The Impact of Thermostat Automation and Retail Rate Designs on Cooling and Heating Flexibility: Balancing Consumer Preferences and an Efficient Grid

by
Zack Schmitz
B.B.A, The University of Texas at Austin

Submitted to the Integrated Design and Management Program and the Department of Electrical Engineering and Computer Science in partial fulfillment of the requirements for the degrees of

Master of Science in Engineering and Management
and
Master of Science in Electrical Engineering and Computer Science
at the
Massachusetts Institute of Technology
May 2024

Authored By:	Zack Schmitz Integrated Design and Management Program Department of Electrical Engineering and Computer Science May 17, 2024
Certified By:	Pablo Duenas Martinez Research Scientist, MIT Energy Initiative Thesis Supervisor
Certified By:	Marija Ilic Adjunct Professor EECS and Senior Research Scientist, LIDS Thesis Reader
Accepted By:	Joan S. Rubin Senior Lecturer Executive Director, System Design & Management Program
Accepted By:	Leslie A. Kolodziejski Professor of Electrical Engineering and Computer Science Chair, Department Committee on Graduate Students

This page is intentionally left blank

The Impact of Thermostat Automation and Retail Rate Designs on Cooling and Heating Flexibility: Balancing Consumer Preferences and an Efficient Grid

by
Zack Schmitz

Submitted to the Integrated Design and Management Program and the Department of Electrical Engineering and Computer Science on May 17th, 2024, in partial fulfillment of the requirements for the degrees of
Master of Science in Engineering and Management
and
Master of Science in Electrical Engineering and Computer Science

Abstract

Flexibility in household energy consumption is crucial for improving grid efficiency and reducing peak electricity demand. The ongoing impact of climate change and the move towards electrification worsen these challenges, emphasizing the need for effective peak demand reduction strategies. Current approaches often involve peak pricing retail tariffs, behavioral responses to grid operator notifications, or expensive technologies such as demand-side batteries. However, these methodologies rely on unpredictable consumer participation or substantial capital investments. On the other hand, the growing use of smart thermostats presents an opportunity for passive, efficient control of household energy consumption. Combining smart thermostats with appropriate price signals creates an opportunity to optimize the balance between energy cost and thermal comfort. This work examines the role of smart thermostat automation and dynamic retail rate designs in maximizing heating and cooling flexibility while ensuring consumer comfort. The research introduces a new approach to demand-side management by using reinforcement learning (RL) to optimize thermostat settings based on individual thermal preferences and price signals. A comprehensive testbed simulation framework was developed to analyze these effects, incorporating bottom-up energy modeling, individualized thermal comfort profiles using smart thermostat data, and advanced thermostat controls to investigate the impacts of various rate designs on residential energy demand. The study evaluates these impacts at a population level, considering the effects on over 80 household archetypes across a localized region. Key findings show that partitioned time-of-use rates with moderate pricing shifts effectively reduce energy usage without creating new peaks, unlike more aggressive pricing strategies that can lead to pre-cooling-induced new peaks. These insights offer valuable guidance for policymakers and utility operators in designing rate frameworks that decrease overall electricity consumption and peak demand without compromising personal comfort.

Thesis Supervisor: Pablo Duenas Martinez
Title: Research Scientist, MIT Energy Initiative

Thesis Reader: Marija Ilic
Title: Adjunct Professor EECS and Senior Research Scientist, LIDS

This page is intentionally left blank

Acknowledgments

As Steve Jobs said, "You can only connect the dots looking backward." When I started at MIT, I had no idea what to expect. If you had given me this thesis on electric retail rate design and smart thermostats three years ago, I would have been amazed that there must have been another Zack Schmitz enrolled at MIT. This path was far from what I had planned. I also would have needed clarification about what electric retail rate design was. But three years later, my time at MIT has exceeded all my expectations. I couldn't have written a better script, and I am confident this was the path I was meant to take. So, thank you, MIT, for being the magical place that made all this possible.

Deep gratitude is extended to those who have advised and motivated this research: Pablo Duenas Martinez, Marija Ilic, Dharik Mallapragada, Morgan Santoni-Colvin, Graham Turk, and Abdelrahman Ayad. Massive appreciation also goes to Tony Hu and Andy MacInnis for taking a chance and allowing the commencement of this journey at MIT. Special thanks to Tim Schittekatte and Carlos Batlle López for teaching the class that ignited a passion for the electric grid. Heartfelt thanks to everyone at IDM and the MIT Energy Initiative, especially Grace Agosto, Jazy Ma, and Bill Foley, for keeping the wheels of IDM turning.

To the old friends who have been the backbone of support: Zach Martin, Eric Condor Samuelson, Nick Eagle Willems, Bahar Iranmehr, Christine Catinis, Abby Gibbs, Ryan Nguyen, Dave Severino, Christian Trudeau, Vig Senthilvel, Matt Tiritilli, Arjun Mocherla, Joe Robbins, Ryan Freed, Brennan Matriano, Tomy Falgui, and Cory Shreffler. You have been invaluable. And to all the new friends in Boston who have been with me on this journey: Zach Sternberg, Jo Katchinoff, Team Helio (Leandra Tejedor, Saloni Bedi, Sahas Gembali), Team Cartpool (Azfar Sulaiman, Akshita Goya, Aidana Daulbayeva), Emma Larsen, Katie Sessa, and Joanne So, thank you.

Coach Kara Lawson, though we have never met, your speech was the rallying call that inspired completing this work. Your words motivated me to handle hard better.

To the family, especially Mom, Dad, and Kendra, thank you for your love. To Koda, thank you for being the most loyal friend. And to Erica, no words can do justice to my appreciation for always sticking by me, but in short, thank you for your patience and unwavering support; you have been the rock through it all.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

The ongoing effects of climate change present an increasingly urgent challenge to household energy consumption, particularly in heating and cooling. Residential and commercial building energy consumption accounts for nearly 75% of all U.S. electricity use, making it a key component in grid demand [1]. This significant energy footprint contributes approximately 40% of all primary energy use and associated greenhouse gases (GHG) [1]. As temperatures rise globally, the strain on energy resources intensifies. An analysis of smart thermostat usage has shown that based on the current heating and cooling preferences of residential consumers, there will be a stark increase in cooling energy requirements—potentially up to 70% higher by 2050 compared to the 2019 baseline [2]. This will directly impact residential consumers' household budgets as heating and cooling systems typically account for 43% of a home utility bill [3]. The Federal Energy Regulatory Commission (FERC) predicts a nationwide increase in electricity demand by 4.7% over the next five years [4].

Due to the increase in demand from climate impacts and the continued push for electrification, one of the pressing concerns of grid operators is the reduction of peak electricity demand. Most of the U.S. grid is summer peaking due to the cooling requirements from summer afternoons when air-conditioning loads are added to relatively constant daytime loads, such as commercial lighting and industrial processes [5] [6] [7] [8]. ERCOT, or the Electric Reliability Council of Texas, has forecasted its peak demand will increase by 13.7% over the next ten years, resulting in an additional 10.9 GW of capacity needed to meet the demand [9].

Since the 1980s, utilities have been implementing demand-side flexibility (also referred to as demand-side management - DSM) strategies to address the challenges they face balancing grid supply and demand [10]. These strategies, which refer to the measures adopted by utilities, consumers, and third parties to adjust the timing and quantity of electric consumption, aim to reduce peak electricity demand during critical times by eliminating some electricity use or shifting it to non-peak times. The electricity saved through DSM is more valuable than electricity generated because after accounting for transmission and distribution along line losses, one unit saved on the consumer side is worth roughly 5% more than the unit saved on the generation side [11]. These strategies have typically been adopted by commercial and industrial customers who can be more easily coordinated than individual residential consumers [12]. Two specific types of DSM - Direct Load Control (DLC) and Demand Response (DR) - concentrate on directly controlling end-use loads or encouraging consumers to modify their usage in response to price signals such as Time of Use (TOU) electricity rates or critical peak event reduction. However, residential participation in peak demand is usually unreliable, especially on sweltering days, highlighting a shortfall if DR programs are not adequately rewarded [13]. Thus, utility companies must design incentives that maximize the tradeoffs in complexity, consumer participation, and program effectiveness. More details on these types of programs will follow in Chapter 2.

One effective way to increase demand flexibility for residential consumers is using smart thermostats that optimize the thermostat settings based on energy price signals. This helps to manage the increased demand on the grid more efficiently. However, incorporating this type of automation requires a complex understanding of the trade-off between individual thermal comfort preferences and price responsiveness. Each residential consumer has a unique idea of what they consider a comfortable temperature and what they are willing to pay for it. This represents an individualized optimization problem for each residential consumer to maximize the price-to-comfort tradeoff. A promising technique in the field of optimization is to use Reinforcement Learning (RL) to have an agent "learn" a model without having the parameters for the tradeoff explicitly defined. Through this learning (which is termed model-free), every end-user thermostat can have a control sequence learned solely based on their preferences. This type of control structure allows for demand flexibility to be personalized and the potential of heating and cooling flexibility to be maximized across participating households.

This work aims to study the impact of maximizing demand-side flexibility through smart thermostat automation alongside various rate designs and measure how they affect residential consumer demand when considering cooling and heating flexibility. To achieve this, a testbed simulator has been created that utilizes smart thermostat control, which considers a population's thermal preferences while trying to deliver optimum flexibility potential. The main contribution of this study is to provide a framework that can simulate and analyze the impacts of rate design and thermostat automation on a population's demand flexibility and energy consumption. This work creates a simulator using bottom-up energy modeling paired with advanced smart thermostat control that considers individualized thermal comfort preferences. It does not focus on technology adoption beyond smart thermostat investment (which most utilities offer steep discounts on to promote adoption [14]) but instead on the operational levers that can be used in the short term to support grid efficiency and influence the climate change transition. The work evaluates how creative rate design can reduce greenhouse gas emissions and mitigate climate impact without requiring immediate substantial investment. Furthermore, given the wide variety of comfort preferences across households, it also assesses the effectiveness of thermal flexibility at a regional scale. By evaluating this potential, the study provides policymakers and utility operators with actionable data to create more effective demand response strategies and rate frameworks that optimize energy consumption in line with the grid's needs while maintaining consumer comfort and affordability. This study contributes a forward-thinking perspective to the contemporary discourse on energy management, innovative technology, and sustainable living.

The rest of this thesis will be laid out in the following structure: Chapter (2) will focus on the background of the critical concepts covered in the thesis, Chapter (3) will introduce related work in the field and extensions on current methods, Chapter (4) presents the methodology and implementation, Chapter (5) provides the technical details about the simulation environment and data, Chapter (6) outlines the key results and discussion points and finally Chapter (7) concludes the work, provides an overview of limitations and addresses future work considerations.

Chapter 2. Background

2.1 Demand Flexibility

Demand flexibility creates substantial opportunities for grid efficiencies. A meta-analysis reveals that flexible operations exhibit varying performance levels depending on the technology and application scope. Peak power reductions range from 1% to 65%, while energy savings can reach 60%. Additionally, flexible operations can lead to operational cost reductions between 1% and 48% and greenhouse gas emission reductions of up to 29% [15]. Another analysis observed that when the outside temperature goes beyond 75°F (24°C), adjusting the thermostat setpoint leads to a proportional increase in the magnitude of zone temperature setpoint adjustments [16]. This is particularly important on days with higher demand for hotter outside temperatures. During such conditions, it is estimated that a 2-degree setpoint change can produce a load reduction potential from 20% to 35%.

As mentioned previously, demand flexibility has two primary strategies used to try and match supply and demand constraints: Direct Load Control (DLC) and Demand Response (DR). Direct Load Control programs allow utilities to control specific appliances directly and remotely in consumers' homes during peak energy demand, such as air conditioners or water heaters. These programs can help reduce the load on the grid by allowing the utility to turn off appliances or even control thermostat setpoints [17]. Participants in DLC programs often receive financial incentives or reduced utility rates to enable the utility to control their appliances during these critical periods. A study conducted on incentives for demand-side management revealed that people value the option to override more than financial incentives [18]. They are willing to participate if they can override based on their preferences and do not want to lose control. Recently, this concern was validated when Xcel Energy prevented manual overrides, causing 22,000 people in Denver, Colorado, to lose control of their home thermostats [19].

On the other hand, DR is a method that encourages consumers to adjust their electricity usage during peak periods in response to price signals or financial incentives. DLC is where the utility directly controls the consumer's energy usage; DR is based on the consumer's choice and flexibility. The users can decide how and when to reduce their energy usage through time-based rates, including time-of-use pricing, critical peak pricing, variable peak pricing, real-time pricing, and critical peak rebates [20]. Demand Response is a more consumer-friendly approach as it gives them autonomy over their consumption and the opportunity to contribute actively to grid reliability and efficiency. However, because of this flexibility, it is often difficult for utility companies to accurately predict or quantify the impact of any consumer-focused action. For instance, in a study examining the implementation of thermostat-based demand response programs, it was observed that the rate at which participants overrode the program varied greatly, ranging from 1% to 39%

[21] [22] [23]. This inconsistency has significant implications for load-shifting efforts and poses a challenge for electrical utilities. The high rate of overrides can lead to missed opportunities for reducing demand, which can significantly shape the effectiveness of a program. This problem can be solved by improving the program's design and delivery and increasing household participation to reduce the variability in aggregate demand response from these loads. This leads to better predictability and reliability when using Thermostat Controlled Loads (TCLs) for grid services [24]. One way to improve predictability is by designing a better retail rate system. This can condition residential users to follow a more consistent pricing schedule rather than relying on reducing consumption during critical peak events.

2.2 Electricity Pricing in the United States

In the United States, the cost of electricity in wholesale markets fluctuates based on various factors, such as the time of day, demand, and the cost of generating electricity from different sources. However, most end-use consumers are not directly exposed to these fluctuating prices. Instead, they typically pay a fixed price per kilowatt-hour (kWh) for their electricity consumption, regardless of the actual cost of producing and delivering that electricity at any given time [25]. This fixed pricing structure does not reflect the true variability in the price of electricity production, which can lead to inefficiencies in the electricity market.

Moreover, the cost of producing electricity isn't the same for every unit of electricity generated. When demand is high, the cost of producing additional electricity increases dramatically. This is because some power plants can't quickly increase their electricity when demand suddenly rises. And even for power plants that can quickly ramp up production, like those that run on natural gas, the cost of producing this additional electricity is often much higher.

Because of these factors, the price of electricity can vary a lot from one hour to the next. Sometimes, especially during certain times of the year when demand is very high, the price of electricity can spike to levels that are more than ten times the average price. It varies depending on the amount of electricity used and how much it costs to produce that electricity at a given time. Different retail rate designs have been implemented to manage fluctuating electricity prices and reduce peak demand. These designs aim to communicate the price to consumers without exposing them to real-time market pricing. The goal is to provide consumers with price signals that encourage them to adjust their energy usage without facing the risks associated with pricing fluctuations, such as during the Texas winter storm in 2021 when the price of electricity skyrocketed to $9/kWh [26].

To that end, there has been some debate regarding the efficacy of different rate structures, but recent reports have shown that thoughtfully crafted Time-of-Use (TOU) rates, especially when paired with a Critical Peak Pricing (CPP) initiative of infrequent periods of elevated costs, are more effective than previous research suggested [27] [28]. There has even been success in exposing residential consumers to modified hourly prices that fluctuate with the real-time cost of electricity but protect them from extreme events. For instance, ComEd's Hourly Pricing program had 37,578 participants in 2022 and provided a net benefit of over $6 million from a societal and environmental perspective. Participants saved an average of $20 (2%) per year compared to ComEd's default fixed-price rate ("Rate BES"). Participants saved an average of $28 on their bills due to conservation efforts, resulting in a total average bill savings of $48. In 2022, the total supply savings from the combined bills of all Hourly Pricing participants was $729,306 [29].

Improved retail electricity rates do more than provide a clear understanding of the cost of electricity at different times. These rates are also crucial to efforts to reduce carbon emissions. By aligning retail rates more closely with the actual costs and demands of the system, it becomes easier to make household electrification affordable on a large scale [30]. This creates an opportunity to speed up the adoption of clean energy, which is both economically and environmentally sustainable. Adopting new end-use technologies at the household level directly impacts electric rates that accurately reflect system costs [30]. When consumers see the financial benefits of rates, they are more likely to adopt energy-saving technologies. This can lead to widespread behavioral change and significant energy savings. At a planning and system level, better-aligned rates can help avoid grid costs that directly impact customers' rates. These include generation, transmission, and distribution capacity costs, subject to different timing and locational constraints. Better alignment can help avoid the need for all three types of investments [31].

Rate treatment	Number of observations	Avg peak demand reduction	Avg reduction in overall consumption	Median peak demand reduction	Median reduction in overall consumption
CPP	13	23%	2.8%	23%	2.6%
PTR	11	18%	2.3%	18%	0.6%
TOU	17	7%	1.2%	6%	1.0%
TOU+CPP	8	22%	2.1%	20%	2.3%
TOU PTR	1	18%	7.4%	18%	7.4%
All	50	16%	2.1%	14%	1.3%

Table 2.1: Results from rate design studies showcasing the potential to reduce consumption and peak load [32]

2.3 Retail Rate Designs

The most common types of retail rates include:

Real-time pricing (RTP)

Under this pricing scheme, the cost of electricity is determined hourly based on the actual cost of delivering it to consumers. This means the prices can fluctuate significantly from one hour to the next and from day to day. For instance, during a heatwave, when the electricity demand is high, the prices may rise to several times the average price. While this scheme can encourage consumers to shift their electricity usage to lower-priced hours, it can also subject them to more significant uncertainty and risk.

Time-of-use (TOU) rates

Time-of-use (TOU) rates divide the day into off-peak, mid-peak, and on-peak periods. Each period has a different price per kilowatt-hour (kWh), with on-peak prices being the highest and off-peak prices being the lowest. Unlike Real-Time Pricing (RTP), TOU rates offer a more predictable and stable pricing structure since the cost for each period stays the same over a more extended period, such as a season or a year. This allows consumers to plan their electricity usage more effectively and take advantage of lower-priced periods.

Fixed rates

This rate is mainly adopted by residential consumers today. They pay a fixed rate for electricity consumption, irrespective of the time of day or the actual cost of electricity production. This rate offers simplicity and predictability for consumers, but it does not encourage changes in consumption behavior based on the varying costs of electricity production.

Critical peak pricing (CPP)

Critical Peak Pricing (CPP) is a combination of Time of Use (TOU) rates and Real-Time Pricing (RTP). Under CPP, customers pay TOU rates for most of the year, but during a limited number of "critical peak" days, when the electricity demand is exceptionally high, prices increase significantly for a short time. Utilities typically notify consumers a day before these critical peak events, allowing them to reduce their consumption and avoid paying higher prices.

Flat Rate

This pricing structure is very straightforward. Consumers pay a fixed monthly fee for their electricity usage, regardless of how much they use. This fixed rate provides consumers with complete predictability as their monthly costs remain constant, allowing them to plan their expenses without worrying about changes in electricity usage or prices. However, this pricing structure does not encourage energy conservation since the payment amount remains constant, irrespective of how much energy is consumed.

Several factors influence the efficacy of retail rate designs in promoting demand flexibility and reducing overall electricity consumption, such as consumer awareness, the availability of enabling technologies such as smart meters and thermostats, and the specific design features of each rate structure.

2.4 Factors that Contribute to Thermal Comfort and Personal Preferences

Thermal comfort is a highly subjective experience that can vary significantly from one individual to another, even when they are in the same environment. A detailed study examining various climates and operational modes suggested that the comfortable temperature ranges can be as comprehensive as 51.3°F to 86.4 °F. In temperate climates, the average comfortable temperature is 73.2 °F; in tropical climates, it is 76.1 °F; in sub-tropical climates, it is 74.1 °F; in continental climates, it is 71.1 °F; and in polar climates, it is 51.3 °F [33]. This variability is due to a complex interplay of physiological differences, personal preferences, behavioral patterns, and environmental conditions [34]. Physiological factors, such as skin temperature, heart rate, and metabolic rate, can influence an individual's sensitivity to heat or cold. Personal factors, like clothing insulation, age, sex, and health status, also play a role in determining comfort. Behavioral aspects, such as fans, heaters, or windows, can further modify an individual's perception of comfort. Finally, environmental factors, including air temperature, humidity, and air velocity, contribute to the overall thermal experience.

Given the wide range of factors that influence thermal comfort, it is essential to recognize and accommodate individual preferences when operating indoor environments. Failing to do so can lead to discomfort, dissatisfaction, and potentially reduced productivity among occupants [35]. Adhering to thermal comfort preferences extends beyond mere comfort. It can also significantly impact energy consumption in buildings. When designed and operated without considering individual preferences, HVAC systems may waste energy by overheating or overcooling spaces. Recognizing the diversity of thermal comfort preferences and developing strategies to accommodate them can lead to more responsive, efficient, and comfortable indoor environments.

2.5 Thermostat Setpoint Control Methods

Controlling heating and cooling systems is crucial for indoor environmental quality. Several control methods have been developed, each with strategies and technologies to manage these systems' operations. The most common types of control algorithms are:

Rule-based controllers (RBC)

These are the most traditional forms of HVAC controls widely used in commercial and residential buildings. Rule-based systems operate on predefined rules or conditions ("if-then-else"), typically using thermostats and timers to control the HVAC components based on specific inputs like temperature and time. For example, a rule might specify that heating should be turned on when indoor temperature drops below a certain point. These controllers are straightforward but can be inefficient as they do not adapt to changing conditions or learn from past performance. Typically, there is little optimization at the entire building level because it would require an incredibly complex RBC controller for each scenario, and it's almost impossible to create general rules for the entire building [36].

Model Predictive Controllers (MPC)

Recently, MPC has emerged as a popular alternative to rule-based control because it can handle multi-input, multi-output systems and readily integrate constraints [37]. MPC is an advanced optimization method that uses a model of the HVAC system to predict future conditions and make decisions that optimize performance over a time horizon [38]. This approach can significantly reduce energy consumption and improve comfort by anticipating future needs and adjusting control actions accordingly. MPC requires a detailed mathematical model of the HVAC system and external factors like weather conditions. Despite its potential for high efficiency, the complexity of developing accurate models and the computational demands have limited its widespread adoption [36]. Therefore, MPC is better suited for regulating lower-level processes and components of HVAC systems where the control was designed along with the system, modeled in advance, and not influenced by dynamic behavior [39].

MPC is based on model-based learning, which relies on a detailed model of the environment to simulate and predict future states. The control actions are then optimized based on the predictions made by the model. This approach is advantageous when the system dynamics are well-understood and can be appropriately modeled. However, the effectiveness of model-based algorithms heavily depends on the model's accuracy, which can be challenging to achieve in complex systems or when considering the need for adaptive tradeoffs for individuals.

Reinforcement Learning Controllers

Reinforcement learning (RL) controllers are algorithms that do not require a pre-existing model of the HVAC system. In contrast to model-based learning, model-free algorithms used in RL do not need a comprehensive model of the system [40]. Instead, they learn directly from their interactions with the environment. They use a form of machine learning where the controller learns to make decisions through trial and error using a reward system. This method is beneficial when it is challenging to model the system accurately or highly dynamic and unpredictable. RL controllers can be adaptive across a range of inputs without having to optimize performance in a particular environment. This allows for more exploration across a range of actions of where personalized dynamics are at play.

Trade-off MPC vs. RL

When considering control algorithms, studies have shown that Model Predictive Control (MPC) slightly outperforms Reinforcement Learning (RL) methods [41] [42] [43]. However, MPC often uses linear or quadratic objective functions, which limits its application to large-scale problems. Every aspect added for optimization increases the dimensionality of the problem space, thus increasing the computation time needed [39]. In contrast, RL does not restrict the reward signal, which allows for considering short-term and long-term reward structures. Work has also been done to combine RL's flexibility with MPC's performance, but this still requires defining an optimization function for MPC and encoding the proper predictive model per environment [44] [45] [46]. Moreover, the studies highlighting MPC's superiority are focused on demonstrating the potential of a single building rather than scaling the implementation regionally.

Given that the outcome of this research is focused on population-level impacts of rate design, the RL controller method provides an adaptive way to differentiate environments without the need for direct domain knowledge encoding. Although MPC is more accurate, the performance of the Reinforcement Learning (RL) method is still reliable for implementing an adaptive control algorithm with similar results (within 5% of the MPC controller for energy savings) [47]. Using the RL approach requires fewer computational resources (it does not require a known model for every implementation), making it a better choice for prototyping and faster iteration of various environments and reward structures (rate designs)[48]. This eliminates the need for an optimized method for each household or end-user thermal comfort preference.

Chapter 3. Prior and Related Work

Model predictive control (MPC) and reinforcement learning (RL) for building control have been a topic of interest in the scientific community for some time. However, only a few studies have been conducted to evaluate how varying time rates may affect these types of automation. Moreover, to the best of my knowledge, very few studies have examined the effects of such controls on the broader population and how individual preferences impact the outcomes of control optimizations. Below is an analysis of the current landscape of research that covers these topics.

3.1 HVAC Control Methods

Model Predictive Control (MPC)

As mentioned before, several studies have applied MPC methods and have seen significant energy savings while maintaining thermal comfort for residents. A study found that using the MPC controller reduced energy consumption by 9.7-25% and cost from 8.2% to 18.2%. Despite this, the controller could maintain the temperature within the personalized comfort band [49]. That research team also integrated demand response, occupancy, and occupant behavior within the controller framework. In addition, a meta-analysis of multiple MPC controllers found that incorporating them as part of control objectives resulted in significant reductions in peak loads, typically around 30% [36]. This was achieved by explicitly formulating the objective function or implementing an indirect variable energy prices policy. However, an excessively conservative MPC algorithm can significantly reduce the anticipated flexibility potential without significantly improving thermal comfort satisfaction [50]. If not executed correctly, the parameters for optimization in the MPC approach, which necessitates a higher level of environment-specific optimization than defining a single reward function (as in RL), may cause issues.

Reinforcement Learning (RL)

Due to the nature of AI research, many researchers have developed methods for utilizing RL in building HVAC control. When applied to commercial and office simulations, results showed that it is possible to achieve significant energy savings by implementing reinforcement learning techniques while keeping thermal comfort levels within acceptable limits. Additionally, when reinforcement learning is applied with demand response, average power consumption can be reduced (or increased) by up to 50% while maintaining a comfortable environment [51]. Another implementation in office buildings that utilized TOU rates showed the RL method can reduce total energy cost by 75.25% over the rule-based methodology [52].

Similar results have been replicated in the residential sector. A study used multiple neural networks to predict comfort and control thermostat usage. The results showed an improved performance of thermal comfort prediction by 14.5%, reduced HVAC energy consumption by 4.31%, and improved the occupants' thermal comfort by 13.6% [53]. An additional application was conducted where an RL control strategy was implemented in a multi-zone household. The results showed that machine learning reduced energy consumption costs by 15% compared to rule-based methods. It also reduced comfort violations by 79% and complaints by 98% [54]. However, these studies have some drawbacks. They focus on single household simulations and static environments and frame the reward function to optimize consumption and comfort without considering price as a variable. Moreover, a meta-analysis of utilizing RL in home energy management showed that over half of the implementations have used a discrete observation space and action space, which oversimplifies the complexity of the building environments where the simulations occur and the available action sequences to smart devices [55].

3.2 Large Impact Analysis Using Simulators

While less developed as a research area, some work has been done to understand the broader scaled-level impacts of control models and alternative methods for demand flexibility.

DyMonDS for ALM [56] [57]

In previous work, Joo and Ilic introduced the Dynamic Monitoring and Decision Systems (DyMonDS) framework for Adaptive Load Management (ALM), which aimed to integrate different energy resources and demand responses in the power grid. The framework focuses on exchanging information at multiple levels and directions to optimize interactions between market participants by using demand bid curves provided by end-users through load aggregators. Similar to the method proposed in this thesis, this system dynamically adjusts electricity consumption based on real-time price signals to optimize user comfort and grid stability. In DyMonDS, the economic value perceived by each end-user is represented through a demand function closely related to the marginal benefit or willingness to pay. However, determining these parameters can be challenging due to limited knowledge of electricity demand price-responsiveness, requiring iterative experiments with real-time pricing to induce and measure price elasticity. In contrast, this thesis proposes an innovative approach by encoding the energy-comfort tradeoff in an RL framework instead of using predefined demand functions. This approach offers several advantages. Firstly, RL does not require explicitly defined functional parameters, making adapting to each household's unique thermal comfort preferences and behaviors easier. Secondly, RL can handle probabilistic thermal comfort profiles, accommodating the inherent variability and uncertainty in user preferences and environmental conditions. This adaptability provides a more robust and efficient HVAC control mechanism for enhancing demand response effectiveness and user satisfaction.

Network Optimized Distributed Energy Systems (NODES)[58]

A recent project developed by the MIT Lincoln Laboratory has introduced a sophisticated multi-layered control model that is specifically designed for HVAC systems. This model focuses on enhancing energy management and operational efficiency at the feeder level. The core of this approach is the integration of two control techniques - Model Predictive Control (MPC) and Sliding Mode Control (SMC). This configuration allows for proactive energy usage planning and real-time operational adjustments rather than modeling price-reactivity, which is where this research effort differentiates itself. In addition, the NODES study was designed for 25 analogous homes in the Pecan Street Community and utilized a fixed energy cost of $10/kWh rather than exploring alternative pricing opportunities.

The impact of energy-efficiency upgrades and other distributed energy resources on a residential neighborhood-scale electrification retrofit (NREL) [59]

The NREL project analyzed the effects of multiple residential building electrification retrofit scenarios, including energy efficiency upgrades, heating and hot water electrification, and distributed resource integration. They used a hypothetical community of 30 single-family homes in Denver to analyze metrics such as energy use, carbon emissions, utility bills, peak demand, and distribution grid voltages for each scenario. The NREL study is based on a centralized optimization scheme with perfect knowledge of building and equipment models using an MPC. On the other hand, this research effort diverges from the NREL study to develop a more personalized and realistic approach using reinforcement learning. For this research, the control agent seeks to learn each household's unique preferences and thermal flexibility through interaction without requiring an explicit forecasting model. Moreover, this work focuses on advanced rate structures rather than a broad home electrification analysis, which requires homeowners to invest capital for house envelope and technology upgrades. This research aims to determine the potential for aggregating thermal demand flexibility through rate-specific incentives while respecting the diversity of individual household preferences and behaviors.

Centralized Planning Agents

Additional efforts have developed learning environments to study group-level impacts through the lens of a centralized planner [60]. For instance, CityLearn is an open-source gym environment for implementing Multi-Agent Reinforcement Learning (RL) for building energy coordination and demand response in cities [61]. CityLearn enables users to implement reinforcement learning agents quickly in single and multi-agent settings to control active energy storage for load shifting and heat pump or electric heater power for load shedding. In a particular scenario utilizing CityLearn, a deep RL controller was developed and tested against a manually optimized RBC. According to the results, the deep reinforcement learning algorithm can potentially reduce clustered populations' overall electricity costs while decreasing the peak energy demand by 23% and the peak-to-average ratio (PAR) by 20%. However, it's worth noting that CityLearn has limited control structures and only allows for manipulating heat pumps, not HVAC equipment. Similarly, AlphaBuilding ResCommunity has implemented a gym environment to train thermostat-controlled load (TCL) algorithms to shift the load and provide grid services [62]. One limitation of their implementation is that they utilize an "ON/OFF" control framework in a discrete action space instead of a continuous action space that allows for more granular control. Lastly, it is also worth noting that these implementations model control from a centralized planner point of view instead of implementing personalized control algorithms that maximize the potential for each environment.

3.3 Demand Flexibility

Various attempts have been made to quantify demand flexibility, but the results are often difficult to quantify outside of real-world implementations due to the stochasticity of consumer responses and the lack of generalizing methodologies to expand to a more significant number of households [63]. A study comparing a Deterministic MPC-based approach versus a Stochastic MPC approach found that flexibility potential was overestimated due to neglecting uncertainties, which is often the case when research attempts to model flexible resources [64] [50]. However, there have been some promising results. One study developed a new design for residential thermostats to eliminate the need for a dead band and use discrete-time control to enhance the aggregate load response to real-time pricing. The proposed thermostat was found to give the residential load an energy demand elasticity between 10% and 25% during peak times [65]. However, it should be noted that this method was only validated on a single simulated residence.

An additional experiment using multi-task DDPG (Deep Deterministic Policy Gradient (DDPG) is a specific algorithm type in Reinforcement Learning) showed lower total energy costs than the rule-based control strategy for both peak/off-peak pricing and PJM market pricing in cooling and heating scenarios. The cost savings ranged from 6.1% to 10.3% across the different pricing schemes and HVAC modes. However, the DDPG method did result in slightly longer temperature violation times in some cases compared to the rule-based approach, but the average violation magnitude was only around 0.6 degrees Celsius [66]. In another simulation with a standardized prototype single-family home model, the authors conducted a simulation to measure the possibilities of demand flexibility in residential space heating under various scenarios of phase change material (PCM) thermal storage and control across the regions of the US with high heating demand. The results showed that peak-load shifting of up to 98.5% was achievable, with electricity cost-saving/revenue potential reaching 338.3% of electricity cost reductions for US residential heating [67]. While encouraging, these studies' limitations are their lack of understanding of each consumer's preferences and the tangible impacts of various reward incentives for demand response. To understand a complete picture, it's essential to consider how end-user preferences will manifest in their behaviors towards any flexibility measures.

3.4 Modeling with Thermal Comfort Implications

When modeling setpoints for thermal comfort, many of the papers reviewed above, regardless of whether they are MPC or RL, choose either a fixed set point band or synthetically generate them. For example, four different probabilistic occupant behaviors were randomly selected in one approach to represent diverse behaviors and test the algorithm under extreme cases [68]. In another study, the authors categorized the comfort temperature range as 23.0°C − 26.0°C in summer and 20.0°C − 23.0°C in winter [69]. The AlphaBuilding ResCommunity uses temperature data from the ASHRAE Global Comfort Database. This database provides objective measurements of the indoor environment, along with subjective evaluations from people who occupy buildings worldwide [70]. While this dataset is better than setting fixed bounds, it is not as regionally condensed to account for local context, distribution, and density differences. In NREL's neighborhood study, they generate synthetic setpoint schedules from Resstock End Use Load Profiles [59] [71]. Still, an analysis comparing ecobee's smart thermostats and the Resstock profiles highlighted some loads were underestimated by as much as 40% [72]. Instead of using fixed schedules or synthetic data, this study incorporates the variation in thermostat setpoints, including seasonality and home / away schedules, from a group of smart thermostats in one region. The data will be used to analyze behavioral patterns that determine the capacity for demand flexibility across the region.

3.5 Novelty

This research builds upon the existing work in HVAC control methods, population-level analyses using simulators, and demand flexibility. It introduces innovative contributions to the field, centered around a novel simulation framework that is designed to push heating and cooling flexibility to its capacity through thermostat automation and rate design and includes:

1. Quantitative analysis of different rate designs on demand flexibility at a population scale and how pricing strategies influence grid efficiency when smart thermostats are controlled

2. Integrating individualized energy-comfort tradeoffs into reinforcement learning models to allow for personalized HVAC control that accounts for unique consumer behavior, preferences, building environment characteristics, and multiple different set point schedules

3. Incorporation of smart thermostat data as thermal comfort profiles to analyze and predict demand flexibility under different retail tariffs across a population while maintaining the variability of consumer responses to energy pricing.

4. Methods

The methodology for exploring the effects of time-varying rates using an RL model is outlined below. The primary implementation consists of 1) a bottom-up resistance capacitor (RC) building simulator, 2) a thermal comfort model that estimates the probability of an individual degree of comfort based on ecobee smart thermostat data, and 3) an RL controller that determines the continuous heating and cooling setpoint for each environment. The methodology contains several tools to encode specificity and domain knowledge that increase the fidelity of the data modeling. The following sections describe the system in more depth.

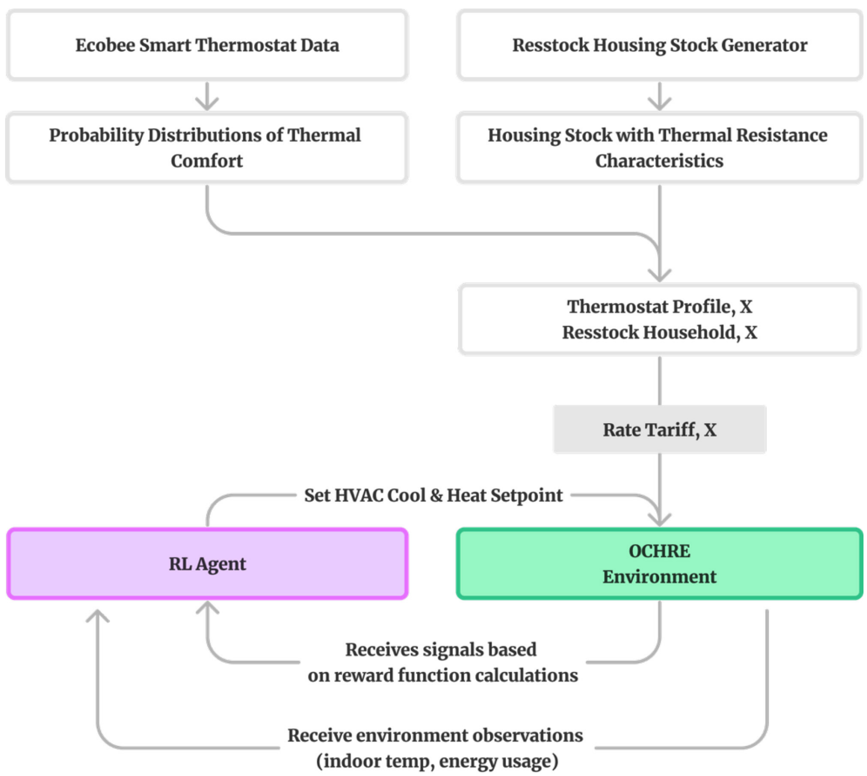


Figure 4.0.1: Overview of the proposed simulation framework

4.1 Bottom-up Energy Modeling

Defining Bottom-up Modeling in Energy Systems

Bottom-up modeling is an analytical method that begins with the smallest level of detail, i.e., individual households and builds a comprehensive representation of the entire energy system. Unlike top-down models that use general assumptions and aggregate data to analyze energy consumption and forecast demand at the macro level, bottom-up models allow for increased fidelity at the individual component level of the system, which, in this case, is the residential consumer.

The granularity of bottom-up modeling allows for high precisions in representing energy consumption. It encompasses detailed energy usage characteristics, including the timing, location, housing build components, and behavior of individual loads and resources. This method is particularly valuable because it captures the complexity and variability inherent in energy use, reflecting the heterogeneity of appliances, human behavior, and system dynamics. Bottom-level modeling provides a robust foundation for developing highly accurate energy management strategies and simulating the potential impact of demand response and other interventions by capturing the nuances of individual devices and usage patterns.

Object-oriented Controllable High-resolution Residential Energy (OCHRE™) Model

Within the scope of this work, the building energy simulator utilized is the Object-oriented Controllable High-resolution Residential Energy (OCHRE™) model. OCHRE™ is a reduced-order residential building model that aims to maintain essential system dynamics without the complexity of detailed simulation models. At its core, it represents the building's envelope through a multi-node equivalent circuit, commonly known as an RC model.

The RC model is an abstraction that uses electrical analogs—resistors (R) and capacitors (C)—to mimic the thermal behavior of a building's envelope. Resistors represent the resistance to heat transfer (akin to the insulation of walls), while capacitors symbolize the building's ability to store heat. This representation creates a simplified yet effective way to model how quickly and effectively a building can respond to temperature changes both internally and from the external environment, which is critical for energy usage predictions. OCHRE™ allows for the control of significant loads and Distributed Energy Resources (DERs), including space conditioning systems, domestic hot water systems, batteries, photovoltaic (PV) systems, and electric vehicles (EVs). Each element can respond to control signals, meaning the model can simulate the response to various demand-side management strategies.

The OCHRE model was validated against the industry standard EnergyPlus simulation tool. To do this, the model developers ran annual simulations at a 1-minute resolution in OCHRE and a 10-minute resolution in EnergyPlus. The results showed that the two models had a reasonable agreement. Regarding HVAC heating, OCHRE underestimated the energy consumption by 8.7% compared to EnergyPlus. On the other hand, OCHRE slightly overestimated the energy consumption of HVAC cooling by 0.8%. The differences in HVAC-delivered heating and cooling were minor (i.e., the temperature management services provided by an HVAC, including heating or cooling of spaces, maintaining indoor air quality, and controlling humidity.). The authors suggest that most of the discrepancy was caused by an overestimation of OCHRE's air-source heat pump efficiency, and differences in solar radiation heat transfer through the windows might contribute to more heating and less cooling in OCHRE compared to EnergyPlus. Based on these results, the differences will not significantly impact the overall outcomes of the study [73].

Resident Schedule Stochasticity in the OCHRE Model Using Resstock Schedules

Accurately reflecting human activities' unpredictable and variable nature is crucial for realistic residential energy modeling. Traditional approaches often rely on deterministic, homogeneous activity schedules that do not capture the complexities of daily human behavior. As a result, inaccuracies can occur in modeling electricity demands and assessing the effectiveness of energy-saving technologies. To enhance the realism and accuracy of simulations, the OCHRE model incorporates the advanced stochastic event generator developed by the National Renewable Energy Laboratory (NREL) [74]. This advanced approach utilizes a combination of time-inhomogeneous Markov chains and probability sampling of event durations and magnitudes. Markov chains are used to model the likelihood of transitioning from one activity state to another based on the current state, capturing the sequential nature of human behaviors throughout the day. Probability sampling complements this by varying the lengths and intensities of each event, thereby introducing realistic randomness into the daily routines modeled in the simulations. These realistic behavior patterns better assess the potential flexibility of energy usage in response to price signals or other demand-response initiatives.

Building Representations

The OCHRE™ model employs a consistent and standardized approach to define building properties by using the Home Performance eXtensible Markup Language (HPXML). HPXML serves as the infrastructure guidelines for all relevant building inputs required for the energy model of a residential building [75]. Resstock (another tool developed by NREL for bottom-up modeling) generates the HPXML input files, which utilize NREL's U.S. Building Stock Characterization Study to generate U.S. home profiles [76]. In this model, the United States housing stock is segmented into 165 distinct subgroups based on critical factors such as climate zone, wall structure, housing type, and year of construction to create building profiles that accurately reflect the diverse conditions of U.S. homes. For each subgroup, thermal energy use is quantified, including energy needed for heating, ventilating, cooling, and water heating, broken down by end use and segment. Resstock uses a down-sampling method to reflect the vast diversity within the U.S. housing stock, comprehensively capturing variations in building types. This approach ensures that simulation-building profiles represent regional diversity, enhancing the model's applicability and accuracy in various geographic and climatic contexts. Using the housing profiles from NREL and Resstock, this research can model unique regional characteristics in building design, construction practices, and climate conditions See Figure 4.1.1 for an overview of different building types and key segmentation characteristics.

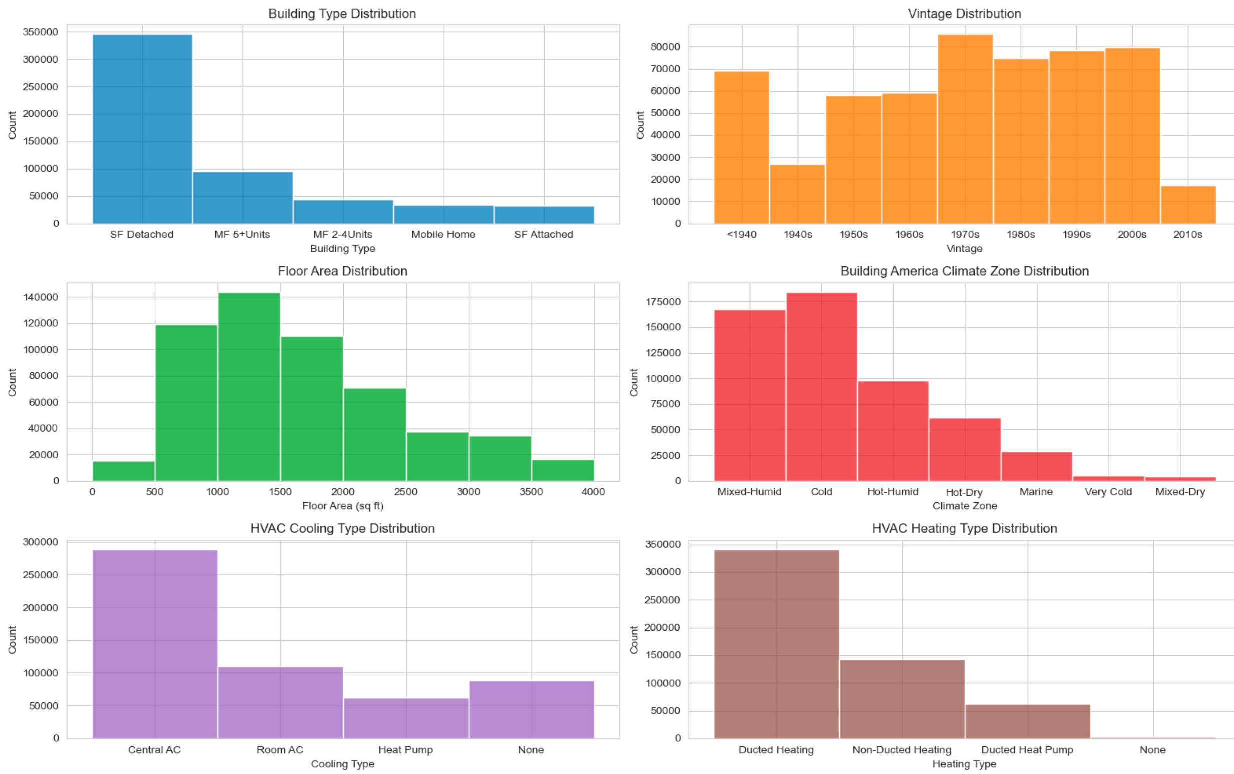


Figure 4.1.1: A high-level summary of the different types of buildings available in the Resstock dataset, along with summary distributions on key characteristics

Preparing OCHRE™ for Reinforcement Learning

To implement the OCHRE model with reinforcement learning (RL), it was necessary to configure it to function within a Gymnasium environment. Gymnasium is a widely used framework that employs a high-level Python class to represent a Markov Decision Process (MDP) flow from reinforcement learning theory and allows for creating custom environments that can be used to train RL algorithms [77]. The team at NREL developed the initial building blocks for using OCHRE within a Gymnasium environment, but significant modifications were needed to tailor the system to the specific needs of this research [78]. These adjustments included expanding the observation space to capture a broader range of environmental inputs, modifying the action space to allow for more nuanced control strategies, and reconstructing the reward function to better align with the desired energy efficiency outcomes. More information about the implementation of RL and how it was applied within this modified framework will be discussed in the subsequent sections.

By utilizing the OCHRE™ model, this research can simulate the behavior of a residential energy system with high accuracy and specificity. This allows for modeling energy usage and reactions to time-varying rates with fidelity that other modeling approaches may not achieve, especially for individualized and adaptive control schemes. A complete description and documentation of the OCHRE™ model and its capabilities are available online [79].

4.2 Thermal Comfort Modeling

About the ecobee Dataset

The modeling of thermal comfort probabilities relies on data from ecobee's Donate Your Data program [80]. ecobee Inc.'s customer base has voluntarily shared anonymized records to support various research endeavors. The dataset contains environmental and operational data obtained from their connected thermostats. The dataset captures five-minute interval readings from thermostats and any linked remote sensors, providing a detailed temporal snapshot of indoor climatic conditions, HVAC system operations, and occupancy. The dataset spans from each thermostat's initial connection to the current day. In addition to the interval data, the dataset is enriched by user-provided metadata describing housing characteristics, HVAC equipment details, and geographical information such as state or province when available. The substantial breadth of data offers a robust foundation for modeling thermal comfort, and the dataset presents an exciting opportunity to model thermal comfort based on the exact indoor temperature instead of just heating and cooling set points. Furthermore, it provides a high density of data points per geographic area representing diverse user preferences and environments. A complete list of the timeseries and metadata points available in the dataset is listed in the Appendix.

The Thermal Comfort Model

Thermal comfort modeling concentrates on understanding the thermal preferences of people occupying a space. Typically, these preferences are acquired through one of two methods: active solicitation of feedback, where occupants explicitly report their comfort levels, or passive inference, where the system deduces preferences based on how occupants interact with their heating, ventilation, air conditioning (HVAC) —for example, adjustments to temperature settings or changes in usage frequency and timing [81].

Active methods involve directly asking residents for feedback or using interfaces to record their comfort levels. In contrast, passive methods track adjustments made to environmental controls, assuming that these adjustments reflect the residents' comfort needs. For example, if an occupant consistently lowers the temperature setting on the thermostat, it is inferred that they prefer cooler conditions. While actively seeking direct feedback can provide accurate data, it requires continuous resident engagement, which can be burdensome and lead to response fatigue. Additionally, occupants may not always be willing or able to provide regular and accurate reports of their comfort levels, which can result in gaps or inaccuracies in the collected data.

To avoid requiring explicit feedback, this model uses a passive inference approach that relies on direct temperature data within households to estimate the comfort levels of the occupants. In the literature, adjustments in thermostat setpoints often indicate an occupant's thermal preference [82] [83] [84]. Although this method has some drawbacks, such as imperfect knowledge of the overall system, varying thermal attire, and individual physiological differences, it is still a good proxy to determine thermal comfort without direct sensation ratings. The primary assumption is that the acceptable temperature range for occupants is the range of temperatures they experience when they are in their household and the HVAC system is inactive, indicating that the ambient temperature is within their comfort zone.

The approach used in this model involves segmenting data based on a household's seasonal and climatic settings, considering the heating and cooling seasons and the home and sleeping climate settings. Looking at the ERCOT's time of year segmentation for their analysis reports, the seasons are segmented where a Peak Season is defined as June - September, and all other months are Non-Peak [85]. Based on prior analysis of smart thermostat data, the home schedule was considered from 6 AM to 10 PM, and the sleep schedule was started at 10 PM and continued until 6 AM [86]. By creating a probability distribution of the internal household temperatures, the model can predict the most likely thermal comfort conditions experienced by the occupants. A Kernel Density Estimation (KDE) with a Gaussian kernel is used to make the distribution. KDE is a non-parametric method used to estimate the probability density function of a random variable. It smooths out the distribution by assigning weights to contributions around each data point using a Gaussian (bell-shaped) curve. This flexible method models the temperature preferences within a household without making rigid assumptions about their distribution. This smoothing is crucial for capturing the nuances in the dataset, allowing for a better representation of the temperature ranges that individual occupants find comfortable. Using KDE, we obtain a probabilistic model that respects the varied nature of comfort across different households and adapts to the subtle shifts in preferences that may occur within individual homes over time.

Segment	**Months**	**Home**	**Away**
Peak	Jun, Jul, August, Sept	6 AM to 10 PM	10 PM to 6 AM
Non-Peak	Jan, Feb, Mar, Apr, May, Oct, Nov, Dec	6 AM to 10 PM	10 PM to 6 AM

Table 4.2.1: Segmentation of thermal comfort profiles based on season and time of day

Gaussian Kernel Distribution

Given a set of n data points $\{x1, x2, \ldots, xn\}$, the Gaussian KDE at any point x is given by

$$\hat{f}(x) = \frac{1}{nh\sqrt{2\pi}} \sum_{i=1}^{n} e^{-\frac{1}{2}\left(\frac{x-x_i}{h}\right)^2}$$

Where **h**=5 is the bandwidth, a positive parameter affecting the kernel's width. Each term in the summation calculates the contribution of the kernel centered at a data point xi to the overall density estimate at the point x.

Figure 4.2.1 provides a visual representation of applying the KDE transformation over the ecobee timeseries data for a year, creating a probability distribution representing an individual's thermal comfort profile segmented by season and climate mode. The probability represents the likelihood that a household occupant will feel comfortable at a given temperature. For the model, a 50% threshold is determined as the lowest point of thermal flexibility to which the control agent can push the thermostat setpoint range before the user's thermal comfort is sacrificed. Due to utilizing the KDE for smoothing, the probability can drop below 0, but this is a mathematical byproduct rather than a direct physical measure. In reality, the lowest probability would be 0. The figure illustrates how climate settings can vary significantly between daytime and nighttime and across cooling and heating seasons. The broader shape of the "Cooling Home" curve suggests a wider acceptable temperature range during active periods, influenced by daily activities and a higher tolerance for moderate indoor temperatures or a desire to reduce energy bills during hotter summer days. The key takeaway is that by modeling individual comfort profiles by season and climate setting, the model can more accurately segment household behaviors and climate preferences, tailoring HVAC control to individual users' needs. Table 4.2.2 provides the summary characteristics of the thermal comfort profile in Figure 4.2.1, including the 50% and 0% temperature bounds.

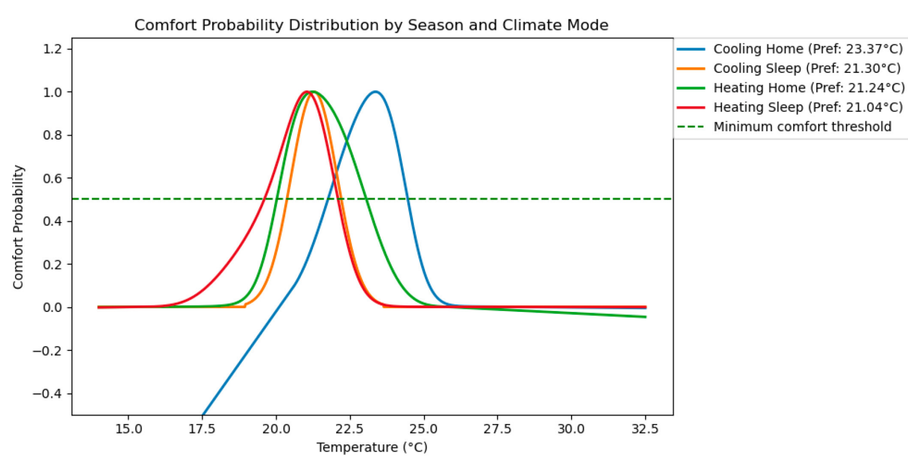


Figure 4.2.1: An example of Thermal Comfort Profile Probability Distribution by Season and Climate

	100% Temp (°C)	50% Temp Range (°C)	0% Temp Range (°C)
Cooling Home	23.4	21.8 - 24.5	20.1 - 27.5
Cooling Sleep	21.3	20.4 - 22.2	15.6 - 27.5
Heating Home	21.2	20.0 - 23.1	15.7 - 25.8
Heating Sleep	21	19.6 - 22.1	15.6 - 26.6

Table 4.2.2: The summary characteristics of a single thermostat profile

Expanding the Comfort Zones

A recent study published in Nature sheds light on how humans perceive changes in environmental temperature and created a metric for the smallest temperature difference that humans can detect 50% of the time, called the Just Noticeable Difference (JND). The JND for temperature was found to be 0.38°C, indicating that individuals could detect a temperature change of 0.38°C with 50% accuracy. However, the threshold where 95% of participants can detect a temperature change, known as JND95, is +/- 0.92°C. Notably, the results were consistent among all participants, regardless of their temperature preferences. These findings are significant in understanding the energy flexibility available while optimizing energy usage and thermal comfort. Therefore, the comfort bands in the model were adjusted based on the empirical insights obtained from the study. The adjustment was made by setting the comfort bands at a minimum of +/- 0.92°C around the highest probability temperature points for those whose 50% comfort thresholds did not meet the 1.84°C comfort band breadth. This adjustment reflects the Just Noticeable Difference with 95% confidence, ensuring that the temperature changes within this range are perceptible yet not discomforting for most individuals and provide a minimum value to which thermal flexibility can be controlled [87].

4.3 Reinforcement Learning

A Quick RL Primer

Reinforcement Learning (RL) is a computational method that learns by interacting with an environment to solve a problem, known as a Markov Decision Process (MDP) [88]. The MDP, which is essential to the RL framework, is based on the Markov Property, meaning that the future state of the environment depends only on its current state and not on the sequence of events that came before it. An MDP consists of four key components: a set of states (S), a set of actions (A), a probability distribution that determines the likelihood of transitioning from one state to another (T), and a reward function (R) that assesses the benefits of actions taken from certain states.

In simple terms, a Reinforcement Learning (RL) agent starts by observing its current state s_t and then chooses to perform an action a_t from the set of possible actions. These actions could be either discrete, as in a game of chess where the agent selects from a predefined set of moves, or continuous, such as a self-driving car making navigational decisions on an open road. The chosen action a_t impacts the environment and causes the state to transition to a new state s_{t+1}, during which the agent receives a reward or penalty. This feedback, represented as reward r, indicates the effectiveness of the chosen action in achieving the desired outcome. The primary goal of the RL agent is to maximize the cumulative reward over time. To accomplish this, the agent continuously refines its decision-making process based on the feedback received from each action. By adapting its policy π through experiential learning, the RL agent enhances its ability to make optimal decisions that align with the overall objectives of the learning process. This iterative feedback and adjustment process allows the RL agent to develop sophisticated policies that effectively navigate the complexities of the environment it is interacting with.

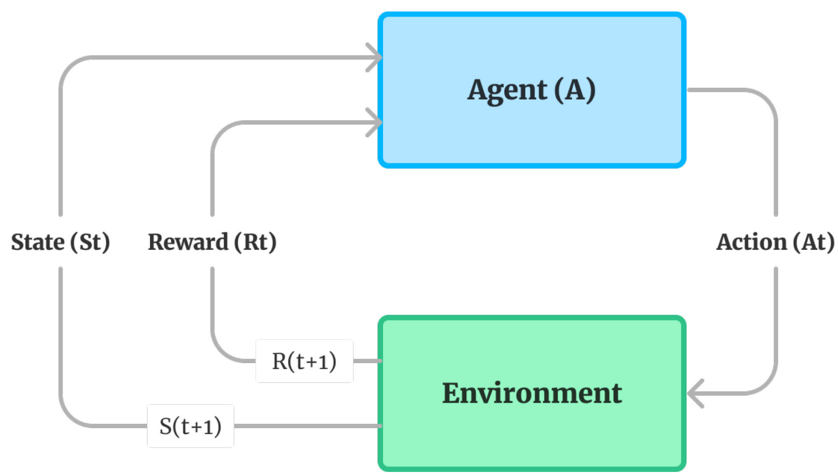


Figure 4.3.1: A high-level overview of how an agent interacts with a reinforcement learning environment

For this study's scenario, the RL agents have been assigned to adjust the thermostat's heating and cooling setpoint at every timestep, affecting the indoor temperature. Each timestep has a 30-minute duration. Different timestep duration settings were tested, including 10 minutes, 15 minutes, 30 minutes, and 1 hour. If the control frequency was too high, the setpoint control changes did not immediately impact the environment. If it were too low, the thermostat would make more drastic changes to overcompensate for the time lag of the larger timesteps. A mid-point of 30 minutes was found to be the most optimal duration. Once the setpoint adjustment has been made, the agent is either rewarded or penalized based on two core factors: the power usage of the HVAC system multiplied by the current electricity price and the level of occupant comfort (a more nuanced breakdown of the reward function is below). Through continuous interaction, this setup teaches the agent about the underlying dynamics of the HVAC system, as well as the tradeoff between energy consumption and electricity costs. As the agent experiences the outcomes of different setpoints, it learns to balance the goal of reducing energy consumption, particularly during peak periods, with maximizing occupant comfort. A detailed list of the Action Space and Observation (state) Space follows.

Action Space

The system utilizes two continuous control actions to adjust the heating and cooling set points. This dual action system allows for independent modification of each set point, providing precise environmental control within the simulated space. To optimize the learning process and ensure practical applicability, the action spaces were individually tailored to each thermostat profile before integration into the learning environment. By predefining the action space boundaries based on realistic and feasible settings from user profiles, the training phase of the model was able to spend less time searching out of bound temperatures and more time refining the control. This approach simplifies the complexity often associated with continuous action spaces by limiting the range of actions to those most relevant and likely to be utilized by actual users. Specifically, the maximum and minimum values from the 'Expected Cool' and 'Expected Heat' settings from each ecobee thermostat were used to delineate the boundaries for the cooling and heating continuous action spaces, respectively. Table 4.3.1 lists the definitions of the continuous action spaces.

Action Type	Parameter	Minimum Value	Maximum Value
HVAC Cooling	Expected Cool	min(x)°C	max(x)°C
HVAC Heating	Expected Heat	min(y)°C	max(y)°C

Table 4.3.1: The action space for the reinforcement learning environment

Observation Space

The reinforcement learning model considers various environmental and operational parameters in its observation space to evaluate the current state and make informed decisions. This space includes environmental and time-related metrics, reflecting the dynamic nature of HVAC operations and energy consumption. A foresight feature was also encoded into the observation space to deliver upcoming energy prices before usage. This feature provides the next 2.5 hours of energy prices, allowing the model to adjust its strategies in response to impending price changes preemptively. This is similar to how a smart thermostat, or an informed user would operate ahead of a time-of-use rate increase. By adding these values to the state space, the model can take actions such as pre-heating or cooling the building to reduce costs during peak periods. Table 4.3.2 summarizes the components of the observation space.

Parameter	Description
Day of Week	Monday - Sunday
Energy Price ($)	Current electricity cost
HVAC Cooling Electric Power (kW)	Power used by the cooling system
HVAC Heating Electric Power (kW)	Power used by the heating system
Hour of Day	Time of day
Month	Month of the year
Rate Type	Type of energy billing
Climate Mode	The operational mode of HVAC
Temperature - Indoor (C)	Indoor ambient temperature
Temperature - Outdoor (C)	Outdoor ambient temperature
Total Electric Power (kW)	Total electricity usage
Net Sensible Heat Gain - Indoor (W)	Heat gain inside the building
Window Transmitted Solar Gain (W)	Solar heat gain through windows
Lookahead - Future Energy Prices (2.5h)	The following six energy rates in the future

Table 4.3.2: The observation space for the control environment

Crafting The Reward Function

The reward function guides the agent toward the desired behavior, balancing thermal comfort and energy efficiency. The overall reward R at each timestep is a composite of a thermal comfort and energy efficiency reward, an energy penalty, and a small action regularization penalty. The weights, w_1, w_2, and w_3, are additional hyperparameters used to balance the trade-off of thermal comfort and energy usage.

$$R = R_{\text{thermal}} + w_1 R_{\text{efficiency}} - w_2 R_{\text{energy}} - w_3 R_{\text{action}}$$

The final weights used are as follows: $w_1 = 0.1 \ w_2 = 0.5 \ w_3 = .01$

Thermal Comfort Reward

The thermal comfort reward is calculated based on the KDE-interpolated probability of the indoor temperature, representing the likelihood that the current temperature is within a comfortable range. A 50% threshold (the comfort range) is determined to be the lowest probability that the agent should allow the indoor temperature to deviate to optimize energy cost and flexibility. The detailed calculation of the thermal comfort reward is as follows:

$$R_{\text{thermal}} = ComfortProbability(T_{\text{curr}}) - \lambda_2\big(\max(T_{\text{min}} - T_{\text{curr}}, 0)\big) + \lambda_3\big(\max(T_{\text{curr}} - T_{\text{max}}, 0)\big)$$

Where T_{curr} is the current temperature, T_{min} is the temperature threshold with a 50% probability of comfort on the left side of the distribution curve, and T_{max} is the temperature threshold with a 50% probability of comfort on the right side of the distribution curve.

The reward is adjusted to increase the penalty for temperatures outside the desired comfort range (thermal comfort probability > 50%), with λ_2 and λ_3 serving as a tuning parameter to control the influence of the trapezoidal penalty when outside of the range. Additionally, the function applies a quartic root transformation to the KDE-derived probability if it exceeds 0.5 to smooth out the reward's curve towards unity while keeping the original shape of the distribution to ensure that the highest rewards are granted for temperatures within the optimal comfort zone. This is a crucial distinction. Instead of providing a fixed or linear reward for probabilities over 50%, the curve gradient allows the model to make informed tradeoff decisions about energy cost vs. thermal comfort.

$$\textbf{ComfortProbability} = 0.5 + (1 - 0.5)\left(\frac{ComfortProbability(T_{\text{curr}}) - 0.5}{0.5}\right)^{\frac{1}{4}}$$

The lambda values for the equation are dynamically calculated for each thermostat to apply a trapezoidal penalty that still accounts for the shape of the KDE distribution but provides a large negative penalty for values past the 50% probability. The calculation for the λ_2 and λ_3 are below:

λ_2

$$\frac{1}{\max\left(T_{\text{min}} - T_{\text{min}_0}, 0.01\right)}$$

λ_3:

$$\frac{1}{\max\left(T_{\text{max}_0} - T_{\text{max}}, 0.01\right)}$$

Where $T_{\text{min}0}$ is the temperature threshold with a 0% probability of comfort on the left side of the distribution curve, and $T_{\text{max}0}$ is the temperature threshold with a 0% probability of comfort on the right side of the distribution curve.

Energy Efficiency Reward

The efficiency term, $R_{\text{efficiency}}$, is calculated to promote energy efficiency based on the current temperature, the season (cooling or heating), the energy price, and the normalized energy use and is only applied when the current temperature is within the desired comfort range (above the 50% probability threshold), otherwise it is set to 0. The efficiency term helps encourage energy-efficient temperature settings while considering the energy price and normalized energy use. Initially, the algorithm tended to optimize solely for the highest probability temperature, quickly learning that any deviation decreased rewards. As a result, the agent did not adequately explore the environment to find a better trade-off between sacrificing positive comfort rewards in favor of reducing energy usage when prices were high. Despite the SAC algorithm being explicitly chosen for its ability to explore complex environments, the thermal inertia in building environments—where it takes time for thermostat changes to manifest—caused the agent to get stuck in local minima. To address this problem, the reward function was redesigned to include efficiency rewards, providing the agent with additional signals of positive actions rather than relying solely on maximizing thermal comfort. Incorporating the efficiency term into the reward function significantly enhanced the reinforcement learning algorithm, enabling it to explore and converge toward policies that balance occupant comfort and energy efficiency. The term is calculated as follows:

Case 1:

When the current temperature is within the desired comfort range ($T_{\text{min}} < T_{\text{current}} < T_{\text{max}}$):

If the season is **Cooling**:

$$R_{\text{efficiency price}} = \min\left((T_{\text{current}} - T_{\text{min}}) \cdot \frac{E_{\text{price,t}}}{\text{E usage}_{\text{normalized},t} + 0.01} \cdot w_{\text{efficiency}}, 2 \right)$$

If the season is **Heating**:

$$R_{\text{efficiency price}} = \min\left((T_{\text{max}} - T_{\text{current}}) \cdot \frac{E_{\text{price,t}}}{\text{E usage}_{\text{normalized},t} + 0.01} \cdot w_{\text{efficiency}}, 2 \right)$$

In both cases, the efficiency term is calculated as the product of the temperature difference (between the current temperature and the minimum or maximum temperature as defined by the individual's thermal comfort range), the ratio of energy price to normalized energy use (with a small constant of 0.01 added to avoid division by zero), and a scaling factor of $w_{\text{efficiency}} = 0.1$. The resulting value is then limited to a maximum of 2.

Case 2

An additional efficiency boost was given when the current temperature is within the desired comfort range but deviates from the highest probability temperature, signaling that it is on the more energy-efficient side of the thermal comfort range based on the season. As mentioned above, this boost encourages the agent to explore temperature settings outside the highest probability when energy prices are higher. If the current temperature is higher than the highest probability temperature for the cooling season, the efficiency reward increases to reflect the improved energy efficiency. Conversely, the reward similarly increases during the heating season if the current temperature is lower than the highest probability temperature. The logic is:

If the season is **Cooling** and $T_{\text{current}} > T_{\text{highest\_prob}}$:

$$R_{\text{efficiency}} = R_{\text{efficiency price}} + \left|T_{\text{current}} - T_{\text{highest\_prob}}\right| \cdot w_{\text{temp}} \cdot E_{\text{price}}$$

If the season is **Heating** and $T_{\text{current}} < T_{\text{highest prob}}$:

$$R_{\text{efficiency}} = R_{\text{efficiency price}} + \left|T_{\text{highest prob}} - T_{\text{current}}\right| \cdot w_{\text{temp}} \cdot E_{\text{price}}$$

The reward is calculated as the absolute difference between the current and highest probability temperatures, multiplied by a factor of $w_{\text{temp}} = 0.2$ and the energy price. This additional term in the efficiency reward signals to the agent that exploring the upper end of the thermal comfort range during high-price periods is more beneficial. If the reward collected from deviating from the highest probability temperature (deviation times energy price) is greater than the loss of reward value from deviating from the value received as a comfort reward, it will make sense to move towards the upper bounds of the thermal comfort range. In contrast, when energy prices are lower, the efficiency reward for deviating from the highest probability temperature is less significant. Therefore, the agent finds it more optimal to maintain the highest probability temperature, balancing comfort and energy efficiency across varying price scenarios. **Note**: If the current temperature is outside the desired comfort range ($T_{\text{current}} \leq T_{\text{min}}$ or $T_{\text{current}} \geq T_{\text{max}}$), the efficiency term is set to zero: $R_{\text{efficiency}} = 0$.

Energy Penalty

The energy penalty reward, on the other hand, is defined as:

$$R_{\text{energy}} = E_{\text{usage}} \times E_{\text{price}}$$

where energy usage is log normalized against a benchmark energy use for the given timestep's month, calculated as

$$\text{Eusage}_{\text{normalized},t} = \frac{log(E_t + 1)}{E_{\text{benchmark}}}$$

$E_{\text{benchmark}}$ is calculated using the mean log normalized energy consumption hours in each month under OCHRE's standard Rule-Based Controller. Before each dwelling was simulated, the setpoints were manually adjusted to reflect the highest probability temperature for each matched simulation building to thermostat profiles. This benchmark also served as the standard operating benchmark for each building simulation and was used as the baseline when comparing the different rate outcomes.

Action Regularization Penalty

L2 regularization is a technique that enhances the effectiveness and efficiency of control signals by providing small negative rewards for significant fluctuations in signals. For thermostat settings, this promotes stable temperature regulation. Stable temperature regulation is crucial for maintaining a consistent indoor environment and reducing the wear and tear on the HVAC system, which results from frequent cycling on and off. Every time HVAC systems are activated, they incur startup energy costs, and the frequent switching can lead to accelerated degradation of system components. The L2 regularization helps to prevent these issues by discouraging large swings in thermostat settings. This approach aids in smoothing out temperature variations throughout the day, leading to a more stable indoor climate from one timestep to the next. Without such regularization, the thermostat might adopt an optimal policy that frequently toggles the system on and off to optimize short-term efficiency rewards, particularly in fluctuating energy prices. Moreover, the inclusion of L2 regularization helps the agent recognize the delayed impact of actions on the environment and helps prevent the agent from creating a false representation of HVAC control settings whereby it determines more significant increases in the control settings will change the internal temperature of the environment faster. This understanding is beneficial for strategies like pre-cooling, where the rewards for actions taken are realized later. This technique improves the agent's ability to manage the temperature consistently and effectively by aligning the system's operation more closely with real-world dynamics and costs. It also avoids the discomfort and annoyance of continuous noise and temperature fluctuations.

The Action penalty is formalized as follows:

$$R_{\text{action}} = \left(H_{current} - H_{previous}\right)^2 + \left(C_{current} - C_{previous}\right)^2$$

Where $H_{current}$ and $C_{current}$ are the current heating and cooling setpoints and $H_{previous}$ and $C_{previous}$ are the setpoints from the previous timestep.

Why Benchmarking

One of the key challenges in reinforcement learning is how to properly shape rewards given to the agent. In the case of this control problem across many building archetypes, the reward signals need to be appropriately scaled to achieve the desired effects across different settings. The core principle is to align the usage penalty with the comfort reward as closely as possible, forcing the algorithm to make calculated tradeoff decisions. However, high energy usage variability among buildings makes applying a consistent reward function to all the simulations challenging. To address this, the reward function needed to be designed to be generalizable, ensuring broader applicability. Thus, normalization of energy usage was necessary. This scaling proves effective for most households and balances energy penalty and comfort rewards around a scaled value of 1. However, in extreme conditions, such as under critical peak pricing or severe weather events leading to high energy usage, the penalty portion of the reward could exceed 1. In such situations, the reinforcement learning agent must evaluate how much violation of energy comfort is justifiable against excessive energy use. This approach enhances the model's adaptability and decision-making in diverse operational environments, and normalization allows for scalable comparison across the buildings.

To normalize the data, an adaptive rolling window approach was first used to calculate the average energy usage of a building. This approach aimed to adjust dynamically to the building's energy patterns. However, this method faced a cold start and a runaway problem. First, the initial observations didn't have enough data to set the rewards accurately. Second, the calculated average escalated continually during the heating season. This led to progressively higher baseline energy assumptions. As described previously; to solve this problem, the algorithm was given benchmarked data that utilized the log normalized mean of the energy usage at every timestep. This was the most effective strategy to familiarize the model quickly with the building's dynamics. Interestingly, it was discovered that the average energy usage at each timestep was enough for the algorithm to determine what a typical energy profile should look like. This simplified approach made the initial learning phase more efficient and stabilized the reward calculations.

Transition Hours

One crucial improvement in this study over previous work involved creating multiple set point schedules tailored to individual sleep and heating preferences. This is an essential factor as it reflects realistic household behaviors. This customization is particularly significant because it accounts for possible end-of-day rate shifts, which could coincide with increased energy demands from nighttime electric vehicle charging activities. A transition period, T, was established to help the reinforcement learning agent understand the transition between typical home and sleep-specific climate settings. This period averaged the sleeping and cooling probability distributions and served as a guide for adjusting between the home and sleep climate settings. Incorporating this transition was crucial to mitigate energy spikes during the set point change, ensuring a smoother and more gradual energy consumption shift. This approach is particularly beneficial for households with significant variations in their thermal preferences between day and night. Without such a transition adjustment, the model could risk aggressive over-corrections, potentially leading to suboptimal thermal comfort and energy inefficiency. Introducing a calibrated transition period allowed the agent to more accurately and gently adapt to the appropriate set point schedules, enhancing energy efficiency and occupant comfort. For this study, the transition period was set as two hours after the sleep schedule and two hours after the home schedule.

Summing Up the Reward Function Design

The reward function was designed to balance the flexibility of the thermostat between the 50% and 100% comfort thresholds. This range sets the minimum target for the reward function to ensure significant energy efficiency while maintaining comfort. A substantial reward reduction was implemented if the temperature exceeded this preferred zone to discourage it. However, the function is not rigidly cut off at this threshold. It can handle exceptional situations involving extreme energy prices, such as critical peak pricing, coupled with high usage. In such cases, the model must decide whether each marginal increase in energy usage can be optimized for thermal comfort and justify deviating from optimal comfort levels. A steeply sloped negative reward is applied for each comfort decrease below the 50% threshold. This compels the model to weigh the energy cost against the decrease in thermal comfort. This design ensures that the model strives to maintain comfort while also dynamically adjusting its strategies based on the severity of the circumstances. By integrating these components, the reward function encapsulates both critical aspects of HVAC control—comfort and cost—providing a quantitative measure that the RL agent aims to maximize over time. Figures 4.3.2-4.3.4 are samples from the training logs used to monitor the agent's progress. They provide visual insights into how the model's training improves from the first rollout to the last before training ends. Figure 4.3.2 breaks down each of the components of the reward function by timestep. Even though the rewards are segmented, the agent only receives the cumulative value and can derive what environment states returned what reward value. Figure 4.3.3 highlights how the agent learns to maximize the internal temperature while utilizing pre-heating and pre-cooling strategies before the rate switch. Figure 4.3.4 displays the internal temperature of the environment across each timestep whereby the agent learns to maximize towards the higher end of the temperature spectrum.

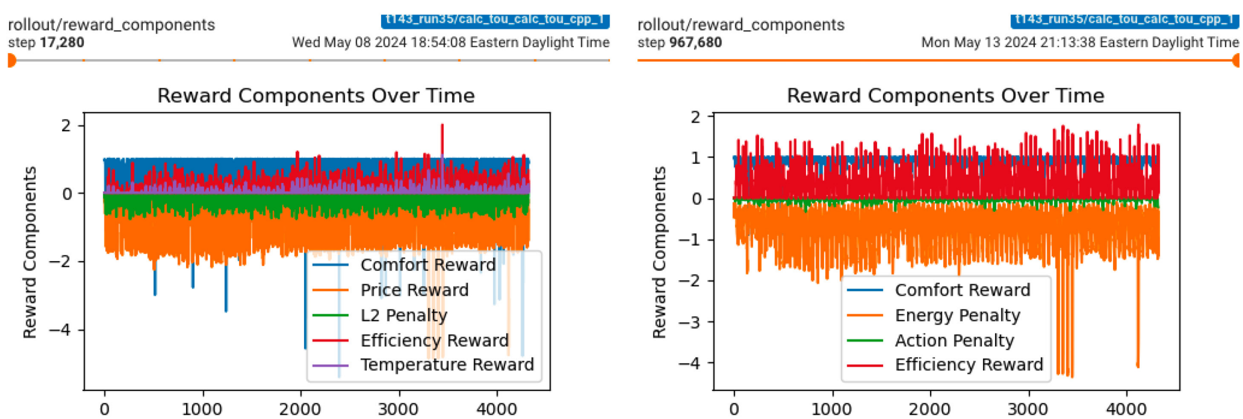


Figure 4.3.2: The reward function's components over the training steps (from timestep 17,280 to 967,860)

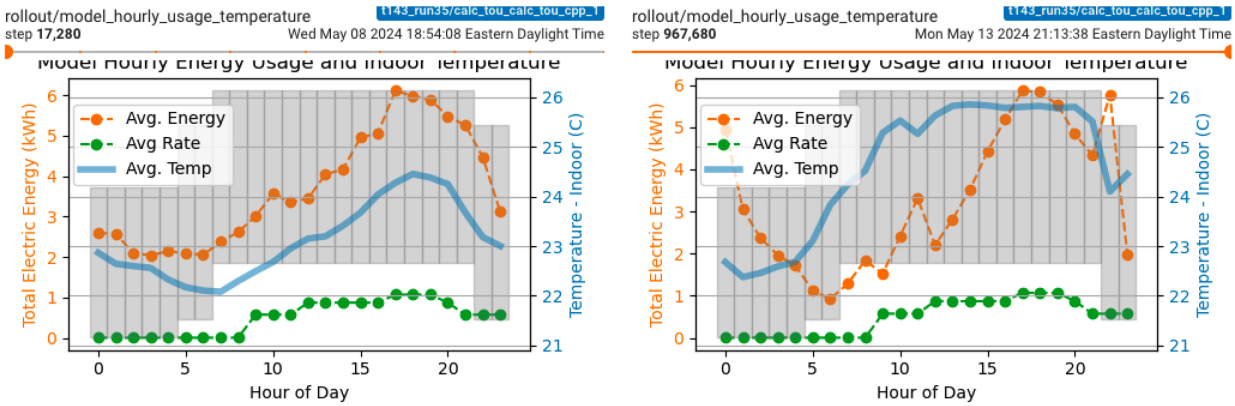


Figure 4.3.3: The model's adjustment of the electricity usage and indoor temperature. The left side of the charts represent the electric energy and the right side the indoor temperature (from timestep 17,280 to 967,860)

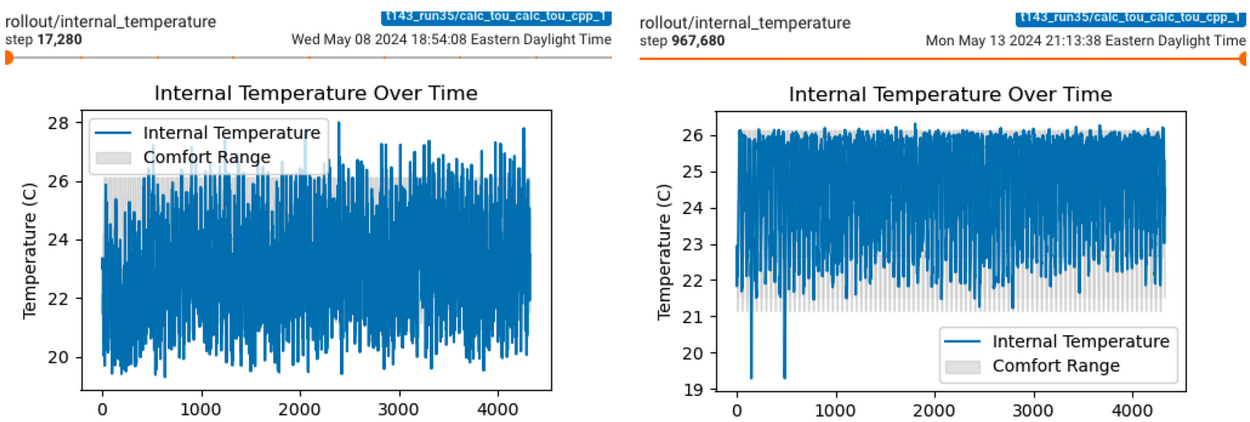


Figure 4.3.4: The internal temperature adjustments over time with the gray zones representing the temperature is within the 50% thermal comfort probability (from timestep 17,280 to 967,860)

Final Steps: Choosing an Algorithm

Off-Policy vs. On-Policy

In reinforcement learning, agents can use two types of algorithms to learn and make decisions: On-Policy and Off-Policy. On-policy algorithms require the agent to learn and improve upon the same policy it uses to make decisions. This means that the agent learns from the direct consequences of its actions under the current policy. On the other hand, Off-Policy algorithms allow the agent to learn a potentially optimal policy while acting on a different, possibly exploratory policy. This method provides greater flexibility and enables the agent to learn from a broader range of actions. The choice between these two algorithms can impact the learning speed, the safety of the learning process, and the agent's ability to handle new scenarios. On-policy algorithms are more conservative, while Off-Policy algorithms allow for a more robust exploration of possible strategies. An On-policy algorithm is more adept for scenarios where safety, stability, and adherence to specific behavior patterns are prioritized, such as when a robot operates alongside human counterparts. However, an Off-Policy algorithm was chosen for this work to allow for extensive exploration of the action space and more aggressive learning of optimal strategies.

Deep RL

Deep Reinforcement Learning (DRL) algorithms represent an advanced subset of reinforcement learning techniques that integrate deep learning. This integration allows DRL to handle highly complex decision-making environments by using deep neural networks to approximate functions, which traditional RL methods may struggle with due to their complexity or the sheer volume of data.

Actor-Critic Methods

Actor-critic methods are a popular type of algorithm used in reinforcement learning. These methods involve two models: an actor and a critic. The actor selects actions, and the critic evaluates how good the actor's action is. This separation enables more efficient processing and learning because the critic's feedback continuously helps refine the actor's policy decisions. The actor learns to perform actions that maximize rewards, while the critic estimates the value of being in a particular state and evaluates the actions' quality.

Soft Actor-Critic Algorithm

One popular Actor-Critic reinforcement learning algorithm variant is the Soft Actor-Critical (SAC) algorithm. SAC adds an entropy term (ambiguity in the decision-making process) to the reward function, incentivizing the actor to choose uncertain actions. This promotes a balance between exploring new actions and exploiting known strategies, leading to improved exploration capabilities and faster convergence. As a result, SAC is a highly effective algorithm for environments with many local optima that could trap less robust algorithms. In optimizing energy cost and consumption while maintaining comfort, the SAC algorithm has proven to be particularly effective, especially when using a continuous action space instead of discrete (ON/OFF) thermostat actions [89]. Some studies have shown that SAC achieves stable performance with ten times less data than on-policy methods [90]. Its ability to continuously explore and exploit makes it adept at finding and maintaining optimal balance. Due to its ability to balance sample efficiency and exploration, the SAC algorithm from the Stablebaselines3 library was utilized [91].

Chapter 5. The Environment and Simulation

Austin, Texas, was chosen as the area of interest due to its high summer peak loads. Oiko Lab provided the 2023 AMY weather data, and ecobee's thermostat data was collected for the same year [92]. The selected simulation timeframe accounts for post-pandemic behavior patterns, especially work habits, and aligns with how people who have settled into permanent work-from-home situations or returned to offices have been operating. Specifically, the simulation ran for 90 days, starting from June 1st, 2023, to cover the summer months. Austin is situated in the Texas Interconnection, operated by the Electric Reliability Council of Texas (ERCOT), an Independent System Operator (ISO) that provides electricity to over 26 million customers in Texas, representing 90% of the state's electric demand [93]. A unique aspect of ERCOT is that most residents must pick their electric supplier upon sign-up for electricity instead of defaulting to the distribution company's electric supply plan. As a result, over 140 active retail electric suppliers offer retail electricity rates to consumers [94]. Interestingly, this system incentivizes market participants to design different electricity tariffs tailored to individual preferences rather than locking all participants into default fixed rates. With that said, Austin, Texas, is unique in that it is serviced by Austin Energy, a publicly owned utility by the City of Austin. It does not provide residents with alternative supply options and uses a progressive tiered rate structure that increases the charge for electricity based on the incremental amounts above a tiered threshold, much like the federal tax system [95].

Lastly, to help the model converge faster, the HVAC heating action was adjusted. This was done because the simulation runs during the summer months in a cooling-dominated climate. The HVAC heating power was set below 20°C during the summer months to avoid providing action noise. This adjustment is based on reality, where smart thermostats have settings that only allow heating or cooling use based on the time of the year.

Location	**Year**	**Start Date**	**Duration**	**Total Timesteps Simulated**
Austin, Texas	2023	June 1st, 2023	90 days	4,320 timesteps per environment episode

Table 5.0.1 Study Design Parameters

5.1 Mapping Resstock to ecobee

To connect the buildings in the Resstock dataset with the thermostats in the ecobee dataset, each home needed to be matched to a thermostat. This mapping allows the simulation environment to establish a link between the simulated building types and the real-world thermostat data. The ecobee dataset underwent a thorough cleansing process to ensure data quality and consistency. The process involved removing duplicate account IDs, thermostats with incomplete information (missing internal temperatures, expected heat, or cool settings), thermostats that did not have building type information, and outliers such as buildings over 120 years old or those with an area of 10,000 square feet. Following the cleansing process, the final ecobee dataset included 597 thermostats suitable for clustering analysis. To pair with these thermostats, 110 homes were sampled for the simulation in Resstock. However, seven homes could not be mapped due to multi-source heat pump compatibility issues within OCHRE. So, the sample was reduced to 103 homes. Resstock uses a down-sampling technique that automatically generates building representations covering the region's segmentation, including various home types, vintages, envelope characteristics, and heating/cooling system types. This approach adds robustness because the sampled homes accurately represent the diverse building stock in the region of interest.

Clustering

To ensure a representative distribution of the ecobee thermostat profile data on the limited number of Resstock samples, a K-Nearest Neighbor (KNN) clustering algorithm was used on the thermostat profiles. This was done by computing each thermostat's ideal comfort profile and analyzing the "Cooling" season settings based on the lower and upper bound 50% threshold for the home and sleep climate schedules. The KNN algorithm was preferred to group similar data points based on their proximity in the feature space. The clustering was performed separately for each thermostat building type after aggregating similar types in the ecobee dataset: Detached, Attached, and Multiunit. Evaluation metrics include the elbow method, Silhouette coefficient, and Davies-Bouldin index to determine each building type's optimal number of clusters. These metrics assess the clusters' compactness, separation, and overall quality. After careful review, the following cluster sizes were chosen: Detached - 6 clusters, Attached - 7 clusters, and Multiunit - 6 clusters. Figures 5.1.1 and 5.1.2 below visualize the clusters based on their Cooling Home Min and Cooling Home Max thermostat settings. The linear bottom shape is due to manually extending the thermal comfort profiles of all residents to be +/- 0.92°C.

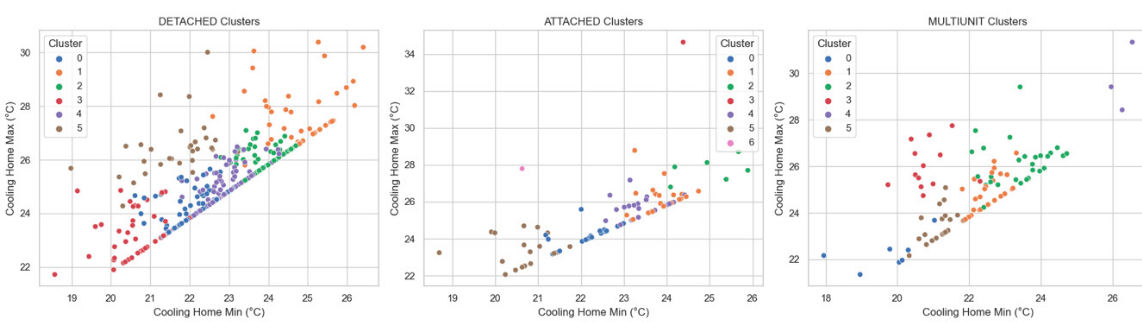


Figure 5.1.1: The grouping of clusters by building type. The y-axis is the maximum cooling temperature for the home climate setting, and the x-axis is the minimum temperature for the home climate setting

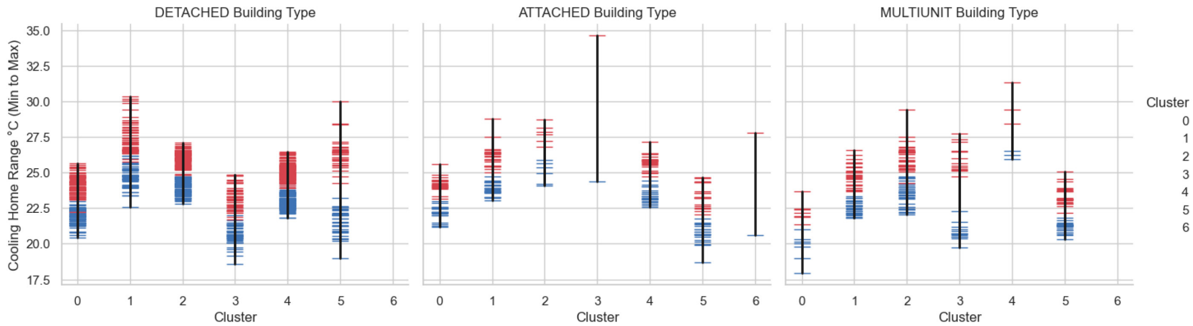


Figure 5.1.2: The temperature range (for Cooling and Home climate setting) for each of the different clusters broken out by building type

Mapping Approach

Clusters were formed based on the ecobee thermostat metadata, and a mapping algorithm was used to match thermostat profiles to corresponding Resstock building profiles. The objective was to ensure that each Resstock building was paired with the most suitable thermostat profile while maintaining the distribution of housing archetypes from Resstock and the thermal preferences captured by the ecobee thermostats. Figure 5.1.2 provides an overview of the methodology. The mapping algorithm employed a distance-based search to find the closest match between each Resstock building and the available thermostat profiles within the same building type and cluster. A combination of normalized features was used to calculate the distance, including the vintage year, number of floors, floor area, and the presence of a heat pump. These features were selected to capture the similarity between the building characteristics and the thermostat settings.

The algorithm used for the mapping process involved iterating over each building type, such as detached, attached, or multiunit, and their respective clusters. The number of buildings assigned within each cluster was determined based on the cluster percentages obtained from the clustering analysis. The algorithm then calculated the distances between each unassigned building and the available thermostat profiles within the cluster. The building-thermostat pair with the shortest distance was considered the best match and assigned accordingly. This process was repeated until all buildings within the cluster were assigned a thermostat profile or until no more suitable matches were found. If any buildings remained unassigned after iterating through all clusters, the algorithm performed an additional assignment round. It searched for the closest match among the unassigned thermostats, regardless of the cluster, to ensure all buildings were paired with a thermostat profile. The distance-based mapping approach linked each Resstock building with the most representative thermostat profile while preserving the distribution of housing archetypes and thermal preferences. The clustering and mapping approach mapped the 597 thermostats to 103 Resstock buildings. See the Appendix for the full breakdown of the clusters.

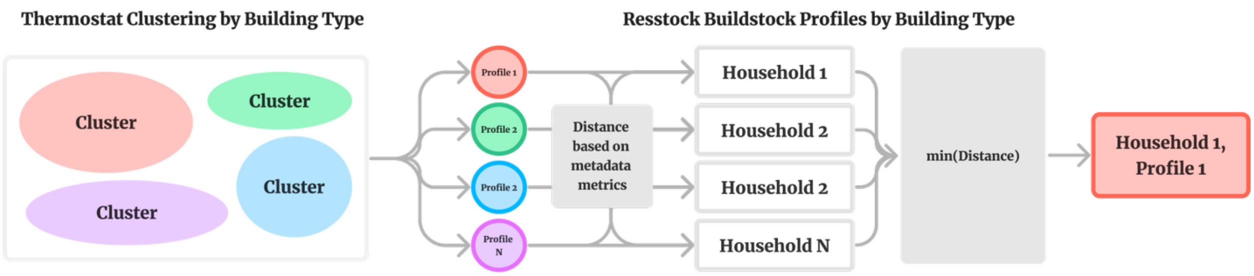
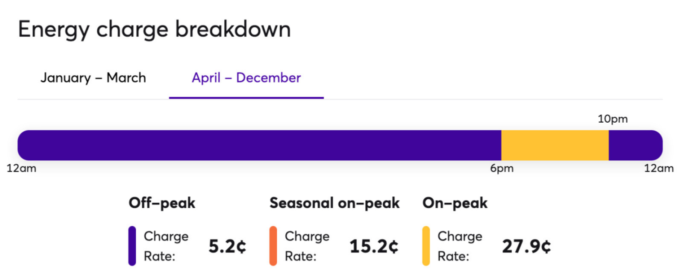


Figure 5.1.2: An overview of the mapping methodology to align thermostat comfort profiles to Resstock housing stock buildings

5.2 Utility Rates

This simulation analyzed four different utility rates to assess their impact when combined with automated thermostat control. This study focuses only on the electricity supply portion of a residential electric bill as this portion is directly exposed to the wholesale retail electric market. It does not consider the distribution, network, or transmission fees associated with an electric bill. Each rate was chosen to represent a distinct pricing structure:

1. **Fixed Rates**: These rates were calculated using the Load Weighted Average Price from historical Day Ahead prices in ERCOT during 2023. The average rate for each hour throughout the year was determined, providing a baseline rate that reflects general market conditions without considering intra-day fluctuations.
2. **Power Shift - Rhythm Energy**: Rhythm Energy uses a power shifting rate structure that applies peak rates from 6 PM to 10 PM daily, from April to December. Georgetown, Texas, was selected due to its location in the South-Central Load Zone of ERCOT, which includes Austin. This decision was made because Austin consumers do not have the option to choose their supplier. The rate mentioned here was accessed from the Rhythm Energy website on May 9, 2024.



[96]

Figure 5.2.1: The Power Shift rate, inspired by Rhythm energies summer rate

3. **TOU Optimized**: This rate structure is constructed using a partitioning algorithm proposed by Yang and Schittekatte [97] [28]. It divides hours into Time of Use (TOU) periods, with a maximum of three daily periods. The algorithm calibrates based on Day Ahead price patterns and a clustering method from Yang et al. (2019). Each period lasts at least three consecutive hours and reflects a more dynamic pricing model designed to encourage energy use that aligns more closely with grid demands. These periods can repeat within the same day, such as off-peak, shoulder, peak, and shoulder. A more detailed overview of the partitioning algorithm is available in the Appendix.
4. **TOU Optimized with CPP**: This rate plan includes five critical peak pricing (CPP) events based on the ERCOT South Central Load Zone peak load and combines them with the optimized TOU structure above. Each critical peak pricing event lasts three hours and is scheduled during the highest load periods observed from June to August 2023. These events signify a time of significant stress on the power grid. To minimize the energy load during these critical times, the hour and the hours preceding and following each peak are designated as the "3-hour" Critical Peak pricing window. If all 5 of the peak hours could be reduced to the 6th highest peak in 2023, it would result in a 364.85 MW reduction across the South-Central zone. Figure 5.2.2 and Table 5.2.1 below show the peak electricity usage for the South-Central Peak Load Zone [98]).

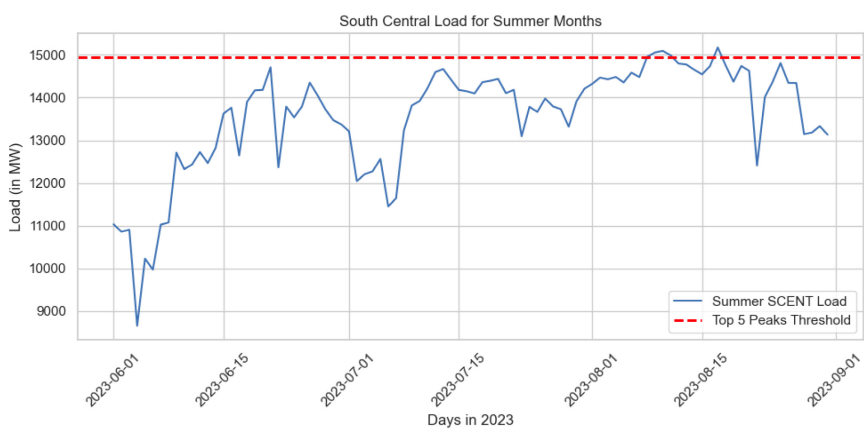


Figure 5.2.2: ERCOT South Central Load for June – September 2023. The red line represents the threshold set for the top 5 peaks in the system

Date	Peak Load (MW)
08-17-2023 @ 17:00	15,174.02
08-10-2023 @ 17:00	15,093.31
08-09-2023 @ 17:00	15,057.03
08-11-2023 @ 17:00	14,987.43
08-08-2023 @ 17:00	14,953.16

Table 5.2.1: Top 5 Peak Load days in (MW) for ERCOT's South Central Region

Final Rates

Each rate was calibrated against the load-weighted average price from the day-ahead prices in the Austin Energy Zone within ERCOT [99] [98]. This calibration process ensured that the rates accurately reflected the South-Central region's energy demand and load characteristics during the simulation period. The final rates are outlined in Table 5.2.2 below after adjusting for the Load Weighted Average scaling.

Rate Type	Day Type	Time Block	Rate ($/kWh)
Fixed	-	-	0.171
Power Shift	-	18:00 - 22:00	0.493
	-	All other times	0.092
TOU	Weekend	00:00 - 08:00	0.0227
	Weekend	09:00 - 23:00	0.2424
	Weekday	00:00 - 08:00	0.0206
	Weekday	09:00 - 11:00	0.0325
	Weekday	12:00 - 20:00	0.3232
	Weekday	21:00 - 23:00	0.0325
CPP	-	17:00-20:00	1.216

Table 5.2.2: Final rate designs used for the simulations

5.3 Soft Actor-Critic Model and Hyperparameter Selection

The success of the simulation heavily depends on tuning hyperparameters for the Soft Actor-Critic (SAC) algorithm. These values influence the learning dynamics of the model, affecting how quickly and accurately the model converges to an optimal policy. If the settings are inappropriate, the model may overshoot the optimal policy or fail to converge, especially in complex environments with nonlinear patterns and varying constraints. This becomes even more critical as the same algorithmic configuration is used in multiple environments with different physical properties and thermostat comfort profiles. Additionally, the environments have stochastic occupant schedules and diverse energy usage patterns that change with every environment reset. Effective hyperparameter tuning ensures that the SAC model can adapt to these dynamics, optimizing energy management strategies across varied scenarios without overfitting to any single environment. The final list of hyperparameters and their values are in Table 5.3.1.

Hyperparameter	Value	Description
Learning Rate	0.003	Moderates update magnitude per learning step, balancing speed and stability.
Batch Size	512	The batch size used for training provides a trade-off between stability and data diversity.
Buffer Size	100,000	The capacity of the memory buffer to store experiences
Tau	0.005	Ensures gradual integration of new knowledge
Gamma	0.99	Discount factor for future rewards, emphasizing the importance of long-term gains.
Train Frequency	32	Balances learning speed and resource use.
Gradient Steps	32	The number of gradient steps per training session is linked to the depth of each learning update.
Entropy Coefficient	auto	Automatically adjusts exploration versus exploitation to prevent suboptimal policy convergence.
Target Entropy	-2	Sets desired entropy to maintain a specific level of exploratory behavior in the policy.

Table 5.3.1: Training hyperparameters for the Soft-Actor Critic (SAC) algorithm

Chapter 6. Results

An efficient computational batching framework was designed to run on the MIT SuperCloud for the final simulation. The simulations ran for 1,000,000 timesteps, at which point the incremental rewards and the actor and critic loss metrics showed diminishing returns. Each simulation was aligned with a single rate structure to enhance the computational efficiency and specificity of the results. Initially, each Time-of-Use (TOU), Power Shift, and Fixed Rate were run simultaneously within a single simulation where a single agent was tasked to learn how to account for each rate. However, this approach did not yield generalizable results because cross-interference between rate structures led to atypical energy usage patterns. For example, the Fixed Rate, typically characterized by stable energy consumption, unexpectedly mirrored the pre-peak patterns of the TOU rate and attempted to pre-cool before a rate change, indicating a spillover effect from one rate structure to another. This observation necessitated the segregation of rate structures into separate simulations to isolate and accurately assess the impact of each agent. As a result, 309 (TOU and TOU CPP were run together) different models were processed in parallel through a batch job framework on the supercomputer, and the total training time lasted under 8 hours. A final data cleansing step was necessary to remove simulations that did not compute correctly, which left the final count of homes to 82.

For the Benchmark, the simulations were compared against the OCHRE model runs with a basic rule-based control (RBC) setting. The purpose of this Benchmark was to replicate a typical scenario of consumer behavior, where the thermostat follows pre-set schedules based on the highest probability temperatures for cooling during day and night. This RBC setup was purposefully designed to be non-responsive to price variations, which mimics a conventional thermostat that functions without dynamic adjustments to fluctuating electricity costs. Additionally, the study uses a set of Key Performance Indicators (KPIs) to measure how well each agent responded to each rate structure. These KPIs include **Thermal Comfort Percentage, Thermal Comfort Maintained over the 50% Threshold, Total Energy Usage, Cost Savings, and Peak Reduction**. Each indicator gives valuable information about different aspects of agent performance, from energy efficiency and cost-effectiveness to maintaining optimal thermal comfort for occupants. The results chapter will begin by evaluating the overall impacts. Then, it will explore how different segments of housing stock are affected. Finally, it will examine two specific examples, one high and one low level of flexibility.

6.1 Thermal Comfort

Thermal comfort is essential in determining the agent's success when combined with different rate structures. The main goal is to ensure that the control algorithm and the rate structure do not negatively impact the comfort of the occupants within their comfort range.

Thermal Comfort Percentage

Using their personal thermal comfort probability distribution, this is the average probability of a user's thermal comfort at any given time relative to indoor temperature. It can be expressed quantitatively as:

$$\text{Thermal Comfort Percentage} = \frac{1}{T}\sum_{t=1}^{T} P\,(\text{comfort at } t)$$

Where T is the total number of timesteps in the simulation, and $P(\text{comfort at } t)$ is the probability of the user being comfortable at timestep t, determined by the indoor temperature.

Rate Type	Mean (%)
Benchmark	93.1
Fixed	81.7
Power Shift	82.9
TOU	81.2
TOU + CPP	81.4

Table 6.1.1: Average thermal comfort probability for each of the rate types

Table 6.1.1 above and Figure 6.1.1 below emphasize the Benchmark having the highest average thermal comfort percentage at 93.1%. This is expected since the Benchmark has set the temperature to the highest probable for comfort. The temperature transition period between sleep and home settings is why this percentage is not 100%. On the other hand, the agents under the different rate types demonstrate substantially lower (mean values range from 81.2% to 82.9%) and more variable thermal comfort percentages compared to the Benchmark. This intentional outcome reflects the effectiveness of the trade-off the agent is making between optimizing costs and achieving 100% thermal comfort.

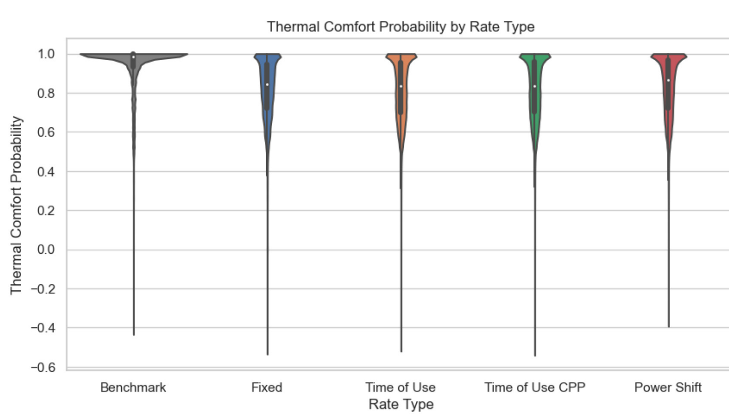


Figure 6.1.1: Thermal comfort probability by rate type

Thermal Comfort Maintained over the 50% threshold

This measures the percentage of time during which the probability of thermal comfort was at least 50%, emphasizing the agent's ability to maintain comfort levels above a minimally acceptable threshold:

$$\text{Thermal Comfort} > 50\% = \left(\frac{\text{Number of timesteps with } P(\text{comfort}) \geq 50\%}{T}\right) \times 100$$

Rate Type	Percentage Above 50%
Benchmark	98.4
Fixed	97.4
Power Shift	97.6
TOU	97.7
TOU CPP	97.6

Table 6.1.2: The percentage for timesteps above the 50% threshold by rate type

When evaluating the ability to maintain comfort, the control algorithms account for less than a 1% difference from the Benchmark. As shown in Figure 6.1.2, when examining the 50% threshold across the distribution, the agent-based controllers have a more consistent distribution than the Benchmark. This validates the control algorithm's ability to maintain thermal comfort under different rate structures.

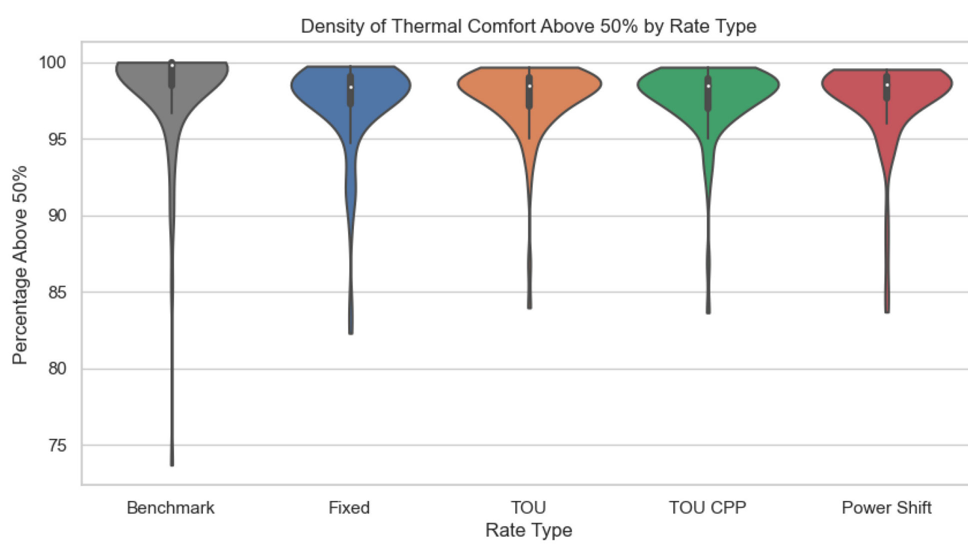


Figure 6.1.2: The distribution of Thermal Comfort Probabilities by rate type

A closer examination of the indoor temperature and thermal comfort probability over the simulation timesteps highlights how the agent opts to push the comfort probability to the 50% threshold to achieve cost savings but very seldom goes below it (Figure 6.1.3). Figure 6.1.4 showcases how this household's internal temperature fluctuates throughout the simulation.

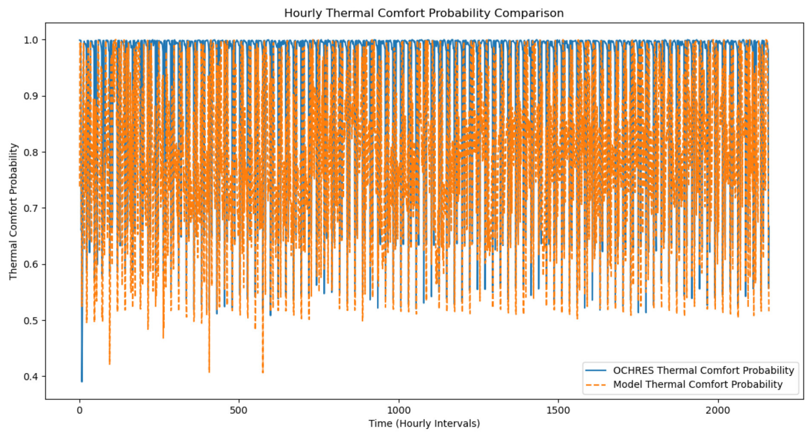


Figure 6.1.3: The Thermal Comfort Probability comparison over every timestep based on the indoor temperature at the timestep (blue for OCHRE Benchmark, orange for TOU agent)

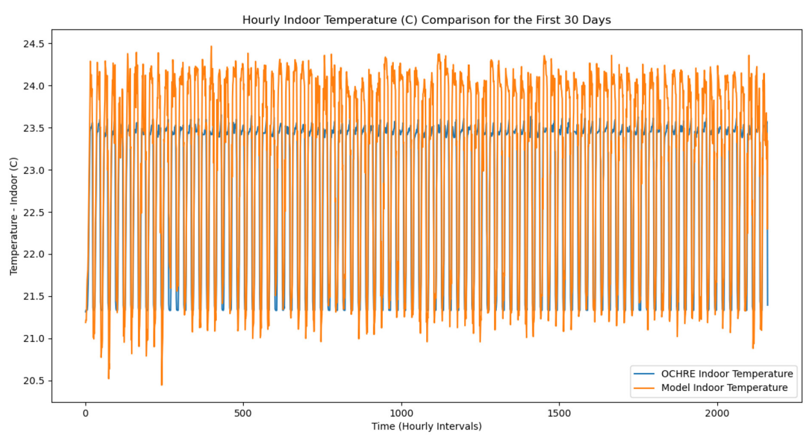


Figure 6.1.4: The internal temperature of the household throughout the simulation (blue for OCHRE Benchmark, orange for TOU agent)

6.2 Total Energy Usage

Total Energy Usage evaluates the agent's effectiveness in reducing the overall amount of energy consumed during the time, providing a direct measurement of the energy impact of different rate designs. It is essential to comprehend how much energy can be saved by optimizing settings within an individual's comfort zone. The formula for calculating Total Energy Usage is straightforward:

$$\text{Total Energy Usage} = \sum_{t=1}^{T} E(t)$$

Where $E(t)$ represents the energy used at each timestep t, and T is the total number of timesteps.

Rate Type	**Percentage Difference (lower is better)**
Fixed	**-2.27%**
Power Shift	**-1.77%**
TOU	**-2.32%**
TOU CPP	**-2.26%**

Table 6.2.1: The Total Energy Reduction from the Benchmark for every rate type

Evaluating the total energy usage across different rate types shows that each rate type can potentially reduce overall energy consumption when paired with an agent. See Figure 6.2.1. In fact, all rate types exhibited a decrease in energy usage compared to a Benchmark. The Time-of-Use (TOU) rate demonstrated the highest reduction in energy usage (2.32%). Additionally, the data indicates that even adding a smart thermostat with a Fixed rate could potentially reduce energy consumption by 2.27%. These results prove the effectiveness of the control agent's optimization goal of reducing energy usage.

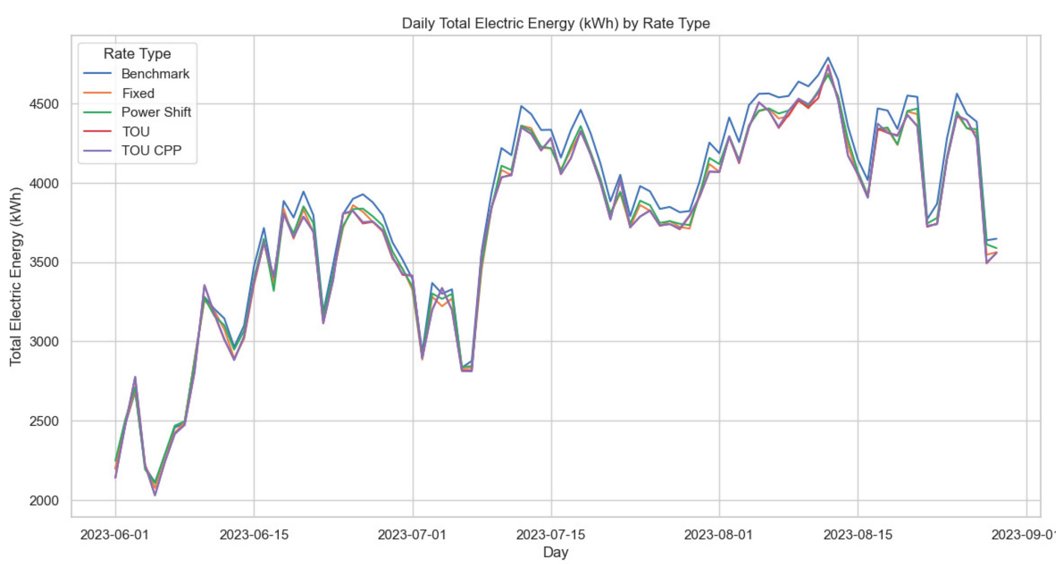


Figure 6.2.1: The Total Electricity Usage (in kWh) throughout the simulation

When analyzing the energy usage data by hour of the day, the agent's different strategies to minimize energy usage become visible (Figure 6.2.2). For TOU and TOU CPP, small pre-cooling bumps are noticeable at 8 AM and 11 AM, which align with the price increase. Similarly, there is a significant increase in energy usage at 5 PM for the Power Shift rate, which corresponds to the hour before the Power Shift's rate increases. The energy usage during the peak rate window (6 PM to 10 PM) significantly decreases before normalizing. The peak usage section below will further analyze the impact of the time varying rates.

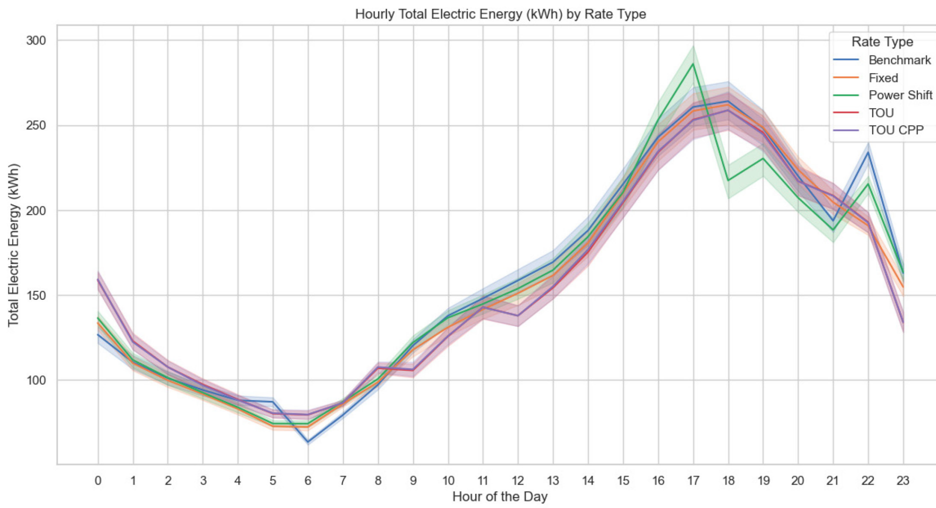


Figure 6.2.2: The hourly Average Total Electric Usage (in kWh) by rate type

6.3 Cost Savings

Cost savings are crucial for encouraging consumers to adopt different tariffs. Savings provide a tangible incentive for users to switch from traditional settings to more dynamic, automated systems. Cost savings are particularly important when consumers are expected to accept minor deviations from their ideal thermal comfort levels in exchange for enhanced efficiency and reduced expenses.

For each rate structure, the cost savings are calculated by integrating the amount of energy used with the corresponding rate at each timestep:

$$\text{Cost during simulation} = \sum_{t=1}^{T} E\,(t) \times \text{Rate}(t)$$

Where $E(t)$ is the energy consumption at timestep t, and $\text{Rate}(t)$ is the cost rate at that same timestep for the different rates (TOU, Power Shift, and Fixed). T represents the total number of timesteps.

The Benchmark total cost is calculated using the tariffs for each rate structure to compare cost savings across different scenarios accurately. The Benchmark total cost represents the rate cost under a non-price-responsive control scenario. The calculation is as follows:

$$\text{Cost in Benchmark scenario} = \sum_{t=1}^{T} E_{\text{Benchmark}}(t) \times \text{Rate}(t)$$

Where $E_{\text{Benchmark}}(t)$ is the energy used at each timestep in the Benchmark scenario. This approach allows for an assessment of how a non-price-responsive system or individual would incur costs under each rate tier, providing a baseline against which to measure the cost-effectiveness of the agent.

Rate Type	Avg Benchmark Cost ($)	Avg Rate Cost ($)	Avg Savings ($)	Savings Percent (%)
Power Shift	876.26	820.83	55.42	**6.32**
TOU	820.42	777.68	42.74	**5.21**
TOU CPP	893.00	849.01	43.99	**4.93**
Fixed	718.79	702.41	16.38	**2.28**

Table 6.3.1: The average savings when utilizing smart thermostat control as opposed to no thermostat control under each of the different rate schemes

Agent-based control with variable rate structures results in significant consumer savings. The data shows that all rate types lead to cost savings compared to a Benchmark scenario. The savings percentage in the table above is based on combining the control agent with the rate instead of non-price responsive usage. Therefore, assuming that the Power Shift rate (6.32% reduction) would provide more savings than the TOU rate (5.21% reduction) is not appropriate. Instead, the incentive for smart thermostat control is higher for the Power Shift rate than the TOU or TOU CPP rates. This is because the agent can better plan for price shifts around the peak windows. As expected, given the lack of a price signal for which to optimize, the Fixed rate provides significantly fewer savings (2.28% reduction). Furthermore, the longer tails in the distribution of cost savings for the time-varying rates indicate a higher variability in the savings achieved by different users. These extended tails suggest that some users may achieve significantly higher savings. Table 6.3.1 above and Figure 6.3.1 below highlight the results.

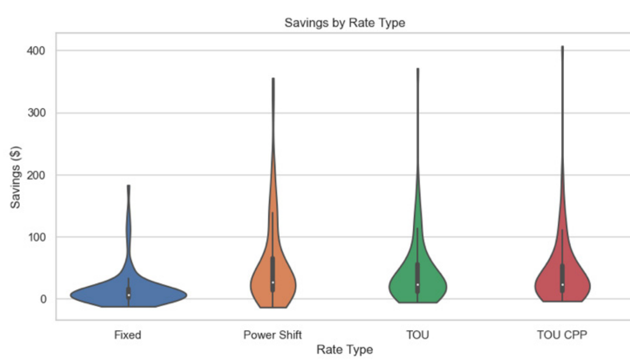


Figure 6.3.1: The distribution of savings when combined with smart thermostat control over the Benchmark with no price-sensitive thermostat control

Figure 6.3.2 demonstrates how cost savings vary by hour for different rate types. The most significant increase in savings occurs just after the rate switches for each time-varying rate, which aligns with expectations. Notably, these savings are not sustained. This suggests that consumers can save the most if they cool their homes before the rate switches, but the savings from pre-cooling quickly diminish as the HVAC returns to normal usage to maintain the environment temperature.

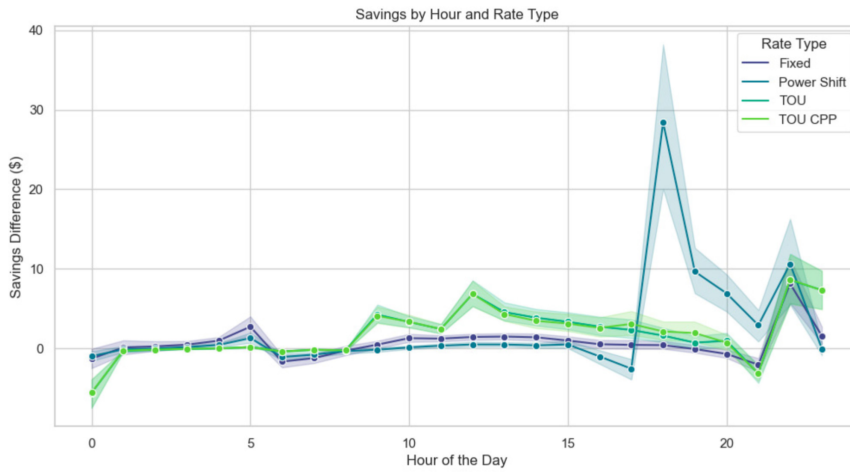


Figure 6.3.2: The savings are broken down by the hour. The effects of saving and pre-cooling are shown in the hours after the rate switches from a high price to lower price

6.4 Peak Reduction

Peak reduction is essential for grid operators and utility companies, who play a critical role in setting rate structures. It is vital for maintaining grid stability by preventing outages, enhancing cost efficiency by reducing reliance on expensive Peaker plants, and decreasing environmental impact by lowering emissions from peak power generation. Additionally, peak reduction helps preserve infrastructure by minimizing strain on the power grid, thereby extending the lifespan and efficiency of energy systems. Peak consumption is determined by identifying the highest cumulative energy usage (Total Electric Power (kW)) at any given timestamp throughout the day. Due to the 30-minute timestep, this approach provides a more precise measure of peak energy demand as it accounts for the continuous operation of power generation and the non-discrete nature of energy consumption cycles instead of considering the entire hour timeframe.

$$\text{Peak Consumption} = \max_i(E(i))$$

Here, $E(i)$ represents the energy usage at each timestep i, and the summation of all the timesteps captures the peak usage within that window.

Rate Type	Max Peak (in kW)	Time	% Difference from Benchmark
Power Shift	372.08	08/10/23 17:30	**3.66**
Fixed	363.23	07/13/23 18:00	**1.19**
Benchmark	358.95	08/10/23 18:00	0
TOU	356.8	08/09/23 18:00	**-0.6**
TOU CPP	351.54	08/12/23 19:00	**-2.06**

Table: 6.4.1 The Max Peak and time for each of the different scenarios

The Power Shift showed the highest peak of the different rate types on August 10, 2023, at 5:30 PM, with a maximum peak of 372.08 kW. Interestingly, this peak occurred half an hour before the second-highest peak from the 2023 ERCOT Load Data. This suggests that this rate type when paired with an optimal cost-minimizing strategy, may overcompensate for price increases and not be effective in reducing critical peaks. The reason is the agent's optimization to pre-cool the building, which encourages shifting consumption to off-peak hours to save money. Figure 6.4.1 confirms this by the significant reduction (18%) in usage half an hour later (during the Benchmark's peak timestamp) when the rate switches to the higher price.

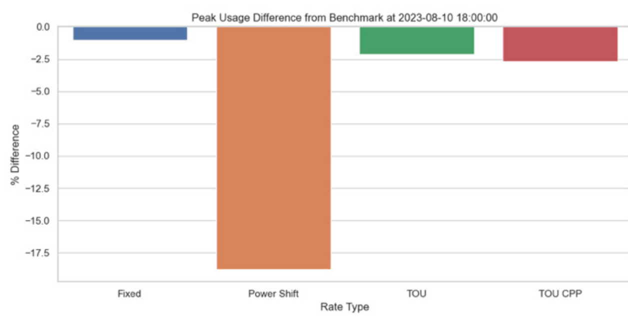


Figure 6.4.1: The percentage difference between the Benchmark scenario's Max Peak (kW) and the rate types of Total Electric Usage (kW) at the same time as the Benchmark's peak (8/23 @ 6 PM).

Furthermore, the Fixed rate with agent control showed a slight peak increase (1.19%) compared with the Benchmark. Given that the Fixed rate's peak date is in July, this result is likely due to the agent overcompensating for temperature changes to maintain thermal comfort. If we remove this outlier date, the second-highest peak for the Fixed rate corresponds with the Benchmark's peak date and time and is 1.04% below the Benchmark's peak. Holding the assumption that the July date was a computational error, this analysis highlights how utilizing fixed rates does not send the appropriate signals to shift usage, but it does point to a slight potential in peak reduction when an agent can adjust to find the optimal point of trade-off between thermal comfort and energy savings. On the other hand, the TOU and TOU CPP demonstrated worthy performance in peak shaving, with the TOU CPP rate showing the lowest peak with a 2% reduction.

This validates the potential of smart thermostat automation and TOU CPP to effectively reduce peak usage through price signals without sacrificing thermal comfort. The key takeaway is understanding that carelessly designed rate structures can produce unintended consequences if participants take advantage of pre-cooling. The agent created new peaks outside the Power Shift peak pricing window by encouraging energy use during off-peak times. Figure 6.4.2 below indicates that the Power Shift rate created higher peaks than the Benchmark rate for nearly every day.

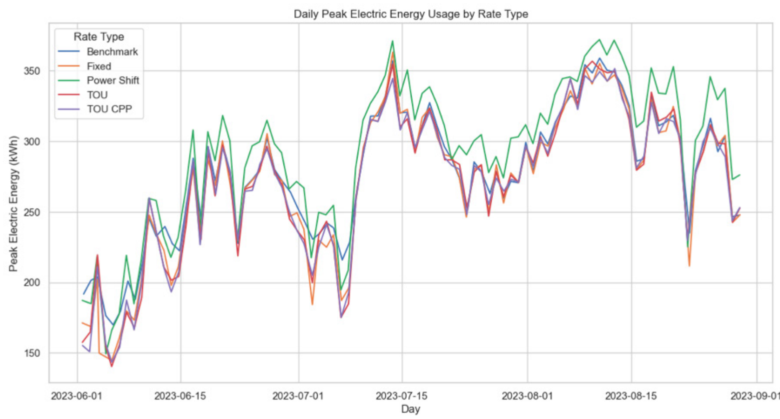


Figure 6.4.2: Daily electric peak (in kW) for each of the different rate types

Further analysis narrowing in on the ERCOT peak dates (2023-08-08, 2023-08-09, 2023-08-10, 2023-08-11, 2023-08-17) during peak times (17:00 - 20:00) highlights the differential impact of CPP and Power Shift pricing. Figure 6.4.3 shows that the CPP maintains a modest peak energy reduction below the Benchmark. Intriguingly, the CPP price does not seem to have a significant pre-cooling effect. This could be because the TOU rate and TOU CPP rate were trained on the same agent, which might have dampened the impact of the CPP rate. To validate this assumption, training the rates on two different agents would be necessary. A slight rebound effect is observed after the demand response (DR) event ends at 20:00, which, though below the daily peak, could pose future challenges if not managed properly. As more renewables are integrated into the grid, energy availability during late evening hours could be affected.

Time	Benchmark	Power Shift (PS)	% Reduction (PS)	TOU CPP	% Reduction (CPP)
17:00:00	332.73	353.32	**-6.19**	317.29	**4.64**
17:30:00	335.46	362.55	**-8.08**	332.27	**0.95**
18:00:00	348.97	276.37	**20.8**	341.39	**2.17**
18:30:00	337.49	307.37	**8.93**	327.01	**3.11**
19:00:00	332.44	315.59	**5.07**	332.33	**0.03**

Table 6.4.1: An analysis of the average of the top 5 peak days and the performance of the time-varying rates to assess the price responsiveness to reduce the peak.

On the other hand, the Power Shift's electric power usage increases dramatically just before the peak rate takes effect (the rate changes at 18:00). This can be visualized in Figure 6.4.3. The Power Shift rate results in a more sustained reduction below the peak compared to the CPP rate. This is likely due to the agent's exposure to the daily increase in peak pricing. There are two potential explanations for this effect. First, the consistent exposure to dramatically increased peak pricing enables the Power Shift agent to handle the peak pricing shift more effectively. On the other hand, the CPP agent experiences extreme price events only for five days in a simulation cycle, resulting in less exposure and suboptimal agent performance. Additionally, the CPP rate was trained alongside the TOU rate for performance efficiency, but this might have reduced the critical peak pricing effect. The other explanation could be that the more moderate step increases in the TOU rates create less dramatic spikes before the peak period than the large price swings associated with the Power Shift rate. This results in a reduced need for significant pre-cooling of the building.

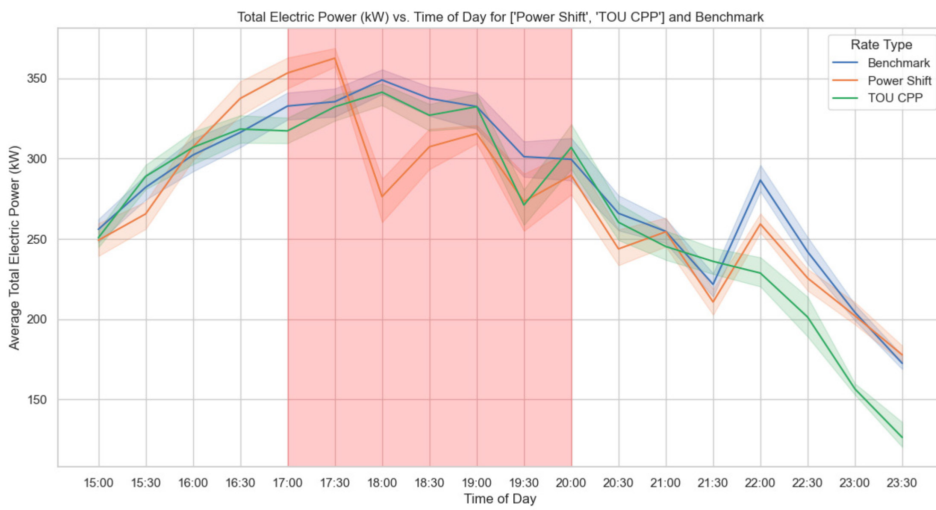


Figure 6.4.3: The average of the top 5 peak days in the system versus the electric usage of the time-varying rates during the same period

6.5 Segmentation Analysis

One of the benefits of the study's simulation approach is the ability to segment households based on characteristics in the metadata available for each Resstock household. To understand how agents and rate designs affect different consumers, two primary characteristics, house size (in square feet) and house vintage (age in years), were analyzed. These factors were chosen because they significantly influence a home's energy usage patterns and thermal "storage" capabilities. Larger homes typically require more energy for heating and cooling. In comparison, older homes often have less efficient insulation, making it challenging to retain heat or cooling, thus requiring a more constant energy flow to maintain desired temperature levels. Each characteristic was divided into three roughly equal-sized groups to facilitate a comparison across groups.

Category	Bucket	Group	Building Count
House Vintage	Old	< 1980	17
	Mid	1980-2000	33
	New	> 2000	32
Floor Area	Small	0-1500 sqft.	37
	Medium	1500-2500 sqft.	29
	Large	2500+ sqft.	16

Table 6.5.1: The breakdown of the different segments and their building counts

House Vintage

Electricity Usage (Figure 6.5.1)

For new homes, all rate types generally show a lower deviation from the Benchmark than middle-aged and older homes, indicating that newer constructions are perhaps already optimized for energy efficiency, resulting in a lesser range of potential flexibility. The mid-category has the highest density of homes, above the Benchmark for electric usage. Interestingly, older homes perform best when paired with an agent based on their thicker, more elongated distribution. Amongst other things, this might indicate that older homes, possibly due to their inherent inefficiencies, have more room for reductions in energy usage, regardless of the rate type.

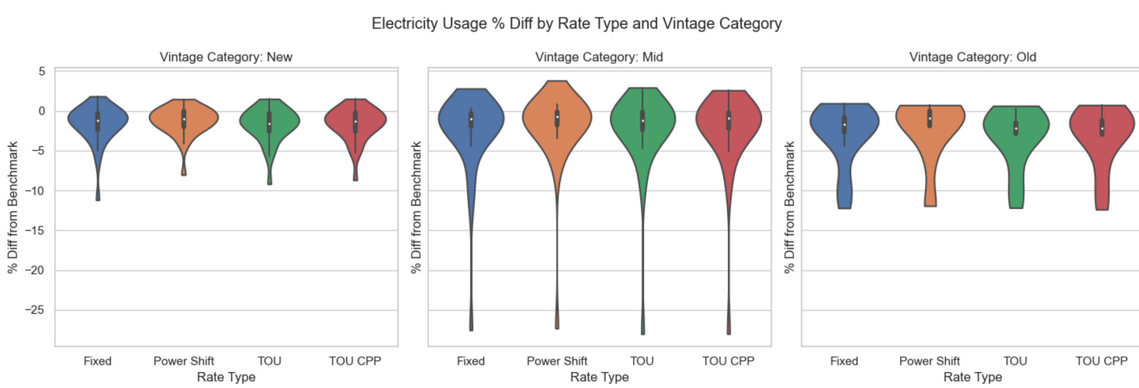


Figure 6.5.1: The distribution of electricity usage by rate type and vintage category

Cost Savings (Figure 6.5.2)

The results show that the savings potential is generally modest for new homes built with modern energy-efficient designs. This indicates that these newer properties are already optimized and have a lower overall savings potential. However, in older homes, using an agent-based smart thermostat has a high potential to help with energy and cost optimizations, resulting in significant savings across all rate types, especially with time-varying rates like Power Shift and TOU.

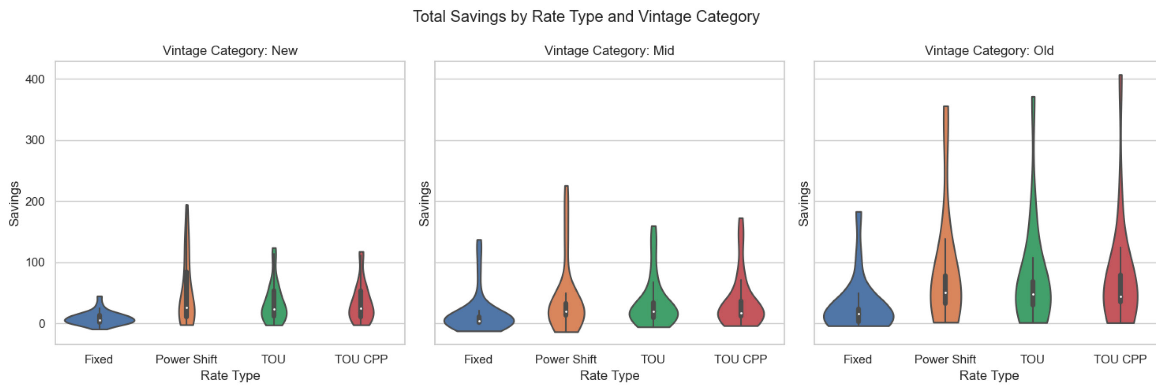


Figure 6.5.2: The distribution of savings over the Benchmark by rate type and vintage category

Floor Area (Square Feet)

Electricity Usage (Figure 6.5.3)

In larger homes, all rate types provide higher energy savings compared to the Benchmark with a similar distribution across rates, implying that larger homes have more capacity to reduce electricity usage below the Benchmark levels. Even a minor adjustment in internal temperature to optimize for the energy-comfort trade-off can lead to significant savings. On the other hand, smaller homes show more variance in energy savings potential and have the highest simulation results that fall above the Benchmark. This pattern intuitively suggests that smaller homes lack the capacity for energy reduction that larger homes have.

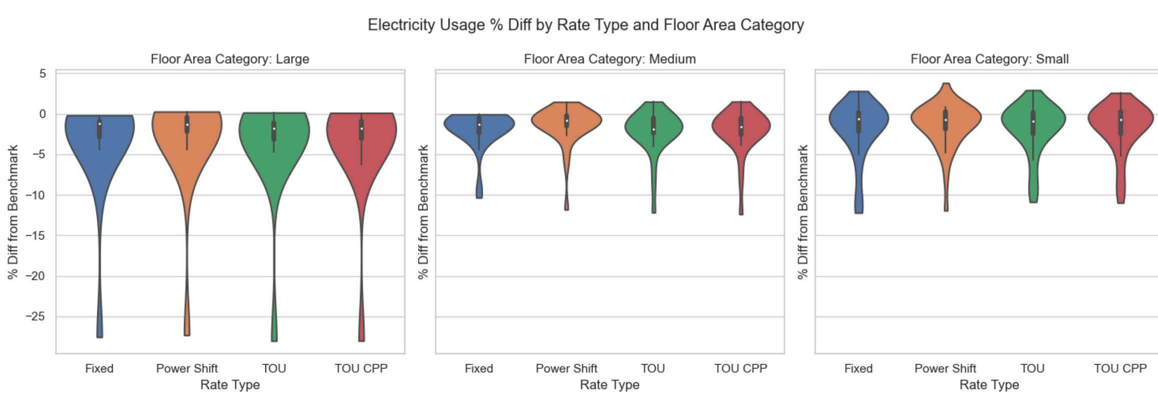


Figure 6.5.3: The distribution of electricity usage by rate type and floor area category

Cost Savings (Figure 6.5.4)

The data shows that larger homes are more likely to benefit from the Power Shift rate, which is more cost-efficient due to pre-cooling to reduce energy costs before peak pricing. The Time of Use (TOU) and TOU Critical Peak Pricing (CPP) rates also offer adequate savings. In medium-sized homes, all rate types show a more extended tail distribution of outcomes, with outliers having higher savings potential than large homes, but the overall mean is below that of larger homes. Small homes exhibit the lowest variability in savings across all rate types, with generally lower savings, reflecting their smaller baseline energy usage.

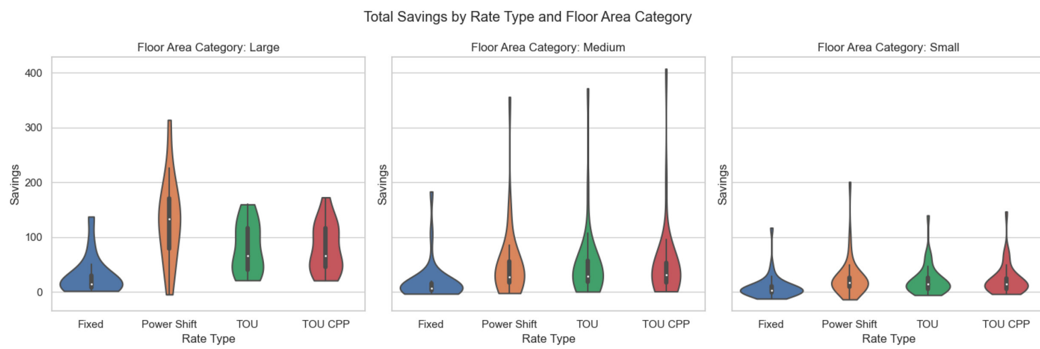


Figure 6.5.4: The distribution of savings over the Benchmark by rate type and floor area category

6.6 High Flexibility

This section analyzes a specific instance of a highly flexible household to showcase the effects of the agent. For the Fixed rate, energy consumption remains constant throughout the day, indicating that the Fixed rate does little to motivate changes in energy consumption habits. However, with the Time-of-Use (TOU) rate, a more dynamic consumption pattern is observed, with visible adjustments in usage that correspond to the varying rates throughout the day. These spikes have a minimal impact on the system, and the scaled TOU rate based on the partitioning algorithm is proven effective. The individual household peak is reduced for the TOU rate, while the benefits of pre-cooling provide cost savings to the household. In contrast, the Power Shift rate significantly differs from the Benchmark during evening peak hours. There is a noticeable reduction in energy use during the peak pricing window, but it is preceded by a significant increase in consumption just before the pricing switch. In Figure 6.6.1, along the second Y-axis in the chart, the thermal flexibility is observed as the indoor temperature exhibits variations throughout the day to optimize for price.

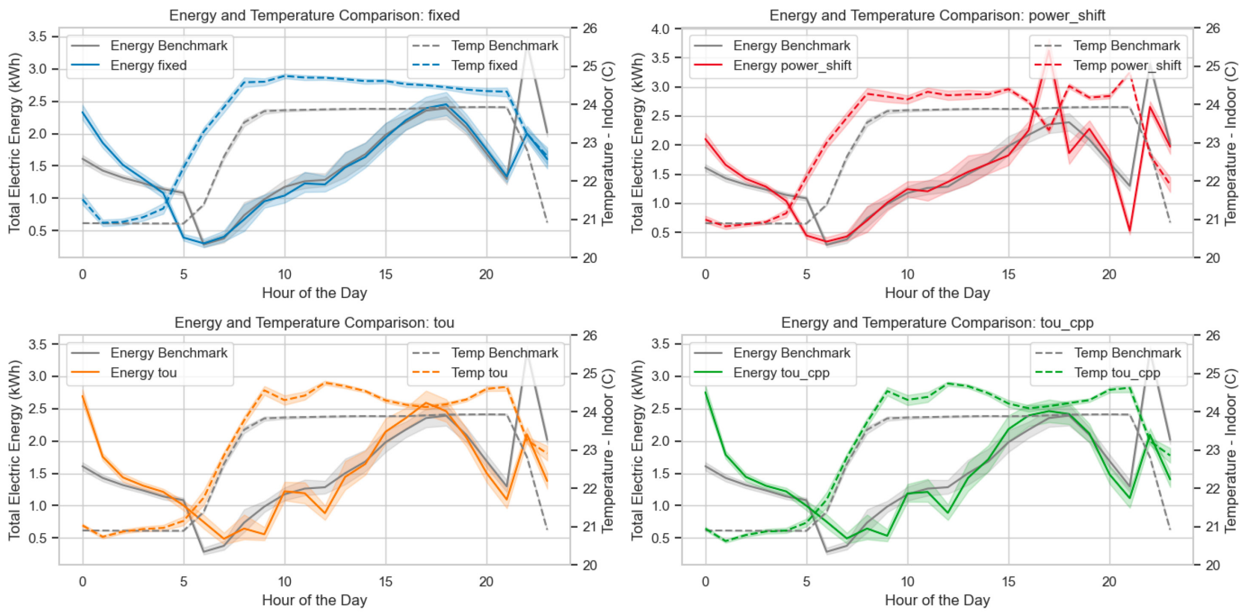


Figure 6.6.1: The average hourly usage (first Y-axis) and indoor temperature (second Y-axis) by rate type for a household with high flexibility

6.7 Low Flexibility

The analysis below considers a household with limited capacity to adjust energy consumption. Figure 6.7.1 illustrates the low level of flexibility due to a smaller energy-use footprint and a more tightly clustered thermal comfort preference (Figure 6.7.2). The hourly electricity usage graph shows that the household's ability to shift or reduce energy consumption is significantly limited regardless of the rate structure implemented. The second y-axis in the chart highlights the limited flexibility in the indoor temperature with minimal opportunity for pre-cooling. The usage patterns largely align with the Benchmarks. The narrow range of thermal comfort offers minimal scope for adjustments that could lead to energy savings without impacting comfort levels. These results highlight households' challenges with low flexibility in leveraging rate structures and technology for energy savings. Energy-saving strategies through rate manipulation may have limited effectiveness for such homes, emphasizing the need for alternative approaches or technologies that can provide benefits without requiring significant deviations from established comfort patterns.

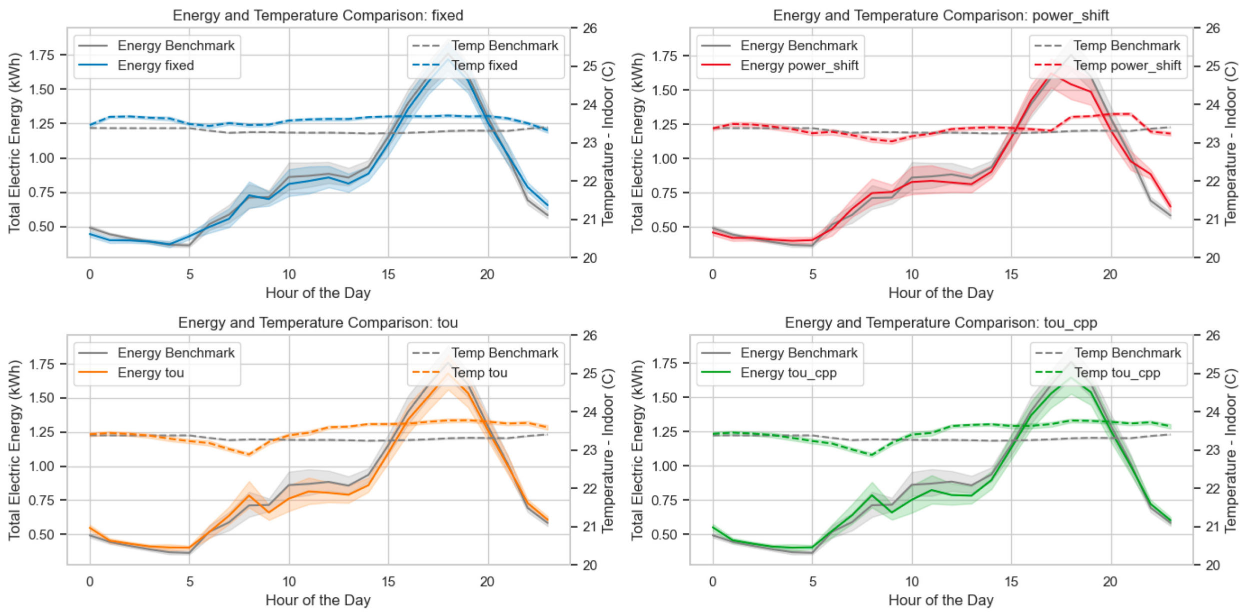


Figure 6.7.1: The average hourly usage (first Y-axis) and indoor temperature (second Y-axis) by rate type for a household with low flexibility

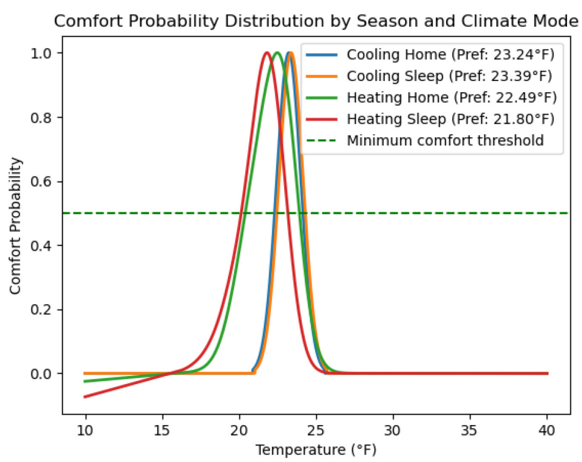


Figure 6.7.2: The Thermal Comfort Probability distribution for the low-flexibility household

7. Discussion and Conclusion

Based on the simulation results, it is evident that an agent that maximizes the balance between thermal comfort and price responsiveness can help reduce peak load if it is paired with the appropriate rate design. The smart thermostat control agents provided comparable comfort to the Rule-Based Controller, with less than a 1% difference in maintaining thermal comfort above 50%. These agents optimized energy consumption without significantly compromising user comfort, particularly during high-demand periods. Additionally, dynamic rate structures have trade-offs, varying effectiveness based on household flexibility. Understanding these nuanced effects is crucial for designing rate structures that maximize efficiency without compromising comfort. The research shows that older and larger homes exhibit greater flexibility and cost-saving potential than newer and smaller homes, owing to their higher thermal comfort capacity and larger scope for energy optimization.

Furthermore, the design of carefully partitioned time-of-use (TOU) rates can encourage a reduction in energy usage without creating new peaks. This is most likely because the pricing shifts during peak events are less dramatic. On the other hand, the Power Shift pricing model can create new peaks as the thermostat agents anticipate price hikes. This finding is critical as it parallels the rebound effect observed in demand response (DR) events, suggesting that encouraging pre-cooling through price incentives must be carefully designed to avoid unintended consequences. In terms of fixed rates, utilizing smart thermostat control can help reduce overall energy consumption, where less energy means fewer carbon emissions, but it fails to support the reduction in the peak, which has a higher impact on grid efficiency.

Lastly, this approach demonstrates how the law of large numbers can help mitigate outlier effects from individual households, resulting in a net positive outcome. The Fixed rate showed a tighter impact distribution, while the time-varying rates displayed a more considerable variance of reactivity to different pricing schemes. It is essential to scale up the simulation setups to include more simulations and environments to understand the full effectiveness of energy efficiency strategies such as smart thermostat automation. It's not enough to view the effects of any mechanism in isolation; the bigger picture must be kept in context. As demonstrated by this work, the net impact of different time-varying rate structures resulted in a higher distribution of electricity consumption and peak usage, but the aggregation of all the households run in the simulation reduced the deviations and smoothed the distribution.

7.1 Limitations

While this study is extensive, it has some limitations to be considered. One primary constraint is that the OCHRE platform models single occupancy zones, which only simulate single setpoint temperatures for the entire house, neglecting the complexities of multi-zone heating and cooling systems that vary temperatures across different rooms. This simplification may not accurately capture energy usage and comfort in residences with more complex layouts. Additionally, the thermal comfort models employed in this study assume that the registered internal household temperature is a representation of their personal comfort preference. However, this generalization might not align with the varied personal comfort preferences based on other factors, such as socio-economic status, where some individuals might set their thermostats outside their comfort range to save money. The study also relied exclusively on data from the ecobee smart thermostat program, further introducing potential biases. This dataset may not fully represent the diversity of thermal preferences across the general population, skewing results toward the behaviors and preferences of smart thermostat owners.

Furthermore, the simulations intentionally had a limited action space to increase model convergence and efficiency. This was done by excluding unnecessary heating actions during summer. Although this restriction helps with computational efficiency, it simplifies the control environment. With that said, in the real world, most thermostats can selectively enable heating and cooling modes, which restrict instances where the heater might attempt to turn on in the summer (a highly, highly unlikely event in Texas). Moreover, initial efforts to train the model simultaneously on four different rate structures did not produce reliable results. Therefore, separate training sessions were conducted for each rate structure. Multiple approaches, including behavioral cloning, initiating the model with a set of "expert" examples to learn from instead of cold starting, and staged learning, were tested to overcome these obstacles, but they were unsuccessful in either improving the model's performance or significantly enhancing its efficiency. It's worth noting that other research in the space has seen some success using imitation learning, but their approach utilized a different type of reinforcement learning algorithm [100]. This difficulty highlights the inherent complexities of creating a model that can adapt to multiple complex rate structures without compromising learning quality.

Lastly, it should go without saying that reinforcement learning is a difficult task. Minor changes in the configuration can significantly impact its performance. Fractional changes in the reward function can lead to unintended behaviors that, while seemingly inappropriate, are to be expected based on the design. For example, at one point, the reward function included a term to reward energy efficiency by dividing the comfort probability by the energy cost (energy usage times price). The goal was to find an energy efficiency metric that weighed the incremental increase in thermal comfort vs. the rise in energy price. However, the model learned to adopt counterintuitive behavior due to the price component. The agent minimized energy usage by letting the temperature rise to the top of the thermal comfort threshold in the building just before an energy price hike to reduce the denominator and maximize the reward. As a result, there was a slight loss in reward for re-cooling the building during a high price period, but a much more significant reward was achieved in the timesteps before. This variability in the performance of reinforcement learning models highlights the need for a balanced approach while tuning and applying these algorithms. These limitations highlight crucial areas for further research, such as developing more sophisticated models capable of handling multi-zone systems, incorporating more diverse datasets, and refining learning algorithms to manage a broader array of complex rate structures more efficiently.

7.2 Future Work

This research sets the groundwork for multiple opportunities for future work. The simulations can expand their geographic scope to include different climates and regions. This would offer more comprehensive insights into the effectiveness of smart thermostats across different regions. The simulation can be extended to cover both winter and summer seasons to better understand the year-round performance and potential energy savings under different thermal demands. Running simulations over a year can provide data on annual energy usage patterns, offering a more detailed view of how different rate structures perform throughout all seasons. Additionally, different rate designs, such as staggering time-of-use rate blocks for different individuals at varying times, could help understand how to create smaller peaks while achieving the benefits of actions such as pre-cooling. Moreover, exploring scenarios that include adopting photovoltaic systems (PV), battery storage, heat pumps, and other technologies could reveal opportunities that enhance energy efficiency and demand flexibility.

The environment can also be fine-tuned to account for occupancy and multiple zones. Utilizing predictive models to gauge occupancy patterns could significantly improve the optimization of building energy usage. Research has shown that HVAC energy savings ranging from 1% to 20% can be achieved with occupancy-based controls, indicating the potential benefits of this approach. Furthermore, expanding the models to manage multi-zone heating and cooling more effectively can lead to substantial improvements in energy efficiency and occupant comfort.

Lastly, as the field of reinforcement learning continues to expand, researchers are exploring new ways to improve the efficiency and generalization of models. One promising approach involves using multi-task learning strategies in reinforcement learning models, which could enable simultaneous learning across various rates and thermostats without attaching a single thermostat to a single building. Another potential avenue is leveraging transfer learning techniques to apply knowledge gained in one setting to others, which could increase the robustness and adaptability of models, allowing them to perform well across different rate structures and technological scenarios. Although transfer learning has not yet shown significant success in HVAC control to date, it is an exciting area to continue exploring, especially as it creates the potential for offline models to be trained and transferred to an online (live thermostat) environment [101].

Wrapping Up

This thesis presents a new simulation framework designed to maximize heating and cooling flexibility through thermostat automation and rate design. The framework builds on existing control methods and introduces innovative contributions to the field. These include a quantitative analysis of different rate designs on demand flexibility at a population scale, integrating individualized energy-comfort trade-offs into reinforcement learning models for personalized HVAC control, and using smart thermostat data as thermal comfort profiles to predict demand flexibility under different retail tariffs across a population. These advancements provide a comprehensive understanding of how pricing strategies affect grid efficiency and consumer behavior, leading to more efficient and personalized flexibility solutions.

Declaration of Use of Generative AI During the Writing Process: While preparing this work, the author used Grammarly and ChatGPT to check grammar and improve readability. After using these services, the content was further reviewed and edited, and the author takes full responsibility for the publication's content.

References

[1] U.S. DOE, "AN ASSESSMENT OF ENERGY TECHNOLOGIES AND RESEARCH OPPORTUNITIES." Department of Energy, Sep. 2015.

[2] M. Blair, S. Alizamir, and S. Wang, "The Impact of Climate Change: An Empirical Analysis of Smart Thermostat Data." Rochester, NY, Nov. 18, 2022. doi: 10.2139/ssrn.4282220.

[3] "Why Energy Efficiency Upgrades," Energy.gov. Accessed: May 15, 2024. [Online]. Available: https://www.energy.gov/energysaver/why-energy-efficiency-upgrades

[4] J. D. Wilson and Z. Zimmerman, "The Era of Flat Power Demand is Over".

[5] "New England's Electricity Use." Accessed: May 15, 2024. [Online]. Available: https://www.iso-ne.com/about/key-stats/electricity-use

[6] "ERCOT Provides New Look at Peak Demand Records." Accessed: May 15, 2024. [Online]. Available: https://www.ercot.com/news/release/2023-09-14-ercot-provides-new

[7] "PJM Publishes 2024 Long-Term Load Forecast | PJM Inside Lines." Accessed: May 15, 2024. [Online]. Available: https://insidelines.pjm.com/pjm-publishes-2024-long-term-load-forecast/

[8] "California ISO Peak Load History 1998 through 2023." Accessed: May 15, 2024. [Online]. Available: https://www.caiso.com/documents/californiaisopeakloadhistory.pdf

[9] "ERCOT Load Forecast." Accessed: May 15, 2024. [Online]. Available: https://www.ercot.com/gridinfo/load/forecast

[10] C. W. Gellings, "The concept of demand-side management for electric utilities," *Proc. IEEE*, vol. 73, no. 10, pp. 1468–1470, 1985, doi: 10.1109/PROC.1985.13318.

[11] "Frequently Asked Questions (FAQs) - U.S. Energy Information Administration (EIA)." Accessed: May 15, 2024. [Online]. Available: https://www.eia.gov/tools/faqs/faq.php

[12] R. Glick, J. Danly, A. Clements, M. C. Christie, and W. L. Phillips, "2021 Assessment of Demand Response and Advanced Metering," 2021.

[13] M. Peplinski and K. T. Sanders, "Residential electricity demand on CAISO Flex Alert days: a case study of voluntary emergency demand response programs," *Environ. Res. Energy*, vol. 1, no. 1, p. 015002, Mar. 2024, doi: 10.1088/2753-3751/ad0fda.

[14] "Mass Save Marketplace| Home." Accessed: May 15, 2024. [Online]. Available: https://www.poweredbyefi.org/masssave

[15] H. Li, Z. Wang, T. Hong, and M. A. Piette, "Energy flexibility of residential buildings: A systematic review of characterization and quantification methods and applications," *Adv. Appl. Energy*, vol. 3, p. 100054, Aug. 2021, doi: 10.1016/j.adapen.2021.100054.

[16] R. Yin *et al.*, "Quantifying flexibility of commercial and residential loads for demand response using setpoint changes," *Appl. Energy*, vol. 177, pp. 149–164, Sep. 2016, doi: 10.1016/j.apenergy.2016.05.090.

[17] D. Li, W.-Y. Chiu, and H. Sun, "Chapter 7 - demand side management in microgrid control systems," in *Microgrid*, M. S. Mahmoud, Ed., Butterworth-Heinemann, 2017, pp. 203–230. doi: https://doi.org/10.1016/B978-0-08-101753-1.00007-3.

[18] X. Xu, C. Chen, X. Zhu, and Q. Hu, "Promoting acceptance of direct load control programs in the United States: Financial incentive versus control option," *Energy*, vol. 147, pp. 1278–1287, Mar. 2018, doi: 10.1016/j.energy.2018.01.028.

[19] "Thousands of Xcel customers locked out of thermostats during 'energy emergency,'" Denver 7 Colorado News (KMGH). Accessed: May 15, 2024. [Online]. Available: https://www.denver7.com/news/contact-denver7/thousands-of-xcel-customers-locked-out-of-thermostats-during-energy-emergency

[20] "Demand Response," Energy.gov. Accessed: May 15, 2024. [Online]. Available: https://www.energy.gov/oe/demand-response

[21] J. S. Lopes and P. Agnew, "FPL Residential Thermostat Load Control Pilot Project Evaluation".

[22] S. Diego, "2005 Smart Thermostat Program Impact Evaluation".

[23] "CONSOLIDATED EDISON COMPANY OF NEW YORK, INC. REPORT ON PROGRAM PERFORMANCE AND COST EFFECTIVENESS OF DEMAND RESPONSE PROGRAMS - 2021."

[24] D. S. Callaway and I. A. Hiskens, "Achieving Controllability of Electric Loads," *Proc. IEEE*, vol. 99, no. 1, pp. 184–199, Jan. 2011, doi: 10.1109/JPROC.2010.2081652.

[25] I. Schneider and C. R. Sunstein, "Behavioral Considerations for Effective Time-Varying Electricity Prices".

[26] B. E. D. and E. Foxhall, "Appeals court says state agency set electricity prices too high during 2021 winter storm," The Texas Tribune. Accessed: May 15, 2024. [Online]. Available: https://www.texastribune.org/2023/03/17/puc-appeals-court-uri-prices/

[27] G. R. Newsham and B. G. Bowker, "The effect of utility time-varying pricing and load control strategies on residential summer peak electricity use: A review," *Energy Policy*, vol. 38, no. 7, pp. 3289–3296, Jul. 2010, doi: 10.1016/j.enpol.2010.01.027.

[28] T. Schittekatte, D. S. Mallapragada, P. L. Joskow, and R. Schmalensee, "Electricity Retail Rate Design in a Decarbonized Economy: An Analysis of Time-Of-Use and Critical Peak Pricing." in Working Paper Series. National Bureau of Economic Research, Oct. 2022. doi: 10.3386/w30560.

[29] "Commonwealth Edison Company's Hourly Pricing," Apr. 2023. Accessed: May 15, 2024. [Online]. Available: https://icc.illinois.gov/downloads/public/edocket/587138.PDF

[30] T. Schittekatte, D. Mallapragada, P. L. Joskow, and R. Schmalensee, "Economy-Wide Decarbonization Requires Fixing Retail Electricity Rates," 1977.

[31] A. Olson *et al.*, "Rate Design for the Energy Transition," *ENERGY Syst.*.

[32] B. Baatz, "Rate Design Matters: The Intersection of Residential Rate Design and Energy Efficiency".

[33] N. Aqilah, H. B. Rijal, and S. A. Zaki, "A Review of Thermal Comfort in Residential Buildings: Comfort Threads and Energy Saving Potential," *Energies*, vol. 15, no. 23, p. 9012, Nov. 2022, doi: 10.3390/en15239012.

[34] R. J. de Dear and G. S. Brager, "Developing an Adaptive Model of Thermal Comfort and Preference".

[35] L. Lan, P. Wargocki, and Z. Lian, "Quantitative measurement of productivity loss due to thermal discomfort," *Energy Build.*, vol. 43, no. 5, pp. 1057–1062, May 2011, doi: 10.1016/j.enbuild.2010.09.001.

[36] G. Serale, M. Fiorentini, A. Capozzoli, D. Bernardini, and A. Bemporad, "Model Predictive Control (MPC) for Enhancing Building and HVAC System Energy Efficiency: Problem Formulation, Applications and Opportunities," *Energies*, vol. 11, no. 3, p. 631, Mar. 2018, doi: 10.3390/en11030631.

[37] S. Taheri, P. Hosseini, and A. Razban, "Model predictive control of heating, ventilation, and air conditioning (HVAC) systems: A state-of-the-art review," *J. Build. Eng.*, vol. 60, p. 105067, Nov. 2022, doi: 10.1016/j.jobe.2022.105067.

[38] R. Kwadzogah, M. Zhou, and S. Li, "Model predictive control for HVAC systems — A review," in *2013 IEEE International Conference on Automation Science and Engineering (CASE)*, Madison, WI, USA: IEEE, Aug. 2013, pp. 442–447. doi: 10.1109/CoASE.2013.6654072.

[39] C. Marantos, C. Lamprakos, K. Siozios, and D. Soudris, "Towards Plug&Play Smart Thermostats for Building's Heating/Cooling Control," in *IoT for Smart Grids*, K. Siozios, D. Anagnostos, D. Soudris, and E. Kosmatopoulos, Eds., in Power Systems. , Cham: Springer International Publishing, 2019, pp. 183–207. doi: 10.1007/978-3-030-03640-9_10.

[40] S. Sierla, H. Ihasalo, and V. Vyatkin, “A Review of Reinforcement Learning Applications to Control of Heating, Ventilation and Air Conditioning Systems,” *Energies*, vol. 15, no. 10, p. 3526, May 2022, doi: 10.3390/en15103526.

[41] C. Zhang, S. R. Kuppannagari, R. Kannan, and V. K. Prasanna, “Building HVAC Scheduling Using Reinforcement Learning via Neural Network Based Model Approximation,” in *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, Nov. 2019, pp. 287–296. doi: 10.1145/3360322.3360861.

[42] N. S. Raman, A. M. Devraj, P. Barooah, and S. P. Meyn, “Reinforcement Learning for Control of Building HVAC Systems,” in *2020 American Control Conference (ACC)*, Denver, CO, USA: IEEE, Jul. 2020, pp. 2326–2332. doi: 10.23919/ACC45564.2020.9147629.

[43] D. Wang, W. Zheng, Z. Wang, Y. Wang, X. Pang, and W. Wang, “Comparison of reinforcement learning and model predictive control for building energy system optimization,” *Appl. Therm. Eng.*, vol. 228, p. 120430, Jun. 2023, doi: 10.1016/j.applthermaleng.2023.120430.

[44] J. Arroyo, C. Manna, F. Spiessens, and L. Helsen, “Reinforced model predictive control (RL-MPC) for building energy management,” *Appl. Energy*, vol. 309, p. 118346, Mar. 2022, doi: 10.1016/j.apenergy.2021.118346.

[45] M. Meimand, V. Khattar, Z. Yazdani, F. Jazizadeh, and M. Jin, “TUNEOPT: An Evolutionary Reinforcement Learning HVAC Controller For Energy-Comfort Optimization Tuning,” in BuildSys ’23. New York, NY, USA: Association for Computing Machinery, Nov. 2023, pp. 265–268. doi: 10.1145/3600100.3623751.

[46] C. Fu and Y. Zhang, “Research and Application of Predictive Control Method Based on Deep Reinforcement Learning for HVAC Systems,” *IEEE Access*, vol. 9, pp. 130845–130852, 2021, doi: 10.1109/ACCESS.2021.3114161.

[47] V. Taboga, A. Bellahsen, and H. Dagdougui, “An Enhanced Adaptivity of Reinforcement Learning-Based Temperature Control in Buildings Using Generalized Training,” *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 6, no. 2, pp. 255–266, Apr. 2022, doi: 10.1109/TETCI.2021.3066999.

[48] J. Drgoňa *et al.*, “All you need to know about model predictive control for buildings,” *Annu. Rev. Control*, vol. 50, pp. 190–232, 2020, doi: 10.1016/j.arcontrol.2020.09.001.

[49] M. Soudari, S. Srinivasan, S. Balasubramanian, J. Vain, and U. Kotta, “Learning based personalized energy management systems for residential buildings,” *Energy Build.*, vol. 127, pp. 953–968, Sep. 2016, doi: 10.1016/j.enbuild.2016.05.059.

[50] A. Amadeh, Z. E. Lee, and K. M. Zhang, “Quantifying demand flexibility of building energy systems under uncertainty,” *Energy*, vol. 246, p. 123291, May 2022, doi: 10.1016/j.energy.2022.123291.

[51] D. Azuatalam, W.-L. Lee, F. De Nijs, and A. Liebman, “Reinforcement learning for whole-building HVAC control and demand response,” *Energy AI*, vol. 2, p. 100020, Nov. 2020, doi: 10.1016/j.egyai.2020.100020.

[52] L. Yu *et al.*, “Multi-Agent Deep Reinforcement Learning for HVAC Control in Commercial Buildings.” arXiv, Jul. 22, 2020. Accessed: Apr. 20, 2024. [Online]. Available: http://arxiv.org/abs/2006.14156

[53] G. Gao, J. Li, and Y. Wen, “DeepComfort: Energy-Efficient Thermal Comfort Control in Buildings Via Reinforcement Learning,” *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8472–8484, Sep. 2020, doi: 10.1109/JIOT.2020.2992117.

[54] Y. Du *et al.*, “Intelligent multi-zone residential HVAC control strategy based on deep reinforcement learning”.

[55] O. Al-Ani and S. Das, “Reinforcement Learning: Theory and Applications in HEMS,” *Energies*, vol. 15, no. 17, Art. no. 17, Jan. 2022, doi: 10.3390/en15176392.

[56] J.-Y. Joo and M. D. Ilic, “An Information Exchange Framework Utilizing Smart Buildings for Efficient Microgrid Operation,” *Proc. IEEE*, vol. 104, no. 4, pp. 858–864, Apr. 2016, doi: 10.1109/JPROC.2016.2526119.

[57] J.-Y. Joo and M. D. Ilic, “A multi-layered adaptive load management (ALM) system: Information exchange between market participants for efficient and reliable energy use,” in *IEEE PES T&D 2010*, New Orleans, LA: IEEE, Apr. 2010, pp. 1–7. doi: 10.1109/TDC.2010.5484311.

[58] R. Jaddivada, M. D. Ilic, M. G.- Sanz, M. Marden, and S. Glavaski, “Energy control of HVAC units for provable ancillary service provision.” arXiv, Oct. 04, 2022. Accessed: May 16, 2024. [Online]. Available: http://arxiv.org/abs/2111.12039

[59] L. Earle, J. Maguire, P. Munankarmi, and D. Roberts, “The impact of energy-efficiency upgrades and other distributed energy resources on a residential neighborhood-scale electrification retrofit,” *Appl. Energy*, vol. 329, p. 120256, Jan. 2023, doi: 10.1016/j.apenergy.2022.120256.

[60] “Gymnasium Documentation.” Accessed: May 15, 2024. [Online]. Available: https://gymnasium.farama.org/index.html

[61] J. R. Vázquez-Canteli, J. Kämpf, G. Henze, and Z. Nagy, “CityLearn v1.0: An OpenAI Gym Environment for Demand Response with Deep Reinforcement Learning,” in *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, New York NY USA: ACM, Nov. 2019, pp. 356–357. doi: 10.1145/3360322.3360998.

[62] Z. Wang, B. Chen, H. Li, and T. Hong, “AlphaBuilding ResCommunity: A multi-agent virtual testbed for community-level load coordination,” *Adv. Appl. Energy*, vol. 4, p. 100061, Nov. 2021, doi: 10.1016/j.adapen.2021.100061.

[63] I. G. Balázs, A. Fodor, and A. Magyar, “Quantification of the Flexibility of Residential Prosumers,” *Energies*, vol. 14, no. 16, p. 4860, Aug. 2021, doi: 10.3390/en14164860.

[64] R. De Coninck and L. Helsen, “Quantification of flexibility in buildings by cost curves – Methodology and application,” *Appl. Energy*, vol. 162, pp. 653–665, Jan. 2016, doi: 10.1016/j.apenergy.2015.10.114.

[65] D. P. Chassin, J. Stoustrup, P. Agathoklis, and N. Djilali, “A new thermostat for real-time price demand response: Cost, comfort and energy impacts of discrete-time control without deadband,” *Appl. Energy*, vol. 155, pp. 816–825, Oct. 2015, doi: 10.1016/j.apenergy.2015.06.048.

[66] Y. Du *et al.*, “Multi-task deep reinforcement learning for intelligent multi-zone residential HVAC control,” *Electr. Power Syst. Res.*, vol. 192, p. 106959, Mar. 2021, doi: 10.1016/j.epsr.2020.106959.

[67] S. Yang, H. O. Gao, and F. You, “Demand flexibility and cost-saving potentials via smart building energy management: Opportunities in residential space heating across the US,” *Adv. Appl. Energy*, vol. 14, p. 100171, Jul. 2024, doi: 10.1016/j.adapen.2024.100171.

[68] H. Elehwany, M. Ouf, B. Gunay, N. Cotrufo, and J.-S. Venne, “A reinforcement learning approach for thermostat setpoint preference learning,” *Build. Simul.*, vol. 17, no. 1, pp. 131–146, Jan. 2024, doi: 10.1007/s12273-023-1056-7.

[69] A. T. Nguyen, D. H. Pham, B. L. Oo, M. Santamouris, Y. Ahn, and B. T. H. Lim, “Modelling building HVAC control strategies using a deep reinforcement learning approach,” *Energy Build.*, vol. 310, p. 114065, May 2024, doi: 10.1016/j.enbuild.2024.114065.

[70] “CenterForTheBuiltEnvironment/ashrae-db-II.” CenterForTheBuiltEnvironment, Apr. 05, 2024. Accessed: May 15, 2024. [Online]. Available: https://github.com/CenterForTheBuiltEnvironment/ashrae-db-II

[71] “End-Use Load Profiles for the U.S. Building Stock.” Accessed: May 15, 2024. [Online]. Available: https://www.nrel.gov/buildings/end-use-load-profiles.html

[72] Z. E. Lee and K. Max Zhang, “Unintended consequences of smart thermostats in the transition to electrified heating,” *Appl. Energy*, vol. 322, p. 119384, Sep. 2022, doi: 10.1016/j.apenergy.2022.119384.

[73] M. Blonsky, J. Maguire, K. McKenna, D. Cutler, S. P. Balamurugan, and X. Jin, “OCHRE: The Object-oriented, Controllable, High-resolution Residential Energy Model for Dynamic Integration Studies,” *Appl. Energy*, vol. 290, Art. no. NREL/JA-5D00-77596, May 2021, doi: 10.1016/j.apenergy.2021.116732.

[74] J. Chen *et al.*, “Stochastic simulation of residential building occupant-driven energy use in a bottom-up model of the U.S. housing stock.” arXiv, Nov. 03, 2021. Accessed: Apr. 18, 2023. [Online]. Available: http://arxiv.org/abs/2111.01881

[75] “HPXML.” Accessed: May 15, 2024. [Online]. Available: https://www.hpxmlonline.com/

[76] J. Reyna *et al.*, “U.S. Building Stock Characterization Study: A National Typology for Decarbonizing U.S. Buildings,” *Renew. Energy*, 2022.

[77] “Gymnasium Documentation.” Accessed: May 15, 2024. [Online]. Available: https://gymnasium.farama.org/content/basic_usage.html

[78] “Basics - OCHRE Gym.” Accessed: May 15, 2024. [Online]. Available: https://nrel.github.io/ochre_gym/Getting%20Started/basics/#the-basic-ochre-gym-rl-loop

[79] “Modeling Approach — OCHRE documentation.” Accessed: May 15, 2024. [Online]. Available: https://ochre-docs-final.readthedocs.io/en/latest/ModelingApproach.html

[80] “Donate Your Data | ecobee.” Accessed: May 15, 2024. [Online]. Available: https://www.ecobee.com/en-us/donate-your-data/

[81] W. Jung and F. Jazizadeh, “Comparative assessment of HVAC control strategies using personal thermal comfort and sensitivity models,” *Build. Environ.*, vol. 158, pp. 104–119, Jul. 2019, doi: 10.1016/j.buildenv.2019.04.043.

[82] J. Kim, Y. Zhou, S. Schiavon, P. Raftery, and G. Brager, “Personal comfort models: Predicting individuals’ thermal preference using occupant heating and cooling behavior and machine learning,” *Build. Environ.*, vol. 129, pp. 96–106, Feb. 2018, doi: 10.1016/j.buildenv.2017.12.011.

[83] T. Peffer, D. Perry, M. Pritoni, C. Aragon, and A. Meier, “Facilitating energy savings with programmable thermostats: evaluation and guidelines for the thermostat user interface,” *Ergonomics*, vol. 56, no. 3, pp. 463–479, Mar. 2013, doi: 10.1080/00140139.2012.718370.

[84] A. Wagner, W. O’Brien, and B. Dong, Eds., *Exploring Occupant Behavior in Buildings*. Cham: Springer International Publishing, 2018. doi: 10.1007/978-3-319-61464-9.

[85] D. Woodfin and K. Ögelman, “Item 6: Summer 2023 Operational and Market Review”.

[86] T. Ueno and A. Meier, “A method to generate heating and cooling schedules based on data from connected thermostats,” *Energy Build.*, vol. 228, p. 110423, Dec. 2020, doi: 10.1016/j.enbuild.2020.110423.

[87] L. Battistel, A. Vilardi, M. Zampini, and R. Parin, “An investigation on humans’ sensitivity to environmental temperature,” *Sci. Rep.*, vol. 13, no. 1, p. 21353, Dec. 2023, doi: 10.1038/s41598-023-47880-5.

[88] R. S. Sutton and A. G. Barto, *Reinforcement learning: an introduction*. in Adaptive computation and machine learning. Cambridge, Mass: MIT Press, 1998.

[89] A. Manjavacas, A. Campoy-Nieves, J. Jiménez-Raboso, M. Molina-Solana, and J. Gómez-Romero, “An experimental evaluation of Deep Reinforcement Learning algorithms for HVAC control.” arXiv, Apr. 10, 2024. Accessed: Apr. 20, 2024. [Online]. Available: http://arxiv.org/abs/2401.05737

[90] M. Biemann, F. Scheller, X. Liu, and L. Huang, “Experimental evaluation of model-free reinforcement learning algorithms for continuous HVAC control,” *Appl. Energy*, vol. 298, p. 117164, Sep. 2021, doi: 10.1016/j.apenergy.2021.117164.

[91] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, “Stable-baselines3: Reliable reinforcement learning implementations,” *J. Mach. Learn. Res.*, vol. 22, no. 268, pp. 1–8, 2021.

[92] "Weather and Climate Data API for Analysts." Accessed: May 15, 2024. [Online]. Available: https://oikolab.com/

[93] "ERCOT_Fact_Sheet.pdf." Accessed: May 15, 2024. [Online]. Available: https://www.ercot.com/files/docs/2022/02/08/ERCOT_Fact_Sheet.pdf

[94] "Alphabetical Directory of Retail Electric Providers." Accessed: May 15, 2024. [Online]. Available: https://www.puc.texas.gov/industry/electric/directories/rep/alpha_rep.aspx

[95] "Austin Energy - Residential Rates." Accessed: May 15, 2024. [Online]. Available: https://austinenergy.com/rates/residential-rates

[96] "Rhythm: Texas Electricity | 100% Renewable Energy Plans," Rhythm | Easy Energy | Electricity Company. Accessed: May 15, 2024. [Online]. Available: https://www.gotrhythm.com/

[97] H. Yang, L. Wang, Y. Zhang, H.-M. Tai, Y. Ma, and M. Zhou, "Reliability Evaluation of Power System Considering Time of Use Electricity Pricing," *IEEE Trans. Power Syst.*, vol. 34, no. 3, pp. 1991–2002, May 2019, doi: 10.1109/TPWRS.2018.2879953.

[98] "Hourly Load Data Archives." Accessed: May 15, 2024. [Online]. Available: https://www.ercot.com/gridinfo/load/load_hist

[99] "Day-Ahead Market." Accessed: May 15, 2024. [Online]. Available: https://www.ercot.com/mktinfo/dam

[100] S. Dey, T. Marzullo, X. Zhang, and G. Henze, "Reinforcement learning building control approach harnessing imitation learning," *Energy AI*, vol. 14, p. 100255, Oct. 2023, doi: 10.1016/j.egyai.2023.100255.

[101] A. Chatterjee and D. Khovalyg, "Dynamic indoor thermal environment using Reinforcement Learning-based controls: Opportunities and challenges," *Build. Environ.*, vol. 244, p. 110766, Oct. 2023, doi: 10.1016/j.buildenv.2023.110766.

Appendix

Appendix A: ecobee Smart Thermostat Data Dictionary

The limited set of data points used in the analysis from the ecobee timeseries data:

Name	Column Name	Description	Units
Datetime	date_time	Date and time that the reading was taken	NA
Serial Number	Identifier	Unique identifier for the device or reading	NA
HVAC Mode	HvacMode	The current mode of the HVAC system (heating, cooling, off, etc.)	NA
Calendar Event Name	CalendarEvent	Items that override the set schedule (e.g., a temperature hold, demand response event, vacation, smart recovery feature)	NA
Climate Name	Climate	User-defined comfort period (e.g., home, away, sleep, etc.)	NA
Control Temperature	T_ctrl	Average indoor temperature based on relevant sensors as defined by the schedule or mode the user is in	ºF
Cooling Setpoint	TemperatureExpectedCool	Indoor cool setpoint	ºF
Heating Setpoint	TemperatureExpectedHeat	Indoor heat setpoint	ºF

The set of metadata points used in the analysis and clustering:

Field	Range	Description	Used For Clustering?
identifier	-	Unique identifier for the data entry	
account_id	-	Unique identifier for the user account	
runtime	-	Total runtime of the HVAC system	
model	-	Model of the ecobee thermostat	
country	-	Country where the thermostat is located	
province_state	-	Province or state where the thermostat is located	
city	-	City where the thermostat is located	
building_type	-	Type of building (e.g., residential, commercial)	Yes
floor_area_sqft	0 - 10500	Total floor area of the building in square feet	Yes
number_floors	0 - 10	Number of floors in the building	
building_age_yrs	0 - 2020	Age of the building in years	Yes
number_occupants	0 - 20	Number of occupants in the building	
number_cool_stages	0 - 2	Number of cooling stages in the HVAC system	
number_heat_stages	0 - 2	Number of heating stages in the HVAC system	

Field	Range	Description	Used For Clustering?
allow_comp_with_aux	False - True	Indicates if the compressor can run with auxiliary heat	
has_electric	False - True	Indicates if the HVAC system has electric heating	
has_heatpump	False - True	Indicates if the HVAC system has a heat pump	Yes
number_remote_sensors	0 - 10	Number of remote sensors connected to the thermostat	
first_connected	-	Date when the thermostat was first connected to the ecobee service	

Appendix B: Cluster Distribution

After ensuring at least one cluster for each dataset was included in the samples, the table below provides a summary of cluster densities and building types:

Building Type	Cluster Distribution	Number of Buildings in Simulation
DETACHED (63)	Cluster 4: 26.07%	14
	Cluster 2: 23.31%	12
	Cluster 0: 19.80%	11
	Cluster 1: 12.03%	6
	Cluster 3: 11.03%	5
	Cluster 5: 7.77%	3
MULTIUNIT (33)	Cluster 2: 31.43%	15
	Cluster 1: 25.71%	12
	Cluster 5: 20.95%	10
	Cluster 3: 1.12%	1
	Cluster 0: 6.67%	2
	Cluster 4: 2.86%	1
ATTACHED (7)	Cluster 1: 24.72%	1
	Cluster 4: 23.60%	1
	Cluster 0: 21.35%	1
	Cluster 5: 21.35%	1
	Cluster 2: 6.74%	1
	Cluster 6: 1.12%	1
	Cluster 3: 12.38%	5

Appendix C: Constructing the Dynamic TOU Rates

This study aimed to evaluate different TOU-based rates. To achieve this goal, one of the rates took inspiration from a refined partitioning algorithm, adapted from the approach outlined initially by Yang and utilized in further studies by Schittekatte [58] [11]. The algorithm optimizes the day into different periods based on historical price and load data. Each period will last a minimum of 3 hours to avoid rates that are impractical to adopt. Specifically, the algorithm was tailored to utilize the ERCOT 2023 Day Ahead prices.

Partitioning Optimization Overview

1. **Input and Initialization**: The first step in the algorithm involves normalizing the prices for each season and day type. The normalization process adjusts the prices relative to the maximum price observed each day.

2. **Step Initialization**: The step length for moving through the price data is set, and the variables (k_1, k_2, k_3) used to define the boundaries of the time blocks are initialized where k_n is the number of desired periods.

3. **Calculation of Moving Variables**: Provisional boundaries for time blocks are established by calculating moving variables as a function of the minimum price (P_{min}) adjusted by the step variables (k_n) for each potential partitioning.

4. **Formation of Time Blocks**: The day is tentatively divided into three periods based on the calculated moving variables.

5. **Filtering Time Block Sets**: The sets are evaluated to ensure each set meets the minimum duration of 3 hours. If not, parameters are adjusted, and the process repeats.

6. **Objective Function Calculation**: After identifying a valid period set, the next step is calculating the objective function, essentially the Root Mean Square Deviation (RMSD) between the normalized price profile and a proposed partitioning. The objective function measures how well the time blocks fit the price data.

7. **Iterative Optimization**: The step variables are adjusted in a loop to explore various configurations in a controlled manner. This iterative process continues until the algorithm arrives at a solution where the period boundaries optimize the objective function or until all configurations within the predetermined limits are tested.

A simplified overview of the partitioning algorithm is below:

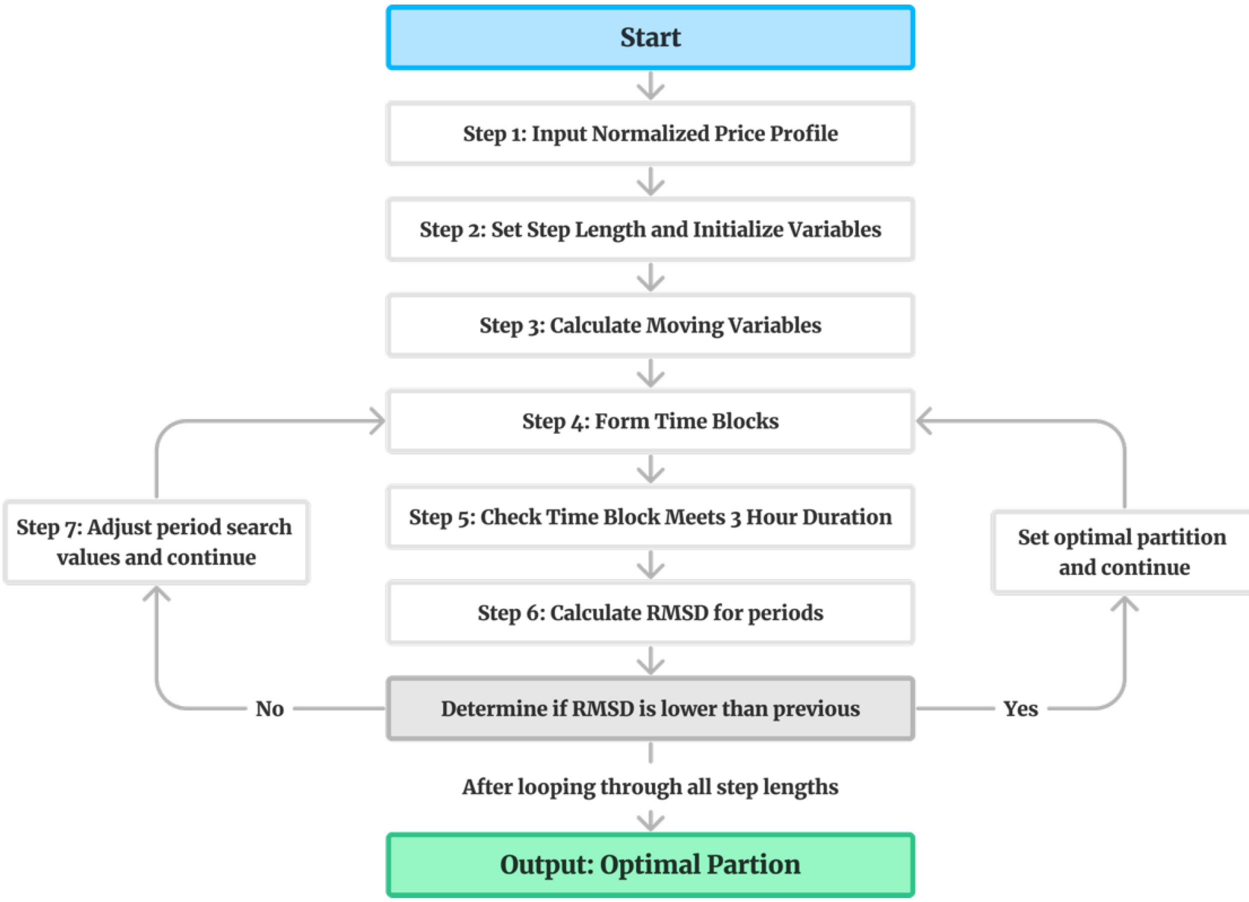


Figure C.1: An overview of the portioning algorithm for creating the TOU rates