Long-term trends in connected thermostat performance

Alan Meier
Lawrence Berkeley National Laboratory
Berkeley, CA
USA
akmeier@lbl.gov

Abigail Daken
U.S. Environmental Protection Agency
Washington, DC
USA
daken.abigail@epa.gov

Leo Rainer
Lawrence Berkeley National Laboratory
Berkeley, CA
USA
lirainer@lbl.gov

Keywords

thermostat, indoor temperature, heating systems, cooling, control and monitoring devices

Abstract

Internet-connected thermostats (CTs) control heating and cooling systems in about 30 % of North American homes, and capture half of annual thermostat sales. In 2017 the U.S. Environmental Protection Agency created a program to certify the performance of ENERGY STAR® Connected Thermostats. To demonstrate compliance with energy-saving criteria, thermostat vendors must submit performance summaries for a representative sample of up to 1250 homes. Vendors must then re-submit results from a new, representative sample every six months in order to maintain their ENERGY STAR certification.

This procedure has created a unique record of each thermostat's long-term performance in response to changes in weather, customer demographics, building stock, and algorithms. Our analysis covers submissions from 13 different vendors, who submitted data up to 7 times over four years. We found that these semi-annual samples generated relatively stable trends for comfort temperatures and Heating, Ventilating and Air Conditioning (HVAC) runtimes over the study period. However, some vendors achieved consistently more energy-conserving comfort temperatures and shorter HVAC runtimes than others. The most recent submission runtimes averaged 700 hours for heating and 1,300 hours for cooling, but some vendors achieved runtimes as much as 17 % below the mean. This implies lower energy consumption is due in large part to more successful algorithms and control strategies used by those vendors.

Introduction: The Connected Thermostat Has Moved from Rare to Commonplace in North America

Internet-connected thermostats (CTs) now control heating and cooling systems in about 30 % of U.S. and Canadian homes (compared to about 4 % in Europe) and capture half of annual thermostat sales. All connected thermostats allow users to remotely view and set their home's temperature and the most popular monitor occupancy, humidity, multiple inside temperatures (and set points), and additional HVAC modes. Popular models also learn patterns of occupancy in a home and establish an energy-saving schedule automatically. Sophisticated control algorithms take into account how outdoor temperature, humidity, and insolation are likely to affect a particular home. Researchers have already tapped this information to derive insights into such diverse aspects as occupancy prediction (Huchuk, Sanner, and O'Brien 2019), load shapes (Vellei, Martinez, and Le Dréau 2021), creation of digital twins (Hosseinihaghighi et al. 2022), thermal comfort preferences (Wang and Hong 2020), and geo-spatial aspects of power outages (Meier, Ueno, and Pritoni 2019).

Internet-connected thermostats especially benefit the designs of U.S. and Canadian buildings and their heating and cooling technologies. Specifically, the region's reliance on wood frame construction means that homes have relatively low thermal mass. Heating and cooling are typically supplied from a central system through ducted air (rather than water or refrigerant). These features enable homes to benefit more from daily temperature set-backs (and set-ups during the summer) than European homes. Nevertheless, at least one thermostat vendor has adapted their products to operate in European homes and heating systems (such as modulating boilers) through the OpenTherm language (Google Nest 2022). Unlike European

homes, North American homes typically use the thermostat to control both heating and cooling.

Vendors of CTs have claimed their products reduce heating and cooling energy use up to 20 % (King 2018)[1]. The CT should in principle save energy because it gives consumers greater flexibility in managing their homes' HVAC systems through enhanced interfaces, geofencing, and motion sensors. In addition, CT vendors can apply algorithms to achieve savings beyond the occupant's settings. Nevertheless, CTs might not achieve their full savings potential for the following reasons:

- the users may misunderstand the interface similar to the difficulties encountered with programmable thermostats (Peffer et al. 2011);

- the vendors' algorithms could incorrectly optimize HVAC operation;

- the vendors' algorithms could prioritize comfort over efficiency; or

- the CTs aren't compatible with the HVAC equipment.

Some of this uncertainty can be traced to a lack of standard laboratory test procedures to assess CTs. Connected thermostat vendors fortunately have data on how their products are operated in the field. Even though the first evaluation dates back to 2006 (Woods 2006) there have been, until recently, few credible field evaluations of energy savings from any CTs (King 2018). One barrier is the unpopularity of truly experimental study designs such as those based on recruit-and-deny programs, leaving most studies vulnerable to self-selection bias. Another major logistical barrier is obtaining matched energy consumption (billing) data from energy suppliers and thermostat data from vendors. Neither group wishes to share their data for both privacy and competitive reasons. Under these conditions, evaluations of energy savings based on metered consumption have had highly variable results. While some studies identified groups achieving up to 15 % savings (Nest Labs 2015), others found little or no savings (Evergreen Economics 2020). The results may have been obscured by mild climates, targeting unique sub-populations (e.g., low-income customers), self-selection biases, and a small energy savings signal across homes with widely disparate consumptions.

The U.S. Environmental Protection Agency (EPA) established the ENERGY STAR certification program for CTs in 2017 to fill this information gap. The antecedents to the program were described in an earlier paper (Daken, Meier, and Frazee 2016). In 2020, the EPA estimates that the program covered sales of about 4 million thermostats, representing 58 % of CTs sold. Uptake has been very rapid and is no longer a gadget for the tech-savvy early-adopters. The stock of homes with connected thermostats already resembles the stock of all single-family homes in the United States. A large group of one vendor's thermostats was compared to single-family homes in the Residential Energy Consumption Survey and found to be similar in many ways, such as geographic distribution, floor area, and number of occupants (Meier et al. 2019).

The energy-savings metric

The new technology of Internet-connected thermostats allowed ENERGY STAR to develop a novel procedure to verify the energy savings from installations of these devices in homes. Instead of a laboratory test procedure prior to release, vendors must demonstrate performance based on analysis and aggregation of data collected from homes in actual operation. This performance data is distilled into a metric that can apply to installations in a wide range of homes and climates. Note that this procedure requires vendors to place their units into the market for at least a year before they can apply for certification.

The ENERGY STAR savings metric estimates the reduction in heating and cooling equipment runtimes from a self-referential baseline that assumes no setback or setup. The ENERGY STAR performance metric is documented on the ENERGY STAR website (U.S. EPA 2022). The metric is calculated using interval data (including set points, indoor temperatures, and equipment runtimes) recorded by the thermostat as well as outdoor temperatures obtained from the national weather service for the thermostat location.

The EPA relies on thermostat vendors to analyse and aggregate data from a sample of homes to determine whether a CT is able to provide sufficient savings. To do this, vendors use the software provided by ENERGY STAR in two steps. First, they analyse the data from each home in a random sample of up to 250 of their thermostats from each of five climate zones (see Figure 1). Vendors input equipment runtime and inside temperatures for every hour of the annual evaluation period. Next, they use the provided software to generate a set of thermostat metrics for each thermostat in heating and in cooling . This set of thermostat metrics is then aggregated into a single file containing summary statistics (including n, mean, standard error of the mean, percentiles, and many more) for each metric. This summary data file is submitted at certification and then the process is repeated every 6 months using a new random sample of customer installations. In this paper all vendor data is anonymous and referred to by a color (e.g. Red). Note that we, the analysts, do not know which vendor submitted which data and have no ability to discuss results with individual vendors.

The procedure requiring all vendors to use the same standardized software package is novel from a regulatory perspective because it places the computational burden on the vendors. To some extent, procedure was dictated by the vendors' insistence that they retain all customer information. This procedure is computationally intensive and requires the vendors to have relatively sophisticated data science capabilities. Some vendors initially lacked these skills, which led to considerable teething problems.

Now, three years on, the seven 6-month data submissions provide a unique insight into how Americans heat and cool their homes as well as the performance of CTs from multiple vendors. Examining long-trends trends will give insights into:

- Occupant behaviour

- CT performance

- Problems with the metric or submissions by vendors

- Comparisons with self-reported data

---

1. The connected thermostat industry consists of three different, overlapping, players: manufacturers, vendors, and service providers. ENERGY STAR mostly deals with vendors, so this term is used here.
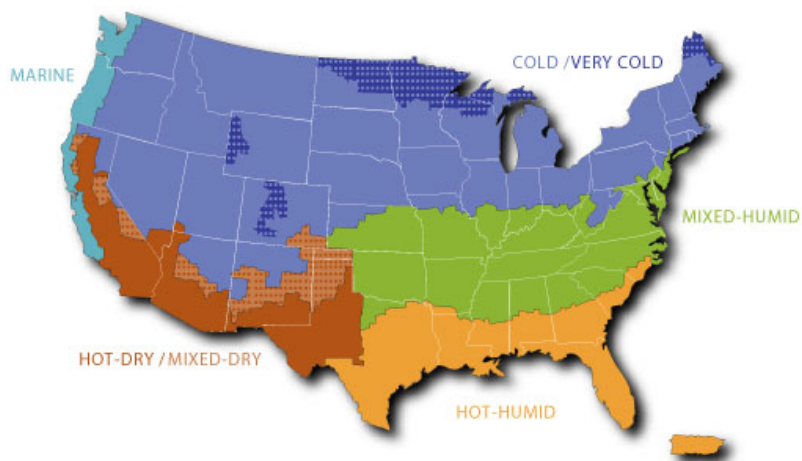
Figure 1. Building America Climate Zones (source: (Baechler et al. 2015)). The dotted areas represent additional climate zones that were merged with zones of the same color.
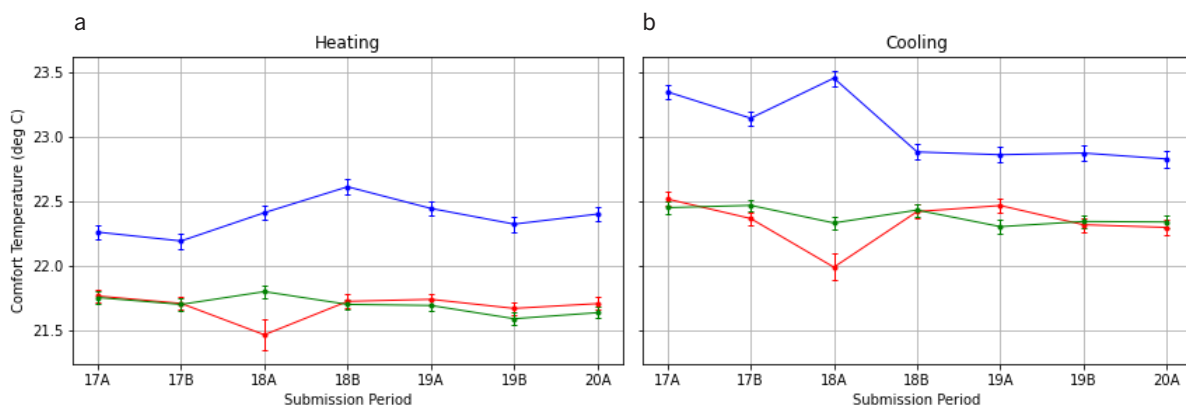

Figure 2. a) Comfort temperatures by vendor for heating and b) for cooling.

These can help shape policies and programs at both the national and regional level. At the same time, the results will inevitably raise more questions that deserve more detailed analyses.

Results

DATA OVERVIEW

Presently, 13 vendors are participating in ENERGY STAR; however only three vendors have continuously resubmitted data since the beginning of the program in 2017. This section presents long-term trends in the CT data submitted to EPA. We begin by presenting trends for comfort temperature, runtime, and core days from those three vendors. Many parts of the ENERGY STAR metric calculation are based on the concept of "core" heating and cooling days. A "core" day occurs when the HVAC system operates for more than 30 minutes in only one mode (no operation of both heating and cooling during the same day). Thus, especially mild days are excluded from calculations which improves regression fits used in the metric calculations.

COMFORT TEMPERATURES

The calculation of the ENERGY STAR metric begins by identifying each home's preferred comfort temperatures for heating and cooling. The comfort temperature for heating is defined as the 90th percentile of indoor temperature during core heating days and the comfort temperature for cooling as the 10th percentile of indoor temperature during core cooling days. The choice of 10 % was arbitrary but provides a transparent "anchor" for analysis and comparison (Roth et al. 2014). The mean comfort temperatures reported by three vendors are shown in Figure 2A for heating and Figure 2B for cooling. The period covers three and half years (that is 7 submissions). We will refer to each vendor by the graph colours, e.g., Blue, Red, Green. to preserve vendor anonymity. The error bars in the plots are the standard error of the mean for each submittal and range from 0.09–0.11 °C

Submissions at a national level conceal wide regional variations. For that reason, we present results for three selected climate zones. Most of the nation's heating energy consumption occurs in the coldest climate zones and most of the cooling in the warmest climate zones (see Figure 3 A–D). The EPA estimates that the Very-Cold/Cold and Mixed-Humid regions
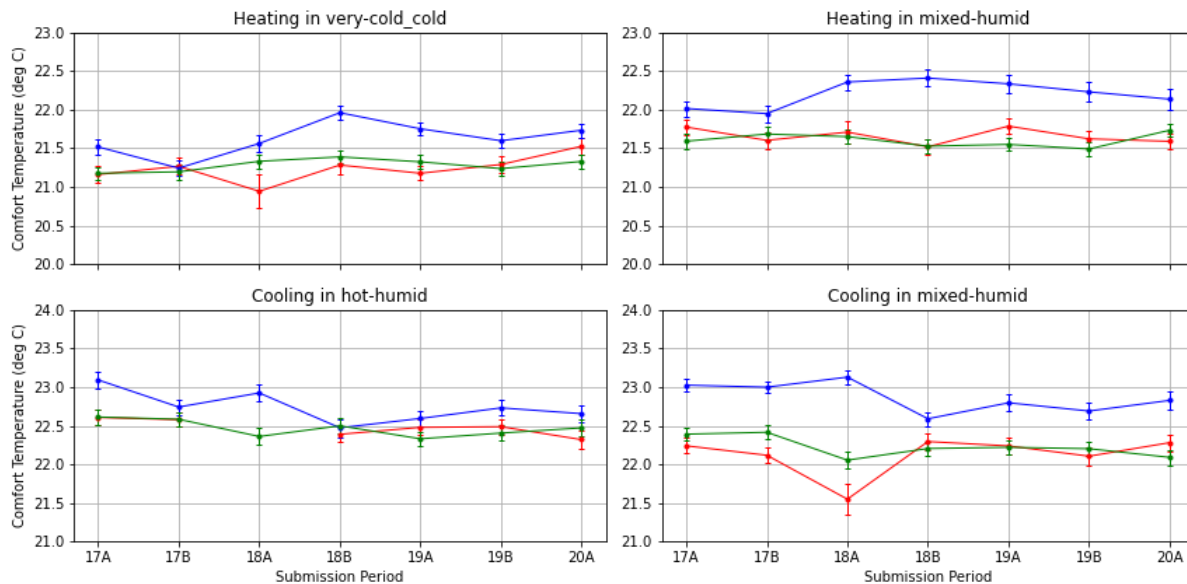
Figure 3. Long-term trends in heating and cooling comfort temperatures for three vendors in three climate zones.
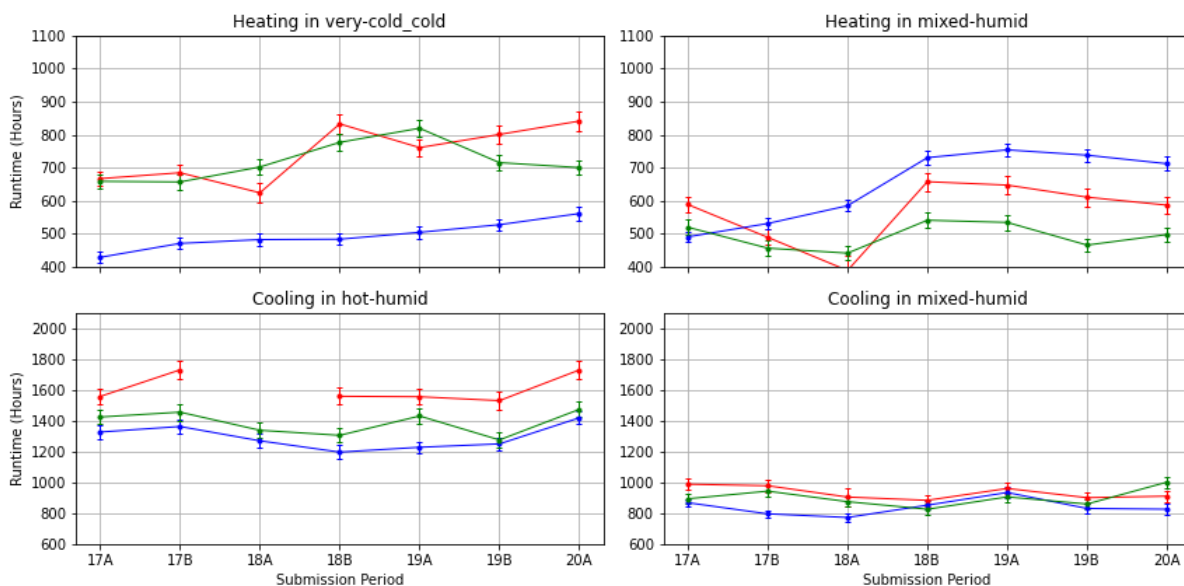

Figure 4. Long-term trends in heating and cooling runtimes for three vendors in three climate zones.

account for roughly 86 % of heating energy use and the Hot-Humid and Mixed-Humid regions account for 76 % of cooling energy use.

HVAC equipment runtimes

The reported heating and cooling runtime is defined as the primary equipment runtime during core days. This does not include heat pump emergency or auxiliary heat or secondary equipment such as stoves or room heaters. Figure 4 A–D shows the long-term trends in heating and cooling runtimes for three vendors in three climate zones. (Note truncated vertical axis.) Runtime is expressed in total hours per season (during core

days). The standard errors in the reported heating values range from 18–27 hours while the standard error in the reported cooling values range from 29–56 hours.

Core heating and cooling days

As defined previously, core days are those days with more than 30 minutes of only heating or cooling. The number of core days depends on the local climate, occupant behaviour, and, to a lesser extent, the thermostat's algorithms. Figure 5 A–D shows the number of core heating and cooling days, respectively. (Note truncated vertical axis.) The standard errors range from 2–5 days.
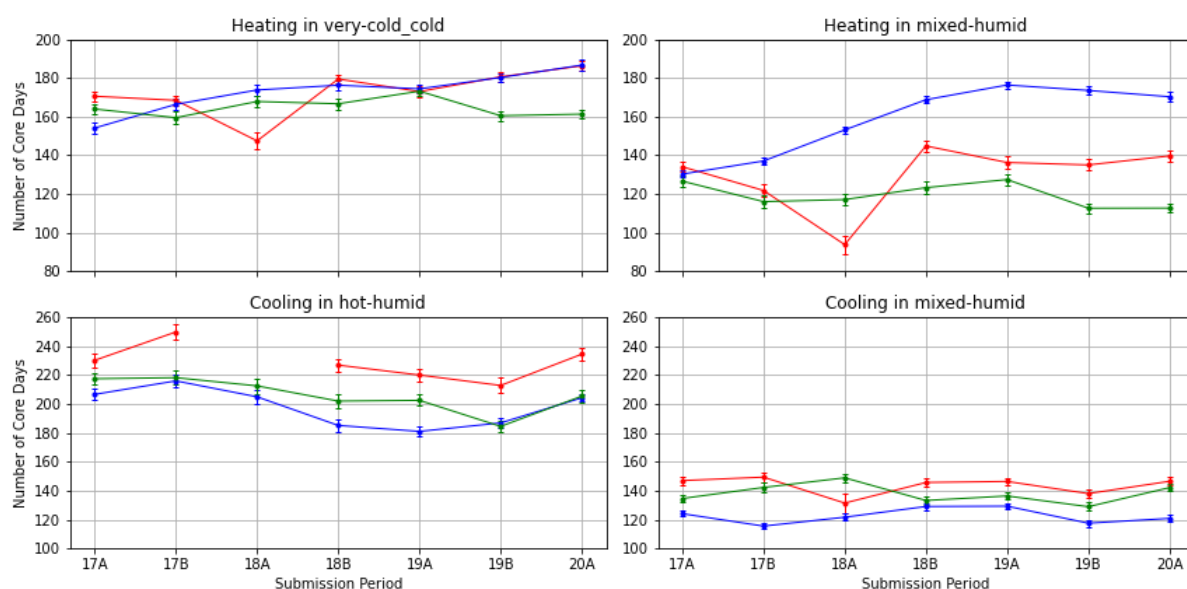
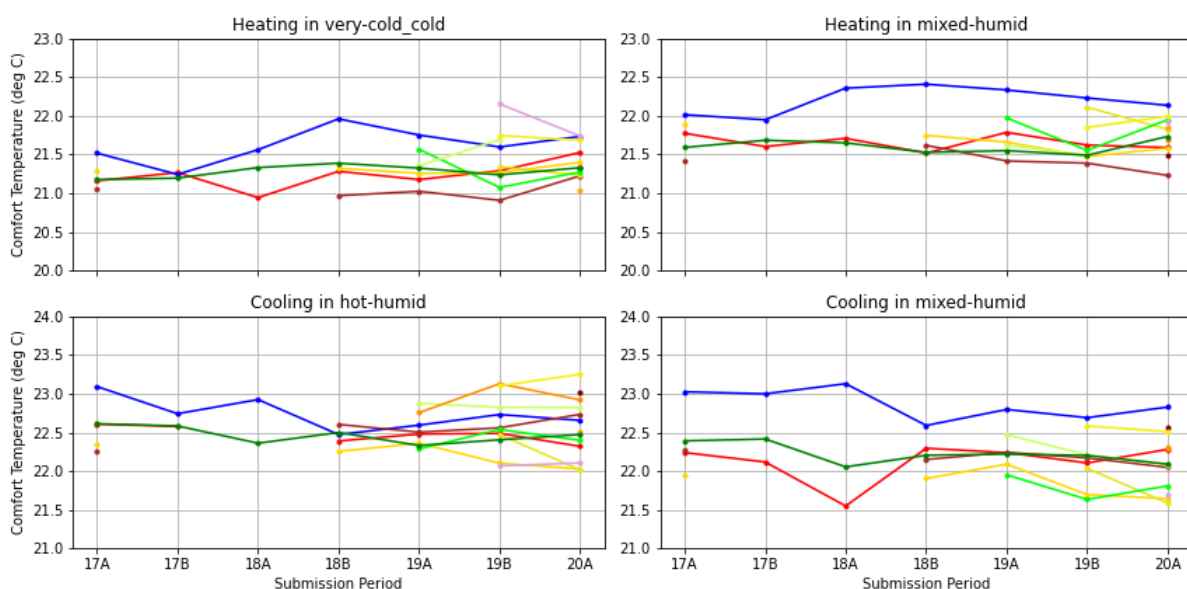*Figure 5. Long-term trends in number of core heating and cooling days for three vendors in two climate zones.*

Figure 6. Comfort temperatures for heating and cooling in key climate zones.

SUBMISSIONS BY ALL VENDORS

Over time more vendors joined the ENERGY STAR program, with 5 in July 2018, 8 in June 2019, 11 in July 2019, and by February of 2020, 13 vendors submitting data. In this section, we present results for all vendors. The submissions cover a shorter period (and sometimes barely qualify as "trends"). All results are shown for the three climate zones mentioned previously.

The comfort temperatures for heating and cooling are displayed in Figure 6 A–D. Standard errors range from 0.1–0.2 °C.

The HVAC runtimes for heating and cooling for the same climate zones are displayed in Figure 7 A–D, respectively. Standard errors range from 18–47 hours for heating runtime and range from 29–121 hours for cooling runtime.

Outside temperatures

Thermostat vendors cater to different geographical markets and might therefore have homes in different climates. Some vendors claimed that climatological bias might affect their thermostats' ENERGY STAR metric relative to others. One indicator of bias would be different outside temperatures. Statistics on the outside temperatures used for calculating the metric were first reported beginning in 2018, so only four submissions are available. Figure 8 A-D shows the average outside temperatures for core heating and cooling days in the key climate zones. The standard errors range from 0.1–0.3 °C. Because the outdoor temperature is calculated by the software using data from the national weather service and not the thermostat itself, only two
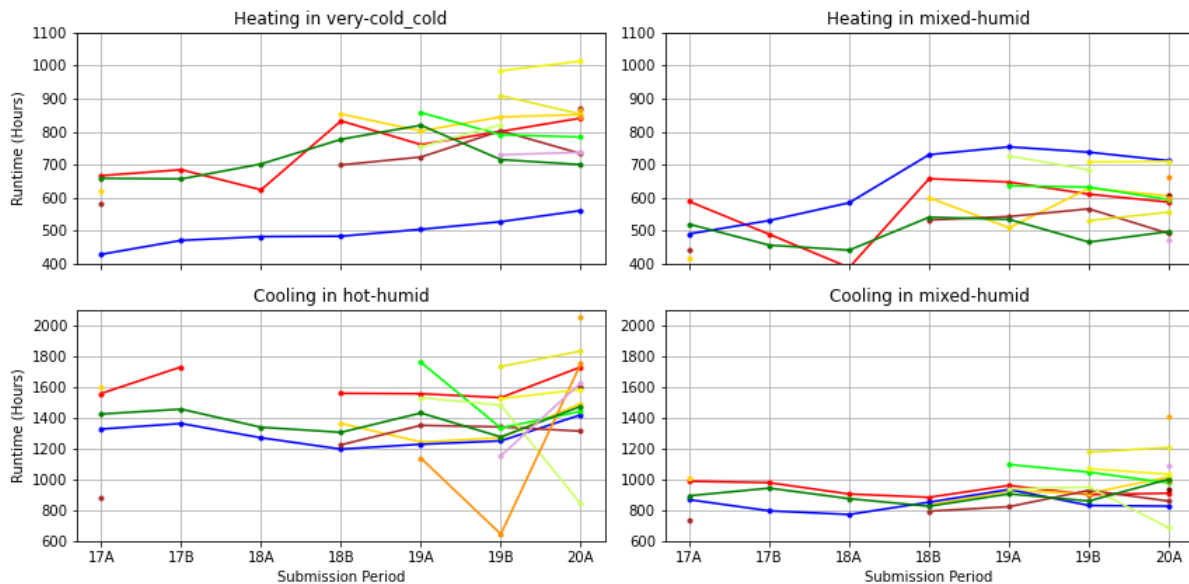
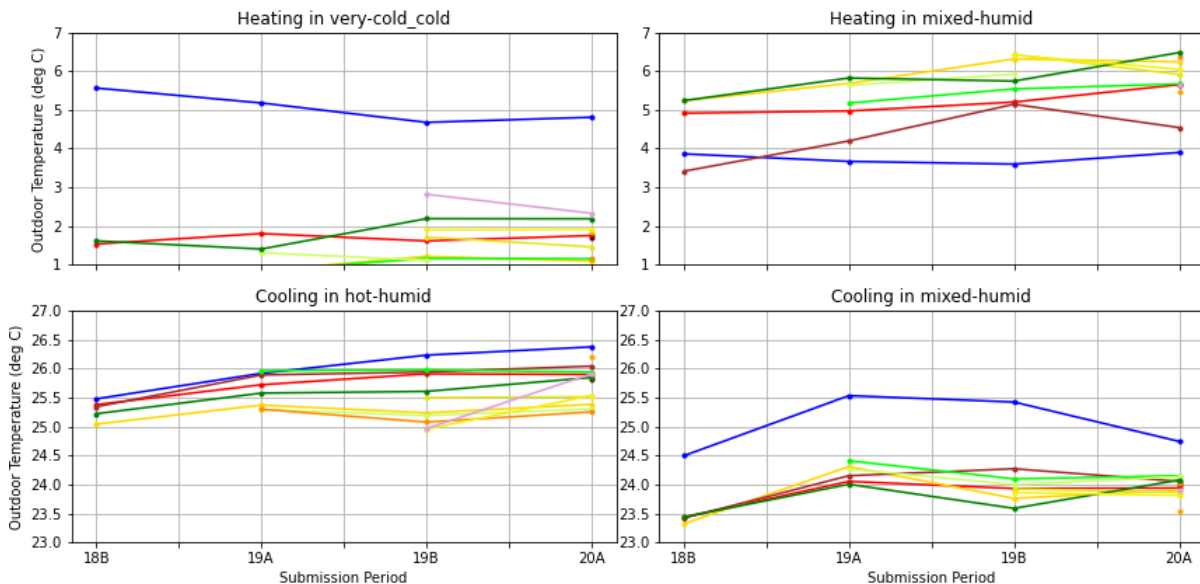*Figure 7 A–D. HVAC runtimes for three climate zones.*

Figure 8. Average outside temperatures for heating and cooling in the key climate zones.

factors can account for differences between vendors: home location, which is related to the vendor stock of homes, and the number or timing of core heating and cooling days.

Discussion

This discussion focuses on the accuracy and consistency of the submitted data. ENERGY STAR relies on the vendors to collect, process, and submit the data. Errors or differences in interpretation of procedures can occur at each stage of this process, and each presents itself differently.

For long term trends, the results from each vendor should parallel the others. The vertical displacements are most likely caused by the vendors' algorithms or their stock of homes. Situations of

concern arise when the trend lines cross each other or trend in opposite directions. These events could occur because the stock of thermostats for one vendor is changing – some of these vendors were experiencing annual growth rates above 20 % – or unusual regional weather conditions. Several examples of crossover appear in Figure 3 A–D (comfort temperatures).

This discussion initially focuses on only two vendors, Green and Red, because Blue's data are so different. The Blue data submission is discussed separately at the end.

Comfort temperatures. Red and Green's comfort temperatures (see Figures 2 and 6) for space heating were similar – roughly 21.6 °C – and were nearly constant over the submission period. Intuitively, this is a reasonable temperature (or perhaps even a little high). Even with the small standard errors (all less

than 0.1 °C) only a few submissions had values that were significantly different. This could be possibly caused by the algorithms or, alternatively, the difference could be biased temperature sensors in the thermostats (because allowable tolerance in ENERGY STAR specifications is 0.3 °C).

These trends continue for cooling comfort temperatures, although the temperatures are about 1.1 °C higher than in the winter. Again, Red and Green comfort temperatures are similar and constant.

In the last four periods, all vendors submitted data on comfort temperatures. Figure 6 A-D show that reported temperatures fluctuated less over time. For heating in very-cold climates, the dispersion in temperatures was getting smaller and converging to an average of about 21.3 °C. In a warmer, mixed-humid climate, however, the comfort temperatures were about 0.3 °C higher and the dispersion was slightly greater. Several vendors' submissions exhibited large fluctuations or had significant crossovers, suggesting that the vendors were still improving their algorithms or their stock of homes changed.

The cooling comfort temperatures were generally flat, with average temperatures at about 22.2 °C in the hot-humid climate and about 0.3 °C higher in the mixed-humid climate. There is greater variation among vendors, but less fluctuations and crossovers.

The comfort temperatures in winter and summer differ by less than 0.8 °C, suggesting that occupants make only modest seasonal adaptations in their clothing. It is also possible that many occupants maintain the same indoor temperature settings throughout the year. This cannot be verified with the available data.

HVAC runtimes. The trends in HVAC runtimes (Figures 4 and 7) are difficult to interpret. Runtimes for heating systems range from 400 to over 900 hours/year. (This corresponds roughly to an annual heating consumption of 70–90 GJ if the system's input is 30 kW). Red and Green begin with similar runtimes but then diverge and cross over – two signs of data irregularities – in the later periods.

For cooling, the runtimes range from about 700–1,800 hours/year. (This roughly corresponds to 2,100–3,600 kWh/year for a 3 kW system.) All three vendors show roughly flat cooling operation over the 3.5 year period—the truncated scale exaggerates their differences—with annual fluctuations less than 8 %.

In the last four periods, all 13 vendors submitted data on heating and cooling runtimes (Figure 7 A–D). For heating in very cold climates, we see considerable fluctuation and crossovers among the vendors. In contrast, the trends for heating in mixed-humid climates are more constant, but with greater variation among the vendors. The variations in runtime among vendors are important because runtime corresponds directly to energy consumption. Clearly some vendors—notably Green—are controlling their thermostats in ways that achieve lower energy use.

The trends in AC runtime (Figures 7c and 7D) exhibit a high degree of consistency. The trends for each vendor parallel each other and rarely cross over. Furthermore, the values are also very close. (The exception is Mustard in the hot-humid zone: its AC runtimes dropped almost 50 % from one submission to the next and then almost tripled. This may be the result of a data processing problem in that climate zone because the hiccup is not duplicated in the mixed-humid zone.) In the hot-humid climate zone, some vendors appear to be achieving consistently lower AC runtimes than other vendors and probably saving their customers electricity. The final submission runtimes averaged 700 hours for

heating and 1,300 hours for cooling, but some vendors achieved runtimes as much as 17 % below the mean. This implies lower energy consumption due in large part to more successful algorithms and control strategies by those vendors.

Many of the vendors' trends for temperature and runtime appeared to rise and dip in tandem, suggesting a response to local variations in weather conditions. Curiously, these weather variations did not appear in the outside temperature trends (Figure 8 A–D). Average temperatures may be a too crude proxy for severity of the weather conditions and requirements for heating and cooling.

Core days in heating and cooling. The calculation of comfort temperatures includes only core days, that is, with more than 30 minutes of heating or cooling. Large differences in core days among vendors might distort other metrics. However, Figure 2A and B show that the core days reported by the vendors are tightly clustered and will have negligible impact on subsequent analyses.

Differing results by climate zone. People adjust their thermostats to reflect regional differences in climate. During the winter, homes in the cold climates zone have temperatures about 0.3 °C lower than mixed-humid climates. Similarly, AC temperatures in hot-humid climate zones are slightly higher than homes in milder climates (although the results from individual vendors are more variable).

Converting runtimes into energy consumption. It is not possible to convert furnace or AC runtimes directly into energy consumption because the thermostat vendors do not have information about equipment capacities. A national survey of furnace and AC capacities would be a valuable supplement to the thermostat data and enable more accurate national estimates of HVAC energy use and savings.

The anomalous behaviour of Blue submissions. The data submitted by Blue appears to be different – sometimes widely – from the other vendors. The trends are individually plausible – which is why we displayed them – but systematically inconsistent. For example, Blue's national comfort temperatures fluctuated as much as 0.8 °C from one submission period to the next (see Figure 2B). Blue's national values for cooling comfort temperature were also surprisingly high, peaking at above 23.3 °C, both in absolute and relative terms. Its comfort temperatures in the very-cold climate zone were significantly higher than other vendors, but the runtimes were much, much lower. These trends violate simple heat loss principles (unless Blue's algorithms were much, much better than the others.) This behaviour might be explained by a unique set of homes located in an especially warm part of the very-cold climate zone but this is not likely to persist for such a long time. Finally, the outside temperatures for Blue's homes in the mixed-humid climate zone (Figure 8D) peaks at near 25.5 °C for two periods, which is unexpectedly high. Ultimately, however, we were unable to identify an explanation for – or even if there was a problem with – Blue's trends.

Conclusions

Indoor temperatures and HVAC equipment behaviour for representative samples from as many as 30 million American homes spanning four years were presented. For the first time, we have insights into the temperature preferences of millions of households, continuing for as long as four years. The procedures de-

veloped here can be extended far into the future or to observe the impact of the pandemic or sudden changes in energy prices.

Examination of long-term trends revealed aspects that were less salient in a single submission. In addition, the results from many vendors give greater confidence regarding both the values and stabilities of the collected data.

Preferred temperatures in each vendor's stock of homes were relatively constant across the years, but there were small (yet significant) differences across the vendors. Those differences (along with other control strategies) translated into lower equipment runtimes and energy savings for some vendors.

There are small but significant regional differences in temperature preferences reflecting the climates, building types, and perhaps even cultural preferences. These results demonstrate that the country should not be treated as a single entity for purposes of heating and cooling behaviours.

The approach described in this paper illustrates the challenge of evaluating future Internet-connected appliances. Laboratory tests are not adequate to capture the impact of external management from the cloud or through software changes, nor are interactions with users necessarily predictable. More and more, vendors will need to demonstrate regulatory compliance or other certification by collecting and evaluating field data. This trend requires addressing consumer privacy, protection of manufacturers' trade secrets (such as their algorithms), and security. ENERGY STAR's development of a common software package is one approach.

To be sure, some problems with data quality and interpretation remain. As mentioned earlier, some vendors had limited data science expertise prior to this program. One vendor submitted temperatures and runtimes that, while technically plausible, cannot be easily explained. Other vendors selected insufficient homes with characteristics to qualify for inclusion. There were also occasional failures in data processing. But a goal of this investigation was to identify these problems and eliminate them. Ultimately, this will create a process to rank and better quantify the savings from Internet connected thermostats.

Acknowledgement

References

Daken, Abigail, Alan Meier, and Douglas Frazee. 2016. "Do Internet-Connected Thermostats Save Energy?" In *ACEEE 2016 Summer Study on Energy Efficiency in Buildings*. Pacific Grove, Calif.: American Council for An Energy Efficient Economy (Washington, D.C.).

Evergreen Economics. 2020. "Evaluation of the California Statewide Smart Thermostat Time of Use Pilot: A Report for Pacific Gas and Electric, Southern California Edison, and San Diego Gas & Electric." Evergreen Economics.

Google Nest. 2022. "OpenTherm and What Nest Thermostats Can Do with It - Google Nest Help." Google Nest Help Centre. 2022. https://support.google.com/googlenest/answer/9259109#zippy=%2Chow-nest-thermostats-make-modulation-better-with-opentherm

Hosseinihaghighi, Seyedehrabeeh, Karthik Panchabikesan, Sanam Dabirian, Jessica Webster, Mohamed Ouf, and Ursula Eicker. 2022. "Discovering, Processing and Consolidating Housing Stock and Smart Thermostat Data in Support of Energy End-Use Mapping and Housing Retrofit Program Planning." *Sustainable Cities and Society* 78 (March): 103640. https://doi.org/10.1016/j.scs.2021.103640

Huchuk, Brent, Scott Sanner, and William O'Brien. 2019. "Comparison of Machine Learning Models for Occupancy Prediction in Residential Buildings Using Connected Thermostat Data." *Building and Environment* 160 (August): 106177. https://doi.org/10.1016/j.buildenv.2019.106177

King, Jen. 2018. "Energy Impacts of Smart Home Technologies." Washington, D.C.: American Council for an Energy Efficient Economy.

Meier, Alan, Tsuyoshi Ueno, and Marco Pritoni. 2019. "Using Data from Connected Thermostats to Track Large Power Outages in the United States." *Applied Energy* 256 (December): 113940. https://doi.org/10.1016/j.apenergy.2019.113940

Meier, Alan, Tsuyoshi Ueno, Leo Rainer, Marco Pritoni, Abigail Daken, and Dan Baldewicz. 2019. "What Can Connected Thermostats Tell Us about American Heating and Cooling Habits?" In *eceee 2019 Summer Study*. Hyères, France: European Council for an Energy Efficient Economy.

Nest Labs. 2015. "Energy Savings from the Nest Learning Thermostat: Energy Bill Analysis Results." Palo Alto, CA: Nest Labs. https://storage.googleapis.com/nest-public-downloads/press/documents/energy-savings-white-paper.pdf

Peffer, Therese, Marco Pritoni, Alan Meier, Cecilia Aragon, and Daniel Perry. 2011. "How People Use Thermostats in Homes: A Review." *Building and Environment* 46 (12): 2529–41. https://doi.org/16/j.buildenv.2011.06.002

Roth, Kurt, Bryan Urban, Victoria Shmakova, and Brian Lim. 2014. "Residential Consumer Electronics Energy Consumption in 2013." In *ACEEE Summer Study on Energy Efficiency in Buildings*. Pacific Grove, CA: American Council for An Energy Efficient Economy (Washington, D.C.).

U.S. EPA. 2022. "ENERGY STAR Certified Smart Thermostats." January 1, 2022. https://www.energystar.gov/productfinder/product/certified-connected-thermostats/results

Vellei, Marika, Simon Martinez, and Jérôme Le Dréau. 2021. "Agent-Based Stochastic Model of Thermostat Adjustments: A Demand Response Application." *Energy and Buildings* 238 (February). https://doi.org/10.1016/j.enbuild.2021.110846

Wang, Zhe, and Tianzhen Hong. 2020. "Learning Occupants' Indoor Comfort Temperature through a Bayesian Inference Approach for Office Buildings in United States." *Renewable and Sustainable Energy Reviews* 119 (March): 109593. https://doi.org/10.1016/j.rser.2019.109593

Woods, James. 2006. "Fiddling with Thermostats." In *2006 ACEEE Summer Study on Energy Efficiency in Buildings*, 10. Pacific Grove, Calif.: American Council for An Energy Efficient Economy (Washington, D.C.).