SimBuild 2024

Eleventh National Conference of IBPSA-USA
Denver, Colorado
May 21 – 23, 2024

Data-Driven Occupant-Thermostat Override Models for Winter Heating in Quebec

Kathryn Kaspar[1], Mohamed M Ouf[1], and Ursula Eicker[1]
[1]Concordia University, Montreal, QC, Canada

Abstract

Understanding occupant thermostat preferences and thermostat setpoint override behavior is critical for exploiting thermostat setpoints or HVAC systems for energy flexibility and demand response potential. This study models occupant thermostat overrides in residential buildings during the heating season in Quebec, Canada. Two distinct occupant types, 'Average' and 'Tolerant', are identified based on their thermostat setpoint preferences. Discrete-time Markov logistic regression models are developed to predict the probability of a thermostat setpoint override during 'Morning/Evening', 'Night', and 'Mid-Day' hours for both occupant types. Random forest models are employed to classify the magnitude of the setpoint changes as either small (less than 0.5 °C) or large (greater than 0.5 °C). Using these models, we are thus able to estimate the probability of a setpoint override and the magnitude of the setpoint override for two distinct occupant types in Quebec. Results indicate good model performance with balanced accuracy, recall, and precision. However, limitations include data availability, model assumptions, and the need for more comprehensive occupancy data. These models offer potential applications in demand response and HVAC control strategies for residential buildings.

Introduction

Background

The electrification of transportation, conversion of existing buildings from natural gas to electricity, and the expansion of industries and infrastructure is expected to cause an increase of 20 TWh (or 12%) in the total electricity demand of Quebec from 2019 to 2029 (Hydro-Quebec 2020). Hydro-Quebec, the government-run energy utility company of Quebec, seeks to curb this increase in demand with a multi-faceted plan, including an expansion of renewable energy supply and energy efficiency measures to reach carbon-neutral operations by 2030. In the Hydro-Quebec Strategic Plan 2022-2026, energy efficiency measures for existing residential buildings including home renovations and Demand Response (DR) programs are projected to achieve savings of 1.1 TWh by 2025, a key component to combating the increasing consumption of electricity (Hydro-Quebec 2022).

DR relies on consumers adjusting electricity consumption based on grid conditions and price signals, and is crucial for energy savings because it empowers homeowners to actively participate in the optimization of their energy consumption. However, encouraging participation in DR is challenging as it requires changes in consumer behavior, such as lowering thermostat temperatures in winter or delaying the use of appliances until off-peak hours. In recent years, smart devices and IoT have facilitated energy flexibility and enhanced DR participation through Virtual Power Plants (VPPs) capable of coordinating distributed energy resources like energy storage systems, EVs, and smart thermostats. This not only contributes to grid reliability but also can lower energy costs for consumers. To better test and develop these VPP algorithms, representative models of occupant behavior and preferences are thus crucial.

As space heating needs represent approximately 60% of annual residential building energy consumption in Canada, adjustments in thermostat temperatures during the cold winter season represents a substantial opportunity for energy savings (Natural Resources Canada 2019). In Quebec specifically, the majority of residential homes use electricity for space heating (Canada Energy Regulator 2019). Hydro-Quebec has launched several programs, such as the Rate Flex D program, that offer financial incentives for adjusting thermostat setpoints during peak hours (Pelletier and Faruqui 2022). Occupants change thermostat settings for a number of reasons, including thermal discomfort, vacation, and energy conservation, and understanding how occupants interact with thermostats remains a difficult task. Smart thermostats can also be difficult to program or interpret, thus causing dissatisfaction in performance and inefficient operations (Tamas, O'Brien, and Quintero 2021). However, the availability of smart thermostat data such as ecobee's Donate Your Data (Luo and Hong 2022) or Google Nest (Yang and Newman 2012) can help us understand what drives occupant thermostat interactions as well as occupant thermal comfort preferences, and is shaping the way in which we study and model occupant thermal comfort.

Related Works on Thermal Comfort Modeling

Initially introduced by P.O. Fanger, the Predicted Mean Vote (PMV)-Predicted Percentage Dissatisfied (PPD) model outlined a quantitative framework for evaluating thermal comfort based on parameters such as air temperature, humidity, air velocity, and clothing insulation (Fanger 1970). However, (Gilani, Khan, and Pao 2015) found that Fanger's PMV-PPD model overestimated thermal sensation by 33% for buildings with an HVAC system and (Du et al. 2022) found that the accuracy of the PMV model never exceeded 60% and performed even worse in cooler sensations. The PMV-PPD model prompted the evolution of standards such as ASHRAE Standard 55 (ASHRAE 2010) and ISO 7730 (ISO 2005), which recognized the need for a broader consideration of factors influencing comfort. These standards continue to provide a reference for appropriate operative temperature levels for varying clothing levels, metabolic rates, and humidity levels. However, ASHRAE 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. Within this range, there is an estimated 10% dissatisfaction from those who exhibit whole body thermal discomfort as well as 10% dissatisfaction from those that may experience partial body thermal discomfort based on Fanger's PMV-PPD index.

While these guidelines remain industry standards, smart thermostat data has been leveraged to inform occupant thermal comfort and is particularly useful for studying the ranges in occupant thermal comfort preferences in the residential sector. Smart home thermostats are programmable devices that allow users to set temperature schedules for their heating and cooling systems, while also providing the flexibility to change these settings manually when needed, thus triggering a thermostat override. (Kane and Sharma 2019) studied the amount of time occupants were present at home prior to making a setpoint override and analyzed the energy consequences of the overrides. (Huchuk, O'Brien, and Sanner 2021) too analyzed thermostat overrides from an energy consequence perspective and found that just a small proportion of data from the 20,000 ecobee thermostats had behaviors resulting in 2-4% increases in energy consumption, but did not attempt to model the drivers or frequency of override of the occupants. (Sarran et al. 2021) 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. In a Quebec-specific study, (Panchabikesan et al. 2021) used the ecobee data to create four clusters of heating setpoint profiles, though did not address thermostat overrides made by the occupants.

The goals of this study are thus to:

1. Identify general thermostat setpoint preferences for multiple occupant types
2. Define two distinct occupant types (Average, Tolerant) in Quebec
3. Model the probability of thermostat setpoint overrides during different times of day for both occupant types
4. Propose a methodology for estimating the magnitude of the setpoint overrides

We firstly discuss the methodology taken to process the data, identify the occupant types, and determine the probability and magnitude of a setpoint override. We subsequently present the results and discuss the limitations of the models produced. The contribution of these models is to have multiple data-driven occupant behavior models to use for developing and testing algorithms for the automated control of HVAC systems in cold climates, as well as to inform general thermostat user behavior and preferences in the region.

Methodology

Data Processing

Thermostat data from 701 houses in Quebec was obtained from ecobee, including house identifiers, timestamps, thermostat setpoints, indoor temperature, HVAC system state, and thermostat events (e.g., 'Morning/Evening') (Luo and Hong 2022). Only data from November to March was considered to focus on the heating season. Using python package *meteostat* (Meteostat 2022) the ecobee thermostat data was then merged with local weather data to consider the significance of outdoor conditions on indoor thermal comfort. A full list of the variables considered is shown in Table 1.

Table 1: Summary of data available for each timestep

Variable	Description
$Identifier$	House ID
$datetime$	Timestamp at 5-minute intervals
$HVACMode$	HVAC system state
$CalendarEvent$	Thermostat event (e.g. 'Home')
$T_{\mathbf{in}}$	Indoor air temperature
$T_{\mathbf{in(t-1)}}$	Indoor air temperature (t-1 hour)
$T_{\mathbf{SP}}$	Setpoint temperature
$T_{\mathbf{SP(t-1)}}$	Setpoint temperature (t-1 hour)
$T_{\mathbf{in}} - T_{\mathbf{SP}}$	Unmet setpoint amount
$(T_{\mathbf{in}} - T_{\mathbf{SP}})_{\mathbf{t-1}}$	Unmet setpoint amount (t-1 hour)
$T_{\mathbf{out}}$	Outdoor air temperature
$T_{\mathbf{out(t-1)}}$	Outdoor air temperature (t-1 hour)
$T_{\mathbf{in}} - T_{\mathbf{out}}$	Temperature difference
$(T_{\mathbf{in}} - T_{\mathbf{out}})_{\mathbf{t-1}}$	Temperature difference (t-1 hour)
$RH_{\mathbf{in}}$	Indoor relative humidity
$RH_{\mathbf{out}}$	Outdoor relative humidity
$hour$	Hour
day	Day of the week

Each household can program various thermostat settings that appear under the 'Calendar Event' column, such as 'Home', 'Away', and 'Vacation', among others, and there is a custom schedule that dictates the hours and days in which thermostat events will occur. A 'Hold' event indicates that the thermostat setpoint was manually changed by the occupant from the original thermostat schedule, and can last for a custom duration set by the user, usually between 2-4 hours. For the 701 houses, all instances in which a thermostat setpoint was changed from the original setpoint schedule, thus initiating a setpoint 'Hold' event, were extracted, herein referred to as a thermostat *override*. It was expected that conditions triggering an increase in the thermostat setpoint (e.g., too cold) would be different than conditions resulting in a decrease in setpoint (e.g., too warm, energy conservation). Setpoint increases and decreases were thus separated in the data analysis.

When modeling the probability of occurrence of a setpoint increase in an office building, (Liu, Gunay, and Ouf 2021) noted how seven out of 20 occupants showed little to no correlation between a setpoint increase and an environmental variable, such as indoor or outdoor temperature, while the remaining 13 occupants showed strong correlations ($p - value < 0.05$). Low correlations could be due to several factors, perhaps due to the presence of multiple drivers behind the setpoint change, including discomfort, habit, environmental savings, and inefficient thermostat use. For example, an occupant could increase the thermostat setpoint in summer to shut off the air conditioning either because they are too cold or because they anticipate leaving the house soon. In a winter scenario, it is anticipated that the lower the indoor air temperature is in the house, the higher the likelihood is that the occupant will increase the setpoint, thus representing an inverse relationship between T_{in} and a setpoint override ('0' if no override, '1' if an override). To best extract discomfort-driven behavior, we thus used Pearson Correlation to select homes with a correlation less than -0.20 (some correlation) between indoor temperature and an increase to the temperature setpoint, thus resulting in a dataset of 119 homes for our study.

Cluster Identification

K-means clustering was then used to identify clusters based on the average indoor temperature during a setpoint override. The clustering with the maximum Silhouette score indicated distinct separation between each cluster as well as small intra-cluster distances (Rousseeuw 1987). Average indoor temperature and setpoint preferences were then determined for the clusters.

Probability of Setpoint Override

The probability of a setpoint override must be determined for the period in which the home is occupied. While some houses' thermostats indicated via the 'Calendar Event' whether the occupant was 'Home' or 'Away', other thermostats did not explicitly specify whether the occupant was home because the event indicated 'Custom123' or 'FirstNameHere', for example. Canadian Time-Use surveys were used to estimate that occupants spend approximately 70% of their time during an average week at home (Statistics Canada 2018). All of the override events were collected for each home, assuming that occupants were at home during this time. Subsequently, all data points where the words 'Away' or 'Vacation' were present in the 'Calendar Event' column were removed from the dataset, indicating that the occupant was not present during this time. However, the number of data points containing 'Away' or 'Vacation' was less than the estimated 30% weekly away period estimated by the time-use surveys. Additional data points were thus randomly sampled until totaling the 30% 'Away' period, and the remaining 70% of the dataset thus represented the occupied period.

Figure 1 shows the average number of setpoint overrides at each hour of the day per house, and demonstrates that the probability of a thermostat override varies depending on the time of day. We thus modeled the probability of setpoint override during three periods of the day: 'Night', 'Mid-Day', and 'Morning/Evening', with the hours of each scenario listed in Table 2. In line with (Karjalainen 2009) and (Gunay et al. 2018), the occurrence of a setpoint override is rare, approximately one override every two days. In a review of variables and models used to predict indoor thermal comfort, indoor and outdoor air temperatures were the variables most frequently used (Ma et al. 2021), with several studies basing their thermal comfort models using air temperature as the sole indoor input variable (Moon, Yoon, and Kim 2013); (Kim, Lee, and Moon 2014). The methodology used in (Liu, Gunay, and Ouf 2021) and (Gunay et al. 2018) was thus used where indoor air temperature is the input variable for univariate discrete-time Markov logistic regression models, an industry standard as referenced in (D'Oca et al. 2019). 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 are determined in the training, and the fit of the curve is evaluated using the p-value.

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

Table 2: 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

Magnitude of Setpoint Override

The distribution of the magnitude of setpoint overrides was first determined, which could be randomly sampled to estimate the magnitude of the setpoint override. In addition, a Pearson Correlation revealed T_{SP}, $(T_{\text{SP}})_{\text{t-1}}$, and $(T_{\text{in}} - T_{\text{SP}})_{\text{t-1}}$ were the variables most correlated to the magnitude of a setpoint override. Random forest models were then developed to classify setpoint changes as either small (less than 0.5 °C) or large (greater than 0.5 °C).

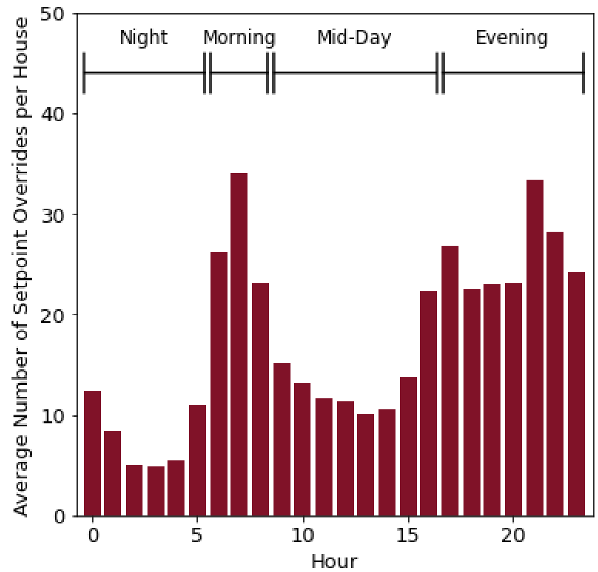

Figure 1: Average number of setpoint overrides at each hour of the day per house

Results

Cluster Identification

K-means clustering led to two distinct clusters being identified by maximizing the Silhouette score as shown in Figure 2. Table 3 shows the average indoor air temperature of each cluster at the time of a setpoint change, and the clusters are herein referred to as *Average* and *Tolerant* based on their comfort preferences.

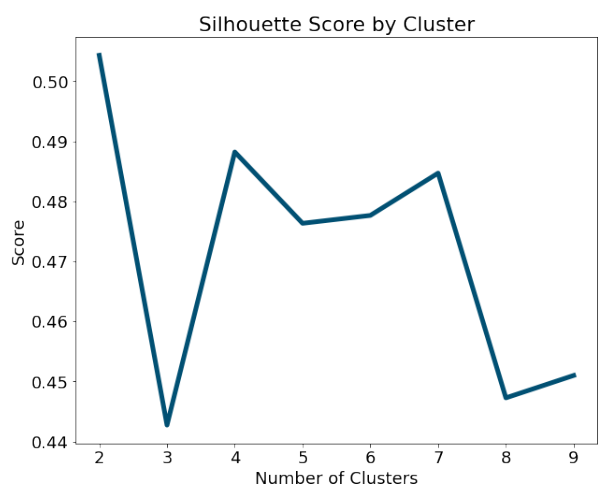

Figure 2: Silhouette score by the number of clusters

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

For the 64 homes in the Average cluster and the 55 homes in the Tolerant cluster, Figure 3 shows the distribution of indoor air temperatures at all timesteps across the winter seasons to show the distinct temperature preferences. Figure 4 also shows the average daily setpoint temperature of each cluster.

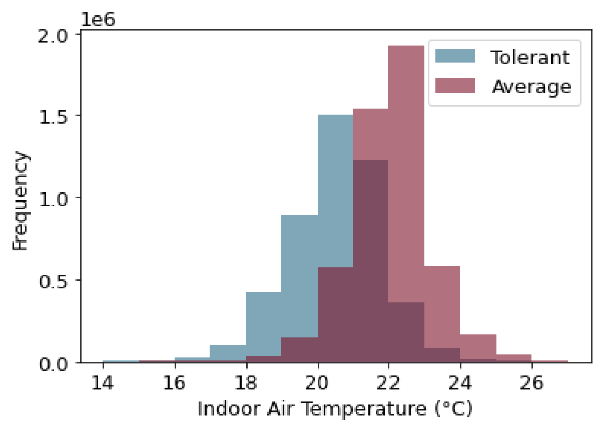

Figure 3: Histograms of indoor air temperature for Tolerant and Average occupants across all timesteps

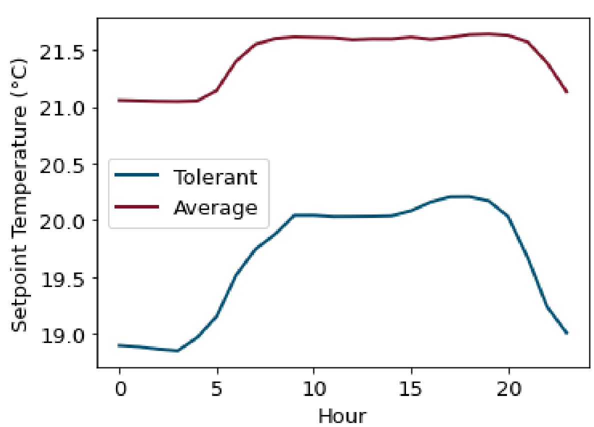

Figure 4: Average setpoint temperatures for Tolerant and Average occupants

Probability of Setpoint Override

Figure 5 shows a sample curve where we plot the probability of a setpoint increase in the next hour for Tolerant occupants during the 'Night' scenario. The size of the scatter plot points represent the number of timesteps of data for which the indoor air was at that temperature, thus shown by the largest blue circles occurring around 18-20 °C. The black dashed line shows the logistic regression equation fitted to the curve. Override data was only available between 14.5 and 23.5 °C for this scenario, shown by the circles in blue. Outside this temperature range, there are some instances of indoor air temperature but no override data. For values less than our data range, we thus assumed a 100% probability of increase of the thermostat and likewise a 0% probability of increase for values greater than our data range, as shown by the red dots on the figure. When modeling the probability of a setpoint decrease, we conversely assumed a 100% probability of decrease for temperatures greater than the range of the override data and a 0% probability of decrease for values less than the range of the override data.

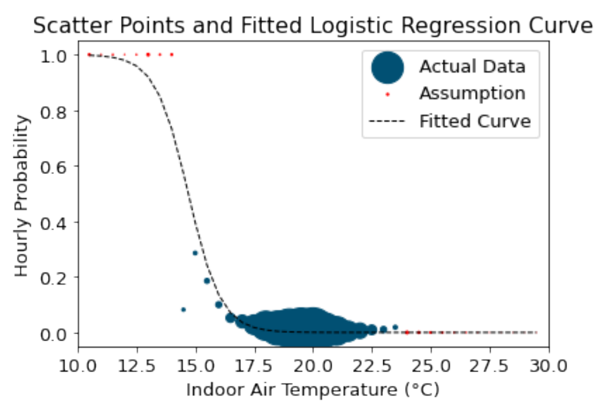

Figure 5: Probability of a setpoint increase in the next hour at various indoor air temperatures, with assumptions where data is not available

Figure 6 shows the discrete-time Markov logistic regression models that were developed for the three times of day for each cluster, indicating the probability of a setpoint override as a function of the indoor temperature in the next hour. Each scenario is modeled as a logistic regression function in the form of Equation (1). 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 Table 4.

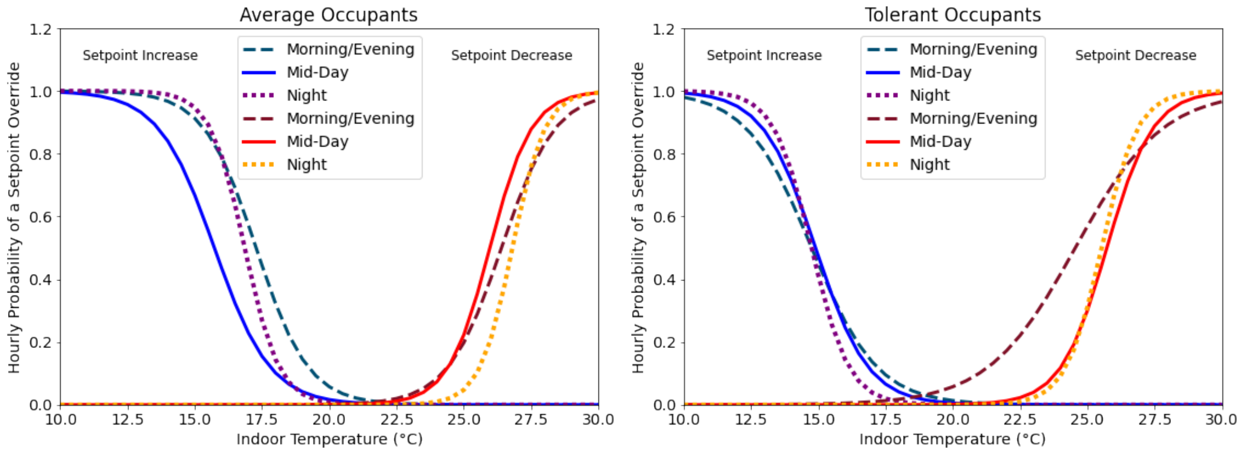

Figure 6: Probability of a setpoint override in the next hour as a function of indoor temperature

Table 4: 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
		Morning/Evening	17.8	-1.03	0.012	0.012
Average	Increase	Night	26.3	-1.56	0.040	0.040
		Mid-Day	15.1	-0.96	0.010	0.010
		Morning/Evening	-26.4	1.00	0.038	0.040
Average	Decrease	Night	-44.5	1.66	0.049	0.049
		Mid-Day	-33.5	1.29	0.028	0.027
		Morning/Evening	12.1	-0.82	0.007	0.007
Tolerant	Increase	Night	21.1	-1.43	0.035	0.035
		Mid-Day	15.2	-1.02	0.016	0.015
		Morning/Evening	-15.1	0.615	0.005	0.006
Tolerant	Decrease	Night	-38.3	1.50	0.037	0.037
		Mid-Day	-30.1	1.17	0.033	0.034

Magnitude of Setpoint Override

Figure 7 shows the distribution of the magnitude of setpoint overrides with an average change of approximately $\pm 0.5\,^{\circ}$C. When implementing the probability models developed in the previous step, the magnitude of the setpoint override could be estimated by randomly sampling from the distribution.

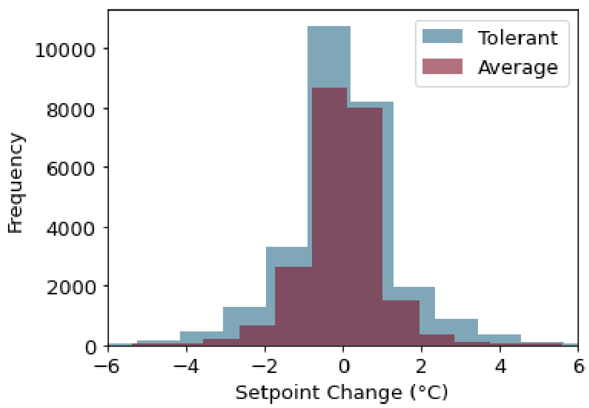

Figure 7: Magnitude of setpoint overrides for Tolerant and Average occupants

Optionally, random forest models can be used to predict the estimated magnitude of the setpoint override. To develop the random forest models, a grid search was performed to determine the optimal random forest parameters as indicated in Table 5. The results of a 5-fold cross-validation using 80/20% training/testing split are presented for the random forest models in Figure 8. Each occupant type (Average, Tolerant) has one model to predict the quantity (small, large) of a setpoint increase and one model to predict the quantity (small, large) of a setpoint decrease. As shown in the results, the classification accuracy was balanced among the two classes and performed well, with average accuracy of approximately 80%, average recall of approximately 85%, and average precision of approximately 77%.

Table 5: 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

Discussion

While the models developed in this study have shown promising results in predicting occupant thermostat overrides, several challenges and assumptions underpin their development. One notable limitation is the reliance on data from only 119 out of 701 houses, which makes the models less representative of the full spectrum of behaviors exhibited in Quebec residences. There is firstly a potential bias in the temperature settings of the houses as well as override behaviors due to the fact that these homes voluntarily chose to install smart thermostats and participate in the ecobee 'Donate Your Data' program, which could result in more environmental habits than the typical population. For example, the distribution of indoor air temperatures shown for the Average households in Figure 3 falls within the optimal comfort range indicated by ASHRAE guidelines, while the Tolerant households often exhibit average indoor air temperatures below the ASRHAE recommendations (ASHRAE 2010). This finding is significant because the indoor air temperature preferences of the Tolerant cluster, representing 46% or 55 homes, would be considered by the ASHRAE Standard 55 to be operating in temperatures where 80% of occupants would be uncomfortable. While additional datasets would be necessary to know for what percentage of the province the Tolerant cluster represents, there is still a subset of the population with a preference for lower indoor air temperatures and a higher tolerance to colder temperatures perhaps due to the long winters in Quebec. This further highlights the importance that individual temperature preferences continue to be studied, especially across different countries, genders, and climates. In addition, because two clusters was found to be the optimal number of clusters to best represent the dataset, a third cluster for 'Sensitive' households was not addressed in this study but could be analyzed with the availability of additional data sources. These models also assume a 70% weekly home occupancy rate, drawing from Canadian Time-Use Survey data, thereby restricting their applicability to homes with similar occupancy rates. The inclusion of more accurate and extensive occupancy data, possibly leveraging ecobee's occupancy sensors, could significantly enhance the reliability of these models. However, it is important to acknowledge that not all houses have occupancy sensors, and the sensors often do not capture all occupancy instances, such as when occupants are asleep.

Using logistic regression to fit the models also introduces certain limitations. Given the sigmoidal nature of the logistic curve, the behavior at 'comfort' temperatures tends

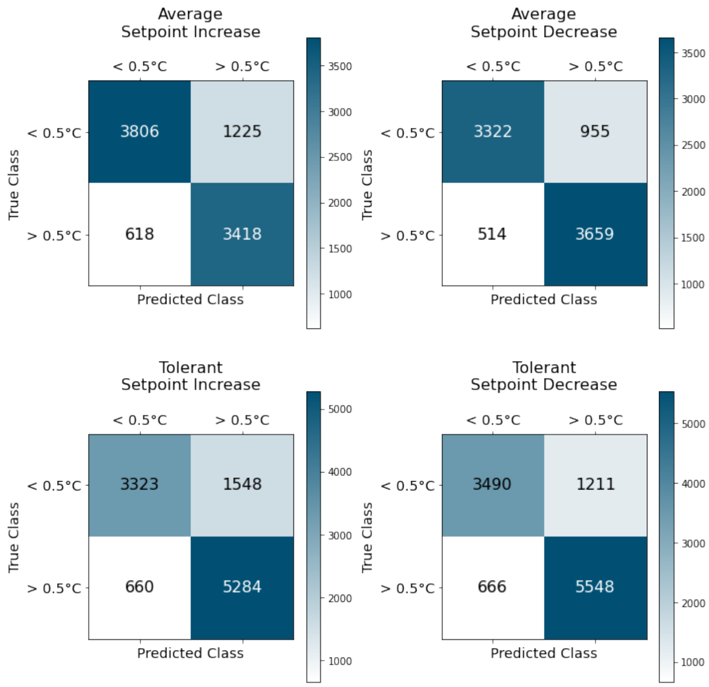

Figure 8: Confusion matrices for the classification of the magnitude of a setpoint override for (a) Average occupants making a setpoint increase, (b) Average occupants making a setpoint decrease, (c) Tolerant occupants making a setpoint increase, (d) Tolerant occupants making a setpoint decrease

to be underestimated while behavior at 'discomfort' temperatures tends to be overestimated. This results in an apparent absence of setpoint overrides at temperatures such as 20 °C. In reality, we see the most thermostat setpoint overrides at these temperatures, thus meaning that our model is underestimating the amount of setpoint overrides at this temperature though the probability of an override remains low. Additionally, the Markov property assumption implies that future thermostat overrides depend solely on the current state of the indoor air temperature and not past history, which may oversimplify the nuanced nature of occupant thermostat use. This neglects other variables that could be drivers for thermostat overrides, such as individuals adjusting the thermostat upon arriving home due to sudden or drastic changes in temperature between the outdoor and indoor environment. Likewise, sharp changes in the indoor air temperature of the home, drafts from windows and other local temperature changes, as well as activity levels are known factors that can contribute to occupants overriding their thermostat setpoint, which our univariate model does not consider. These factors can lead people to interact with the thermostat frequently over a short duration as well while trying to calibrate the temperature, and the time-series nature of these events leading to additional thermostat overrides is not considered due to the Markov property. Consequently, while the results in classifying the magnitude of a setpoint override are promising, they are also contingent on having a comprehensive model for the probability of a thermostat override.

In line with this idea, it is imperative to note that not all thermostat overrides are driven by discomfort alone as we assume in this study, and extracting discomfort-driven behavior from the dataset poses inherent challenges. It is possible that as the usability of thermostats improves, the frequency (and thus probability) of thermostat overrides could decrease. However, the methodology proposed in this study whereby extracting occupants with high correlations between setpoint overrides and indoor air temperature could provide a useful solution to this problem.

Despite the limitations of the modeling process, the application of these models could prove beneficial in testing algorithms for DR. For example, Hydro-Quebec has direct load control programs in which the winter heating thermostat setpoint of each house can be lowered to achieve reductions in energy demand during DR events in exchange for monetary compensation. Depending on which houses are in which clusters, the setpoints could be tailored such that the Tolerant houses undergo larger decreases in the setpoint while the Average houses would experience smaller decreases to their thermostat setpoint during the DR event to best balance energy savings and thermal comfort. By providing two types of occupant models, it is possible to test algorithms that control the setpoint changes during DR events in their ability to manage the different occupant preferences for different homes. The development of two distinct profiles for occupant-thermostat behavior and three daily scenarios in which the behavior changes is a noteworthy contribution of this study. Future work entails integrating these models as an occupant behavior module for HVAC system control during the heating season, validating their performance in a control scheme, and testing their real-world applicability.

Conclusion

This study presented probability and magnitude models of setpoint overrides for two occupant types during the winter heating season in Quebec, Canada. The primary goal was to develop models that could predict when and to what extent occupants would override thermostat setpoints given various indoor air temperatures, with the ultimate aim of deriving models that could be used to test various occupant behaviors in automated HVAC or setpoint control algorithms.

One key outcome and novel contribution of this research is the data-driven identification of two distinct occupant types, Average and Tolerant, during three times of day: 'Morning/Evening', 'Night', and 'Mid-Day'. The occupant types exhibit different thermal comfort preferences which are reflected in their thermostat setpoint override behaviors. The models demonstrated good performance in capturing the relationship between indoor temperatures and setpoint overrides, and the coefficients of the logistic regression equations were statistically significant, indicating that the models provided a reasonable fit to the observed data. However, it is important to note that these models assume a Markov property, implying that future thermostat overrides depend solely on the current indoor temperature. While this assumption is a limitation, it allows for a simplified modeling approach that prioritizes implementation of these models in simulations.

Random forest models were employed to classify the magnitude of setpoint overrides, distinguishing between small (less than 0.5 °C) and large (greater than 0.5 °C) adjustments to the setpoint. The results showed that these classification models achieved a balanced accuracy of approximately 80%, indicating their ability to classify the setpoint changes with relatively high accuracy. However, these results are contingent upon having reliable models for predicting the occurrence of a setpoint override.

In conclusion, the developed models offer valuable insights into occupant thermostat overrides and have potential applications in various contexts. They can be used to simulate and test DR strategies, where thermostat setpoints are adjusted automatically during peak load events while considering occupant comfort, in lieu of traditional one-size-fits-all thermal comfort models presented in (ASHRAE 2010) and (ISO 2005). Future work should focus on integrating these models into practical control schemes and conducting real-world validation to assess their applicability and performance. Overall, this research contributes to the understanding of occupant thermal comfort preferences in residential buildings such that both energy consumption and occupant thermal comfort can be optimized.

Acknowledgments

This research was supported by the Fonds de Recherche du Québec Nature et technologies (FRQNT) Research Support for New Academics (Grant #315109), the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (RGPIN-2020-06804), and the Canada Excellence Research Chairs Program with grant number CERC-2018-00005. Kathryn Kaspar is additionally supported by the Fonds de recherche du Québec – Nature et technologies (FRQNT) Doctoral Research Scholarship (B2X).

References

ASHRAE. 2010. "ASHRAE Standard 55: Thermal Environmental Conditions for Human Occupancy." Technical Report, Atlanta, GA.

Canada Energy Regulator. 2019. "Canada's Energy Future Data Appendices." Technical Report.

D'Oca, Simona, H Burak Gunay, Sara Gilani, and William O'Brien. 2019. "Critical review and illustrative examples of office occupant modelling formalisms." *Building Services Engineering Research & Technology* 40 (6): 732–757.

Du, Heng, Zhiwei Lian, Dayi Lai, Lin Duanmu, Yongchao Zhai, Bin Cao, Yufeng Zhang, Xiang Zhou, Zhaojun Wang, Xiaojing Zhang, and Zhijian

Hou. 2022. “Evaluation of the accuracy of PMV and its several revised models using the Chinese thermal comfort Database.” *Energy and Buildings* 271:112334.

Fanger, P.O. 1970. *Thermal Comfort*. Copenhagen: Danish Technical Press.

Gilani, Syed Ihtsham ul Haq, Muhammad Hammad Khan, and William Pao. 2015. “Thermal Comfort Analysis of PMV Model Prediction in Air Conditioned and Naturally Ventilated Buildings.” *Energy Procedia* 75:1373–1379.

Gunay, H. B., William O’Brien, Ian Beausoleil-Morrison, and Jayson Bursill. 2018. “Development and implementation of a thermostat learning algorithm.” *Science and Technology for the Built Environment* 24 (1): 43–56.

Huchuk, Brent, William O’Brien, and Scott Sanner. 2021. “Exploring smart thermostat users’ schedule override behaviors and the energy consequences.” *Science and Technology for the Built Environment* 27 (2): 195–210.

Hydro-Quebec. 2020. “Faits saillants de l’État d’avancement 2021 du Plan d’approvisionnement 2021-2029.” Technical Report, Hydro-Quebec.

Hydro-Quebec. 2022, March. “Strategic Plan 2022-2026.” Technical Report, Hydro-Quebec. ISBN 978-2-550-90996-5 (PDF).

ISO. 2005. “ISO 7730: Ergonomics of the Thermal Environment - Analytical Determination and Interpretation of Thermal Comfort Using Calculation of the PMV and PPD Indices and Local Thermal Comfort Criteria.” Technical Report.

Kane, Michael, and Kunind Sharma. 2019. “Data-driven Identification of Occupant Thermostat-Behavior Dynamics.”

Karjalainen, Sami. 2009. “Thermal comfort and use of thermostats in Finnish homes and offices.” *Building and Environment* 44 (6): 1237–1245.

Kim, Sooyoung, Ji-Hyun Lee, and Jin Woo Moon. 2014. “Performance evaluation of artificial neural network-based variable control logic for double skin enveloped buildings during the heating season.” *Building and Environment* 82:328–338.

Liu, Weihao, H. Burak Gunay, and Mohamed M. Ouf. 2021. “Modeling window and thermostat use behavior to inform sequences of operation in mixed-mode ventilation buildings.” *Science and Technology for the Built Environment* 27 (9): 1204–1220 (October).

Luo, Na, and Tianzhen Hong. 2022. “Ecobee Donate Your Data 1,000 homes in 2017.” 3.

Ma, Nan, Dorit Aviv, Hongshan Guo, and William W. Braham. 2021. “Measuring the right factors: A review of variables and models for thermal comfort and indoor air quality.” *Renewable and Sustainable Energy Reviews* 135:110436.

Meteostat. 2022, August. meteostat 1.6.5 package.

Moon, Jin Woo, Sung-Hoon Yoon, and Sooyoung Kim. 2013. “Development of an artificial neural network model based thermal control logic for double skin envelopes in winter.” *Building and Environment* 61:149–159.

Natural Resources Canada. 2019. “Energy Use Data Handbook 1990-2017.” Technical Report, Natural Resources Canada.

Panchabikesan, Karthik, Mohamed Ouf, Ursula Eicker, Guy Newsham, and Heather Knudsen. 2021, September. “Investigating thermostat setpoint preferences in Canadian households.”

Pelletier, Frederic, and Ahmad Faruqui. 2022. “Does dynamic pricing work in a winter-peaking climate? A case study of Hydro Quebec.” *The Electricity Journal* 35 (2): 1–7.

Rousseeuw, Peter J. 1987. “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis.” *Journal of Computational and Applied Mathematics* 20:53–65.

Sarran, Lucile, H. Burak Gunay, William O’Brien, Christian A. Hviid, and Carsten Rode. 2021. “A data-driven study of thermostat overrides during demand response events.” *Energy Policy* 153:112290.

Statistics Canada. 2018. “General Social Survey (Time Use) 2015: Public Use Microdata File (PUMF).” Technical Report.

Tamas, Ruth, William O’Brien, and Mario Santana Quintero. 2021. “Residential thermostat usability: Comparing manual, programmable, and smart devices.” *Building and Environment* 203 (October): 108104.

Yang, Rayoung, and Mark W Newman. 2012, September. “Living with an intelligent thermostat: advanced control for heating and cooling systems.” Pittsburgh, PA, USA, 1102–1107.