POLITECNICO DI TORINO
Repository ISTITUZIONALE

Effects of occupant thermostat preferences and override behavior on residential demand response in CityLearn

Original
Effects of occupant thermostat preferences and override behavior on residential demand response in CityLearn / Kaspar, K., Nweye, K., Buscemi, G., Capozzoli, A., Nagy, Z., Pinto, G., Eicker, U., Ouf, M.M.. - In: ENERGY AND BUILDINGS. - ISSN 0378-7788. - ELETTRONICO. - 324:(2024). [10.1016/j.enbuild.2024.114830]

Availability:
This version is available at: 11583/2995378 since: 2024-12-16T16:31:42Z

Publisher:
Elsevier Ltd

Published
DOI:10.1016/j.enbuild.2024.114830

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

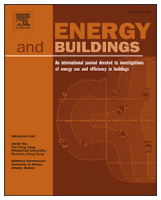
17 September 2026

Energy & Buildings 324 (2024) 114830

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enbuild



Effects of occupant thermostat preferences and override behavior on residential demand response in CityLearn

Kathryn Kaspar *, Kingsley Nweye, Giacomo Buscemi, Alfonso Capozzoli, Zoltan Nagy, Giuseppe Pinto, Ursula Eicker, Mohamed M. Ouf

ARTICLE INFO

Dataset link: https://github.com/kkaspar10/Occupant_Thermostat_Int/

ABSTRACT

As space heating accounts for 54% of annual residential electricity consumption in Quebec, demand response programs specifically target load shifting through the automated control of thermostat setpoints during peak hours. On a district scale, varied thermostat preferences and setpoint override behaviors can have an impact on the success of the demand response program. This study examines two unique occupant types (Average, Tolerant) in terms of thermostat setpoint preferences as well as three different occupant Levels-of-Detail (LoDs) and analyzes their effects on the energy flexibility provided during demand response periods. For our baseline scenario, LoD 1, a static setpoint schedule is used and there is no control of the heat pump, while LoD 2 and LoD 3 incorporate thermostat setbacks during demand response events. LoD 2 assumes the occupant is comfortable within 2 °C from the setpoint while LoD 3 allows the occupant to override the DR setbacks. We estimate the flexibility services provided by a ten-house residential community through the automated control of heat pumps during a three-month winter period, and we implement and simulate our study in CityLearn using reinforcement learning-based control for district-level energy management. When comparing LoD 3 to LoD 1, electricity cost was reduced by approximately 12% and net electricity consumption was reduced by approximately 17% during demand response periods. Likewise, we find that LoD 2 could overestimate savings in net electricity consumption, cost, and peak demand by 5% compared to LoD 3. The number of hours where the indoor temperature deviated more than 2 °C from the setpoint occurred for less than 5% of the timesteps on average for the 10 buildings for LoD 3 while still achieving significant net electricity consumption reductions, thus highlighting that we can provide energy flexibility services to the grid while balancing occupant thermal comfort. Finally, the agents learned optimal decision-making that reduced the number of overrides across the training episodes for both Average and Tolerant occupants. We thus present a multi-agent framework as a means for addressing various occupant setpoint preferences and override behaviors for the control of heat pump systems at the neighborhood level.

1. Introduction

1.1. Background

Buildings of the 21st century, increasingly equipped with renewable energy generation, electrical storage systems, smart appliances, meters, and sensors, are not only significant consumers of electricity but now hold the potential of providing grid services through energy flexibility. As the electrical grid becomes inundated by the electrification of the transportation, building, and industrial sectors, morning and evening peaks in energy demand cause strain on the electrical grid and sometimes require back-up power plants (e.g. gas plants in Ontario [1]) to supplement usual production. The building sector, however, can provide energy flexibility through Demand-Side Management (DSM) for enhanced stability and reliability of the grid.

As approximately 94% of the energy production in Quebec comes from hydropower [2] and space heating represents approximately 54% of annual residential electricity consumption [3], energy flexibility can be exploited through load shifting of space heating loads in the residential sector. Furthermore, approximately 77% of residential buildings in the U.S. [4] and approximately 54% in Canada [5] are single-family homes, so there is untapped potential in aggregating the smaller loads of a neighborhood or district to create substantial impact on energy flexibility. Demand Response (DR) programs offer buildings a way to implement DSM, and these programs rely on customers to either a) change their behavior during peak-hours to reduce energy consumption; b) shift

* Corresponding author.
E-mail address: kathryn.kaspar@mail.concordia.ca (K. Kaspar).

https://doi.org/10.1016/j.enbuild.2024.114830
Received 3 June 2024; Received in revised form 6 September 2024; Accepted 21 September 2024
Available online 27 September 2024
0378-7788/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Nomenclature

Abbreviations

AMY	Actual Meteorological Year
COP	Coefficient of Performance
DHW	Domestic Hot Water
DLC	Direct Load Control
DR	Demand Response
DSM	Demand-Side Management
EUI	Energy Use Intensity
EULP	End-Use Load Profiles
FF	Flexibility Factor
HVAC	Heating, Ventilation and Air Conditioning
KPI	Key Performance Indicator
LoD	Level of Detail
LSTM	Long Short-Term Memory
MAPE	Mean Absolute Percentage Error
MDP	Markov Decision Process
MILP	Mixed Integer Linear Programming
MPC	Model Predictive Control
PAR	Peak-to-Average
RBC	Rule-Based Control
RL	Reinforcement Learning
RMSE	Root Mean Squared Error
SAC	Soft Actor-Critic
SP	(Thermostat) Setpoint
WWR	Window-to-Wall Ratio

Symbols

a	agent action
c	electricity cost
e	net electricity consumption
p_{avg}	average daily peak
p_{max}	maximum peak demand
r	agent reward
s	agent environment state
T_{in}	Indoor Air Temperature
T_{SP}	Thermostat Setpoint Temperature

energy-intensive activities to off-peak hours; or c) incorporate and manage on-site renewables that help reduce demand during peak hours [6]. Various DR programs exist to encourage participation in DSM, including price-based strategies that encourage energy-efficient behavior in exchange for financial incentives as well as Direct Load Control (DLC) by the utility [7]. The availability of smart devices such as thermostats, dishwashers, and electric vehicles can also allow for additional levels of control and coordination of building energy consumption. While manual adjustments of the thermostat can be effective, an automated control algorithm seeking to reduce energy consumption while maintaining occupant comfort could lead to increased participation in the DR programs and thus services provided to the grid in exchange for financial remunerations for the customer. The following sections thus present the state-of-the-art regarding thermal comfort for occupants as well as the control of building energy systems for residential buildings.

1.2. Occupant thermal comfort in buildings

When automatically controlling building heating systems, it is paramount to characterize and quantify occupant thermal comfort. P.O. Fanger's Predicted Mean Vote (PMV)-Predicted Percentage Dissatisfied (PPD) model presents a quantitative framework for evaluating thermal comfort based on parameters such as air temperature, humidity, air velocity, and clothing insulation, and remains an industry standard for estimating occupant thermal comfort in buildings [8]. However, Du et al. [9] found that the accuracy of the PMV model was less than 60% and performed even worse in cooler sensations while Gilani et al. [10] similarly noted that Fanger's PMV-PPD model overestimated thermal sensation by 33% for buildings with an HVAC system. Likewise, the ASHRAE Standard 55 acknowledges that the ideal operative temperature ranges (approximately 21-24 °C when wearing full clothing) still only result in thermal comfort for approximately 80% of the occupants, thus 20% of the occupants experience whole body or partial body thermal discomfort within this range [11]. While these standards remain applicable in many settings, they were not originally intended for residential use. There is thus a shift towards occupant-centric approaches as a way to increase the quality of the indoor environment by analyzing and considering a variety of occupant thermal comfort ranges, especially for applications in residential buildings.

Smart thermostat data have been leveraged to inform occupant thermal comfort and are particularly useful for studying the ranges in occupant thermal comfort preferences in the residential sector. Programmable smart home thermostats enable users to establish temperature setpoint schedules for their heating and cooling systems while offering the convenience of manual adjustments when required. When a user temporarily changes the setpoint directly without modifying the program, this action is referred to as a setpoint *override*, usually lasting between 2-4 hours depending on the user's setting. Automated thermostat control represents an avenue that could allow for significant energy savings, though doing so requires an understanding of both thermostat setpoint preferences as well as occupant-thermostat override behaviors. Several studies have been performed that address occupant thermostat preferences or occupant override behavior in North America. Huckuk et al. [12] presented a comprehensive examination of ecobee's smart thermostat data across 10,000 homes in the U.S. and Canada while noting the importance that climate plays in regional setpoint preferences, thus underlying the importance of further analyzing local setpoint preferences and override behaviors. In a Quebec-specific study, Panchabikesan et al. [13] used the ecobee data to create four clusters of heating setpoint profiles, though did not address thermostat overrides made by the occupants. Kane and Sharma [14] studied the amount of time occupants were present at home prior to making a setpoint override and analyzed the changes in energy consumption due to the overrides, while Huchuk et al. [15] found that occupant overrides led to negative energy consequences for just 2% of the timesteps. In a DR application, Sarran et al. [16] studied thermostat overrides during summer cooling DR events and identified several factors that may have led to overrides, such as quickly changing indoor air conditions and the magnitude of pre-cooling setpoint changes prior to the DR event. Using an agent-based approach, Vellei et al. [17] created a novel dynamic thermal comfort model which firstly simulates skin temperature and body core under uniform conditions and then uses these parameters to estimate the thermal sensation and thermal comfort of the occupants. This study modeled an occupant's probability of a thermostat override given various occupancy rates and at-home activities but considers only occupants with a default setpoint temperature of 21°C. Many studies, such as Hazyuk et al. [18] and Razmara et al. [19], assume thermal comfort so long as the indoor air temperature of the building remains within a band (e.g. 2°C) of the setpoint temperature. While this assumption is reasonable for many commercial or institutional building applications, it could result in an overestimation of occupant thermal comfort and does not consider a residential application where the occupant could decline or override DR signals. Analyzing the effects of varied occupant preferences and override behaviors at the neighborhood or district scale on energy flexibility remains a key gap in literature.

1.3. Methods for the automated control of building energy systems

Most commercial and residential buildings with a building energy management system function via Rule-Based Control (RBC), whereby the temperature setpoint of the building is dictated by pre-determined schedules. RBC falls short in that it is unable to account for factors such as extreme or sudden weather changes, varied occupant thermal preferences, changing occupancy patterns, and changes to building energy consumption patterns. Model Predictive Control (MPC) was thus developed, which leverages physics-based building models to make forecasts for energy consumption, renewable energy production, and occupancy, among other variables [20]. The goal of the MPC control schemes can vary, though the optimization function often involves minimizing energy cost [21], minimizing energy consumption [22], or maximizing consumption of on-site renewable energy production [23]. For example, Zhang et al. [24] was able to achieve approximately 18% reduction in energy consumption through MPC-based control of office cooling systems in comparison to a standard on/off RBC. While MPC can be powerful, its application in the industry remains limited as barriers to its adoption include expert knowledge and time required to build reliable physics-based models, dependency on advanced and costly softwares and solvers such as MATLAB [25] and CPLEX [26], and computational complexity to run the optimization [27].

To avoid the cumbersome nature of constructing physics-based models and running intense optimization solvers, data-driven approaches have been leveraged to facilitate the coordination of building energy systems, renewable energy production, and on-site energy storage. Reinforcement Learning (RL), a branch of machine learning, has gained recent popularity due to the agent-based system's adaptability to changing conditions and ability to infer relationships and optimal decision-making from data rather than from physics-based models [28]. RL centers around an agent taking an action and receiving a reward (or penalty) for its action depending on the objectives. At each timestep, the agent tries to maximize the reward function and receives feedback on its performance using the updated conditions of the environment, thus learning optimal decision-making through trial and error [29]. A potential drawback of using RL is that it can take additional time for the agent to discover and learn optimal decision making, and it is not guaranteed that an agent will always be capable of making optimal decisions [29]. However, the advantage is that the agents make decisions without performing system-wide optimization at every timestep, thus making the approach less computationally intensive compared to heuristic optimization methods [30]. One study tested an RL approach against a Mixed Integer Linear Programming (MILP)-based optimization scheme and found comparable results for minimizing electricity cost for a building cluster, which is promising given the efficiency of building and training the RL agent compared to MILP [31].

To perform the analysis of building energy management, several popular emulators exist such that algorithms and applications can be tested and optimized in a simulated setting. CityLearn is an open-source, Gymnasium environment that serves as an emulator for simulating the control of building energy systems, including storage and HVAC systems [32]. While other emulators such as Boptest [33] and Gym-Eplus [34] focus on single buildings, CityLearn allows for the control and coordination of multiple buildings to assess the effects on the grid of energy management applications, and has been leveraged in this study to analyze more realistically the potential energy flexibility of a residential neighborhood under DR signals.

1.4. Literature gap and research objectives

To extend the lifespan of our current grid infrastructure while we upgrade the province-wide network, DR via DLC can be implemented to best optimize energy consumption and related costs, renewable energy use, thermal comfort, and user satisfaction. We noted previously the growing research related to DR through thermostat setpoint flexibility, and we identified three key gaps in the literature as (a) a lack of variation in thermostat setpoint preferences and profiles; (b) a comparison between allowing and not allowing occupants to override the thermostat during DR events; (c) neighborhood-level control of building energy systems as opposed to single-building analysis.

The aim of this study is thus to address these key gaps, and as such the primary objectives are to:

1. Determine more realistic flexibility services provided through the automated control of residential heating systems
2. Analyze the ability of the energy management scheme to handle varied thermostat setpoint profiles
3. Compare three different levels-of-detail in which we model and consider the occupant in the DR scheme

To achieve these goals, we expand on the existing CityLearn RL environment by including representative and varied occupant thermostat preferences as well as occupant thermostat override models to explore a realistic application of DR participation.

2. Methodology

2.1. Overview

This study analyzes the flexibility provided by a ten-house residential community through the automated control of heat pumps under DR signals during the heating season in Montreal, Quebec. We implement our study in CityLearn, and the source code is available on GitHub[1] for reproducibility. For our study we include two occupant types, herein referred to as *Average* and *Tolerant* occupants based on different thermostat temperature setpoint preferences. We consider the role of occupants in the control scheme by comparing three different Levels of Detail (LoD) in which we incorporate the occupants into the control scheme as shown in Table 1 and Fig. 1.

LoD 1 represents the simplest LoD in which we use fixed setpoint schedules for each occupant type with no control of the HVAC system, thus representing our baseline scenario. The ideal heating load needed to reach the setpoint at each timestep is fed into the Long Short-Term Memory (LSTM)-based Building Thermal Dynamics Models to determine the resulting indoor air temperature of the building, and this helps us establish the reliability of our LSTM models for the subsequent LoDs. In the case of LoD 2 and LoD 3, the heat pump heating supply power is varied according to the CityLearn RL agent action. Firstly, each house has a representative RL agent in CityLearn which determines the heat pump heating supply power given the current state of the environment. Based on the heating supply power of the heat pump, the new indoor air temperature of each home after one hour is calculated using LSTM-based building thermal dynamics models. For LoD 2, this new indoor air temperature is sent back to the agent to continue with the next timestep. In LoD 3, the occupant-thermostat override models are used to determine whether a setpoint change is made based on the indoor air temperature of each home. The thermostat setpoint is either updated corresponding with the schedule or with the occupant change that was made. In the event that the setpoint is overridden, the new setpoint determined by the occupant initiates a 'Hold' whereby the setpoint is changed for the next four hours before reverting back to the setpoint schedule. As in LoD 2, the RL agent then makes a decision based on this information what the heat pump supply power should be for the next one-hour timestep. Finally, LoD 2 and LoD 3 incorporate thermostat setpoints that reflect a DR scenario to incentivize energy-efficient behavior by the agent as well as note the effects of the occupant overrides on the effectiveness of the DR scenario. By comparing the three LoDs, we are able to show

[1] https://github.com/kkaspar10/Occupant_Thermostat_Int/.

Table 1
Occupant representation across three LoDs for comparison.

Occupant Representation Level of Detail (LoD)

LoD	Occupant Type	Thermostat Setpoint	Thermostat Interaction	Demand Response	Occupancy	Description
1	Average, Tolerant	Fixed schedule	No	No	Fixed schedule	No HVAC control with assumed comfort at default setpoint and ideal heating load to meet setpoint at all timesteps
2	Average, Tolerant	Fixed schedule	No	Yes	Fixed schedule	HVAC control with thermostat setbacks during DR events and assumed thermostat comfort band (± 2°C from SP) for flexibility
3	Average, Tolerant	Schedule with dynamic occupant changes	Yes (via occupant-thermostat override models)	Yes	Fixed schedule	HVAC control with thermostat setbacks during DR events and possible SP override action by the occupant

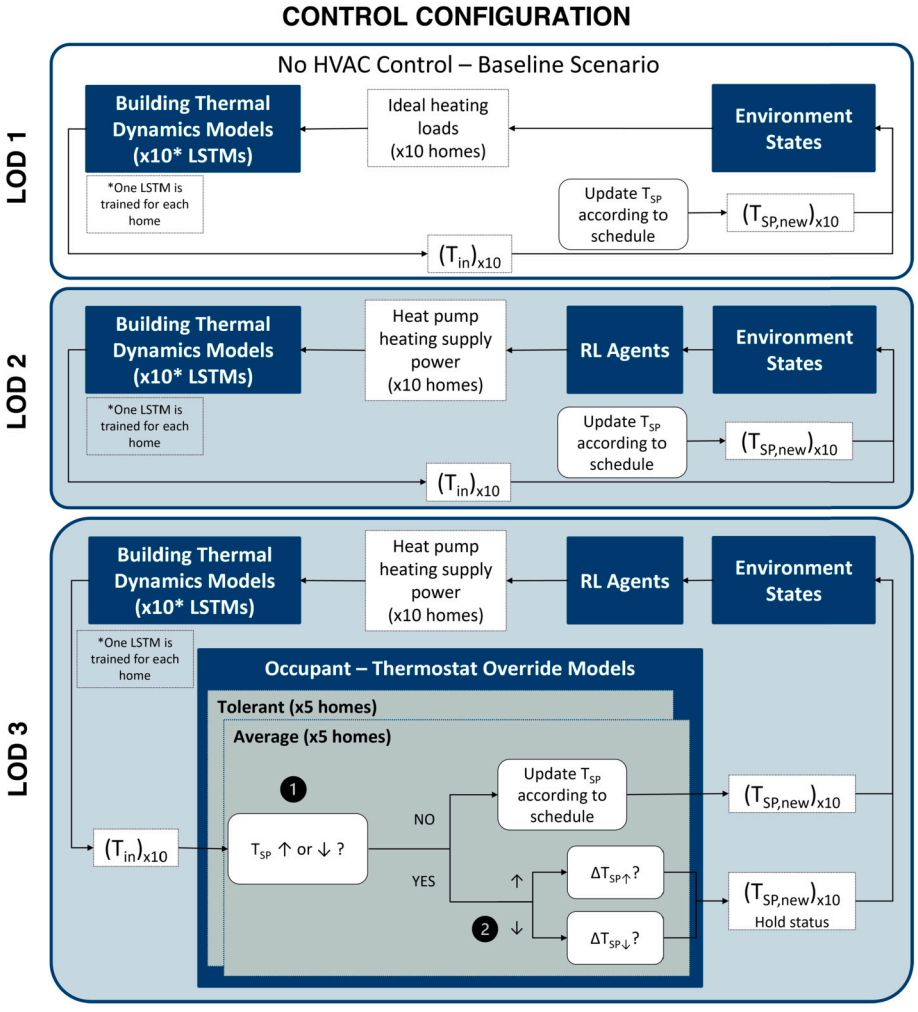


Fig. 1. Control configuration for the three LoDs.

the effects of introducing realistic occupant behavior into the control scheme in LoD 3 as well as provide a methodology for how the automated control of heat pumps could be structured using an independent multi-agent approach.

The coming sections of the methodology address firstly the occupant-thermostat override models as outlined in Section 2.2 followed by the process taken to develop the building models used in this study as described in Section 2.3. Section 2.4 details the control environment we use in this study, and Section 2.5 lists the metrics used to evaluate the control performance.

2.2. Occupant-thermostat override models for DR

Thermostat data from 701 houses in Quebec was queried from ecobee's Donate Your Data dataset, including readings at 5-minute intervals from December, 2016 to December, 2021 [35]. A summary of

Table 2
Data available for each timestep.

Variable	Description
$Identifier$	House ID
$datetime$	Timestamp at 5-minute intervals
$HVACMode$	HVAC system state (e.g. 'Heat')
$CalendarEvent$	Thermostat event (e.g. 'Home', 'Away', 'Hold')
T_{in}	Indoor air temperature
T_{SP}	Setpoint temperature

Table 3
Average indoor air temperature during a setpoint increase and decrease for both clusters.

Cluster	T_{in} during setpoint increase	T_{in} during setpoint decrease
Average	21.0 °C	22.0 °C
Tolerant	19.2 °C	20.4 °C

the variables extracted from this dataset is shown in Table 2, which includes the thermostat setpoint and indoor air temperature readings, the current state of the HVAC system (e.g. 'Heat', 'Off'), as well as specific events or settings (e.g. 'Home', 'Away'). Only readings from November to March inclusively were used in this study as we analyze just the potential of energy flexibility during the heating season in Quebec.

For each home's thermostat, the user specifies the setpoint at various times of day and programs different settings, such as 'Home', 'Work', or 'Vacation', each with a unique schedule applied to specific hours of the day and days of the week. In the ecobee data, the thermostat program for each timestep is listed under a column called 'Calendar Event', which also includes the setting 'Hold'. A 'Hold' event indicates that the program setting was manually changed by the occupant for a certain duration of time, usually lasting between 2-4 hours by default [36]. The timestep at which a 'Hold' was initiated, thus overriding the program via a manual change, is herein referred to as a thermostat *override*, which is discussed further in Section 2.2.2.

2.2.1. Thermostat setpoint schedules

We use k-means clustering based on the average indoor temperature during a setpoint override to consider different thermostat preferences and behaviors in this study. K-means clustering led to two distinct clusters, identified by maximizing the Silhouette score [37]. Table 3 shows the average indoor air temperature when a setpoint override was made (both increases or decreases to the setpoint) for the two clusters, with the average setpoint temperature at each hour of the day presented in Fig. 2 for the two clusters. The cluster with the indoor air temperature preferences of approximately 21-21.5 °C is herein referred to as the *Average* cluster, as these setpoints are typical for the region, whereas the cluster herein referred to as the *Tolerant* cluster showed average setpoints between 19-20 °C, slightly cooler than the Average occupants. We use the hourly setpoints shown in Fig. 2 as the baseline thermostat setpoint schedules for all LoDs.

For LoD 2 and LoD 3, we consider a scenario whereby the baseline setpoint schedule is decreased during certain periods to reflect DR events, as shown in Fig. 2. We model the setpoint changes based on Quebec's Rate Flex D Demand Response program [38], which consists of between 25-33 events from December 1 to March 31 in which participants receive significant decreases in electricity rates during off-peak periods and conversely steep rates during the DR events, as shown in Table 4. Each event in this program lasts either three hours (6:00-9:00 AM) or four hours (4:00-8:00 PM), and we implement 25 events during the three-month period. For the 2020, 2021, and 2022 weather datasets, we identified the 25 periods (either 6:00-9:00 AM, or 4:00-8:00 PM) in which the average outdoor air temperature was the coldest, thus selecting these as our DR events. During the event, we modify the setpoint schedule as shown in Fig. 2 by reducing the setpoint by 2°F (approximately 1.1 °C) for the entirety of the DR event. This reduction falls in line with acceptable occupant setback adjustments during a DR event [39]. Finally, for just LoD 3, the occupants could override the setpoint reductions during all DR events according to the override models described in the following sections.

Table 4
Electricity rate under Hydro-Quebec's Rate Flex D Demand Response Program [38].

Period	Cost
Off-Peak	0.04582 $/kWh
During DR Event	0.53526 $/kWh

Table 5
Hours corresponding to each scenario.

Scenario	Hours
Morning/Evening	6:00-9:00 AM, 5:00 PM-12:00 AM
Night	12:00-6:00 AM
Mid-Day	9:00 AM-5:00 PM

2.2.2. Probability of a setpoint override

As shown previously in Fig. 1, the occupant-thermostat override models for both the Average and Tolerant occupants consist of two separate models: a) the probability of a setpoint override at each timestep, as discussed in this section, and b) the magnitude of the change of the setpoint as described in Section 2.2.3. Occupants override thermostats for various reasons, including thermal discomfort as well as changing the thermostat directly before leaving for vacation or work rather than changing the program settings. Thermostat overrides represent less than 1% of all data, so a deviation of the thermostat setpoint from the default schedule is an unlikely occurrence. As such, we based our methodology off of the work found in [40] where we use discrete-time logistic regression curves to approximate the probability of an override at each timestep. A more detailed explanation of the models described in Section 2.2.2 and Section 2.2.3 can be found in our previous work [41].

To develop the discrete-time logistic regression curves, the probability of a setpoint increase in the next hour is determined for each occupant type as well as each time of day across a range of indoor air temperatures from 14.5 to 23.5 °C. We then model the probability of a setpoint increase during 'Morning/Evening', 'Night', and 'Mid-Day' hours using discrete-time Markov logistic regression for each cluster, with Table 5 outlining the hours corresponding to each scenario. These three scenarios were chosen as the time of day affected the frequency of thermostat overrides. A fourth scenario considered the probability of a setpoint decrease due to occupants' discomfort, which is less likely to occur in winter though possible in the event of inefficient operations. The form of the logistic regression equation is shown in Equation (1), with T_{in} representing the input variable and p representing the probability of a setpoint increase or decrease (override = 1, in either case) for the given model. Coefficients a and b were determined in the training, and the final plots are shown in Fig. 3. All p-values for a and b coefficients were less than 0.05, thus indicating a good fit, and a summary of the coefficients and p-values can be found in the Appendix Table 13.

$$p(override = 1) = \frac{1}{1 + e^{-(a+bT_{in})}} \tag{1}$$

2.2.3. Magnitude of the setpoint change

Fig. 4 shows the distribution of the magnitude of setpoint changes at the time of a thermostat override, with an average of approximately ±0.5 °C. We thus trained random forest models to predict whether the magnitude of the setpoint change would either be small (less than 0.5 °C) or large (greater than 0.5 °C) during a thermostat override. A Pearson Correlation revealed T_{SP}, $(T_{SP})_{t-1}$, and $(T_{in} - T_{SP})_{t-1}$ were the variables most correlated to the magnitude of a setpoint override, and thus these variables were used as inputs to the random forest models.

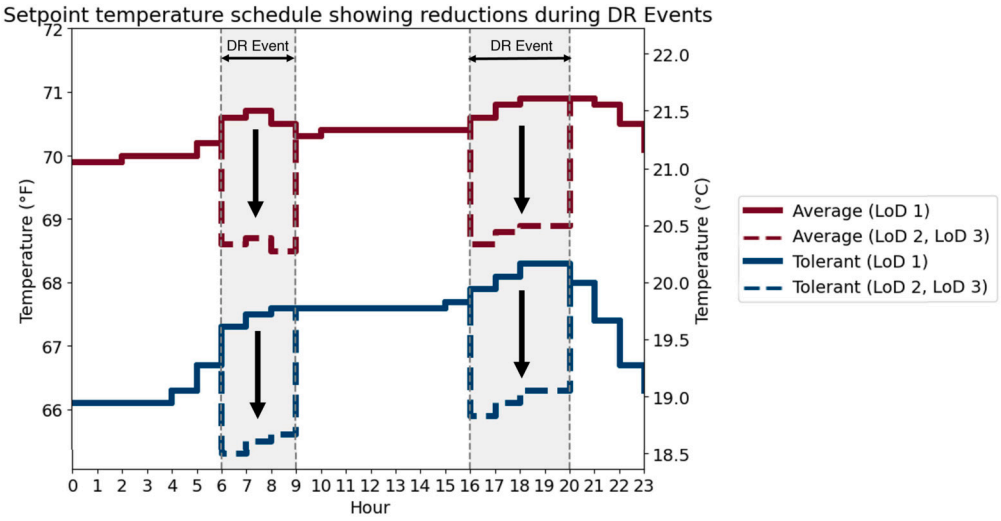


Fig. 2. Thermostat setpoint temperatures for Average and Tolerant occupants according to the normal schedule and during DR events. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

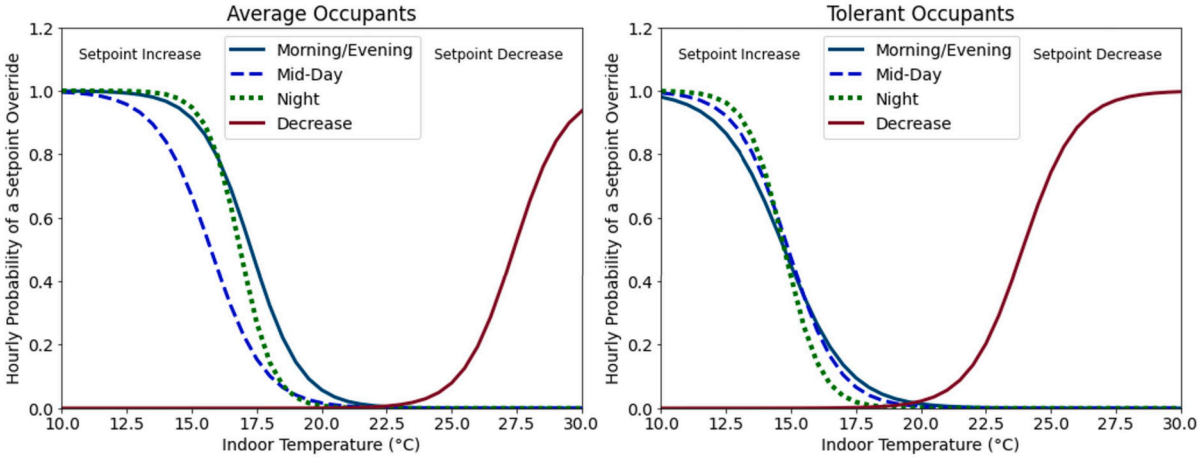


Fig. 3. Probability of a setpoint override as a function of the current indoor air temperature for Average and Tolerant occupants.

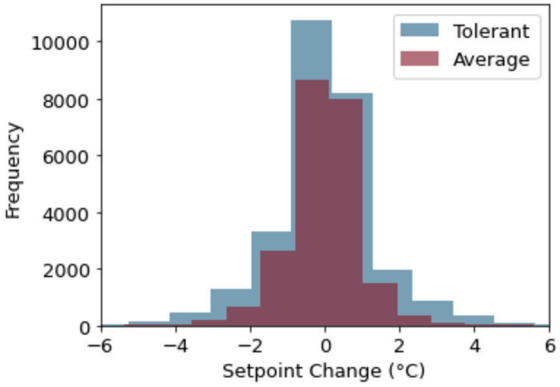


Fig. 4. Distribution of the magnitude of the setpoint change during an override.

The average accuracy was 80%, the average recall was 85%, and the average precision was 77%, and the classification accuracy was balanced between the two classes. We present the full confusion matrix of the model testing in Appendix Fig. 15 alongside with the final model parameters in Appendix Table 14.

2.3. Case study buildings

In addition to the occupant-thermostat override models, we employ EnergyPlus models as described in detail in Section 2.3.1 as well as thermal dynamics models discussed in Section 2.3.2. The data used to produce or train the models is shown in Fig. 5. For our study, a ten-house neighborhood was selected due to this being the average number of houses typically on one transformer in Quebec [42]. The province of Quebec is situated within ASHRAE Climate Zones 6A and 7A, typically characterized by cold, long winters. The simulation duration thus spans three consecutive months during the heating season, running from January 1st to March 31st, for a total of three consecutive years. The data from 2020 and 2021 were used to develop and train the RL agent models which was then subsequently tested on the data from 2022, thus resembling a deployment scenario where prior data informs training and calibration prior to deployment. NASA's Langley Research Center (LARC) POWER Hourly API [43] was used to obtain both the *.epw* files for EnergyPlus as well as the weather states for the CityLearn agent environment. The decision-making for the control of the HVAC systems takes place at hourly timesteps, and for this study we exclude additional energy storage systems and renewable energy generation to isolate the interplay between the control agent actions and occupant behavior on comfort and performance.

2.3.1. Home EnergyPlus models

We utilize the *resstock-amy2018-2021-release-1* version of the End-Use Load Profiles (EULP) for the U.S. Building Stock [44] dataset to a) provide input data for building energy models in our control environment, CityLearn, using the methodology outlined in [45]; b) generate

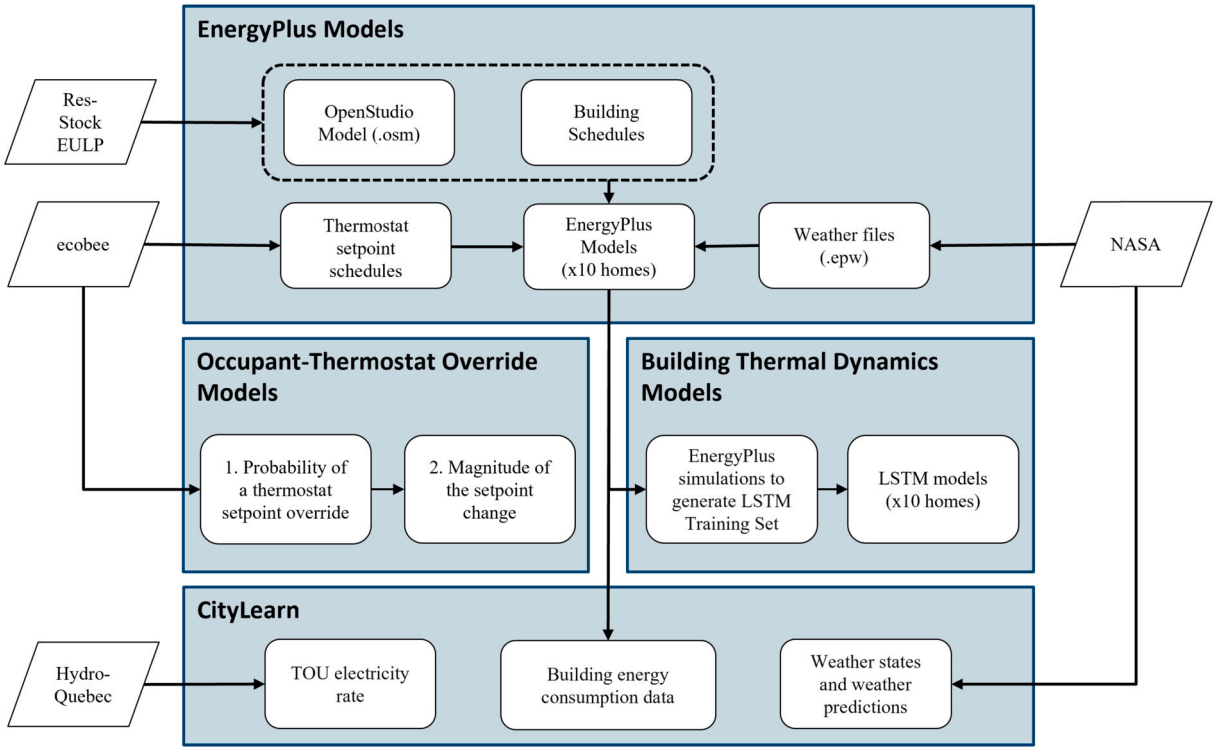


Fig. 5. Databases used to generate and train models.

training data for the building thermal dynamics models based on the work by Pinto et al. [46]. This dataset is the 2021 release of approximately 500,000 OpenStudio energy models and their simulation results using the 2018 Actual Meteorological Year (AMY) weather data, where the energy models are assumed to be representative of the U.S. residential building stock in 2018.

While the EULP dataset is based strictly on U.S. building stock, we identify the Chittenden County, Vermont as a close match to Quebec in terms of climate and construction types and filter our building selection pool in the EULP dataset to this location. We use a cluster-frequency-based sampling of the 117 occupied single-family detached buildings available in the dataset to select a 10-building subset following the methodology in [45]. The cluster model is based on metadata similarities across buildings where the metadata fields include orientation, decade of construction, number of occupants, infiltration rate, ceiling, slab and wall insulation, Energy Use Intensity (EUI), and Window-to-Wall (WWR) ratio, and the ten buildings and relevant characteristics are summarized in Table 6. Because each building in the EULP dataset is representative of a number of homes specified in the metadata, we calculate that our clustering leads to these 10 buildings representing approximately 25% of residential buildings in Chittenden County.

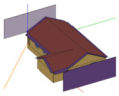
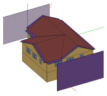
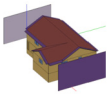
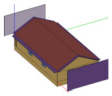
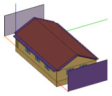
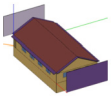
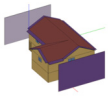
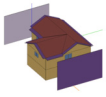
We convert the OpenStudio models of the ten selected buildings described in Table 6 into EnergyPlus models to generate the training data for the LSTM-based building thermal dynamics model for each building as well as other input data for CityLearn. To curate the training data for the building thermal dynamics models described further in Section 2.3.2, we first modify the thermostat setpoint schedule in each as-provided EULP building energy model such that half the buildings are assigned either Tolerant or Average occupant type and their associated setpoint schedule, and each building cluster in Table 6 has a mix of both occupant types. For 2020, 2021, and 2022, we use the EnergyPlus Weather (*.epw*) data from the beginning of January to the end of March and run eight simulations to obtain building ideal heating loads as well as the effect of heating below and above the ideal loads on the indoor dry-bulb temperature. A description of the eight simulations is presented in the Appendix, with the goal being to vary the heat pump heating supply load (under-heating and overheating) such that the LSTM is able to be trained on a wide range of varying conditions and heat pump actions. We visually summarize the results from these eight simulations for the 2020 data in Appendix Fig. 16 to assert the validity of the as-provided EULP energy models as well as the quality of the LSTM training data. Finally, we query the results of all 2020 and 2021 generic internal heat gain/loss equipment simulations for all buildings to extract features that make up the building thermal dynamics model training data, including the following variables: direct and diffuse solar irradiance, outdoor dry-bulb temperature, occupant count, heating energy, month, day of the week, hour, and indoor dry-bulb temperature (target variable).

2.3.2. *Building thermal dynamics models*

Co-simulating in conjunction with a software like EnergyPlus can require extensive computational and technical resources. As a result, we develop models that capture the indoor air temperature dynamics of the buildings as heating supply power changes at each timestep. Long Short-Term Memory networks [47], LSTMs, are a powerful type of neural network widely used for processing sequential data and time series analysis. LSTMs introduce a new type of neuron called the LSTM cell. This cell has several gates that control the flow of information, which allows the LSTM to selectively forget or remember information from previous timesteps.

The sliding window method used in LSTM training partitions historical data into sequences with a lookback of 12 hours. Each sequence represents a window of observations, such as indoor temperature measurements over consecutive hours. For training, each sequence is paired with its corresponding target observation, forming input-output pairs. This approach captures temporal dependencies crucial for LSTM models to learn patterns and make accurate predictions. By sliding the window along the time axis, multiple input-output pairs are generated, covering the entirety of the historical data. Through this iterative process, the LSTM model learns to predict the indoor temperature value at a given timestep (t), leveraging the lagged input sequence leading up to the previously *t-12* timesteps. This enables the model to effectively capture temporal dynamics and furnish reliable forecasts based on past observations.

Table 6
Building clusters and characteristics of selected buildings.

Cluster ID	Bldg. ID	Geometry	Orient.	Built	Num. Occ.	Infiltration (ACH50)	Insulation (R-Value)			EUI ($\frac{kWh}{m^2}$)	WWR	Occ. Type (Assigned)
							Ceiling	Slab	Wall			
1	1		W	2000s	2	6	30	-	19	122	0.11	Tolerant
	2		NE	1990s	1	10	30	-	15	289	0.05	Average
	3		SW	2010s	3	8	-	-	19	235	0.13	Tolerant
	4		SW	1980s	2	15	30	-	19	359	0.07	Average
2	5		S	1980s	2	15	30	-	11	329	0.08	Average
	6		S	1940s	1	8	13	-	-	310	0.11	Tolerant
3	7		W	1940s	1	25	7	-	-	478	0.13	Tolerant
	8		NW	1970s	2	20	19	-	7	349	0.15	Average
4	9		S	1990s	3	10	30	10	19	221	0.05	Average
	10		E	2000s	1	8	30	10	11	291	0.04	Tolerant

One LSTM was developed for each building and the hyperparameter setting for train the networks are summarized in Appendix Table 15. The performance of the neural network was evaluated using the Mean Absolute Percent Error (MAPE) and Root Mean Squared Error (RMSE) metrics displayed in Table 16 and Table 17 respectively in the Appendix. We found an average MAPE of 1.39% and an average RMSE of 0.75 °C for the 10 buildings in Closed Loop testing, which is in line with findings in [48] and [49]. Fig. 6 shows an example of the indoor air temperature predictions of the trained LSTM model for Building 3 during a 7-day period compared to the indoor air temperature when simulated with EnergyPlus, thus showing our model is able to reasonably capture the building's thermodynamic evolution as a result of the hourly heat pump action.

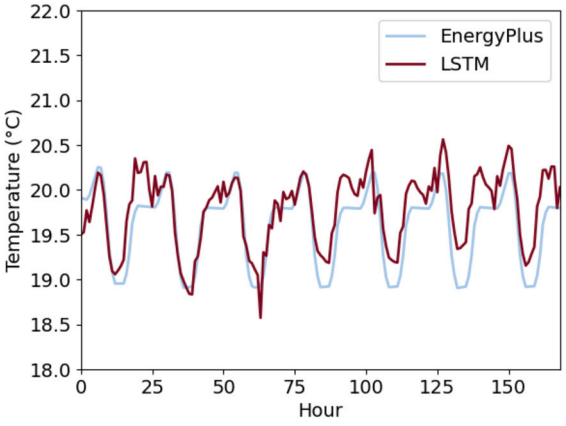


Fig. 6. LSTM predicted temperature compared with EnergyPlus simulation.

2.4. Control environment

Our CityLearn environment is made up of the ten buildings described in Table 6 where each building is made up of controllable and non-controllable loads. Each building is equipped with distributed energy resources that consume electricity from the grid including a heat pump to meet space heating loads and an electric resistance heater to meet Domestic Hot Water (DHW) heating loads. We summarize these resources in Table 7 where the heat pump is sized to meet the maximum hourly space heating load with a 1.3 factor of safety and an ideal cycle Coefficient of Performance (COP) $\in [2.23, 3.18]$ driven by outdoor dry-bulb temperature and a target temperature between 45 °C-48 °C. We have sized the heater to meet the maximum hourly DHW heating load with a 1.1 factor of safety and efficiency $\in [0.85, 0.95]$. These sizing consid-

Table 7
Building distributed energy resource specification in CityLearn environment.

Bldg. ID	1	2	3	4	5	6	7	8	9	10
Heat pump (kW)	23.23	18.67	40.41	46.77	36.57	30.38	44.90	37.44	28.94	17.53
Heat pump max. COP	2.42	2.52	2.87	3.18	2.23	2.61	2.63	2.61	2.75	2.47
Heater (kW)	21.34	6.46	9.73	19.32	17.52	15.97	17.80	35.13	6.49	19.70
Heater, η	0.88	0.92	0.87	0.92	0.87	0.87	0.91	0.90	0.86	0.93

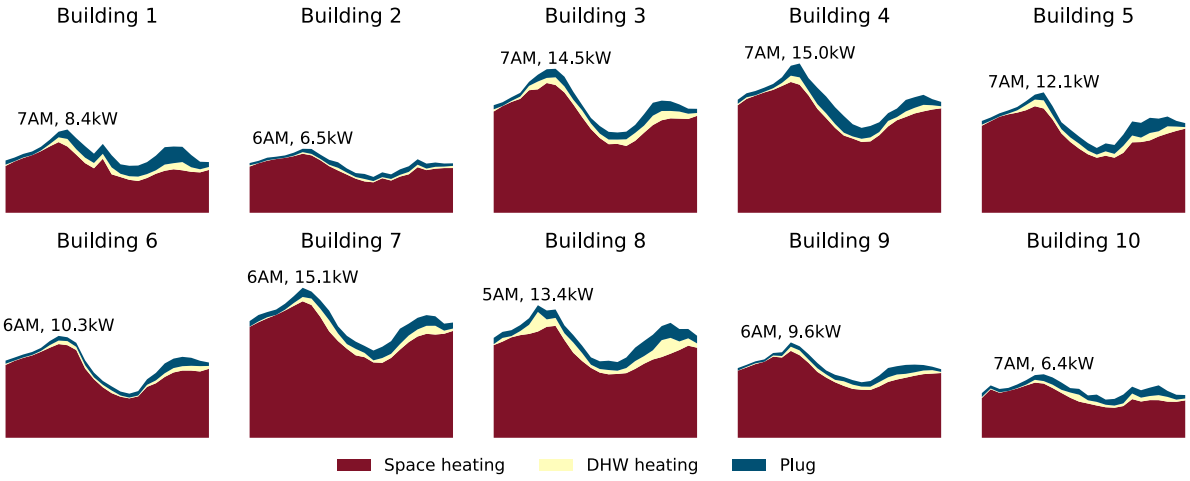


Fig. 7. Average daily electricity consumption profile disaggregated by load type with peak load and hour annotated.

erations ensure that the capacities of the resources are adequate for observed space and DHW heating loads.

The space heating load is controllable through managing the heat pump power which affects the building indoor dry-bulb temperature. In contrast, the DHW heating loads are non-controllable and the as-simulated loads from EnergyPlus must be met at all timesteps. Other non-controllable loads are plug loads including lighting and any electric household appliances that are met directly by the grid.

Fig. 7 shows the variance in the average daily electricity consumption profiles across all ten buildings for the sized resources when all loads are uncontrolled i.e., ideal loads are maintained by the resources. The space heating load makes up 86.6±3.8% of the average daily load compared to only 4.6±1.4% and 8.8±2.7% represented by DHW heating and plug loads.

In LoD 1 there is no controlled resource and the results from this simulation serve as the baseline. For LoD 2 and LoD 3, we use the Soft Actor-Critic (SAC) RL algorithm [50] to control the heat pump power in order to deliver adequate heating energy to each building to maintain indoor temperature in the comfort range for the occupant while providing flexibility measured by the criteria discussed in the next section.

For each RL agent, we have the action space, observation space, and the reward function. The control action, a_t, defines the proportion ($\in [0, 1]$) of the heat pump nominal power available at any given timestep as chosen by the agent. We adopt a decentralized approach where each building's heat pump is controlled independently as this allows us to tailor the control policy to unique comfort preferences. The observation space shown in Table 8 presents the environment states available to each agent. To generate a 6-hour prediction of the outdoor dry-bulb temperature, the methodology in [51] was used whereby the actual outdoor dry-bulb temperature for 6 hours ahead of each timestep was obtained and noise of ±2.5% of the reading was added. The observations are transformed to aid learning by applying cyclical transformation, one-hot encoding and min-max normalization to periodic, discrete, and continuous observations respectively.

Equation (2) shows the reward function calculated at each timestep as a function of the indoor air temperature and the setpoint temperature, with Table 9 providing the exponents for each building that are used in Equation (2). Subsequently, Fig. 8 represents a visualization of

Table 8
Agent observation space used for LoD 2 and LoD 3.

Observation	Unit	Transformation
Temporal		
Day	-	One-hot
Hour	-	Cyclical
Weather		
Outdoor Dry-Bulb Temperature	°C	Min-max norm.
Outdoor Dry-Bulb Temperature (+6 hr)	°C	Min-max norm.
Building		
Net Electricity consumption	kWh	Min-max norm.
Indoor Air Temperature	°C	Min-max norm.
Thermostat Setpoint Temperature	°C	Min-max norm.
Indoor Air/Setpoint δ	°C	Min-max norm.
Electricity Rate	$/kWh	Min-max norm.
Electricity Rate (+1 hr.)	$/kWh	Min-max norm.
Electricity Rate (+2 hr.)	$/kWh	Min-max norm.
Electricity Rate (+3 hr.)	$/kWh	Min-max norm.
LoD 3 Only		
Occupant setpoint override δ	°C	Min-max norm.

Table 9
Reward function exponents used in Equation (2) where a and b values minimize discomfort for each building.

Bldg. ID	1	2	3	4	5	6	7	8	9	10
a	1.0	1.8	2.0	1.0	1.0	2.0	1.2	1.6	1.0	1.0
b	1.0	1.8	2.0	1.0	1.0	2.0	1.2	1.6	1.0	1.0

the reward function to better illustrate which agent actions lead to increased rewards for the agent. Fig. 8(a) shows the case in which T_{in} is more than 2 °C less than the setpoint temperature, which results in a penalty equal to the deviation between T_{in} and T_{SP} raised to the exponent a. In this case, the indoor air temperature is too far below the setpoint and thus could negatively affect the occupant comfort, hence the penalty enforced. Fig. 8(b) shows the case where T_{in} at or less than 2 °C from the setpoint temperature, which is the ideal scenario during the heating season, thus the reward is equal to 0 (the maximum possible value in our reward structure). The third case in Fig. 8(c) represents

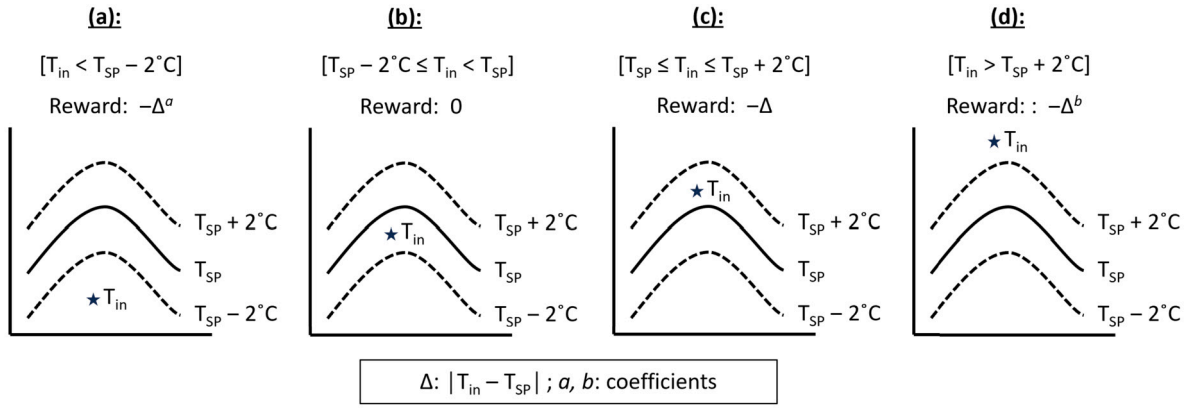


Fig. 8. Calculation of the agent reward as a function of the indoor air (T_{in}) and setpoint (T_{SP}) temperatures as follows: (a) $T_{in} < (T_{SP} - 2\,°C)$; (b) $(T_{SP} - 2\,°C) \leq T_{in} < T_{SP}$; (c) $T_{SP} \leq T_{in} \leq (T_{SP} + 2\,°C)$; (d) $T_{in} > (T_{SP} + 2\,°C)$.

Table 10
KPIs used to evaluate the performance of the RL controller.

KPI	Description	Equation
Net electricity consumption, (e)	Total electricity consumption of the cluster	$e = \sum_{t=0}^{T} e_t$
Electricity cost, (c)	Electricity cost of the cluster during the test period	$c = \sum_{t=0}^{T} \text{price}_t \times e_t$
Average daily peak, (p_{avg})	Average daily peak electricity demand of the cluster	$p_{avg} = \frac{\sum \text{Daily peak demand}}{90 \text{ days}}$
Maximum peak demand, (p_{max})	Maximum electricity demand of the cluster during the testing period	$p_{max} = \max(e_t)$
Ramping	Change in electricity demand from one timestep to the next	$\text{ramping} = \sum_{t=0}^{T} e_t - e_{t-1}$
1 - load factor	Improvement to the ratio of the average load to the peak load	$1 - \text{load factor} = \frac{p_{avg}}{p_{max}}$
Flexibility Factor (FF) [52]	Ability to shift load on days with DR events	$FF = \frac{\sum_{t=0}^{T} e_t^{Non-DR} - \sum_{t=0}^{T} e_t^{DR}}{\sum_{t=0}^{T} e_t^{Non-DR} + \sum_{t=0}^{T} e_t^{DR}}$
Occupant discomfort	Number of unmet setpoint hours **(LoD 3 Only)** Number of thermostat overrides	

when the setpoint is between 0 and 2 °C greater than the setpoint, which is penalized proportional to the deviation from the setpoint, as this results in inefficient operations in the heating season. Likewise, the last case shown in Fig. 8(d) represents the worst-case scenario where the indoor air temperature is more than 2 °C greater than the setpoint, and is thus penalized more severely.

Finally, we make use of the 2020 and 2021 weather, thermostat setpoints, pricing, and ideal heating load data to train the control agent by sequentially alternating between the two years for ten episodes. The 2022 version of these data are then used to evaluate (test) the trained agent.

$$r = \begin{cases} -|T_{in} - T_{SP}|^{a}, & \text{if } T_{in} < (T_{SP} - 2\,°C) \\ 0, & \text{if } (T_{SP} - 2\,°C) \leq T_{in} < T_{SP} \\ -|T_{in} - T_{SP}|, & \text{if } T_{SP} \leq T_{in} \leq (T_{SP} + 2\,°C) \\ -|T_{in} - T_{SP}|^{b}, & \text{if } T_{in} > (T_{SP} + 2\,°C) \end{cases} \quad (2)$$

2.5. Evaluation metrics

To evaluate the efficacy of the automated control scheme, we evaluate and compare the simulations performed in the three LoDs using the Key Performance Indicators (KPI) outlined in Table 10. We present the results from the three LoDs across the entire simulation duration to note the effects on electricity cost, net electricity consumption, average daily peak, maximum demand, ramping, and load factor. We subsequently analyze the net electricity consumption and flexibility factor specifically during the DR events for LoD 1 as a comparison to the DR applications presented in LoD 2 and LoD 3. Finally, we quantify the levels of occupant discomfort for the three LoDs via the number of unmet setpoint hours (LoD 2 and LoD 3) and the number of thermostat overrides (LoD 3).

3. Results

3.1. Visual inspection of results

To ensure that our inter-operating models were behaving as expected, we firstly visualized the setpoint temperature, indoor air temperature, and heating demand together as shown in Fig. 9. The figure shows a sample four days of the testing period with six DR events, where Building 1 has Tolerant occupants as shown by the temperature setpoints of approximately 19-20 °C and Building 5 has Average occupants as shown by the temperature setpoints of approximately 21-22 °C. In Fig. 9(a) and (b), we see the default setpoint schedule used in LoD 1 in black while the setbacks during the six DR events can be seen in LoD 2 in blue. The dashed red lines of LoD 3 show the periods where the occupant overrides the setpoint due to discomfort, and otherwise follows the same schedule as LoD 2. Fig. 9(c) and (d) show the indoor air temperature during the same period. Our reward function is structured to keep the indoor air temperature at or within 2 °C below the setpoint temperature, so the indoor air temperatures we see for LoD 2 and LoD 3 are as expected. We see that at approximately timesteps 425, 435, and 485 for Building 1, the indoor air temperature for LoD 3 falls well below comfort levels, which triggers an increase to the setpoint that we see in 9(a), thus our override models are working as expected. We also see that during the same periods where the indoor air temperature decreases (e.g. timesteps 425, 435, and 485 for Building 1), the corresponding heating demand shown in from Fig. 9(e) is also low, whereas the high indoor air temperature seen for Building 1 at around timestep 495 corresponds with an increase in heating demand at the same time. Likewise, we see in Building 5 that the indoor air temperatures of LoD 2 and LoD 3 remain consistently lower than LoD 1, as reflected in the heating demand as well, thus confirming that our models are behaving as expected.

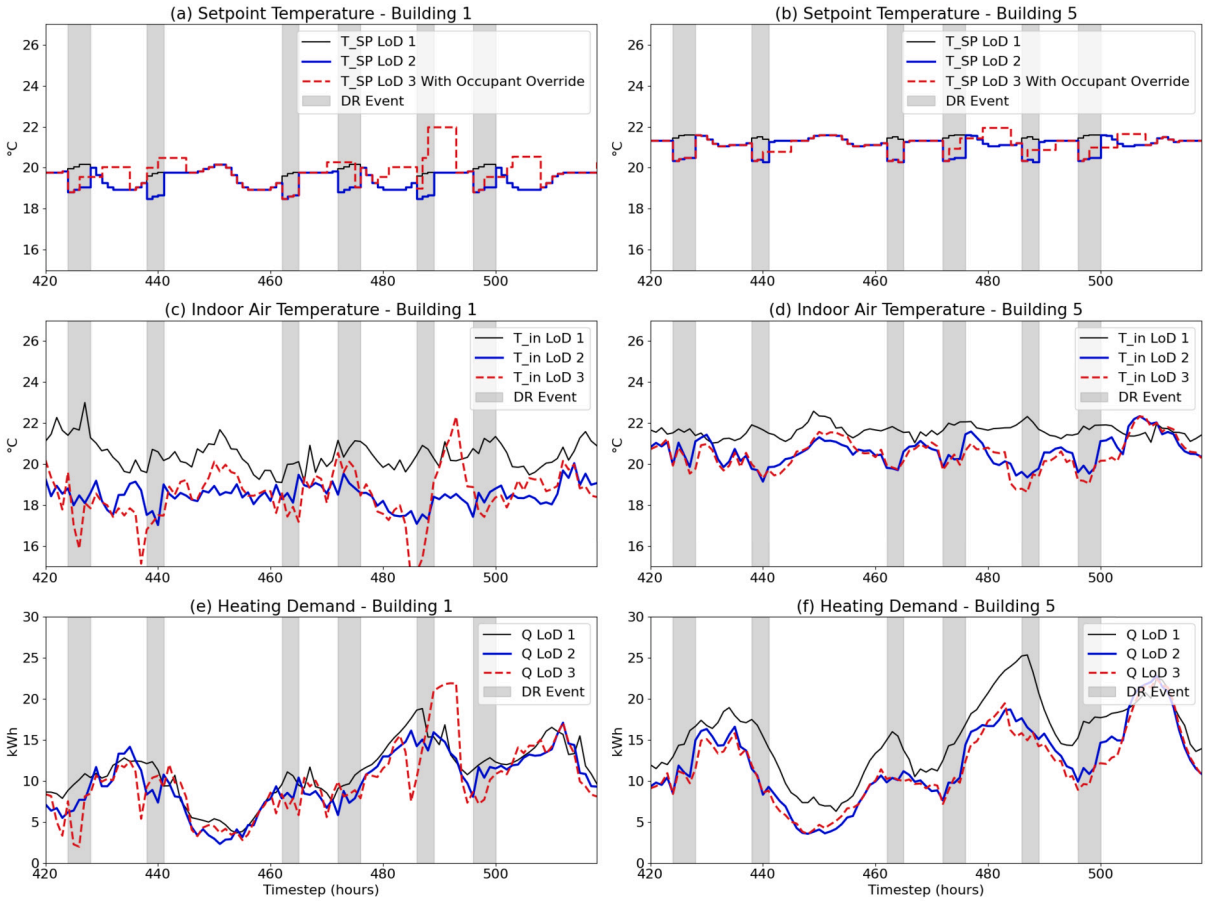


Fig. 9. Indoor air temperature, setpoint temperatures, and corresponding heating demand for fours days of the testing period for Building 1 with Tolerant occupants and Building 5 with Average occupants.

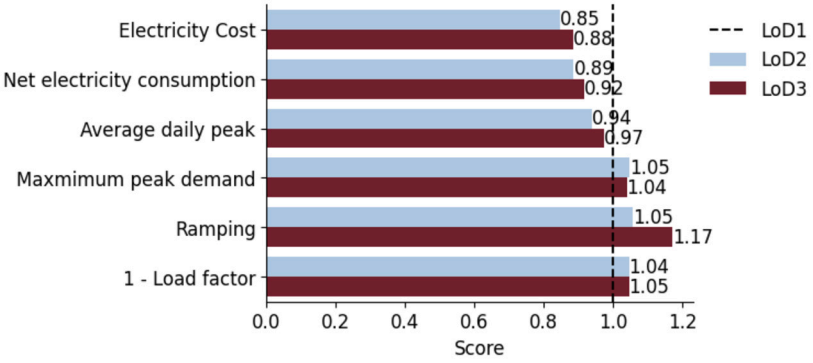


Fig. 10. KPI results for LoD 2 and LoD 3 compared with the baseline LoD 1.

3.2. Results summary

We subsequently analyzed the cumulative results of the 10-house neighborhood for the three-month testing period, which are presented in Fig. 10. LoD 1 serves as the baseline, and we compare the results of LoD 2 and LoD 3. We firstly see that with the agent being encouraged to keep the indoor air temperature slightly below or at the setpoint temperature, as well as the setpoint adjustments made during the high-price DR periods, the cost of electricity was reduced by 15% and 10% respectively for LoD 2 and LoD 3 in comparison to LoD 1 for the three-month duration. We also see that the total electricity consumption was reduced by 11% and 8% for LoD 2 and LoD 3 respectively. There were just slight reductions in the average daily peak by approximately 6% and 3% for LoD 2 and LoD 3 respectively; however, the maximum peak electricity demand during the three-month period was increased in both LoD 2 and LoD 3 with respect to the baseline. Likewise, ramping and 1 - load factor were increased in both LoDs, thus the electricity demand curve became more variable and less flat under the two control scenarios compared to the baseline.

3.3. DR events analysis

In Fig. 11, we look more closely at the electricity consumption specifically during the DR events to note the effects of the reductions to the setpoint temperature during these periods. For LoD 2, each building had an average reduction of 22% in net electricity consumption during the DR periods compared to 11% when looking at the total testing period, while for LoD 3 each building had on average an approximately 17% re-

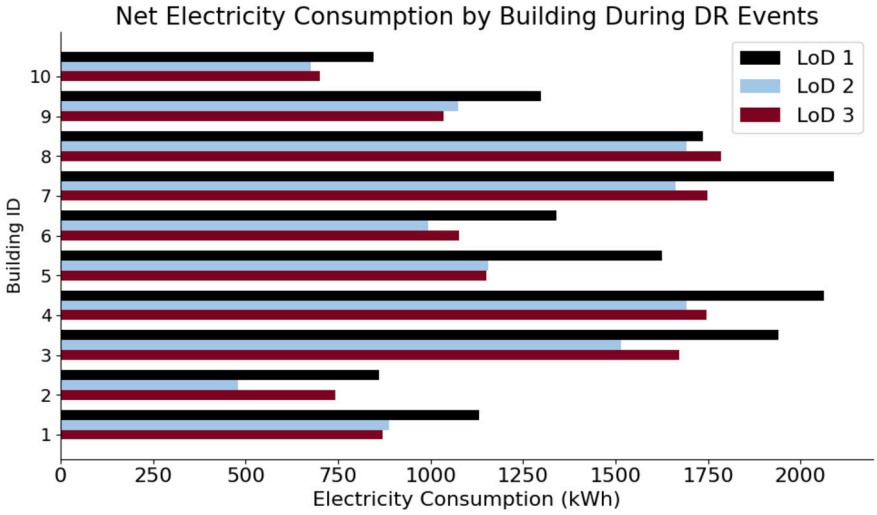


Fig. 11. Total electricity consumption during DR Hours for LoD 2 and LoD 3 with respect to LoD 1.

Table 11
Flexibility Factor results for LoD 1, LoD 2, and LoD 3 during the days with DR events.

	LoD 1	LoD 2	LoD 3
FF	0.26	0.33	0.32

duction in electricity consumption during the DR periods compared to an 8% reduction when looking at the entire testing period. As such, our targeted setpoint reductions during the DR periods were very effective at shifting electricity consumption during these peak periods. Likewise, all buildings saw reductions in the electricity consumption during the DR periods except for Building 8 in LoD 3, which saw an increase in net electricity consumption of approximately 2%. We see more significant reductions in net electricity consumption for LoD 2 when compared to LoD 3, which is expected as the occupant overrides are not considered for LoD 2. More importantly, we saw that the occupant overrides caused a difference ranging from less than 1% up to 34% depending on the building considered in the amount of net electricity consumption used, thus showing the range of impact that including the occupant overrides in LoD 3 had when comparing directly to LoD 2.

We also include in Fig. 12 the average hourly net electricity consumption load curves of the district for days with DR events as well as days without any DR events. Of note, the days with DR events were selected as the days with the lowest temperature forecasts, and as such had higher overall electricity consumption compared to the non-DR days. From the figure, we see that on days without the DR events (represented by the dashed lines) the shape of the curve for LoD 2 and LoD 3 aligns similarly to our curve for LoD 1; however, the net electricity consumption drops significantly for LoD 2 and LoD 3 during the DR hours (6-9 AM, 4-8 PM) on days that include the DR events, where we have imposed thermostat setbacks. As such, the models were able to effectively reduce the net electricity consumption during the targeted DR events.

Lastly, we look at the Flexibility Factor which we calculated using the equation shown in Table 10, and the results are shown in Table 11. We calculated the flexibility factor considering only the days in which a DR event occurred. A higher flexibility factor means more electricity is shifted to off-peak periods, where 1 is the maximum score and -1 is the minimum score. Once again, we see that the DR setpoints and control schemes of LoD 2 and LoD 3 achieve an increase to the flexibility factor when compared to the baseline LoD 1, with little difference seen between LoD 2 and LoD 3.

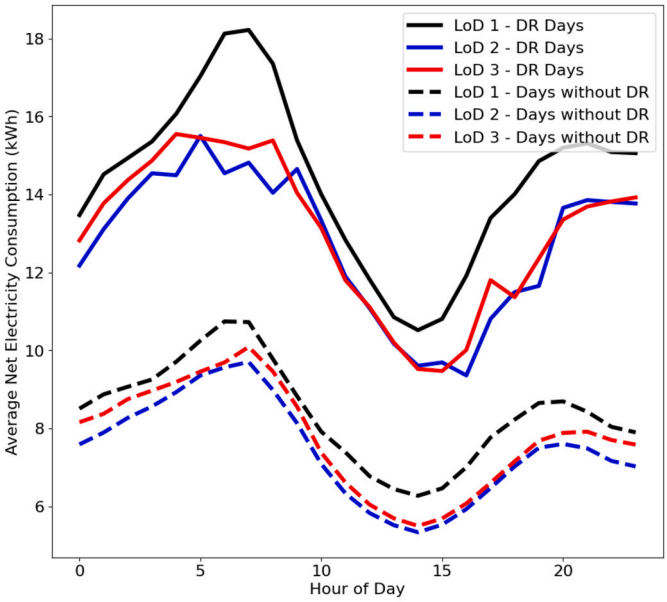


Fig. 12. Average net electricity consumption by hour for days with DR events and days without DR events.

3.4. Occupant thermal comfort

We quantified and analyzed occupant discomfort using two metrics: firstly, the total number of unmet setpoint hours (>2 °C deviation from the setpoint) for LoD 2 and LoD 3 as well as the number of thermostat overrides for each household for LoD 3. Fig. 13 shows a heat map of the number of hours in which the deviation from the setpoint is equal to the value in the box. The darker colors are concentrated around -1.5 to 0 °C, thus indicating that the indoor air temperature often fell slightly below the setpoint temperature but rarely deviated more than 2 °C from the setpoint, as was encouraged by our reward function. We see visually as well that the indoor air temperature remained closer to the setpoint for timesteps of LoD 2 when compared to LoD 3, and this confirmed with the number of unmet setpoint hours displayed in Table 12. However, the number of unmet setpoint hours cannot be taken directly as a proxy for occupant thermal discomfort because allowing for the overrides (which adjust the thermostat by up to 1.5 °C) causes fluctuations in the setpoint that do not occur in LoD 2, thus resulting timesteps where the indoor air temperature has yet to reach the desired new setpoint. As such, while

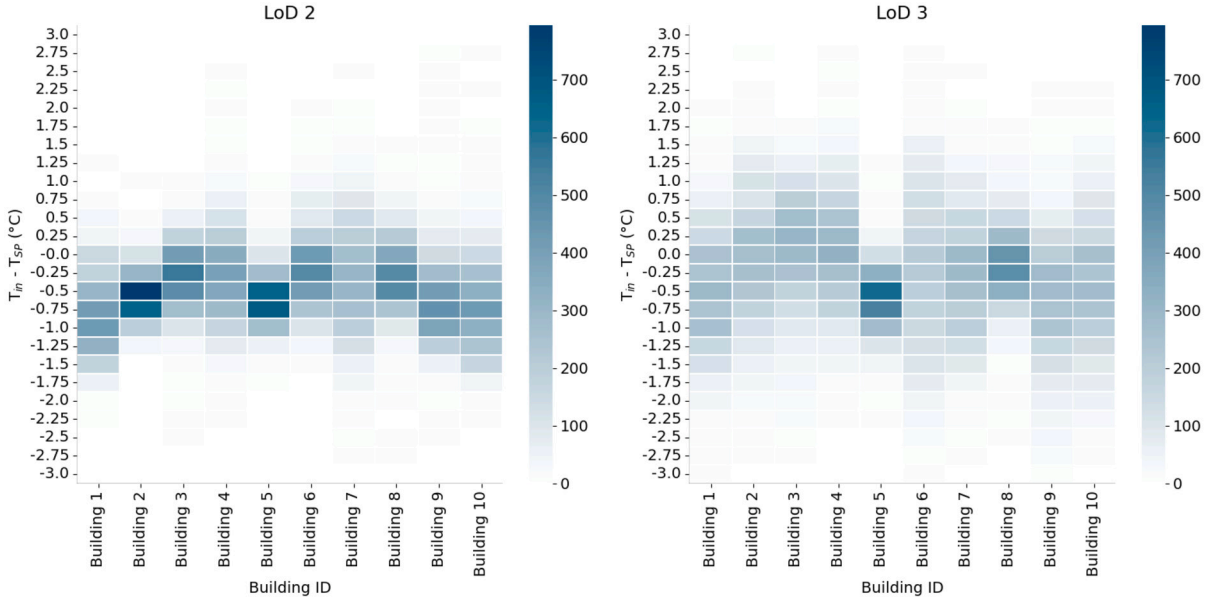


Fig. 13. Difference between indoor air temperature and setpoint temperature by frequency.

Table 12
Total number of unmet setpoint hours during 2160-hour (three-month) testing periods and percent of total hours.

Building ID	Unmet SP Hours		Unmet SP Hours (%)		Number of Overrides during DR (LoD 3)	Total Overrides LoD 3
	LoD 2	LoD 3	LoD 2	LoD 3		
1	14	122	0.6	5.6	33	106
2	22	99	1.0	4.6	6	35
3	18	61	0.8	2.8	7	65
4	31	42	1.4	1.9	7	29
5	16	25	0.7	1.2	10	47
6	22	168	1.0	7.8	19	105
7	52	83	2.4	3.8	13	79
8	39	37	1.8	1.7	10	36
9	53	207	2.5	9.6	18	101
10	31	111	1.4	5.1	20	89

the setpoint may deviate further from the indoor air temperature on average for LoD 3 when compared to LoD 2, we note the added level of control and adjustment to the thermostat as an important factor.

Also presented in Table 12 is the number of overrides by the occupants for LoD 3. While the DR events represented about 4% of the total timesteps, approximately 21% of the total occupant overrides occurred during the periods of DR, likely due to the decreased setpoints during these periods.

Finally, Fig. 14 shows the number of occupant overrides and the magnitude of the overrides throughout the 10 training episodes (each training episode being one cycle through the 2020 and 2021 winter heating seasons). The darker red and blue colors indicate the larger overrides that occur, predominately in the first few episodes, and subsequently we see that the later episodes show overrides of smaller magnitude and frequency. We also note that by using the methodology of one agent per household, the agent is able to minimize the overrides over time and learn the setpoint preferences regardless of Occupant Type (Average or Tolerant), thus presenting a suitable multi-agent methodology for handling occupants of differing preferences and override behaviors.

4. Discussion

With many interesting results to deliberate on, we firstly discuss our choice of methodologies for the respective LoDs. LoD 2 represents a common assumption in literature (see for example [19], [21]) whereby occupant comfort is assumed so long as the indoor air temperature is maintained within certain temperature bounds, usually between 1-3 °C of the setpoint. In our case, the RL agent reward is structured to keep the indoor air temperature of each building up to 2 °C less than the setpoint temperature without penalty, as described in detail in Section 2.4. Because there is no interaction from the occupant included in this LoD, energy-efficient behavior is thus inherent within this reward function, though it is possible that this configuration overestimates the real energy savings that we could expect in a real-world application. We see this in the results of the KPIs, where LoD 2 shows reductions in electricity cost, net electricity consumption, and average daily peak of 3% more than LoD 3, which is not surprising given that there are no occupant overrides. When looking specifically at the periods of DR events, on average LoD 2 reduced the net electricity consumption by 5% more than LoD 3. However, some individual houses showed discrepancies between LoD 2 and LoD 3 as large as 30%, as shown in Building 2. This insight is important for understanding the efficacy of various DR programs and the implications behind assumptions we make with regard to occupant thermal comfort and thermostat override behavior. Importantly, we also demonstrate that even with realistic occupant overrides, we were able to achieve approximately 17% reductions in net electricity consumption during the DR events for LoD 3. We thus show that it is possible to allow occupants to override the thermostat setpoints as well as achieve energy savings, and targeting setpoint reductions during peak periods could be one way to do so.

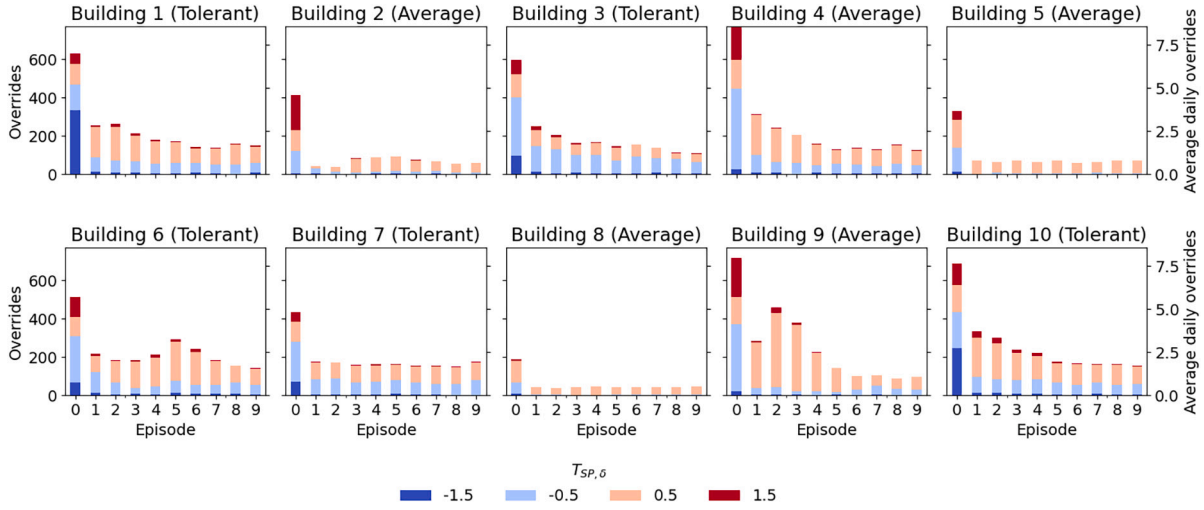


Fig. 14. Number of occupant overrides and magnitude of the override for the 10 training episodes.

When we compare our method to an MPC-based optimization, we note the advantages of the RL method. By using the multi-agent reinforcement learning approach, our agents learned optimal decisions that considered their occupants' thermostat preferences and maintained the indoor air temperature at or slightly less than the setpoint temperature. Because the agent learns through trial and error via the reward function, this approach does not require a physics-based model of the home for optimal decision making and thus can be less computationally intensive than MPC. It is difficult to directly compare results from one study to another due to varying control parameters and regional cost differences, but [53] showed the FF to be equal to 0.27 and 0.37 for two different MPC approaches while controlling the heating system of a single home in the Netherlands, thus showing that our results of 0.32 fall within a similar range.

The RL approach also becomes increasingly more practical when deployed at scale, as an agent can be assigned to each individual house rather than developing, testing, and calibrating physics-based models for each home in the district. Furthermore, while we do use LSTM models to simulate the indoor air temperature at each timestep, this is only to have an estimation for the changing indoor air temperature as a result of the heat pump action and the readings are not used for optimization or forecasting purposes. If implementing our methodology in practice, the LSTMs would no longer be needed as the actual reading from the thermostat of the indoor air temperature would be used. However, a drawback of the RL method is that sufficient training data is needed for the agents to be able to learn optimal decision making prior to deployment. In this case, we show that two seasons' worth of data was sufficient, but the MPC method means that the model could be immediately deployed. Despite these differences, the RL approach proves optimal when considering a larger scale of houses and occupant preferences as the agents can learn each house's optimal decisions, and showing this represents the key contribution of our work.

While the RL methodology used in this study showed promising results, we note several limitations in the approach we took for this study. Firstly, the results from LoD 2 and LoD 3 show increases to ramping and 1 − load factor. As mentioned in Table 10, ramping represents the change in electricity demand from one timestep to the next while 1 − load factor represents the improvement to the ratio between the average and peak loads, thus these factors characterize the slope of the demand curve. When referring back to Fig. 9, we can see that during periods of many setpoint overrides (e.g. timesteps 485-495 for Building 1), the heating demand also oscillated significantly. At timestep 485, the indoor air temperature dips down to below 16 °C, causing several subsequent increases to the thermostat setpoint. Because of the lag in time from the setpoint change to the more optimal indoor air temperature, the setpoint is increased up to 22 °C at timestep 489 while the indoor air temperature of the home does not reach this temperature until several hours later at timestep 494, at which point the setpoint restores to the original schedule at just below 20 °C. This means that now the agent is penalized for the indoor air temperature being slightly too high, and thus we see an oscillating pattern of increases and decreases as the agent tries to restore the indoor air temperature to ideal settings. This is likely a contributing factor as to why the ramping and 1 − load factor were increased for LoD 2 and LoD 3. Of note, the reward function used did not explicitly seek to flatten the demand curve over each timestep. However, it is an important consequence of the methodology that while we achieved promising results for other principal KPIs, additional work could focus on flattening the aggregate load demand curve. We did not include future setpoint schedules in our environment state variables (e.g. the thermostat setpoint in one hour or two hours as dictated by the schedule), which could be one way to help the agent both meet the current heating needs of the building as well as plan for future needs simultaneously.

In this study, we also only address the winter heating season during which there is a considerable difference between outdoor air temperature and indoor air temperature. As such, the estimated percentages in electricity savings as a result of the minor decreases to the thermostat setpoint may not translate to as significant of savings in a region with more mild temperatures. Furthermore, thermostat setpoint preferences and override behaviors are likely to change in shoulder seasons and summer cooling seasons, and thus additional occupant-thermostat override behavior models would need to be developed to apply this methodology in other climates.

Additionally, the reward function used in this study maintains occupant comfort by penalizing large deviations from the setpoint. Incorporating occupant thermostat overrides into the reward function, such as penalizing for each timestep when an override occurs, was considered but not applied in this study. This could potentially further exacerbate the oscillating effect of the agent actions as the agent accumulates additional override penalties. However, because our reward function does not directly penalize the overrides, it is necessary that the occupant manually changes their thermostat program in the event that their setpoint preferences have changed over time, rather than the agent learning their new setpoint preferences. Future work could focus on the agent selecting the thermostat setpoint directly rather than adjusting the heat pump heating supply power, which could enable the agent to learn setpoint preferences over time.

Finally, this study does not focus on occupancy as a potential factor as the EULP buildings (and associated occupant parameters assigned to each building) resulted in very high occupancy rates of approximately

90% or more for all buildings. However, Pinto et al. [46] exploited unoccupied periods to achieve additional energy savings with automated HVAC control. Obtaining more realistic occupancy data and leveraging unoccupied periods could also lead to even higher performance, which is another key element that could be pursued in future work.

5. Conclusions

This study controls the HVAC heating systems of 10 residential buildings in Quebec for neighborhood-level energy management. We consider within the control scheme two distinct occupant types: Average and Tolerant, and three Levels-of-Detail for the occupants in the control scheme. LoD 1 represents our baseline scenario with no control of the HVAC system, while LoD 2 assumes that the occupant is comfortable if the indoor air temperature is within 2 °C of the setpoint temperature. LoD 2 and LoD 3 include reductions to the thermostat setpoint during simulated DR events for reduced electricity consumption during these periods compared to LoD 1. Our highest LoD, LoD 3, also incorporates occupant-thermostat override models in which the occupant can override the thermostat at any timestep, including during the DR events. To the best of the authors' knowledge, this study represents the first time that data-driven occupant-thermostat override models have been incorporated into an HVAC control scheme, thus adding a level of detail not seen in previous works. The second major contribution of this work lies in incorporating multiple occupant types in a multi-agent approach to show how we might handle a real-world scenario where each household may have unique thermostat setpoint preferences and override behaviors.

We conclude that using a multi-agent approach, the agents are able to learn to keep the indoor air temperature near the setpoint temperature for both the Average and Tolerant setpoint schedules. The most significant savings came in the total electricity consumption, where 11% and 8% reductions were achieved for LoD 2 and LoD 3 respectively. Likely due to the setpoint reductions, LoD 2 and LoD 3 also achieved 22% and 17% reductions respectively in electricity consumption when looking specifically at the DR periods, with LoD 2 likely overestimating the amount of electricity savings while LoD 3 presents a more realistic scenario where the thermostat overrides are allowed. We found that approximately 20% of the total overrides occurred during the DR events, which represented 4% of the total timesteps, thus showing the significance that lowering the setpoints had on the override occurrence. Finally, we see that while we were able to achieve our objectives for several key metrics, including reducing electricity cost and consumption, both ramping and 1 − load factor increased under LoD 2 and LoD 3, and thus our control scheme did not help flatten the load curve of the district. While our RL reward function did not specifically target these KPIs, future work could be focused on how to also achieve better aggregate load shaping. However, the methodology used in LoD 3 still managed to balance both energy efficiency and occupant thermal comfort while integrating realistic occupant override behavior, and underscores the importance of integrating more realistic occupant behavior models to best estimate the return of DR programs.

CRediT authorship contribution statement

Kathryn Kaspar: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Kingsley Nweye:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Giacomo Buscemi:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Alfonso Capozzoli:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Zoltan Nagy:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition, Formal analysis, Conceptualization. **Giuseppe Pinto:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Ursula Eicker:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Conceptualization. **Mohamed M. Ouf:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Kathryn Kaspar reports financial support was provided by Quebec Research Fund - Nature and Technology Doctoral Research Scholarship (B2X). Kingsley Nweye and Zoltan Nagy received funding from the Climate Change AI Innovation Grants program Contract IG-2023-32. Mohamed M. Ouf reports financial support was provided by Natural Sciences and Engineering Research Council of Canada Discovery Grant RGPIN-2020-06804. Ursula Eicker reports financial support was provided by Canada Excellence Research Chairs Program with Grant CERC-2018-00005. Mohamed M. Ouf reports financial support was provided by Quebec Research Fund - Nature and Technology Research Support for New Academics Grant #315109. Giacomo Buscemi reports financial support was provided by the Italian Ministry of University and Research under project PRIN 2020: OPTIMISM-Optimal refurbishment design and management of small energy micro-grids. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Some source data is confidential, but code is available and has been shared at https://github.com/kkaspar10/Occupant_Thermostat_Int/.

Appendix A. Occupant-thermostat override model testing results

Table 13
Equations of logistic regression curves and resulting p-values to evaluate the fit of the curve.

Occupant Type	Probability of...	Scenario	a	b	p-value, a	p-value, b
Average	Increase	Morning/ Evening	17.8	-1.03	0.012	0.012
		Night	26.3	-1.56	0.040	0.040
		Mid-Day	15.1	-0.96	0.010	0.010
Average	Decrease	-	-28.2	1.03	0.021	0.022
Tolerant	Increase	Morning/ Evening	12.1	-0.82	0.007	0.007
		Night	21.1	-1.43	0.035	0.035
		Mid-Day	15.2	-1.02	0.016	0.015
Tolerant	Decrease	-	-23.2	0.97	0.010	0.010

Table 14
Random forest model parameters.

Parameter	Setting
Number of trees	300
Max. depth of tree	10
Min. samples for to split internal node	2
Min. samples for leaf node	2

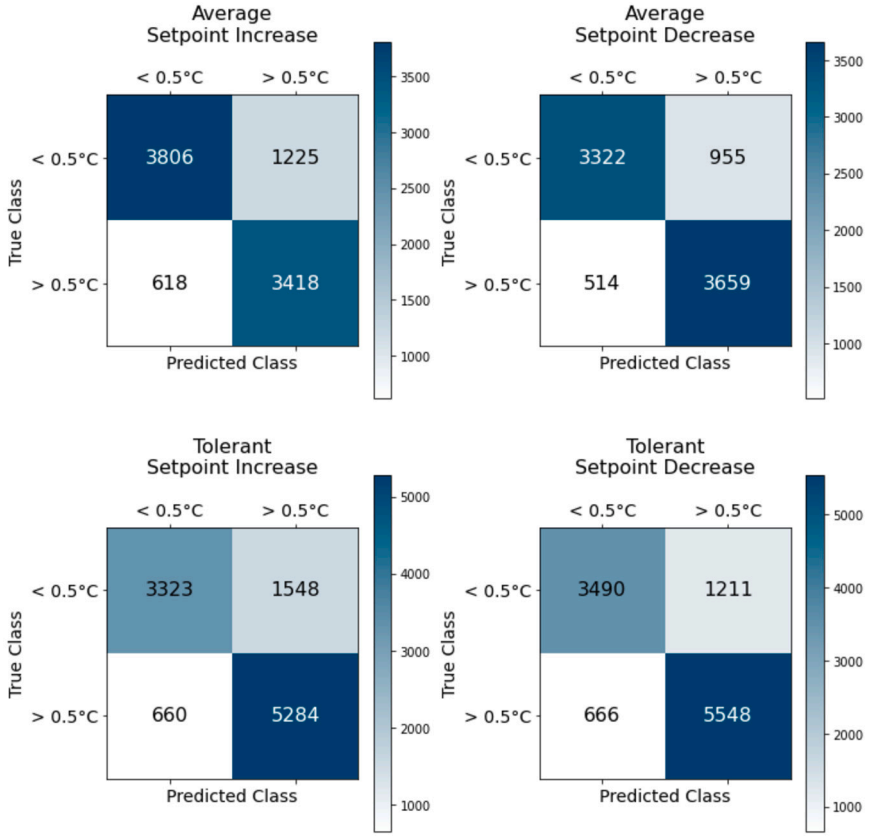


Fig. 15. Confusion matrices for the classification of the magnitude of a setpoint override for a) Average occupants making a setpoint increase; b) Average occupants making a setpoint decrease; c) Tolerant occupants making a setpoint increase; d) Tolerant occupants making a setpoint decrease.

Appendix B. Training and validation of building models

B.1. Validation of building energy system

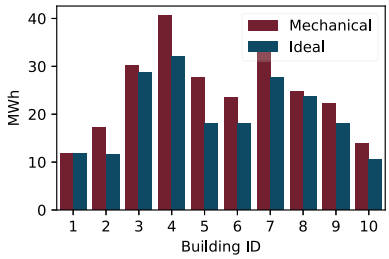
Fig. 16(a) shows the difference between total mechanical and ideal system supplied heating energy where the difference is 18.4±11.8% on average and 34.8% at maximum. The mechanical system delivers more energy compared to the ideal system due to the efficiency constraints of the mechanical system causing the indoor dry-bulb temperature to stray farther away from the setpoint thus needing more energy to maintain the setpoint. This is evident in Fig. 16(b) where we compare the number of unmet hours from using mechanical system, ideal system, or generic internal heat gain/loss equipment (supplying ideal load) to satisfy the heating loads in the buildings when using a 2 °C throttle range. There are no unmet hours when the ideal system is used compared to the mechanical system that results in six unmet hours on average from six of ten buildings. Only one building has unmet hours when the ideal system is replaced with the generic equipment. Nevertheless, the recorded unmet hours from using any of the system approach to meet heating load are well below the ASHRAE Standard 90.1 limit of 300 hours for every 8,760 hours (one year) [54]. Given the minimal difference in supplied energy by either mechanical or ideal system and the satisfactory number of unmet hours when switching from mechanical to ideal to generic equipment, we assert that we maintain the integrity of the as-provided energy model despite our modifications.

We show the distribution of indoor dry-bulb temperature difference between under-heating or overheating with generic internal heat gain/loss equipment and ideal system in the final four EnergyPlus simulations in Fig. 16(c). For all buildings, the difference is centered around zero with a majority of hours having a difference of less than 1.0 °C. On average under-heating and overheating result in 0.39±0.38 °C deviation from the ideal temperature. The distribution of temperature change between consecutive hours from under-heating or overheating with respect to ideal heating energy with heat generating equipment is shown in Fig. 16(d) with an absolute average of 0.56±0.52 °C. We assert that the indoor dry-bulb temperature and thermostat setpoint difference distribution as well as the temperature change distribution between consecutive hours have adequate variance that can be used to model the temperature dynamics of all of the buildings.

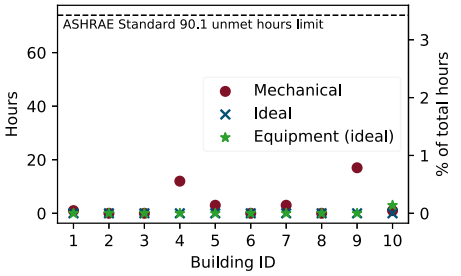
B.2. Generation of LSTM training data

The first simulation (*energyplus.simulation.1*) uses the as-provided mechanical HVAC system in each building model as a reference point for subsequent simulations. A second simulation (*energyplus.simulation.2*) where the mechanical HVAC system is replaced with an ideal load system is then run to determine the ideal heating load that satisfies the indoor dry-bulb temperature setpoint, and the result from this simulation is validated against that of *energyplus.simulation.1*. The validation ensures that the as-provided building energy model integrity is not tampered with when replacing mechanical systems with ideal load systems. Our check for integrity is that for the ideal load case, heating energy delivered to the building is equal or slightly less than that for the mechanical case. Indoor dry-bulb temperature in the building using the ideal load system should be similar to that from using the mechanical system.

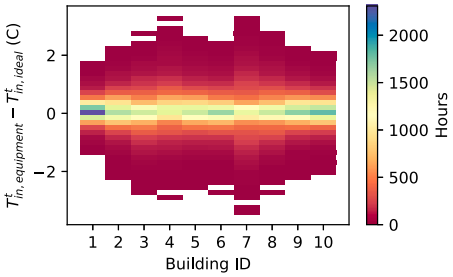
The remaining six simulations make up the LSTM training dataset where we replace the ideal load system with generic internal heat gain/loss equipment to simulate under-heating and overheating with respect to the building's ideal heating loads. The first of these six simulations (*energyplus.simulation.3*) supplies the ideal load from *energyplus.simulation.2* to validate the operation of the generic equipment compared to the ideal load system, as the effect on temperature should closely match. The next simulation (*energyplus.simulation.4*) free-floats



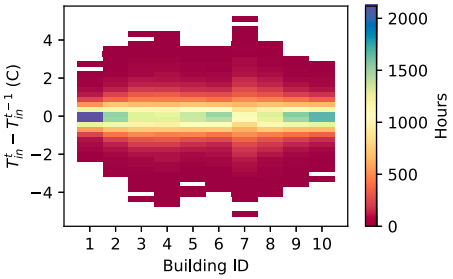
(a) Comparison between total mechanical and ideal system heating energy delivered to each building.



(b) Unmet hours (2°C throttle range) comparison from using mechanical system, ideal system or ideal heat generating equipment to meet the thermostat setpoint. The dashed reference line is the maximum allowable unmet hours prescribed in ASHRAE Standard 90.1 [54].



(c) Distribution of difference between temperature from under-heating or overheating with heat generating equipment and temperature from using ideal system.



(d) Distribution of temperature change between consecutive timesteps from under-heating or overheating with respect to ideal heating energy with heat generating equipment.

Fig. 16. EnergyPlus simulation summary for January-March 2020.

the temperature by supplying no heating energy to the building while the remaining simulations (*energyplus.simulation.5-8*) vary the percentage of ideal load for each simulation timestep between 80% and 120% with a 60% probability that the ideal load is varied. To estimate the building's internal temperature trend at a given instant *t*, temporal variables, meteorological variables, and the internal temperature related to previous timesteps were selected, along with the thermal load provided at *t*. The temporal variables were coded using sine and cosine transformation to enhance the continuity effect of the data. The selected variables, translated in time through the sliding window technique, are then merged with the internal temperature to provide the model with supervised input and output couples.

B.3. LSTM hyperparameters and testing results

Table 15
LSTM hyperparameter configuration. All models share the same number of epochs (70), batch size (168), and optimizer (Adam).

Building ID	Hidden Size	Num Layer	Dropout	Learning Rate
1	64	3	0.4	0.003042
2	56	3	0.0	0.003054
3	24	3	0.0	0.005248
4	16	2	0.1	0.007600
5	56	2	0.0	0.004720
6	40	3	0.2	0.004380
7	24	3	0.4	0.006086
8	64	3	0.2	0.007294
9	48	2	0.0	0.004284
10	56	3	0.4	0.002717

Table 16
MAPE Metrics.

Building ID	MAPE Train	MAPE Validation	MAPE Test in OL	MAPE Test in CL
1	24.73	50.66	3.87	1.21
2	7.26	16.33	1.63	1.31
3	7.45	6.88	1.93	1.18
4	816.63	33.68	4.58	1.98
5	9.02	5.94	1.38	0.94
6	64.62	66.82	1.57	1.17
7	29.94	52.36	1.91	1.52
8	8407.41	24.70	2.81	1.78
9	5.56	6.79	4.64	1.25
10	43.95	28.11	3.80	1.56

Table 17
RMSE Metrics.

Building ID	RMSE Train	RMSE Validation	RMSE Test	RMSE CL
1	0.92	1.07	0.34	0.97
2	0.47	0.46	0.35	0.43
3	0.29	0.32	0.30	0.48
4	1.06	1.20	0.62	1.28
5	0.41	0.32	0.27	0.39
6	0.71	0.88	0.31	0.41
7	1.50	1.80	0.40	0.50
8	0.62	0.78	0.46	0.73
9	0.43	0.38	0.40	1.30
10	1.01	1.20	0.44	1.00

References

[1] O.C.A. Alliance, How Ontario Can Avoid New Gas Plants and Lower Electricity Costs, Tech. Rep., 2023, Available from: https://www.cleanairalliance.org/wp-content/uploads/2023/01/Avoid-New-Gas-Plants-INTERACTIVE-jan-16-copy.pdf.

[2] Canada Energy Regulator, Canada's Energy Future Data Appendices, Tech. Rep., Canada Energy Regulator, 2023.

[3] Statistics Canada, The heat is on: how Canadians heat their home during the winter, 2023, https://doi.org/10.25318/3810028601-eng, Available from: https://www.statcan.gc.ca/o1/en/plus/2717-heat-how-canadians-heat-their-home-during-winter.

[4] U.S. Energy Information Administration, Residential energy consumption survey, Available from: https://www.eia.gov/consumption/residential/data/2020/, 2020. (Accessed 11 November 2023).

[5] S. Canada, Dwellings in Canada, Tech. Rep., Statistics Canada, 2017 May, 9, Available from: https://www12.statcan.gc.ca/census-recensement/2016/as-sa/98-200-x/2016005/98-200-x2016005-eng.cfm.

[6] M.H. Albadi, E.F. El-Saadany, A summary of demand response in electricity markets, Electr. Power Syst. Res. (2008) 8, https://doi.org/10.1016/j.epsr.2008.04.002.

[7] J. Han, M.A. Piette, Solutions for summer electric power shortages: demand response and its applications in air conditioning and refrigerating systems, Refrig. Air Cond. Electr. Power Mach. 19 (1) (2008) 1–4, Available from: https://www.smartgrid.gov/files/documents/Solution_for_Summer_Electric_Power_Shortages_Demand_Respon_200806.pdf.

[8] P. Fanger, Thermal Comfort, Danish Technical Press, Copenhagen, 1970.

[9] H. Du, Z. Lian, D. Lai, L. Duanmu, Y. Zhai, B. Cao, Y. Zhang, X. Zhou, Z. Wang, X. Zhang, Z. Hou, Evaluation of the accuracy of PMV and its several revised models using the Chinese thermal comfort database, Energy Build. 271 (2022) 112334, https://doi.org/10.1016/j.enbuild.2022.112334, Available from: https://www.sciencedirect.com/science/article/pii/S0378778822005059.

[10] S.I.u.H. Gilani, M.H. Khan, W. Pao, Thermal comfort analysis of PMV model prediction in air conditioned and naturally ventilated buildings, Energy Proc. 75 (2015) 1373–1379, https://doi.org/10.1016/j.egypro.2015.07.218, Available from: https://www.sciencedirect.com/science/article/pii/S1876610215009868.

[11] ASHRAE, ASHRAE Standard 55: Thermal Environmental Conditions for Human Occupancy, Tech. Rep., Atlanta, GA, 2010.

[12] B. Huchuk, W. O'Brien, S. Sanner, A longitudinal study of thermostat behaviors based on climate, seasonal, and energy price considerations using connected thermostat data, Build. Environ. 139 (2018) 199–210, https://doi.org/10.1016/j.buildenv.2018.05.003, Available from: https://www.sciencedirect.com/science/article/pii/S0360132318302634.

[13] K. Panchabikesan, M. Ouf, U. Eicker, G. Newsham, H. Knudsen, Investigating thermostat setpoint preferences in Canadian households, 2021 Sep, https://doi.org/10.26868/25222708.2021.30433, Available from: https://publications.ibpsa.org/conference/paper/?id=bs2021_30433. (Accessed 20 January 2023).

[14] M. Kane, Sharma K. Data-driven, Identification of occupant thermostat-behavior dynamics, https://doi.org/10.48550/arXiv.1912.06705, 2019.

[15] Brent Huchuk, W. O'Brien, S. Sanner, Exploring smart thermostat users' schedule override behaviors and the energy consequences, Sci. Technol. Built Environ. 27 (2021) 195–210, https://doi.org/10.1080/23744731.2020.1814668.

[16] L. Sarran, H.B. Gunay, W. O'Brien, C.A. Hviid, C. Rode, A data-driven study of thermostat overrides during demand response events, Energy Policy 153 (2021) 112290, https://doi.org/10.1016/j.enpol.2021.112290, Available from: https://www.sciencedirect.com/science/article/pii/S0301421521001592.

[17] M. Vellei, S. Martinez, J. Le Dréau, Agent-based stochastic model of thermostat adjustments: a demand response application, Energy Build. 238 (2021) 110846, https://doi.org/10.1016/j.enbuild.2021.110846, Available from: https://www.sciencedirect.com/science/article/pii/S0378778821001304.

[18] I. Hazyuk, C. Ghiaus, D. Penhouet, Model Predictive Control of thermal comfort as a benchmark for controller performance, Autom. Constr. 43 (2014) 98–109, https://doi.org/10.1016/j.autcon.2014.03.016, Available from: https://www.sciencedirect.com/science/article/pii/S0926580514000685.

[19] M. Razmara, G.R. Bharati, D. Hanover, M. Shahbakhti, S. Paudyal, R.D. Robinett, Building-to-grid predictive power flow control for demand response and demand flexibility programs, Appl. Energy 203 (2017) 128–141, https://doi.org/10.1016/j.apenergy.2017.06.040, Available from: https://www.sciencedirect.com/science/article/pii/S0306261917307936.

[20] H. Zhang, S. Seal, D. Wu, F. Bouffard, B. Boulet, Building energy management with reinforcement learning and model predictive control: a survey, IEEE Access 10 (2022) 27853–27862, https://doi.org/10.1109/ACCESS.2022.3156581.

[21] A. Berouine, R. Ouladsine, M. Bakhouya, M. Essaaidi, A predictive control approach for thermal energy management in buildings, Energy Rep. 8 (2022) 9127–9141, https://doi.org/10.1016/j.egyr.2022.07.037, Available from: https://www.sciencedirect.com/science/article/pii/S2352484722013038.

[22] X. Zhang, G. Schildbach, D. Sturzenegger, M. Morari, Scenario-based MPC for energy-efficient building climate control under weather and occupancy uncertainty, in: 2013 European Control Conference (ECC), 2013, pp. 1029–1034.

[23] H.F. Scherer, M. Pasamontes, J.L. Guzmán, J.D. Álvarez, E. Camponogara, J.E. Normey-Rico, Efficient building energy management using distributed model predictive control, J. Process Control 24 (2014) 740–749, https://doi.org/10.1016/j.jprocont.2013.09.024, Available from: https://www.sciencedirect.com/science/article/pii/S0959152413001935.

[24] T. Zhang, M.P. Wan, B.F. Ng, Yang S. Model, Predictive control for building energy reduction and temperature regulation, in: 2018 IEEE Green Technologies Conference (GreenTech), 2018, pp. 100–106.

[25] MATLAB, Available from: https://www.mathworks.com/products/matlab.html.

[26] IBM ILOG CPLEX Optimizer, Available from: https://www.ibm.com/products/ilog-cplex-optimization-studio/cplex-optimizer.

[27] G. Serale, M. Fiorentini, A. Capozzoli, D. Bernardini, A. Bemporad, Model predictive control (MPC) for enhancing building and HVAC system energy efficiency: problem formulation, applications and opportunities, Energies 11 (2018), https://doi.org/10.3390/en11030631, Available from: https://www.mdpi.com/1996-1073/11/3/631.

[28] Z. Wang, T. Hong, Reinforcement learning for building controls: the opportunities and challenges, Appl. Energy 269 (2020) 115036, https://doi.org/10.1016/j.apenergy.2020.115036, Publisher: Elsevier.

[29] Richard S. Sutton, Andrew G. Barto, Reinforcement Learning: An Introduction, second edition, The MIT Press, Cambridge, MA, ISBN 9780262039246, 2018.

[30] E. Mocanu, D.C. Mocanu, P.H. Nguyen, A. Liotta, M.E. Webber, M. Gibescu, J.G. Slootweg, On-line building energy optimization using deep reinforcement learning, IEEE Trans. Smart Grid 10 (2019 Jul) 3698–3708, https://doi.org/10.1109/TSG.2018.2834219, Available from: https://ieeexplore.ieee.org/document/8356086/. (Accessed 16 June 2021).

[31] B.V. Mbuwir, K. Paridari, F. Spiessens, L. Nordstrom, G. Deconinck, Transfer learning for operational planning of batteries in commercial buildings, in: 2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), IEEE, Tempe, AZ, USA, 2020 Nov, pp. 1–6, Available from: https://ieeexplore.ieee.org/document/9303016/. (Accessed 31 March 2021).

[32] J.J.R. Vázquez-Canteli, J. Kämpf, G. Henze, Z. Nagy, CityLearn v1.0: an OpenAI gym environment for demand response with deep reinforcement learning, in: BuildSys 2019 - Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, ISBN 9781450370059, 2019, pp. 356–357.

[33] D. Blum, J. Arroyo, S. Huang, J. Drgoňa, F. Jorissen, H.T. Walnum, Y. Chen, K. Benne, D. Vrabie, M. Wetter, L. Helsen, Building optimization testing framework (BOPTEST) for simulation-based benchmarking of control strategies in buildings, J. Build. Perform. Simul. 14 (2021 Sep) 586–610, https://doi.org/10.1080/19401493.2021.1986574, Available from: https://www.tandfonline.com/doi/full/10.1080/19401493.2021.1986574.

[34] Z. Zhang, K.P. Lam, Practical implementation and evaluation of deep reinforcement learning control for a radiant heating system, in: Proceedings of the 5th Conference on Systems for Built Environments, BuildSys'18, Shenzen, China. New York, NY, USA, 2018, pp. 148–157.

[35] N. Luo, T. Hong, Ecobee Donate Your Data 1,000 homes in 2017, 2022 Mar, https://doi.org/10.25584/ecobee/1854924, Available from: https://www.osti.gov/biblio/1854924.

[36] ecobee, User Guide ecobee3, Tech. Rep., 2014, Available from: https://storage.googleapis.com/article_attachments/pdfs/ecobee3_UserGuide.pdf.

[37] P.J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math. 20 (1987) 53–65, https://doi.org/10.1016/0377-0427(87)90125-7, Available from: https://www.sciencedirect.com/science/article/pii/0377042787901257.

[38] Hydro Quebec, Electricity Rates, Hydro Quebec, 2023, Available from: https://www.hydroquebec.com/documents-data/official-publications/electricity-rates-conditions-electricity-service.html.

[39] M.M. Ouf, M. Osman, M. Bitzilos, B. Gunay, Can you lower the thermostat? Perceptions of demand response programs in a sample from Quebec, Energy Build. 306 (2024) 113933, https://doi.org/10.1016/j.enbuild.2024.113933, Available from: https://www.sciencedirect.com/science/article/pii/S0378778824000495.

[40] W. Liu, H.B. Gunay, M.M. Ouf, Modeling window and thermostat use behavior to inform sequences of operation in mixed-mode ventilation buildings, Sci. Technol. Built Environ. 27 (2021 Oct) 1204–1220, https://doi.org/10.1080/23744731.2021.1936629, Available from: https://www.tandfonline.com/doi/full/10.1080/23744731.2021.1936629. (Accessed 7 August 2023).

[41] K.E. Kaspar, M.M. Ouf, U. Eicker, Data-driven occupant-thermostat override models for winter heating in Quebec, in: Proceedings of SimBuild Conference 2024, vol. 11. IBPSA-USA Building Simulation Conference. IBPSA-USA, IBPSA-USA, Denver, Colorado, 2024 May, pp. 725–734, Available from: https://publications.ibpsa.org/conference/paper/?id=simbuild2024_2167.

[42] C.Z. El-Bayeh, I. Mougharbel, M. Saad, A. Chandra, D. Asber, L. Lenoir, S. Lefebvre, Novel soft-constrained distributed strategy to meet high penetration trend of PEVs at homes, Energy Build. 178 (2018 Nov) 331–346, https://doi.org/10.1016/j.enbuild.2018.08.023, Available from: https://linkinghub.elsevier.com/retrieve/pii/S0378778818307588. (Accessed 24 September 2021).

[43] (LARC) NLRC, NASA LARC power hourly API, Available from: https://power.larc.nasa.gov/docs/services/api/temporal/hourly/, 2023.

[44] E.J.H. Wilson, A. Parker, A. Fontanini, E. Present, J.L. Reyna, R. Adhikari, C. Bianchi, C. CaraDonna, M. Dahlhausen, J. Kim, A. LeBar, L. Liu, M. Praprost, L. Zhang, P. DeWitt, N. Merket, A. Speake, T. Hong, H. Li, N. Mims Frick, Z. Wang, A. Blair, H. Horsey, D. Roberts, K. Trenbath, O. Adekanye, E. Bonnema, R. El Kontar, J. Gonzalez, S. Horowitz, D. Jones, R.T. Muehleisen, S. Platthotam, M. Reynolds, J. Robertson,

K. Sayers, Q. Li, End-Use Load Profiles for the U.S. Building Stock: Methodology and Results of Model Calibration, Validation, and Uncertainty Quantification, 2022 Mar, Available from: https://www.osti.gov/biblio/1854582.

[45] K. Nweye, K. Kaspar, G. Buscemi, G. Pinto, H. Li, T. Hong, M. Ouf, A. Capozzoli, Z. Nagy, A framework for the design of representative neighborhoods for energy flexibility assessment in CityLearn, in: Proceedings of Building Simulation 2023: 18th Conference of IBPSA. Building Simulation, IBPSA, Shanghai, China, 2023, Available from: https://publications.ibpsa.org/conference/paper/?id=bs2023_1404.

[46] G. Pinto, D. Deltetto, A. Capozzoli, Data-driven district energy management with surrogate models and deep reinforcement learning, Appl. Energy 304 (2021 Dec) 117642, https://doi.org/10.1016/j.apenergy.2021.117642, Available from: https://www.sciencedirect.com/science/article/pii/S0306261921010096. (Accessed 28 April 2023).

[47] Y. Yu, X. Si, C. Hu, J. Zhang, A review of recurrent neural networks: LSTM cells and network architectures, Neural Comput. 31 (2019 Jul) 1235–1270, https://doi.org/10.1162/neco_a_01199.

[48] J. Vivian, E. Prataviera, N. Gastaldello, A. Zarrella, A comparison between grey-box models and neural networks for indoor air temperature prediction in buildings, J. Build. Eng. 84 (2024) 108583, https://doi.org/10.1016/j.jobe.2024.108583, Available from: https://www.sciencedirect.com/science/article/pii/S2352710224001517.

[49] B.K. Park, C.J. Kim, Short-term prediction for indoor temperature control using artificial neural network, Energies 16 (2023), https://doi.org/10.3390/en16237724, Available from: https://www.mdpi.com/1996-1073/16/23/7724.

[50] T. Haarnoja, A. Zhou, P. Abbeel, S. Levine, Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor, 2018, https://doi.org/10.48550/ARXIV.1801.01290, Available from: https://arxiv.org/abs/1801.01290. (Accessed 14 April 2023).

[51] J.R. Vazquez-Canteli, S. Dey, G. Henze, Z. Nagy, CityLearn: standardizing research in multi-agent reinforcement learning for demand response and urban energy management, arXiv 2020 Dec. Available from: http://arxiv.org/abs/2012.10504.

[52] J. Le Dréau, P. Heiselberg, Energy flexibility of residential buildings using short term heat storage in the thermal mass, Energy 111 (2016) 991–1002, https://doi.org/10.1016/j.energy.2016.05.076, Available from: https://www.sciencedirect.com/science/article/pii/S0360544216306934.

[53] C. Finck, R. Li, W. Zeiler, Optimal control of demand flexibility under real-time pricing for heating systems in buildings: a real-life demonstration, Appl. Energy 263 (2020) 114671, https://doi.org/10.1016/j.apenergy.2020.114671, Available from: https://www.sciencedirect.com/science/article/pii/S0306261920301835.

[54] ASHRAE, ASHRAE Standard 90.1 - Energy Standard for Sites and Buildings Except Low-Rise Residential Buildings, Atlanta, GA, 2022.