Do Connected Thermostats Save Energy?

Abigail A. Daken, United States Environmental Protection Agency
Alan K. Meier, Lawrence Berkeley National Lab
Douglas G. Frazee, ICF International

ABSTRACT

Connected thermostats (CTs) manage HVAC systems in over four million homes. Widely varying strategies are used by these thermostats to reduce HVAC energy use. Thermostat vendors claim savings of up to 20%; however, there is no accepted procedure to evaluate the effectiveness of these strategies. Presently, consumers (and utilities) have no way to identify the most effective CT products. We developed a method to quantify HVAC energy savings from a CT and assign a savings metric to CT products based on the method. The method collects indoor temperature and HVAC run time data from thermostats, plus publicly available local weather data. Temperature data is then regressed against HVAC run time to develop a unique HVAC-thermal model for each home. CT savings are expressed as percentage HVAC run time reduction from that with an assumed constant temperature baseline. To assign a metric value to a product (hardware plus service), savings from a large number of homes using the product are aggregated via a specific procedure. The method is being tested on large groups of thermostats from several vendors. Many of the strengths and weaknesses of this approach have been identified and will be discussed, along with anticipated future improvement of the method.

Introduction and Background

The residential thermostat is arguably the most important device controlling home energy use. Together, residential thermostats control nearly 10% of national energy use. (Peffer et al. 2011) A correctly programmed thermostat can easily reduce a home's heating and cooling energy over 20% compared to constant temperatures and 14% for the average home. (Moon and Han 2011; Sanchez et al. 2008) For these reasons, the Environmental Protection Agency (EPA) became interested in ENERGY STAR™ labeling of advanced thermostats. At first, EPA focused on programmable thermostats, establishing specifications that required certain features and performance. In addition, the specification required that the thermostats be pre-programmed with a schedule of temperatures that were likely to save heating and cooling energy. ENERGY STAR labeling almost certainly accelerated the widespread adoption of higher-quality, programmable thermostats.

Over time, however, it became clear that the energy-saving potential of programmable thermostats was not being realized. (Nevius and Pigg 2000) Field research found that consumers disabled many of the energy-saving features, and on the whole, the presence of a device that *could* enable savings was poorly correlated with actual energy bill reductions. As a result, in 2009, EPA terminated ENERGY STAR labeling for thermostats until it could develop a means of ensuring higher utilization of the energy-saving features. EPA explored several strategies to encourage higher consumer use of energy-saving features, but ultimately none of these approaches proved feasible.

Around 2012, several thermostat manufacturers began offering devices with Internet connectivity and other means of communicating with devices beyond the furnace and AC, such

as smart phones and home energy management systems. The Internet connection enabled entirely new approaches to controlling home temperatures. First, residents could remotely control their thermostats through a web portal and their mobile phones. Second, thermostat manufacturers could track and help manage temperatures through the Internet connection. This feature opened up many new energy-saving opportunities that skirted problems with conventional thermostats. It also changed the thermostat from a relatively boring device on the wall to one of the first residential applications of the Internet of Things (connected devices), Big Data (analysis of large amounts of data to yield new insights) and Software as a Service (users interact with software not as a single purchase but as a service hosted elsewhere which they interact with as needed).

The market grew rapidly so that by 2013, roughly two million Internet-connected thermostat were sold, increasing to four million in 2015; a growth rate of roughly 25% per year (Tweed 2015). This allowed EPA to focus its program on connected thermostats (CTs). These products allow visibility into a key gap in previous efforts to label thermostats: how they are used in the home. The thermostat service provider generally has access to data that is indicative of how the thermostat is being used in the homes of their subscribers.

Thermostats have never easily fit within EPA's conventional labeling framework. A typical ENERGY STAR product specification begins with the assumption that a product's energy efficiency can be measured with a test procedure performed in a test laboratory. Typically, the products with an energy efficiency in the top 25% of the category will earn the label. This framework works well for furnaces and air conditioners but not the device that controls them. To be sure, some performance characteristics of the thermostat can be tested, such as temperature sensitivity, but these don't significantly affect the fundamental heating and cooling operation. The energy savings due to CTs are even more difficult to assess and no energy test procedures exist today.

This paper describes EPA's progress towards developing an ENERGY STAR specification for CTs. It focuses on the technical problem of assessing the ability of CTs to save energy.

Goals of the Metric and Program

Every ENERGY STAR program relies on a "metric" to express energy performance or savings. Metrics are typically expressed in kWh/year, energy factor (EF), EER, or similar, and are based on a recognized energy test procedure. For CTs, however, no test procedure exists and a traditional approach, e.g., laboratory test, is not feasible. A metric must nevertheless adhere to Energy Star's principles: namely, it must be technology neutral, fair and transparent, able to assure consumers of an acceptable payback time, and the metric must be obtained through a procedure that is fast and reasonably easy to determine in order to keep pace with changing technologies. Ideally, it would allow testing of a broad range of thermostat products, including the variety of unique combinations of hardware, software, and services. On one end of the spectrum, some businesses sell little more than the Internet-connected thermostats (pure hardware). On the other end of the spectrum, businesses sell only the web services that a consumer needs to manage the thermostat that he or she obtains elsewhere. The product to test, then, should be understood as a combination of hardware, software, and service.

How Connected Thermostats Save Energy

Before developing a metric, it is also important to understand how CTs can reduce heating and cooling energy consumption. In its simplest form, a CT saves energy by reducing HVAC run time compared to what would have occurred with a conventional[1] thermostat. The primary means for a CT to do this is by minimizing the inside-outside temperature difference. In the winter, this means lowering the average inside temperature and, in the summer, raising it. The plots below demonstrate this savings in a representative way.
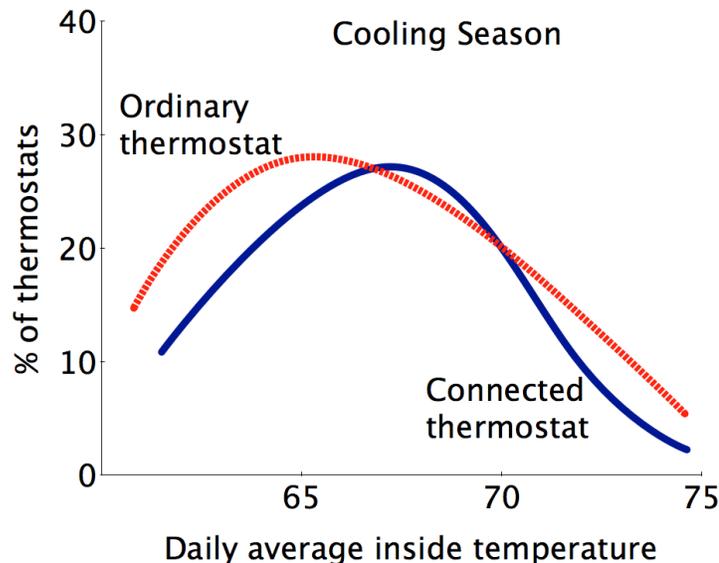
Figure 1: A conceptual illustration of how connected thermostats (CTs) save energy in heating season and cooling season, by influencing the daily average indoor temperature.

However, the CT must manage temperatures such that occupants do not experience thermal discomfort. Vendors often employ cloud-based algorithms to achieve these goals. Techniques employed by a CTs during the heating season include:

- allowing temperatures to begin floating downwards slightly before programmed time
- optimizing morning recovery from setback
- optimizing HVAC control for weather conditions
- lowering temperatures when occupants are home and/or when away[2]
- minimizing use of electric resistance auxiliary heating for homes with heat pumps

---

[1] Being interested in aggregated savings, we use "conventional thermostat" to refer to the mix of manual and programmable thermostats that are used in US households.

[2] Some CTs employ sensors to detect vacancy or occupancy (depending on the system). This data stream was not used in the analysis because not all products employed the sensors. Some CTs also capture humidity data; this data stream was also not used in the analyses.

Similar strategies are employed during the cooling season. Thus, the primary means of saving energy is by closing the gap between indoor and outdoor temperatures. A metric must capture the CT's proficiency in managing temperatures to achieve energy savings.

Data Available to Determine a Metric

The obvious source of performance data are from the utility electricity and gas meters. Unfortunately, CT vendors cannot reliably obtain utility data and utilities cannot reliably obtain CT data. Third-party verification entities have great difficulty obtaining both data sets. In the ENERGY STAR program, EPA works with the providers of products - in this case, the providers of CTs. The data common to all CTs and available to CT service providers are:

- thermostat set points (every 30 minutes or less)
- indoor temperatures (every 30 minutes or less)
- run-time of controlled HVAC equipment
- if the unit is a heat pump
- geographic location
- outside temperatures from nearby weather stations

Other data collected by some CTs include: occupancy, humidity, multiple inside temperatures (and set points), and additional HVAC modes. None of these were included in our analyses. A fragment of a typical data stream is shown in Figure 2, illustrating hourly temperatures, set points, and furnace run times. The daily operating patterns are evident. There are also frequent divergences between the set points and the inside temperatures (in both directions), presumably caused by appliance gains, solar gain, and thermal mass.
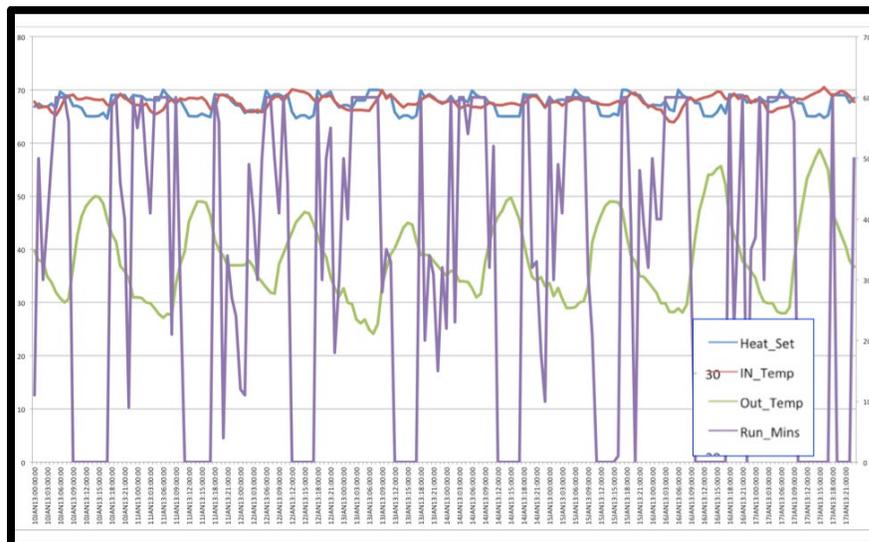

Figure 2: 10 days of the data stream from week in January for a connected thermostat located in Central California, illustrating the density and variety of data.

It is also important to understand the data that are not widely available, including thermostat settings, indoor temperatures and run times prior to CT installation; electricity and gas consumption both prior and subsequent to CT installation; or the capacity of the heating and cooling systems and other technical features of the system, beyond what can be determined by the thermostat wiring & settings.

Many techniques of extracting building thermal performance metrics from energy data have been developed. The best known is perhaps PRISM (Fels and Goldberg 1986), which compares metered energy consumption to outdoor daily average temperatures in order to derive thermal parameters. Field studies beginning with PRISM in the 1970s have found a roughly linear relationship between a home's heating or cooling energy use and ΔT, the differential between inside and outside temperatures (Fels 1986). The linear relationship includes a temperature offset ("free heat") expressing the tendency of the home to warm by several degrees even in the absence of active heating. This rise is driven by body heat, heat from appliances, and solar gain. Run time for single-speed HVAC exhibits a similar linear relationship to ΔT. The goal of modeling each thermostat is to derive the unique linear relationship for each home.

The CT data sets are both richer and poorer than those required by PRISM. The CT data sets include set points and room temperatures at least every hour. In contrast, PRISM and similar models typically assume constant inside temperatures or some sort of variable degree-day base. In the case of CTs, the variation of inside temperatures arising from hour-by-hour management of set points are in fact the independent variable. PRISM uses metered energy data as an input, which is not generally available to CT providers. At hourly data rates, thermal mass effects and random fluctuations are readily apparent, while they are smoothed out or invisible in PRISM analyses. Finally, PRISM treats energy use for the whole home and must include a term to account for energy consumption not dependent on outside temperature. No equivalent use term is needed for a CT since all HVAC run time affects inside temperature. Furthermore, PRISM-type analyses cannot always distinguish between heating and cooling energy; with CTs the run times for heating and cooling are tracked independently. We concluded that PRISM (or its successors) could not be easily adapted for the derivation of a CT performance metric.

Choosing a Baseline

The CT performance metric must be directly related to energy savings. However, any estimate of energy savings first requires identification of a baseline. In broad terms, the baseline is the home with a conventional thermostat instead of the CT. Since these thermostats do not track temperature data, it is difficult to measure the baseline directly. In practice, several ways of estimating this baseline are possible, depending on data availability and program goals. Possible baselines include:

- A constant indoor temperature or constant thermostat set point, or a schedule of thermostat set points, determined without reference to the particular home
- A run time derived from population energy use data, such as DOE technical support documents or EIA data, without reference to temperature choices

- Data from a "test period" when the CT's features are switched off and on (to create an A/B test)[3]
- Analysis of CT data to infer comfort preferences that were likely relevant both before and after the CT was installed

Each of these baselines is a compromise between simplicity and realism. A constant temperature or a constant setpoint is simple to implement; however, field studies have illustrated variance in comfort preferences between households and between regions. In contrast; individual, per-home baselines derived from comfort preferences captured by CTs may more accurately represent the variance of baseline conditions across households. The performance metrics described below rely on different baselines and have their respective strengths and weaknesses.
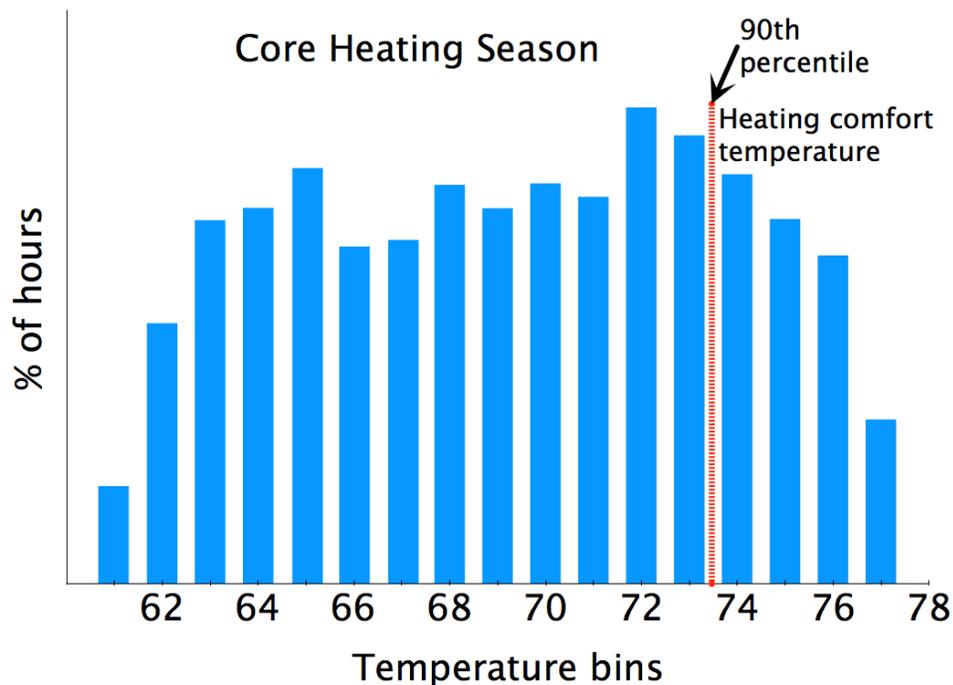

Figure 3: A conceptual illustration of the selection of 90th percentile set temperature in heating season as a baseline indoor temperature, and how it would compare to a years' worth of indoor temperatures.

EPA explored one particular baseline in detail, assuming persistent use of comfort temperatures as the baseline condition. These individual, per-home comfort temperatures are extracted from the CT set point history based on the method of Urban and Roth (2014):

---

[3] A baseline derived from periods when the CT features are switched off is analytically attractive but introduces numerous technical and behavioral interactions. For example, most thermostats have algorithms that "learn" the occupants' behavior, so the weeks subsequent to a baseline period would mostly reflect the transitional period while the algorithms seek to re-learn occupant behavior and re-optimize operating schedules, rather than the CT's ultimate performance.

Heating Comfort Temperature - During the core heating season (days with > 1 hour of heating run time), the 90th percentile of the set point history is used as the preferred heating comfort temperature for the home.

Cooling Comfort Temperature - Similarly, the 10th percentile of the setpoint history in the core cooling season is used as the preferred cooling comfort temperature.

See Figure 3 for a graphical explanation of these points.

Performance Metrics for Connected Thermostats

Based on the constraints outlined above, two approaches to evaluating the performance of CTs and calculating a metric were identified:

1. Savings Degree-Hours
2. HVAC run time

These two approaches are introduced below and their merits discussed.

The Savings Degree-Hours (SDH) Metric

The "Savings Degree-Hours" metric seeks to capture the extent to which thermostat management maximizes the difference between the measured indoor temperature and an arbitrary reference temperature. It compares the history of indoor temperatures to the reference temperature, multiplying the temperature difference by the number of hours the difference exists, then summing:

$$SDH = \sum_{each\ heating\ hour} (T_{ref} - T_{obs})$$

Where $T_{ref} = reference\ temperature$
$T_{obs} = observed\ indoor\ temperature$

The primary advantage of a SDH metric is simplicity. It seeks to rank products by the numerical quantity of accumulated SDHs over a period of time, as higher SDHs equate to less energy use. This strength, however, is also its largest weakness: there is no obvious way to assess the magnitude of energy or consumer bill savings. [4]

HVAC Run Time Metric

This metric seeks to measure the reduction in equipment run time resulting from better thermostat management. It compares the actual run time of the controlled HVAC equipment to a baseline run time, expressing the savings as % run time reduction.

---

[4] There are situations in which savings from improved control of HVAC equipment are not captured by SDH metric. These include situations where the vendor reduces the amount of auxiliary heat (e.g. avoiding the use of electric resistance heat as backup to a heat pump) or from prompting residents to open windows and shut off the AC when it is cool outside.

$$RS = (R_{baseline} - R_{obs})/R_{baseline} \; x \; 100\%$$

Where $R_{baseline} = run\ time\ that\ would\ have\ occurred\ without\ CT$
                     $R_{obs} = observed\ run\ time\ with\ CT$
                     $RS = \%\ reduction\ in\ run\ time\ due\ to\ the\ CT$

The primary advantage of the HVAC run time metric is that it is closely tied to energy and cost savings – a given percent reduction in run time is clearly related to percent HVAC energy use reduction (at least for single speed equipment).[5] In addition, the metric is potentially capable of capturing savings from a variety of strategies, not just from more energy conserving set points. However, estimating the baseline run time (in the absence of the connected thermostat) is not straightforward.

The hybrid temperature-run time metric

Based on stakeholder and expert input, EPA selected a hybrid run time approach. The ultimate goal of the metric is to characterize the energy performance of the entire US deployment of a CT model. Thus, the metric must capture both the energy performance of individual thermostats and the mean performance of a representative sample of the entire US deployed population. As with previous methods, a metric is first calculated for an individual thermostat, then these results are aggregated over a large sample of homes.

The Hybrid Metric for a Single Home

Generating savings metric scores for a particular home is a three step process:
1. *Develop the Home's Thermal/HVAC Model* - construct a model of the relationship between HVAC run time, outside temperature, and temperature choices in the home;
2. *Calculate the Baseline Run Times* - use that model to calculate baseline heating and cooling run times; i.e. what run times would have been under baseline temperature conditions;
3. *Calculate Savings* - output savings as % run time reduction of actual vs modeled.

Develop the home's thermal/HVAC model. We explored in detail a model where varying inside temperature explains energy savings. The calculation summarizes data daily, as this seemed to be the most robust. The model is similar to previous models, using a simple linear relationship with a balance temperature:

$$RT_c = \left((\Delta T - \tau_c)/\alpha_c\right)_+$$
$$RT_h = \left((\Delta T - \tau_h)/\alpha_h\right)_+$$

Where $(\quad)_+$ indicates that the term is zero if its value would be negative

---

[5] Special treatment will be required for variable capacity equipment because run time does not directly correlate with energy consumption.

τ_h = the reference temperature *difference* of the home for heating, which is maintained without use of heating equipment

τ_c = the reference temperature *difference* of the home for cooling , which is maintained without use of cooling equipment

ΔT = indoor minus outdoor temperature = $T_{in} - T_{out}$

T_{in} = indoor temperature reported by the CT

T_{out} = outdoor temperature reported by the closest NOAA weather station

α_h = the responsiveness of the home to heating equipment run time

α_c = the responsiveness of the home to cooling equipment run time

Note that ΔT as defined will be negative for most of the cooling season. τ is the reference ΔT which would result in the absence of running the thermostatically controlled heating and cooling equipment, reflecting "free heat" from solar gain and from activities and appliances in the home. Because of its physical cause, τ is expected to typically fall in the range of 5 to 15°F. It will be different for heating and for cooling.
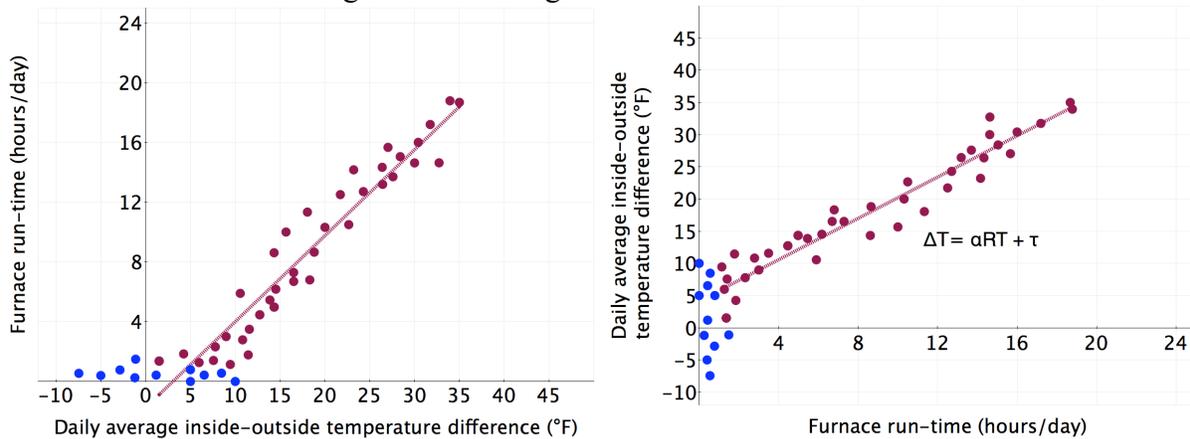

Figure 4: A conceptual illustration of how this method would be used with field data for the heating season: A linear regression of ΔT vs. RT using only those days that have more than 60 minutes of heating equipment run time, and no cooling equipment run time.

In order to calculate the fit, the equation is reversed and a straightforward linear regression of ΔT as a function of RT is performed. The regression is limited to "core cooling/heating days" by excluding points with less than one hour of HVAC run time per day, or with both heating and cooling run time.

$$\Delta T = \alpha_c \cdot RT + \tau_c \text{ (core cooling days)} \qquad \Delta T = \alpha_h \cdot RT + \tau_h \text{ (core heating days)}$$

Calculate baseline run time. Baseline run times are calculated by first calculating ΔT in the baseline condition, then calculating baseline run times from that. Run times will be longer to the extent that ΔT is larger in the baseline condition.

$$\Delta T^{base,h} = T_{in}^{base,h} - T_{out}$$
$$\Delta T^{base,c} = T_{in}^{base,c} - T_{out}$$

$$HRT^{base} = \left((\Delta T^{base,h} - \tau_h)/\alpha_h\right)_+$$
$$CRT^{base} = \left((\Delta T^{base,c} - \tau_c)/\alpha_c\right)_+$$

Where $T_{in}^{base,h}$ is the indoor temperature in the heating season baseline condition

$T_{in}^{base,c}$ is the indoor temperature in the cooling season baseline condition

HRT^{base} is the heating run time in the period of study for the baseline indoor temperatures, as estimated using the previously derived model

CRT^{base} is the cooling run time in the period of study for the baseline indoor temperatures, as estimated using the previously derived model

Calculate savings. Savings are calculated in percent run time reduction, with the following formulas:

$$HS = 100 \times (HRT^{base} - HRT^{obs})/HRT^{base}$$
$$CS = 100 \times (CRT^{base} - CRT^{obs})/CRT^{base}$$

where HRT^{obs} is the observed heating run time in the period of study

CRT^{obs} is the observed cooling run time in the period of study

HS is the heating savings metric (in %)

CS is the cooling savings metric (in %)

It should be noted that EPA does not consider the result predictive of savings for any individual thermostat installation. There are a number of reasons that the results could vary in any given installation, and the results are only indicative of performance when averaged over a large sample of homes for which these factors tend to average out.

Aggregation

In other analyses such as PRISM and its descendants, the savings calculation for one home or a group of homes relies on a pre-retrofit data or a group of homes that received no treatment. For CTs, there is generally no similar data available from a control group of homes without CTs. Nevertheless, the homes receiving CTs represent a very large – tens of thousands—and diverse set of homes. Factors that affect metric scores for specific homes can be expected to vary randomly, including:

- Occupancy patterns
- Household response to the particular strategies used by the product to achieve savings. For instance, turning off or on particular features, response to behavioral prompts, etc.
- Changes in occupancy patterns or thermal needs of residents (e.g. due to illness)
- Variations in solar gain over the course of a season

Thus, aggregating scores over a large sample of homes will better reflect the capabilities of the product. However, savings scores vary widely from climate to climate, largely because a

1°F change in ΔT will equate to a large percentage of ΔT in mild climates and a small one in more extreme climates. Furthermore, different vendors may have very different distribution of subscribers across climate zones. To fairly represent savings of all vendors, it is likely necessary to collect scores aggregated within climate regions. The Energy Information Administration (EIA) climate zones [https://www.eia.gov/consumption/residential/maps.cfm] provide a convenient broad-brush distinction, and allow comparison of scores with public data from EIA and RECS.

Discussion

In this paper we outlined a procedure to calculate a hybrid performance metric of connected thermostats (CTs) using data from the installed base of thermostats. It can be used with any baseline average daily indoor temperature. We explored one such baseline in some detail: a constant comfort temperature, with the comfort temperature derived from analysis of each home's set point history. This baseline partially corrects for variation in user population between products which would otherwise tend to skew metric scores. However, savings from products that successfully encourage more energy-saving comfort set points will not be captured. A regional baseline would capture these savings, but unless the regions are small enough, may also introduce bias between products based on their geographical spread of deployments. While public data for a highly granular regional baseline do not exist, CT service provider data could themselves be used to develop such a baseline. Regional baselines developed from this data may not reflect the true consumer savings from purchasing a CT compared to other types of thermostats. In future work, EPA intends to investigate the implications of using regional baselines.

One fundamental issue with the hybrid run time metric is that it deals poorly with HVAC systems where energy use is not roughly proportional to run time. Currently, true variable capacity systems are a small percentage of installations, though we expect their popularity to rise; already staged systems are more common. The metric might be modified for staged systems by weighting run time with an estimated proportion of energy use by each stage. Truly variable capacity systems may be able to provide estimated energy consumption information directly. Heat pump systems with electric resistance auxiliary heat are similar to staged systems. However, the metric as described might allow adequate comparison of CTs controlling these systems, and we look forward to exploring the results with stakeholders in the near future.

Another challenge is accounting for float - the tendency for homes to be warmer than the average set temperature in the heating season or cooler than the average set temperature in the cooling season, particularly in shoulder seasons. High solar gain and opening windows during cool mornings may cause this kind of temperature variation. Generally, it results in high comfort with low energy use. Fortunately, most HVAC run time occurs on days when float is not an issue, and indoor temperatures track set temperatures closely. We plan to evaluate the implications of ignoring float on our results.

The hybrid metric does not account for saving strategies that are not reflected in indoor temperature, e.g. humidity control, air movement, and ventilation. A run time reduction metric using a baseline purely referring to HVAC run time could capture these saving. Such a baseline is complicated significantly by home to home variation in HVAC sizing relative to heating and cooling loads, a datum that is invisible to the thermostat.

Conclusions

We have presented a first effort to quantify savings using the wealth of data available from connected thermostats installed in homes. We look forward to the evolution of such methods to take full advantage of that data, including hourly information. Our hope is that engagement of CT providers with EPA in the context of the ENERGY STAR program will encourage data sharing and openness and hasten the advent of such methods.

We are still unable to unequivocally answer the question posed in this paper's title, "Do connected thermostats save energy?" However, we described many of the steps and procedures that will be used as the data become available. We expect that, in the next few years, the answer will become clear. Moreover, we expect that it will be possible to identify which vendor's products save the most energy.

Acknowledgments

References

Fels, M. F., and M. L. Goldberg. 1986. "Using the Scorekeeping Approach to Monitor Aggregate Energy Conservation." *Energy and Buildings* 9 (1–2): 161–68. doi:10.1016/0378-7788(86)90017-4.

Fels, M. F. 1986. "PRISM: An Introduction." Energy and Buildings 9 (1-2): 5–18.

Moon, J. W., and S. Han. 2011. "Thermostat Strategies Impact on Energy Consumption in Residential Buildings." *Energy and Buildings* 43 (2–3): 338–46. doi:10.1016/j.enbuild.2010.09.024.

Peffer, T., M. Pritoni, A. Meier, C. Aragon, and D. Perry. 2011. "How People Use Thermostats in Homes: A Review." *Building and Environment* 46 (12): 2529–41. doi:16/j.buildenv.2011.06.002.

Tweed, K.. 2015. "Smart Thermostats Begin to Dominate the Market in 2015." *Greentech Media*. July 22. http://www.greentechmedia.com/articles/read/smart-thermostats-start-to-dominate-the-market-in-2015.

Urban, B., and K. W. Roth. 2014. "A Data-Driven Framework For Comparing Residential Thermostat Energy Performance." Cambridge (MA): Fraunhofer Center for Sustainable Energy Systems.