Energy & Buildings 249 (2021) 111201

Contents lists available at ScienceDirect

Energy & Buildings

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

Development and evaluation of data-driven controls for residential smart thermostats

Brent Huchuk [a,*], Scott Sanner [a], William O'Brien [b]

[a] Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Canada
[b] Department of Civil and Environmental Engineering, Carleton University, Ottawa, Canada

ARTICLE INFO

Article history:
Received 13 February 2021
Revised 11 May 2021
Accepted 14 June 2021
Available online 19 June 2021

Keywords:
Residential building
Model predictive control
Reinforcement learning
Smart thermostat

ABSTRACT

The advent of smart thermostats with real-time sensing raises the question of how to preemptively control heating, ventilation, and air conditioning (HVAC) systems to minimize energy usage while maintaining occupant comfort. To this end, we empirically compare a standard reactive deadband control to two new smart thermostat HVAC control methods: (1) a model-free reinforcement learning (RL) approach and (2) a novel model predictive control (MPC) method, whose solution is optimal with respect to its data-driven linear model. We evaluated the controls with 500 unique energy models of houses located in the United States. The models were modified to facilitate the short-term performance simulation required for residential HVAC systems. Overall, we found the MPC controller offers three distinct advantages over the RL and deadband methods: (1) MPC had the lowest average cost (defined as a custom weighted combination of runtime and comfort) of the evaluated controllers; (2) the MPC control's linear model was able to reliably extrapolate from the sparse sample of training observations, thus enabling it to adapt quickly to recent data; and (3) in contrast to RL methods, MPC did not subject the houses or occupants to the discomfort of system exploration.

1. Introduction

Approximately a tenth of North America's total energy usage is for space conditioning of residential buildings [1–3]. Despite the magnitude of this energy expenditure, and the associated financial and environmental costs, residential buildings are typically managed by reactive and heuristic-driven thermostats [4]. Even the most advanced smart thermostats tend not to meet the definition of an advanced building controller; i.e., an advanced controller should balance multiple and often conflicting objectives (e.g., lower costs or reducing emissions while improving comfort), adapt controls to environmental conditions, and include the occupants as part of the decision [5].

The development of advanced controls for residential buildings has generally been either model-based, through paradigms such as model predictive control (MPC) [6–9], or model-free reinforcement learning (RL) methods (e.g., Q-learning) [10–16]. Model-free RL methods do not explicitly have a model for state transition and instead learn their policy (i.e., how to operate) directly through online exploration or experience replay in cached data [17]. Some controllers have relied on a combination of model-based and model-free methods by first building a data-driven model and used the trained model for exploration with the model-free method instead of using the actual system [15,18].

Regardless of the method, previous development and evaluation of control methods have had limitations. Prior work has generally looked at optimizing specific buildings under tightly controlled conditions. However, to be a viable solution for mass adoption, we posit that a thermostat control method should be able to leverage the sensing data for a specific building to automatically learn and optimize its control. Furthermore, beyond the issue of automation, the fact that past evaluations have focused on only a few buildings or spaces [6,13,14,15,19] leads to an incomplete picture of how well these approaches may generalize (and how robust they will be) across a variety of building types and climate conditions. The importance of testing control strategies on a large and diverse population was illustrated by Pang et al. [20]. They found large variation in the effectiveness of occupant-centric control strategies across the United States. While their study did simulate the variation in location, foundation type, heating system, and building energy codes, they did not focus on optimizing methods for each individual house, nor were their models adjusted to

* Corresponding author.
E-mail address: brent.huchuk@mail.utoronto.ca (B. Huchuk).

https://doi.org/10.1016/j.enbuild.2021.111201
0378-7788/

reflect how residential heating, ventilation, and air conditioning (HVAC) equipment is operated (Section 2.1.2).

With the latest smart residential thermostats, individual houses now possess large troves of high-quality historical data that captures the indoor and outdoor temperature, equipment runtime, and users' defined preferences. Through this historical data, it is possible to build customized solutions for individual houses. In this paper, we seek to empirically compare a model-free RL approach and a novel model predictive control method (optimal w.r.t. its linear model) for smart residential thermostats and compare them to standard reactive deadband controls. This paper makes the following major contributions:

(i) a demonstration of constructing and modifying a large and diverse sample of energy models that are representative of houses in the United States for smart thermostat control research;
(ii) a novel optimal model-based MPC approach is developed using a mixed integer linear program solver to globally optimize thermostat controls over a finite horizon w.r.t. a learned linear deterministic model of thermal dynamics;
(iii) a direct comparison of deadband, model-based MPC, and model-free RL controls trained using typical smart thermostat data; and
(iv) a discussion of the challenges in applying a model-free RL controller in comparison to a MPC approach.

The remainder of the paper is as follows. In Section 2 we provide an overview of how our testbed was built and how we defined our control methodologies and cost function. Section 3 and Section 4 presents the results and associated discussion of the controllers and where both model-free RL and the MPC method succeeded and where the model-free RL approach failed. In addition, we provide a discussion on how the control methods could be implemented in the field and what limitations of our analysis would need to be addressed. Finally, Section 5 summarizes our findings on testing and developing data-driven control strategies with existing smart thermostat data and discusses the challenges of using model-free RL methods as part of a commercially available thermostat controller.

2. Methodology

Our testing methodology required us to first construct a testbed consisting of representative building energy models (Section 2.1). Next, we selected and formalized a series of control methodologies (Section 2.2); namely, our baseline deadband controller, our model predictive controller, and tabular Q with batch experience replay (model-free RL).

2.1. Building simulation testbed

We constructed a testbed to simulate the reaction of a building to the proposed HVAC controllers. In addition to seeing how the control strategy affected the comfort of the occupants and HVAC runtime of a single residence, we wanted to know how the controllers would perform on a wide variety of houses (i.e., different locations, geometries, sizes, orientations, equipment types, window areas, envelope constructions, etc.). We built a large collection of energy models that are representative of single-family building stock in the United States (Section 2.1.1) and modified the simulations to accurately reflect the temperature response of the building and the on/off actuation of residential HVAC equipment at five-minute resolution (Section 2.1.2). The on/off actuation was particularly important to be able to control individual equipment cycles.

2.1.1. Building sample

We used ResStock[1] to generate a unique and diverse collection of single-family houses located across the United States. We selected ResStock because of the large number of characteristics from various data sources that are sampled for each generated model. ResStock samples based on conditional probabilities of nearly 100 housing characteristics and over 1400 total options [21]. The characteristics include building elements such as location, floor area, window area, envelope construction, vintage, and HVAC type in addition to variables such as the number of occupants and their schedules, appliance types and schedules, and thermostat setpoints. While these features lead to models that are more diverse than other archetype models, the generated models are still a simplification of the complex interactions between humans and buildings as well as the diversity of housing envelopes and HVAC system properties.

The distributions for the different building characteristics come from numerous sources, including the Residential Energy Consumption Survey, field studies, and census data [21]. Based on the sampled characteristics, ResStock constructs OpenStudio[2] models and their associated EnergyPlus[3] models for each house in the sample.

We modified the housing characteristic sampling to ensure that houses had at least a centralized heating or cooling system, otherwise no conditional probabilities were adjusted. This modification means some houses may only have a single (i.e., heating or cooling) system and may be uncontrollable during certain times of year. Having houses with uncontrollable periods during the year will reduce the amount of improvement achievable with any proposed control strategy, but it accurately reflects the realities of applying controls broadly to any house with a smart thermostat.

We constructed 500 distinct energy models using ResStock. All house models used single-stage equipment and the models used five-minute timesteps for simulation; otherwise, each model had a unique combination of features. An example of two different geometries of houses generated are shown in Fig. 1. The two sample houses (Fig. 1a and 1b) are seen to have different sizes, layouts, and foundations.

The housing generation process first sampled the region for each model and then sampled cities from each region based on the available weather data. Fig. 2 shows the number of models in each location across the United States. The size of the bubble indicates the number of houses at each location, while the color of the bubble indicates the associated Building America Climate Zone [22].

Fig. 3 shows the number of models generated for each vintage (age) of the house. The number of models of each vintage are further broken down by the floor area of the house. It is apparent that more modern houses tend to be larger. The average house in the sample was built in the 1980s and the average house is 2100 ft^2 (195 m^2). The newest houses in the sample were built in 2009. Provided new and updated distributions from which to sample, the houses could be resampled to include the years 2010–2019 in future investigations.

2.1.2. Energy model modification

While ResStock provided a collection of unique houses to test the controls on, the models needed to be further refined for our intended control application. The refinements addressed how the energy from the HVAC equipment was provided to the thermal zones by EnergyPlus and how the indoor air temperature responded to the internal loads.

[1] https://github.com/NREL/resstock.
[2] https://www.openstudio.net.
[3] https://energyplus.net.

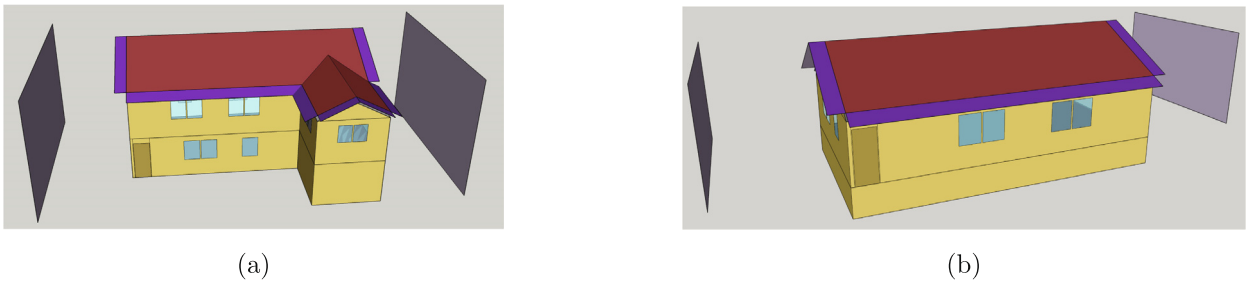

Fig. 1. Example of two randomly sampled houses from the generated sample of 500 models.

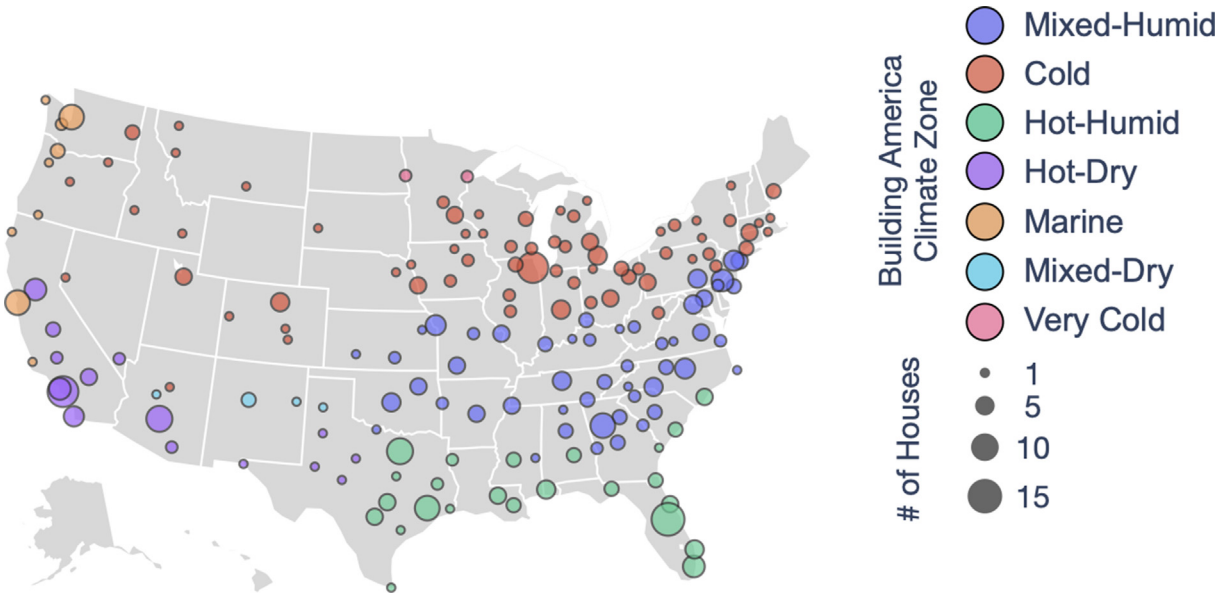

Fig. 2. The number of models in each location for the generated energy models. The associated Building America Climate Zone is indicated for each location bubble.

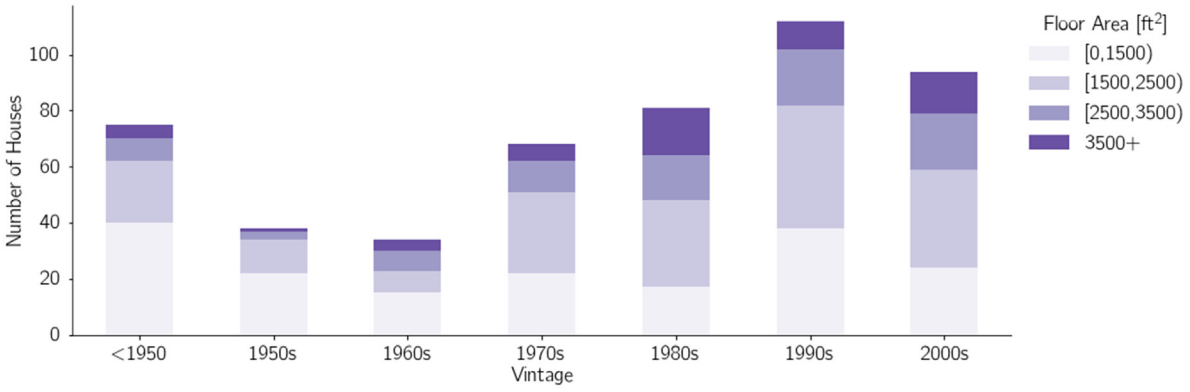

Fig. 3. Number of modeled houses broken down by vintage and by floor areas.

North American residential HVAC equipment typically does not operate with an adjustable thermal energy rate, but rather operates using binary (i.e., on/off) control. When the equipment is turned on, it provides its maximum available capacity. Multi-stage equipment is similar; however, the stages have a fixed portion of the total available capacity. In contrast, EnergyPlus calculates and delivers the energy from the equipment to the zone to precisely meet the setpoint. The equipment in each model built by ResStock was sized using ANSI/ACCA Manuals J and S [23,24]; meaning the equipment is appropriately sized for the house size and location. We suspect our sample of houses will deviate from the true population as real houses may have improperly sized or improperly running equipment. This variation may change the magnitude of the achievable results presented later. The HVAC equipment of the models consisted of an on/off fan and a single speed direct expansion (DX) cooling coil (when the house had cooling). For heating, the equipment was either a DX heating coil with heat pumps, or a furnace used with various heating fuel types.

Fig. 4 illustrates the default behavior of an EnergyPlus model over eight days of simulation. The figure shows the binary mapping of delivered HVAC thermal energy to the zone as on/off (top), the delivered HVAC thermal energy (middle), and the indoor air temperature response (bottom) of the building. It is observed that the EnergyPlus simulation calculates and provides just enough thermal energy to maintain the setpoint. The precise energy calculation and delivery results in the HVAC equipment nearly always being on. This HVAC behavior is unrealistic in comparison to how real houses are controlled (example Fig. 5). A methodology similar to that of Cetin et al. [25] was implemented for each of the 500 thermal models. At the beginning of each timestep (and before the predictor in the simulation), the next control decision was made based on the indoor air temperature from the previous timestep and the current timestep's outdoor air temperature. Based on the control decision, the heating and cooling setpoints for the thermal zone were overridden to be extremely high if heating was required, extremely low if cooling was required, or left unaltered if neither were required. The different controllers (Section 2.2) determined which system (i.e., heating or cooling) to activate differently. The adjustments to the setpoints forced the equipment into a binary on/off state with the maximum available capacity. The control modifications were implemented using the Python interface in EnergyPlus; an option available as of version 9.3.

Once the HVAC system was adapted to mimic the binary on/off nature of real residential systems, a secondary concern was the temperature response of the energy models when only using the full HVAC capacity. Fig. 5 shows the on/off cycles (top), the measured equipment runtime (middle), and the indoor temperature measurements (bottom) from a real thermostat in the Donate Your Data dataset [26]. It is visible in Fig. 5 that the indoor air temperature with a real thermostat is never stable at the setpoint – unlike in the unmodified EnergyPlus model (Fig. 4). We determined that the models from ResStock did not account for enough thermal mass of furnishings which caused rapid and dramatic temperature changes to the indoor air when the equipment was running; unlike what is seen in Fig. 5. To be able to better approximate a real building, the temperature capacity multiplier of each model (i.e., the `ZoneCapacitanceMultiplier:ResearchSpecial` field in EnergyPlus) was increased. We tested values ranging from one to 10; ultimately selecting a value of five. Based on our observations (e.g., Fig. 7) the selection of five as the multiplier provided multi-timestep equipment cycle lengths of a similar length to real thermostats (approximately 15 min on average) and the average temperature change of around 0.5°C for a cycle when maintaining a static setpoint.

2.2. Control methods

As previously mentioned, we sought to evaluate different methods of data-driven control; specifically, a model-based and model-free method. We developed and evaluated a model predictive control (Section 2.2.2) and the tabular Q method with batch experience replay (Section 2.2.3). The data-driven control methods were compared to themselves and to a baseline deadband controller (Section 2.2.1). All control methods were controlled at the same five-minute interval as the energy simulation timesteps. In reality, HVAC equipment in a residential building is able to actuate faster than five minutes, but with minimum on and off times of the equipment, few actions actually occur much faster than five minutes.

2.2.1. Deadband control

Deadband control provided multiple functions in our analysis. The deadband controls were the baseline control method for evaluation, fallback operation in other controllers, and the method used when building the training datasets for the data-driven controllers. Deadband control is a standard control method for equipment with on/off actuation. Even with commercially available smart thermostats, deadband control remains a standard method to control a residential building's HVAC equipment. Deadband controls are able to keep the air temperature close to the desired temperature setpoint while also reducing how often the equipment is turning on/off. Reducing the on/off cycling helps to increase the longevity of the equipment. A deadband controller has an inherent trade-off between comfort and runtime. The controller lets the indoor air temperature move outside of the temperature preference band (i.e., the range of indoor temperatures between the scheduled temperature setpoints) before engaging the HVAC equipment by the magnitude of the deadband.

We implemented a controller with a single-sided deadband. This is similar to the deadband strategy found on ecobee thermostats. For heating, when indoor air temperatures were 0.5°C or more below the heating setpoint the heating system was turned on. Once the air temperature went above the heating setpoint, the equipment was turned off. The cooling system used similar logic except the equipment turned on when the indoor temperature went 0.5°C above the cooling setpoint and turned off once the temperature had dropped below the cooling setpoint. For those houses with air conditioning, we included a 10°C outdoor temperature lockout; preventing the air conditioning from being used and protecting the compressor. The outdoor temperature at which the air conditioner (or any compressor) is locked out varies based on the smart thermostat manufacturer and can typically be further adjusted by the user. We ran each energy model using the deadband control, implemented in Python and connected to EnergyPlus, for an entire year of simulation.

2.2.2. Data-driven model predictive control

Model predictive controls (MPC) are a popular methodology for controlling complex buildings or systems when models of the dynamics are available or can be learned. It has previously been demonstrated that accurate thermal models can be easily *learned* using the data available from smart thermostats [27]. Using the historical data for each house under deadband control (Section 2.2.1), we trained customized thermal models for each house. A new model was trained and available on the 1^{st} and 15^{th} of each month using the previous two weeks of historical data. The latest model was used until a new model was made available. Based on the findings of Huchuk et al. [27], the customized thermal model selected was a linear ridge regression model; as shown in Eq. (1). The model used indoor and outdoor temperatures (T_{in}, T_{out}), the binary on/off value for heating and cooling (q_{heat}, q_{cool}), and direct horizontal solar irradiance (q_{solar}). All values were provided at the current timestep t and with history ($t-m$) where $m \in \{1,2,3,4\}$. The ridge regression model was implemented in Scikit-learn [28] and used threefold cross validation to select the strength of the ℓ_2 regularization (λ) for each model. Values of $\lambda \in \left\{10^{-3}, 10^{-2}, 10^{-1}, 1\right\}$ were selected during training based on the lowest mean squared error.

With the following trained linear model

$$T_{in,t+1} = \sum_{m=0}^{4}\left[\beta_{1,m}T_{in,t-m} + \beta_{2,m}T_{out,t-m} + \beta_{3,m}q_{heat,t-m} + \beta_{4,m}q_{cool,t-m} + \beta_{5,m}q_{solar,t-m}\right], \quad (1)$$

a mixed integer linear problem (MILP) was formulated in Gurobi[4] and solved at every timestep of the yearly simulation for each model. The MILP, defined as follows

[4] https://www.gurobi.com.

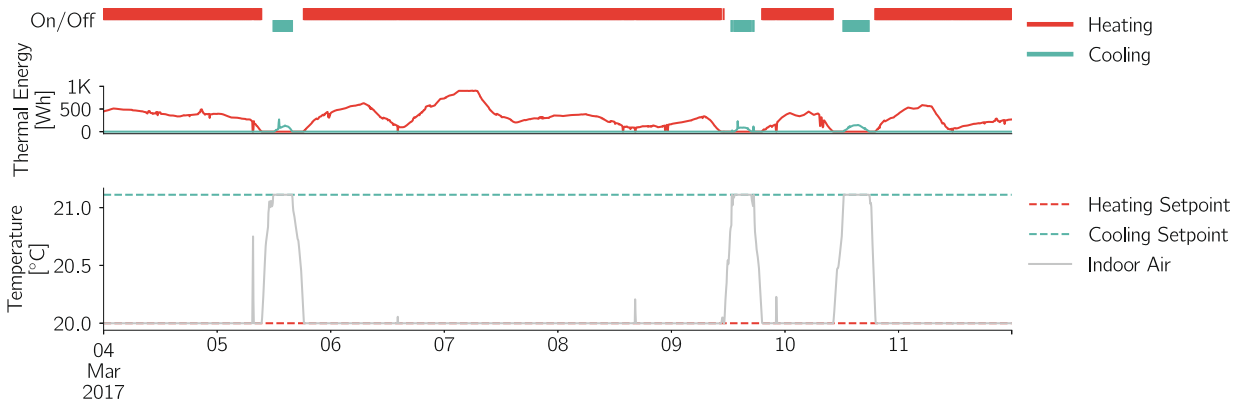

Fig. 4. The on/off cycling, delivered HVAC thermal energy, indoor air and setpoint temperatures of an example smart thermostat taken from our generated ResStock sample before modification.

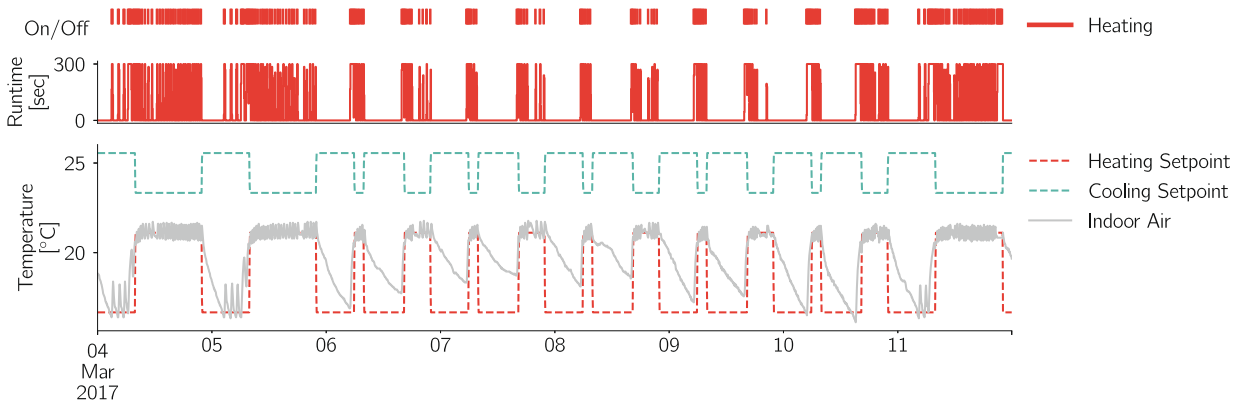

Fig. 5. The on/off cycling, equipment runtime, indoor air and setpoint temperatures of an example smart thermostat taken from the Donate Your Data dataset [26].

$$
\begin{aligned}
\min \quad & \sum_{h=1}^{H} \left(q_{heat,t+h} + q_{cool,t+h}\right) + \gamma\left(d_{heat,t+h} + d_{cool,t+h}\right) \\
\text{s.t.} \quad T_{in,t+1} &= \sum_{m=0}^{4} \left[\beta_{1,m} T_{in,t-m} + \beta_{2,m} T_{out,t-m} + \beta_{3,m} q_{heat,t-m} + \right. \\
& \left. \beta_{4,m} q_{cool,t-m} + \beta_{5,m} q_{solar,t-m}\right] \\
& q_{heat,t+h}, q_{cool,t+h} \in \{0,1\} \\
& q_{heat,t+h} + q_{cool,t+h} \leqslant 1 \qquad (2) \\
d_{heat,t+h} &= \begin{cases} 1000(T_{heat} - T_{in,t+h}), & T_{in,t+h} \leqslant T_{heat,t+h} - 2.0 \\ T_{heat,t+h} - T_{in,t+h}, & T_{in,t+h} \leqslant T_{heat,t+h} \\ 0, & T_{in,t+h} > T_{heat,t+h} \end{cases} \\
d_{cool,t+h} &= \begin{cases} 1000(T_{in,t+h} - T_{cool,t+h}), & T_{in,t+h} \geqslant T_{cool,t+h} + 2.0 \\ T_{in,t+h} - T_{cool,t+h}, & T_{in,t+h} \geqslant T_{cool,t+h} \\ 0, & T_{in,t+h} < T_{cool,t+h} \end{cases}
\end{aligned}
$$

optimized the use of heating and cooling (q_{heat}, q_{cool}) in order to minimize the total cost over a prediction horizon H. The cost objective was the γ-weighted sum of discomfort (d) and runtime (q); a detailed description of the choice of $\gamma = 1$ is provided in Section 2.2.4. In a real application, the γ value trading off equipment running time and user discomfort could be customized for each house or individual occupant, similar to what was done by Jin et al. [29]. However, we opted not to add the complication of modeling a distribution of user trade-off preferences for γ and instead used a fixed value for all houses.

The discomfort values (d_{heat}, d_{cool}) penalized overheating or cooling when the indoor air temperature was outside of the comfort band of the heating and cooling setpoint. Defined individually for a heating and cooling component (see Eq. (2)), discomfort had both a zero-cost region, a linear cost region, and an extreme discomfort cost region (with a penalty multiplier of 1000) incurred when the indoor temperature deviated more than 2\°C from the heating or cooling setpoint. Similar to the runtime and discomfort trade-off, both the size of the deviation (i.e., the 2.0°C value) and the extreme penalty (i.e., the 1000 multiplier) could be customized per house if desired. We note that our definition of discomfort only considers air temperature and not all the factors considered in ASHRAE 55 [30]; however, air temperature is the primary control variable for smart thermostats. We leave a more holistic approach to discomfort and the development of potentially multiple additional predictive models as potential future work.

During the optimization, exogenous variables such as outdoor temperature (T_{out}) and solar irradiance (q_{solar}) were treated as deterministic. The horizon H was limited to only 12 timesteps (or one hour) given the practical constraints of running the optimization every timestep (105,120 instances per year of simulation) for all 500 models. Based on the MILP's solution, the first predicted control action was executed and the EnergyPlus model continued to the next timestep. Just as with the deadband controller, the air conditioning was locked out when the outdoor temperature was below 10°C.

2.2.3. Tabular Q with batch experience replay

In contrast to the model-predictive control methodology discussed previously, model-free RL methods offer an attractive alter-

native for training an HVAC control policy directly from historical data through a method known as batch (offline) learning [31]. Ostensibly, this permits us to avoid the model training required in model-predictive control methods (e.g., the linear model we use in the MPC method described previously), thus also avoiding control errors induced by inherent inaccuracies and approximations in a learned model.

We selected a standard tabular Q-learning method (TQBER) [17] as the basis for our model-free RL method. Given the relatively simple observation space, there is no need to use more recent (and more difficult to train) function approximation methods such as deep Q-learning with neural networks (DQN) [32] discussed in Section 4.2. The key advantage of Q-learning is that it can be trained offline using batch experience replay, i.e., it is possible to train it on historically collected data using a different control policy so long as the control policy in the historically collected data adequately explores all control actions in all states (i.e., it will randomly deviate from the control policy to explore alternative controls). For the batch experience replay, we used the same two weeks of historical (deadband control based) data as the ridge regression model training received (Section 2.2.2).

Unlike the deadband control data used for the baseline (Section 2.2.1) and ridge regression model training (Section 2.2.2), a new ϵ-deadband controller was used to generate a new historical dataset. The ϵ-deadband controller took a random action 10% (ϵ) of the timesteps to meet the exploration requirements for convergence of Q-learning. To understand this exploration requirement, it is important to consider that for offline training on historical data without a model, a model-free RL method must be able to evaluate the repercussions of all control actions in each state. Quite simply, it is impossible to know if there is another better action in a state if the controller that generated the historical training data only ever makes the same decision in that state. We selected a 10% value for random action exploration to minimise the effects of exploration on equipment runtime and occupant discomfort. While a higher ϵ value may lead to better learning performance for TQBER, we conjecture that it would significantly degrade occupant comfort and unnecessarily increase runtime – we deemed $\epsilon = 10\%$ to be the maximum random exploration that a user would tolerate with a smart thermostat. The random action selection obeyed the same temperature lockout on the air conditioner as the other control methods.

We defined our state space (S) using two values S_1 and S_2. The S_1 component of the state was the difference between the indoor temperature and the heating setpoint ($T_{in} - T_{stp,heat}$). The S_2 component was the difference between the indoor temperature and the cooling setpoint ($T_{in} - T_{stp,cool}$). The values for both S_1 and S_2 were rounded to the nearest tenth of a degree Celsius. The action space (A) had three states: off, heat on, and cool on. The rewards (R) were the same magnitude as the costs in Section 2.2.2, only negated. Runtime of heating or cooling was assigned a value of -1 and had discomfort (d_{heat} and d_{cool}) subtracted. Similar to the model predictive control strategy, the relative weights of the rewards could be further customized in other applications.

Using our historical cache of data from a yearly simulation of each house using the ϵ-deadband controller, a new Q-value table (containing the learned control policy and cost estimates) was learned by TQBER using the following iterative update

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha\left[R_{t+1} + \phi \max_a Q(S_{t+1}, a) - Q(S_t, A_t)\right]$$

A new Q-value table was constructed for the 1^{st} and 15^{th} of each month using the previous two weeks of data. Eq. 3 was trained with $\alpha = 0.2$ and $\phi = 0.9$ which were both manually selected. During control, the trained Q-value table was used to decide the best next action to take, based on which action given S_1 and S_2 had the highest value. The state (S_1, S_2) needed to have been visited at least five times in the training data, else the controller fell back to the standard deadband control (Section 2.2.1) for that timestep.

2.2.4. Comfort versus runtime trade-off

Past researchers have applied diverse definitions of comfort when incorporating it into control optimization. Some researchers chose to make comfort bands (i.e., setpoints) hard constraints in their problem definition [14]. With that approach, it is possible to have completely infeasible solutions that the system must be properly engineered to gracefully handle. Other researchers have relied on the predicted mean vote [16,33] or the adaptive comfort model [34] as their definition of comfort. Unfortunately, neither of these standard models translate well to the residential domain [35–37]. The inability to transfer these models is attributable to both a lack of datasets related to the comfort of occupants in their homes, and the difference in occupant behavior at home. For instance, the range of activities happening and the demographics of the occupants in a house is far more varied than in commercial buildings.

We utilized the existing thermostat setpoints (which were selected as part of the ResStock sampling) as the basis for the understanding of the comfort of the occupants and considered the linear deviation outside of the comfort band (i.e., the temperatures between the heating and cooling setpoints) as discomfort. We tuned the γ in our cost function (see Eq. (2)) by looking at a random sample of houses and how the magnitudes in both runtime and discomfort changed as γ was adjusted. Fig. 6 shows, for only two of the sampled building models after a year of simulation, how comfort and runtime were affected by different values of γ. In both cases (Fig. 6a and b) the MPC controller is more responsive to changing γ compared to the TQBER. For the MPC in particular, the effect of changing γ stabilizes as it is increased above 1.0. Based on the extension of this analysis to multiple houses from our sample, we selected $\gamma = 1.0$. This value seemed to result in only minor violations of the thermostat setpoints but also still seemed to achieve HVAC runtime savings.

3. Results

Our initial results confirm that our testbed was performing as expected (Section 3.1). We then present a detailed breakdown of the deadband controller since it is highly critical for comparison in both absolute and relative terms (Section 3.2). The evaluation of the controllers was conducted based on the total evaluated costs and on the discomfort and runtime components (Section 3.3). We investigated the relative per-house difference in controller performance between the model-based (MPC) and model-free RL (TQBER) methods over traditional deadband controls (Section 3.4). Finally, leveraging our diverse sample, we investigated if houses with specific characteristics performed differently than others (Section 3.5).

3.1. Specific control demonstration

The three control strategies (defined in Section 2.2) were simulated for an entire year with each of the 500 thermal models. Fig. 7 shows the indoor and outdoor air temperature, temperature setpoints, and heating and cooling equipment runtimes for three days for two different simulated houses (Fig. 7a and b). The energy models are observed to have realistic (compared to Fig. 5) temperature deviations when the HVAC equipment is turned on. The deadband controls appear to follow the setpoints and are reactive to changing setpoints. The MPC is seen to make decisions that differ from the standard deadband control. For instance, it is able to predict set-

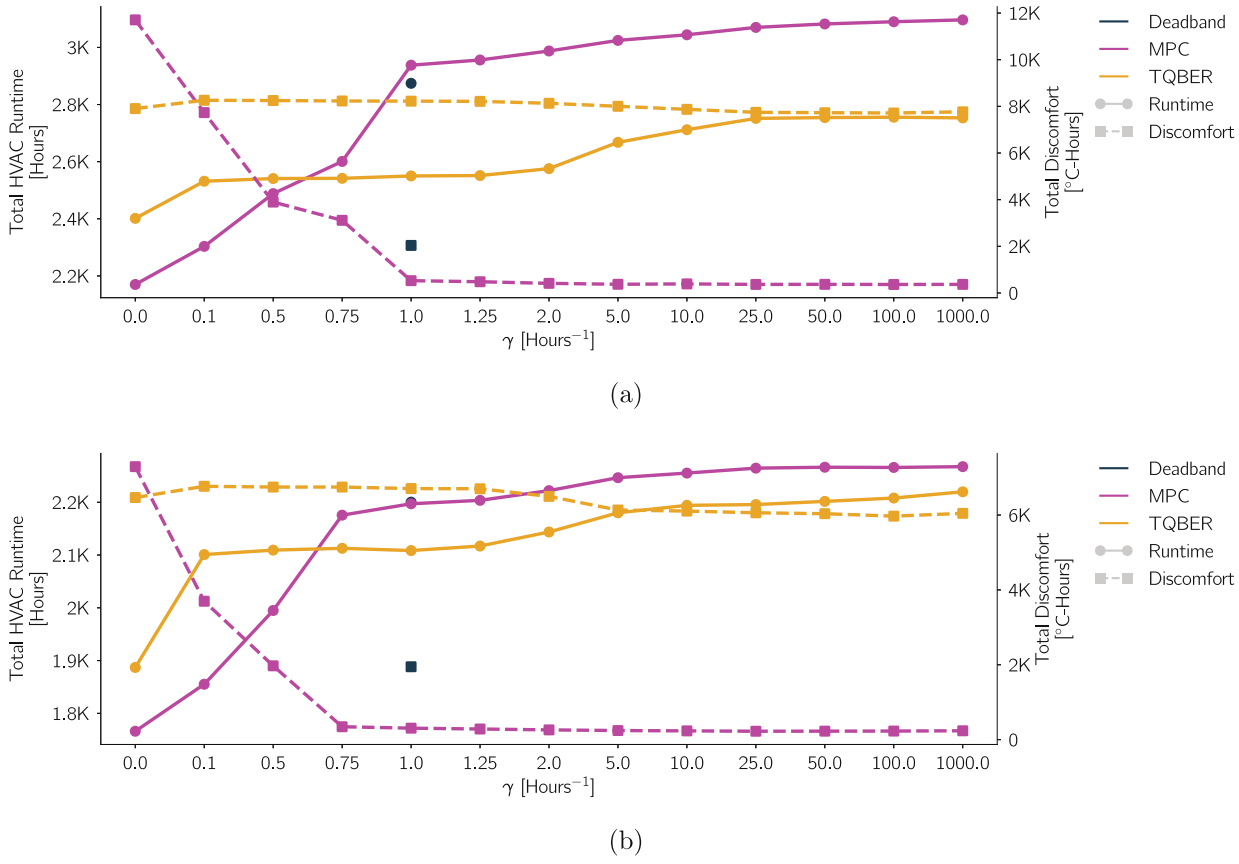

Fig. 6. Total HVAC runtime (left y-axis) and discomfort (right y-axis) for two different modeled residential buildings with different values of γ.

point changes (through the ability to calculate future predicted discomfort) and avoid discomfort. Simultaneously the MPC's solution finds times when the setpoints should be violated just a little to save equipment runtime. The TQBER method often mimics the deadband control (as it was a fallback state) but also appears to make poor decisions overall. These trends allude to many of the aggregated results we will observe in the coming sections.

3.2. Deadband controller

The deadband controller was the baseline for gauging the effectiveness of both our candidate data-driven control methods. Across our sample of 500 houses, there was considerable variation in the amount of runtime and discomfort experienced under the baseline control. Fig. 8 shows the distribution in heating and cooling runtime hours for all simulated residential buildings, broken down by month. The boxes represent the quartiles of the distributions, while the whiskers are 1.5 times the upper and lower portions of the interquartile range.[5] The majority of houses are not continually running their equipment, meaning there are savings opportunities – there are no opportunities for advanced control when the system is constantly required to run.

Fig. 9 shows the distributions in total runtime and discomfort for all house models broken down by month. The discomfort and runtime appear to positively correlate, meaning months with more runtime appear to have more discomfort. The discomfort values appear to have more outliers compared to the runtime. The difference in the number of outliers is attributed to the runtime having a smaller range of potential values. The median amount of discomfort appears fairly consistent over the year. The stability makes sense since the deadband control is always allowing a 0.5°C discomfort.

3.3. Absolute cost for each controller

Fig. 10 shows the distributions for all house models in (a) total cost and (b) separated out between runtime and discomfort for each control method over one year of simulation. In Fig. 10a the MPC is shown to have the lowest costs compared to both the deadband and the TQBER controls. The TQBER is observed to have the highest cost of any of the control methods. The two extreme outliers (a total cost of more than 80,000) for all control methods were found to be two small (1000 ft^2) houses in relatively hot climates (see Section 3.5) that have no central air conditioning. In the breakdown of the cost components, as shown in Fig. 10b, the discomfort value is seen to be where MPC has a distinct advantage over the deadband baseline. This ability to improve comfort is visible in both examples in Fig. 7. The MPC made the transition to setpoints sooner and would often remain just above the setpoints. Meanwhile, for TQBER, the discomfort is where the majority of the increased cost appears, with actually some reduction of runtime visible.

3.4. Relative performance over the deadband control

The absolute costs, while informative, mask the trends on a per house basis. To better understand the impacts of the controllers per house, we calculated the percentage change in discomfort and runtime from the deadband controls for each house. The distributions of the percentage change by month and by year are shown

[5] All boxplots presented in the later figures have this same form.

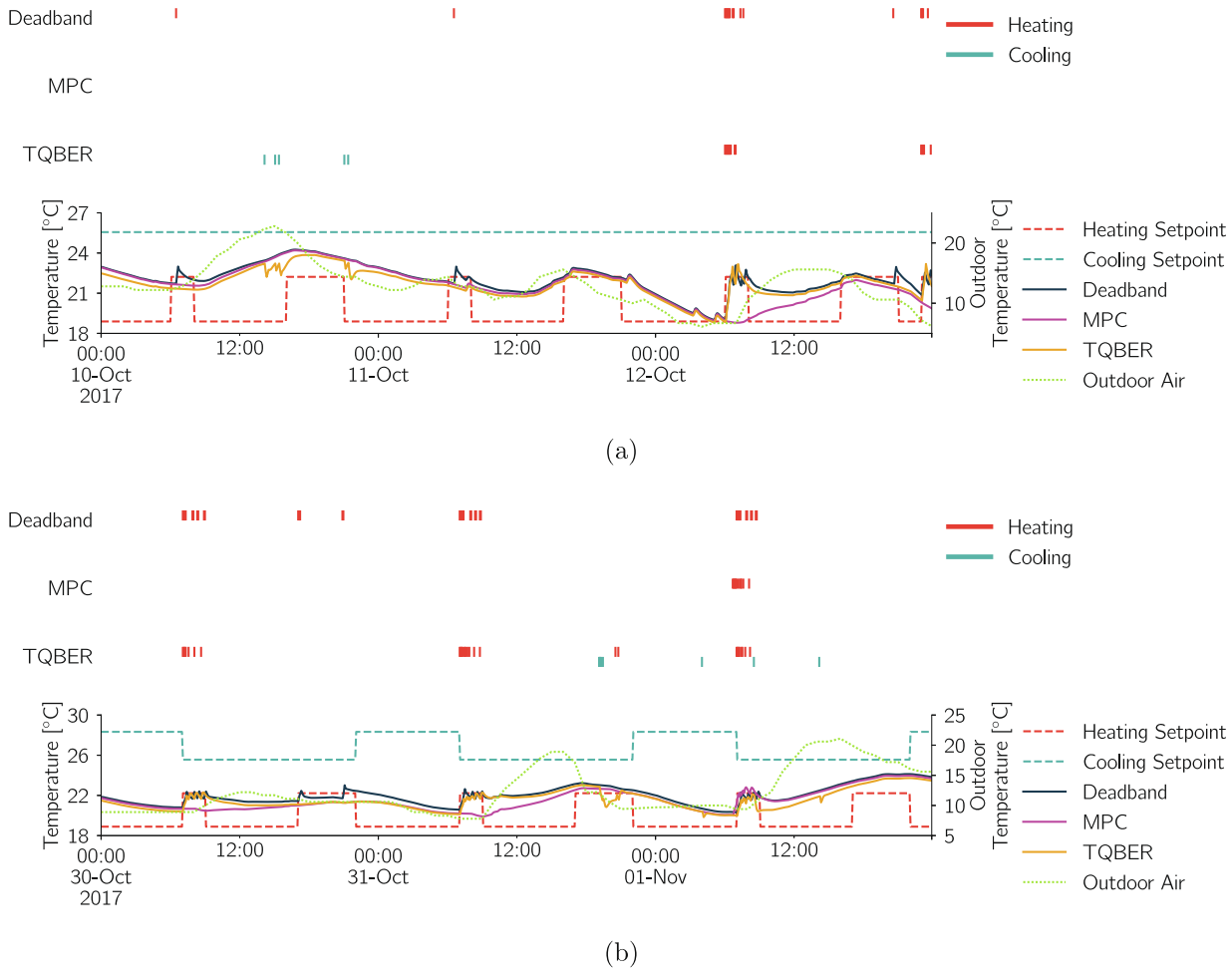

Fig. 7. Runtime and temperatures for three days using the three controllers for two (a and b) modeled houses.

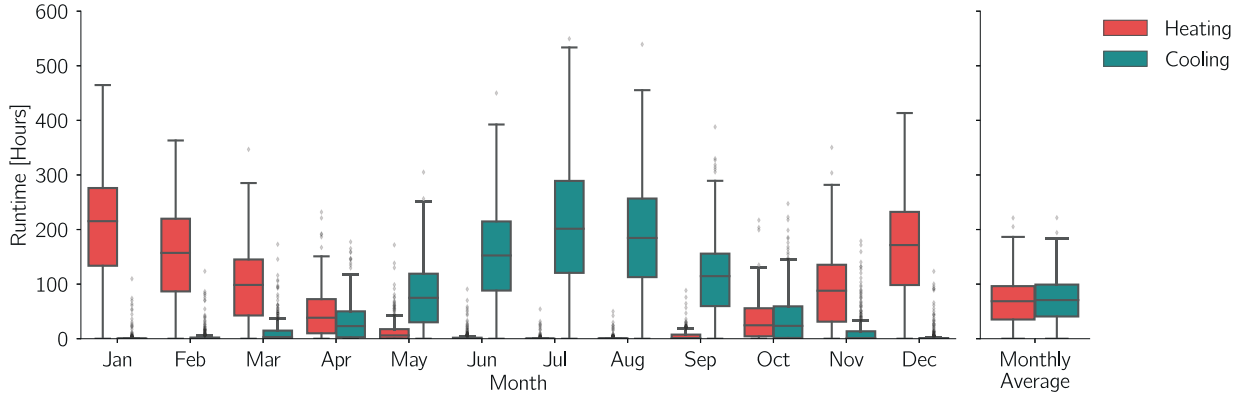

Fig. 8. Distribution in the number of heating and cooling hours each month for all house models simulated using deadband control.

for MPC in Fig. 11 and for TQBER in Fig. 12. For MPC, discomfort was often almost fully eliminated in comparison to the deadband controls. The variation in discomfort savings is highest in the shoulder months. In contrast, there is relatively little variation in the savings in runtime during the year. Unlike the MPC, the TQBER performed much worse than the deadband controller; particularly in terms of discomfort. The relative amount of discomfort is the least in the summer, but the summer is also when the absolute amount of discomfort (Fig. 9) is the highest.

3.5. Latent factors affecting control

One intention of using a diverse sample of modelled houses was to be able to see if the latent (i.e., underlying and physical) characteristics of the house made the data-driven control methods more or less effective. We looked specifically at the floor area (Section 3.5.1), the climate zone (Section 3.5.2), and the vintage of the house (Section 3.5.3).

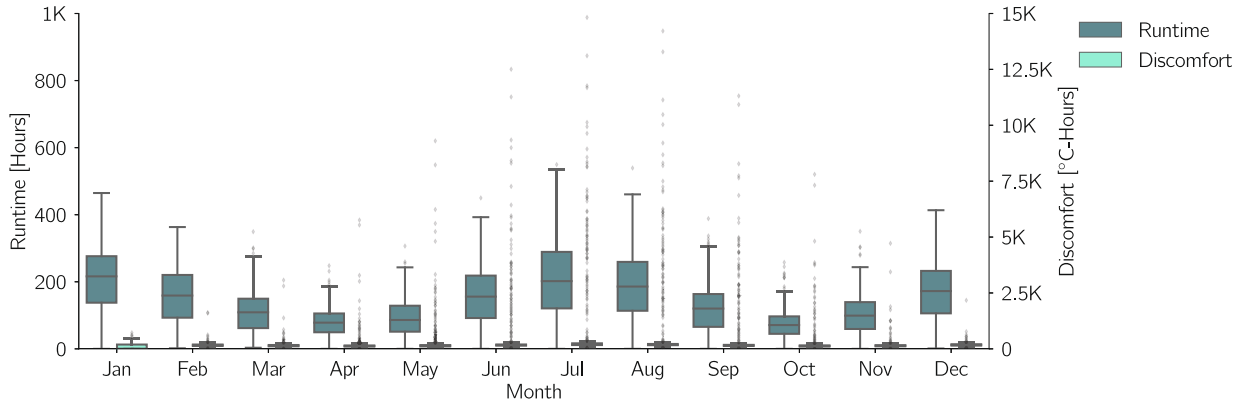

Fig. 9. Distributions in runtime and discomfort when using deadband controls for all house models broken down by month.

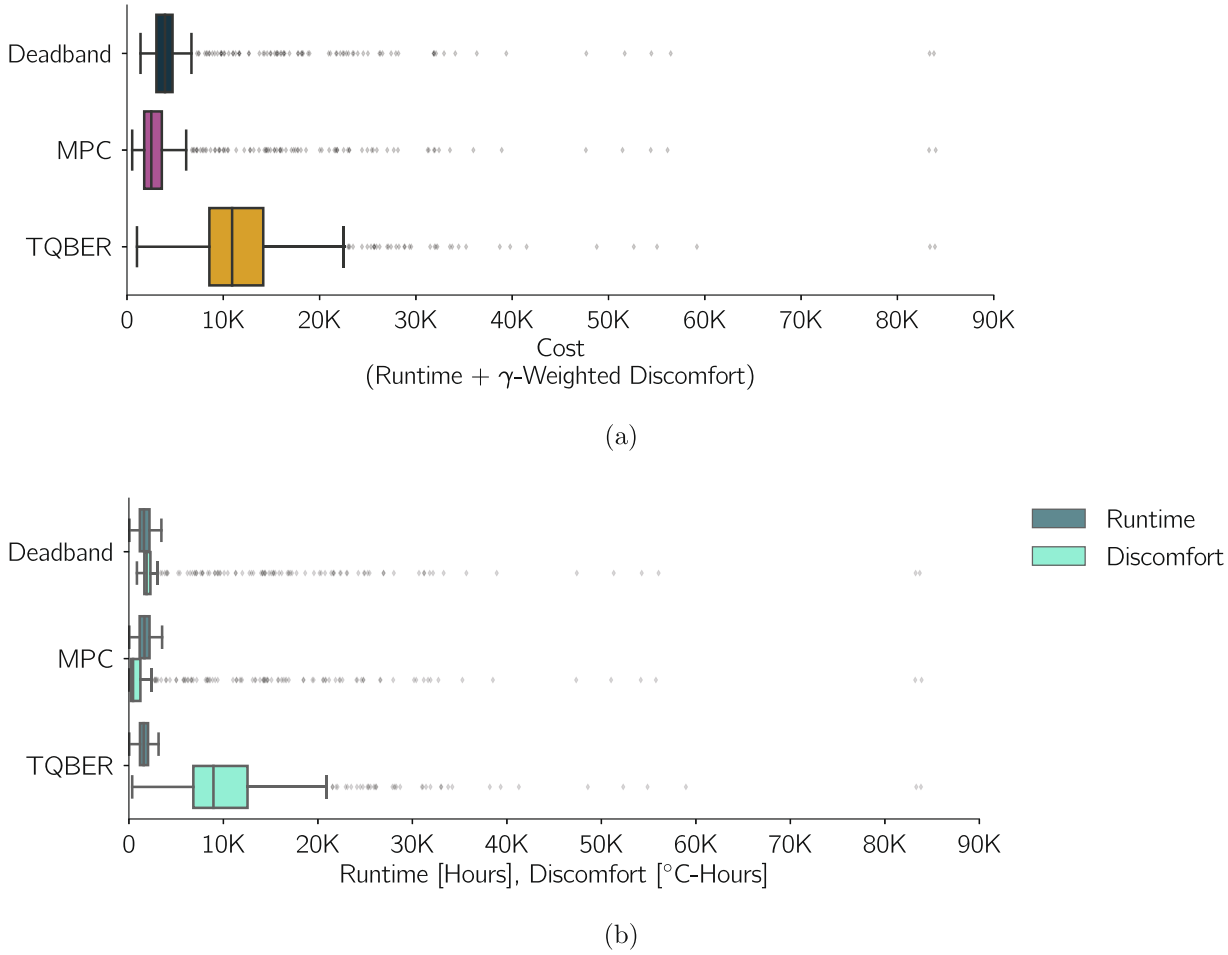

Fig. 10. Distribution for each control method across the entire sample for the (a) total cost and (b) separate components of runtime and discomfort.

3.5.1. Performance based on floor area

The floor area of the houses was expected to have some effect on the ability of the house to be controlled by various control methods. For instance, larger homes may have lower floor area to wall (and window) ratios that change their responsiveness to outdoor conditions. Fig. 13 shows the distribution in total cost (the γ-weighted sum of runtime and discomfort) for each controller over a year of simulation, broken down by the floor area of the building. The smallest houses appear to have the largest variation of value, implying that smaller buildings are more difficult to control. We postulate that these houses are influenced more by exogenous disturbances such as occupants or ambient conditions. The two worst performing buildings (across all control methods) are both 1000 ft^2 in size.

3.5.2. Performance based on climate zone

The location of each house dictates the prevailing climate conditions each building is exposed to. Houses in different locations would have different opportunities for optimizing the controls based on changing weather patterns and may face different limits

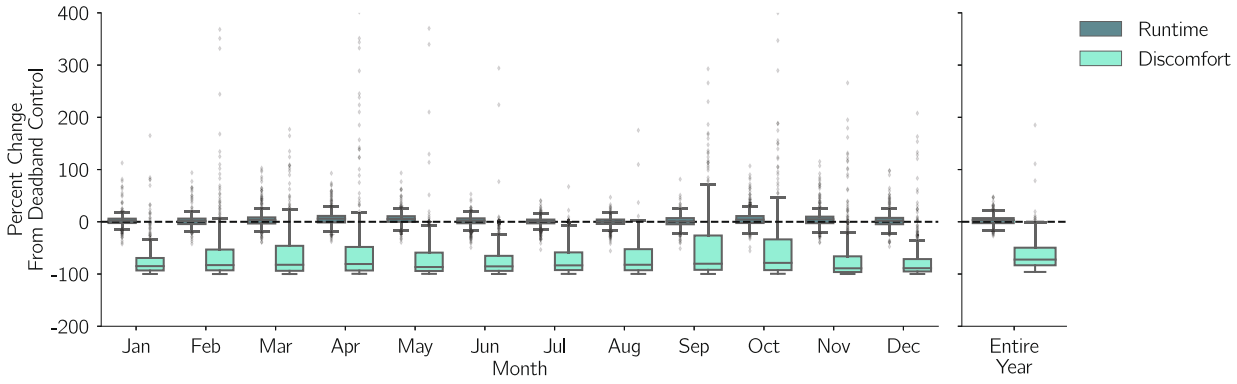

Fig. 11. Percent change in runtime and discomfort from deadband to the MPC controls broken down by month.

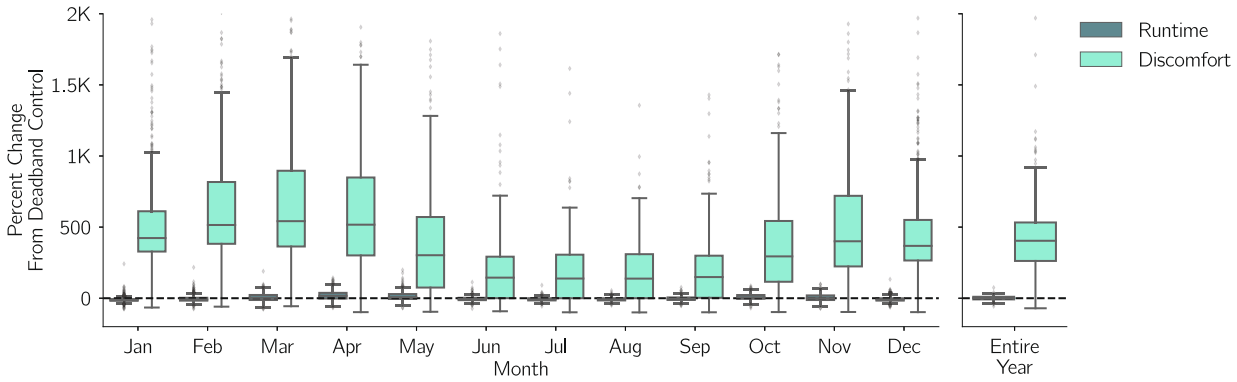

Fig. 12. Percent change in runtime and discomfort from deadband to the TQBER controls broken down by month.

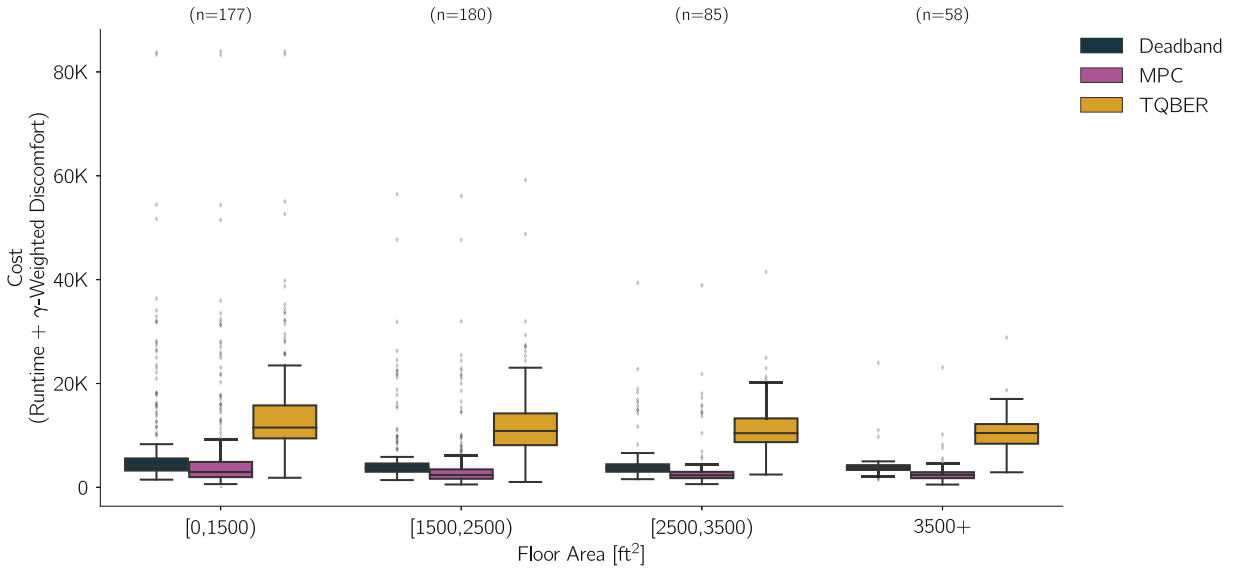

Fig. 13. Distribution in total costs over an entire year of simulation using the three controllers, broken down by the floor area of the modeled house. The number of houses with each floor area is shown above each floor area group.

on how much saving is possible. Fig. 14 shows the distribution of total cost based on each control method broken down by Building America Climate Zone [22]. The two smallest climate zones (Very Cold and Hot-Dry) have been grouped with their similar climates. It appears that regardless of the climate zone, MPC is able to reduce the total cost compared to deadband control. The Mixed-Dry/Hot-Dry region has a much larger distribution in the values compared to the other climate zones. This variation implies that control in the Mixed-Dry/Hot-Dry region is harder to optimize across the population. We note that the Mixed-Dry/Hot-Dry region did have one of the lowest installation rates of central air conditioners, with only the Marine climate having a lower installed fraction. This would lead to high discomfort during the summer months for some houses; reducing the overall effectiveness of any of the con-

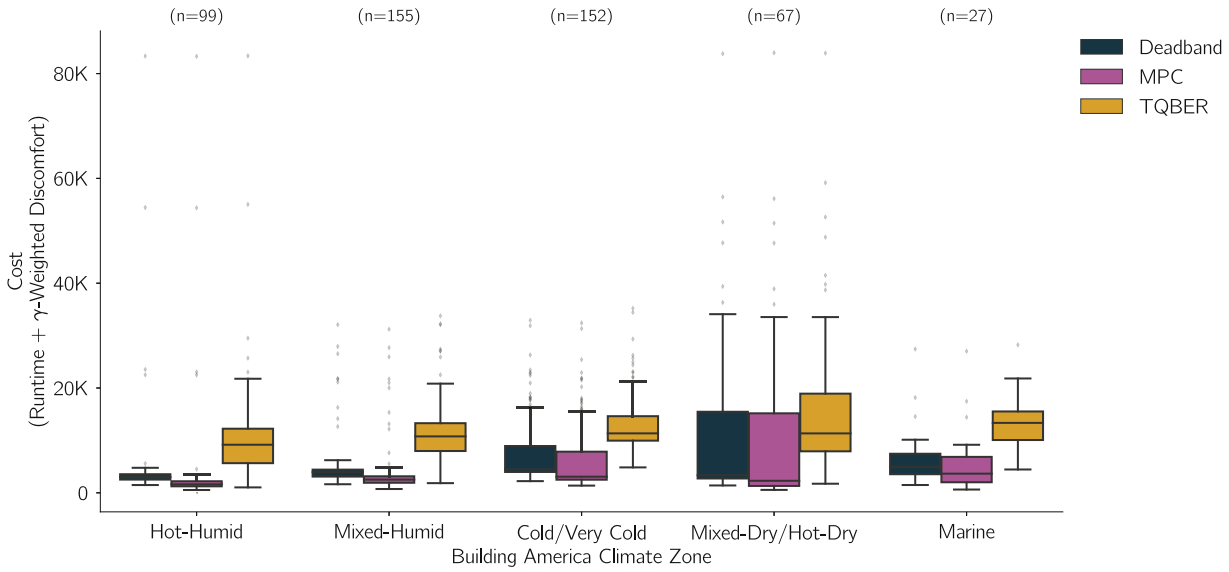

Fig. 14. Distribution in total costs over an entire year of simulation using the three controllers, broken down by the Building America Climate Region of the modeled house. The number of houses in each climate zone is shown above each climate zone.

trollers. However, these results reflect the actual opportunity for improvement based on the underlying housing distribution.

3.5.3. Performance based on the vintage

Similar to both the size and location of the houses being modelled, we wanted to explore if the age of the house affected its ability to be controlled. The vintage of the house affects a number of characteristics such as the materials used, installed HVAC, and internal loads. Fig. 15 shows the total cost for each controller based on the vintage of the house that was modeled. The oldest houses (from before 1950) have the largest variation in the controllers, while the newest houses have the lowest variation. The median performances across the vintages for each controller remain relatively consistent.

4. Discussion

Based on our investigation into the various data-driven control methods, we present three main points of discussion.

4.1. Effectiveness of model-based methods

Compared to both deadband and model-free RL (TQBER), the model-based MPC had the best performance when both runtime and discomfort were considered. The deadband control often made runtime-efficient decisions and only ran the HVAC equipment a minimal amount (see Fig. 7). The MPC generally reduced the number of setpoint violations and made the controls more preemptive to setpoint changes.

One major concern with implementing the MPC strategy is whether an accurate model of the system can be acquired. It generally appeared that our models were able to reasonably capture the dynamics of the systems with our testing methodology. However, given the deterministic nature of the exogenous variables in our predictions (i.e., our predictions of weather conditions were perfect[6] and occupants did not change their setpoints), the estimated median improvement of 36% over the deadband controller is an upper limit of savings and only with the same costing structure and weighting values in the optimization.

4.2. Challenges with model-free methods

For our model-free method which we trained offline, we encountered a number of issues which generally plague the implementation of RL in real-world applications. The first issue was how often the controller needed to fall back to the deadband controller because it had not seen the state often enough in training. Fig. 16 shows the distribution of the fraction of timesteps where the TQBER controller was falling back to deadband for each house. On average, over a third (36%) of the control actions per thermostat defaulted to the deadband control because the state did not meet the minimum number of observations during training to use the Q-value from the table. Even with the modest exploration (10%) in the ϵ-deadband training data, which comes with additional energy and discomfort for the occupants of the house, being able to quickly train a model using existing data appears to be infeasible. The narrow range of observed states is particularly evident in the visualization of the tabular Q-values for one house model after two weeks of training (Fig. 17a). The observations based on the two states (S_1 and S_2) exist only in a tiny band, and some of these values are based on less than five samples. If during testing there is any change in setpoints or shift in values (referred to as distribution shift), the table quickly becomes inadequate. Our simple state space does not take into account outdoor temperature nor any future states – two advantages the MPC method had. Addressing this lack of predictive features with a more detailed state space would only further exacerbate the existing data sparsity issues.

We did explore using a function approximation of the Q-values with a deep neural network. The deep Q network (DQN) [32] leverages all given observational data by learning a deep neural network function approximator to avoid a blow-up in space and sample complexity that occurs when a tabular method is used instead. The learned Q-values (i.e., both tabular and deep) for one house are shown in Fig. 17. Both used the same two weeks of training data from ϵ-deadband control. While the function approximation of the DQN managed to get some extrapolations correct, certain extrapolations (such as heating to on the right side) would be a terrible control decision in actuality. To be able to properly use DQN,

[6] Given the short one hour prediction horizon, weather forecasts could be generally quite accurate [38].

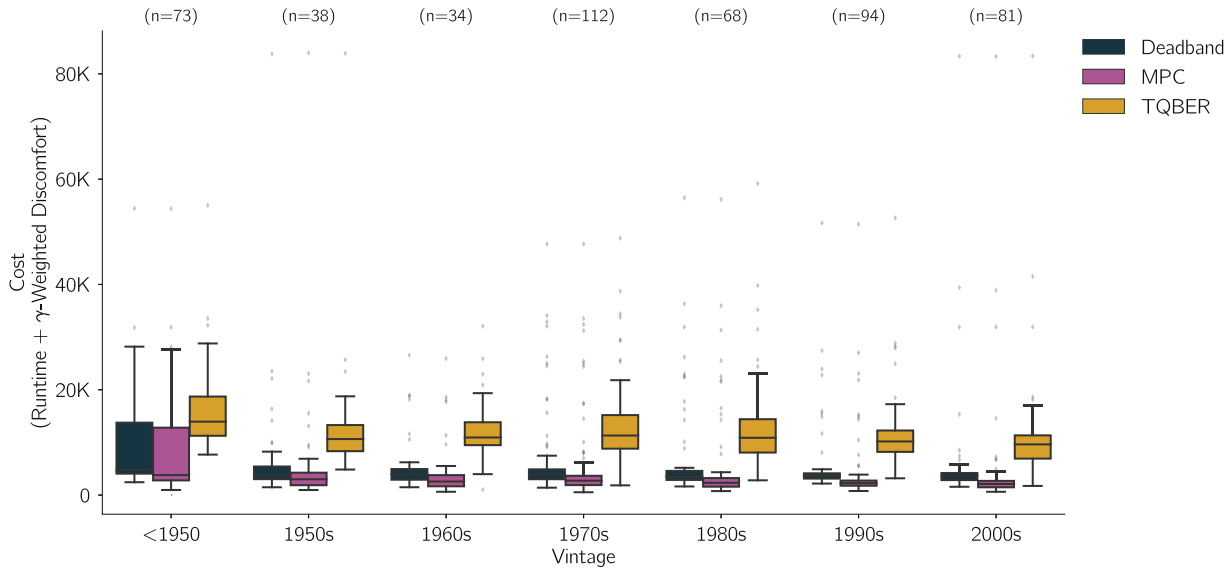

Fig. 15. Distribution in total costs over an entire year of simulation using the three controllers, broken down by the vintage of the modeled house. The number of houses of each vintage is shown above.

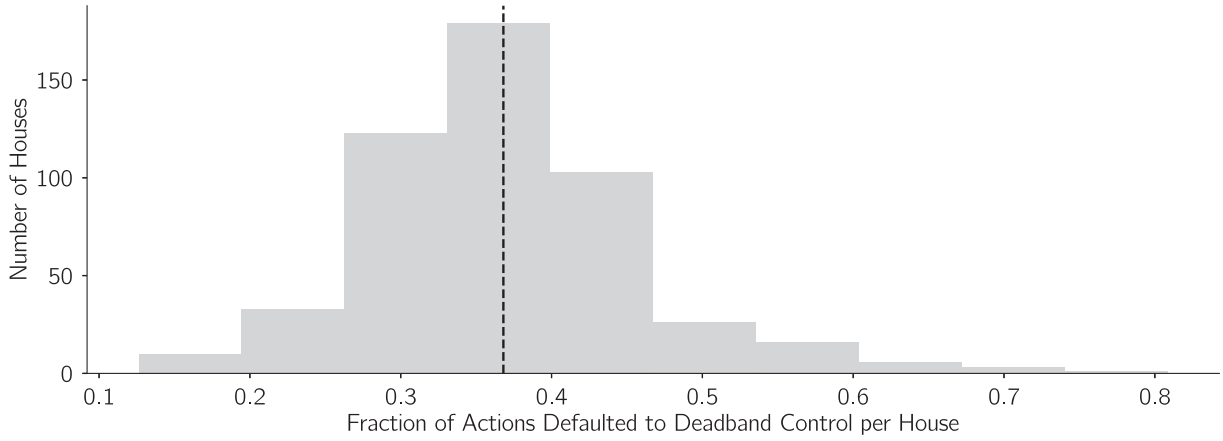

Fig. 16. Fraction of decisions per thermostat that had to default back to deadband controls. The average value is indicated by the dashed line.

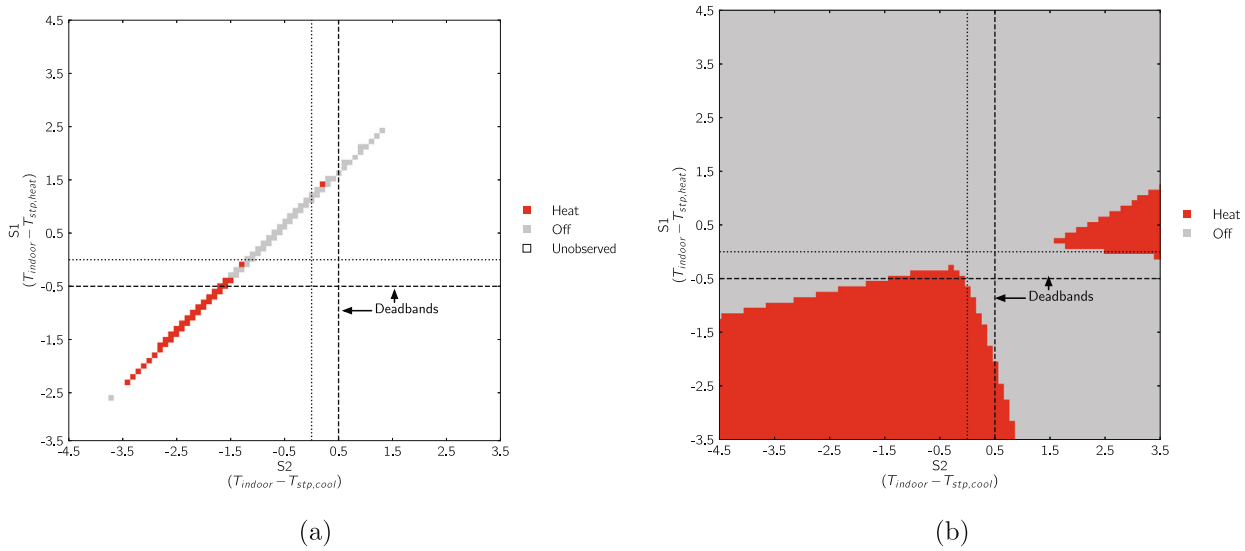

Fig. 17. Learned Q-values with (a) Tabular Q and (b) a Deep Q Network for one house trained using two weeks of data with 10% exploration in the ϵ-deadband controller.

the function approximation would need to be more tightly controlled during training or a larger and more diverse dataset is required.

4.3. Implementing data-driven control in the field

The intention of our work was to determine how best to utilize typical and existing smart thermostat data for improving the control over individual houses. We used modeling methods and a testing strategy that replicated near real-world scenarios and could easily be scaled to hundreds of thousands of devices. Of course, to utilize these methods in a fleet of real smart thermostats, a number of additional considerations would need to be made. In a real system, more safeguards would need to be in place. For instance, based on our model training methodology, if a controller performed poorly, this could persist for almost two weeks before it had a chance to be corrected. During seasonal changes, it was possible a model could be trained with no knowledge of cooling or heating equipment that during active control may be required. Evaluation of model training would need to be better monitored to prevent the use of a bad model for optimization. In our investigation, we purposefully avoided this screening to be able to see how poor the control experience may be in a simulated environment; this is not an option in a real house. The developed controller would need to account for more complicated equipment, such as multi-stage cooling or heating. In the case of multi-stage equipment, the relative costs of the different stages would need to be included in the cost function. Finally, the adjustment of the relative trade-off of comfort versus runtime (γ) should be adjusted per user or house.

5. Conclusion

With the availability of rich historical data from the latest generation of residential thermostats, we sought to understand how the data could be leveraged to improve the standard heuristic controls generally implemented by residential thermostats today. To test the data-driven controls, we built a large and diverse sample of residential energy models and simulated a year of operation. The energy models were modified to accurately recreate the temperature response, equipment actuation, and data provided by commercially available smart thermostats.

We evaluated three varieties of controllers: a baseline controller, a model-based (model-predictive) controller, and a model-free (reinforcement learning) controller. The baseline control strategy was a deadband controller; typical of the controls found on standard smart or programmable thermostats. Our model-based method was a model predictive controller using a linear ridge regression model trained using only the observational data from the deadband controller. Finally, our model-free method, a tabular Q with batch experience replay, was trained offline using the same amount of data as our model predictive controller. Both data-driven methods were intended to minimize the cumulative costs of discomfort (based on distance from the users' comfort band) and total HVAC runtime. Based on year-long simulations on a diverse collection of 500 residential models, the model predictive controller was able to generally reduce the combined cost of runtime and discomfort in comparison to the deadband controller. The tabular Q method's performance was poor due to sparse observational data and the limited amount of exploration we allowed due to HVAC equipment runtime cost and occupant comfort considerations.

Ultimately, to achieve decentralized on-device control of smart thermostats, the controller needs to be robust, reliable in the control actions, and have good performance across a range of settings. We conclude that optimal data-driven MPC methods leveraging the linear learned models introduced in this work provide a promising and effective method for the control of smart residential thermostats across a wide variety of housing stock and climate conditions.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

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

References

[1] Natural Resources Canada, Canada's Secondary Energy Use, https://oee.nrcan.gc.ca/corporate/statistics/neud/dpa/showTable.cfm?type=HB&sector=aaa&juris=ca&rn=2&page=0, 2019. .
[2] U.S. Energy Information Administration, Use of Energy Explained, https://www.eia.gov/energyexplained/use-of-energy/, 2019. .
[3] U.S. Energy Information Administration, Use of Energy Explained: Energy use in homes, https://www.eia.gov/energyexplained/use-of-energy/homes.php, 2019. .
[4] U.S. Energy Information Administration, Residential Energy Consumption Survey (RECS), https://www.eia.gov/consumption/residential/data/2015/, 2015. .
[5] R. May, The reinforcement learning method: A feasible and sustainable control strategy for efficient occupant-centred building operation in smart cities, ISBN 9789188679031, 10.13140/RG.2.2.29921.86883, http://urn.kb.se/resolve?urn=urn:nbn:se:du-30613, 2019. .
[6] A. Afram, F. Janabi-Sharifi, A.S. Fung, K. Raahemifar, Artificial neural network (ANN) based model predictive control (MPC) and optimization of HVAC systems: A state of the art review and case study of a residential HVAC system, Energy and Buildings 141 (2017) 96–113, ISSN 0378-7788, 10.1016/j.enbuild.2017.02.012, http://linkinghub.elsevier.com/retrieve/pii/S0378778816310799 https://doi.org/10.1016/j.enbuild.2017.02.012. .
[7] D. Lindelöf, H. Afshari, M. Alisafaee, J. Biswas, M. Caban, X. Mocellin, J. Viaene, Field tests of an adaptive, model-predictive heating controller for residential buildings, Energy and Buildings 99 (2015) 292–302, https://doi.org/10.1016/j.enbuild.2015.04.029, ISSN 03787788.
[8] C.D. Corbin, G.P. Henze, Predictive control of residential HVAC and its impact on the grid. Part I: simulation framework and models, Journal of Building Performance Simulation 10 (3) (2017) 294–312, ISSN 19401507, doi: 10.1080/19401493.2016.1231220, http://www-tandfonline-com.ezproxy.library.ubc.ca/doi/abs/10.1080/19401493.2016.1231220. .
[9] J. Drgona, M. Klauco, M. Kvasnica, MPC-based reference governors for thermostatically controlled residential buildings, Proceedings of the IEEE Conference on Decision and Control 54rd IEEE (Cdc), 2015, pp. 1334–1339, ISSN 07431546, 10.1109/CDC.2015.7402396. .
[10] B.J. Claessens, D. Vanhoudt, J. Desmedt, F. Ruelens, Model-free control of thermostatically controlled loads connected to a district heating network, Energy and Buildings 159 (2018) 1–10, ISSN 03787788, doi: 10.1016/j.enbuild.2017.08.052. .
[11] J.R. Vázquez-Canteli, Z. Nagy, Reinforcement learning for demand response: A review of algorithms and modeling techniques, Applied Energy 235 (October 2018) (2019) 1072–1089, ISSN 03062619, doi: 10.1016/j.apenergy.2018.11.002. .
[12] C. Lork, W.T. Li, Y. Qin, Y. Zhou, C. Yuen, W. Tushar, T.K. Saha, An uncertainty-aware deep reinforcement learning framework for residential air conditioning energy management, Applied Energy 276 (January) (2020) 115426, https://doi.org/10.1016/j.apenergy.2020.115426, ISSN 03062619.
[13] E. Barrett, S. Linder, Autonomous hvac control, a reinforcement learning approach, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9286, Springer, Cham, 3–19, ISBN 9783319234601, ISSN 16113349, 2015, doi: 10.1007/978-3-319-23461-8_1. .
[14] O. Kotevska, K. Kurte, J. Munk, T. Johnston, E. McKee, K. Perumalla, H. Zandi, RL-HEMS, Reinforcement learning based home energy management system for HVAC energy optimization, ASHRAE Transactions 126 (2020) 421–429, ISSN 00012505.
[15] K. Kurte, J. Munk, O. Kotevska, K. Amasyali, R. Smith, E. McKee, Y. Du, B. Cui, T. Kuruganti, H. Zandi, Evaluating the adaptability of reinforcement learning based HVAC control for residential houses, Sustainability (Switzerland) 12 (18) (2020) 1–38, ISSN 20711050, doi: 10.3390/su12187727. .

[16] A. Javed, H. Larijani, A. Ahmadinia, R. Emmanuel, Comparison of the robustness of RNN, MPC and ANN controller for residential heating system, in: Proceedings - 4th IEEE International Conference on Big Data and Cloud Computing, BDCloud 2014 with the 7th IEEE International Conference on Social Computing and Networking, SocialCom 2014 and the 4th International Conference on Sustainable Computing and C, IEEE, 2015, pp. 604–611, ISBN 9781479967193, doi: 10.1109/BDCloud.2014.20, http://ieeexplore.ieee.org/document/7034849/. .

[17] R.S. Sutton, A.G. Barto, Reinforcement Learning: An Introduction, MIT Press, 2018.

[18] G.T. Costanzo, S. Iacovella, F. Ruelens, T. Leurs, B.J. Claessens, Experimental analysis of data-driven control for a building heating system, Sustainable Energy, Grids and Networks 6 (2016) 81–90, ISSN 23524677, https://doi.org/10.1016/j.segan.2016.02.002. .

[19] F. Bünning, B. Huber, P. Heer, A. Aboudonia, J. Lygeros, Experimental demonstration of data predictive control for energy optimization and thermal comfort in buildings, Energy and Buildings 211, ISSN 03787788, DOI: 10.1016/j.enbuild.2020.109792. .

[20] Z. Pang, Y. Chen, J. Zhang, Z.O. Neill, H. Cheng, B. Dong, How much HVAC energy could be saved from the occupant-centric smart home thermostat: A nationwide simulation study, Applied Energy (November) (2020) 116251, ISSN 0306-2619, 10.1016/j.apenergy.2020.116251, https://doi.org/10.1016/j.apenergy.2020.116251. .

[21] E. Wilson, C. Christensen, S. Horowitz, J. Robertson, J. Maguire, E. Wilson, C. Christensen, S. Horowitz, J. Robertson, J. Maguire, Energy Efficiency Potential in the U.S. Single-Family Housing Stock, Tech. Rep. January, www.nrel.gov/publications. (2017).

[22] M.C. Baechler, T.L. Gilbride, P.C. Cole, M.G. Hefty, K. Ruiz, Guide to Determining Climate Regions by County, Tech. Rep. August, Pacific Northwest National Laboratory (2015), https://www.energy.gov/sites/prod/files/2015/10/f27/ba_climate_region_guide_7.3.pdf.

[23] ANSI/ACCA, Manual J: Residential Load Calculation, 8th edn., 2016. .

[24] ANSI/ACCA, Manual S: Residential Equipment Selection, 3 edn., 2014. .

[25] K. Cetin, M.H. Fathollahzadeh, N. Kunwar, H. Do, P.C. Tabares-Velasco, Development and validation of an HVAC on/off controller in EnergyPlus for energy simulation of residential and small commercial buildings, Energy and Buildings ISSN 0378-7788, doi: 10.1016/J.ENBUILD.2018.11.005. .

[26] ecobee Inc., Donate Your Data, https://www.ecobee.com/donateyourdata/, 2019. .

[27] B. Huchuk, S. Sanner, W. O'Brien, Evaluation of data-driven thermal models for multi-hour predictions using residential smart thermostat data, Journal of Building Performance Simulation. .

[28] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.

[29] X. Jin, K. Baker, D. Christensen, S. Isley, Foresee: A user-centric home energy management system for energy efficiency and demand response, Applied Energy 205 (2017) 1583–1595, https://doi.org/10.1016/j.apenergy.2017.08.166, ISSN 03062619, https://linkinghub.elsevier.com/retrieve/pii/S0306261917311856.

[30] ASHRAE, ANSI/ASHRAE Standard 55-2017: Thermal Environmental Conditions for Human Occupancy, American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Inc., 2017. .

[31] S. Lange, T. Gabel, M. Riedmiller, Batch Reinforcement Learning – Reinforcement Learning: State-of-the-Art, Springer, Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 45–73, ISBN 978-3-642-27645-3, doi: 10.1007/978-3-642-27645-3_2. .

[32] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, M. Riedmiller, Playing Atari with Deep Reinforcement Learning, in: NeurIPS-13 Deep Learning Workshop, 2013. .

[33] C. Turley, M. Jacoby, G. Pavlak, G. Henze, Development and evaluation of occupancy-aware HVAC control for residential building energy efficiency and occupant comfort, Energies 13 (20) (2020) 1–30, ISSN 19961073, doi: 10.3390/en13205396. .

[34] C. Wang, K. Pattawi, H. Lee, Energy saving impact of occupancy-driven thermostat for residential buildings, Energy and Buildings 211, ISSN 03787788, doi: 10.1016/j.enbuild.2020.109791. .

[35] F. Durrani, S. Samsuddin, M. Eftekhari, Y. Uno, Performance Assessment of Fanger's PMV in a UK residential building in Heating Season, in: ASHRAE Winter Conference, 2017.

[36] B. Huchuk, W. O'Brien, S. Sanner, A longitudinal study of thermostat behaviors based on climate, seasonal, and energy price considerations using connected thermostat data, Building and Environment 139 (2018) 199–210, ISSN 0360-1323, https://doi.org/10.1016/j.buildenv.2018.05.003, http://www.sciencedirect.com/science/article/pii/S0360132318302634. .

[37] H. Stopps, M.F. Touchie, Managing thermal comfort in contemporary high-rise residential buildings: Using smart thermostats and surveys to identify energy efficiency and comfort opportunities, Building and Environment 173 (February) (2020) 106748, ISSN 03601323, 10.1016/j.buildenv.2020.106748, https://doi.org/10.1016/j.buildenv.2020.106748. .

[38] A.R. Florita, G.P. Henze, Comparison of short-term weather forecasting models for model predictive control, HVAC and R Research 15 (5) (2009) 835–853, ISSN 10789669, doi: 10.1080/10789669.2009.10390868. .