimal policies from interactions with the environment without explicitly modelling system dynamics (Sutton and Barto, 2018). Moreover, considering the dynamic environment where parameters are realized in real-time makes the results more robust and closer to real-world scenarios, ensuring that identified CPEs are better aligned with the evolving energy landscape, leading to more effective peak load reduction strategies.

CPP programs can be categorized into two types: penalty-based, which imposes higher electricity rates during peak events, and rebate-based, which offers rebates for reduced consumption during these periods. Both mechanisms are in use in Quebec (Canada), the location chosen as a context for this analysis. This chapter investigates their effectiveness in financial savings and peak load reduction. By identifying the optimal pricing strategy for the distributor, the study calculates the contribution of each prosumer profile to peak load shaving and the financial benefits they can achieve by responding optimally to the CPEs.

As the prevalence of prosumers increases, understanding their impact on the electricity grid and the effectiveness of demand response programs becomes vital (Angelus, 2021). This research delves into how the growing prosumer population, equipped with energy-generating and storage technologies, can benefit distributors. Specifically, it examines the

effectiveness of mass CPEs, where all consumers receive the same incentive to reduce usage simultaneously. The study introduces a novel concept of targeted dynamic pricing designed to efficiently manage prosumers' load adjustments in response to CPEs, ensuring a more balanced distribution of demand response efforts across the grid. Furthermore, it investigates the critical point at which offering the same incentive to all becomes ineffective due to varying capabilities of prosumer profiles. This analysis is crucial for policymakers to determine the most beneficial strategies for encouraging consumers to adopt prosumer technologies and behaviors that support grid stability and efficiency.

As some prosumers, given their profile, are inherently more capable of adjusting their energy consumption and production in response to pricing signals, they naturally stand to receive more offers and achieve greater financial savings. This discrepancy can foster a perception of unfairness among consumers with less adaptive capacity or fewer resources to become prosumers. To address this concern, the study explores targeted CPP as a solution to optimize demand response while incorporating a fairness constraint into its proposed algorithms. By balancing the benefits distributed among all network participants, the study ensures that transitioning towards a more prosumer-driven model does not inadvertently create disparities. This balanced approach can help policymakers and distributors identify when and how to encourage consumer transition towards prosumer status, considering the broader implications for grid stability and equitable access to energy savings.

The remainder of the study comprises a literature review, the formulation and description of mathematical models, a discussion of the implemented solution approaches, and an analysis of real-world data to evaluate the efficacy of the proposed strategies. Through these components, the study contributes to the existing body of knowledge on DR, prosumer behavior, and CPP, providing insights and tools for utility companies to manage electricity demand better and maintain grid stability in an increasingly complex and dynamic energy landscape.

2.2 Literature Review

Peak load shaving is crucial for sustainable energy systems, supporting SDG 7's goal of universal access to modern energy.[1] It reduces electricity grid costs and enhances reliability by minimizing peak demand, thus making energy more affordable and preventing supply disruptions (Parker et al., 2019). Additionally, it facilitates the integration of renewable energy, contributing to a sustainable and environmentally friendly energy mix (Silva et al., 2020).

Dynamic pricing emerges as a promising strategy for peak load shaving, adjusting electricity prices based on real-time supply and demand (Liang et al., 2013). It encourages consumers to shift usage to off-peak times, reducing the load and pressure on the grid during peak events. Key dynamic pricing approaches include TOU, CPP, and RTP, each with unique mechanisms for managing consumption during high-demand periods. CPP has the advantage to be more flexible than TOU while more stable for consumers than Real-Time Pricing. It charges consumers higher prices or offers rebates for consumption reduction during pre-identified events, when the grid is under stress. For example, CPP encourages consumers to reduce their electricity consumption during heat waves or cold snaps.

This section delves into the application of dynamic pricing for managing peak electricity loads, focusing on four critical aspects: the deployment of CPP; how consumer behavior and willingness to participate in demand response programs are affected; the use of RL techniques to fine-tune pricing and scheduling for residential loads; and identifying the challenges and open questions in implementing effective dynamic pricing strategies. By examining these areas, we aim to illuminate the current research landscape on CPP's role in promoting demand flexibility for peak reduction.

2.2.1 Critical Peak Pricing

CPP has become a popular demand response strategy for managing peak loads in electricity grids. Early studies by Herter (2007) and Wolak (2007) provided initial evidence that CPP can motivate significant reductions in peak demand in the residential sector. Herter

[1]Sustainable Development Goals (SDGs) are goals adopted by the United Nations, see https://sdgs.un.org/goals/goal7

and Wayland (2010) later quantified CPP impacts for households across different usage levels, climate zones, and building types, finding that high-use customers reduce the most in absolute terms while low-use customers achieve greater percentage savings. More advanced strategies have been developed to optimize residential CPP response. Javaid et al. (2018) designed algorithms to efficiently schedule appliances and minimize costs for households facing CPP rates.

Beyond residential applications, CPP has also been studied extensively for commercial and industrial users. Jang et al. (2015) empirically demonstrated substantial heterogeneity in CPP response across different types of commercial/industrial customers. They found a strong correlation between electricity expenditure shares and price responsiveness. Building on this, Park et al. (2015) formulated a mathematical model to design optimal CPP rates that maximize retailer profits based on the heterogeneous price sensitivity across customer segments. Additionally, Piette et al. (2006) showed that automated CPP strategies for commercial buildings can yield significant peak demand reductions .

Several studies have focused specifically on implementation strategies and incentives in CPP program design. Zhang (2014) developed an optimization model that accounts for wind generation commitments when scheduling CPP events. Their goal was to minimize total system costs spanning energy, CPP rebates, and wind imbalance penalties. Zhang et al. (2009) similarly co-optimized CPP parameters to balance bill savings for consumers against cost reductions for utilities . And more recently, Aurangzeb et al. (2021) proposed differentiated CPP prices for low versus high energy users to reduce cross-subsidization.

Multiple papers have cited enabling technologies, dynamic pricing, and better customer segmentation as key facilitators for CPP. For example, Fitzpatrick et al. (2020) demonstrated that optimized control logic for heat pump-storage systems can significantly increase responsiveness to CPP rates. Silva et al. (2020) described various peak load management strategies in their broad review and highlighted opportunities for automation and customer targeting . Moreover, Lavin and Apt (2021) specifically argue that realizing system-level benefits from distributed storage requires peak pricing incentives during the highest load hours of the year. Finally, Faruqui and Sergici (2010) extensively surveyed multiple pilots. They found a consistent trend that more dynamic CPP designs, along with technologies like

programmable thermostats, deliver the greatest peak reductions, in the range of 27-44%. Jessoe et al. (2014) also provide evidence that non-price factors influence household choices, indicating additional complexities behind consumer behavior.

2.2.2 Consumer Behavior and Willingness to Adopt DR

Understanding and predicting consumer behavior is critical for designing effective demand response programs. As an introduction to this issue, Vassileva et al. (2012) highlighted the need to tailor demand response information and incentives to consumers' heterogeneous characteristics and preferences. Gao et al. (2020) further proposed incorporating risk attitudes and learning processes into consumer response models. These insights provide a foundation for evaluating demand response proposals.

The first key question is the level of price signals needed to trigger consumer responsiveness. Using Monte Carlo simulation on survey data, He et al. (2012) found consumers only respond to substantial peak-price increases of 20-40% through TOU rates. Complementarily, Sundt et al. (2020) estimated via a choice experiment that offering bill discounts can elicit demand shifting from most consumers on TOU tariffs. However, Annala (2015) suggested, based on focus groups, that the discounts would need to be relatively high to significantly impact behavior beyond simple load control that doesn't require habit changes.

In exploring the challenges associated with adopting voluntary TOU tariffs, Choi et al. (2020) highlight the significant impact of consumer heterogeneity in preferences and the resulting trade-offs for utility firms. This complements the findings on consumer responsiveness to price signals and incentives, further explaining why TOU tariffs have not achieved widespread adoption despite their potential benefits.

Predicting the extent of consumer responsiveness is also crucial. Liu et al. (2019) put forward long short-term memory neural networks to predict consumer demand response patterns based on analyzing historical data. Alternatively, Kwag and Kim (2014) developed reliability models capturing uncertainty in consumer behavior for demand response planning. Zeng et al. (2017) specifically incorporated the correlation between consumers' past experiences with profitability and future willingness to participate.

There are also important program design considerations around consumer incentives and

types of loads targeted. Ming et al. (2023) found social inefficiencies in current decentralized incentive designs and proposed modifications to improve welfare. Wang et al. (2020) Used regression analysis on surveys to reveal income level and energy saving attitudes as key participation drivers, with customers focused more on direct incentives over dynamic prices. Sridhar et al. (2023) similarly estimated via a comparitive analysis that consumers require higher compensation for heating versus appliance control.

Furthermore, Mirzaei et al. (2023) developed a model for large industrial consumers showing integrated demand response can provide additional electricity bill savings by strategically shifting loads to influence market prices. Relatedly, Naeem et al. (2015) suggested differentiated multi-tier programs based on both financial and environmental motives among heterogeneous customers.

2.2.3 Applications of RL in Peak Load Management

Optimization techniques are essential in smart grid management to enhance efficiency, reduce operational costs, and balance supply and demand. These techniques analyze various variables, such as energy consumption and resource availability, to devise strategies for peak load management (Akkara and Selvakumar, 2023). However, traditional optimization approaches often struggle with the dynamic and uncertain nature of power systems, such as unpredictable renewable energy sources and changing consumer behaviors. These limitations underscore the need for more adaptive and powerful solutions, paving the way for the application of RL in peak load management (Vázquez-Canteli and Nagy, 2019).

RL has emerged as a promising data-driven technique for tackling challenges in smart grid management and peak load reduction. Sheikhi et al. (2016) developed an RL-based approach that enhanced system adaptability and responsiveness to fluctuations in energy supply and demand, beyond just improving efficiency and reducing peak loads. Building on this, several other studies have formulated the residential load scheduling problem as RL tasks and developed algorithms that effectively reduce electricity bills and peak demands (Remani et al., 2019; Mathew et al., 2020).

Another major application area has been using RL for dynamic pricing in smart grids. Facing uncertainties about customer demand patterns and volatility in wholesale electricity

prices, Kim et al. (2014, 2016) proposed RL algorithms that can learn pricing and consumption scheduling policies without needing upfront system knowledge. Following this work, RL techniques have been widely adopted for optimizing dynamic pricing strategies to balance grid supply and demand. For example, Zhong et al. (2021) applied a deep RL method to determine dynamic subsidies for aggregators managing clusters of residential electric heating systems. And Zhang et al. (2022) developed a distributed RL pricing approach that aligns the interests of individual users and the power supplier.

More recent work has focused on applying RL specifically for electric vehicle charging infrastructure and integrating pricing incentives to shift charging loads. Moghaddam et al. (2020) proposed an online RL charging station pricing model to flatten the duck curve effect from renewables while increasing station revenue. Wang et al. (2021) developed an online RL algorithm for a public EV charging station to jointly optimize pricing and charging schedules to maximize total station profit.

On the demand side, accurately modelling and influencing flexible user consumption via pricing signals poses challenges. To address this, Ghasemkhani and Yang (2018) applied RL to learn models of users' responses to prices for demand response management without assuming known and fixed response functions. Vázquez-Canteli and Nagy (2019) surveyed the use of RL for various demand-side management applications while discussing the importance of integrating human comfort preferences. Expanding on this, Ismail and Baysal (2023) recently employed actor-critic RL to determine optimal dynamic pricing and control user demand response simultaneously.

2.2.4 Summary and Research Gap

The review highlights the critical role of dynamic pricing, specifically CPP, in peak load management and the nuances of consumer behavior, alongside the use of RL for optimizing energy consumption. It reveals a need for deeper research into CPP's impact in the residential sector, the integration of DERs, and the transition to prosumer models affecting pricing strategies. Notably, there's a gap in developing dynamic pricing models that consider prosumer diversity and ensure equitable demand response. Moreover, empirical validation of theoretical models in real-world settings is essential for bridging academic research with

Profile #	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Flexibility options		PV				PV	PV	PV				PV	PV	PV		PV
			EV			EV			EV	EV		EV	EV		EV	EV
				Bat			Bat		Bat		Bat	Bat		Bat	Bat	Bat
					DR			DR		DR	DR		DR	DR	DR	DR

Figure 2.1: Prosumers' profiles

practical implementation, underscoring the importance of tailored and equitable dynamic pricing schemes for grid stability and efficiency.

2.3 Mathematical Modeling

In this study, two entities are considered: a group of prosumers and a distributor. Prosumers might be equipped with PV, EV, BAT, or be able to modify their demand contingent on the pricing signals they encounter (DR). DR can be either a technical device (like a programmable thermostat) or a behavioral commitment (like the willingness to reduce consumption). As Figure 2.1 represents, by amalgamation of the four flexibility options, 15 distinct profiles are delineated, in addition to a baseline profile. This chapter assumes a uniform distribution of these profiles among the prosumer group since we do not have information on the actual distribution of profiles within the consumer population.

Prosumers manage their electrical energy flow within their domiciles to optimize cost efficiency and minimize the inconvenience of demand adjustments. Conversely, the distributor aims at increasing the net revenue from electricity sales to consumers while minimizing the monthly peak demand, which incurs capacity-related expenditures. For this objective, as illustrated in Figure 2.2, after the distributor obtains the required information (including consumers' load, prosumers' best response function, market and weather forecast, among others), the distributor determines if a CPE will be announced for the next day or not. This cycle repeats throughout the month, and at the end of the month, the monthly peak demand is realized. For the Quebec case, two strategies (Winter Credit Option - WCO, and Flex D - FXD) are in place to incentivize consumers to reduce or shift their consumption during peak periods (Hydro-Quebec, 2024a).

Winter Credit Option (WCO)

WCO is a dynamic pricing strategy that works with the standard Rate D, the standard residential rate in Quebec. Under Rate D, consumers are charged based on a two-tiered pricing structure:

- 6.319¢/kWh for energy consumption up to a threshold of 40 kWh per day multiplied by the number of days in the consumption period (first tier)
- 9.749¢/kWh for any subsequent energy consumption (second tier).

Upon enrollment in the WCO, consumers continue to be billed at the foundational Rate D. However, this mechanism incentivizes reductions in electricity consumption during times of peak demand. Consumers are notified the day before a forecasted peak demand event (CPE). For every kWh they curtail (i.e., do not consume compared to their typical energy use, or reference consumption — which will be later named L_t^{SQ}) during these CPEs, they receive a credit of 51.967¢. This strategy poses no financial risk to the consumers as their bills can only decrease, with no penalties for non-reduction in consumption during peak periods. CPEs occur between December 1 and March 31, from 6 to 9 a.m. and 4 to 8 p.m., with an estimated 25 to 33 events each winter, not exceeding a total of 100 hours (as indicated in Hydro-quebec rates).

Flex D rate (FXD)

FXD introduces another dynamic pricing structure, distinct from the base Rate D. From December to March, FXD provides reduced rates outside of peak demand events, charging:

- 4.449¢/kWh for energy consumption up to 40 kWh/day (a 30% saving)
- 7.650¢/kWh for consumption exceeding 40 kWh/day (a 22% saving).

This structure offers potential savings to consumers compared to the base rate. However, during peak demand events, the electricity rate substantially escalates to 51.967¢/kWh. Consumers are also warned a day prior to these events, encouraging them to postpone non-essential electricity consumption or to minimize overall usage. Outside of the winter

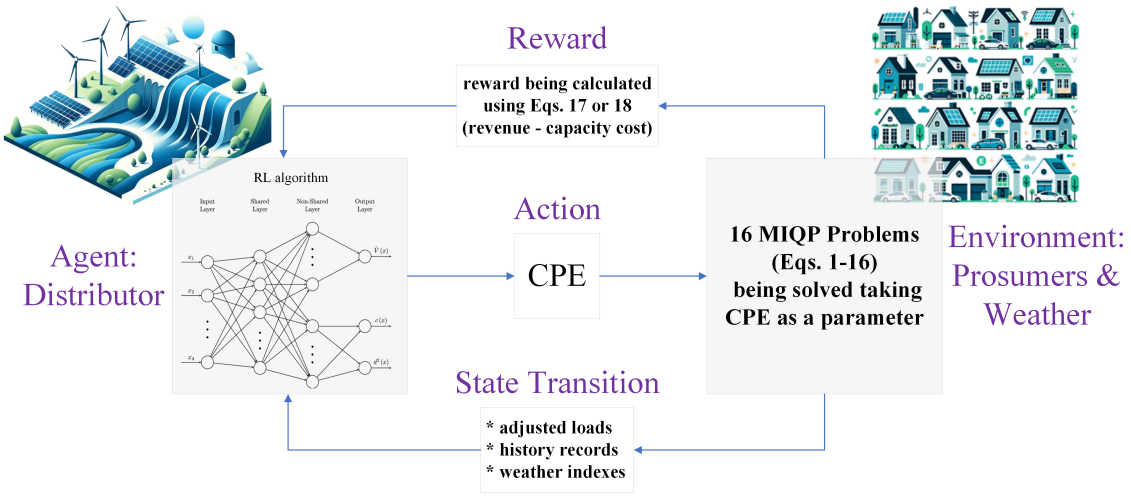

Figure 2.2: Interaction between distributor and prosumers

months, the standard Rate D charges are applicable. While FXD can yield substantial savings, it also carries a risk. If consumers fail to adjust their consumption habits during peak demand events, their bills under FXD might surpass those under the base Rate D due to the significantly higher charges during CPEs.

For this study, considering the monthly planning horizon, we impose a cap of 25 hours per month for both WCO and FXD to account for peak demand events.

Indexes and Parameters

Below are the indexes used throughout the models:

$i \in \{\mathcal{N}\}$ consumer i

$d \in \{\mathcal{D}\}$ day d in the planning month

$t \in \{\mathcal{T}\}$ episode t in peak hours $\{\mathcal{T}_{\mathcal{P}}\}$

and in off-peak hours $\{\mathcal{T}_{\mathcal{O}}\}$ during the day $\{\mathcal{T}_{\mathcal{P}} \cup \mathcal{T}_{\mathcal{O}} = \mathcal{T}\}$

And the list below introduces the parameters of the models:

$B^{max} \doteq$ capacity (kWh) of battery

$B^{min} \doteq$ minimum level of the stored electricity (kWh) in battery

$E^{max} \doteq$ capacity (kWh) of EV battery

$E^{min} \doteq$ minimum level of stored electricity (kWh) in EV

$D_t^{max} \doteq$ maximum rate (%) of load curtailment at time t

$U^c \doteq$ charging capacity (kW) of battery/EV

$U^d \doteq$ discharging capacity (kW) of battery/EV

$1 - \eta^c \doteq$ charging loss rate

$1 - \eta^d \doteq$ discharging loss rate

$A_t \doteq$ availability of EV at time t

$V_t \doteq$ EV usage (kWh) (for vehicle riding) at time t

$R_t \doteq$ electricity generation (kWh) from solar panel at time t

$L_t^C \doteq$ consumer's load (kWh) at time t

$L_t^{SQ} \doteq$ consumer's reference consumption record (kWh) at time t

$L^L \doteq$ daily load limit (kWh) to purchase at lower rate t

$P^R \doteq$ electricity **r**etail price (\$/kWh) to buy/sell from/to grid ($\leq L^L$)

$P^H \doteq$ **h**igher electricity retail price (\$/kWh) to buy/sell from/to grid ($\geq L^L$)

$P^P \doteq$ WCO's rebate/FXD's higher price (\$/kWh)

$P^D \doteq$ cost coefficient (\$/kWh^2) of curtailed load

$P^C \doteq$ cost coefficient (\$/kW) of (peak load) capacity

$P_t^M \doteq$ real-time electricity price (\$/kWh) in wholesale market at time t

$\alpha \doteq$ regularization coefficient of prosumers' remaining CPE hours

To obtain L_t^{SQ}, the prosumer's model is solved for each prosumer profile and each day by offering the basic D rate and setting $D_t^{max} = 0, \forall t$

Decision and State Variables

The decision variables below represent the electricity flows and curtailed load within the consumer's problem.

$$
\begin{aligned}
x_t^{GL} &\doteq \text{flow (kWh) from grid to load at time } t \\
x_t^{GB} &\doteq \text{flow (kWh) from grid to BAT at time } t \\
x_t^{GE} &\doteq \text{flow (kWh) from grid to EV at time } t \\
x_t^{RL} &\doteq \text{flow (kWh) from PV to load at time } t \\
x_t^{RB} &\doteq \text{flow (kWh) from PV to BAT at time } t \\
x_t^{RE} &\doteq \text{flow (kWh) from PV to EV at time } t \\
x_t^{RG} &\doteq \text{flow (kWh) from PV to grid at time } t \\
x_t^{BL} &\doteq \text{flow (kWh) from BAT to load at time } t \\
x_t^{EL} &\doteq \text{flow (kWh) from EV to load at time } t \\
x_t^{BG} &\doteq \text{flow (kWh) from BAT to grid at time } t \\
x_t^{EG} &\doteq \text{flow (kWh) from EV to grid at time } t \\
x_t^{D} &\doteq \text{deducted consumption (kWh) at time due to DR } t \\
x_t^{G} &\doteq \text{total electricity (kWh) taken from grid at time } t
\end{aligned}
$$

On the other side, the only decision variable of the distributor is choosing some time windows as CPEs. Note that the distributor does not set the load deduction price, but it just indicates which time windows are chosen as CPE, and, the prosumers take this as a parameter in their model. Moreover, the energy stored in the storage devices are defined as the state variables of the prosumers' problems and the remaining CPE hours is included in the state variables of the distributor's model:

$b_t^B \doteq$ State Of Charge (SOC = available electricity (kWh)) in BAT at time t

$b_t^E \doteq$ SOC in EV at time t

$h_{it} \doteq$ Remaining CPE hours for prosumer i at the end of the month

Prosumer's Model

The models below assume that CPE announcement is realized and reflected in P^P and then minimize the daily cost of the prosumers; each dynamic pricing offer comes with a different objective function, while the constraints are the same. A fixed daily charge (43.505¢/day) is equal for both rates, so it's been removed from the optimization problem.

WCO's objective function:

$$\min \Bigg[\sum_{t\in\mathcal{T}} P^R x_t^G + (P^H - P^R) \left(\sum_{t\in\mathcal{T}} x_t^G - L^L \right)^+ - \sum_{t\in\mathcal{T}_\mathcal{P}} P^P \left(L_t^{SQ} - x_t^G \right)^+ + P^D \sum_{t\in\mathcal{T}} (x_t^D)^2 - P^R (b_T^E + b_T^B) \Bigg] \tag{2.1}$$

FXD's objective function:

$$\min \Bigg[\sum_{t\in\mathcal{T}_\mathcal{O}} P^R x_t^G + (P^H - P^R) \left(\sum_{t\in\mathcal{T}_\mathcal{O}} x_t^G - L^L \right)^+ + \sum_{t\in\mathcal{T}_\mathcal{P}} P^P x_t^G + P^D \sum_{t\in\mathcal{T}} (x_t^D)^2 - P^R (b_T^E + b_T^B) \Bigg] \tag{2.2}$$

subject to:

$$x_t^{GL} + x_t^{GB} + x_t^{GE} - x_t^{RG} - \eta^d(x_t^{BG} + x_t^{EG}) = x_t^G \qquad \forall t \qquad (2.3)$$

$$x_t^{GL} + \eta^d(x_t^{BL} + x_t^{EL}) + x_t^{RL} + x_t^D = L_t^C \qquad \forall t \qquad (2.4)$$

$$x_t^D \leq D_t^{max} L_t^C \qquad \forall t \qquad (2.5)$$

$$x_t^{BL} + x_t^{BG} \leq b_t^B \qquad \forall t \qquad (2.6)$$

$$x_t^{EL} + x_t^{EG} \leq b_t^E A_t \qquad \forall t \qquad (2.7)$$

$$x_t^{GE} + x_t^{RE} \leq E^{max} A_t \qquad \forall t \qquad (2.8)$$

$$x_t^{BL} + x_t^{EL} + x_t^{BG} + x_t^{EG} \leq U^d \qquad \forall t \qquad (2.9)$$

$$x_t^{GB} + x_t^{GE} + x_t^{RB} + x_t^{RE} \leq U^c \qquad \forall t \qquad (2.10)$$

$$B^{min} \leq b_t^B \leq B^{max} \qquad \forall t \qquad (2.11)$$

$$E^{min} \leq b_t^E \leq E^{max} \qquad \forall t \qquad (2.12)$$

$$x_t^{RL} + x_t^{RB} + x_t^{RE} + x_t^{RG} \leq R_t \qquad \forall t \qquad (2.13)$$

$$b_t^B + \eta^c(x_t^{GB} + x_t^{RB}) - (x_t^{BL} + x_t^{BG}) = b_{t+1}^B \qquad \forall t \qquad (2.14)$$

$$b_t^E + \eta^c(x_t^{GE} + x_t^{RE}) - (x_t^{EL} + x_t^{EG}) - \frac{V_t}{\eta^d} = b_{t+1}^E \qquad \forall t \qquad (2.15)$$

$$x_t^{RL}, x_t^{RB}, x_t^{RE}, x_t^{RG}, x_t^{BG}, x_t^{EG}, x_t^{BL}, x_t^{EL},$$
$$x_t^{GL}, x_t^{GB}, x_t^{GE}, x_t^G, x_t^D, b_t^B, b_t^E \geq 0 \qquad \forall t \qquad (2.16)$$

Objective Function 2.1 consists of five terms; load covering costs (below and beyond L^L) - rebate taken from the distributor for load curtailment + dissatisfaction cost - terminal values of stored energy in batteries. Similarly, Objective Function 2.2 has five elements representing the load covering costs below and beyond L^L during off-peak and peak hours and adds dissatisfaction cost and subtracts terminal values of the stored energy in the batteries.

Incorporating a quadratic form into the objective functions means dissatisfaction costs escalate faster with larger load reductions; minor reductions trigger low dissatisfaction, but as reductions grow, dissatisfaction increases more rapidly. The coefficient P^D adjusts the dissatisfaction cost's importance in the objective function, with a higher P^D making dissatisfaction more significant and likely resulting in smaller load reductions to reduce overall dissatisfaction.

Constraint 2.3 calculates the net electricity taken from grid at each time step. Constraint 2.4 makes sure the load is covered and constraint 2.5 indicates the maximum amount of load that can be reduced. Constraints 2.6 to 2.8 state that electricity taken out of BAT and EV must be firstly less than SOC in them and secondly, electricity can be taken from or sent to EV, when it is available. Constraints 2.9 and 2.10 indicate the charging and discharging limits. Constraints 2.11 and 2.12 make sure the SOC in BAT and EV are in the allowed range of stored energy in them. Constraint 2.13 checks electricity sent from PV would be less than or equal to the generated electricity and constraints 2.14 and 2.15 calculate the SOC in BAT and EV in the next episode. Finally, constraint 2.16 specifies the domain of the decision variables. This study also considers $b_0^B = b_0^E = 0$. While some studies remove this condition and the last terms in Equations 2.1 and 2.2 and set $b_0^B = b_T^B, b_0^E = b_T^E$.

in case where a prosumer does not posses one or some tools, the following adjustment is applied:

- EV $\rightarrow E^{min} = E^{max} = 0$
- BAT $\rightarrow B^{min} = B^{max} = 0$
- PV $\rightarrow R_t = 0, \forall t$
- DR $\rightarrow D_t^{max} = 0, \forall t$

Distributor's Model

Given the stochastic nature of P_{dt}^M and x_{idt}^G parameters in the distributor's model, the objective functions under both WCO and FXD scenarios aim to maximize the expected value. For WCO, the objective 2.17 encapsulates revenue from selling electricity, compensation for consumption beyond L^L, reimbursement to prosumers, and capacity costs due to peak load. Under FXD, Equation 2.18 accounts for revenue from consumption below and above L^L during off-peak, revenue during peak hours, and monthly capacity costs.

WCO's objective function:

$$\max \mathbb{E}\Bigg[\sum_{i\in\mathcal{N}}\sum_{d\in\mathcal{D}}\Bigg\{\sum_{t\in\mathcal{T}}(P^R-P^M_{dt})x^G_{idt}+(P^H-P^R)\left(\sum_{t\in\mathcal{T}}x^G_{idt}-L^L\right)^+ - \sum_{t\in\mathcal{T}_\mathcal{P}}P^P\left(L^{SQ}_{idt}-x^G_{idt}\right)^+\Bigg\}-P^C x^{peak}\Bigg] \tag{2.17}$$

FXD's objective function:

$$\max \mathbb{E}\Big[\sum_{i\in\mathcal{N}}\sum_{d\in\mathcal{D}}\{\sum_{t\in\mathcal{T}_\mathcal{O}}(P^R-P^M_{dt})x^G_{idt}+(P^H-P^R)(\sum_{t\in\mathcal{T}_\mathcal{O}}x^G_t-L^L)^+ + \sum_{t\in\mathcal{T}_\mathcal{P}}(P^P-P^M_{dt})x^G_{idt}\}-P^C x^{peak}\Big] \tag{2.18}$$

Moreover, constraint 2.19 calculates the peak load in the planning horizon

$$\sum_{i}^{N}x^G_{idt}\leq x^{peak} \qquad \forall\{d,t\}. \tag{2.19}$$

In the context of targeted offers within a peak load shaving program, a disparity where certain prosumers receive more CPE offers than others can lead to an imbalance in revenue generation opportunities. This issue can be exacerbated when the utility company is state-owned, potentially fostering perceptions of unfairness. To address this concern and enhance the fairness of the program, this study integrates the variance of the remaining CPE hours into the distributor's objective functions, 2.17 and 2.18. This integration is represented by the following Equation:

$$-\alpha\sum_{i\in\mathcal{N}}\frac{(h_i-\bar{h})^2}{N}. \tag{2.20}$$

In this Equation, $\bar{h}$ denotes the average remaining CPE hours across all prosumers at the end of the month, and N signifies the total number of prosumers. This addition aims to quantify and minimize the disparity in CPE hour distribution, thereby fostering a more equitable environment within the peak load shaving initiative.

2.4 Methodology

This section outlines the study's methodology, highlighting the use of a primary dataset from Hydro-Quebec featuring hourly electricity consumption records for Quebec from 2013 to 2019. The study models consumer consumption to align peak and valley loads with the actual distribution, aiding in managing aggregated demand crucial for power system efficiency and stability under CPP strategies. Consumption per consumer is derived by

dividing total consumption by the consumer count. To ensure broad applicability, we adjust the Prosumer Presence Percentage (PPP) from 0.01% to 100%, affecting the mix of prosumers and conventional consumers in our analysis—for instance, a 0.01% PPP results in 16 prosumers among 160,000 individuals. Additional datasets, including PV generation, temperature, and EV usage data, with data preparation and normalization methods, are discussed in the Appendix.

The problem investigated in this study is inherently dynamic, as the decisions made in one day can influence the future state of the system. Key parameters such as temperature and energy consumption, which are part of the state space, are challenging to forecast accurately over extended periods. Conventional operations research methods often struggle to effectively handle such dynamic and uncertain environments (Vázquez-Canteli and Nagy, 2019). Moreover, due to the continuous nature of the state space approximate dynamic programming techniques may lose their applicability due to the curse of dimensionality, which exponentially increases computational and storage demands. Additionally, the reliance on function approximation introduces errors that can accumulate, affecting the quality of the derived policies. Lastly, effective exploration and generalization in these vast spaces are challenging, impacting the stability and convergence of the algorithms. On the other hand, RL is well-suited for sequential decision-making problems with dynamic and stochastic characteristics, so it is an attractive approach for this study.

Markov Decision Processes (MDPs) are the standard framework in RL, representing sequential decision-making problems as a set of states, actions, state transition probabilities, and rewards (Sutton and Barto, 2018). At each time step, the agent observes the current state and selects an action, which leads to a new state and a corresponding reward. The goal is to learn an optimal policy that maximizes the cumulative rewards over time. MDPs assume the Markov property, which states that the future state depends only on the current state and action, and not on the complete history.

We explore the application of three standard and state-of-the-art RL algorithms, each tailored for discrete action spaces. These include the Double Dueling Deep Q Network with Prioritized Experience Replay (D3QN), Soft Actor Critic with Discrete actions (SACD), and Proximal Policy Optimization with Discrete actions (PPOD), offering a diverse range of ap-

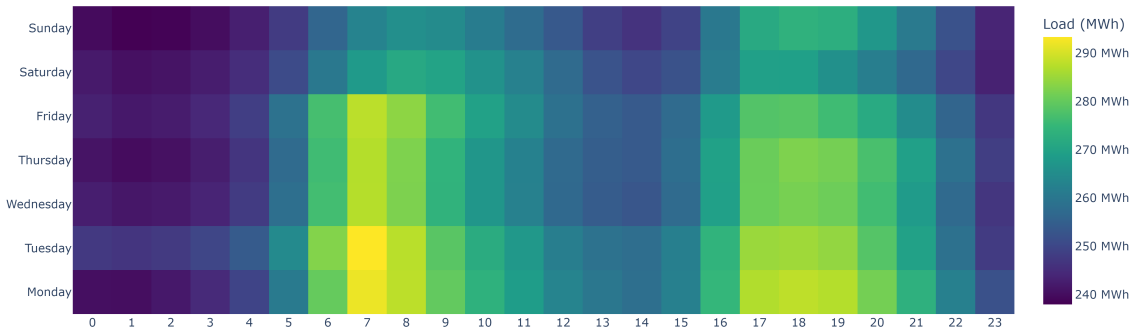

Figure 2.3: Average winter load heat map and peak loads (2013-2019)

proaches to our analytical framework. The Python code for data preprocessing and formatting, algorithm implementation, execution, and results visualization is provided alongside this chapter and can also be accessed on GitHub (To preserve the anonymity of the authors, the link to the repository has been omitted and will be made available upon request).

In this study, the distributor serves as the (RL) agent interacting with the environment (consumers, prosumers, and market). It decides on actions (CPEs) based on the current state to maximize its cumulative expected reward over time. This section defines the state and action spaces, alongside the reward function, which are central to defining this interaction.

Action Space

Figure 2.3 represents the hourly average electricity consumption for each day of the week during winter in Quebec (for 160,000 households). Two peak loads emerge, one during the early morning "work" days and the other one during the evening when people return home. Therefore, CPEs can target only these two daily time periods. The action space is a scalar that indicates whether a CPE is announced for the next day or not and if announced, what type of CPE is chosen. An action is an integer number ranging from zero to three, where a value of zero means no CPE is announced, one indicates a CPE for 6–9 am, two denotes a CPE for 4–8 pm, and three signifies two CPEs announced for both 6–9 am and 4–8 pm.

State Space

Under the mass offer scenario, the state space is a vector that consists of the next day's highest load during 6–9 am, the next day's highest load during 4–8 pm, the next day's highest load during other hours, the monthly forecasted lowest 6–9 am and 4–8 pm temperatures, the remaining CPE hours, the peak load so far, and the current day of the month.

Under the targeted scenario, the state space is a vector that includes the respective prosumer's highest consumption during 6–9 am, 4–8 pm, and other hours (obtained after prosumers' MIQP problems are solved), the lowest temperature forecast during 6–9 am, the lowest temperature forecast during 4–8 pm, remaining CPE hours for the respective prosumer and average remaining CPE hours for all prosumers, current aggregated monthly peak load so far, the current day in the month, and four binary values indicating if the respective prosumer possesses PV, EV, BAT, and DR.

To normalize the state space, consumption records are divided by 1,000, temperature forecasts are normalized using the minimum and maximum temperature forecasts in the month, the remaining CPE hours are divided by 25, and the day in the month is divided by the number of days in that month.

Reward Function

The reward function of the distributor is calculated according to Equations 2.17 and 2.18 for WCO and FXD options. Under the targeted offer with fairness Equation 2.20 is used. Also, since the distributor's monthly revenue differs for each month, rewards for each month are normalized using the distributor's revenue in status quo mode in that month.

Algorithm Operation

In the mass offer scenario, the distributor indicates the CPE type (0, 1, 2, or 3) for the next day, and this offer is the same for all prosumers. Based on this offer, 16 different Mixed-Integer Quadratic Programming (MIQP) prosumer models are solved, and the aggregated and adjusted load for the whole population is calculated for the next day. The daily net revenues/cost of the distributor and prosumers are calculated, continuing until the end of

the month, when the monthly peak load is realized, and its respective cost is subtracted from the distributor's monthly net revenues to obtain the distributor's total reward.

In the targeted offer scenario, the distributor indicates the CPE type for prosumer profile 1 for the next day. Based on this offer, prosumer 1'a MIQP model is solved, and his/her load for the next day is calculated. The distributor repeats this process for all consumers. After calculating prosumer 16's load for the next day, the distributor's net revenues for day 1 can be calculated. This process is repeated for all days in the month, and at the end of the month, when the monthly peak load is realized, its respective capacity cost is subtracted from the distributor's monthly revenue.

The action space remains the same under both mass and targeted scenarios, but the targeted scenario has 16 times more episodes than the mass offer scenario.

Next, the algorithms developed for this study are discussed. Their pseudocodes are in the Appendix.

2.4.1 D3QN: Double Dueling Deep Q Network with Prioritized Experience Replay

The D3QN algorithm integrates the Deep Q Network (DQN) with enhancements for improved efficiency in learning. DQN, introduced by Mnih et al. (2015), uses a neural network to approximate the Q-function, essential for evaluating actions in a state s based on the expected cumulative reward. The Q-function $Q(s, a)$ updates as:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t) \right]. \tag{2.21}$$

Double DQN, Dueling Network architecture, and Prioritized Experience Replay constitute the core of D3QN. Double DQN, by Van Hasselt et al. Van Hasselt et al. (2016), minimizes overestimation by using two networks for action selection and evaluation, updating Q-values as:

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha \left[R_{t+1} + \gamma Q(S_{t+1}, \arg\max_{a} Q(S_{t+1}, a; \theta); \theta') - Q(S_t, A_t) \right]. \tag{2.22}$$

The Dueling Network, proposed by Wang et al. Wang et al. (2016), splits the network into state value and advantage streams, enabling precise Q-value estimation:

$$Q(s, a; \theta, \beta, \alpha) = V(s; \theta, \beta) + \left(A(s, a; \theta, \alpha) - \frac{1}{|\mathcal{A}|} \sum_{a'} A(s, a'; \theta, \alpha) \right). \tag{2.23}$$

Prioritized Experience Replay, by Schaul et al. Schaul et al. (2015), focuses on significant experiences, measured by TD error, for efficient learning, with prioritization as:

$$P(i) = \frac{p_i^\alpha}{\sum_k p_k^\alpha}. \tag{2.24}$$

D3QN's efficacy comes from the synergy of these components, addressing overestimation, enabling precise value estimation, and optimizing learning from experiences.

2.4.2 SACD: Soft Actor Critic with Discrete Action Space

Soft Actor Critic (SAC), initially proposed by Haarnoja et al. (2018) for continuous action spaces, is adapted to discrete action spaces as SACD. SAC's effectiveness stems from balancing exploration and exploitation by maximizing expected return alongside policy entropy, promoting robust policy learning.

Actor-critic methods are a class of RL algorithms that utilize two components: an actor, which learns to select actions based on the current policy, and a critic, which evaluates the actions taken by the actor by learning the value function. This structure allows simultaneous refinement of both the policy and value estimates through continuous feedback from the critic. Compared to Q-learning approaches, Actor-critic methods integrate learning and exploration more seamlessly, allowing for nuanced policy updates and potentially faster, more reliable convergence in complex scenarios.

SACD translates continuous outputs to discrete actions through a probability distribution over possible actions, calculated via a neural network that outputs probabilities for each action, normalized with a softmax function.

Action selection methods include:

- **Deterministic Action Selection:** Choosing the action with the highest probability, suitable for evaluation phases.
- **Stochastic Action Selection:** Sampling actions based on the categorical distribution of probabilities, encouraging exploration during training.

SACD employs an actor-critic architecture and a soft policy update mechanism, enhancing efficiency in discrete action spaces:

Actor-Critic Architecture: Comprises an actor proposing actions, and a critic evaluating them, enabling effective policy updates through direct feedback.

Soft Policy Update: Incorporates entropy into the objective, preventing premature convergence by maintaining policy stochasticity. The update rule is:

$$\pi^* = \arg\max_{\pi} \sum_t \mathbb{E}_{(s_t,a_t)\sim\pi} \left[r(s_t, a_t) + \alpha \mathcal{H}(\pi(\cdot|s_t)) \right] \tag{2.25}$$

where α adjusts policy stochasticity and $\mathcal{H}$ represents entropy, ensuring a balance between exploration and exploitation in discrete action environments.

2.4.3 PPOD: Proximal Policy Optimization with Discrete Action Space

Proximal Policy Optimization (PPO) represents a significant advancement in the field of RL, particularly in policy-gradient methods. Originally developed by (Schulman et al., 2017), PPO is favored for its simplicity and stability compared to SAC, which also aims to maximize entropy. While SAC is more sample-efficient due to its off-policy nature, PPO outperforms traditional on-policy methods like Q-learning in terms of sample efficiency by reusing data across multiple updates. PPO's robust performance across diverse environments, especially in both discrete and continuous control tasks, and its straightforward implementation make it a popular choice in RL, offering a good balance between simplicity, efficiency, and efficacy.

PPOD is an adaptation of PPO for discrete action spaces. This adaptation broadens the applicability of PPO to environments and problems where actions are distinct and non-continuous. Such environments are common in various domains, including gaming and decision-making processes in discrete systems.

PPOD inherits the fundamental principles of PPO and adapts them for discrete action scenarios. The two primary components of PPOD are the policy gradient approach and the clipping technique.

Policy Gradient Approach: PPOD, like its predecessor PPO, utilizes a policy gradient method for optimizing the policy. This approach directly maximizes the expected return by adjusting the policy parameters in the direction of the gradient. The objective function is given by:

$$J(\theta) = \mathbb{E}_t \left[\min(g_t(\theta)A_t, \text{clip}(g_t(\theta), 1-\epsilon, 1+\epsilon)A_t) \right] \tag{2.26}$$

where $g_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$ is the probability ratio and A_t is the advantage at time t.

Clipping Technique: The clipping mechanism in PPOD helps maintain the updates within a reasonable range, thus preventing destructive large policy updates. The clipping function is applied to the ratio $g_t(\theta)$, ensuring that it stays within the interval $[1-\epsilon, 1+\epsilon]$. This technique stabilizes the training process by avoiding large deviations from the old policy.

2.5 Results and Discussion

This section presents the implementation results of the suggested algorithms and first introduces Figure 2.4, which illustrates the convergence patterns of various algorithms towards their best policies. These algorithms have undergone separate training processes for WCO and FXD programs and varying PPPs. Figure 2.4 reports the training process of algorithms implemented for FXD with PPP of 0.8% and the rewards are the average normalized rewards for the months in the validation set. While D3QN demonstrates the greatest variation, all algorithms eventually attain their maximal rewards which are greater than 1. The distributor's net revenues vary, occasionally yielding negative returns across different months; considering that the training dataset spans 21 months, the rewards have been normalized. A reward of 1 signifies the baseline Status Quo net revenues and any value exceeding 1 indicates an increase in the distributor's net revenues. The dataset used in this study includes 27 months; 21 months are selected for training, three months for validation, and three months for test. All results presented in this section, report the test performance (i.e., the agent's performance on the held-out test set). We will start by presenting the results of mass CPP implementation where all prosumers receive the same offer and then will move to the targeted CPP cases where profiles come into play and extend CPP effectiveness boundary.

2.5.1 Mass offer

To streamline the comparison of outputs and figures despite monthly variations in parameters and variables, all figures in this section are the PPOD implementation results on February 2017, a month from the test set.As an example of the CPP implementation effect

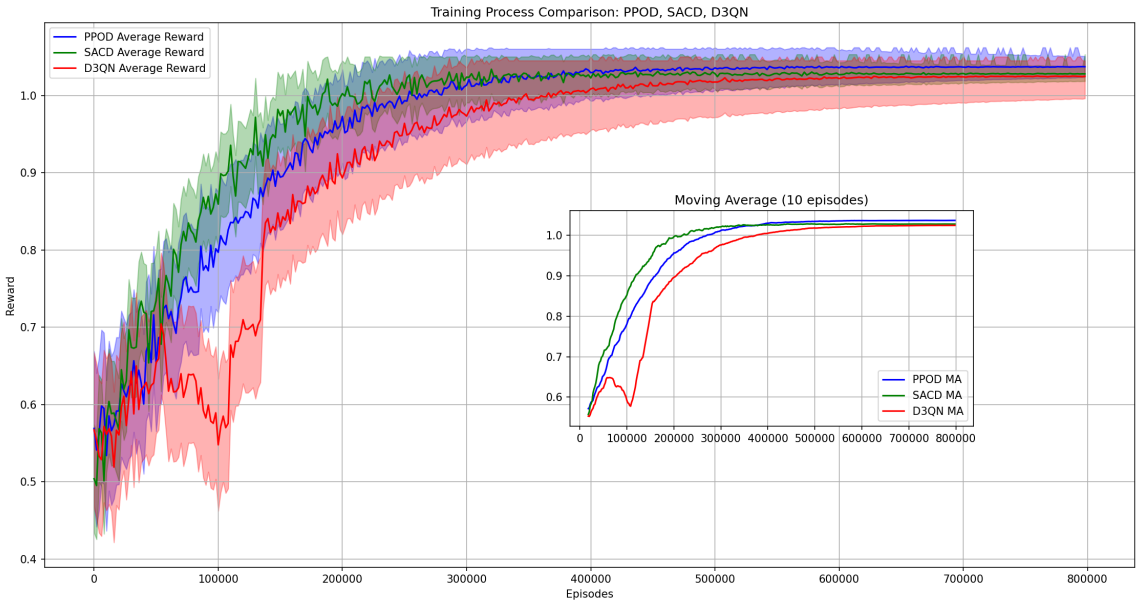

Figure 2.4: RL algorithms' training process – stability and convergence comparison

on the load cures, Figure 2.5 shows the aggregated monthly load for 160,000 households with a PPP of 1%. The blue line indicates the reference load without CPE announcements, with the dashed blue line marking the monthly peak load in this case. In February 2017, the peak load reached 335 MW, on Feb 10th between 6–9 am, as shown in Figure 2.5. After implementing PPOD, four CPEs are announced for the top four peak loads in the month. Pink vertical stripes mark the two morning peak CPEs on February 1st and 10th, and green stripes show two evening peak CPEs on February 7th and 10th, reducing the new peak load (dashed red line) to 321 MW, a 4% decrease. This reduction in peak load leads to a 4.4% rise in the distributor's monthly net revenues. Also, the red line represents the actual realized load at the end of the month.

While peak loads are reduced during designated periods (where the blue line falls below the red), there are intervals where increased activity, such as EV and BAT charging by prosumers preparing for peak times, pushes the blue line above the red. Savings also stem from PV generation and demand response load reduction, yet this off-peak increase does not lead to new peak loads, as reported in Figure 2.6. Despite the "do nothing" strategy being better (compared to RL) for PPPs over 2%, an experiment with one CPE during the highest peak shows a 19% reduction in load during previously peak period but also a 5%

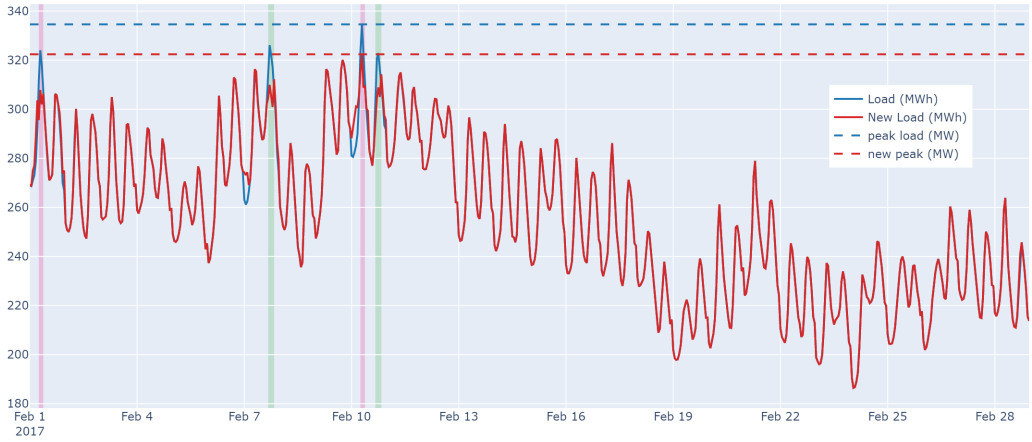

Figure 2.5: Change in monthly aggregated and peak loads by Mass offer (Option: FXD, PPP: 1%)

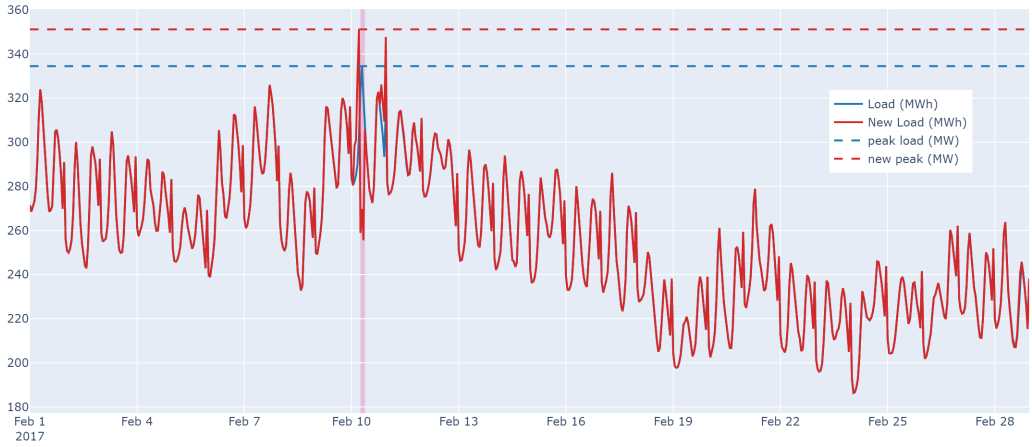

Figure 2.6: Change in monthly aggregated and peak loads by Mass offer (Option: FXD, PPP: 4%)

increase in new peaks, culminating in a 16% net revenue decrease for the distributor due to higher costs from new peaks. This isn't a limitation of the RL approach or modeling, but rather of the Mass Offer option, because under Mass Offer, the volume of the shifted load is not controllable and new peaks emerge inevitably.

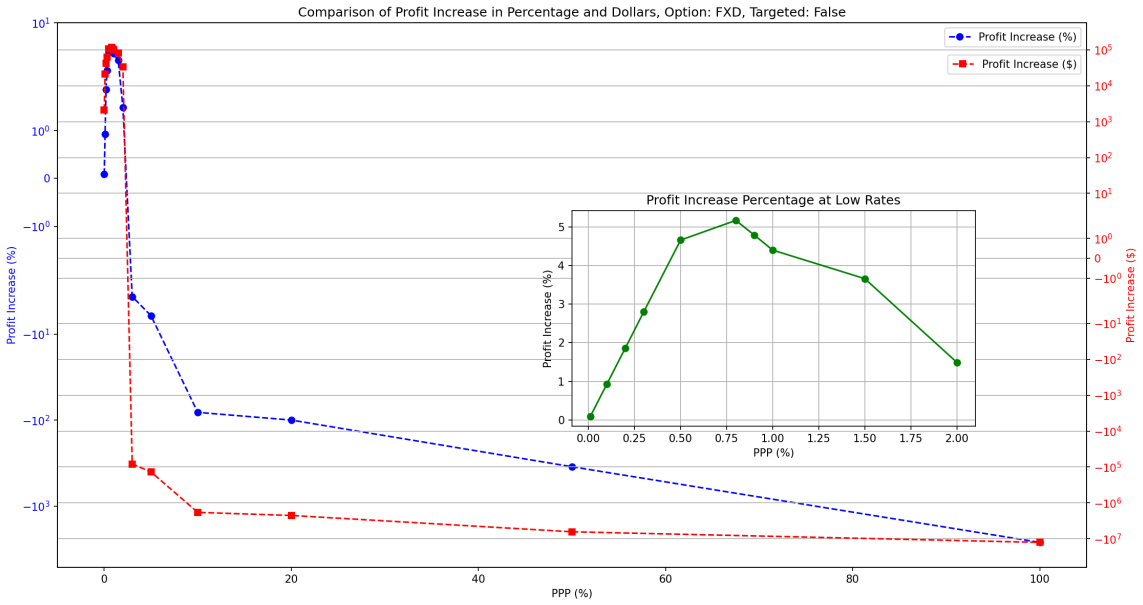

Figure 2.7: Change in distributor's net revenues with increase in PPP under FXD option

Similar outcomes are observed with the WCO option; beyond a 2% PPP, shifted loads create new peaks, diminishing mass offer effectiveness. Table 2.1 compares average test set results over three months, affirming the robustness of these findings. Figures 2.7 and 2.8 illustrate net revenue changes with rising PPP. Both WCO and FXD show net revenue growth with PPP increase up to a point, detailed further for low PPP values in embedded figures, indicating maximal distributor motivation around 1% PPP. Beyond this, revenue increases wane as more prosumers participate, though CPEs can boost net revenues up to 2% PPP. These analyses use the distributor's standard monthly billing as a baseline, showing that while WCO maintains status quo revenues, FXD rate discounts for off-peak prosumers progressively diminish net revenues as PPP increases.

Figures 2.7 and 2.8 analyze distributor net revenue changes with PPP increases under FXD and WCO schemes, whereas Figure 2.9 presents total net revenues in Canadian dollars for 160,000 households at various PPP levels (Points are connected with dashed lines for easier visualization and do not imply generalization). Net revenues initially rise with PPP, peaking at around 1%, then decline. With WCO, net revenue decreases but stays positive, whereas FXD results in negative revenues beyond a 20% PPP. Despite this, the comparison between WCO and FXD in Figure 2.9 shows WCO's superiority for most PPP levels, with

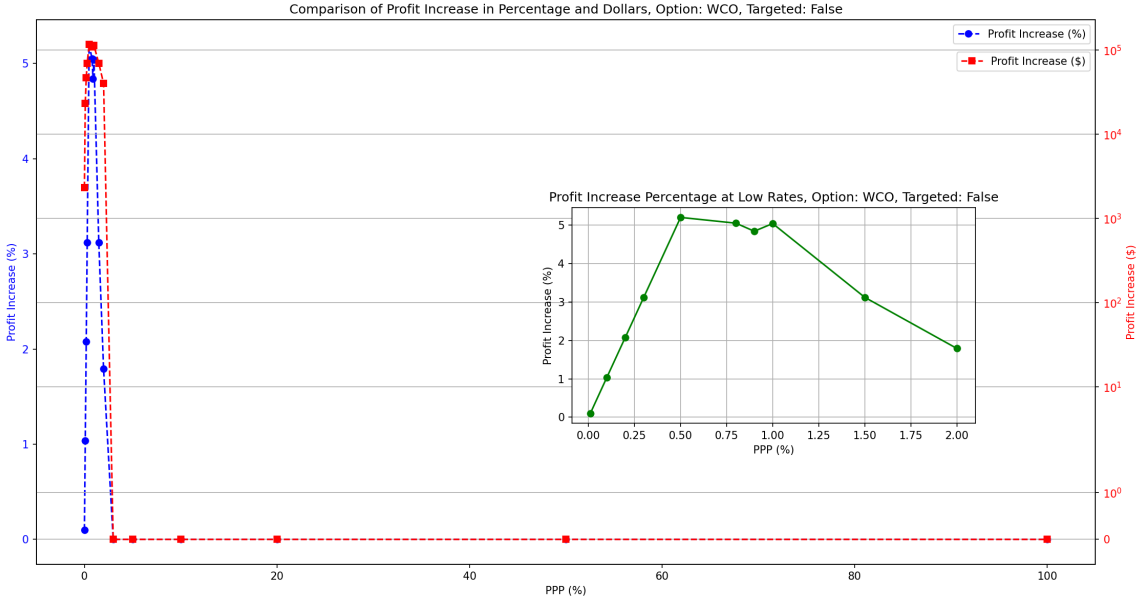

Figure 2.8: Change in distributor's net revenues with increase in PPP under WCO option

Table 2.1: Mass offer results

	WCO		FXD	
PPP(%)	mean net revenue increase(%)	reduced peak load(MW)	mean net revenue increase(%)	reduced peak load(MW)
0.01	0.06	0.164	0.06	0.162
0.10	0.66	1.641	0.63	1.618
0.20	1.26	3.212	1.17	2.813
0.30	1.72	4.923	1.72	4.853
0.50	2.80	8.204	2.84	8.089
0.80	2.65	8.415	3.12	10.597
0.90	2.42	9.044	2.79	9.597
1.00	2.32	9.972	2.47	9.360
1.50	1.08	7.029	1.45	7.341
2.00	0.66	3.926	0.61	5.867

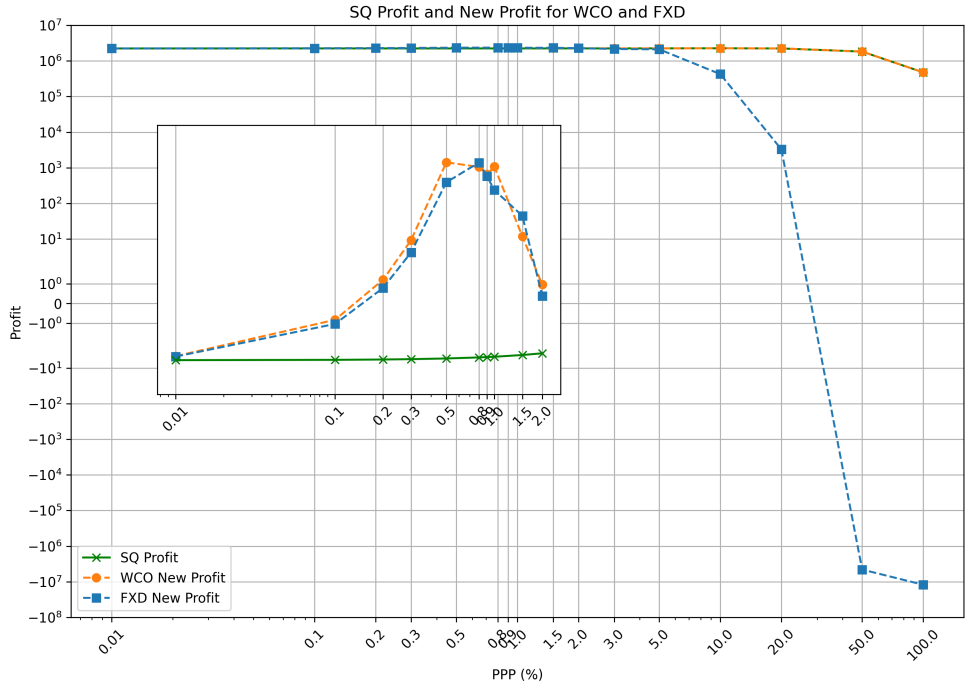

Figure 2.9: Distributor's net revenue comparison under two options

FXD only outperforming at 0.8% and 1.5% PPPs, as depicted in the embedded figure.

Figure 2.10 depicts the monthly billing totalities for each of the sixteen profiles. The subsequent analyses employ data from an instance of the problem with 1% PPP and WCO mass offering. The figure also reports the comparison between the two options regarding their respective impacts on prosumer expenditures. The results exemplify that WCO not only constitutes a risk-free selection for prosumers (since in the absence of peak load shifting responses, their monthly invoice would not become inflated, whereas FXD program can lead to spikes in costs for peak hours without proper reactions), but additionally, suitably responsive prosumers can attain superior financial outcomes under WCO compared to FXD alternative.

Figure 2.11 shows peak hour consumption reductions for technology combinations (PV, EV, BAT, DR), based on their operational status (*True* or *False*). For the *True* scenario, savings are computed when technologies are present and active. For the *False* scenario,

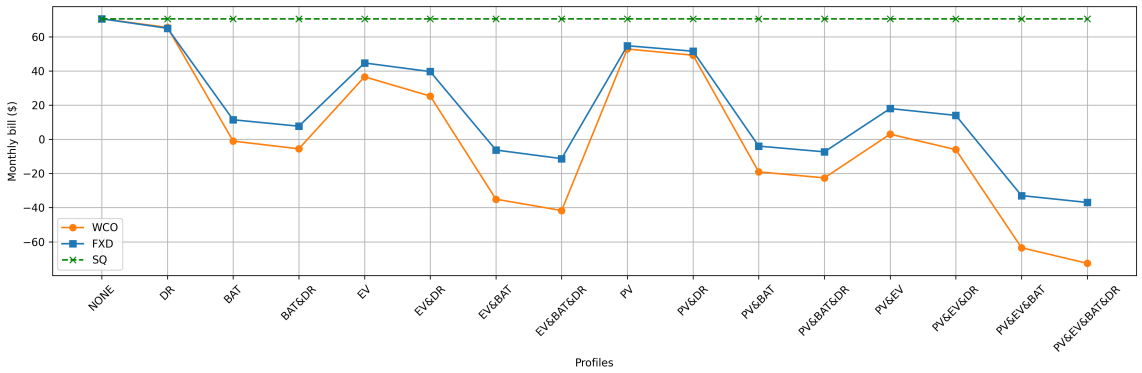

Figure 2.10: Monthly bills of profiles under mass offer (Option: WCO, PPP: 1%)

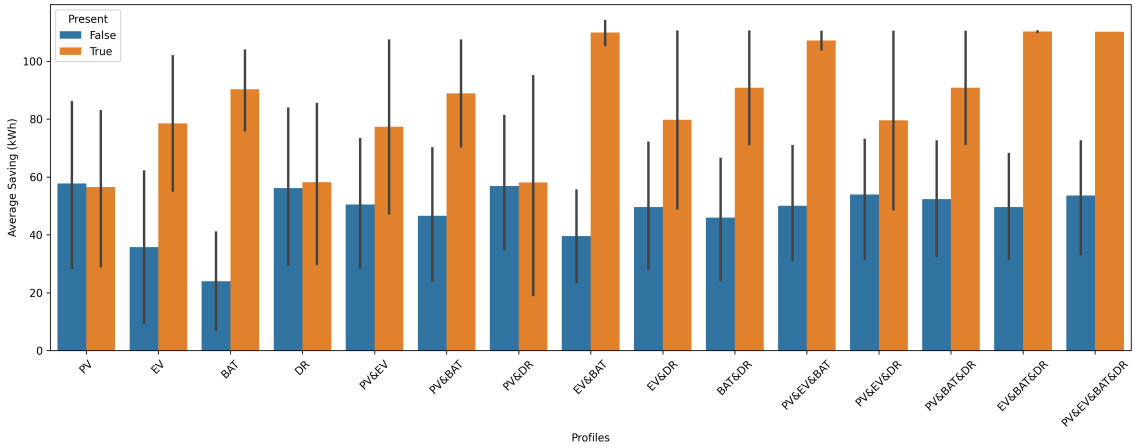

Figure 2.11: Peak load shaving power of each profile under mass offer (Option: WCO, PPP: 1%)

savings are determined when technologies are absent or inactive:

$$S_{\text{True/False}} = \frac{1}{N_{\text{True/False}}} \sum_{i=1}^{N_{\text{True/False}}} s_{i,\text{True/False}}$$

where S_{True} and S_{False} represent average savings with technologies present or absent. N_{True} and N_{False} denote the number of instances, and $s_{i,\text{True}}$ and $s_{i,\text{False}}$ are savings for the ith instance. The data supports that EV and BAT have the most substantial impact, with BAT being more influential due to higher capacity. DR and PV offer only slight improvements.

2.5.2 Targeted offer

As shown in Figure 2.6, upon exceeding a PPP threshold under mass offer program, even one CPE announcement can constitute new peak loads during previously off-peak intervals.

Table 2.2: Targeted offer results (without fairness consideration)

	WCO			FXD		
PPP(%)	mean net revenue increase(%)	mean reduced peak load(MW)	mean disparity score	mean net revenue increase(%)	mean reduced peak load(MW)	mean disparity score
1	3.27	12.391	1.91	4.87	10.194	1.16
2	3.06	16.602	3.37	5.26	18.591	1.46
3	2.68	17.199	2.72	6.21	15.088	1.35
5	1.04	9.479	3.00	1.35	3.357	0.13
10	1.73	4.806	0.89	-	-	-
20	1.66	7.615	0.84	-	-	-
50	0.75	3.178	0.17	-	-	-
100	-	-	-	-	-	-

Therefore, managing the magnitude of shifted loads by targeting specific prosumer respondents becomes imperative. Figures 2.12 and 2.13 represent targeted CPE events for FXD and WCO options at 3% and 5% PPPs, respectively. Darker vertical bands indicate more CPE recipients in that timeframe. In Figure 2.12, most events take place during February 7th evenings and February 10th mornings. Similarly, Figure 2.13 concentrates CPEs on February 10th morning and evening intervals. Specifically, for FXD case, the algorithm utilizes CPE strategy of {1:{5:1, 2:1, 16:1}, 7:{2:2, 4:2, 8:2}, 9:{4:2, 10:2}, 10:{14:3, 1:1, 3:3, 6:3, 7:2, 10:3, 11:1, 12:2, 9:3, 13:1, 16:3}} while Figure 2.13 employs {1:{6:1}, 7:{15:3}, 10:{12:1, 7:1, 8:1, 2:3, 16:3}}. In the given coding, "1:{5:1, 2:1, 16:1}" represents that on day 1, profiles 5, 2, and 16 each received a morning CPE announcement. By restricting audiences, newly forming peaks become governable, instituting consistently beneficial programs for both prosumers and distributor via targeting all profiles, yet on different days.

Table 2.2 presents the outcomes of using PPOD (the highest-performing algorithm on our validation data), covering WCO and FXD results. The table details how the mean net revenue of the distributor increases with the implementation of a targeted offering approach, displayed in the first column. In this table, the distributor's objective function does not factor in the disparity score (Equation 2.20). A lower disparity score indicates a more equitable distribution of CPE offers among prosumers, presented in the third column. Additionally, the table shows the average reduction in peak load over three months in the test dataset, listed in the second column.

To tackle the issue of disparity, Equation 2.20 is incorporated into the distributor's objective function in RL algorithms, with updated findings shown in Table 2.3. Comparing

Table 2.3: Targeted offer results (with fairness consideration)

	WCO			FXD		
PPP(%)	mean net revenue increase(%)	mean reduced peak load(MW)	mean disparity score	mean net revenue increase(%)	mean reduced peak load(MW)	mean disparity score
1	3.07	13.275	1.15	4.77	10.139	0.71
2	2.96	16.624	2.12	4.81	18.006	1.26
3	2.60	16.740	1.50	5.96	15.643	0.90
5	1.02	10.102	1.79	1.23	3.295	0.13
10	1.61	4.815	0.81	-	-	-
20	1.57	7.309	0.73	-	-	-
50	0.69	3.100	0.15	-	-	-
100	-	-	-	-	-	-

the data from two tables shows a lower disparity score, indicating a more uniform distribution of CPE offers, leads to a decrease in both the average net revenue increase for the distributor and the reduction in peak load. This outcome highlights a trade-off between minimizing disparity and maximizing net revenue. This balance can be adjusted through the coefficient α. this result indicates that including the disparity term in the calculation lowers the disparity score by approximately 46%, while the mean net revenue increase sees a reduction of about 5%.

Additionally, Table 2.3 provides two insights — targeted CPEs further profitability beyond mass offerings' limits, with FXD reaching 5% PPP and WCO remaining productive until 50% PPP. Mass offerings increased the distributor's net revenue by 5.04% and 1.79% for a 1% PPP, and by 4.40% and 1.49% for a 2% PPP under the WCO and FXD programs, respectively. However, targeted CPEs further enhanced these gains, boosting revenue by 7.05% and 8.00% for a 1% PPP, and by 7.26% and 5.92% for a 2% PPP for WCO and FXD programs, respectively. Secondly, except 1% PPP, WCO dominates FXD for all other values of PPP.

Figure 2.14 illustrates the scenario when PPP reaches a magnitude where existing options and programs are ineffective. In such instances, introducing new devices alters the historically stable load patterns, leading to new peak loads. These peak loads manifest during periods that are not only outside of traditionally recognized peak times but are also in varying time and more challenging to predict. Additionally, the increased demand on the grid to accommodate EV and BAT charging has significantly altered load patterns. Notably, new peak loads are observed between 2 AM and 6 AM, when current programs fail to ad-

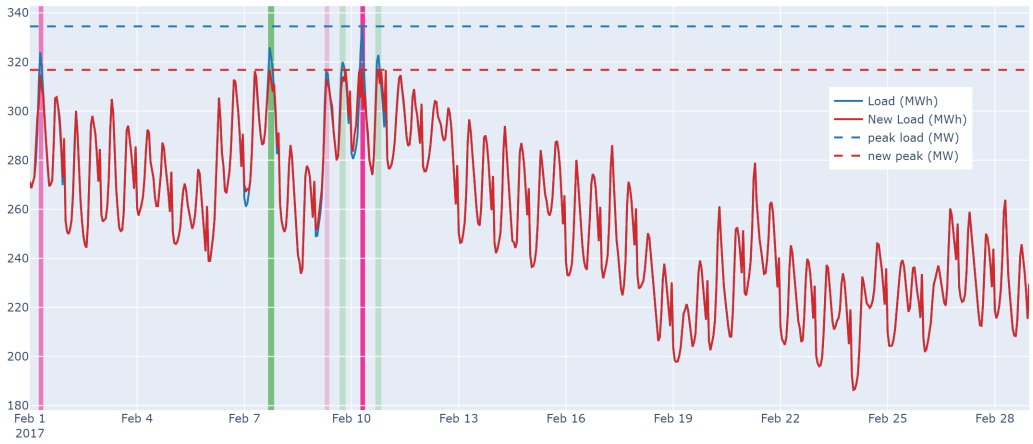

Figure 2.12: Change in monthly aggregated and peak loads by Targeted FXD offer (PPP: 3%)

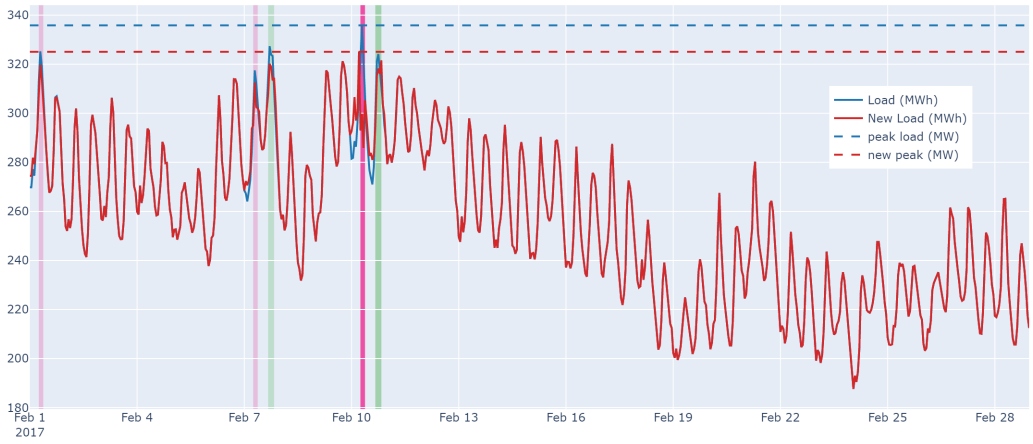

Figure 2.13: Change in monthly aggregated and peak loads by Targeted WCO offer (PPP: 5%)

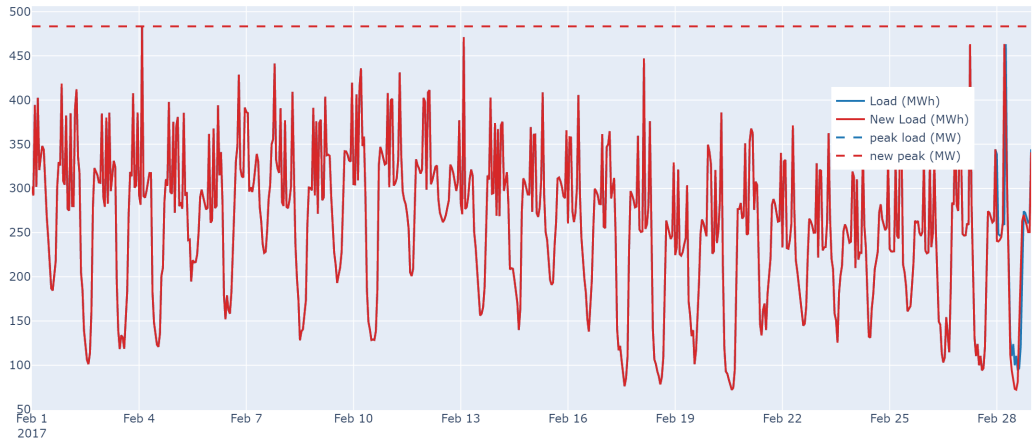

Figure 2.14: Change in monthly aggregated and peak loads by Targeted WCO offer (PPP: 100%)

dress them effectively. This observation underscores the need for a transition towards more dynamic peak load management strategies, such as dynamic CPPs where announcement periods are flexible. The proposed methodology in this chapter with minor adjustments would be able to address dynamic CPPs.

2.6 Conclusion and Future Directions

This chapter presented strategies to advance grid efficiency and peak load reduction through optimized dynamic pricing programs informed by prosumer analytics. The analysis focused on CPP, evaluating impacts under a reward-based (WCO) and a penalty-based (FXD) variations of CPP programs.

The key findings demonstrate that CPP events, when appropriately targeted, can incentivize significant peak load reductions from prosumer populations equipped with DERs and DR. However, mass program offerings become ineffective beyond low PPPs around 1-2%, as load shifting creates problematic new peaks. This motivates a transition towards more selective, differentiated incentives.

Accordingly, targeted dynamic pricing was introduced, restricting CPEs to subsets of

prosumers and guiding controlled, predictable DR. RL-based algorithms determined near-optimal policies for event scheduling, significantly extending profitable program viability to 3–5 times higher prosumer adoption. Moreover, the assessment of technology configurations found batteries and EVs as principal contributors, while PV provided minimal direct peak reduction. Furthermore, to address the fairness issues that arise when some profiles receive relatively more offers, the algorithms are adjusted to consider disparity and distribute the offers more uniformly among the prosumers with a reduction in disparity score by 33.82%.

In conclusion, advanced prosumer analytics and intelligent incentives coordination unlock the flexibility promised by distributed resources to shave peaks and enhance system efficiency. However, this study has some shortcomings and simplifying assumptions that provide avenues for future investigation to enhance realism: Firstly, the analysis assumed all prosumers opt-in to either the WCO or FXD program, but in practice, heterogeneous adoption across rate structures would affect aggregated load profiles, and analyzing mixed program enrollment could offer further insights. Moreover, complete prosumer responsiveness to CPP signals was assumed, but considering partial passive response would better match real-world behavior, which could significantly impact peak reduction viability. Additionally, despite political difficulties, exploring real-time pricing in Quebec would provide valuable understanding, given comparable situations elsewhere. Multi-agent RL has the potential to make prosumer decisions more dynamic beyond fixed best response functions. Furthermore, a single cost-focused objective was defined for prosumers here. However, expanding for diverse preferences like simplicity over profit or risk aversion would offer realistic heterogeneity, as consumer aspirations likely affect technology adoption and price signal reactions.

The study shows that WCO generally outperforms FXD in reducing peak usage in the short term (though in a few cases, FXD performs better), but problems can arise over time due to the way WCO adjusts the baseline for calculating rebates: if people use less power during peak times in their first year with WCO, their baseline for getting rebates will be lower the next year, which might discourage them from continuing in the program unless they sometimes use more power during peak times. Users might also purposely use more power during peak times when they are not in WCO, so they have a higher baseline and get

bigger rebates later. Studying these long-term effects and the potential for cheating in the system can help create stronger policies, and adding features to WCO to fix these issues is worth looking into, as the best solution likely finds a middle ground between WCO's careful management of peak demand and FXD's long-term rewards.

Moreover, this research uses the specific consumption patterns in Quebec, where peak loads occur during the winter and within particular time frames, so it would be beneficial to extend the model to accommodate various load patterns, such as those experienced during the summer months in the United States, thereby enhancing its applicability across different geographical and climatic conditions. Furthermore, the current model assumes a uniform distribution of prosumer profiles, which may not accurately reflect the reality where certain profiles are more prevalent, so investigating the actual distribution of these profiles or examining the impact of varying distributions on the study's outcomes could open new avenues for research. Moreover, in the realm of prosumer behaviour modeling, transitioning from an optimization-based approach to developing a predictive model grounded in historical data on prosumer responses to price signals could offer more dynamic and realistic insights into prosumer behavior, which could further refine the strategies for peak load management and the effectiveness of dynamic pricing programs.

Finally, assessing environmental sustainability impacts of peak load shaving would provide crucial perspective as attention increases on strategic decisions' planetary consequences, as quantifying emissions reduction potential could motivate additional programs on climate grounds. Overall, relaxing simplifying assumptions through multi-disciplinary expansions offers significant potential to enhance model realism and policy insights on coordinating distributed flexibility for efficient, renewable electricity systems.

2.7 Appendix

2.7.1 Data Preparation

The electricity consumption data obtained from Hydro-Quebec spans over seven years from January 2013 to March 2019, encapsulating the winter months of January, February, March, and December Hydro-Quebec (2024b). These months are of particular interest as they

coincide with the application of the dynamic pricing program and are also the periods during which yearly peak loads are observed. Given the nature of the region's climate, with extreme cold temperatures necessitating continuous operation of electric heating systems and an increased use of lighting and electrical appliances, there is a substantial increase in electricity demand. The initial consumption dataset was recorded on 15 minutes time steps and then the dataset is transformed to an hourly resolution to facilitate a more granular analysis of consumption patterns.

For the calculation of Hydro-Quebec's net revenues, the wholesale electricity price in the New York Independent System Operator (NYISO) energy market is used as a proxy for the generation cost. This approximation also serves as an alternative price, representing the potential revenue that could be generated from selling the electricity elsewhere. The price data is retrieved from the NYISO energy market & operational data system.[2]

The generation data for the PV solar panels, alongside temperature records, are acquired from the Photovoltaic Geographical Information System (PVGIS) provided by the European Commission's Joint Research Centre website.[3]. The specific details for the PV setup include the use of Copper Indium Selenide technology, an installed peak PV power of 2 kWp, and a system loss of 14%. The solar radiation data is sourced from the PVGIS-ERA5 database. The geographic coordinates for the installation are at Latitude 45.551, Longitude -73.602, and it employs a two-axis mounting type. The data retrieved from PVGIS is subsequently formatted to match the temporal resolution of the electricity consumption dataset.

The study incorporates the usage of Electric Vehicles (EVs) during working days, between 8:45 AM and 5:15 PM. On days when the EV is utilized, it consumes a total of 10 kWh/day, and it requires charging for the subsequent working day. The battery parameters for the EV include a maximum energy storage capacity of $E^{max} = 60$ kWh, a minimum energy threshold of $E^{min} = 15$ kWh, and maximum charging/discharging rates of $U^d = U^c = 10$ kW. The charging/discharging efficiency rates are both $\eta^c = \eta^d = 0.95$.

Additional parameters and consumer profiles are defined to aid in the modeling and analysis of the electricity market and consumption behaviors. These include a Dissatisfaction Coefficient for the Curtailed Load set at $P^D = 1$, a total of $T = 24$ hours in a day, and a

[2]https://www.nyiso.com/custom-reports
[3]https://re.jrc.ec.europa.eu/pvg_tools/en/toolshtml

Daily Load Limit for Purchasing at a Lower Rate of $L^L = 40$ kWh. The Cost Coefficient of Peak Load Capacity is $P^C = \$15.75$/kW. The BAT capacity is specified as $B^{max} = 60$ kWh, with a Minimum Level of Stored Electricity in the BAT at $B^{min} = 0$ kWh, and a Maximum Rate of Load Curtailment at $D^{max} = 25\%$. The consumer's reference consumption record (L^{SQ}) is derived by solving the prosumer's model for each profile, assuming participation in the basic D rate tariff, thereby enabling the calculation of status quo consumption values.

2.7.2 Pseudocodes of the RL algorithms

The pseudocode representations of the RL algorithms discussed in this manuscript are provided herein.

Algorithm 1 SACD: Soft Actor-Critic with Discrete Action Space

Initialize replay buffer $\mathcal{D}$
Initialize policy network π_ϕ with weights ϕ
Initialize two Q-value networks Q_θ^1 and Q_θ^2 with weights θ^1 and θ^2
Initialize target Q-value networks $\bar{Q}_{\bar{\theta}}^1$ and $\bar{Q}_{\bar{\theta}}^2$ with weights $\bar{\theta}^1 \leftarrow \theta^1$, $\bar{\theta}^2 \leftarrow \theta^2$
for episode $= 1, \ldots, M$ **do**
    Initialize sequence $s_1 = \{s_1\}$
    for $t = 1, \ldots, T$ **do**
        Sample an action $a_t \sim \pi_\phi(\cdot|s_t)$ from the current policy
        Execute action a_t and observe reward r_t and next state s_{t+1}
        Store transition (s_t, a_t, r_t, s_{t+1}) in $\mathcal{D}$
        Sample a minibatch of transitions (s_j, a_j, r_j, s_{j+1}) from $\mathcal{D}$
        Compute target values $y_j = r_j + \gamma \bar{Q}_{\bar{\theta}}(s_{j+1}, \pi_\phi(s_{j+1}))$
        Update Q-value networks by minimizing the loss:
        $\mathcal{L}_Q(\theta^1, \theta^2) = \mathbb{E}_{(s_j,a_j)\sim\mathcal{D}} \left[\frac{1}{2} \left(Q_{\theta^1}(s_j, a_j) - y_j \right)^2 + \frac{1}{2} \left(Q_{\theta^2}(s_j, a_j) - y_j \right)^2 \right]$
        Update policy network by maximizing the expected return and entropy:
        $\mathcal{L}_\pi(\phi) = \mathbb{E}_{s_j\sim\mathcal{D}} \left[\log \pi_\phi(a_j|s_j) - \alpha \log \pi_\phi(a_j|s_j) Q_{\theta^1}(s_j, a_j) \right]$
        Update target Q-value networks:
        $\bar{\theta}^1 \leftarrow \rho\bar{\theta}^1 + (1-\rho)\theta^1$
        $\bar{\theta}^2 \leftarrow \rho\bar{\theta}^2 + (1-\rho)\theta^2$
    end for
end for

Algorithm 2 D3QN: Double Dueling Deep Q Network with Prioritized Experience Replay

Initialize replay buffer $\mathcal{D}$
Initialize action-value function Q with random weights θ
Initialize target action-value function Q' with weights $\theta' \leftarrow \theta$
for episode $= 1, \ldots, M$ **do**
    Initialize sequence $s_1 = \{s_1\}$ and preprocessed sequence $\phi_1 = \phi(s_1)$
    for $t = 1, \ldots, T$ **do**
        With probability ϵ select a random action a_t
        Otherwise, select $a_t = \arg\max_a Q(\phi(s_t), a; \theta)$
        Execute action a_t and observe reward r_t and next state s_{t+1}
        Set $\phi_{t+1} = \phi(s_{t+1})$ and add transition $(\phi_t, a_t, r_t, \phi_{t+1})$ to $\mathcal{D}$
        Sample random minibatch of transitions $(\phi_j, a_j, r_j, \phi_{j+1})$ from $\mathcal{D}$
        Compute importance sampling weights w_j for each transition
        Set $y_j = r_j + \gamma Q'(\phi_{j+1}, \arg\max_a Q(\phi_{j+1}, a; \theta); \theta')$
        Perform a gradient descent step on $(y_j - Q(\phi_j, a_j; \theta))^2$ with weights w_j
        Every C steps, reset $Q' = Q$
    end for
end for

Algorithm 3 PPOD: Proximal Policy Optimization with Discrete Action Space

Initialize policy network π_θ with weights θ
Initialize value function network V_ϕ with weights ϕ
for iteration $= 1, \ldots, N$ **do**
    Collect a batch of transitions $\mathcal{D} = \{(s_t, a_t, r_t, s_{t+1})\}$ by running the current policy π_θ
    for epoch $= 1, \ldots, K$ **do**
        Compute advantage estimates $\hat{A}_t$ using $\mathcal{D}$ and V_ϕ
        Compute the policy ratio $g_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}$
        Update the policy network by maximizing the clipped surrogate objective:
        $\theta \leftarrow \arg\max_\theta \frac{1}{|\mathcal{D}|} \sum_t \min\left(g_t(\theta)\hat{A}_t, \text{clip}(g_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t\right)$
        Update the value function network by minimizing the mean squared error:
        $\phi \leftarrow \arg\min_\phi \frac{1}{|\mathcal{D}|} \sum_t \left(V_\phi(s_t) - V_{\text{target}}(s_t)\right)^2$
    end for
end for

2.7.3 Detailed results of CPE announcements in February 2017

Table 2.4 exhibits the results of implementing mass offer announcement for different levels of PPP in February 2017. 'SQ ($)' column shows the reference net revenue of the distributor where no offer is announced and consumers receive basic D rate. 'net revenue ($)' column represent the new net revenue gained under respected program and 'inc(%)' column

Table 2.4: Detailed results of mass offer announcement in February 2017

		WCO				FXD				
PPP(%)	SQ ($)	net revenues($)	acts	inc(%)	rpl(MW)	net revenues($)	acts	inc(%)	inc'(%)	rpl(MW)
0.01	2,239,743	2,242,074	{10:1}	0.10	0.164	2,241,831	{10:1}	0.09	0.11	0.162
0.10	2,239,923	2,263,236	{10:1}	1.04	1.642	2,260,809	{10:1}	0.93	1.06	1.618
0.20	2,240,123	2,286,749	{10:1}	2.08	3.283	2,281,895	{10:1}	1.86	2.11	3.235
0.30	2,240,323	2,310,262	{10:1}	3.12	4.925	2,302,980	{10:1}	2.80	3.18	4.853
0.50	2,240,723	2,357,288	{10:1}	5.20	8.208	2,345,152	{10:1}	4.66	5.30	8.089
0.80	2,241,323	2,354,466	{10:1}	5.05	8.474	2,357,120	{10:1,7:2,1:1}	5.17	6.20	11.688
0.90	2,241,523	2,349,957	{10:1}	4.84	8.337	2,348,784	{10:1,7:2,1:1}	4.79	5.95	11.688
1.00	2,241,723	2,354,710	{10:3,7:2,1:1}	5.04	13.691	2,340,448	{10:3,7:2,1:1}	4.40	5.69	12.177
1.50	2,242,724	2,312,614	{10:3,7:2,1:1}	3.12	14.213	2,324,904	{10:3,7:2,1:1}	3.66	5.60	14.496
2.00	2,243,724	2,283,923	{10:3}	1.79	8.478	2,277,197	{10:3,7:2,1:1}	1.49	4.03	14.496
3.00	2,245,722	2,245,722	{}	-	-	2,163,581	{}	-3.66	-	-
5.00	2,249,726	2,249,726	{}	-	-	2,112,824	{}	-6.09	-	-
10.00	2,259,729	2,259,729	{}	-	-	1,572,868	{}	-30.40	-	-
20.00	2,239,928	2,239,928	{}	-	-	3,282	{}	-99.85	-	-
50.00	1,827,270	1,827,270	{}	-	-	-4,538,111	{}	-348.35	-	-
100.00	476,396	476,396	{}	-	-	-12,107,100	{}	-2641.40	-	-

Table 2.5: Detailed results of targeted offer for February 2017

		WCO			FXD				WCO - FXD
PPP(%)	SQ ($)	net revenue($)	inc(%)	rpk(MW)	net revenue($)	inc(%)	inc'(%)	rpl(MW)	net revenue
1.00	2,241,723	2,399,831	7.05	13.277	2,404,376	7.26	8.58	10.133	-4,546
2.00	2,243,724	2,423,312	8.00	16.625	2,413,862	7.58	10.27	18.057	9,450
3.00	2,245,722	2,403,606	7.03	16.939	2,378,699	5.92	9.94	17.711	24,908
5.00	2,249,726	2,316,359	2.96	10.802	2,191,094	-2.61	3.70	5.776	125,265
10.00	2,259,729	2,365,342	4.67	7.736	1,572,868	-30.40	-	-	792,474
20.00	2,239,928	2,338,222	4.39	10.242	3,282	-99.85	-	-	2,334,941
50.00	1,827,270	1,864,017	2.01	7.853	-4,538,111	-348.35	-	-	6,402,129
100.00	476,396	476,396	-	-	-12,107,100	-2641.40	-	-	12,583,495

calculates the increase in net revenue. The "acts" columns denote the optimal CPE announcements at various PPP levels, where for example, 10:3, 7:2, 1:1 signifies two morning CPEs on February 1st and 10th plus two evening CPEs on 7th and 10th for that PPP. Furthermore, the "inc' %" column under the FXD rate structure represents net revenue differences between the listed strategy and taking 'no action' under FXD rates. Finally, 'rpl(kW)' portrays the reduced peak load in kW. Also, Table 2.5 represents the detailed results implementing the targeted options in Febraury 2017 without considering fairness term in the objective function.

2.7.4 Grid Profitability and Prosumer Savings

The rationale behind the differences in net revenue changes between the WCO and FXD options, particularly regarding Quebec's grid profitability concerns, is further examined in Figure 2.15. This figure contrasts Hydro-Quebec's wholesale electricity prices (blue dots)

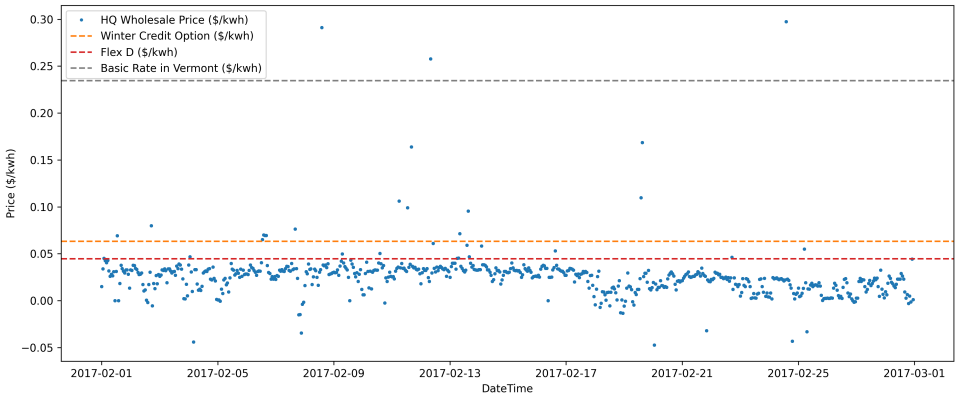

Figure 2.15: Rates comparison

with Vermont's residential uniform rate (grey dashed line) and the base rate for WCO and FXD. It underscores Quebec's considerably lower rates, which are among the lowest in North America.[4] This comparison suggests potential profitability challenges for Quebec's grid, attributed to the minimal gap between retail prices and generation costs.

Moreover, Figure 2.16 exhibits the monthly financial savings attained by each profile. The results demonstrate that profiles consisting of both EV and BAT, comprising 25% of the prosumer demographic, accrued approximately 50% of the aggregate financial savings realized by the prosumers. Additionally, profiles containing either EV or BAT (but not both) obtained 7% of the total savings. Furthermore, winter declines in PV generation quantities, aligned with peak PV production timeframes existing external to the morning and evening consumption peak periods, explains negligible savings gained through PV ownership. Moreover, in numerous cases, augmenting DR contributes only marginally to improving a profile's monthly financial savings.

2.7.5 Heuristic Algorithms

RL algorithms consistently outperformed the 'do nothing' approach in numerous instances of this problem. To thoroughly assess the effectiveness of our developed RL algorithms, we incorporated heuristic algorithms as an additional point of comparison beyond the simple 'do nothing' strategy. Our implementation encompassed several concepts, outlined as follows:

[4]https://www.hydroquebec.com/data/documents-donnees/pdf/comparison-electricity-prices.pdf

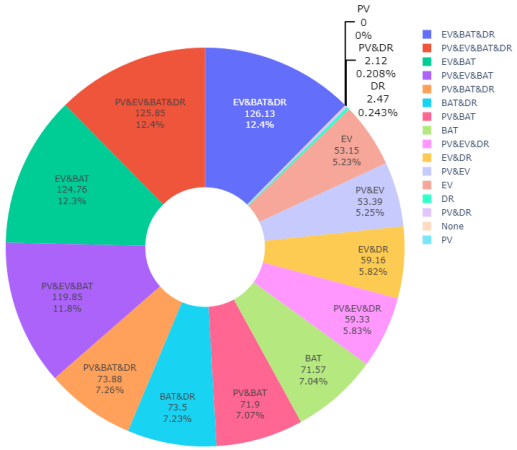

Figure 2.16: Monthly saving gained by each profile under mass offer (Option: WCO, PPP: 1%)

1. We aggregated all consumption data within the training and validation datasets to compute their mean μ and standard deviation σ. The heuristic algorithm then applies a threshold, determined by the formula $x = \mu + \theta\sigma$, to announce a CPE if the consumption value during peak hours surpasses this threshold. By adjusting θ between 2.7 and 4.3 in increments of 30 steps, we achieved a 0.36% performance enhancement in the training dataset as the highest score among all θ values.

2. A second heuristic was developed based on temperature forecasts, announcing a CPE when the normalized minimum temperature during the upcoming day's peak hours falls below a certain threshold, ranging from 0 to 0.15 in 30 steps. This method resulted in a 3.47% improvement as the highest enhancement among all threshold values.

Enhancements were made to these heuristics by incorporating additional criteria for CPE declaration:

3. An advanced version of the first heuristic introduced two supplementary conditions: (a) the maximum consumption during peak hours must exceed the highest consumption in off-peak periods within the same day, and (b) the peak consumption during peak

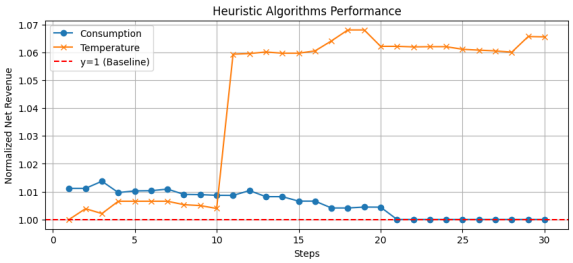

Figure 2.17: Heuristic algorithms performance in training sets (Option: WCO, PPP: 1%)

hours must surpass the current peak. Although this refined heuristic exhibited a 0.56% uplift in the training set, it underperformed compared to the 'do nothing' strategy in the test set.

4. The refined variant of the second heuristic imposed an extra criterion, declaring a CPE if the maximum consumption in a given period exceeds the current peak load. This adjustment led to an average enhancement of 4.3% in training datasets and successfully boosted the distributor's net revenue by 4.46% in February 2017. Despite these gains, the algorithm did not match the performance and consistency of the PPOD algorithm in test set evaluations or overall average improvement.

Figure 2.17 visualizes the outcomes for the enhanced algorithms in the training set. After determining the optimal threshold for each heuristic, they were executed at these ideal settings for performance benchmarking.

References

Akkara, S. and Selvakumar, I. (2023). Review on optimization techniques used for smart grid. *Measurement: Sensors*, 30:100918.

Angelus, A. (2021). Distributed renewable power generation and implications for capacity investment and electricity prices. *Production and Operations Management*, 30(12):4614–4634.

Annala, S. (2015). Households' willingness to engage in demand response in the finnish retail electricity market: an empirical study. Master's thesis, Lappeenranta University of Technology.

Aurangzeb, K., Aslam, S., Mohsin, S. M., and Alhussein, M. (2021). A fair pricing mechanism in smart grids for low energy consumption users. *IEEE Access*, 9:22035–22044.

Chan, S.-C., Tsui, K. M., Wu, H., Hou, Y., Wu, Y.-C., and Wu, F. F. (2012). Load/price forecasting and managing demand response for smart grids: Methodologies and challenges. *IEEE signal processing magazine*, 29(5):68–85.

Choi, D. G., Lim, M. K., Murali, K., and Thomas, V. M. (2020). Why have voluntary time-of-use tariffs fallen short in the residential sector? *Production and Operations Management*, 29(3):617–642.

Faruqui, A. and Sergici, S. (2010). Household response to dynamic pricing of electricity: a survey of 15 experiments. *Journal of regulatory Economics*, 38(2):193–225.

Fitzpatrick, P., D'Ettorre, F., De Rosa, M., Yadack, M., Eicker, U., and Finn, D. P. (2020). Influence of electricity prices on energy flexibility of integrated hybrid heat pump and thermal storage systems in a residential building. *Energy and Buildings*, 223:110142.

Gao, J., Ma, Z., Yang, Y., Gao, F., Guo, G., and Lang, Y. (2020). The impact of customers' demand response behaviors on power system with renewable energy sources. *IEEE Transactions on Sustainable Energy*, 11(4):2581–2592.

Ghasemkhani, A. and Yang, L. (2018). Reinforcement learning based pricing for demand response. In *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, pages 1–6.

Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., Tan, J., Kumar, V., Zhu, H., Gupta, A., Abbeel, P., et al. (2018). Soft actor-critic algorithms and applications. *arXiv preprint arXiv:1812.05905*.

He, Y., Wang, B., Wang, J., Xiong, W., and Xia, T. (2012). Residential demand response behavior analysis based on monte carlo simulation: The case of yinchuan in china. *Energy*, 47(1):230–236. Asia-Pacific Forum on Renewable Energy 2011.

Herter, K. (2007). Residential implementation of critical-peak pricing of electricity. *Energy Policy*, 35(4):2121–2130.

Herter, K. and Wayland, S. (2010). Residential response to critical-peak pricing of electricity: California evidence. *Energy*, 35(4):1561–1567. Demand Response Resources: the US and International Experience.

Hydro-Quebec (2024a). Electricity rates – 2023 edition. `https://www.hydroquebec.com/residential/customer-space/rates/`. Accessed: 2024-02-26.

Hydro-Quebec (2024b). Open data. `https://www.hydroquebec.com/open-data/`. Accessed: 2024-02-26.

International Energy Agency (IEA) (2024). Electricity 2024. `https://www.iea.org/reports/electricity-2024`. Licence: CC BY 4.0.

Ismail, A. and Baysal, M. (2023). Dynamic pricing based on demand response using actor–critic agent reinforcement learning. *Energies*, 16(14).

Jang, D., Eom, J., Kim, M. G., and Rho, J. J. (2015). Demand responses of korean commercial and industrial businesses to critical peak pricing of electricity. *Journal of Cleaner Production*, 90:275–290.

Javaid, N., Ahmed, A., Iqbal, S., and Ashraf, M. (2018). Day ahead real time pricing and critical peak pricing based power scheduling for smart homes with different duty cycles. *Energies*, 11(6):1464.

Jessoe, K., Rapson, D., and Smith, J. B. (2014). Towards understanding the role of price in residential electricity choices: Evidence from a natural experiment. *Journal of Economic Behavior & Organization*, 107:191–208.

Kim, B.-G., Zhang, Y., van der Schaar, M., and Lee, J.-W. (2014). Dynamic pricing for smart grid with reinforcement learning. In *2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 640–645.

Kim, B.-G., Zhang, Y., van der Schaar, M., and Lee, J.-W. (2016). Dynamic pricing and energy consumption scheduling with reinforcement learning. *IEEE Transactions on Smart Grid*, 7(5):2187–2198.

Kwag, H.-G. and Kim, J.-O. (2014). Reliability modeling of demand response considering uncertainty of customer behavior. *Applied Energy*, 122:24–33.

Lavin, L. and Apt, J. (2021). The importance of peak pricing in realizing system benefits from distributed storage. *Energy Policy*, 157:112484.

Liang, X., Li, X., Lu, R., Lin, X., and Shen, X. (2013). Udp: Usage-based dynamic pricing with privacy preservation for smart grid. *IEEE Transactions on smart grid*, 4(1):141–150.

Liu, D., Sun, Y., Qu, Y., Li, B., and Xu, Y. (2019). Analysis and accurate prediction of user's response behavior in incentive-based demand response. *IEEE Access*, 7:3170–3180.

Mathew, A., Roy, A., and Mathew, J. (2020). Intelligent residential energy management system using deep reinforcement learning. *IEEE Systems Journal*, 14(4):5362–5372.

Ming, H., Meng, J., Gao, C., Song, M., Chen, T., and Choi, D.-H. (2023). Efficiency improvement of decentralized incentive-based demand response: Social welfare analysis and market mechanism design. *Applied Energy*, 331:120317.

Mirzaei, M. A., Mehrjerdi, H., and Saatloo, A. M. (2023). Robust strategic behavior of a large multi-energy consumer in electricity market considering integrated demand response. *IEEE Systems Journal*, 17(4):6346–6356.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533.

Moghaddam, V., Yazdani, A., Wang, H., Parlevliet, D., and Shahnia, F. (2020). An online reinforcement learning approach for dynamic pricing of electric vehicle charging stations. *IEEE Access*, 8:130305–130313.

Naeem, A., Shabbir, A., Ul Hassan, N., Yuen, C., Ahmad, A., and Tushar, W. (2015). Understanding customer behavior in multi-tier demand response management program. *IEEE Access*, 3:2613–2625.

Parag, Y. and Sovacool, B. K. (2016). Electricity market design for the prosumer era. *Nature energy*, 1(4):1–6.

Park, S., Jin, Y., Song, H., and Yoon, Y. (2015). Designing a critical peak pricing scheme for the profit maximization objective considering price responsiveness of customers. *Energy*, 83:521–531.

Parker, G. G., Tan, B., and Kazan, O. (2019). Electric power industry: Operational and public policy challenges and opportunities. *Production and Operations Management*, 28(11):2738–2777.

Piette, M. A., Watson, D., Motegi, N., Kiliccote, S., and Xu, P. (2006). Automated critical peak pricing field tests: Program description and results. *Lawrence Berkeley National Laboratory*.

Remani, T., Jasmin, E., and Ahamed, T. I. (2019). Residential load scheduling with renewable generation in the smart grid: A reinforcement learning approach. *IEEE Systems Journal*, 13(3):3283–3294.

Schaul, T., Quan, J., Antonoglou, I., and Silver, D. (2015). Prioritized experience replay. *arXiv preprint arXiv:1511.05952*.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Sheikhi, A., Rayati, M., and Ranjbar, A. (2016). Dynamic load management for a residential customer; reinforcement learning approach. *Sustainable Cities and Society*, 24:42–51.

Silva, B. N., Khan, M., and Han, K. (2020). Futuristic sustainable energy management in smart environments: A review of peak load shaving and demand response strategies, challenges, and opportunities. *Sustainability*, 12(14):5561.

Sridhar, A., Honkapuro, S., Ruiz, F., Stoklasa, J., Annala, S., Wolff, A., and Rautiainen, A. (2023). Toward residential flexibility—consumer willingness to enroll household loads in demand response. *Applied Energy*, 342:121204.

Sundt, S., Rehdanz, K., and Meyerhoff, J. (2020). Consumers' willingness to accept time-of-use tariffs for shifting electricity demand. *Energies*, 13(8).

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

U.S. Energy Information Administration (2023). Annual electric power industry report, form eia-861 detailed data files. `https://www.eia.gov/electricity/data/eia861/`. Accessed: 2024-02-26.

Van Hasselt, H., Guez, A., and Silver, D. (2016). Deep reinforcement learning with double q-learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30, pages 2094–2100.

Vassileva, I., Wallin, F., and Dahlquist, E. (2012). Understanding energy consumption behavior for future demand response strategy development. *Energy*, 46(1):94–100. Energy and Exergy Modelling of Advance Energy Systems.

Vázquez-Canteli, J. R. and Nagy, Z. (2019). Reinforcement learning for demand response: A review of algorithms and modeling techniques. *Applied Energy*, 235:1072–1089.

Wang, B., Cai, Q., and Sun, Z. (2020). Determinants of willingness to participate in urban incentive-based energy demand-side response: An empirical micro-data analysis. *Sustainability*, 12(19).

Wang, S., Bi, S., and Zhang, Y. A. (2021). Reinforcement learning for real-time pricing and scheduling control in ev charging stations. *IEEE Transactions on Industrial Informatics*, 17(2):849–859.

Wang, Z., Schaul, T., Hessel, M., Hasselt, H., Lanctot, M., and Freitas, N. (2016). Dueling network architectures for deep reinforcement learning. In *International conference on machine learning*, pages 1995–2003. PMLR.

Wolak, F. A. (2007). Residential customer response to real-time pricing: The anaheim critical peak pricing experiment. *UC Berkeley: Center for the Study of Energy Markets*.

Zeng, B., Wu, G., Wang, J., Zhang, J., and Zeng, M. (2017). Impact of behavior-driven demand response on supply adequacy in smart distribution systems. *Applied Energy*, 202:125–137.

Zhang, L., Gao, Y., Zhu, H., and Tao, L. (2022). A distributed real-time pricing strategy based on reinforcement learning approach for smart grid. *Expert Systems with Applications*, 191:116285.

Zhang, Q., Wang, X., and Fu, M. (2009). Optimal implementation strategies for critical peak pricing. In *2009 6th International Conference on the European Energy Market*, pages 1–6.

Zhang, X. (2014). Optimal scheduling of critical peak pricing considering wind commitment. *IEEE Transactions on Sustainable Energy*, 5(2):637–645.

Zhong, S., Wang, X., Zhao, J., Li, W., Li, H., Wang, Y., Deng, S., and Zhu, J. (2021). Deep reinforcement learning framework for dynamic pricing demand response of regenerative electric heating. *Applied Energy*, 288:116623.

Chapter 3

Towards Sustainable Energy Use: Reinforcement Learning for Demand Response in Commercial Buildings

Abstract

This study presents a new demand response framework for optimizing electricity consumption in small and medium-sized commercial buildings. We develop a mathematical model categorizing building loads into non-controllable, controllable, and HVAC consumption, aiming to minimize costs, CO_2 emissions and dissatisfaction and maximize peak load shaving. We implement and compare three reinforcement learning (RL) algorithms for controllable loads and HVAC systems against traditional heuristic approaches, using eight weeks of winter consumption data from a commercial building. Results show that RL algorithms, particularly PPOD & TD3 combinations, achieve peak load shaving ratios exceeding 25%, with significant cost reductions and environmental benefits. Our analysis incorporates the impact of outdoor temperature variations and risk assessments using VaR and CVaR metrics. This study contributes to developing efficient and environmentally conscious energy management strategies for commercial buildings, with implications for policy and industry practices in sustainable energy transitions.

3.1 Introduction

In the face of growing energy demands and increasing environmental concerns, demand response (DR) has emerged as a critical strategy in modern power systems. DR offers significant benefits in terms of cost savings, infrastructure investment deferral, and GreenHouse Gas (GHG) emission reduction. The importance of DR is underscored by its potential to reduce costs across power systems. For instance, the U.S. Department of Energy estimates that Grid-interactive Efficient Buildings (GEBs) could unlock untapped opportunities valued between $8 billion and $18 billion annually by 2030, representing 2–6% of total U.S. electricity generation and transmission costs (Satchwell et al., 2021).

The impact of DR extends beyond cost savings. By flattening peak demand curves and improving overall system efficiency, DR can significantly reduce the need for additional capacity, leading to substantial long-term savings in infrastructure investments. Moreover, DR plays a crucial role in reducing GHG emissions. Through demand flexibility and energy efficiency, GEBs could decrease U.S. CO_2 emissions by 80 million tons per year by 2030, equivalent to 6% of total U.S. power sector CO_2 emissions (Langevin et al., 2019).

Within the broader context of DR, commercial buildings represent a significant untapped resource. In the United States, small and medium-sized commercial buildings account for 50% of the commercial building floor area (Energy Information Administration, 2021), indicating vast potential for DR implementation. The commercial sector as a whole represented 35% of U.S. electricity sales in 2020 (EIA, 2021), further emphasizing the importance of targeting this sector for DR initiatives.

Electricity consumption in commercial buildings can be broadly categorized into three groups: non-controllable loads, controllable loads, and HVAC (Heating, Ventilation, and Air Conditioning) consumption. While non-controllable loads, such as security devices, cannot be adjusted for DR purposes, controllable loads like lighting systems offer opportunities for reduction and optimization. However, it is the HVAC systems that present the most significant potential for DR in commercial buildings.

HVAC systems are particularly suitable for DR applications due to their substantial energy consumption and inherent thermal inertia. In commercial buildings, HVAC systems account for more than 50% of energy consumption (Tang et al., 2018; Yuan et al., 2021).

This high proportion of total building energy use makes HVAC an ideal target for DR strategies. Furthermore, the thermal mass of buildings allows for temperature adjustments without immediate impact on occupant comfort, providing flexibility in energy consumption patterns. This is particularly effective not only over day-night cycles but also within shorter time frames when there is a significant difference between indoor and outdoor temperatures. In commercial buildings, this allows for hourly adjustments during peak and off-peak periods while maintaining comfort levels through the buffering capacity of the building's thermal mass.

The capacity for the top 100 hours of peak demand in a year accounts for nearly 20% of electricity costs for power grids (Arnold, 2011). This highlights the importance of optimizing electricity consumption, particularly during peak hours, to reduce costs and improve grid efficiency. By leveraging the thermal characteristics of buildings and implementing intelligent control strategies, a well-designed DR program can achieve significant energy reductions while maintaining acceptable levels of occupant comfort.

This study introduces a mathematical model aimed at optimizing electricity consumption in a group of commercial buildings. The model categorizes building consumption into three distinct groups: HVAC consumption, adjustable loads, and non-controllable loads. In addition to minimizing electricity costs, the model considers three other objectives: minimizing consumer dissatisfaction due to reductions in consumption or deviations from the desired temperature, minimizing CO_2 emissions, and maximizing peak load shaving and load reduction during peak hours.

To solve this optimization problem, we employ heuristic and Reinforcement Learning (RL) algorithms. The heuristic algorithms include setting the HVAC set point at different levels below the desired temperature, with and without considering peak hours. For the RL approach, we implement three algorithms for HVAC systems: Soft Actor-Critic for Discrete action spaces (SACD), Proximal Policy Optimization for Discrete action spaces (PPOD), and Double Dueling Deep Q-Network with Prioritized Experience Replay (D3QN). For controllable loads, we use Deep Deterministic Policy Gradient (DDPG), Twin Delayed Deep Deterministic Policy Gradient (TD3), and Proximal Policy Optimization (PPO).

Our results demonstrate that RL algorithms, particularly PPOD & TD3 and SACD

& TD3, achieve the highest peak load shaving ratios, exceeding 25%, with notable cost reductions and environmental benefits. The implementation of TD3 alters the realized controllable load consumption, with the majority of load reduction occurring during evening peak hours. The application of PPOD on HVAC load management showcases a strategic pre-heating of buildings prior to peak hours, maintaining temperatures just above the lower admissible bound during these critical periods.

Furthermore, this study considers the impact of outdoor temperature variations on the potential peak load shaving volume that an aggregator can provide. We conduct risk analyses to assess the robustness of our proposed strategies and incorporate CO_2 emission reduction as a key objective in our DR framework.

By addressing these critical aspects of DR in commercial buildings, this paper aims to contribute to the development of more effective, efficient, and environmentally conscious energy management strategies. The findings and methodologies presented here have the potential to inform policy decisions, guide industry practices, and ultimately accelerate the transition towards a more sustainable and resilient energy future.

The remainder of this paper is structured as follows: Section 3.2 explores the current literature on DR in commercial buildings and identifies the research gap addressed by this study. Section 3.3 introduces the problem indexes, parameters, and decision variables, and formulates both the reference and main optimization models. Section 3.4 presents the solution approaches, justifying the use of RL, and introduces the RL elements and data collection methodology. Finally, Section 3.5 discusses the results, summarizes the key findings, and suggests future research avenues in this field.

3.2 Literature Review

The landscape of energy management has undergone a significant transformation in recent years, with DR emerging as a critical tool for balancing the power grid and reducing costs. This review traces the evolution of DR implementation in commercial buildings, highlighting key developments in strategies, technologies, and methodologies.

The journey of DR in commercial buildings began in the early 2000s when Watson

et al. (2006) laid the groundwork by describing strategies to temporarily reduce electric load during grid emergencies or high-price periods. These initial efforts primarily focused on adjusting HVAC and lighting systems, aiming to achieve energy savings while minimizing negative impacts on occupants.

As the potential of DR became increasingly apparent, researchers like Hao et al. (2012) made strong claims about its capabilities. They suggested that the HVAC systems in just 90,000 medium-sized commercial buildings could provide the entire regulation service needed by PJM, a major power grid operator. This revelation highlighted the untapped potential of commercial buildings in grid stabilization.

The implementation of DR strategies took a significant leap forward with automation. Kiliccote (2010) reported on seven years of field performance data for automated DR in California, demonstrating that commercial buildings could reduce peak demand by an average of 13%. This success story paved the way for wider adoption of automated DR systems.

As the Internet of Things (IoT) began to permeate our lives, it also found its way into building energy management. Zhao et al. (2016) explored how IoT could improve energy efficiency and DR in commercial buildings. By integrating wireless networks with existing building automation systems, researchers gained unprecedented insights into energy consumption patterns at the zone level.

The quest for optimal DR strategies continued, with researchers exploring various approaches. Huang et al. (2018) proposed integrating residential and commercial users into a single DR system, utilizing particle swarm optimization to improve energy efficiency. Meanwhile, Li et al. (2020) developed a comprehensive energy management system for commercial building clusters, incorporating distributed generation and energy storage equipment.

As DR strategies became more sophisticated, researchers began to pay closer attention to occupant comfort. Liang et al. (2019) proposed a DR framework that considered both operating costs and comprehensive comfort levels, using stochastic programming to handle uncertainties in renewable energy output, electricity prices, and flexible energy demand.

The importance of building-specific strategies became evident as the field progressed. Christantoni et al. (2016) used whole-building energy simulation to develop and evaluate DR strategies for different zones within a multi-purpose commercial building. Their findings

emphasized the need to tailor strategies based on thermal and usage profiles of individual spaces.

Recognizing the significant role of small commercial buildings, Cai and Braun (2019) conducted a comprehensive assessment of DR potential across various building types, vintages, and locations. Their results showed promising cost savings and demand reductions, particularly for buildings with time-of-use energy rates and demand charges. As the field matured, the focus shifted towards integrating multiple aspects of DR. Darwazeh et al. (2022) provided a holistic review of peak load management strategies, combining DR programs, strategies, load forecasting models, and occupant comfort considerations.

In parallel with these developments, the evolution of control and optimization methods for DR has been equally remarkable. Hao et al. (2017) proposed a novel approach called transactive control for commercial building HVAC systems, while ho Lee and Braun (2016) developed an inverse building model to study the performance of a demand-limiting control strategy. The complexity of DR strategies grew, as did the mathematical models used to optimize them. Aussel et al. (2020) introduced a trilevel energy market model for load shifting induced by time-of-use pricing, showcasing the challenges of solving multi-leader-multi-follower games.

Artificial intelligence has also made its mark on DR research. Chen et al. (2021) explored the use of RL for optimal DR strategy in a commercial building-based virtual power plant. Building on this theme, Liang et al. (2021) proposed a safe RL for resilient proactive scheduling in commercial buildings.

Traditional control methods continued to evolve alongside AI approaches. Tang et al. (2019) developed a model predictive control (MPC) strategy for optimizing the operation of central air-conditioning systems with active cold storage during fast DR events. More recently, Hosseini et al. (2022) proposed a robust MPC approach for online energy scheduling of multiple commercial buildings. The integration of Energy Management Systems (EMS) and Building Automation Systems (BAS) has played a crucial role in enabling effective DR strategies. Piette et al. (2009) laid the groundwork for automated DR infrastructure in commercial buildings, introducing the concept of an open, interoperable communications infrastructure. Cybersecurity concerns in building automation systems were addressed by

Jones and Carter (2017), who developed and tested a building automation intrusion detection system. This innovative system provided a cyber-secure connection between public and private BAS networks. The scope of DR management in commercial buildings expanded with the work of Wang et al. (2018), who presented an optimal operation model that considered not only traditional flexible loads but also emerging technologies such as concentrating solar power plants.

Understanding and predicting load profiles for commercial buildings has become increasingly crucial. He and Liu (2017) delved into load profile analysis for commercial buildings microgrids under DR conditions, while Pallonetto et al. (2022) incorporated machine learning models for electricity demand forecasting. The economic and market aspects of DR in commercial buildings have also been a focus of research. Yoon et al. (2020) developed an optimal pricing strategy for price-based DR in HVAC systems, while Kim and Norford (2016) explored the price-based DR of energy storage resources in commercial buildings within the context of wholesale electricity markets.

The integration of renewable energy sources and energy storage systems has opened up new frontiers in DR strategies. Hu et al. (2012) envisioned a future where buildings operate as part of an interconnected cluster, while Hossain et al. (2024) focused on optimal peak-shaving for dynamic DR in smart Malaysian commercial buildings. Wang et al. (2021) proposed an optimization framework for low-carbon oriented integrated energy system management in commercial buildings, with a focus on electric vehicle (EV) DR.

Advancements in modeling and simulation techniques have significantly influenced DR strategies. Gao et al. (2015) developed a robust DR control strategy that could maintain effectiveness under uncertain load conditions. Christantoni et al. (2015) created a comprehensive simulation model for a multi-purpose commercial building specifically designed for DR analysis. Yin et al. (2016) developed a novel DR estimation framework applicable to both residential and commercial buildings.

HVAC systems have played a crucial role in DR strategies. Beil et al. (2016) explored the potential of commercial HVAC systems in frequency regulation, while MacDonald et al. (2020) critically examined the efficiency impacts of such DR strategies. Aghniaey and Lawrence (2018) brought attention to the impact of increased cooling setpoint tempera-

tures during DR events on thermal comfort.

Recent research has also focused on quantifying energy flexibility and DR potential in commercial buildings. Afroz et al. (2023) presented a comprehensive analysis of the energy flexibility and DR potential of schools, offices, and data centers in Australia.

Finally, case studies and pilot programs have provided valuable insights into the practical implementation of DR strategies. Li et al. (2006) presented a case study of a DR pilot program in a large commercial office building in Shanghai, while Son et al. (2014) explored the potential of combining chiller systems and energy storage systems for DR participation. Mutule et al. (2017) conducted a feasibility study for DR in a commercial building in Latvia, demonstrating the initial steps in implementing a commercial building automation system with automated energy consumption scheduling units.

While significant progress has been made in developing DR strategies for commercial buildings, several key areas remain underexplored. First, most existing studies focus on either HVAC systems or controllable loads independently, without considering their combined optimization potential. Second, the application of advanced RL algorithms to simultaneously manage both HVAC and controllable loads in commercial buildings is still limited. Third, there is a lack of comprehensive frameworks that integrate multiple objectives such as cost minimization, CO_2 emission reduction, peak load shaving, and occupant comfort in a single model.

This study aims to address these gaps by introducing a novel DR framework that optimizes both HVAC systems and controllable loads in small and medium-sized commercial buildings. By implementing and comparing multiple RL algorithms against traditional heuristic approaches, we provide insights into the most effective strategies for DR in this context. Furthermore, our model uniquely incorporates multiple objectives, including cost reduction, environmental impact, peak load shaving, and occupant satisfaction. This comprehensive approach, combined with our analysis of outdoor temperature variations and risk assessments using VaR and CVaR metrics, contributes to the development of more robust and practical DR strategies for commercial buildings.

3.3 Modeling

This study introduces a mathematical model aimed at optimizing electricity consumption in a group of commercial buildings. The buildings' consumption is categorized into three distinct groups: HVAC consumption, which accounts for the energy used for heating or cooling the buildings; adjustable loads, such as lighting, that are reducible; and non-controllable loads, like refrigerators, security and alarm services. In addition to minimizing electricity costs, the model considers three other objectives: minimizing consumer dissatisfaction due to reductions in consumption or deviations from the desired temperature, minimizing CO_2 emission and maximizing peak load shaving and load reduction during peak hours.

The planning horizon for this model is weekly with five-minutes time steps, and it employs a simplified version of HVAC systems that operate in on and off states. Temperature dynamics within the buildings are modeled using thermal capacitance (C_i), thermal resistance (R_i), and HVAC power rating (h_i). Specifically, thermal capacitance represents the ability of the building to store heat, thermal resistance indicates the building's insulation effectiveness, and HVAC power rating defines the rate at which the HVAC system can change the indoor temperature.

3.3.1 Indexes and Parameters

The models in this study use two indexes for time steps (peak and off-peak periods) and buildings:

$i \in \{\mathcal{N}\}$	building i in the targeted group of buildings
$t \in \{\mathcal{T}\}$	episode t in horizon $[1, 2, 3, ..., T]$
$t \in \{\mathcal{H}\}$	episode t in peak period, $(\mathcal{H} \subset \mathcal{T})$

Below is the list of parameters:

$C_i \doteq$ Thermal capacitance of building i (W/°C)

$R_i \doteq$ Thermal resistance of building i (°C/W)

$h_i \doteq$ HVAC power rating at building i (kW)

$T_{i,t}^{des} \doteq$ Desired indoor temperature in building i at time $t(^\circ C)$

$T_t^{out} \doteq$ Outdoor temperature forecast at time $t(^\circ C)$

$T^{dev} \doteq$ Maximum allowable deviation from the desired temperature $(^\circ C)$

$e_{i,t}^N \doteq$ Non-controllable load of building i at time t

$e_{i,t}^C \doteq$ Controllable load of building i at time t

$P_t \doteq$ Electricity price ($/kWh) to buy from wholesale market at time t

$\alpha_i \doteq$ Building i's discomfort weight factor for the temperature difference ($/$^\circ C^2$))

$\beta_i \doteq$ Building i's discomfort weight factor for the reduced power ($/$^\circ C^2$)

$\gamma \doteq$ Weight factor (rebate rate) for the reduced load ($/kWh)

$\delta \doteq$ Weight factor (Carbon tax) for the CO_2 emission ($/g$CO_2$eq)

$\epsilon \doteq$ Total carbon intensity of electricity consumed in Quebec (gCO_2eq/kWh)

In this list, parameters T_t^{out}, P_t, $e_{i,t}^N$ and $e_{i,t}^C$ are realized in real-time.

3.3.2 Decision and State Variables

The decision variables below indicate HVAC status, electricity flows and curtailed loads in the buildings:

$B_{i,t} \doteq$ HVAC status (on/off) in building i's at time t

$T_i^{in} \doteq$ Indoor temperature $(^\circ C)$ in building i's at time t

$E_{i,t}^N \doteq$ Realized non-controllable load (kWh) of building i at time t

$E_{i,t}^C \doteq$ Realized controllable load (kWh) of building i at time t

$E_t^B \doteq$ Total energy consumption (kWh) in the reference model at time t

3.3.3 Reference Model

The reference model is introduced below to calculate the reference consumption record of the buildings when they, regardless of DR objectives, only minimize their dissatisfaction.

$$\min \sum_{i,t} (T_{i,t}^{des} - T_{i,t}^{in})^2 \tag{3.1}$$

subject to:

$$T_{i,t}^{in} = T_{i,t-1}^{in} + \left(\frac{T_t^{out} - T_{i,t-1}^{in}}{R_i} + B_{i,t}.h_i \right) \frac{\Delta t}{C_i} \qquad \forall i,t \tag{3.2}$$

$$|T_{i,t}^{des} - T_{i,t}^{in}| \leq T^{dev} \qquad \forall i,t \tag{3.3}$$

$$E_{i,t}^{N} = e_{i,t}^{N} \qquad \forall i,t \tag{3.4}$$

$$E_{i,t}^{C} = e_{i,t}^{C} \qquad \forall i,t \tag{3.5}$$

$$E_t^B = \sum_i E_{i,t}^N + E_{i,t}^C + B_{i,t}.h_i \qquad \forall t \tag{3.6}$$

$$0 \leq T_{i,t}^{in}, B_{i,t}, E_{i,t}^N, E_{i,t}^C, E_t^B \qquad \forall i,t \tag{3.7}$$

In the reference model, objective function 3.1 keeps the indoor temperature as close as possible to the desired temperature, with no cost term included. Equation 3.2 calculates the indoor temperature in the next time step where Δt represents time step. Equation 3.3 makes sure the inside temperature remains in the allowed temperature bound. Constraints 3.4 and 3.5 ensure all load types are covered at their preferred power. Finally, Equation 3.6 calculates the total energy consumed at time t and Equation 3.7 indicates the decision variables' domain.

3.3.4 Main Model

Once the reference model is solved and E_t^B is calculated, the main model uses E_t^B as a reference point to calculate the reduced consumption during peak hours. The main model is presented below:

$$
\begin{aligned}
\min & \sum_{i,j,t\in\mathcal{T}} P_t(B_{i,t}\cdot h_i + E^N_{i,t} + E^C_{i,t}) \\
& + \sum_{i,j,t\in\mathcal{T}} \alpha_i(T^{\text{in}}_{i,t} - T^{\text{des}}_{i,t})^2 + \beta_i(E^C_{i,t} - e^C_{i,t})^2 \\
& -\gamma \sum_{t\in\mathcal{H}} \left(E^B_t - \sum_{i,j} B_{i,t}\cdot h_i + E^N_{i,t} + E^C_{i,t} \right) \\
& +\delta\epsilon \sum_{t\in\mathcal{T}} \left(\sum_{i,j} B_{i,t}\cdot h_i + E^N_{i,t} + E^C_{i,t} \right)
\end{aligned} \tag{3.8}
$$

Subject to:

$$T^{in}_{i,t} = T^{in}_{i,t-1} + \left(\frac{T^{out}_t - T^{in}_{i,t-1}}{R_i} + B_{i,t}.h_i \right) \frac{\Delta t}{C_i} \qquad \forall i,t \tag{3.9}$$

$$|T^{des}_{i,t} - T^{in}_{i,t}| \leq T^{dev} \qquad \forall i,t \tag{3.10}$$

$$E^N_{i,t} = e^N_{i,t} \qquad \forall i,t \tag{3.11}$$

$$E^C_{i,t} \leq e^C_{i,t} \qquad \forall i,t \tag{3.12}$$

$$0 \leq T^{in}_{i,t}, B_{i,t}, E^N_{i,t}, E^C_{i,t} \qquad \forall i,t \tag{3.13}$$

Objective function 3.8 consists of four terms. The first term minimizes the electricity cost, the second term minimizes the dissatisfaction arising from adopting DR and the difference between the desired and actual indoor temperature. The third term maximizes the energy consumption reduction during the peak period and the forth term minimizes CO_2 emission. Constraint 3.9 calculates the indoor temperature in the next time step, and Constraint 3.10 ensures the indoor temperature does not deviate from the desired temperature more than allowed extent. Constraints 3.11 and 3.12 ensure that the non-controllable load is covered and the deducted load is less than or equal to the allowed bound. Constraint 3.13 indicates the admissible values for the decision variables. For the dissatisfaction terms, a quadratic form is chosen, indicating that the dissatisfaction arising from DR increases quadratically as the deviation increases. Also, this study considers heating commercial buildings during winter (when peak loads occur in Quebec); however, this model could also be used for summer scenarios where HVAC cools the building. In that case, a negative sign must be used before $B_{i,t}\cdot h_i$ in Equations 3.2 and 3.9.

3.4 Methodology

The DR problem in commercial buildings presents a complex and dynamic challenge that requires an adaptive and intelligent approach. Traditional optimization methods, such as linear programming or dynamic programming, are often inadequate for addressing this problem due to the real-time availability and constant updates of critical parameters, such as energy consumption, outdoor temperature, and electricity market prices, which are realized dynamically as the system operates. Long-term forecasting of these parameters is challenging, and relying on forecasts can lead to sub-optimal solutions and reduced effectiveness. Figures 3.1 and 3.2 demonstrate the high fluctuations in both electricity price and carbon emissions intensity (part of the information needed for decision making), highlighting the need for a more flexible and responsive approach.

To overcome these challenges, we formulate the DR problem as a Markov Decision Process (MDP). An MDP is a mathematical framework for modeling sequential decision-making problems, where an agent interacts with an environment by taking actions and observing the resulting states and rewards. The DR problem can be naturally represented as an MDP, with the states capturing the current conditions of the buildings and surroundings (e.g., indoor temperatures, energy consumption levels), the actions representing the adjustments to controllable loads and HVAC systems, and the rewards reflecting the objectives of minimizing energy costs, maintaining occupant comfort, and achieving consumption reduction during peak hours and emission reduction.

The MDP formulation allows us to leverage the power of RL algorithms, which are particularly well-suited for solving complex, dynamic problems like the DR scenario. RL algorithms enable an agent to learn optimal policies through trial-and-error interactions with the environment, without requiring a complete model of the system dynamics (indoor temperature dynamics in this study). This adaptability is crucial in the DR context, where the real-time realization of parameters creates a constantly evolving environment.

RL algorithms balance the exploration-exploitation trade-off, allowing the agent to explore different actions and learn from their consequences, while also exploiting the knowledge gained to make more informed decisions. Through this iterative process, the RL agent can discover near-optimal policies that effectively manage the trade-offs between energy cost

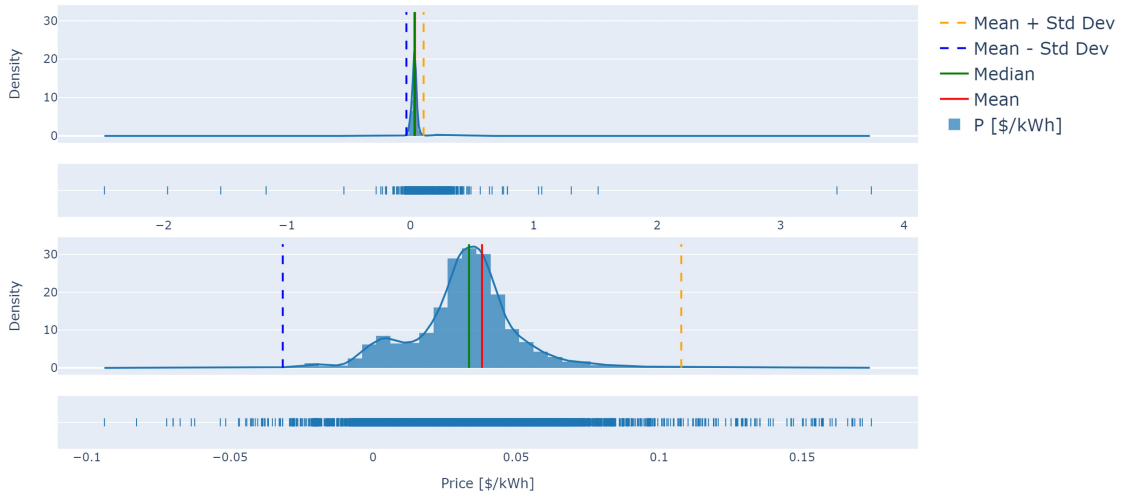

Figure 3.1: Distribution of Hydro-Quebec Wholesale Electricity Price Proxy (Bottom Panel: Zoomed View of Top Panel)

minimization, occupant comfort, and peak load and emission reduction.

In this study, we employ a range of RL algorithms to address the DR problem, considering both discrete and continuous action spaces. For the discrete action space, where the HVAC systems operate in an on/off state, we utilize SACD, PPOD, and D3QN. These algorithms are adapted from their originally designed continuous action space counterparts to handle the discrete nature of the HVAC control problem. For the continuous action space, where the controllable ligthing loads can be adjusted within a range, we employ the DDPG, TD3 and PPO algorithms. These algorithms are well-suited for handling the continuous nature of the controllable load adjustments. data collection, pre-processing and visualization and algorithms implementation are available on GitHub [1] and a brief introduction of the developed algorithms is presented in the Appendix.

3.4.1 RL Elements

Figure 3.3 depicts the process and elements of RL algorithms. The RL elements, including the state space, action space, reward function, and state transition, are defined separately for the HVAC control problem and the controllable load adjustment problem.

HVAC Control Problem

[1] `https://github.com/srmadani/Demand-Response-in-Commercial-Buildings`

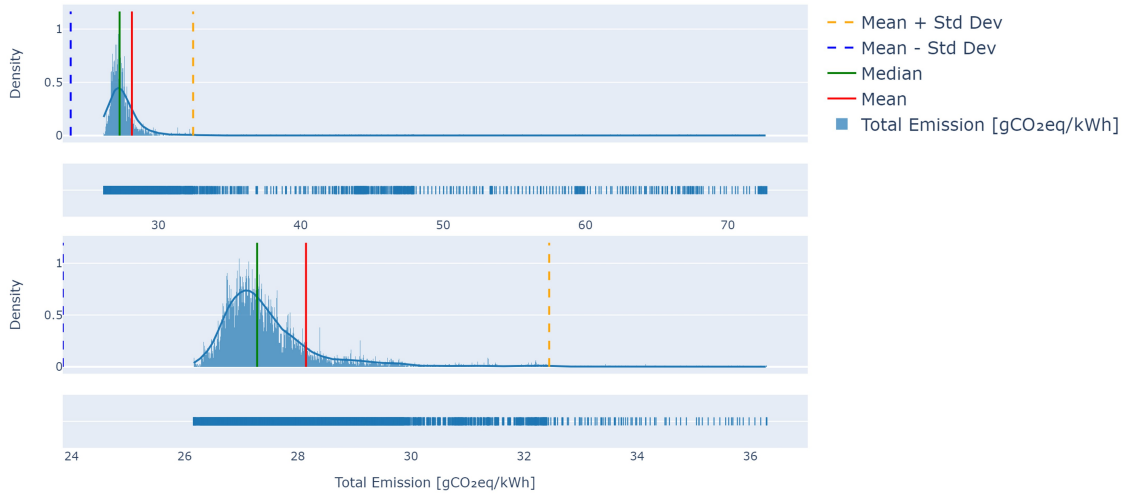

Figure 3.2: Distribution of Hydro-Quebec electricity total CO_2 emission (Bottom Panel: Zoomed View of Top Panel)

- **State Space:** Features include the next time step's wholesale price proxy, sine and cosine transformations of the hour (to capture time dependencies more smoothly instead of providing the hour directly), the next time step's reducible load, emission rate, and indicator variables for peak hours and working days.

- **Action Space:** Discrete, with a binary variable representing the HVAC system's state: 0 for off and 1 for on.

- **Reward Function:** Derived from the objective function in Equation 3.8, aiming to minimize electricity costs, dissatisfaction from deviations from desired indoor temperatures and emission, and maximize consumption reduction during peak hours.

- **State Transition:** Realized by obtaining data for the problem parameters (e.g., outdoor temperature, market prices), calculating the next time step's indoor temperature using Equation 3.9, and updating the indicator variables and placeholders accordingly.

Controllable Load Adjustment Problem

- **State Space:** Features include the next time step's wholesale price proxy, sine and cosine of the hour, the next time step's reducible load, emission rate, and indicator variables for peak hours and working days.

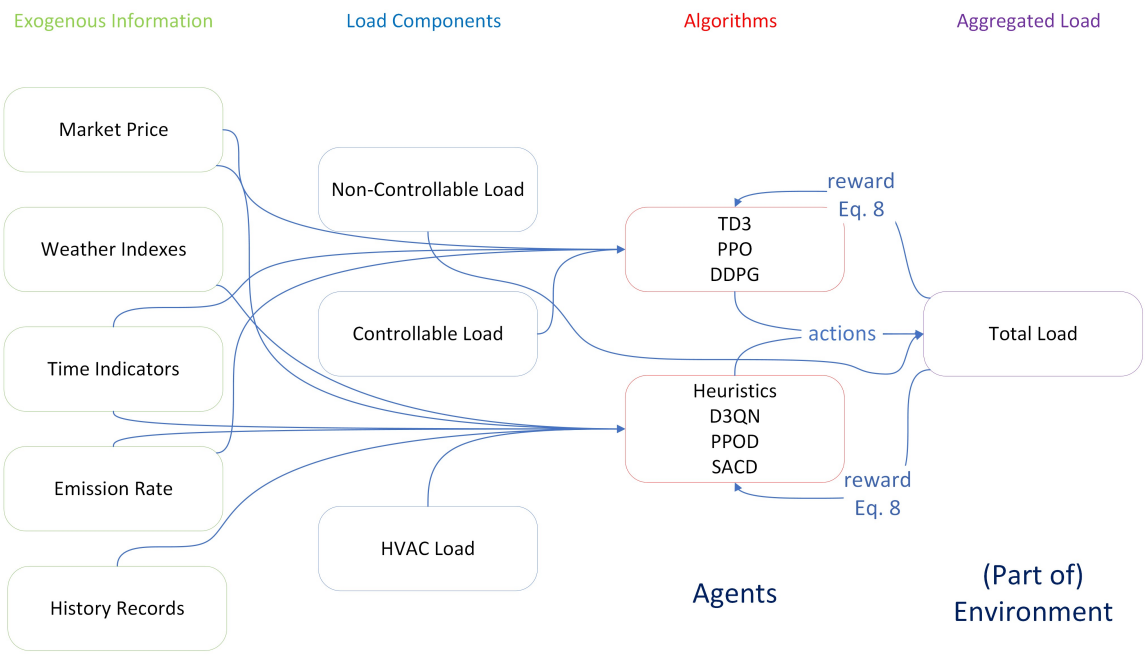

Figure 3.3: Methodology

- **Action Space:** Continuous, represented by a value between 0 and 1, indicating the ratio by which the controllable load is reduced.
- **Reward Function:** Similar to the HVAC control problem, calculated using Equation 3.8, aiming to minimize electricity costs, dissatisfaction from reduced power consumption and emission, and consumption during peak hours.
- **State Transition:** Determined by obtaining the realized parameters, updating the indicator variables and placeholders, and calculating the next time step's reducible load based on the chosen action.

3.4.2 Data Collection

This study uses real consumption data from the Varennes Library at Concordia University in Varennes, Quebec, Canada (Sartori et al., 2023), and the outdoor temperature records are sourced from Canada's Environment and Natural Resources[2], while the wholesale market price for electricity in Quebec is approximated using Hydro-Quebec's electricity rate at

[2] `https://climate.weather.gc.ca/historical_data/search_historic_data_e.html`

NYISO[3]. Additionally, the emission rate of electricity generated in Quebec is derived from data obtained from Electricity Maps[4], and the HVAC configuration details are acquired from Brainbox AI.[5]

Considering the climatic conditions of Quebec, where peak electricity demands occur during winter, this study focuses on the months of January and February. During these months, there are distinct morning (6-9 AM) and evening (4-8 PM) peak periods when people are commuting to and from work, prompting grid operators to aim at minimizing electricity consumption and peak loads during these critical times[6].

Our dataset spans eight weeks, with six weeks designated for training, one week for validation, and one week for testing. The desirable indoor temperature is set to 22°C during working hours and 18°C during non-working hours, with a linear transition between these temperatures. Data is collected at 5-minute intervals, ensuring a detailed and precise representation of consumption patterns.

3.5 Results and discussion

We start this section by comparing the results of applying heuristic and RL algorithms on HVAC load management. The non-controllable load is added to the total load unchanged, while for the controllable load, the TD3 algorithm demonstrated the best overall performance.

- **Heuristic 0**: The set point for turning on the HVAC is equal to the desired temperature. The HVAC starts working as soon as the indoor temperature falls below the desired level.

- **Heuristic 1**: The set point is one degree below the desired temperature.

- **Heuristic 2**: The set point is two degrees below the desired temperature.

[3] `https://www.nyiso.com/custom-reports`
[4] `https://app.electricitymaps.com/map`
[5] `https://brainboxai.com/en/`
[6] `https://www.hydroquebec.com/residential/customer-space/rates/winter-credit-option.html`

Figure 3.4: Algorithms' performance

- **TD3**: This result reflects the use of the reference optimization model for HVAC and TD3 algorithm for controllable load.
- **Heuristic 2 peak**: The set point is equal to the desired temperature during off-peak hours and two degrees below the desired temperature during peak hours.
- **Heuristic 2 peak & TD3**: Similar to Heuristic 2 peak, additionally TD3 is used for the controllable load.
- **D3QN & TD3**: The D3QN algorithm is applied to HVAC, and TD3 is applied to the controllable load.
- **PPOD & TD3**: The PPOD algorithm is applied to HVAC, and TD3 is applied to the controllable load.
- **SACD & TD3**: The SACD algorithm is applied to HVAC, and TD3 is applied to the controllable load.

Figure 3.4 presents a comparison of the heuristic and RL algorithms across four metrics: total cost decrease, peak load shaved, dissatisfaction increase, and CO_2 emission decrease. Heuristic 2 demonstrates the highest cost reduction, 5.82%, and the greatest decrease in CO_2 emissions, 4.89%. However, it also causes the most significant dissatisfaction increase, 266.07%, highlighting a trade-off between cost savings and user comfort. Among the RL algorithms, PPOD & TD3 achieves the highest peak load shaving ratios, exceeding 27%, with notable cost reductions and moderate environmental benefits. SACD & TD3 also shows strong performance across all metrics, making it a balanced choice for effective HVAC load management. Overall, RL algorithms excel in optimizing peak load shaving and cost reduction while maintaining environmental benefits, albeit with varying impacts on user satisfaction. While it is true that optimizing for cost reduction and peak load shaving can sometimes negatively impact user satisfaction, this outcome is not a given. The model aims to strike a balance between these objectives, as exemplified by the inclusion of dissatisfaction costs in its calculations, ultimately offering financial incentives in exchange for slight deviations from desired comfort levels.

TD3 Implementation alters the realized controllable load consumption, as illustrated in Figure 3.5. The blue area represents the demand, while the red line indicates the actual consumption. Analysis of the results reveals that the majority of load reduction occurs during evening peak hours.

Figure 3.6 demonstrates the impact of applying PPOD on HVAC load management, showcasing the variation in average indoor temperature. The green line represents outdoor temperature, the blue line indicates desired temperature, and the red line depicts actual indoor temperature. The trend observed in the red line suggests a strategic pre-heating of buildings prior to peak hours, maintaining temperatures just above the lower admissible bound during these critical periods. To minimize CO_2 emissions and circumvent high energy prices, the actual indoor temperature occasionally drops below the desired level during off-peak hours. Figure 3.7 presents the aggregated load across the studied buildings. The combined utilization of TD3 and PPOD yields compelling results: across the 32 buildings analyzed, the average indoor temperature is maintained at 1.66 °C below the desired temperature. Moreover, the total energy savings during peak hours amount to 65,068.68

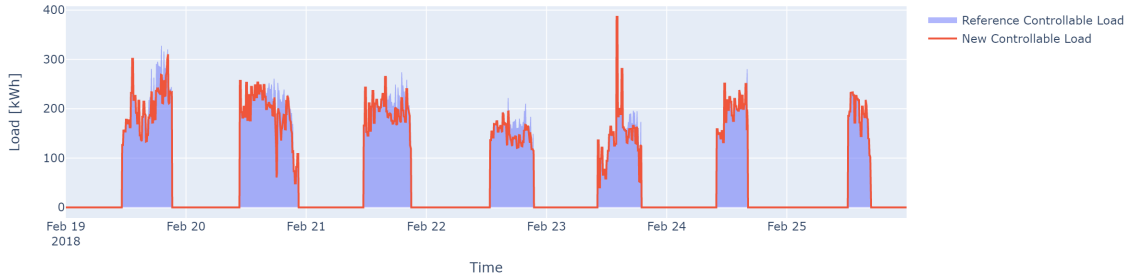

Figure 3.5: Change in controllable load

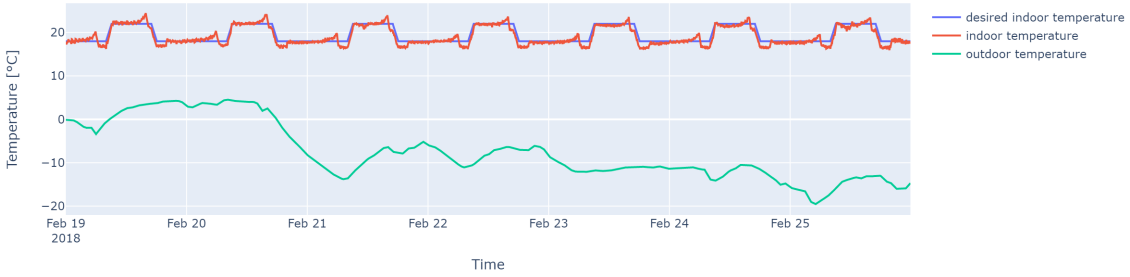

Figure 3.6: Inside temperature vs. desired and outdoor temperatures

kWh, representing a 13.22% reduction in consumption. Furthermore, the peak load shaving achieves 614.89 kW reduction, equivalent to a 27.07% decrease in maximum demand.

Figure 3.8 illustrates the variation in key performance metrics for DR potential in commercial buildings as the outdoor temperature is decreased/increased from -15 °C to +15 °C. The total consumption during peak hours shows a median reduction of approximately 14%, with a range from about 8% to 22%, indicating a significant potential for energy savings during peak demand periods, though the exact reduction can vary considerably depending on the outdoor temperature. Similarly, the reduction in peak load presents a median value around 15%, with a spread from roughly -20% to 30%. This variability suggests that while there can be substantial peak load reduction, in some scenarios, the load may not reduce as expected or might even increase slightly, possibly due to the increased HVAC demands at extreme temperatures.

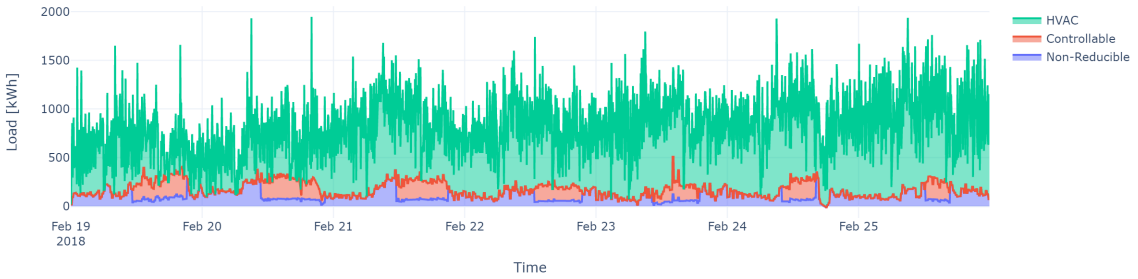

Figure 3.7: Aggregated optimized load

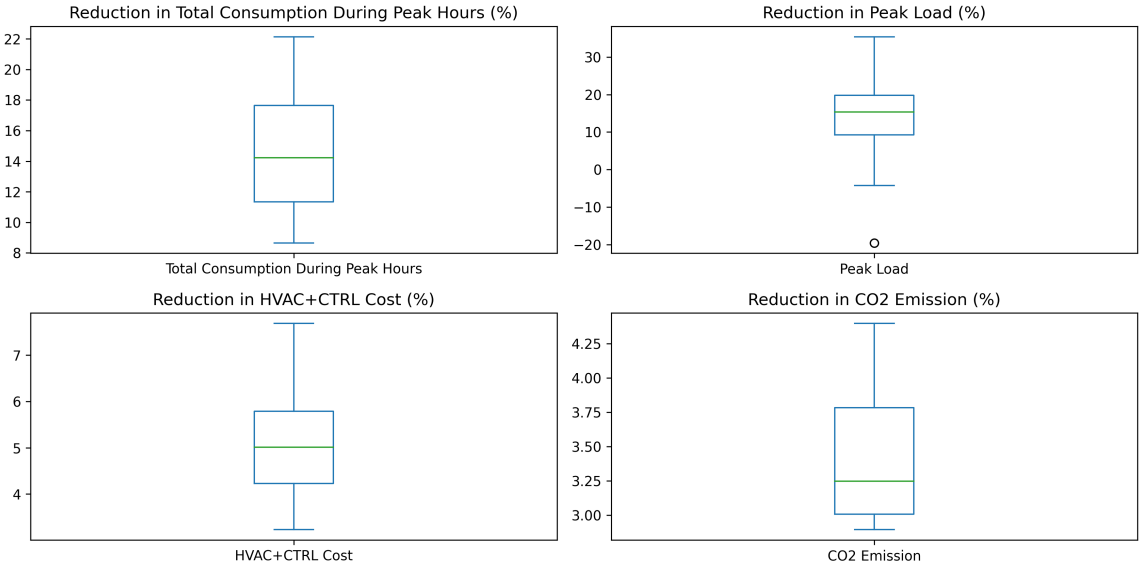

Figure 3.8: Impact of change in outdoor temperature

Furthermore, the cost reduction for HVAC and controlled loads shows a median of around 5%, with values ranging from about 4% to 7%, highlighting the economic benefits of implementing DR strategies, even as outdoor temperatures fluctuate. Additionally, the CO_2 emission reduction varies from 3% to about 4.25%, with a median around 3.25%, demonstrating the environmental benefits of DR initiatives, as reducing energy consumption directly correlates with lower CO_2 emissions.

To complement the average metrics, risk assessment in energy markets requires consideration of worst-case scenarios, which are particularly valuable for policymakers. To address this, we calculated the Value at Risk (VaR) and Conditional Value at Risk (CVaR) for Total

Consumption During Peak Hours, Peak Load, HVAC+CTRL Cost, and CO_2 Emission.

VaR represents the maximum expected loss (or equivalently minimum expected return) at a given confidence level (95%), over a specific time horizon. In our context, it indicates the worst-case scenario for each metric that we can expect with 95% confidence. CVaR, also known as Expected Shortfall, provides the expected value of the loss (or return) given that it exceeds the VaR threshold, offering insight into the severity of extreme events beyond VaR.

The results in Figure 3.9 reveal interesting patterns across the analyzed metrics. For Total Consumption During Peak Hours, both VaR and CVaR show substantial positive values around 9.5%, indicating potential decreases in consumption during extreme scenarios. This suggests that DR strategies may face challenges in consistently reducing peak hour consumption under all conditions. While the median for this metric is around 14%, with 95% confidence we can expect to have 9.45% reduction and in the 5% worst case scenarios, we expect to have 9.12% reduction in total consumption during peak hours.

Peak Load exhibits a notable discrepancy between VaR and CVaR. While VaR shows a slight reduction (0.45%), CVaR indicates a negative value of -3.50%. This implies that although there is a 95% chance of peak load shaving with a minimum reduction of 0.45%, there is a limited 5% chance that peak load could actually increase under certain conditions, which must be considered for grid stability.

HVAC+CTRL Cost and CO_2 Emission both demonstrate positive VaR and CVaR values, albeit lower than those for Total Consumption. For HVAC+CTRL Cost, VaR and CVaR are around 3.5%, suggesting that even in worst-case scenarios, cost saving is expected. Similarly, CO_2 Emission shows VaR and CVaR values of about 2.9%, indicating potential decreases in emissions even under extreme conditions.

These risk metrics provide a more comprehensive understanding of the variability and potential extremes in DR outcomes, offering valuable insights for robust policy formulation and risk management in energy systems.

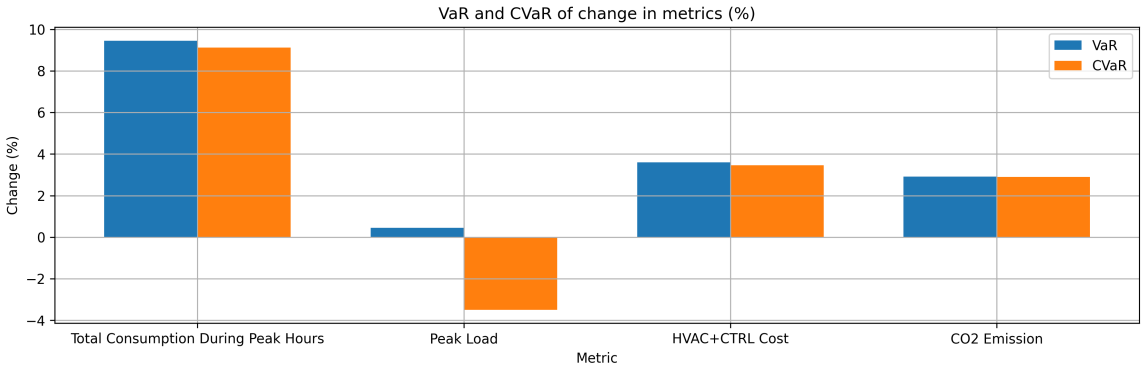

Figure 3.9: VAR and CVAR for sensitivity analysis on outdoor temperature

3.5.1 Conclusion and Future Research Directions

This study has demonstrated the efficacy of RL algorithms in optimizing DR strategies for small and medium-sized commercial buildings. Our novel approach, combining PPOD & TD3 and SACD & TD3 algorithms, outperformed traditional heuristic methods in managing HVAC systems and controllable loads.

The proposed framework achieved substantial improvements in peak load shaving, energy savings, and cost reduction, while maintaining acceptable indoor temperature levels. Moreover, our risk assessment using VaR and CVaR metrics provided insights into the framework's robustness under varying environmental conditions.

These findings underscore the potential of intelligent DR strategies to contribute significantly to grid stability and energy efficiency in the commercial building sector. However, several areas warrant further investigation to enhance the applicability and effectiveness of RL-based DR strategies:

- **Multi-zone building modeling**: Extend the current framework to account for the complexities of multi-zone buildings, including thermal interactions between zones and varying occupancy patterns. This can be achieved within a high-level model by representing each zone with simplified thermal and occupancy parameters, allowing for efficient handling of inter-zone heat transfer and aggregated occupancy effects. While micro-scale modeling offers granular insights, a high-level approach provides a scalable and practical alternative, capturing key interactions and enabling effective

energy management across multiple zones.

- **Advanced HVAC systems**: Investigate the application of RL algorithms to more sophisticated HVAC systems, such as variable air volume systems or chilled beam systems, which offer greater flexibility in temperature control.

- **Integration of EV batteries**: Explore the potential of using parked EVs' batteries as additional energy storage and grid balancing resources within the DR framework.

- **Safe RL**: Implement safe RL algorithms that prioritize constraint satisfaction (e.g., maintaining indoor temperature within strict comfort bounds) over maximizing expected rewards, ensuring occupant comfort and system safety.

- **Multi-building coordination**: Investigate strategies for coordinating DR across multiple buildings to maximize aggregate benefits at the grid level.

- **Long-term adaptation**: Develop techniques for continuous learning and adaptation of RL algorithms to account for seasonal variations and long-term changes in building characteristics or occupancy patterns.

By addressing these research directions, we can further enhance the effectiveness and real-world applicability of RL-based DR strategies, facilitating their widespread adoption in the commercial building sector and contributing to a more sustainable and resilient energy future.

3.6 Appendix

This section provides a brief overview of the RL algorithms implemented in this study for HVAC management and controllable load optimization.

3.6.1 Soft Actor-Critic (SAC)

SAC is an off-policy algorithm that aims to maximize both the expected return and the entropy of the policy. It consists of two main components:

1. Actor (Policy): Learns a stochastic policy that maximizes the expected return while also maximizing entropy.

2. Critic: Learns the Q-function to evaluate the actor's actions.

SAC introduces a temperature parameter α that balances the trade-off between exploration (high entropy) and exploitation (maximizing returns). The algorithm updates the policy to maximize the expected value of the action while also maximizing its entropy:

$$\pi^* = \arg\max_{\pi} \mathbb{E}_{s_t \sim \rho_\pi, a_t \sim \pi}[Q(s_t, a_t) + \alpha H(\pi(\cdot|s_t))] \tag{3.14}$$

Where $H(\pi(\cdot|s_t))$ is the entropy of the policy.

For our HVAC management problem, we adapted SAC to work with discrete actions (SACD) by using a categorical distribution for the policy instead of a continuous distribution.

3.6.2 Proximal Policy Optimization (PPO)

PPO is an on-policy algorithm that aims to improve the stability of policy gradient methods. PPO updates the policy in a way that ensures the new policy doesn't deviate too much from the old policy, which helps in avoiding catastrophic performance drops.

The key idea in PPO is to use a clipped surrogate objective:

$$L^{CLIP}(\theta) = \mathbb{E}_t[\min(r_t(\theta)A_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)A_t)] \tag{3.15}$$

Where $r_t(\theta)$ is the ratio of the probability of the action under the new policy to the probability under the old policy, A_t is the advantage function, and ϵ is a hyperparameter that controls the clipping.

For our HVAC problem, we adapted PPO to work with discrete actions (PPOD) by using a categorical distribution for the policy.

3.6.3 Double Dueling Deep Q-Network with Prioritized Experience Replay (D3QN)

This algorithm combines several improvements to the basic Deep Q-Network:

1. Double Q-learning: Uses two networks to reduce overestimation bias in Q-value estimates.

2. Dueling architecture: Separates the value and advantage functions to better estimate state-action values.

3. Prioritized Experience Replay: Samples more important transitions more frequently during training.

The Q-value is estimated as:

$$Q(s, a) = V(s) + (A(s, a) - \text{mean}(A(s, a'))) \tag{3.16}$$

Where $V(s)$ is the state value function and $A(s, a)$ is the advantage function.

Prioritized Experience Replay assigns priorities to transitions based on their TD error:

$$p_i = |\delta_i| + \epsilon \tag{3.17}$$

Where δ_i is the TD error for transition i and ϵ is a small constant to ensure non-zero probability.

3.6.4 Deep Deterministic Policy Gradient (DDPG)

DDPG is an off-policy algorithm that combines ideas from DQN and deterministic policy gradients. It consists of four neural networks:

1. Actor: Learns a deterministic policy $\mu(s)$.

2. Critic: Learns the Q-function $Q(s, a)$.

3. Target Actor: Slowly updated version of the actor for stability.

4. Target Critic: Slowly updated version of the critic for stability.

The actor is updated using the deterministic policy gradient:

$$\nabla_\theta J(\theta) \approx \mathbb{E}_s[\nabla_a Q(s, a)|_{a=\mu(s)} \nabla_\theta \mu(s)] \tag{3.18}$$

DDPG uses a replay buffer and soft target updates to improve stability and sample efficiency.

3.6.5 Twin Delayed Deep Deterministic Policy Gradient (TD3)

TD3 is an extension of DDPG that addresses some of its shortcomings. It introduces three key modifications:

1. Clipped Double Q-learning: Uses two critics to reduce overestimation bias.
2. Delayed Policy Updates: Updates the policy (and target networks) less frequently than the Q-function.
3. Target Policy Smoothing: Adds noise to the target action to make it harder for the policy to exploit Q-function errors.

The target Q-value is calculated as:

$$y = r + \gamma \min_{i=1,2} Q_{\theta'_i}(s', \mu_{\phi'}(s') + \epsilon) \tag{3.19}$$

Where ϵ is clipped noise added to the target action.

These algorithms were implemented and adapted for our specific HVAC management and controllable load optimization problems, allowing us to compare their performance in the context of DR in commercial buildings.

References

Afroz, Z., Goldsworthy, M., and White, S. D. (2023). Energy flexibility of commercial buildings for demand response applications in australia. *Energy and Buildings*, 300:113533.

Aghniaey, S. and Lawrence, T. M. (2018). The impact of increased cooling setpoint temperature during demand response events on occupant thermal comfort in commercial buildings: A review. *Energy and Buildings*, 173:19–27.

Arnold, G. W. (2011). Challenges and opportunities in smart grid: A position article. *Proceedings of the IEEE*, 99(6):922–927.

Aussel, D., Brotcorne, L., Lepaul, S., and von Niederhäusern, L. (2020). A trilevel model for best response in energy demand-side management. *European Journal of Operational Research*, 281:299–315.

Beil, I., Hiskens, I., and Backhaus, S. (2016). Frequency regulation from commercial building hvac demand response. *Proceedings of the IEEE*, 104:745–757.

Cai, J. and Braun, J. E. (2019). Assessments of demand response potential in small commercial buildings across the united states. *Science and Technology for the Built Environment*, 25:1437–1455.

Chen, T., Cui, Q., Gao, C., Hu, Q., Lai, K., Yang, J., Lyu, R., Zhang, H., and Zhang, J. (2021). Optimal demand response strategy of commercial building-based virtual power plant using reinforcement learning. *IET Generation, Transmission & Distribution*, 15:2309–2318.

Christantoni, D., Flynn, D., and Finn, D. P. (2015). Modelling of a multi-purpose commercial building for demand response analysis. *Energy Procedia*, 78:2166–2171. 6th International Building Physics Conference, IBPC 2015.

Christantoni, D., Oxizidis, S., Flynn, D., and Finn, D. P. (2016). Implementation of demand response strategies in a multi-purpose commercial building using a whole-building simulation model approach. *Energy and Buildings*, 131:76–86.

Darwazeh, D., Duquette, J., Gunay, B., Wilton, I., and Shillinglaw, S. (2022). Review of peak load management strategies in commercial buildings. *Sustainable Cities and Society*, 77:103493.

EIA (2021). Electric power annual 2020. Tech. rep, US Energy Information Administration.

Energy Information Administration (2021). U.s. commercial buildings energy consumption survey. `https://www.eia.gov/consumption/commercial/data/2018/bc/html/b1.php`. Online; accessed March 31, 2022.

Gao, D., Sun, Y., and Lu, Y. (2015). A robust demand response control of commercial buildings for smart grid under load prediction uncertainty. *Energy*, 93:275–283.

Hao, H., Corbin, C. D., Kalsi, K., and Pratt, R. G. (2017). Transactive control of commercial buildings for demand response. *IEEE Transactions on Power Systems*, 32:774–783.

Hao, H., Middelkoop, T., Barooah, P., and Meyn, S. (2012). How demand response from commercial buildings will provide the regulation needs of the grid. pages 1908–1913.

He, L. and Liu, N. (2017). Load profile analysis for commercial buildings microgrids under demand response. pages 461–465.

ho Lee, K. and Braun, J. E. (2016). Development and application of an inverse building model for demand response in small commercial buildings. *Proceedings of SimBuild*, 1.

Hossain, J., Saeed, N., Manojkumar, R., Marzband, M., Sedraoui, K., and Al-Turki, Y. (2024). Optimal peak-shaving for dynamic demand response in smart malaysian commercial buildings utilizing an efficient pv-bes system. *Sustainable Cities and Society*, 101:105107.

Hosseini, S. M., Carli, R., and Dotoli, M. (2022). Robust optimal demand response of energy-efficient commercial buildings. pages 1–6.

Hu, M., Weir, J. D., and Wu, T. (2012). Decentralized operation strategies for an integrated building energy system using a memetic algorithm. *European Journal of Operational Research*, 217:185–197.

Huang, T.-H., Tai, C.-S., and Fu, L.-C. (2018). Demand response in residential and commercial community considering user comfort using improved particle swarm optimization. pages 1215–1220.

Jones, C. B. and Carter, C. (2017). Trusted interconnections between a centralized controller and commercial building hvac systems for reliable demand response. *IEEE Access*, 5:11063–11073.

Kiliccote, S. (2010). Findings from seven years of field performance data for automated demand response in commercial buildings.

Kim, Y.-J. and Norford, L. K. (2016). Price-based demand response of energy storage resources in commercial buildings. pages 1–5.

Langevin, J., Harris, C. B., and Reyna, J. L. (2019). Assessing the potential to reduce us building co2 emissions 80% by 2050. *Joule*, 3(10):2403–2424.

Li, S., Zhou, X., Zhang, Z., Gao, M., Yang, W., and Shi, J. (2020). Research on energy management scheme of commercial buildings cluster considering demand response. pages 754–761.

Li, W., Xu, P., and Ye, Y. (2006). Case study of demand response in a large commercial building-a pilot program in shanghai.

Liang, Z., Bian, D., Zhang, X., Shi, D., Diao, R., and Wang, Z. (2019). Optimal energy management for commercial buildings considering comprehensive comfort levels in a retail electricity market. *Applied Energy*, 236:916–926.

Liang, Z., Huang, C., Su, W., Duan, N., Donde, V., Wang, B., and Zhao, X. (2021). Safe reinforcement learning-based resilient proactive scheduling for a commercial building considering correlated demand response. *IEEE Open Access Journal of Power and Energy*, 8:85–96.

MacDonald, J. S., Vrettos, E., and Callaway, D. S. (2020). A critical exploration of the efficiency impacts of demand response from hvac in commercial buildings. *Proceedings of the IEEE*, 108:1623–1639.

Mutule, A., Obushev, A., Grebesh, E., and Lvovs, A. (2017). Feasibility study for demand response in commercial buildings. pages 12–14.

Pallonetto, F., Jin, C., and Mangina, E. (2022). Forecast electricity demand in commercial building with machine learning models to enable demand response programs. *Energy and AI*, 7:100121.

Piette, M. A., Ghatikar, G., Kiliccote, S., Watson, D., Koch, E., and Hennage, D. (2009). Design and operation of an open, interoperable automated demand response infrastructure for commercial buildings. *Journal of Computing and Information Science in Engineering*, 9:21004.

Sartori, I., Walnum, H. T., Skeie, K. S., Georges, L., Knudsen, M. D., Bacher, P., Candanedo, J., Sigounis, A.-M., Prakash, A. K., Pritoni, M., et al. (2023). Sub-hourly measurement datasets from 6 real buildings: Energy use and indoor climate. *Data in Brief*, 48:109149.

Satchwell, A., Piette, M. A., Khandekar, A., Granderson, J., Frick, N. M., Hledik, R., Faruqui, A., Lam, L., Ross, S., Cohen, J., et al. (2021). A national roadmap for grid-interactive efficient buildings. Technical report, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States).

Son, J., Hara, R., Kita, H., and Tanaka, E. (2014). Energy management considering demand response resource in commercial building with chiller system and energy storage systems. pages 96–101.

Tang, R., Wang, S., Shan, K., and Cheung, H. (2018). Optimal control strategy of central air-conditioning systems of buildings at morning start period for enhanced energy efficiency and peak demand limiting. *Energy*, 151:771–781.

Tang, R., Wang, S., and Xu, L. (2019). An mpc-based optimal control strategy of active thermal storage in commercial buildings during fast demand response events in smart grids. *Energy Procedia*, 158:2506–2511. Innovative Solutions for Energy Transitions.

Wang, J., Wen, F., Wang, K., Huang, Y., and Xue, Y. (2018). Optimal operation of commercial buildings with generalized demand response management. pages 265–270.

Wang, Z., Li, X., Li, Y., Zhao, T., Xia, X., and Zhang, H. (2021). An optimization framework for low-carbon oriented integrated energy system management in commercial building under electric vehicle demand response. *Processes*, 9.

Watson, D. S., Kiliccote, S., Motegi, N., and Piette, M. A. (2006). Strategies for demand response in commercial buildings.

Yin, R., Kara, E. C., Li, Y., DeForest, N., Wang, K., Yong, T., and Stadler, M. (2016). Quantifying flexibility of commercial and residential loads for demand response using setpoint changes. *Applied Energy*, 177:149–164.

Yoon, A.-Y., Kang, H.-K., and Moon, S.-I. (2020). Optimal price based demand response of hvac systems in commercial buildings considering peak load reduction. *Energies*, 13.

Yuan, X., Pan, Y., Yang, J., Wang, W., and Huang, Z. (2021). Study on the application of reinforcement learning in the operation optimization of hvac system. In *Building Simulation*, volume 14, pages 75–87. Springer.

Zhao, P., Peffer, T., Narayanamurthy, R., Fierro, G., Raftery, P., Kaam, S., and Kim, J. (2016). Getting into the zone: how the internet of things can improve energy efficiency and demand response in a commercial building.

General Conclusion

This thesis has explored the application of advanced optimization techniques and RL algorithms to address critical challenges in the transition towards a more sustainable and efficient energy future. The three chapters presented focus on distinct yet interconnected aspects of this transition, introducing insights and methodologies for policymakers, industry practitioners, and researchers alike. Each chapter tackles a specific aspect of these challenges, from optimizing investment strategies in DER and V2G technologies to implementing effective DR programs for residential and commercial consumers.

By exploring these interconnected themes, the thesis provides a holistic perspective on the potential solutions and strategies for managing the complex dynamics of the evolving energy landscape. The insights gained from these studies can collectively inform the development of comprehensive energy policies and practices that promote the adoption of clean technologies, encourage prosumer participation, and foster more sustainable energy consumption patterns across various sectors.

The first chapter investigates the optimal investment strategies for DER and V2G technologies from the perspectives of both distributors and prosumers. By considering different tariff structures and stakeholder objectives, the study provides a comprehensive cost-and-benefit analysis of various investment scenarios. The findings underscore the significance of incorporating V2G technology to enhance the profitability of DER investments and highlight the importance of designing appropriate tariff structures to align the interests of different stakeholders.

The second chapter delves into the application of CPP as a DR strategy for peak load shaving in electricity grids. By integrating diverse prosumer profiles and employing RL

algorithms, the study analyzes the effectiveness of both mass and targeted CPP offerings. The results reveal the limitations of mass CPP as prosumer participation increases and propose targeted dynamic pricing strategies to counteract these challenges. The chapter also emphasizes the influential role of BATs and EVs in peak load reduction, suggesting the need for focused policy and incentive structures to encourage their adoption.

The third chapter presents a novel DR framework for optimizing electricity consumption in small and medium-sized commercial buildings. By categorizing building loads and implementing RL algorithms, the study demonstrates the potential for significant peak load shaving, cost reductions, and environmental benefits. The analysis incorporates the impact of outdoor temperature variations and risk assessments, contributing to the development of robust and efficient energy management strategies for commercial buildings.

The thesis also highlights the crucial role of batteries and electric vehicles in enabling more effective peak load management, suggesting that policymakers should prioritize these technologies in their efforts to decarbonize the electricity grid. Notably, the research finds that while photovoltaic systems are beneficial for overall energy generation, their direct contribution to peak reduction could be limited due to their dependence on solar availability, which may not always align with peak demand periods.

This thesis finds that the increasing complexity of the future grid, characterized by high penetration of DERs and decentralization, necessitates advanced management techniques. The traditional grid's predictability contrasts sharply with the variability and intermittency of renewable sources like solar and wind power in the future grid. Furthermore, the integration of prosumers, equipped with technologies like batteries and EVs, introduces diverse and dynamic energy consumption and generation patterns, adding another layer of complexity. The thesis demonstrates that effectively managing this complex and dynamic future grid environment requires moving beyond traditional optimization methods to leverage advanced optimization techniques and RL algorithms. Unlike their traditional counterparts, RL algorithms excel in handling large numbers of variables and uncertainties. The thesis shows that this adaptability is essential for real-time decision-making in the future grid, allowing for efficient peak load management even with the unpredictability of renewable energy sources and prosumer behavior. By learning from real-world data, including energy consumption

patterns and pricing signals, these algorithms can predict peak demand periods, optimize DER scheduling, and design targeted dynamic pricing strategies. This approach, proven effective in the thesis, not only ensures grid stability but also allows prosumers to maximize their benefits, marking a significant step towards a sustainable and efficient energy future.

Collectively, these chapters contribute to the growing body of knowledge on DR, prosumer behavior, and the application of advanced optimization techniques in the energy sector. The findings and methodologies presented here have the potential to inform policy decisions, guide industry practices, and ultimately accelerate the transition towards a more sustainable and resilient energy future.

However, it is important to acknowledge the limitations of these studies and the need for further research. Future work could explore the scalability and generalizability of the proposed frameworks across different geographical contexts and energy market structures. Additionally, the integration of other emerging technologies, such as hydrogen storage and smart grid infrastructure, could provide new avenues for optimizing energy consumption and production.

Moreover, the social and behavioral aspects of energy transitions warrant further investigation. Understanding the factors that influence consumer adoption of DER and V2G technologies, as well as their responsiveness to dynamic pricing signals, could help design more effective incentive structures and educational campaigns.

In conclusion, this thesis demonstrates the immense potential of advanced optimization techniques and RL algorithms in addressing the complex challenges of the energy transition. By providing a comprehensive analysis of DER and V2G investment strategies, exploring the effectiveness of CPP programs, and proposing a novel demand response framework for commercial buildings, these chapters contribute to the development of more sustainable, efficient, and equitable energy systems. As the world continues to grapple with the urgent need to mitigate climate change and ensure energy security, The insights and methodologies presented here have the potential to contribute to shaping the future of the energy sector.