Science and Technology for the Built Environment, (2020) **0**, 1–16

ISSN: 2374-4731 print / 2374-474X online
DOI: 10.1080/23744731.2020.1814668

Taylor & Francis
Taylor & Francis Group

Check for updates

Exploring smart thermostat users' schedule override behaviors and the energy consequences

BRENT HUCHUK[1]*, WILLIAM O'BRIEN[2], AND SCOTT SANNER[1]

[1]*Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Canada;*
[2]*Department of Civil and Environmental Engineering, Carleton University, Ottawa, Canada*

Programmable thermostats with frequent schedule overrides (i.e., "holds") are widely assumed to be detrimental to potential energy savings. We explored whether this assumption generalized to a population of smart thermostats using a longitudinal dataset of 20,000 devices from a single manufacturer. We observed a majority of devices were infrequently overridden, most often to less energy-efficient setpoints; nonetheless, the short duration of each override resulted in minimal increase in energy use. When we instead looked at the minority of thermostats consistently overridden (accounting for approximately 12% of the thermostats), we observed that a large fraction were being repeatedly manually adjusted, but often in an energy saving manner yielding energy performance on par with rarely overridden thermostats. Only the consistently overridden but infrequently adjusted population of thermostats showed poor energy performance, but these accounted for less than 5% of thermostats with energy performance roughly 2% worse on heating and 4% worse on cooling than the remaining population. While there may be effective ways to modify thermostat user interfaces to dissuade users from long-duration energy-inefficient hold settings, we conclude from our study of smart thermostats that user holding behavior may not be as detrimental to collective energy savings as commonly believed.

Introduction

Programmable thermostats have been seen as an ideal solution to a common problem; reducing energy consumption without affecting comfort (US Department of Energy 2017; Natural Resources Canada 2015). By allowing the thermostats to follow a programmed schedule that captures unoccupied or dormant periods in the home, temperature setpoints can be set back (or set up) at times when occupants will not notice. Substantial savings (over 30%) was estimated from applying these schedules (Moon and Han 2011). However, when installed programmable thermostats were evaluated, it became clear that the homes with programmable thermostats were often not saving or even spending more compared to homes using a non-programmable thermostat (Pritoni et al. 2015; Peffer et al. 2011; Nevius and Pigg 2000).

The savings from a thermostat are directly related to how the thermostat is used by the user and are not achieved simply from the installation of the device. Researchers investigated why there were drastic differences between results in the field and their initial expectations. Peffer et al. (2011; 2013) identified that poor usability was a common problem that led to sub-optimal operation. Alternatively for many users, schedules did not regularly provide savings potential because the occupancy of the homes never changed or was too sporadic to be scheduled (Nevius and Pigg 2000; Pritoni et al. 2015). Finally, several studies observed that large fractions of the sample population (30-50%) had their thermostat manually overridden. The overrides prevented even a properly programmed schedule from adjusting setpoint temperatures (Pritoni et al. 2015; Lopes and Agnew 2010; Sachs et al. 2012; Peffer et al. 2011, 2013). Additionally, the overrides were observed to be set to maintain comfortable temperatures and not to lower-energy setpoints (Sachs et al. 2012).

While the dialog around thermostat overrides appears damning, other factors should be considered. First, the majority of the investigations were limited by the available methods. The studies often relied on self-reported survey data (Sachs et al. 2012; Pritoni et al. 2015) that often gave conflicting answers. Furthermore, the researchers relied on proxy information such as temperature and runtimes to be able to infer overriding behaviors over short observation windows or on

Received April 20, 2020; accepted August 21, 2020
Brent Huchuk, **M.A.Sc.**, is a Ph.D. Candidate.
William (Liam) O'Brien, Ph.D., is an Associate Professor.
Scott Sanner, Ph.D., is an Assistant Professor.
*Corresponding author e-mail: brent.huchuk@mail.utoronto.ca

small samples (Sachs et al. 2012). Second, the schedule overrides were observed to be brief (Sachs et al. 2012) so over the long term the effects on energy efficiency might be minimal. Finally, users were found to often keep a similar control strategy to the one they used when managing their previous thermostats (Nevius and Pigg 2000). This finding indicates that individuals familiar with actively managing a static setpoint based on their occupancy may not require a schedule. Perhaps then the resulting narrative regarding schedule override behavior and energy-saving potential remains incomplete or more nuanced than previously discussed – requiring better visibility into a large population and high-quality human-thermostat-interaction data to be fully understood.

Programmable thermostats have continued to have wide market adoption with recent estimates indicating nearly 60% of homes have them installed (Energy Information Administration 2015). Programmable thermostats remain the specified requirement in energy savings guidelines (International Code Council 2015). However, a new class of thermostats, smart thermostats (ST), have entered the market. While still being a relatively recent technology, adoption has been quick. Forecasts have shown that STs will be installed in over 40 million homes in the U.S. by 2020 (Parks Associates 2017). These new devices, through their continual transmission of data, can provide new insights into the operation of thermostats by residential customers. For example, because of access to ST data, analyses have been possible on equipment runtimes (Touchie and Siegel 2018), motion/occupancy (Huchuk, Sanner, and O'Brien 2019), and setpoints over multiple years of observation (Huchuk, O'Brien, and Sanner 2018). These analyses for programmable thermostats would have been extremely challenging to implement and required sacrifices to study populations, duration, or geographic regions.

This article is the first to explore a large-scale dataset (exceeding 20,000 households of runtime and temperature data) to quantify the effects of holds and to conduct a nuanced analysis of user types and behaviors. The overall objective and key contribution of this paper is to understand if the override behaviors of ST users (based on a particular manufacture) are as detrimental to energy usage as the observations made previously on programmable thermostat users (Sachs et al. 2012; Nevius and Pigg 2000). In particular, the data from STs allow us to answer the following questions that inform this understanding:

Q1 What is the frequency of overrides by the entire sample?

Q2 Are there different types of user override behavior?

Q3 What are the energy impacts for each behavior type?

Q4 Are there potential mitigation techniques to reduce the frequency or length of overrides, or is it necessary to mitigate hold behavior at all?

This paper explores those four questions utilizing a large, open dataset containing multiple years of data for a single manufacturer of STs. The remainder of the paper is structured as follows. First (in Approach and data overview) we present an overview of the method and dataset, how data selection was conducted, and high-level summary statistics that addresses **Q1**. The section further motivates specific explorations. The following section (Exploratory analysis) explores the different overriding behaviors of ST users and addresses **Q2**. Next (The impacts of overrides), specifically quantifies the relative effects and energy impacts as a result of the behaviors of various user behavior types identified (**Q3**). We use the Mitigation strategies section to discuss the potential predictability and mitigation of the thermostat overrides (**Q4**) before outlining the limitations of this study (Discussion of limitations). Finally in the last section (Conclusions and future work), the main conclusions and future directions are presented.

Approach and data overview

Our study was conducted on an open dataset from the Donate Your Data (DYD) program administered by the ST manufacturer ecobee (ecobee Inc. 2018). We selected a subset of households who met minimum data length requirements and possessed specific equipment configurations (Thermostat selection). Using the selected sample of thermostats, we explored the behaviors over the entire sample (Schedule overrides) before breaking users down into further user types (The ratio of time in a schedule override and Frequency of override initialization or adjustment). The single infrequent-utilization user type was used to better understand how hold behavior is dependent on season, time of day, and outside temperature (Contextual factors to override behaviors). These investigations lead to the conclusion that the prediction of holds for the infrequent user type is difficult due to lack of consistent patterns, which makes hold preemption strategies difficult to apply (Mitigation strategies). But this brings us back to question of whether preemption or holds prevention strategies are necessary (i.e., how much energy loss is due to existing behaviors)? The remaining investigation of the article then, is where we quantify the runtime impacts of different hold behaviors (The impact of overrides).

Thermostat selection

The DYD dataset was made up of over 70,000 thermostats, found predominately in North America, with up to three years of measurements. The dataset had five generations of devices, with the oldest released in 2008 and the most recent released in 2017. The dataset consists of data both measured and saved in the cloud by the thermostat and user-provided metadata regarding the building (e.g., age of home, size of home, number of stories, etc.). Measurements made by the thermostat were recorded at five-minute intervals. The records include temperature, HVAC equipment runtimes, setpoints, motion, and schedule overrides. Specific information such as thermostat or sensor placement in the home are not provided. Previous work has provided more documentation and characterization of the dataset (Huchuk, O'Brien, and Sanner 2018; Touchie and Siegel 2018).

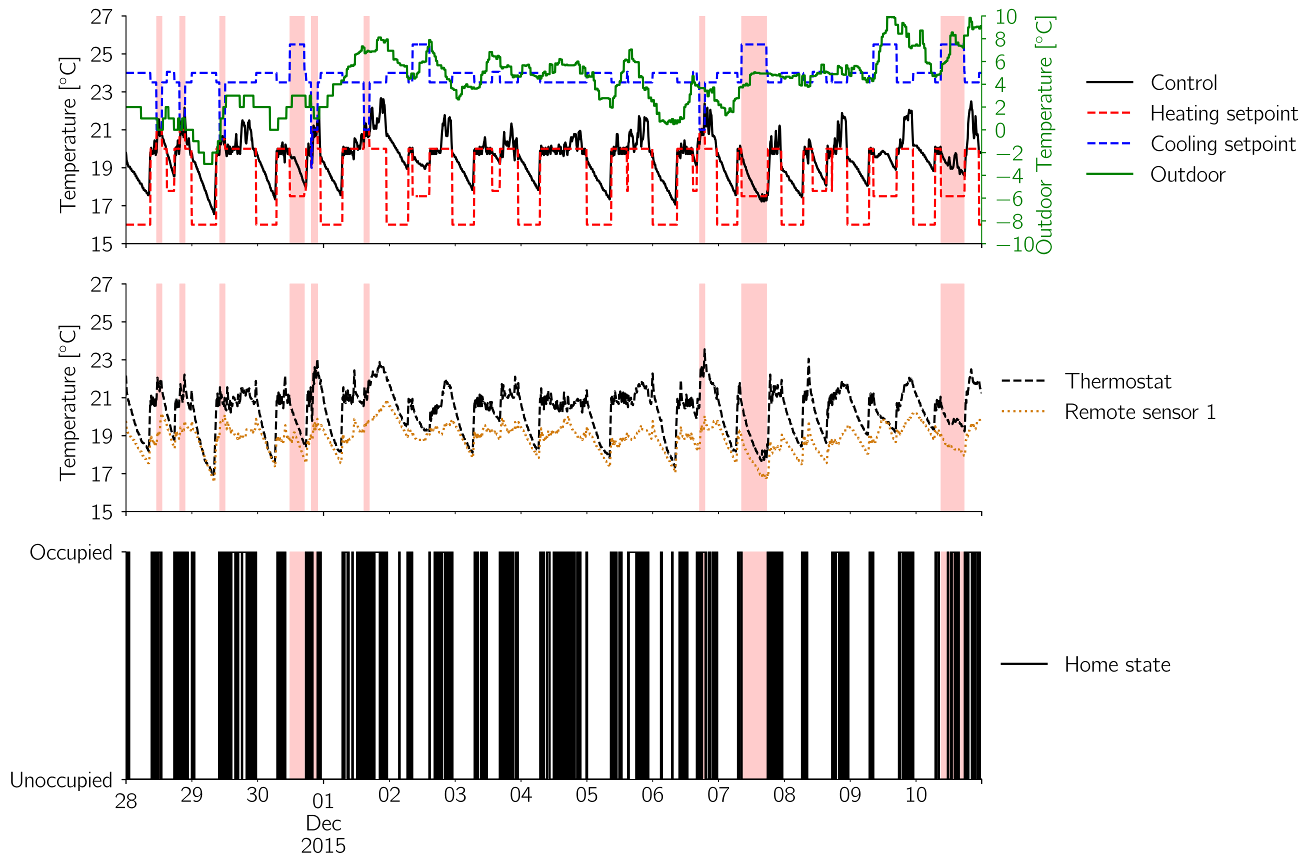

Fig. 1. Data from a single thermostat over a two-week period. Setpoint, control (based on weighted averages of sensors), and outdoor temperatures are shown (top) along with variation between remote sensor and thermostat temperatures (middle) and motion state within the home (bottom). Nine schedule overrides are indicated on each plot by the under-laid boxes (red).

Our investigation considered a subset of the thermostats in the data. The thermostat sample was limited to those whose listed household was only associated to one thermostat, that had over 300 days of data, and was active past August 1, 2018. Thermostats with a heat pump were excluded because their uncertain equipment configurations in heating seasons (i.e., compressor and auxiliary heating) and variable coefficient of performance exclude them from later runtime-based analyses. This screening narrowed the available sample down to 23,624 thermostats which henceforth will be considered the complete sample of devices (S1).

Schedule overrides

The thermostats from this particular manufacturer (ecobee) allow users to program a schedule with a different set of timings and setpoints each day. If a user has not programmed a schedule the thermostat is defaulted to a "sleep" and "home" period each day based on predefined default start and end times. Each scheduled period on the thermostat (i.e., sleep, home, away, or a custom one) is associated with both heating and cooling setpoints. Left alone by the user, a thermostat follows the setpoints associated with the period at the current time in the schedule – the standard behavior of any programmable thermostat. The STs in this dataset have additional reactive "smart" setpoint features. For example, when anticipating setpoint changes located in the schedule, the setpoints are adjusted to smooth transitions so the household is at the correct temperature at the time of the schedule change. Additionally, modest adjustments (within a couple of degrees) to the scheduled setpoints can be made on perceived occupancy or absence on some generations of devices. The occupancy state of the home is determined by the motion sensors both on the thermostat and located around the home in paired remote sensors. All of these feature-actions were marked in the interval data.

Users are able to override the schedule (which often colloquially is referred to as a "hold") on their thermostat only when the thermostat is in an active system mode (i.e., heating, cooling, or auto) and not turned off. The action can come from an intervention on the thermostat, through a mobile device, using the Web portal, through voice assistants, or other third-party integrations (e.g., with a registered app). With this dataset, it was not possible to tell the method by which the thermostat override was initiated. The interval data simply indicates that a hold was applied and the setpoint values indicated the new temperature setpoints being controlled against. The interval data indicated a hold was in place over each of the five-minute intervals until it was removed. During the override's duration, other smart features are prevented from operating and the thermostat cannot adjust its setpoints based on the schedule.

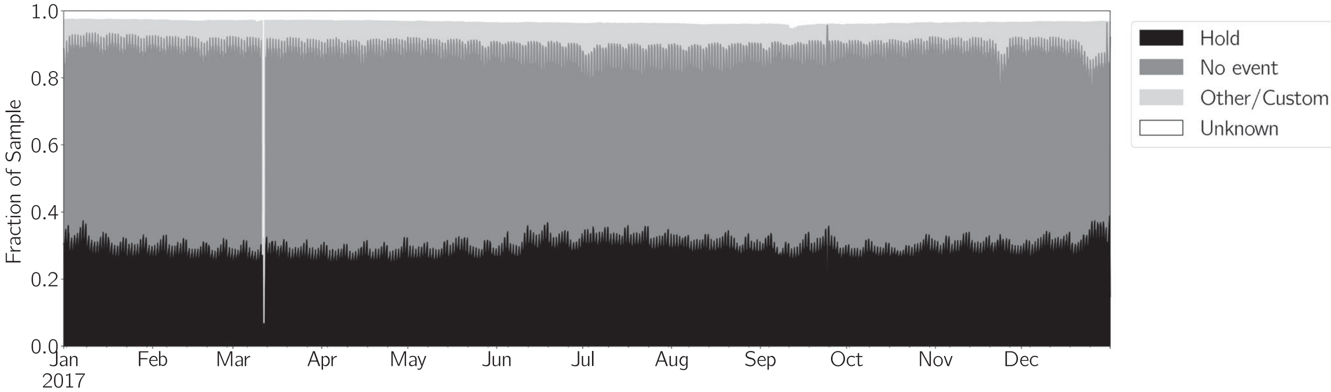

Fig. 2. The fraction of thermostats in hold, no event, other/custom and unknown system states for a sample of thermostats online over all of 2017 (n = 12,627).

The duration of a hold is based both on user behaviors and the preferences by which the thermostat has been configured. While a user is able to remove a hold at any time, thermostats from this manufacturer are configured to remain overridden for a predefined duration. The options include: (i) a two-hour duration, (ii) a four-hour duration, (iii) until the next scheduled period, (iv) indefinite (the default option), or (v) to be prompted with the choice of behavior (from the previously listed options) each time a hold is initiated or adjusted. The user-defined preference was not included in the DYD program's dataset and could be changed at any time by the user as part of the thermostat configuration. Without a reliable source for the user preference on duration, analyses of the reason for canceling a hold or the duration of an individual hold were not considered.

Figure 1 illustrates a pattern of overrides for a single thermostat's two-week period. This thermostat was placed into a hold nine times. The conditions at which the holds were set and the specific characteristics (i.e., selected setpoints and durations) of the holds are varied. The thermostat setpoints (top plot) are adjusted in both positive and negative directions, the overrides appear to be of differing lengths, at differing outdoor temperatures (top, secondary y-axis), indoor temperatures (middle), and with varying motion patterns (bottom) near the time of the hold being applied. The diversity of conditions when the holds are set on a single thermostat foreshadows the challenges in predicting when a hold was initiated across the sample; discussed in the section entitled Contextual factors to override behaviors.

To address **Q1**, a smaller selection of thermostats that had data over all of 2017 was selected from S1. At each five-minute timestamp the event status of every device in the sample was mapped to one of four states: (i) *hold* indicating an override, (ii) *no event* indicating the thermostat was operating just on the schedule, (iii) *other/custom* indicating an event other than an override by manual adjustment (i.e., smart features, utility events, or vacations), and (iv) *unknown* indicating data was missing. Figure 2 shows the fraction of the sample population in each state at each time interval.

Overall the fractions of each state are relatively stable. The spike in March is the result of many thermostats coming offline together. This was because of a widespread service issue by the entire sample. Missing data can be attributed to many factors that, unfortunately, are all indistinguishable from each other given the data. Common reasons for lost data include WiFi-connectivity issues within the home, disrupted electricity in the area, and occasionally Internet service provider outages in an area. Potentially the data could, in this observed spike, be missing due to server-side issues from the manufacturer given that it was so wide-spread. Typically, missing data observations for a thermostat are seen as short-duration anomalies (less than an hour) though longer events are seen. During any Internet-based outages, the thermostats would have continued to function.

Figure 2 indicates about 60% of the population consistently relied on a schedule (either programmed or default). Approximately 5% of the sampled population had missing data at any time resulting in an *unknown* state. Interestingly, it appears that roughly 30% to 35% of the population was engaged in a hold at any time. This value is on the lower end of the estimate range from studies on programmable thermostats (Pritoni et al. 2015; Peffer et al. 2011), but similar to some utility evaluations (Schellenberg, Lemarchand, and Wein 2017). Unlike those studies, Figure 2 shows that the fraction of people in a hold remains fairly constant over a calendar year. Small variations in the magnitudes can be seen occurring at a daily frequency which motivates the idea (Contextual factors to override behaviors) that there are consistent trends dictating when people adjust their overrides during a day.

Exploratory analysis

Addressing **Q2** requires a more detailed analysis of the available longitudinal data than was needed for answering **Q1**. Specifically, we want to know the amount of time individual thermostats are overridden (The ratio of time in a schedule override), and the frequency that prolonged holds (Frequency of override initialization or adjustment) are initiated or adjusted. Finally, we explore the contextual elements present at the time of a hold action (Contextual factors to

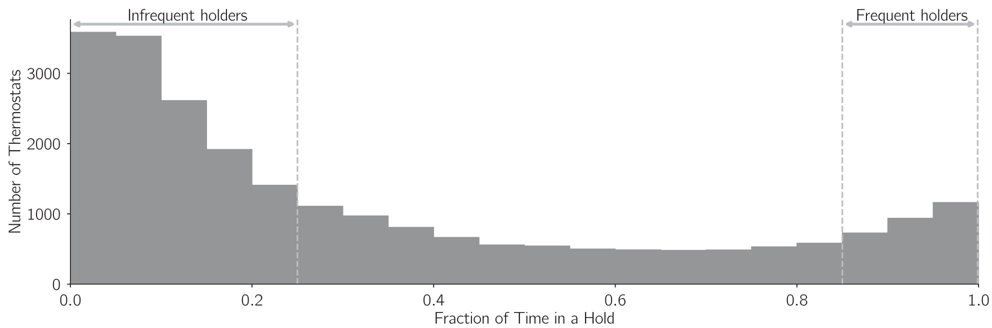

Fig. 3. For each thermostat in our sample S1, the distribution of the fraction of time spent in a hold (n = 23,624). The defined *infrequent* and *frequent holder* groups are shown on the distribution.

override behaviors) to try and understand the factors contributing to the initializing or adjusting of holds.

The ratio of time in a schedule override

While the fraction of the population in a hold at any time (similar to Figure 2 viewed at only a single timestamp) has often been the measurement made by previous studies (Pritoni et al. 2015; Peffer et al. 2011), the cumulative amount of time spent in an override is also important. Time spent in an override begins to illustrate differing user types, based on their preferred control strategy, in a way that the previous observational data did not natively capture. For example, are the number of people in an override at any given time the result of the same individuals putting their thermostat in a hold, or are a large number of people adjusting their thermostat in and out of a hold? In the literature, surveys have previously been employed to determine and separate the general control method employed by users with their thermostats. In the Residential Energy Consumption Survey (Energy Information Administration 2015), users are asked if their heating behavior is to "set one temperature and leave it there most of the time". Unfortunately relying on surveys leaves responses subject to personal interpretation and self-reporting accuracy. Pritoni et al. (2015) observed that 50% of users found to be using a hold had claimed to be using a scheduled program. Given the interval data for each ST, however, it was possible to accurately quantify the time in a hold. For a given thermostat, the fraction of time in a hold was calculated as the ratio of the number of timesteps in a hold over the total number of logged timesteps.

The fraction of time was calculated for all of the thermostats in our sample S1. The resulting distribution of ratios is shown in Figure 3. The median fraction of time in a hold for the sample is 21%. Approximately 12% of users were in a hold 85% of the time or more, and are referred to as *frequent holders*. Meanwhile, 55% of thermostats were in a hold for 25% of the time or less. These users will be referred to as *infrequent holders*.

Frequency of override initialization or adjustment

The frequency of adjustments to a thermostat into or in a hold was of interest as it indicates how active a user is in their thermostats operation. For example, two thermostats which spent 90% of the observation time in a hold are shown in Figure 4. One thermostat, Figure 4(a), was kept at a single heating and cooling setpoint over the observation window; the other (Figure 4(b)) had the setpoints adjusted actively. We classified both thermostat users as *frequent holders*, yet the behaviors imply very different users and control decisions. The motion (bottom of figure) detected by the thermostat and sensors in Figure 4(b) had very repeatable patterns while the observed motion of the first thermostat had a less defined pattern. The user(s) of the second thermostat (Figure 4(b)) appears to have decided to actively actuate the setpoints similarly to a scheduled program. The user(s) in Figure 4(a) may have felt their occupancy was too random or that they never were away long enough to justify the adjustments to their setpoints. The method these two users chose to manage their times in holds would result in very different energy implications (Peffer et al. 2013; Moon and Han 2011).

A new metric was used to quantify the frequency of setting (or changing) a hold each day. The number of initiated or adjusted holds were counted for each thermostat in the population S1. Holds being initialized or adjusted were identified based on changes to the thermostat setpoints during labeled overrides. The holds that lasted only one timestep were not counted as they most likely represented either a transitional state between two setpoints or a mistaken input by the user. For each day and each thermostat, the number of holds was counted and the average of all days taken for a thermostat.

The average daily number of holds being initiated or adjusted is shown as a distribution in Figure 5. The median value of 0.53 holds per day implies that not many adjustments were made across the population each day. The low number of interactions is consistent with the low fraction of time spent in a hold by a majority of the population seen in Figure 3.

Figure 6 shows the same distribution as Figure 5 with only thermostats identified as *frequent holders*. For this group, the median value of holds being set per day is considerably higher at 1.3. This indicates a large portion of the *frequent holders* were regularly adjusting their thermostats and not simply remaining in long and potentially inefficient setpoints.

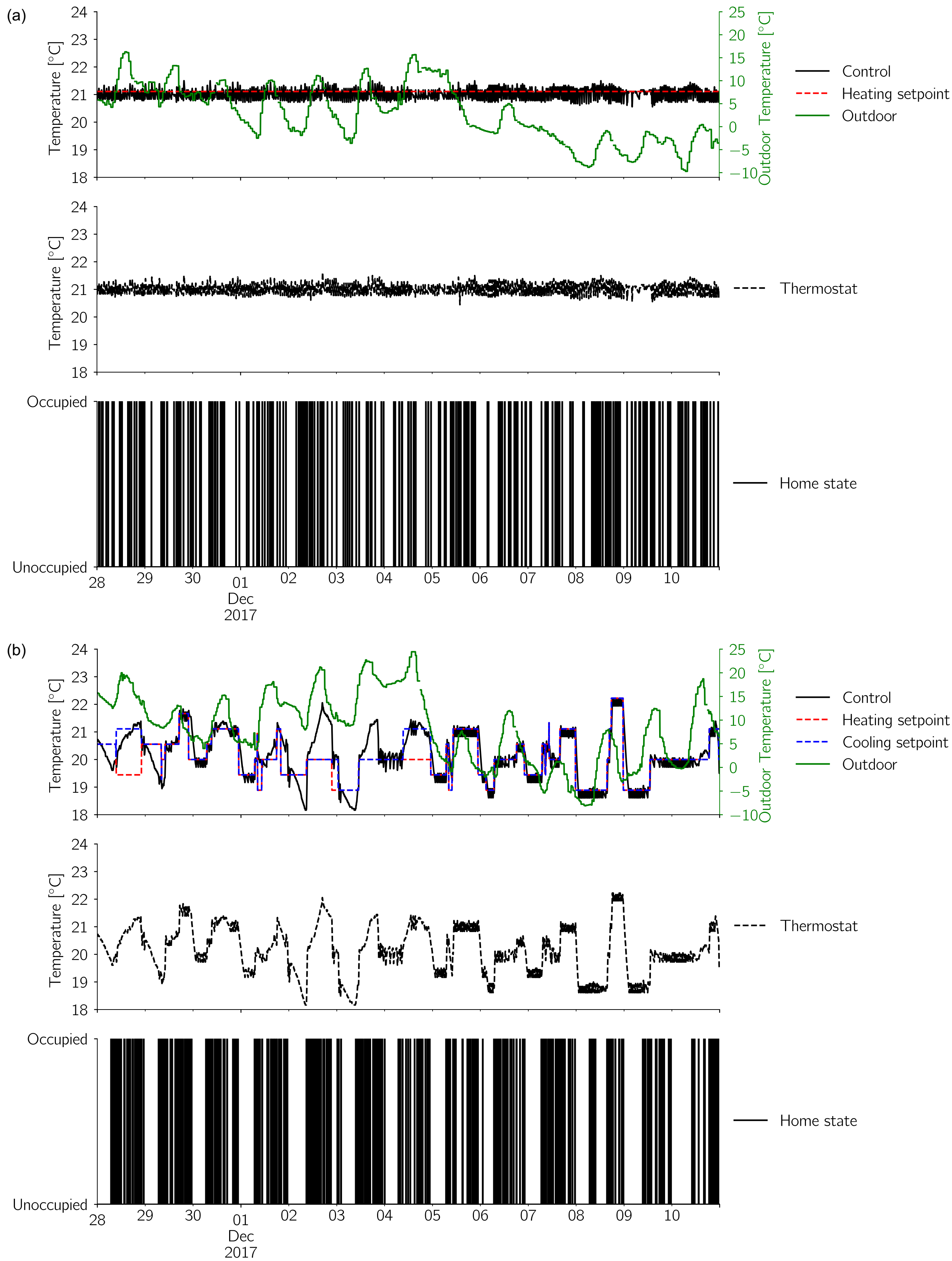

Fig. 4. Two thermostats who both appear in holds more than 90% of the time but with (a) users who do not adjust the setpoints and (b) users who are actively managing their comfort/climate. For each thermostat, indoor and outdoor temperatures (top and middle) plots in addition to the motion state across the home (bottom) are shown.

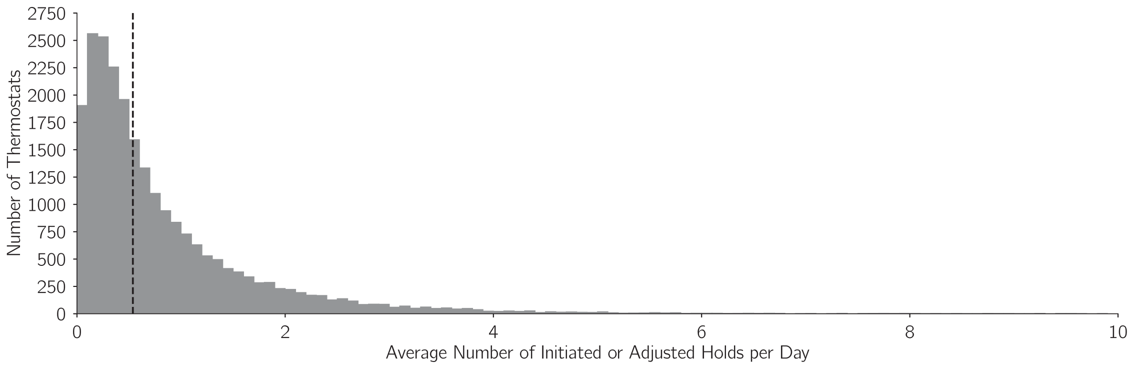

Fig. 5. Daily average number of holds which lasted more than five minutes for the sample S1 (n = 23,624). The median value of 0.53 is indicated by a dashed line.

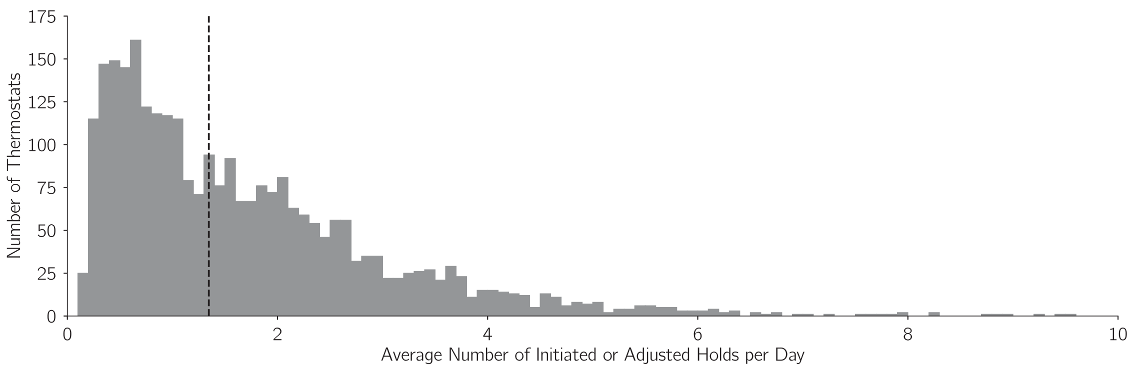

Fig. 6. The average number of holds set per day for individuals who were in a hold more than 0.85 of the time (i.e., *frequent holders*, n = 2,831). The median value of 1.3 is shown with a dashed line.

Contextual factors to override behaviors

The above analyses have dealt with how the schedule overrides were being used by our population of thermostat users, but it remains unclear if the users of these thermostats were consistent in their behaviors for setting the overrides. For example, was a certain and specific set of conditions causing the actions? If users are predictable in their behavior, it would be possible to automate a response from the thermostat, unique to each user(s), that could mitigate potentially energy-intensive behaviors and answer **Q4**. The following sections examine potential patterns in when thermostat schedules were being overridden. We selected three variables to investigate: time of day, time of year, and outdoor temperature conditions.

Thermostats that were predominantly in a hold were not ideal candidates for this part of the study. The behavior of being predominately in a hold indicates the user was utilizing holds as their predominant control method and not as an override to their existing schedule. The analysis was performed only on the *infrequent holders* population (i.e., thermostats who were in a hold for a quarter of the time or less).

Time of day

The first considered variable was the time of day. Figure 7(a) show the fraction of the holds initiated or adjusted by the *infrequent holders* for every hour of the day. Figure 7(b) shows the distributions for four randomly selected thermostats. Figure 7(a) and the sub-figures of Figure 7(b) show the fraction of holds normalized over the 24 hours of a day. For the aggregated group of *infrequent holders* (Figure 7(a)), few holds were initiated early in the morning (i.e., before 07:00). In contrast, most of the overrides appear to be initiated or adjusted in the evening. This pattern of low hold rates in the early hours of the day and higher rates in the evening seems logical — assuming that conscious and present occupants would be setting holds. Unfortunately ground-truth occupancy data was not available to compare against and only certain generations of devices contained in the DYD dataset have embedded PIR sensors to measure motion. Previous analysis had shown relatively stable frequency of overrides across different smart thermostat populations based on time of day (Schellenberg, Lemarchand, and Wein 2017). However, that analysis did not separate *infrequent holders* and was conducted on groups of users who had participated in a utility rebate program.

The consistent and generally expected behaviors of the users are seen less in the individual thermostats (Figure 7(b)). None of these example thermostats appear to follow the trends of the aggregated group for all hours of the day meaning a global model would not translate well to all users. Meanwhile, generating an individual model based on time of

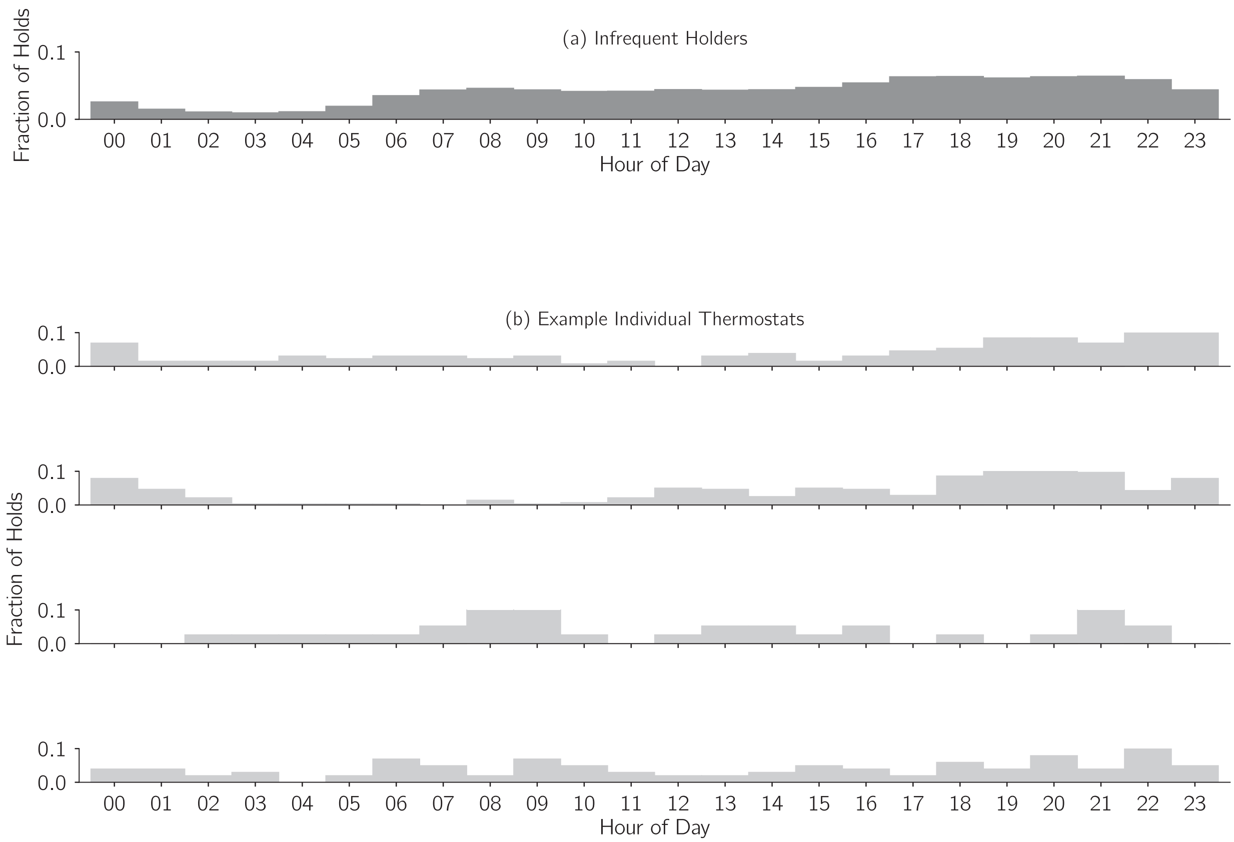

Fig. 7. Fraction of holds initiated or adjusted by the *infrequent holder* population based on hour of day. The top plot (a) shows the distribution for the entire group. The four figures below (b) show examples for four thermostats randomly selected thermostats from the sample. For each plot, the sum of all hours equals one.

day may be informative for some users who have times of day where a hold would be unexpected (e.g., second example from the top of Figure 7(b)), but it would not be useful for other users who have little variation based on the hour of the day.

Time of year

The distinction in behaviors based on the time of year has not been thoroughly investigated by others because of a lack of observational studies over a calendar year or more. Figure 8 shows the fraction of holds initiated or adjusted each month for the *infrequent holder* population. Values were normalized by calendar month for the individual user to avoid discrepancies with users who had more than a calendar year of data. The highest fractions of time in holds appear in typical cooling (June, July, and August) and heating months (December and January). The lowest rates appear in the spring and autumn (April, September, and October). It is postulated that with the more moderate temperatures associated with the shoulder seasons, users are less dependent on mechanical space condition in maintaining acceptable thermal conditions and less prone to initiate holds. During shoulder seasons, systems are also more likely to be turned off and it is not possible to initiate or adjust a hold on these devices.

Outdoor temperature

Outdoor temperature and associated comfort were expected to show some connection to the initiating or adjusting of a thermostat override. Each thermostat in the DYD dataset had outdoor temperature provided from a local weather station. For the thermostats in the *infrequent holder* population, the outdoor temperatures were binned in 2°C increments along with absorbing bins below −15°C and above 35°C. Thermostats were grouped by their state or province to account for similar outdoor conditions and potential preference adaptations. For each thermostat, the frequency in each bin was normalized to prevent infrequent temperature bins (i.e., extreme high and low temperatures) from being underrepresented against more common moderate outdoor temperatures.

The fraction of holds in each temperature bin is shown in Figure 9 for the 10 most frequent state and provinces in the *infrequent holder* population. Each region shares a similar decrease in frequency around the 13-18°C temperature range. The relative increase in frequency on each side would be a response to entering heating and cooling season temperatures. Each of these geographical regions appears to have a steady increase in holds being initiated or adjusted as temperature increase into the low to mid 20's. It is speculated that as temperature increases along with potential discomfort, people are choosing to set more holds.

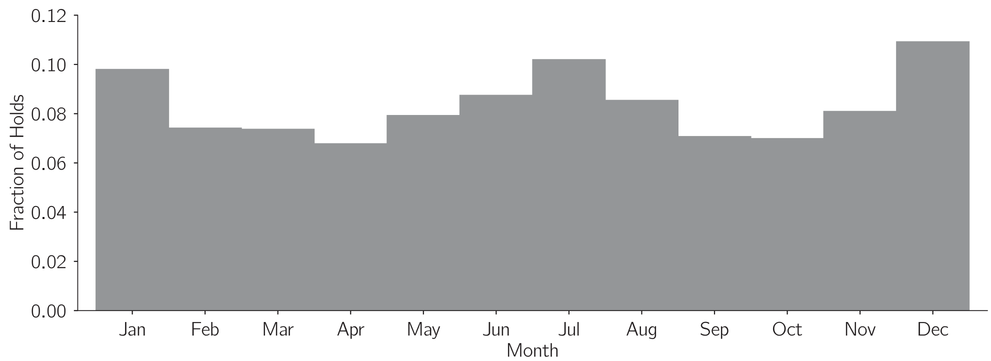

Fig. 8. The frequency at which overrides were set or adjusted for each month of the year for the *infrequent holder* population.

Thermostats in warm climates (Florida, California, and Texas), were overridden as temperatures got below 10°C. Users in California and Florida quickly set increasing numbers of holds as temperatures began to fall. Nearly all holds were initiated before the temperature had even gone below 0°C. In comparison, residents in TX were much slower in their reactions and have increased rates of hold initiation until around −3°C. Users in these three states appear distinct to the other, typically colder, states which could imply different regional expectations and control strategies.

The hold-related behavior patterns explored above highlighted their diverse nature. While some users manage their thermostat with singular prolonged holds, many elected to frequently manually adjust a prolonged hold. For those users who rather rely on utilizing holds infrequently, the contextual information of their actions showed time of day, time of year, and external temperatures to have some effect on aggregate. However, on the individual level the relations appear less definitive.

The impact of overrides

We have managed to show that hold-related behavior of thermostat users are typically complicated and diverse. The question (**Q3**) remains of what the actual impacts of these overrides are in terms of equipment runtimes and, by extension, energy. Two different methods were attempted to quantify impacts. The first method asks if we could counterfactually[1] compute the energy usage of a household if they never initiated a hold, "how much energy savings would they have used?" This counterfactual method captures the absolute effects of the offsets in setpoint temperature, but does not directly translate to effects on runtime; it also was only applicable to *infrequent holders*. The second method, derived from the Energy Star connected thermostat metric (Environmental Protection Environmental Protection Agency 2016), was more universally applicable and estimated the effect of setpoints on equipment runtime against a fictitious personalized static baseline temperature. This approach captures relative changes between user groups but does not capture the real deviation from their own schedules. Both methods are merely estimates of the effects of overrides and actually getting a ground-truth to the energy usage on an individual thermostat level is impossible as it would require testing two situations in a home at a given time.

[1]Defined as what the setpoints would have been had the thermostat not been overridden.

Counterfactual estimation of setpoint behavior

For the group of *infrequent holders*, their logged data had long periods where no holds were in place. Using these available periods, we produced a counterfactual estimate for what the setpoint temperatures may have been had an override not been applied. For each month, and each schedule period (e.g., home, sleep, custom, etc.) a different counterfactual value was calculated. The counterfactual setpoint for each month/schedule combination was assumed to be the most frequent setpoint (both heating and cooling) under that combination but when not in a hold. This assumption is not without potential issues. For example, a behavior change may have occurred that resulted in completely new setpoint preferences from the beginning to the end of the month. However, utilizing smaller windows of time could lead to more periods where a counterfactual baseline may not be available.

The days of data for each thermostat were split into heating and cooling seasons to more easily separate changes to the heating and cooling setpoints. Days were classified using the same approach as the Energy Star connected thermostats specification (Environmental Protection Environmental Protection Agency 2016). The connected thermostat metric classifies days as either in *core heating* or *core cooling*. A day that falls into neither category (e.g., days the thermostat was completely off) is excluded from further analysis. A *core day* is defined as having 30 minutes or more of equipment runtime for the appropriate system and zero minutes of the other runtime. For each thermostat, and for each core day, the difference between the actual heating or cooling setpoint and the counterfactual setpoint at each timestep while a hold was in effect was taken. The median of these temperature differences was taken per thermostat. Figure 10 shows the median temperature difference between actual and counterfactual (i.e., our assumed what would have been scenario) setpoint per thermostat for heating (top) and cooling (bottom) days. The solid colored lines in each figure

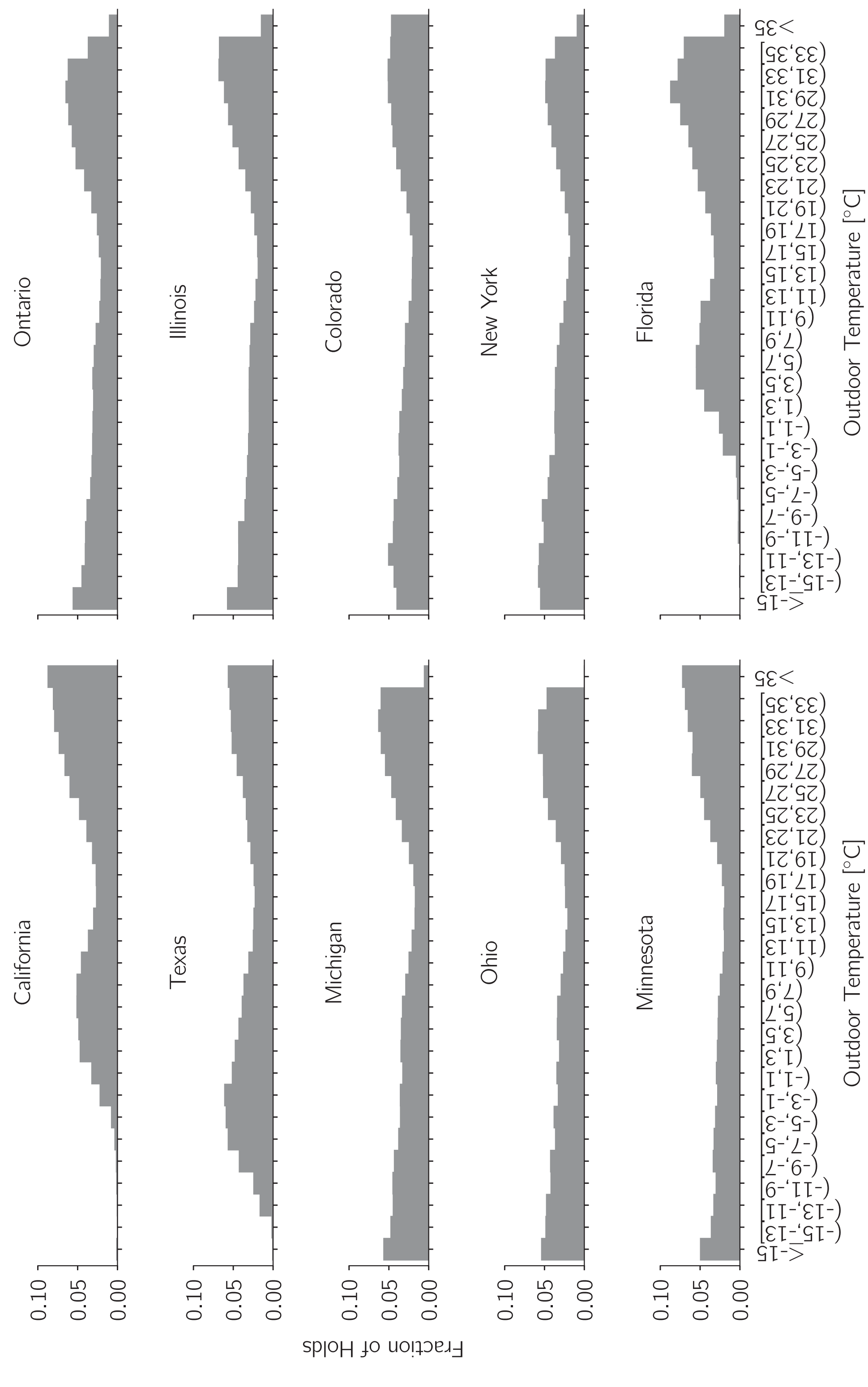

Fig. 9. The fraction of holds initialized at each outdoor temperature bin for the *infrequent holder* population in the 10 most common provinces or states in the sample.

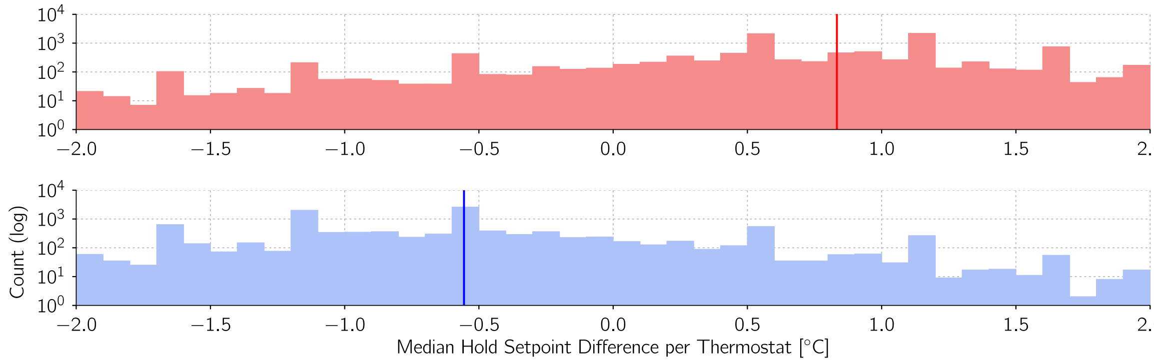

Fig. 10. Distribution of the median relative offsets (i.e., difference between actual and counterfactual value) for heating (top) and cooling (bottom) setpoints for the *infrequent holders*. For each plot the median value is indicated by the solid vertical line.

indicates the median value of the distributions. The spikes are a result of the precision of the thermostat and their location on both plots are a remnant from the conversion of temperatures from a predominantly Fahrenheit-utilizing population to Celsius. The heating and cooling distributions are shifted on either side of zero, with heating being generally positive and cooling generally negative. This indicates predominantly comfort-driven actions in which thermostat settings were adjusted to more energy-intensive setpoints; a similar finding in earlier programmable thermostat studies (Sachs et al. 2012). The median value of the heating setpoint difference is about 0.85 °C while the magnitude of the cooling difference is about 0.55 °C.

To better quantify and relate to the effects of the cumulative holds by the *infrequent holders*, the relative offset based on the counterfactual setpoints was translated into degree-hours per day. The degree-hours were found by summing the setpoint difference of each timestep when a hold was overriding the thermostat schedule and normalizing timesteps to hours by dividing by 12 (i.e., the number of timesteps each hour). The summation of these degree-hours was divided by the number of heating and cooling days for each thermostat respectively. The distributions for the average number of degree-hours offset for both heating and cooling are shown in Figure 11. The median number of degree-hours for both heating (left) and cooling (right) are 0.05 and −0.05 respectively. Both median values were found to be statistically different than zero (p-values < 0.0001). Similar to the offsets reported in Figure 10, the directionality indicates less efficient setpoints being used in both circumstances. Utilizing holds for comfort considerations predominately was similarly observed by Sachs et al. (2012). For comparison, a 3 °C set back for 8 hours a day is expected to save over 5% of heating energy (US Department of Energy 2017). Meanwhile 0.05 degree-hours per day would be an effect of 0.15 °C for 8 hours a day, or a change of only 0.3% to heating energy per year. The effects of these holds are then much smaller on average than even a moderate repeated schedule set back. Assuming that the heating and cooling loads are linearly proportional to the degree-hours (Kissock, Haberl, and Claridge 2002), the energy effects for a single thermostat may be minimal; however, when evaluated across a population, it could still have appreciable negative energy ramifications.

Runtime savings estimation

While the counterfactual approach worked for the *infrequent holders*, it was not appropriate for those thermostats who were predominantly in a hold and whom a counterfactual setpoint could not be easily ascertained. As such, the same counterfactual method could not be used with the *frequent holder* group. To compare the *frequent* and *infrequent holders* required another method.

The Environmental Protection Agency's Energy Star specification for connected thermostats (Environmental Protection Environmental Protection Agency 2016) was designed to assess the runtime savings of connected thermostats (similar to smart thermostats) from different manufacturers. The metric bases savings on how a thermostat is able to be operated with setpoints away from a personalized *comfort temperature*. The *comfort temperatures* are assumed to be what a traditional non-programmable thermostat would have manually been set to control against. The *comfort temperatures* are defined individually for each thermostat as the 90^{th} and 10^{th} percentiles of indoor temperatures on *core heating* and *core cooling* days respectively. The runtime difference between the modeled amount of runtime when keeping the home at the *comfort temperature* and the measured runtime, which was based on all of the overrides, is deemed as savings a smart or connected thermostat was able to provide.

A model was trained for each thermostat and heating and cooling seasons independently. The runtime model calculated the ratio (α) between actual runtime and the thermal demand. As an example in the heating season (denoted by subscript h), for each core heating day d and each hour n the temperature difference between indoor and outdoor temperatures was calculated using Equation 1.

$$\Delta T_{d,n} = T_{indoor_{d,n}} - T_{outdoor_{d,n}} \quad (1)$$

Since the interval data from the DYD dataset was in five-minute intervals, the average of the hourly indoor temperature each hour was taken. Next, the hourly ΔT values were used for calculating the daily heating thermal demand ($DHTD$) for each day d as in Equation 2.

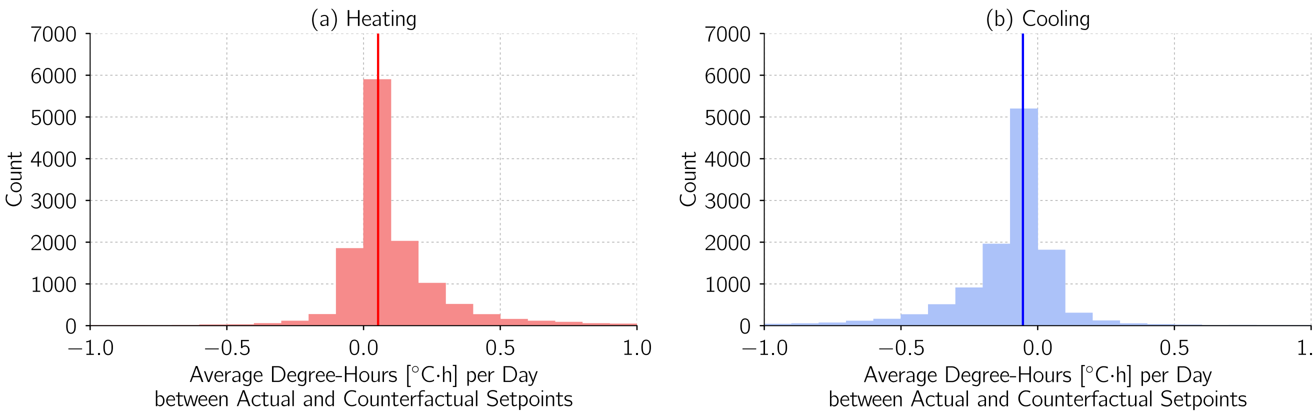

Fig. 11. Distribution of degree-hours per day as a result of the hold s as estimated by a counterfactual estimation on the *infrequent holders* population for (a) heating and (b) cooling seasons. The median value for each season is indicated by the solid line.

$$DHTD_d = \frac{\sum_{n=1}^{24} \max(\Delta T_{d,n} - \tau_h, 0)}{24} \tag{2}$$

The heating balance point (τ_h) value in Equation 2 is the difference between indoor and outdoor temperatures associated with a *DHTD* of zero (i.e., no runtime). It represents the self-heating effects found in homes. Finally, the heating thermal demand was calculated according to Equation 3 in which each heating day d all core heating days (D) are utilized.

$$\alpha_h = \frac{\sum_{d=1}^{D} Runtime_d}{\sum_{d=1}^{D} DHTD_d} \tag{3}$$

The implemented open-source software package (US Department of Energy 2019) found the appropriate τ_h and by extension α_h for each thermostat's model by iteratively minimizing the sum-squared error between actual and predicted runtime. Once fit, the module was able to calculate the baseline temperature deltas ($\Delta T_{base_{d,n}}$) using that thermostat's unique heating *comfort temperature* as opposed to the indoor hourly temperature. With the baseline temperature differences, the baseline *DHTD* values were determined. The base *DHTD* was used to estimate runtime by multiplying by the α_h. A similar methodology was applied to cooling season data which resulted in cooling equivalents to all the calculated values. The software module automatically determined core days for each thermostat based on the runtime. The classification of days was based on the same previously used runtime-based definitions (Counterfactual estimation of setpoint behavior).

We restricted the sample of thermostats to those from S1 that were located in Ontario, Canada. This was done to reduce climate variability and to remove the need to normalize savings results by energy usage in each Building America climate zone as prescribed by the module/standard. From the thermostats in Ontario who possessed only single-stage heating and cooling (a requirement of the metric), 1500 thermostats were sampled; the same as the sample specified by the Environmental Protection Agency. Each thermostat had all of its available data reformatted to be of a form readable by the software module. The software module typically requires a valid American zip code to be able to map to a weather station. The open source code was modified to be able to use the temperature data provided in the DYD dataset. Following the result of the module for each thermostat, thermostats were filtered based on the same screening criteria as specified by the standard. The filtering was based on the determined balance point temperature (τ) being between 0 and 25°C and the cross-validated root mean square error being less than 0.6.

Table 1. Fraction of the Ontario-based population in each behavior group for cooling and heating seasons. Seasons were determined by thresholds on cooling and heating runtime.

Group	Cooling	Heating
infrequent holders	0.53	0.58
diverse holders	0.32	0.31
frequent holders - frequently adjusted	0.09	0.08
frequent holders - rarely adjusted	0.05	0.03

After filtering, the remaining thermostats were split into one of four groups for both heating and cooling seasons based on their users' behaviors. The behavior groups were:

- infrequent holds
- frequent holders - frequently adjusted
- frequent holders - rarely adjusted
- diverse holders

Thermostat users in the *infrequent holders* group had the same definition as previously used. The *infrequent holder* population of thermostat users were the most likely to benefit from any "smart" features enabled and provided by their thermostat. The *frequent holders - frequently adjusted* had an average fraction of time spent in holds greater than or equal to 0.85 while the average number of daily holds being set was greater than 1.5. The *frequent holders - rarely adjusted* were those users who had the same fraction of time in a hold as the *frequent holds - actively adjusted* but who were setting fewer than 1.5 holds per day. Finally, members of the population who were not in any of the other three categories were *diverse holders*.

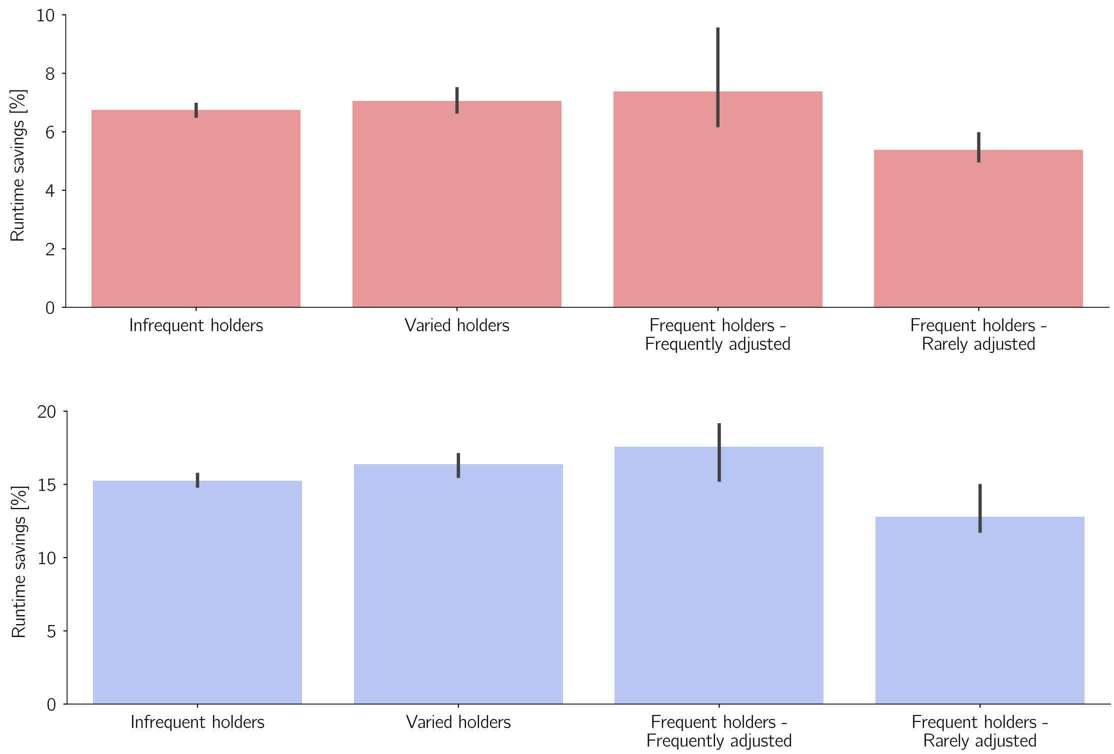

Fig. 12. Median average daily runtime percentage savings with indicated confidence intervals for Ontario-based thermostats in the four specific user groups. Savings are based on a regression analysis and comfort temperature baseline for both (a) heating season and (b) cooling season.

Table 1 breaks down the fraction of the Ontario-based sample based on the four groups and the heating and cooling seasons. Fractions of the population in each group for both seasons remain relatively consistent. The *infrequent holders* have the largest divergence with a five percent change between seasons.

Figure 12(a) shows the median heating runtime percentage savings for the four user groups. Meanwhile, for the cooling season, the same metric and breakdown is shown in Figure 12(b). For each group, the 95% confidence interval of the median is indicated by a line. Unsurprisingly, the *frequent holds - rarely adjusted* group appears to have the lowest savings in both heating and cooling at 5 and 12% respectively. The long, static holds would provide little variation in temperature and therefore little savings relative to the comfort temperature. The magnitude of savings achieved by the *frequent holds - frequently adjusted* groups is surprising. Higher savings from thermostats with frequent overriding patterns had been observed by Schellenberg, Lemarchand, and Wein (2017); however, we found that the users actively managing their thermostats with holds achieved savings at least of the same level as a relatively undisturbed thermostat. Based on average energy expenditures for Ontario residential customers (Natural Resources Canada 2018), the best savers (*frequent holders - frequently adjusted*) are estimated to save over 2000 kWh per year, or about 8% of the average household energy usage in Ontario (Statistics Canada 2015).

It is important to note that these estimates are not necessarily actual achieved savings by the thermostat. An actual savings analysis would require more energy usage data and information regarding previous thermostat installation and configurations. Regardless, the relative ranking of the user types by savings is expected to remain consistent while magnitudes of values could be subject to changes should a more formal measurement and verification process take place.

Based on the comfort temperatures, the average daily runtime savings potential compared to their calculated comfort baselines on both heating and cooling days are estimated by the software. Figure 13(a) and Figure 13(b) show the median runtime savings for heating and cooling respectively along with 95% confidence intervals of the median values. Heating runtime savings are about one-third of the value of cooling savings which, for the highest savers, is savings of about 20 minutes per day.

Mitigation strategies

The previous sections have illustrated that even with recent ST technology, users still utilize holds in multiple ways and that there are some consequences to their utilization. How then might these overrides be better mitigated or even prevented?

Mitigating the use of holds by users of these thermostats may be approached differently for each identified user type used the Runtime savings estimation section. It was found that users who manage their thermostat with infrequent adjustment and a near-continuous hold have reduced energy savings potential compared to the other groups by at least

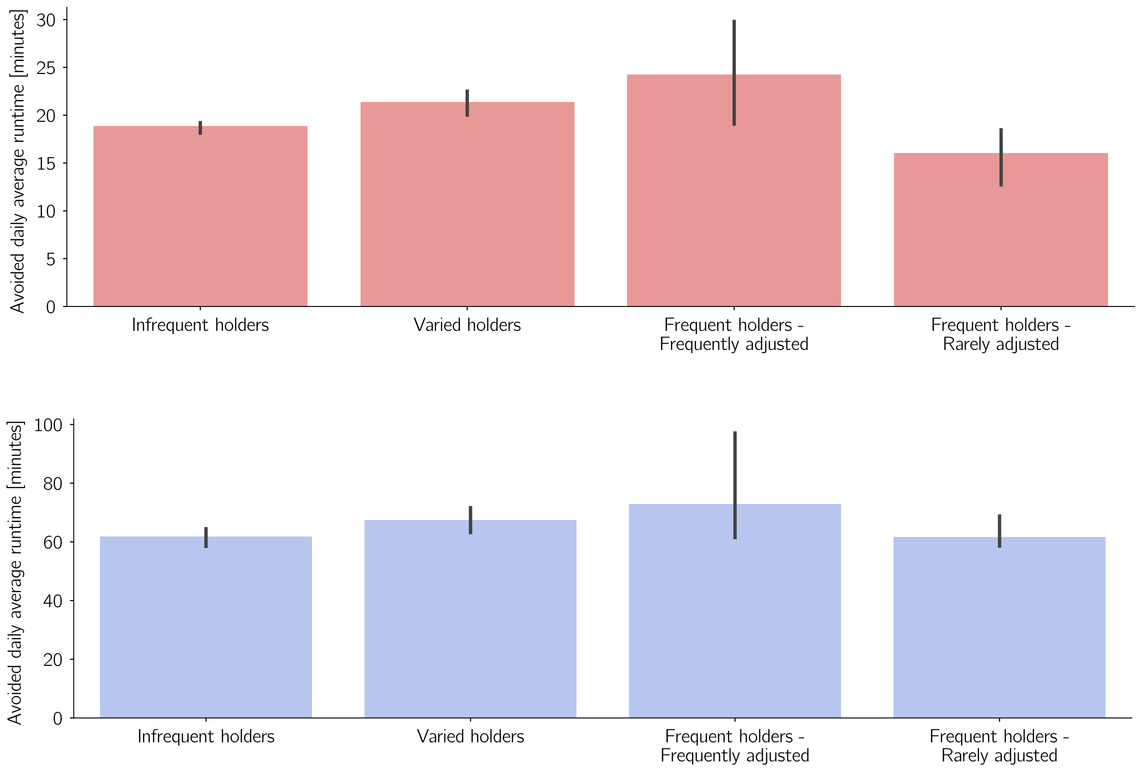

Fig. 13. Median average daily runtime savings values with indicated confidence intervals for Ontario-based thermostats in the four specific user groups based on a regression analysis and comfort temperature baseline for both (a) heating season and (b) cooling season.

2%. If this group was to be targeted by the manufacturer with methods to increase their energy savings, the easiest solution would be to change the default action from indefinite holds to something shorter in duration. One solution could be the thermostat providing users a selection of behavior at the time of initializing a hold; an option of which may include keeping indefinite duration. Ensuring that the indefinite duration was a deliberate selection would undoubtedly reduce the number of individuals in that state. How best to appropriately message or present these options remains an open question of investigation. For those individuals in the *frequent holders - frequently adjusted* grouping, both their behaviors and relative energy savings are hard to discourage as they are effectively managing their devices. However, they would still potentially benefit from being similarly prompted and provided an option at the time of hold initialization. One method to encourage manufacturers to adopt a method of mitigating holds would be through codes or standard requirements. A similar approach has been taken with lighting in the Canadian National Energy Code for Buildings which requires lights to turn off after 20 minutes of vacancy (National Research Council Canada 2011). Unfortunately, relying solely on a similar occupancy-based method in homes for removing overrides is risky in residential applications. Occupancy in homes can be difficult to accurately know because of the range of activities happening in a home and low sensor density.

For the two remaining groups (i.e., *infrequent* and *diverse holder* populations), adjustment to the user experience would be expected to reduce some of the time in a hold and potentially increase savings. Alternatively, if a hold could be preempted and avoided by the thermostat, additional savings may also be achieved. Inspection of Figures 7–9 suggests some predictable patterns exist. Unfortunately, in our attempts to build predictors of a hold event utilizing available temperature and motion data, the methods failed to achieve a level of prediction accuracy that was considered actionable. This could be a result of unpredictable exogenous events, missing measurements of environmental factors driving the action, or the difficulty in developing a single model for all occupants in a home.

Discussion of limitations

The usage practices of thermostat users are known to vary based on their preferences (Nevius and Pigg 2000; Kelsven, Weber, and Urbatsch 2016) but are also based on usability of the devices itself (Meier et al. 2011). Inherently the override behavior exhibited in this sample from a single manufacturer will likely not reflect the entire smart thermostat population. With any similar smart thermostat which is installed by a similar group of users, there is expected to be a similar variety of user behaviors (i.e., *infrequent holders, diverse holder, frequent holders-rarely adjusted, and frequent holders-frequently adjusted*) though the fractions of the population for that device may be different. For example, it has been observed that rates of overrides differed between different smart thermostat manufacturers (Apex Analytics LLC 2016; Schellenberg, Lemarchand, and Wein 2017). As for the energy impacts of these various groups, these would be expected to be similar for the same user types in other thermostats though the overall savings of a devices (i.e., all groups inclusive) would be affected by the size of each group.

Conclusions and future work

Understanding how people operate the thermostats in their homes has been a goal of researchers for decades. The way by which they utilize a setpoint schedule and potentially override it has only been shallowly quantified given data and collection methods available to previous studies. Furthermore, there is widespread belief that user overrides in the form of holds are a major problem and a leading cause of underwhelming programmable thermostat performance. Now, given the latest generation of smart thermostats (STs), continuous longitudinal data for thermostat usage is available to researchers. Hence, our objective in this article was to analyze this longitudinal data to better dissect the complex override behaviors of users and to assess the relative energy impact of these behaviors.

We analyzed an ST population of over 20,000 devices from the same manufacturer. First, we showed these ST users have similar levels of time spent in holds compared to previously observed behavior for programmable thermostats. However, a deeper analysis showed thermostats were overridden for less time than those previous studies may have indicated. In observing the potential differing user types, some ST users did choose to "set and forget" and remain consistently at a fixed thermostat setting, which resulted in approximately 2%-4% less savings in heating and cooling than the other user behavior groups. A considerable fraction of the population frequently set multiple holds per day — in essence manually actuating what an automatic schedule may have done. The effects of holds on runtime savings showed the behavior of this population of users had minimal impact on energy usage and actually achieved savings that were similar or better than the *infrequent* or *varied* population of hold ers (i.e., individuals who mostly left their thermostat operating on a schedule).

With this behavioral analysis and its energy impact complete, we next considered recommendations to mitigate the potentially negative energy impact of user holds. Since accurately predicting individual user hold events from observable data remains an unsolved problem, this makes it difficult to consider an effective strategy based on preempting unreliable predictions of user holds. Thus, given the observed population distributions for different hold ing behaviors and their relative energy impacts, simple modifications to the user interface to discourage energy-inefficient holds (or decrease their duration) would appear to be the simplest and most effective mitigation strategy. Any other strategies may only target a subset of the user population with substantially less broad impact on energy savings. Finally, one trivial strategy would simply be to take no action to mitigate user holds — this could be warranted given our overall results suggesting user hold behavior is a relatively minor source of ST energy efficiency.

In future work, several additional questions could be considered given additional ST data. For example, the question of duration of a hold was not considered because the programmed default duration was not known. Alternatively, in a more structured study, pairing hold data with the ability for users to say why they had set the override would give deeper insights to the motivations as to whether they were purely discomfort-driven or energy motivated; this information in turn could potentially could help with prediction and mitigation efforts. Finally, this analysis only looked at a family of STs from a single manufacturer. Similar data from other manufacturers who had potential different strategies around overrides may highlight the most (or least) effective approaches at dealing with overrides.

Acknowledgments

Two of the authors are active participants of International Energy Agency Energy in Buildings and Communities (IEA-EBC) Programme Annex 79 and wish to thank their colleagues for their inspiration.

Funding

This work is funded by a research grant provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) and ecobee Inc.

ORCID

Brent Huchuk http://orcid.org/0000-0002-7720-7830
William O'brien http://orcid.org/0000-0002-0236-5383
Scott Sanner http://orcid.org/0000-0001-7984-8394

References

Apex Analytics LLC. 2016. Energy Trust of Oregon Smart Thermostat Pilot Evaluation. Technical Report 303. http://assets.energytrust.org/api/assets/reports/Smart_Thermostat_Pilot_Evaluation-Final_wSR.pdf.

Ecobee Inc. 2018. Donate Your Data. https://www.ecobee.com/donateyourdata/.

Energy Information Administration. 2015. Residential Energy Consumption Survey (RECS). Washington, DC.

Huchuk, B., W. O'Brien, and S. Sanner. 2018. A longitudinal study of thermostat behaviors based on climate, seasonal, and energy price considerations using connected thermostat data. *Building and Environment* 139:199–210. doi:10.1016/j.buildenv.2018.05.003

Huchuk, B., S. Sanner, and W. O'Brien. 2019. Comparison of machine learning models for occupancy prediction in residential buildings using connected thermostat data. *Building and Environment* 160: 106177. doi:10.1016/j.buildenv.2019.106177

International Code Council. 2015. International Energy Conservation Code - Residential Provisions. Technical Report.
Kelsven, P., R., Weber, and E. Urbatsch. 2016. Nest Learning Thermostat Pilot Program Savings Assessment Bonneville Power Administration & Franklin Public Utility District. Technical Report.
Kissock, J. K., J. S. Haberl, and D. E. Claridge. 2002. Development of a Toolkit for Calculating Linear, Change-point Linear, and Multiple-Linear Inverse Building Energy Analysis Models. Technical Report. ASHRAE.
Lopes, J. S., and P. Agnew. 2010. FPL residential thermostat load control pilot project evaluation. *In ACEEE Summer Study on Energy Efficiency in Buildings* 2:185–92.
Meier, A., C. Aragon, T. Peffer, D. Perry, and M. Pritoni. 2011. Usability of residential thermostats: Preliminary investigations. *Building and Environment* 46 (10):1891–8. http://www.sciencedirect.com/science/article/pii/S0360132311000874. doi:10.1016/j.buildenv.2011.03.009
Moon, J. W., and S. H. Han. 2011. Thermostat strategies impact on energy consumption in residential buildings. *Energy and Buildings* 43 (2-3):338–46. doi:10.1016/j.enbuild.2010.09.024
National Research Council Canada2011. National Energy Code of Canada for Buildings.
Natural Resources Canada2015. Home heating and cooling equipment.https://www.nrcan.gc.ca/energy/products/reference/publications/14498.
Natural Resources Canada2018. National Energy Use Database. Accessed September 21, 2019. http://oee.nrcan.gc.ca/corporate/statistics/neud/dpa/data_e/databases.cfm.
Nevius, M. J., and S. Pigg. 2000. Programmable thermostats that go berserk? Taking a social perspective on space heating in Wisconsin. *Proceedings ACEEE Summer Study on Energy Efficiency in Buildings* 8: 233–244. https://www.aceee.org/files/proceedings/2000/data/papers/SS00_Panel8_Paper19.pdf.
Parks Associates. 2017. More Than 100 Million U.S. Households Do Not Have a Smart Home Device.
Peffer, T., D. Perry, M. Pritoni, C. Aragon, and A. Meier. 2013. Facilitating energy savings with programmable thermostats: Evaluation and guidelines for the thermostat user interface. *Ergonomics* 56 (3):463–79. doi:10.1080/00140139.2012.718370
Peffer, T., M. Pritoni, A. Meier, C. Aragon, and D. Perry. 2011. How people use thermostats in homes: A review. *Building and Environment* 46 (12):2529–41. doi:10.1016/j.buildenv.2011.06.002
Pritoni, M., A. K. Meier, C. Aragon, D. Perry, and T. Peffer. 2015. Energy efficiency and the misuse of programmable thermostats: The effectiveness of crowdsourcing for understanding household behavior. *Energy Research & Social Science* 8:190–7. doi:10.1016/j.erss.2015.06.002
Sachs, O., V. Tiefenbeck, C. Duvier, A. Qin, K. Cheney, C. Akers, and K. Roth. 2012. *Field Evaluation of Programmable Thermostats*. Technical Report December.
Schellenberg, J., A., Lemarchan, and A. Wein. 2017. Xcel Energy Colorado Smart Thermostat Pilot – Evaluation Report. Technical Report. Nexant Inc.
Statistics Canada. 2015. “Household energy consumption, Canada and provinces.” https://www150.statcan.gc.ca.
Touchie, M. F., and J. A. Siegel. 2018. Residential HVAC runtime from smart thermostats: Characterization, comparison, and impacts. *Indoor Air* 28 (6):905–15. http://doi.wiley.com/10.1111/ina.12496. doi:10.1111/ina.12496
US Department of Energy. 2019. Connected Thermostat draft field savings metrics module.https://github.com/EPAENERGYSTAR/epathermostat.
Environmental Protection Environmental Protection Agency. 2016. “Method to Demonstrate Field Savings Version 1.0 (rev. Dec-2016).” https://www.energystar.govhttps://www.energystar.gov/sites/default/files/Version1.0Methodto DemonstrateFieldSavingsofENERGYSTARConnectedThermostats.pdf.
US Department of Energy. 2017. Thermostats. https://www.energy.gov/energysaver/thermostats.