Fraunhofer Center for Sustainable Energy Systems

A DATA-DRIVEN FRAMEWORK FOR COMPARING RESIDENTIAL THERMOSTAT ENERGY PERFORMANCE

DRAFT FINAL REPORT | CO-DEVELOPED WITH NEST
JULY 2014

by Bryan Urban and Kurt Roth

PI
Bryan Urban, Building Energy Efficiency Group
burban@fraunhofer.org 617-714-6514

Disclaimer

This report was commissioned by Nest Labs on terms specifically limiting Fraunhofer USA's liability. Our conclusions are the results of the exercise of our best professional judgment, based in part upon information provided to us by Nest Labs. Use of this report by any third party for whatever purposes should not, and does not, absolve such third party from using due diligence in verifying the report's contents.

Any use which a third party makes of this document, or any reliance on it, or decisions to be made based on it, are the responsibility of such third party. Fraunhofer USA accepts no duty of care or liability of any kind whatsoever to any such third party, and no responsibility for damages, if any, suffered by any third party as a result of decisions made, or not made, or actions taken, or not taken, based on this document.

This report may be reproduced only in its entirety.

1 Introduction

Thermostats control the thermal environment of building spaces and often include features that can save energy if properly implemented. Historically, the mere presence of these features has not been sufficient to guarantee savings, in part because of differences in human behavior and usage. Prior attempts to understand, characterize, and generalize thermostat energy performance have met with varying degrees of success, and savings estimates vary considerably (Nevius and Pigg 2000, RLW 2007, Peffer et al. 2011, Sachs et al. 2012). As a result, there is confusion among industry stakeholders – consumers, utilities, energy efficiency program managers, and standards bodies – about the energy savings that different types of thermostats actually deliver. Critically, the lack of an industry-accepted performance assessment methodology makes it difficult to effectively compare thermostat products based on their expected real-world energy performance.

To address these shortcomings, we present a new data-driven framework for comparing thermostat energy performance. This report is organized in three main sections: Background and Motivation, Assessment Framework, and a Data-Driven Example. To provide context, Section 2 describes the energy-saving pathways of different thermostat features and identifies key challenges associated with existing methods. To overcome these challenges, we propose a new framework in Section 3 that aims to fairly and consistently assess energy performance, based in part on user data collected from communicating thermostats. Because parts of the method are fairly new, we present an illustrative example in Section 4 based on data from existing thermostat users.

1.1 Goals, Guiding Principles, and Scope

Before presenting this new energy performance assessment framework, we must define the goals, scope, and guiding principles that led to its development.

1.1.1 Goals

The goal of this thermostat assessment framework is to produce accurate, consistent, representative, and verifiable energy performance comparisons. Assessments should not unfairly burden manufacturers or interfere with innovation. Since new features are continually being developed, it is important to make the framework flexible and expandable to include new savings mechanisms as they are invented.

1.1.2 Guiding Principles

The primary function of a thermostat is to allow occupants to control their thermal environment, and by doing so, to help occupants achieve their preferred balance between thermal comfort and energy consumption. We believe that energy savings should be realized without compromising a household’s ability to achieve a comfortable environment. Several studies suggest that most thermostat users are satisfied with their ability to achieve comfort, even if it comes at the expense of energy savings (Peffer et al. 2011, Sachs et al. 2012). Consequently, we do not attempt to evaluate thermal comfort explicitly for individual thermostats. The framework we propose, however, does consider that individual households have subjective preferences about thermal comfort.

Energy performance depends on many factors, including climate, building characteristics, HVAC systems, and lifestyle. Specific features of a thermostat may only apply to a subset of users. For example, some HVAC control features apply only to heat pumps, so homes that do not own or plan to own heat pumps would not realize any energy performance gains from those features. Similarly, households that are vacant during a large portion of the day have a different savings potential (through setbacks and setups) than do households that are occupied most of the day. By taking into account the most important factors, energy performance can be assessed for each category of end-users. Results can then be

averaged (weighted by occupancy profiles, building characteristics, HVAC system types, and climate region) to identify typical savings across the entire population.

1.1.3 Scope

The scope of this document is limited to comparing the energy performance of thermostats that control central HVAC systems in residential buildings. Thermostats in commercial buildings may have different features, control different types of HVAC systems, and can be used in different ways. Specifically, this work focuses on comparing the energy performance of thermostat products relative to a specified baseline. This framework was not developed to meet the particular requirements of utility measurement and verification protocols, which have many constraints and tend to evaluate "energy savings" relative to historic energy consumption. Instead, this framework is intended to assess thermostat "energy performance", defined as the expected difference in actual HVAC energy consumption relative to a specified baseline. Historic energy consumption is just one possible baseline. Nevertheless, the proposed framework may be applicable to other building types and could be adapted for other kinds of assessments.

Thermostats could further influence household energy consumption in ways that are not directly related to HVAC system control, for example, by also controlling non-HVAC related equipment (e.g., lighting control, home automation, etc.), or by providing users with energy feedback on non-HVAC related end-uses (e.g., home energy display feedback, home benchmarking tools, energy efficiency tips, etc.). Features that fall outside of the core functionality of a thermostat, i.e., HVAC system control, are not considered in this framework.

Two-way communicating thermostats present opportunities for data-driven analysis that can enable more robust performance assessment methods. In this report we focus primarily on two-way communicating programmable thermostats, though with sufficient data, the proposed methods could be applied equally to more basic thermostats.

In addition to energy performance, thermostats could be compared across other dimensions, such as user satisfaction, energy cost savings, and peak load shifting and reduction. Because these goals may not be fully aligned (e.g., optimizing for peak load reduction or cost savings may not yield the highest energy savings), we constrain our attention in this report on energy performance, while recognizing that these methods may eventually be adapted to assess other parameters. Such parameters could be addressed independently, potentially using a framework similar to the one we describe.

1.2 A Hybrid Data-Driven Approach

To address the shortcomings with existing methods, we propose an approach that separately considers behavioral attributes (setpoint-related) from non-behavioral attributes (HVAC control strategies and Fault Detection and Diagnostics [FDD]), see Figure 1. The approach applies separate assessment techniques, informed by thermostat field data, and integrates results to determine typical energy performance relative to a specified baseline.

Because behavior-driven attributes can be highly variable and more challenging to evaluate, we propose using a data-driven simulation-based approach. Temperature setpoint data from regional samples of installed thermostats will be used in energy models of prototypical homes. Applying observed real thermostat settings directly as input to building energy simulations helps to isolate the effect of the thermostat setpoint selections on energy consumption. Energy performance may be calculated relative to a user-specific baseline temperature that is determined from each user's seasonal history of

setpoints. To demonstrate the feasibility of this approach, Section 4 presents a hypothetical performance assessment using data from over 50 Nest thermostat users in the Washington D.C. area.

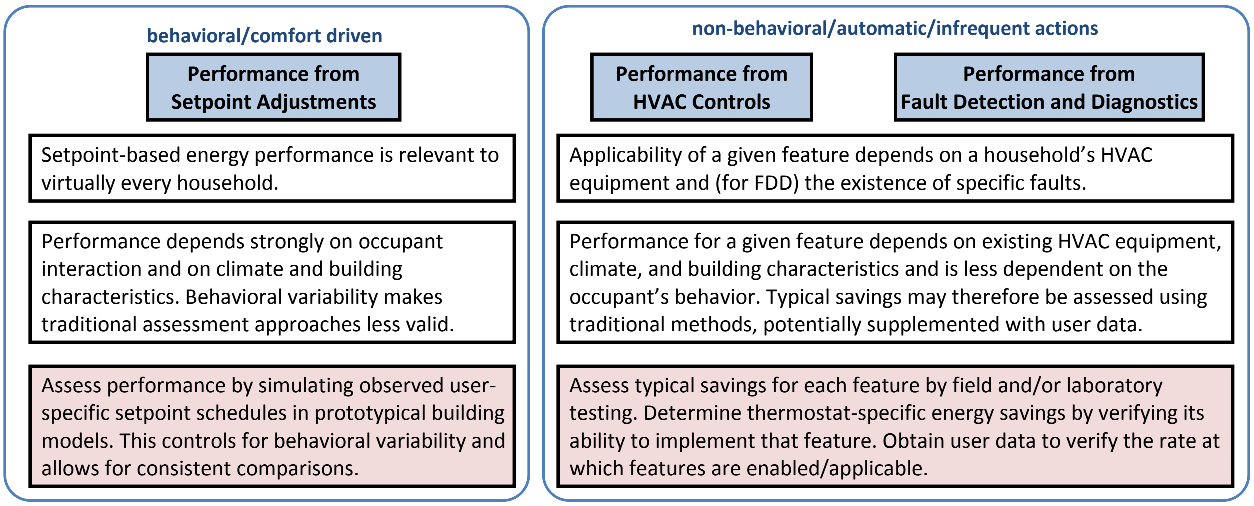


Figure 1: Assessing thermostat energy performance differently based on the influence of occupant behavior.

Assessments of other HVAC control strategies and other features may be handled with more traditional methods, as they are less sensitive to user preferences and often take place without the user's knowledge or input. For these kinds of savings, we recommend using scientific field studies, laboratory testing, and/or whole-building energy simulation to assess the performance of individual features. The results of such field studies may be generalized to other regions with simulation as appropriate. In many cases, well-controlled studies can be performed directly on HVAC systems to assess the impact of specific features on typical systems and under specific conditions. When there is consensus about the typical performance of selected features, thermostats may automatically receive a specified energy performance credit by including those features in their products. Additionally, field data from communicating thermostats could be used to verify that these energy features remain enabled among the user base and function as intended. This is an important feature, since persistence of savings is a major uncertainty associated with thermostat performance. If a manufacturer believes that its product offers additional savings beyond the typical savings attributed to a specific feature, it could submit additional data as part of a structured review process to petition for additional credit.

Because actual thermostat energy performance depends on many factors, we can perform assessments under several operating scenarios, repeating the analysis by climate region, for different common HVAC system types, and types of housing construction (e.g., code-compliant new construction vs. leaky, old construction). Once the energy performance is calculated for each component, the savings estimates can be combined into a single figure of merit, such as percent savings of heating and cooling energy relative to a specified baseline, that can be used to compare products.

Turning this framework into a practical analysis procedure will require additional input from industry stakeholders and additional scientific study to ensure the ultimate results are fair, consistent, and representative.

2 Thermostat Performance Background and Motivation

2.1 Thermostat Device Classes and Features

In recent years, thermostats have come to market with a range of new features, many with significant energy implications. At the highest level, we can define three general categories of thermostats:

1. **Manual Thermostats (MTs)** require human input to change the setpoint or to implement setbacks.
2. **Programmable Thermostats (PTs)** allow users to define a setpoint program, or time schedule of setpoints, to automate the setpoint control.
3. **Programmable Communicating Thermostats (PCTs)** can be controlled or programmed externally by the user, through a website, smartphone, or by a utility in the case of a demand side management (DSM) program.

This general classification can be misleading from an energy perspective. That is, it would be convenient if all thermostats in a given class outperformed those of another class; however, this is not necessarily the case. In reality, the actual energy performance of a given thermostat depends on a combination of factors including climate, building and HVAC system characteristics, thermostat usability, thermostat control algorithms, and occupant behavior. Within each device category, thermostats offer different feature sets and may implement similar features in different ways. Thus, it would be overly simplistic to assess thermostats by general category. Similarly, just because a thermostat has specific energy-saving features does not mean the features can or will be used to save energy. The following is an overview of the key features and attributes that may affect thermostat energy consumption.

2.1.1 User Interface and Usability

User interface design and usability can play an important role in how thermostats are used to save energy (Meier et al. 2010, 2011). Thermostats now offer a range of interface options including touch screen displays, web or smartphone interface, remote control, and voice commands. These features are changing the way people use their thermostats (CADMUS 2012b). Some products attempt to automate, facilitate, or encourage the use of temperature setbacks through features like occupancy detection, geolocation, away button, interview-based programming, machine learning, and user feedback (Wooley et al. 2012, Gupta et al. 2009, Gao et al. 2009, Lu et al. 2010). Some feedback techniques invoke social norms, allowing users to compare their usage or settings with their peers. Others estimate and display energy or cost savings associated with a particular setback program and use this as a motivator. To discourage temperature overshoot (e.g., selecting a high setpoint temperature to cause the building to heat faster), thermostats may display the expected time-to-temperature. Feedback may be delivered via alerts through email or smart-phone notifications or on the thermostat itself.

A thermostat's schedule of setpoints can change over time with user interaction or through automated adjustment. Even though many thermostats ship with energy-saving default setpoint programs, field studies of existing programmable thermostats have found that many households eventually change these settings, opting instead for permanent holds or programs without setbacks (Nevius and Pigg 2000, Sachs et al. 2012). In contrast, thermostats with learning features or with built-in behavioral feedback (e.g., monthly energy reports, automatic schedules) may lead to more energy efficient setpoint schedules over time. Thus, an early comparison between two thermostats may show no difference in energy savings, while comparisons two years later may show a large difference. Because of potential changes in performance over time, testing should be performed on homes whose thermostats have been in place for different lengths of times so that changes in performance over time may be estimated.

This issue may be critical for accurate ratings but poses many challenges. For instance, connected thermostats can receive software updates, and providers can revise their web interfaces, which can change their feature set or the details of how features work. Consequently, performance evaluations must be tied to a specific software version, and a decision must be made about what level of software revision requires a new performance rating. Changes in performance over time will be especially difficult to address for newly developed devices or thermostats that undergo significant software updates. More research is needed to understand how setpoints change over time and to evaluate the persistence of savings from thermostats.

2.1.2 HVAC Controls

Thermostats can include algorithms for improving energy performance, thermal comfort, or both. Some are able to operate specific HVAC system types in ways that reduce energy consumption. These include control strategies for heat pumps that reduce the use of auxiliary electric resistance heating, fan overrun for air conditioning systems, and weather forecast-based operational optimization. Many of these features can work independently of user-defined setpoints, and without the user's knowledge. The applicability of such features depends on what HVAC system(s) exist in the home.

2.1.3 Fault Detection and Diagnostics

Some thermostats provide fault detection and diagnostics (FDD) that could lead to energy savings. Examples include feedback or maintenance reminders suggesting the user replace their HVAC air filter periodically. Such actions may have varied energy implications. More serious HVAC faults may be detected if they cause significant changes in the heating and cooling dynamics of a building, such as low refrigerant charge, coil fouling, duct leakage, and over or undersized HVAC detection. It may be possible for thermostats to detect opportunities to improve the building envelope, e.g., by identifying poor insulation or leaky construction. For these to result in savings, the faults must be accurately identified, and the occupants must be convinced to take action.

2.1.4 Utility-Based Control

With growing utility interest in peak load shedding strategies, thermostats are coming to market with the capability to react to demand response events (Surles and Henze 2012, Lopes and Agnew 2010). Such programs normally aim to curtail peak demand and not to optimize energy consumption, and these goals may not always be compatible. Utility-based control may be attractive to certain stakeholders, though their implications on energy savings can be mixed. Accordingly, demand response capabilities should be assessed independently of the energy-saving capabilities of a thermostat.

2.1.5 Other Features

Thermostats may address aspects of home energy management that go well beyond thermal environment control. For instance, some devices can be integrated with home automation systems that can control lights or plug loads. We consider such additional features and their energy impact outside the scope of this framework development and instead restrict performance to those features directly pertaining to the core functionality of thermostats, i.e., control of the thermal environment through the building's HVAC systems.

2.1.6 Summary

Because of the ever-increasing and evolving combinations of features and implementations, it is important to develop an evaluation procedure that is agnostic to the specific techniques or features and that can include new savings opportunities. While it may be desirable to know which features are generally responsible for energy savings, this is not always possible (or necessary), due to the potentially complex interaction of features and users. Additionally, since users can change which features are

enabled or disabled over time, the framework should account for variation in energy performance, by (1) considering households who have owned their thermostat for different lengths of time, and/or by (2) enabling periodic re-evaluation or longitudinal studies to monitor performance over time.

2.2 How Thermostats Can Save Energy

Although thermostats may aim to save energy using a combination of features, we focus on the two primary pathways for energy savings: (1) setpoint-based savings, and (2) HVAC control strategies. This distinction is made on the basis that setpoint-based savings are strongly dependent on human behavior and comfort preferences, whereas HVAC control strategies are not as dependent. Both pathways lend themselves to different assessment approaches. The applicability and energy-savings potential of both types of strategies depends on factors such as climate region, building characteristics, HVAC system type, and occupant behavior.

2.2.1 Setpoint Temperature Based Savings Mechanisms

The most common and well-known energy savings mechanism of thermostats is achieved by maintaining lower heating temperature setpoints and higher cooling temperature setpoints. This can be done in two distinct ways. First, users can sacrifice comfort during periods of occupancy. Second, users can implement temperature setbacks, typically when occupants are away or asleep, to achieve savings with minimal impact on comfort. We do not suggest that occupants should sacrifice comfort to achieve savings; however we recognize that some households do employ this strategy. Both approaches can be implemented regardless of thermostat type.

Thermostat setpoints affect energy use in two primary ways:

1. The setpoint schedule affects the indoor temperature which, in turn, affects the building's rate of heat gain or loss through the building envelope from conduction, convection, radiation, and air leakage. It can also affect whether heating or cooling is needed at all during mild conditions.
2. The setpoint schedule can affect the efficiency of the HVAC system and its distribution system through how it affects the use of secondary HVAC stages such as the indoor and outdoor temperature and humidity conditions (and resulting temperature lift of a vapor-compression cycle), heat pump auxiliary heat, and the length of HVAC cycles during system operation.

Changes in thermostat setpoints may occur for several reasons, due to a combination of user interaction and automation. The thermostat program is a time-varying schedule of temperature setpoints that can be programmed with a default setting, created or modified by the user, or adjusted by the thermostat itself through automated feature(s). Automated features may adjust the setpoint based on various criteria, such as time of day, occupancy (e.g., using motion detection or GPS-based away detection) or to accomplish other goals, (e.g., responding to demand response signals from the utility). Other examples of automation include dynamic start and stop times to reach setpoints at specific times, setpoint timing adjustments to control slower-responding systems (e.g., thermally massive construction, radiant floor heating), and specialized controls for heat pumps or multi-stage HVAC systems. Users may elect to disable or override an existing thermostat program, and they may disable or modify the settings of certain automated features that control setpoints. These adjustments may be temporary or permanent.

Because the change in energy consumption from a change in setpoint does not depend on *why* the setpoint was changed, the energy performance of all setpoint changes may be quantified using a single approach. While it may be of interest to assess the impacts of each feature (e.g., auto-away, time-to-temperature, remote control via app or web, etc.), those feature assessments are not required to

determine overall energy performance, and the resulting feature-driven impacts will likely be interactive and not additive.

2.2.2 HVAC Control Strategies

Beyond the selection and scheduling of temperature setpoints, thermostats may also save energy through specific HVAC system control strategies. These savings often occur without the occupant's knowledge or interaction and are therefore usually far less sensitive to human behavior than setpoint-driven savings. In contrast, their savings potential or applicability may depend more strongly on the type of HVAC systems in a given home. The level of savings can also depend on how well the control strategies are implemented, and what sensors and data are available to the thermostat. Many of these features can be enabled, disabled, or adjusted by the users, which could influence energy savings capabilities. Here we indicate a few examples of these opportunities, including air conditioner fan overrun, heat pump auxiliary heat control, zoned heating and cooling, and other strategies.

Air conditioners often remove moisture from the air even in hot and dry conditions, leading to unnecessary energy expenditure. By continuing to run the AC fan for a brief period after the compressor has stopped running (fan overrun), it is possible to provide additional cooling by recovering some of the latent cooling capacity, thereby saving electricity. This strategy is most effective when indoor relative humidity is low, and its applicability depends on climate region (Proctor 2010, Conant 2008, Shirey 2008). Thermostats can use humidity sensors and algorithms to intelligently apply this feature only when it does not adversely affect comfort conditions, and they may optimize when the compressor is switched off to maximize energy savings.

Buildings that use heat pumps for their primary source of heating often include a backup or supplemental heating source (usually electric resistance strip heating) that is used when the outdoor air temperature becomes too low for the heat pump to meet heating loads (WSU 2010). Electric resistance heat is inefficient for warming a building and the thermostat may trigger its use to help accelerate temperature recovery from a set-back when it is not necessarily needed. Special thermostats exist for controlling heat-pump systems, often by using outdoor air temperature sensors and algorithms to help determine when to start running the heat pump to achieve the desired temperature at a given time without the need for resistance heat. Some thermostats offer additional avenues for energy optimization using web-based weather data forecast information.

HVAC zoning permits heating or cooling delivery to selected portions of the home and may save upwards of 20% in heating and cooling energy (Sookoor et al. 2011). While thermostats cannot create a multi-zone system from a single zoned system – zoning requires special HVAC infrastructure – they can impact the energy used by such systems. Some thermostats are designed specifically for controlling zoned systems and may rely on occupancy sensors and predictions to intelligently heat and cool specific zones, thus saving more energy than traditional thermostats.

Other HVAC related strategies may be used to improve thermal comfort. Some thermostats are able to estimate the time to temperature based on prior system performance. This information may be used to pre-heat or pre-cool spaces in anticipation of an upcoming setpoint change or occupancy event, so that the building is comfortable at the desired time. This capability may make aggressive setbacks more palatable, as occupants will return to an extremely hot or cold building less frequently. Such features may increase or decrease energy consumption depending on the building characteristics, occupancy patterns, and control algorithms (Malnick et al. 2012).

For certain HVAC control features, the energy performance may be consistent among different thermostat models, while for others it may depend on the specific implementation of the control algorithm. The assessment approach for HVAC related savings, should consider each unique feature independently and in the context of typical buildings and systems.

2.3 Performance Assessment Challenges

When weighing the benefits of different assessment methods, we considered several important challenges and goals.

2.3.1 Baseline Selection

Determining savings requires the establishment of a baseline, against which energy reductions are calculated. In the case of thermostats, it can be challenging to define an appropriate or 'typical' baseline. Setpoint schedule histories tend to be irregular, and using average setpoint schedules (e.g., across a sample of users) as a baseline condition does not yield representative results.

In the context of pre-/post- retrofit evaluations, where participating households receive a new thermostat to replace an existing one, historical energy use is often chosen as a baseline, despite the fact that each households may start out with a different thermostat and use them in different ways. This is problematic when using results to compare thermostat products to one another. For instance, some households may have been using temperature setbacks with their existing thermostats, and thus, their baseline historical energy use would be lower than another household that does not use setbacks at all. Consequently, historical pre-retrofit energy consumption may not be the most consistent and representative baseline for comparing the energy performance of different thermostat products.

One approach that solves both challenges is to develop an artificial baseline that represents the energy consumption associated with a thermostat that has no energy-saving capabilities at all. This is akin to a fictional thermostat that can only maintain one fixed temperature setpoint during the entire heating season and another fixed setpoint for the entire cooling season. In such a case, no savings would be possible through setbacks or special HVAC control. In fact, it is likely that a large portion of homes currently operate their thermostats in precisely this manner (e.g., MTs which are never or rarely adjusted, and PTs that are on permanent hold; Sachs et al. 2012, Peffer et al. 2011). These artificial baselines can be determined from observed user-specific setpoint preferences. We expand on this concept in Section 3 and show how it could be used to calculate energy performance.

2.3.2 Sampling Requirements for Energy Analysis

Insufficient and non-representative sampling of households or participants can compromise the statistical validity of energy assessments, and this has caused much confusion about thermostat energy savings in particular. Energy performance depends on many factors, including thermostat features and implementation, user behavior, building characteristics, HVAC system characteristics, and climate. As a result, the number of combinations required to achieve a representative sample may be quite large.

Taking a simple view, suppose households could take on only a few discrete values for each of the following parameters:

1. Motivation to save energy (high, medium, low)
2. Household vacancy rate (high, medium low)
3. Building type (single-family detached, single-family attached, multi-family)
4. Building thermal characteristics (well vs. poorly insulated; leaky vs. tight construction)
5. HVAC system type (e.g., heat pump, furnace) and characteristics (efficient, inefficient)
6. Climate (hot, cold, mixed, humid)

Motivation to save energy could represent a household's willingness to use nighttime or vacancy setbacks or to accept slightly less comfortable temperatures to achieve energy savings. Vacancy rate – when and for how long the home is unoccupied – determines the maximum potential for away-driven setbacks. Households that are occupied most of the time, for example, will have limited opportunity to achieve setback related savings. In this simplistic view, there are at least nine combinations of behavioral factors: three for motivation (low, medium, and high) and three for vacancy rate (low, medium, high). In reality, there will be a continuous spectrum that is much more complex.

Similarly, building type and thermal characteristics can affect a thermostat's ability to save energy. Building type affects the amount and type of exposed surfaces through which heat can be gained and lost throughout the day, and building characteristics, such as air tightness, insulation levels, and thermal mass all influence how heat is gained, stored, and lost. In a highly simplistic view of these factors, there are at least 24 combinations of building characteristics: three for building type (single-family detached, single-family attached, and multi-family), four for building thermal characteristics (well vs. poorly insulated and leaky vs. airtight), and two for HVAC system condition (efficient vs. inefficient).

To sample each of the simplified combinations of behaviors and building types exactly once – in just one climate region and for only one HVAC system type – would require at least 216 participants [=(9 behaviors) x (24 buildings)]. In reality, the situation is more complex and considering all possible permutations of behavior, climate, and construction could lead to very large sampling requirements that make truly representative field studies prohibitive. RLW (2007), for example, determined that for a specific survey/billing analysis study, at least 3,000 participants would be required to show statistically significant energy savings, if the average savings were about 6%.

2.3.3 Self-Selection and Response Bias

The use of field data to assess thermostats can lead to potential issues with self-selection and self-report. For instance, if a particular thermostat is favored by people who frequently implement setbacks and/or are frequently away, then such a thermostat could appear to have better energy performance, regardless of any cause and effect relationship. Implementing user-specific baseline temperatures should reduce the larger self-selection effects of overall settings, but potential bias still exists. Care must be taken when generalizing savings to populations that differ from the existing user-base. Response bias is most relevant to methodologies that depend on self-reported data, such as surveys. In these cases, people may report information that varies from their actual experiences. For instance, if test subjects believe that they ought to use temperature setbacks, they may exaggerate how frequently they do use them (social desirability bias).

2.3.4 Non-Thermostat Behavior Variability

Certain user behaviors can influence HVAC energy consumption independently of thermostat usage. Window operation and thermal loads from household activities (cooking, appliances, miscellaneous electric loads, etc.), for instance, can affect heating or cooling energy demand. Wide variation in such behaviors could confound results of small population studies. To overcome the experimental noise introduced by these variables may require very large sample sizes, or may be addressed using special analysis techniques that can control for these variations. Building energy simulation is one way of holding all non-relevant behaviors constant, while the energy savings associated with other variables is considered.

2.3.5 Differences in Temperature Measurements

Thermostat settings and measurements are not universal from one device to another or from one home to another. One thermostat's 70°F may correspond to 72°F on another device. Measured temperatures

can differ because of variation in sensor calibration, the physical design of the device, and placement within the room (e.g., solar incidence from windows, proximity to heating or cooling equipment, etc.). In one study of air temperature within an unoccupied apartment unit, instantaneous air temperature measurements varied by as much as 5°F depending on sensor type, sampling rate, and sensor position within the room (Urban et al. 2013). A simple data logger commonly used in field studies demonstrated poor measurement sensitivity[1] due to its small plastic enclosure. These differences can be exaggerated by behaviors that affect local space temperatures, such as the operation of windows or blinds. A consequence is that direct comparisons of measured space temperature or measured temperature settings (either from the thermostat or from external data loggers) from one user to another may not be equivalent.

2.4 Review of Performance Assessment Methods

Prior attempts to assess and generalize the energy performance of thermostats have met with varying degrees of success. Here we will consider several common approaches that may work well in specific cases, but that may be poorly suited for making generalized comparisons about energy savings. This consideration is what led to the approach we propose in Section 3.

2.4.1 Utility Billing Analysis

The majority of the literature assesses thermostat performance based on billing analyses, user surveys, focus groups, and occasionally field studies (Peffer et al. 2011), and most efficiency program billing analyses rely on a pre-/post- experimental design (Michaud et al. 2009, RLW 2007). Large-scale billing analyses typically calculate average per-household savings (absolute [kWh] or relative [%]) heating and/or cooling. Normally, a large number participants in one region (n>1,000) is selected for study. Monthly electricity and gas data are normalized by cooling and heating degree data and compared, and savings are calculated using statistical analysis, based on treatment vs. control group and/or pre- and post- thermostat upgrade. Savings are then attributed to the thermostat.

Since average observed savings relative to historic consumption have been on the order of 5%, this method can require a large number of participants to obtain statistically valid data. Randomized control trials (RCTs) with utility billing analysis are often considered the gold standard of evaluation, though practical considerations can make RCTs difficult to implement. Achieving a truly random and unbiased sample is often difficult in practice due to challenges with recruitment. These studies are often limited to a specific utility's service region, which may not be representative of the larger population. Unless billing analyses are repeated annually, questions about the persistence of observed energy savings will remain unanswered.

2.4.2 Field Studies

Field studies can provide a more detailed picture of energy performance by studying a selected number of households (typically n<100) in greater detail. Households can be instrumented and monitored for a period of time ranging from months (e.g., one heating or cooling season) to a year or more (see, for example, CADMUS 2012a-b and Sachs et al. 2012). Site visits may yield additional information about thermostat usage through occupant interviews or surveys. Detailed knowledge of indoor space temperature, conditions, and submetered energy usage can improve understanding of savings mechanisms in these specific homes. Field studies can also be used to assess the energy performance of isolated features under specific conditions.

[1] The simple data logger measured cyclical temperature swings of 2°F during air conditioner cycling events, while the nearby exposed thermocouple probes measured swings of 7°F.

Studies involving in-home monitoring and sub-metering can provide far richer insights into usage and behavior. Unfortunately, owing to the relatively high cost of project administration, relatively few detailed studies about thermostat usage and energy consumption exist. Those that do are normally confined to a few buildings within small regions, have small sample sizes (typically n<100), and span relatively brief time periods (months). Introducing in-home monitoring may also unintentionally alter the subjects' behavior (e.g., due to the Hawthorne "observer" effect). Consequently, results cannot always be generalized with a high degree of certainty. In some cases, detailed data collected by field studies can be used to validate or inform energy simulations, which can then be used to generalize performance (e.g., Lu et al. 2010; Urban et al. 2012, 2013).

2.4.3 Survey-Based Methods

By conducting interviews or surveys, researchers attempt to identify thermostat ownership, usage, setpoint preferences, and related behavior (e.g., EIA 2009, Nevius and Pigg 2000, Andersen et al. 2009). These results can then be used to calculate expected energy performance. In some cases, surveys are used in conjunction with field studies to establish a baseline condition for each specific household.

Self-report survey results of this kind may be unreliable for several reasons. People generally have a poor understanding of thermostat terminology and how thermostats work. Additionally, people tend to adjust setpoints frequently, meaning that settings described during an interview are unlikely to capture this range of variability. As with any survey-based study, there are risks of response bias and difficulty recalling past events accurately. Since people are often confused by thermostat terminology, the reliability of self-report survey-based studies remains questionable (Peffer et al. 2011, Sachs et al. 2012).

2.4.4 Laboratory-Based User Testing

In this approach, human subjects are tasked with performing specific actions on thermostats in a laboratory environment. Researchers measure the time it takes for subjects to complete – or fail to complete – specific tasks that are essential for achieving schedule-based energy savings (e.g., setting the clock, changing the temperature, implementing a program). From these tests, is possible to rate thermostats according to their ease of use (Meier et al. 2011).

While there is ample evidence that people often find thermostats difficult to use (Peffer et al. 2011; Meier et al. 2010, 2011), ease of use is not a sufficient condition for achieving energy savings (Sachs et al. 2012). Just because a person is able to program a thermostat to save energy when prompted in a laboratory does not mean they will choose to do so in practice. People must also be motivated and willing to implement those energy-saving features. Occupants who value comfort and control more than energy savings, for instance, may not adopt energy-saving features even if they are easy to use. That is, laboratory tests fail to capture motivation and willingness to use a thermostat's energy saving features. Moreover, laboratory tests of usability do not recognize that new users may struggle to figure out how to use a thermostat at first use, but over time, they may seek instructions or get help from others.

In practice, the efficiency gains from a thermostat depend on the interaction between the end user(s), the thermostat, and its features, and not simply whether features are easy to use or adjust. Certain features may enhance the user experience, but may also compromise energy-savings. For example, the permanent hold feature enhances the user experience by giving users more direct control of their space temperature, yet is a prime culprit for the lack of savings from programmable thermostats (Peffer et al. 2011, Sachs et al. 2012, Nevius and Pigg 2000). Additionally, the implementation of features can affect the adoption rate of specific features, e.g., an auto-away feature that does a poor job of identifying occupancy may be disabled by irritated users. Thus, a high-usability feature that is easy to enable may not achieve savings in practice because of poor implementation. It is therefore not sufficient to assess

typical thermostat energy performance in a laboratory or by assigning values to the presence or absence of certain features.

2.4.5 Whole Building Energy Modeling

Building energy simulation is a common approach for calculating the energy performance of specific energy conservation measures. A prototypical home model can be used to simulate the idealized impact of different thermostat schedules. First, a detailed building model is created to represent typical building characteristics. The inputs are chosen to represent a baseline case, which is then simulated using a regional weather file to calculate the annual or seasonal energy performance. Next, one or more inputs are altered to represent the test scenario and the simulation is repeated. Energy performance is calculated as the difference in simulated energy consumption between the baseline and test scenarios. To model the impact of thermostat settings, one can model different setpoint schedules, while holding all other variables fixed. The baseline case might include a fixed setpoint schedule without temperature setbacks, and the test case might apply a particular energy-saving setback schedule. Simulations can be repeated to calculate performance in different scenarios (e.g., climate regions, HVAC systems, building construction, etc.). Examples of thermostat simulation studies include Pilati (1976), Ingersoll and Huang (1985), Guo and Nutter (2010), and Moon and Han (2011), Urban et al. (2012, 2013).

The main advantages of simulation are consistency, control, and the flexibility to compare strategies under different scenarios. The primary weakness is that the inputs (thermostat setpoint schedules) are normally based on non-validated assumptions or limited and site-specific field data. Often simulations apply rigid schedules, whereas people tend to vary their setpoints irregularly with time. Two-way communicating thermostats present an opportunity to obtain a larger and more representative data set of thermostat usage behavior. This data can be used to obtain more meaningful comparisons. The approach we describe in Section 3 will apply data-driven simulation to assess the impact of observed user setpoints relative to an appropriate baseline setpoint schedule.

2.4.6 Data Driven Approaches

Many communicating thermostats provide access to a larger and more inclusive set of data that could be used to assess thermostat performance in a more representative way than the limited studies of the past. This opens the door for different analysis strategies. Thus, we propose to use data that could be easily obtained from most existing communicating thermostats. We discuss a few possibilities here and describe our motivation for the approach we ultimately selected.

Energy Consumption

In the ideal evaluation case, thermostats would measure and provide user-specific time series data of HVAC system energy consumption, history of temperature setpoints, and detailed information about which features are enabled and how they are used. Unfortunately, few (if any) thermostats directly measure the energy consumption of the devices they control, and communicating thermostats may have access to different types of data and sensors. Even if energy consumption were directly available, there are still important behavioral factors that could make a direct comparison among users difficult. These behaviors include factors that affect building heating and cooling loads, like operable windows, blinds, cooking, other internal heat loads, and use of supplemental heating or cooling systems that are not controlled by the thermostat (e.g., space heaters, etc.). Two identical homes with identical thermostat settings would still consume different amounts of energy if these behavioral factors differed appreciably among users.

Proxies for Energy Consumption
HVAC energy consumption may be approximated using algorithms on data collected by the thermostat; however, there is not currently an accepted or validated way of doing this. One approach is to consider HVAC system runtime and to use this as a proxy for energy use, comparing the weather-normalized energy performance of different thermostats, perhaps by region. By measuring HVAC runtime reduction relative to a baseline condition or to other thermostats, the effect of certain features could be determined. Several practical challenges make this difficult to implement. First, different HVAC systems operate and perform differently and their runtime is not always proportional to energy consumption. Notably, air conditioner and heat pump performance can have highly variable energy consumption during runtime, depending on ambient conditions, part load efficiency, and the presence of multi-stage or variable-speed components. This means that directly comparing HVAC runtime for different thermostats may not be meaningful. By combining runtime estimates with detailed submetering field studies, it may be possible to create models of HVAC performance for various typical systems. This strategy would require knowing details about each household's HVAC system to correctly infer savings. Ultimately, comparing runtime reduction provides some information about savings, and may be useful in detailed field studies, however because of these challenges we believe it is not the best method for comparing thermostat product performance on a wide scale.

Thermostat Setpoint Histories
Another simple approach is to directly compare historic thermostat setpoint data among users. This could be used to help identify patterns of energy-saving behavior among users of different thermostats. Since setpoint selection affects when and for how long HVAC systems can run, this could be a useful approach for inferring performance associated with behavior and comfort. For instance, if average setpoint temperatures were lower for users of one kind of thermostat compared with another, some energy savings could be inferred.

Unfortunately, there are serious flaws with directly comparing setpoint histories from one user to another. As discussed earlier, thermostats could measure temperature differently, owing to differences in product, manufacture, and placement within a room, so direct comparisons of setpoint data among different users can be of limited validity. Experimental testing and developing improved standards and requirements for testing thermostat temperature measurement accuracy could help address these concerns.

Additionally, comparing basic statistical measures, such as users' average seasonal daily setpoint temperatures can be strongly misleading. For instance, imagine two users who have identical heating setup temperatures (e.g., 75°F) and different setback temperatures: one selects 50°F while the other selects 60°F. In mild climates, the heating system might never turn on in either case, so although the average setpoint temperatures differ, the energy consumption during these periods would not. Similarly, the setpoint *time* profiles can also have a strong effect on energy performance. Imagine three users, all with an average setpoint temperature of 60°F. The first selects 50°F during the night and 70°F during the day; the second selects 70°F during the night and 50°F during the day; and the third keeps a fixed schedule of 60°F. All else equal, the energy consumption for these scenarios could differ owing to the specific timing of loads. Compounding the issue, HVAC system efficiency also can depend on when and for how long the systems run. Most notably, air conditioner and heat pump efficiency varies strongly as a function of outdoor temperatures.

Energy Simulation of Thermostat Setpoint Histories
To overcome the practical challenges associated with the previous method, we propose to analyze user setpoint data with building energy simulation of observed temperature setpoints. Simulation solves

several problems. First, it isolates all non-thermostat variables including unknown behaviors (e.g., window operation, cooking, internal heat gains, etc.). Second, simulation protocols can be defined to make the analysis consistent and repeatable. Third, by developing specific simulation scenarios, it is possible to calculate performance under a range of different building, HVAC, and climate conditions. This flexibility allows for a better appreciation for the variability inherent in households' real behaviors and preferences. Using observed thermostat schedules directly in the simulations can provide a more realistic estimate of energy consumption.

Section 3 presents the proposed data-driven simulation approach in detail.

3 A New Framework for Comparing Thermostat Performance

No single approach is perfectly capable of comparing thermostat performance in all situations. As a result, we present a new framework that uses a combination of methods and techniques to overcome the challenges presented in Section 2. This approach assesses the energy performance achieved through temperature setpoint control separately from other HVAC control and fault detection and diagnostic (FDD) attributes. Setpoint-related energy impacts are determined using a data-driven simulation-based approach, while HVAC control and FDD impacts are determined using more traditional testing methods.

3.1 Assessing Behavioral Performance by Simulating Observed Setpoints

We propose an approach that applies actual user data to the energy simulation of prototypical building models to calculate energy performance associated with specific thermostat products. This type of analysis can control for variations in behavior and building characteristics that undermine other types of assessments. Relying on actual user data ensures that performance calculations are based on realistic assumptions and does not ignore the variability inherent in thermostat usage behavior. Observed setpoint histories are simulated across multiple prototypical building energy models to account for the variation in the U.S. building stock (i.e., performance may differ depending on building type, building characteristics, HVAC system type, and HVAC efficiency).

The two data requirements are (1) user setpoint histories and (2) location (e.g., zip code or climate region for weather data selection), and most communicating thermostats can collect this data. The approach involves three main steps that must be repeated for each thermostat model, climate region, and season (see Figure 2):

1. **Sample Thermostat Users and Collect Thermostat Data**
   From a sampling of households, obtain the zip code and hourly averaged temperature setpoint histories for a suitable period during the cooling season and during the heating season. Flag times when the HVAC system was manually turned off or otherwise unavailable (e.g., power interruptions, equipment failure, etc.).
2. **Define a Baseline Setpoint Schedule for Each User**
   The energy performance of a user's setpoint adjustments need to be assessed relative to a meaningful baseline setpoint behavior. Later in this section we propose a method of calculating a user-specific baseline temperature to represent their preferred seasonal setpoint.
3. **Simulate Energy Impacts**
   Using a building energy simulation tool, two simulations are performed for each user. Each sampled thermostat is simulated first using the baseline setpoint, and second using the actual recorded setpoints. The difference in energy consumption between these two cases represents the energy impact related to seasonal setpoint adjustments. This process is repeated for each user in the sample.

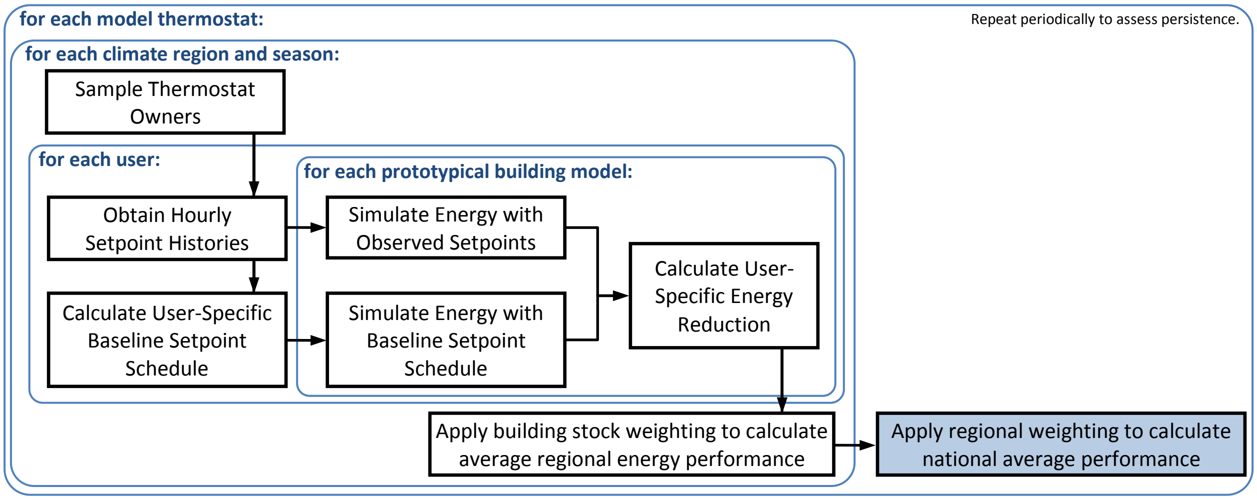


Figure 2: Flow chart of the data-driven simulation process.

The result of this simulation process is a calculation of typical heating and cooling energy reductions associated with each of the sampled temperature setpoint schedules, as applied to each of the prototype home models, by climate. National or regional performance may then be obtained by performing a weighted average of user-specific savings. Regional performance could be weighted according population, end-use energy consumption, and/or building stock.

When implementing this modeling approach in a formal procedure, several details must be determined to ensure that the comparisons fairly characterize energy performance, see Table 1.

Table 1: Simulation parameters needing further development.

	Parameter	Notes / Examples
Define Climate Regions	How many climate regions should be considered?	cold/very cold, mixed-humid, hot-humid, mixed-dry/hot-dry, and marine (e.g., Building America or ASHRAE)
	How will climate regions be defined?	heating and cooling degree days thresholds
	How will regional results be aggregated?	weighted by population, performance reported regionally
User Sampling and Data Handling	How many users per thermostat model and/or climate region are required to form a representative sample?	likely more than 200 users
	How should users be selected?	randomly, based on demographics, based on length of time thermostat has been installed, etc.
	How should incomplete user data be handled?	specify criteria for excluding poor quality data sets
	How to ensure data are accurate and free of manipulation?	third party data audits
	How to comply with data privacy agreements?	anonymize user data
Simulation Protocol	What modeling software should be used?	validated whole building simulation tool, e.g., EnergyPlus
	How should prototypical models be created?	BEopt software, RECS data, etc.
	How many prototype models are needed?	depends on combination of parameters below
	What HVAC systems should be represented?	central AC, forced air furnace, heat pump, etc. typical new equipment vs. typical installed base
	What building types should be represented?	single-family, multi-family, etc. typical new construction vs. typical installed base
	What building characteristics should be represented?	floor area, insulation, air leakage, windows, etc.
	What should be simulated and reported?	heating energy, cooling energy, fan energy
	How should the temperature baseline be defined?	90^{th} and 10^{th} percentile of user setpoint histories, for heating and cooling seasons, respectively
	What weather data should be used in the simulation?	typical year data, actual year data

While each of these attributes is important, none is especially challenging to determine. We expect that through additional targeted research and stakeholder review, protocols may be developed that clearly define how to perform each step of this assessment. Additional research and testing for sensitivity of key parameters would help inform specific decisions.

3.1.1 Defining Prototypical Building Models

The characteristics of the prototype home models (efficiency, thermal mass, and HVAC systems) can have a significant impact on the energy consumption impact of setpoint adjustments. The prototype models should be representative of the common construction styles in each climate, and energy performance should be weighted to reflect the building stock. The EIA RECS survey data may be useful in defining prototypes. Notably, LBNL (2011) has worked on developing a simulation tool that automates the running of calibrated simulation models for the entire RECS sample in order to model the impacts of various efficiency changes on the larger population. Additionally, NREL (2010, 2013a) has developed a freely available research-based software tool called BEopt that follows a prescribed detailed procedure to create consistent prototypical residential building energy models.

The selected simulation tool should be validated using published studies, and must be able to properly model the dynamic interactions between setpoint changes and building thermal mass, HVAC, and distribution system efficiencies. EnergyPlus is another freely-available tool that has been well validated and satisfies these requirements (DOE 2013). Ideally, large data sets from connected thermostats could be used for additional validation studies.

3.1.2 Temperature Baseline Selection

Of central importance is the selection of an appropriate user-specific temperature baseline. Ordinarily, simulation studies presume unrealistic and fixed thermostat setpoint schedules. In reality, people may adjust their setpoints frequently and irregularly (Sachs et al. 2012), and this can lead to unrealistic consumption estimates (Urban et al. 2012, 2013, NREL 2013b). It may be tempting, for instance, to calculate the average temperature among all users, and use this as a fixed-temperature baseline setpoint for everyone, but this would yield misleading findings.

To understand why, suppose that one household prefers an unusually high heating setpoint of 80°F while they are home and is willing to use a 60°F setback while they are away. Also suppose this household has a thermostat that automatically implements setbacks whenever the house is unoccupied. If we simulate performance relative to an arbitrary fixed setpoint that is the same for every household (e.g., 72°F), then we might calculate positive savings (if the house is frequently vacant) or negative savings (if the house is frequently occupied). This baseline incorrectly rates the user's occupancy behavior instead of its utilization of temperature setbacks.

Suppose now that the same household previously had a non-programmable thermostat and always kept the temperature set to 80°F and never used setbacks (e.g., out of convenience). By using an automated thermostat instead, this household could achieve significant energy savings, despite the fact that it still might use more energy than an 'average' household. *Consequently, a comparative calculation of performance based on a universal comfort setpoint fails to represent the comfort preferences of that particular household*.

We think that setpoint-based performance should be assessed relative to each household's seasonal preferred setup temperature. This baseline represents the fixed seasonal temperature that the household would prefer to maintain if setbacks were not available, and corresponds to the second hypothetical case we have just described. This no-setback baseline scenario represents the most energy that a particular household would be likely to consume, given their temperature preferences, and is largely independent of their occupancy behavior. By comparing all setpoint-related performance relative to a baseline condition where setbacks are not allowed, it is possible to assess the energy performance consistently for *any* thermostat.

Importantly, calculating setpoint-related performance using a preferred temperature baseline does not represent the savings that a household will achieve by switching from one kind of thermostat to another.[2] This is a slightly different and more complex question. Studies that compare the savings of switching thermostats tend to generalize about the pre-existing thermostats, e.g., incorrectly assuming that all programmable thermostats are created equally, when in reality usability may strongly influence results. Nonetheless, if all thermostats are assessed with this proposed framework, conceivably it would be possible to compare the energy performance of any thermostat to any other thermostat, without needing to re-normalize results to some arbitrary or artificial baseline setpoint temperature.

A user's preference baseline temperature can be determined directly from their historical setpoints. The approach we recommend is to start with a user's seasonal hourly setpoints and calculate the 90th percentile value for the heating season and the 10th percentile value for the cooling season. This method is a simple and fairly robust way of identifying the likely setpoint temperatures during occupied periods. More basic statistics, such as maxima and minima, are not always representative, since a single overshoot event (e.g., if user sets the temperature to 90°F for one hour) could skew the baseline. Additional study may be done to verify that this method produces reasonable estimates for the setpoints during occupied hours.

We have applied this approach to the user data example in Section 4. Figure 3 shows the heating setpoint histories from thermostats in three different households, and the horizontal gray line indicates the household-specific 90th percentile temperature baseline. Notice that each user appears to prefer very different setup and setback temperatures, and each applies setbacks at different times. For the first and second households, users adhere to fairly consistent setup and setback temperature values. In the third panel, the user has a more dynamic setpoint strategy, with very deep setbacks, and more variable setup temperatures. The first household has one extended period where no setbacks are used, while the third household has an extended period where setbacks are implemented continuously. Nevertheless, in all three cases, the 90th percentile calculation does a good job of representing the typical setup temperature, despite differences in the underlying schedule patterns. Section 4 provides additional information about this approach providing data for a larger sample of users.

[2] Although, up to half of households use their thermostats in a way that is consistent with our baseline definition, namely by using fixed temperature setpoints or never using setbacks (Nevius and Pigg 2000, Meier et al. 2010, Sachs et al. 2012).

Figure 3: Heating setpoint histories for three users: 90th percentile of heating setpoint values indicated by horizontal line.

Another benefit of this user-driven baseline approach is that it computes setpoint-related performance in a realistic and intuitive way. For instance, if a household maintains one fixed setpoint for an entire season, we would calculate no savings for that household because the preference temperature baseline is identical to the actual temperature schedule, and both simulations would yield identical energy consumption results.

If energy consumption depended only on the setpoint history values, then simulation would not be required, and one could calculate performance directly by taking the time-weighted average difference between the 90th or 10th percentile values and the actual schedule. In practice, because buildings are dynamic and can store and release heat at different rates, it is necessary to use simulation to calculate performance in a more realistic way. For instance, deeper setbacks that occur for brief periods may be just as effective as less aggressive setbacks, depending on how quickly the building cools down.

3.1.3 Example Based on User Data

The framework is designed to be flexible to allow different stakeholders to define their own evaluation criteria specific to their needs. For example, a utility may wish to quantify energy performance in one specific region and for a specific building type, whereas a national energy efficiency program such as ENERGY STAR may wish to characterize performance more broadly.

To demonstrate the feasibility of this approach, we present an example in Section 4 based on a small sample of about 50 Nest thermostat users in one region. While this process may seem complex to those unfamiliar with building energy simulation, once the prototypical building models are created and the modeling procedure is formalized, the simulations can be performed readily (each simulation takes less than one minute on an ordinary laptop computer, and simulations can be run in batches). The approach of modeling observed setpoint histories with prototypical building model has been previously demonstrated by Urban et al. (2012, 2013) using setpoint data from programmable thermostats.

3.2 Assessing Non-Behavior (Other HVAC) Energy Performance

This section describes the proposed methodology to assess the HVAC energy performance derived from thermostat features that are not due to temperature setpoint changes. These features fall into two main categories, those related to enhanced HVAC system control and HVAC fault detection and diagnostics (FDD). On a high-level, we propose to assess these measures through feature-specific validation through field or laboratory testing. The overall process is indicated in Figure 4.

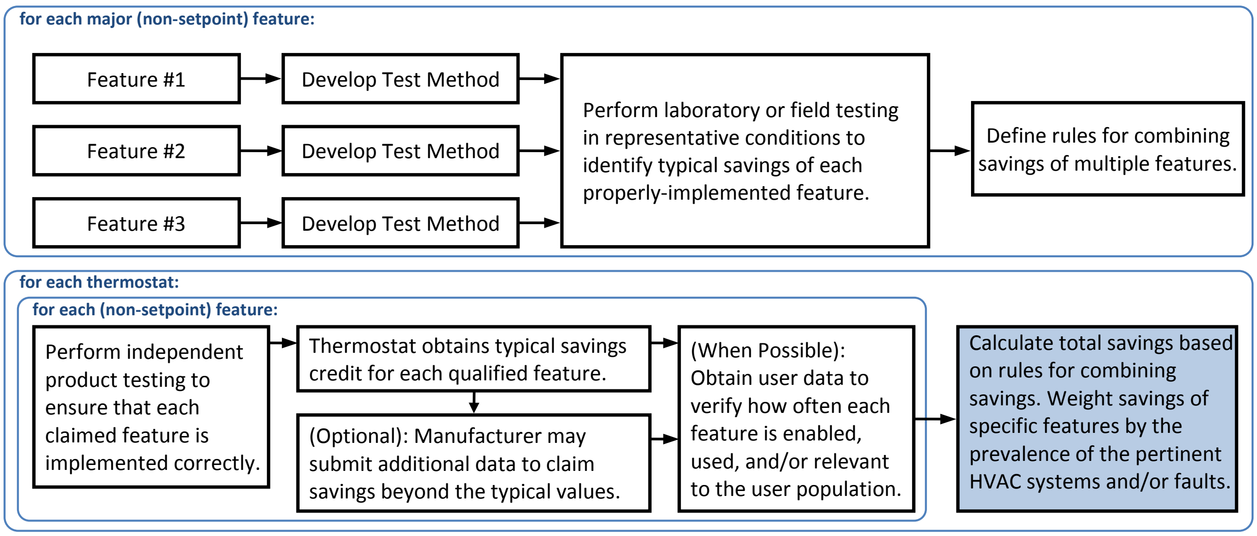


Figure 4: Flow chart for assessing performance of non-setpoint related HVAC control strategies or FDD features.

First, testing protocols must be defined for each key feature. Next, laboratory or field tests must be performed to characterize the typical energy savings associated with each feature under representative conditions. Testing procedures and testing results must be reviewed and approved by an independent review board to ensure their adequacy. Once typical savings are established for a number of features, thermostats may qualify to receive credit for the features they include. To qualify for savings, an independent evaluator must verify that the thermostat implements each designated feature to a minimum, pre-defined specification. If a manufacturer believes their thermostat is capable of delivering savings beyond the typical level for a given feature, they may submit additional data for review.

Once a thermostat's energy savings features are established, typical savings will be calculated based on the prevalence of households for which the feature is applicable. For instance, if a feature only applies to heat pumps with auxiliary electric resistance heat, a weighting factor would be applied to its typical savings to account for the portion of households with this specific system. Similarly, in the case of FDD measures, it is necessary to (1) identify the portion of households that have each specific fault, and (2) to identify the rate at which the thermostat can resolve the fault. When appropriate, user data from communicating thermostats may be used to identify how frequently specific features are enabled or are relevant to a thermostat's user-base. This step is important since users may disable or modify specific energy-savings features over time.

Since any savings from these features will be achieved relative to any savings already achieved through temperature setpoint schedules, the energy savings will typically be multiplicative[3] (and not additive) with the savings resulting from the use of temperature setbacks. Similarly, the energy savings from

[3] For example, if setpoint savings were 15% and HVAC control savings were 10%, the actual combined savings would *not* be additive (e.g., 25%=15%+10%), but would instead be multiplicative (e.g.: 23.5% = 100% – [(100%–15%) x (100%–10%)]).

these features will also tend to multiplicative with each other unless they interact with each other; this will need to be assessed for specific combinations of features. As discussed earlier, we exclude energy savings derived from features that are not related to the core functionality of a thermostat, i.e., HVAC system control, in this framework.

For both Control and FDD savings, a key aspect is to validate that the thermostat can actually effectively achieve the savings promised by the feature. There are two basic approaches to determine the effectiveness of such strategies: field tests and laboratory tests. Field testing could be performed to monitor the response of HVAC systems in households with and without the feature to determine whether or not the HVAC system shows the expected energy-saving response. Alternatively, laboratory testing could be performed under controlled conditions specific to how the feature reduces HVAC energy consumption, with the testing validating that the feature achieves the expected reduction in energy consumption. Given that there are a relatively limited number of basic control strategies and FDD opportunities that save appreciable energy in the residential HVAC systems that account for the largest portion of HVAC energy consumption, we expect that test protocols could be developed for generic versions of many strategies.

3.2.1 Control Strategies that Save Energy without Modifying Temperature Setpoints

Examples of ways that thermostats can control an HVAC system to improve its operational efficiency include reducing use of heat pump auxiliary (i.e., resistance) heating during temperature recovery, fan overrun, and optimal operation of multi-stage/variable-capacity systems. We expect that most control strategies will fall under existing strategies to improve HVAC operational efficiency. In practice, there will be, inevitably, differences between the precise feature implemented in a thermostat and the more generic control strategies (discussed further below).

Once the effectiveness of a feature has been validated, we would propose to award a basic energy-savings credit for the feature based on the typical energy savings attained by that strategy. This credit would be determined using at least one of three approaches to assess its annual energy savings: field testing; laboratory testing, or energy simulations. The strategy to assess the impact of a specific feature will depend upon how much supporting field and laboratory data already exist. For example, several studies have evaluated the energy savings potential of fan overrun, in different climates (e.g., Conant et al. 2008, Shirey 2008, Proctor et al. 2010). Prior or new laboratory testing could also be performed under controlled conditions specific to how the feature reduces HVAC energy consumption. The findings of those tests could then be extrapolated to annual energy savings potentials for different types of equipment in different climate zones based upon a model for the annual energy savings impact of the feature. For example, reducing HP auxiliary heat operation primarily saves energy during temperature recovery periods. A climate-specific model could be used to scale the savings from recovery periods to total seasonal HP heating energy consumption. Finally, energy modeling may be a viable way to assess savings in cases where a validated whole-building energy model can accurately model the operational strategies of both the base-case and feature strategy (e.g., the fan overrun studies of Henderson [2007]).

As noted earlier, there will be, inevitably, differences between the precise feature implemented in a thermostat and the more generic control strategies. If a manufacturer feels that a system provides further benefits above and beyond the typical performance of the feature, we envision that it could submit additional test, laboratory, and/or simulation data to substantiate superior performance. For example, a manufacturer could submit data that compared similar households with and without the feature (or with the featured turned on and off under similar conditions in a designed experiment) in a

sufficient number of households to determine its effectiveness. An independent review board would review the submission to determine if a higher savings credit than the basic credit is warranted.

3.2.2 Fault Detection and Diagnostics

Fault detection and diagnostics (FDD) algorithms assess HVAC actual performance relative to expected performance to find instances of energy-wasting faults. The effectiveness of an FDD algorithm depends upon its ability to detect a specific fault and the likelihood that households will act to remediate the fault. Previous research suggests that duct leakage and low refrigerant charge are the two HVAC faults that waste the most energy in homes (Rossi 2010, Roth et al. 2006). Other energy-wasting faults include insufficient indoor-coil air flow (typically due to AC or HP duct system design, but also filter clogging), AC and HP indoor and outdoor coil fouling, and high water temperatures on condensing boilers (that limit or preclude condensing heat transfer) (Palani et al. 1992). The technical literature (see above) provides estimates for the distribution of the prevalence and severity of the AC and HP faults, as well as the energy impact of faults of varying severity for more common system types. We recommend using these values to assess the technical energy savings potential of FDD.

Each model thermostat that has FDD features would require laboratory testing to assess its ability to detect faults that have a significant energy impact (e.g., improper refrigerant charge at meaningful charge thresholds). The test results would be evaluated not only for the likelihood of detecting a fault, but also for the frequency of false positives, which, we believe, erode households' confidence in the FDD feature. Then the FDD function could be evaluated, e.g., by requiring it to surpass an appropriate threshold on a receiver operating characteristic (ROC) curve. Figure 5 represents a hypothetical ROC for a model thermostat that has the ability to detect and report three different faults: A, B, and C. This thermostat is only able to detect Fault A sufficiently well (above the minimum threshold) to be eligible for energy savings credit. The thermostat identifies Fault C no better than a random guess, and Fault B slightly better than a random guess but still below the minimum threshold, so these features would be ineligible for energy savings credit.

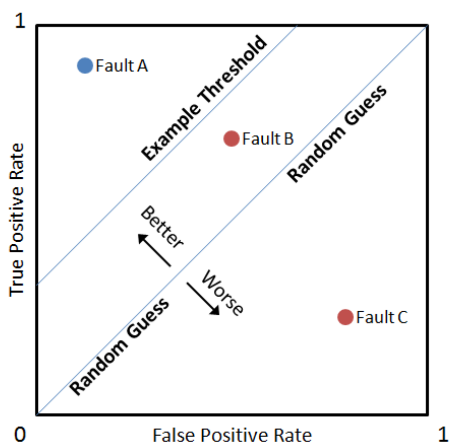


Figure 5: Receiver operating characteristic for a hypothetical thermostat.

Much less well understood is the likelihood that households will act upon this information to remediate a fault. The fault correction rate (faults resolved per faults detected) will likely depend on how the information is communicated to the user and is therefore also likely to vary by thermostat model. In the case of faults that are not addressed by most HVAC contractors (e.g., duct leakage) or that require significant investment to fix (e.g., insufficient indoor coil airflow due to suboptimal duct layout), we suspect that most households would not decide to fix the fault. Thus, we propose that the system

should not receive energy savings credit for FDD for simply detecting those faults. In contrast, for faults that could be more readily addressed, such as improper refrigerant charge and coil fouling, we expect that a significant portion of households would address the fault, and propose that the thermostat should receive partial credit for the technical energy savings potential, for instance 50%.

Although FDD is still an emerging field for residential thermostats, this framework is able to consider savings driven by FDD related features.

3.3 Combining Performance into One Metric

The quantified energy performance from setpoints and other HVAC savings will ultimately be combined into a single metric that expresses total energy reduction compared to the baseline case of a thermostat that has no energy savings features. The energy reductions calculated for the sub-metrics may be additive or multiplicative depending on the design and implementation of HVAC control algorithms and FDD measures and must be integrated accordingly.

Depending on the intended purpose, the energy impact of features can be assessed in different ways to reflect different HVAC system types and climates. For example, a national program such as ENERGY STAR may prefer to weight the thermostat cooling efficiency metric to reflect the most common HVAC systems deployed in regions with the greatest cooling energy consumption, e.g., central heat-pump systems and central AC systems in the South, Southeast, and Southwest. More regional applications, such as state-level utility energy efficiency programs, could use more typical equipment and weather conditions to determine appropriate credits specific to that situation.

Results will need to be weighted to reflect the defined population of interest (e.g., all single family homes in U.S.) and also to reflect the relative proportion of the target population with the specific HVAC system types and climates, see Figure 6. Rules for weighting and combining sub-metric results will need to be defined in the assessment procedure.

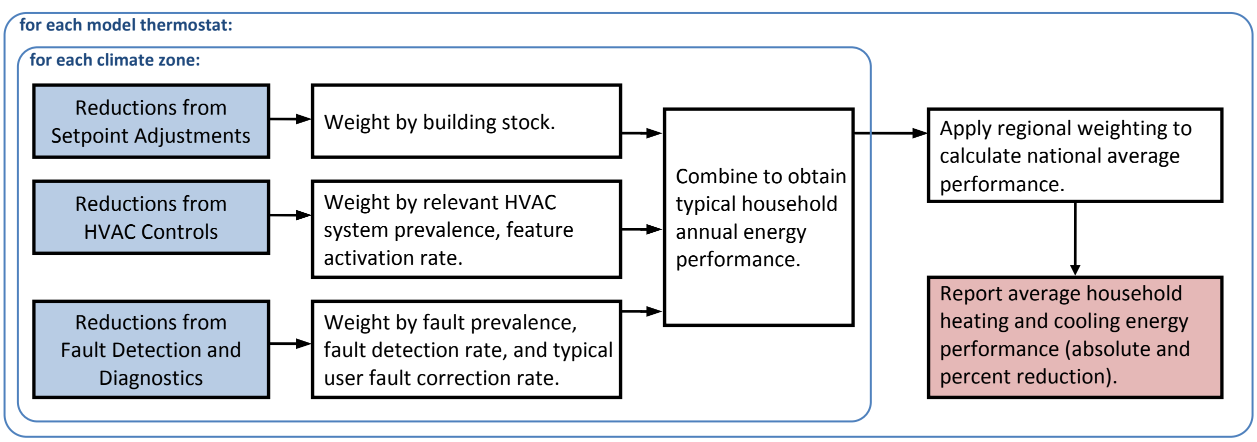


Figure 6: Flow chart for combining energy reductions assessed through different measures.

4 Data-Driven Thermostat Setpoint Assessment Example

To demonstrate the feasibility of this data-driven simulation-based approach, and to help identify potential challenges with its implementation, we have constructed an example using thermostat data from a small sample of Nest thermostat users in one region[4]. Note that this exercise is intended only to illustrate the process of calculating energy performance in a single region. We do not claim that the performance calculated from this example is representative or statistically significant. In a full-fledged application of this method, data from more users would likely be required to form a representative sample in each region, and the entire simulation process would need to be repeated with user data from multiple climate zones.

Nest provided historical hourly-averaged setpoint temperatures for over 50 users in the Washington D.C. area for a two-month period in winter (Jan-Feb) and summer (Jun-Jul). We chose this location since it has significant heating and cooling seasons and allows us to readily demonstrate results for both.

The setpoint data are shown for one user in Figure 7, and for all users in Figure 8. The y-axis represents hour of the day, the x-axis represents the date, and the color represents the observed hourly setpoint temperature. Regions in white indicate times when the HVAC system was disabled for at any portion of the hour. For specific users, solid *vertical* color bands represent days where temperatures setpoints were held constant. Solid colors that stretch over multiple consecutive days may indicate periods of extended vacancy or continuous occupancy. Solid *horizontal* color bands may indicate a routine setup or setback behavior. From this sample, it is evident that temperature preferences and the implementation of setbacks vary significantly by user, reinforcing the need to include enough users to create a statistically representative sample.

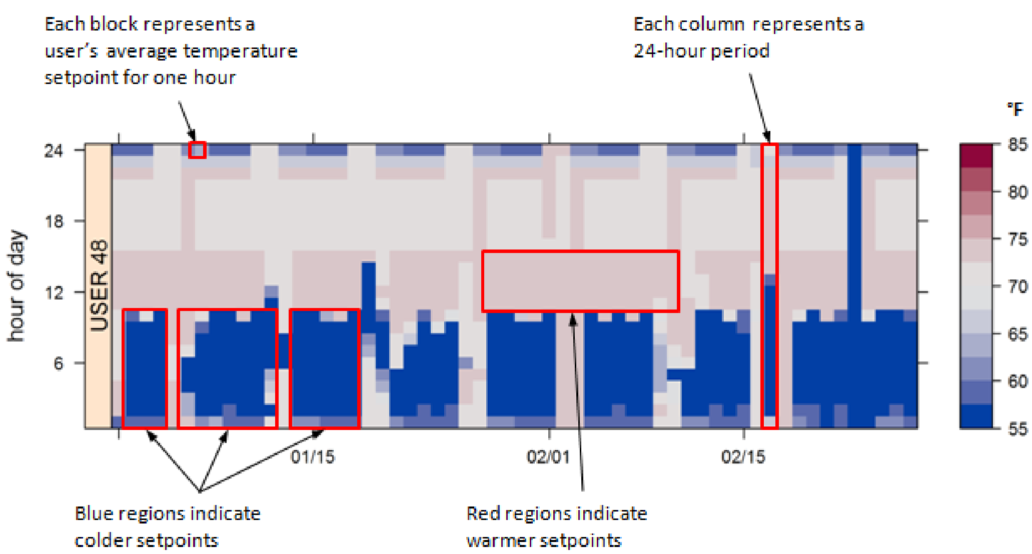


Figure 7: One user's hourly heating setpoints for a two-month period.

[4] The data provided to Fraunhofer did not disclose any personal identifying information of the users, and included only a two month temperature setpoint history and geographical region identified by zip code.

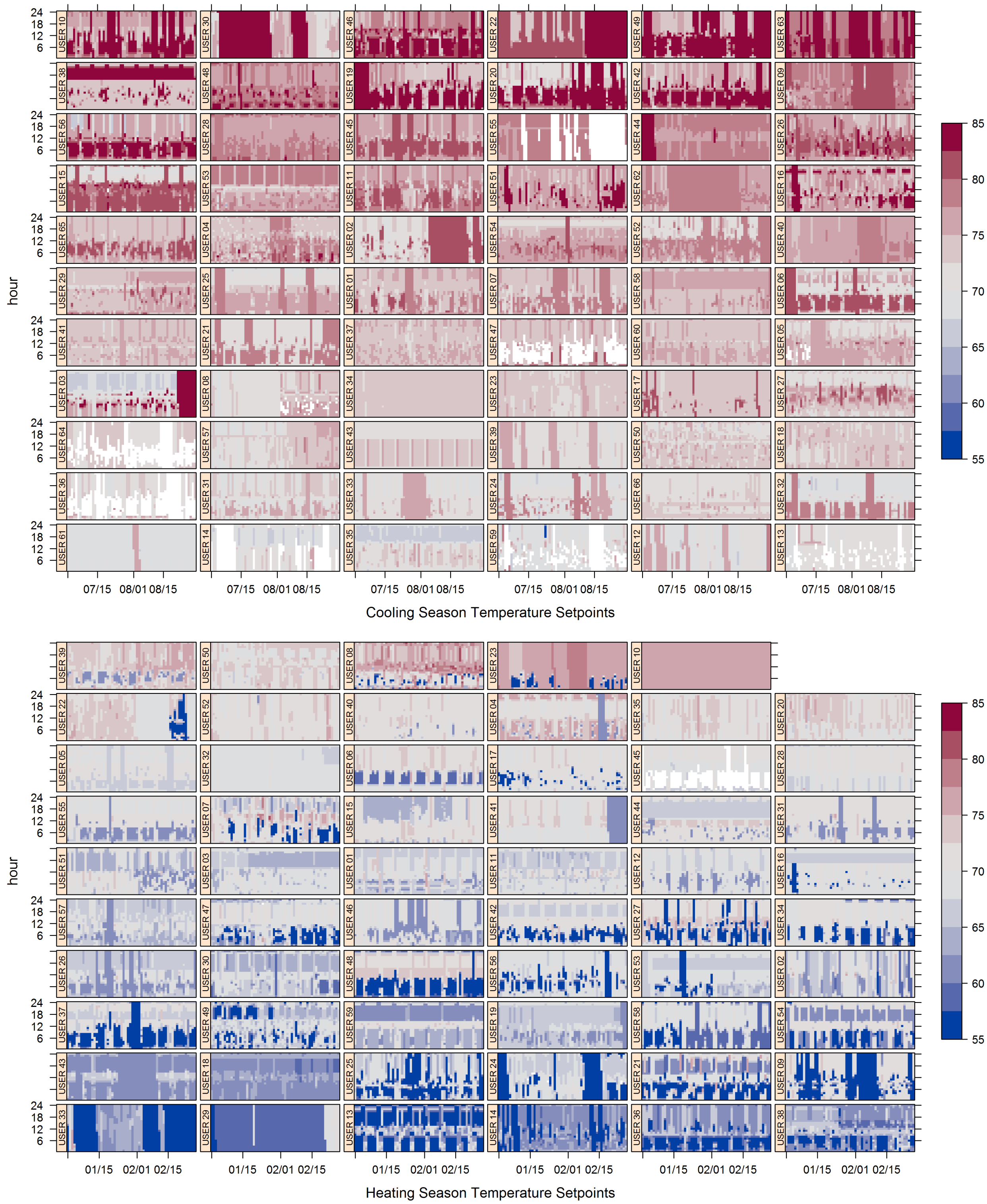


Figure 8: Observed cooling and heating setpoints (°F) for each user by time of day and date. Users sorted by mean setpoint temperature.

We calculated a baseline temperature for heating and cooling specific to each user, based on the 90th and 10th percentile values of their historic setpoints during the corresponding season. Percentiles are calculated based on the time spent at or above a given setpoint temperature. These are used as a fixed setpoint in the comparative simulations and serve as the baseline case. Observed and preference temperatures are shown in Figure 9 for selected users. See Appendix C for similar plots of all users.

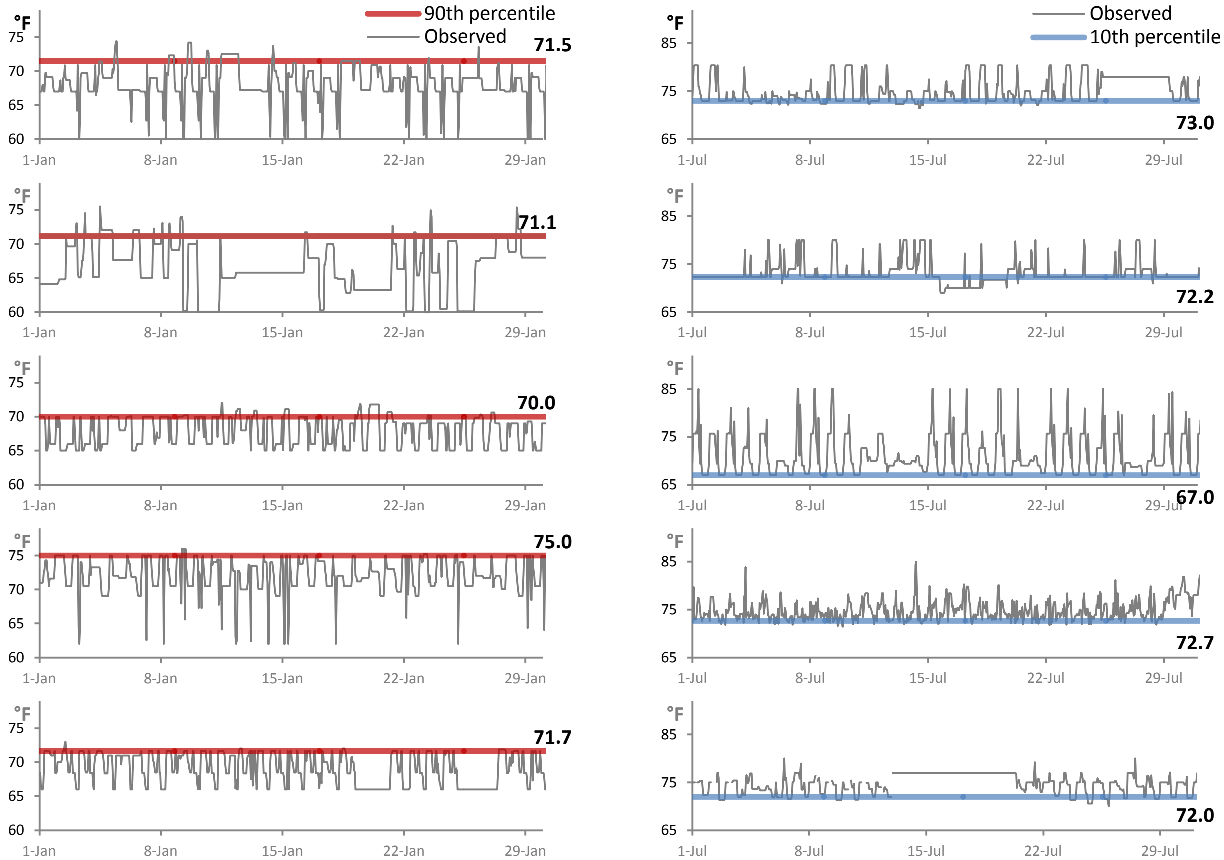


Figure 9: Comparing one month of observed setpoints with 90th and 10th percentile baseline setpoints for five users, LEFT=heating and RIGHT=cooling.

Figure 10 shows the distribution of baseline preference temperatures among users in the sample. Preference temperatures ranged from 67 to 78°F in summer (mean=72°F, n=66), and 64 to 79°F in winter (mean=71°F, n=59). The range of temperature distributions among users is consistent with findings from a field study that measured indoor space temperatures of 60 U.S. homes in several climates (NREL 2013b). Note that preference temperature does not necessarily correlate with mean setpoint temperature because people use setbacks irregularly and inconsistently.

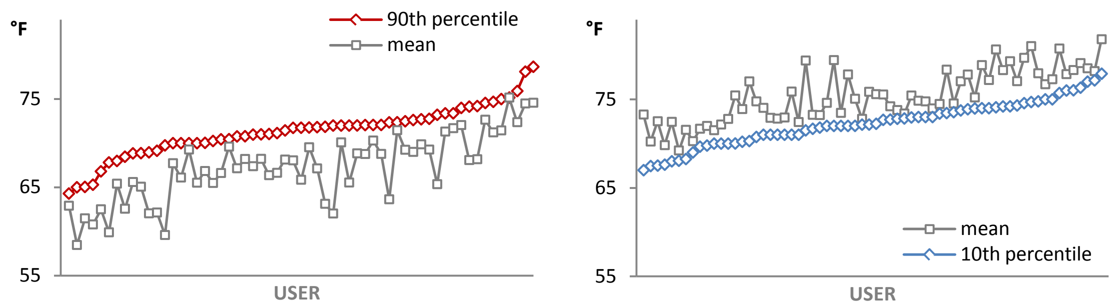


Figure 10: Distribution of baseline preference temperatures among users. LEFT=winter, RIGHT=summer.

Next, we generated a prototypical energy model of a 1,000 ft^2 single-family home to be used as the common basis for all simulations (see Appendix A for more details). We chose to model a typical new construction home with a single-stage central air conditioner and a natural gas furnace (the most common residential HVAC systems); however, other construction types and HVAC systems could readily be modeled instead. We used BEopt and its default settings to produce the prototypical residential input files for simulation in EnergyPlus.

For each user, we simulated energy consumption for two cases per season. First, we simulated the user's *observed* hourly setpoints directly to calculate the energy consumption associated with their actual thermostat usage profile. Second, we simulated a baseline setpoint case corresponding to a situation when the setpoint was constantly set to that user's preferred (i.e., baseline) temperature. The results of these simulations are shown in Figure 11 and Figure 12. Fan energy is shown separately from heating and cooling energy here to show their relative importance, however, in practical assessments they may be considered together. In each case, the heating, cooling, and fan energy of the observed setpoints case was lower than for the baseline. Actual energy reductions relative to the baseline case depend on each user's application of setbacks.

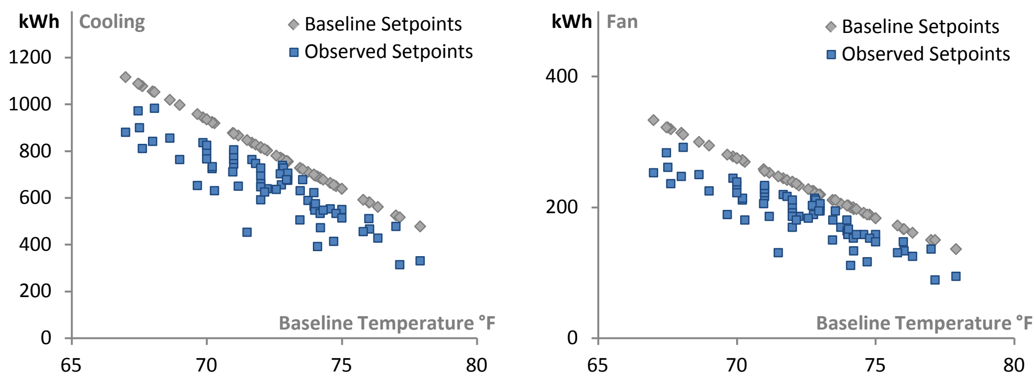


Figure 11: Simulated cooling season energy vs. baseline setpoint temperature.

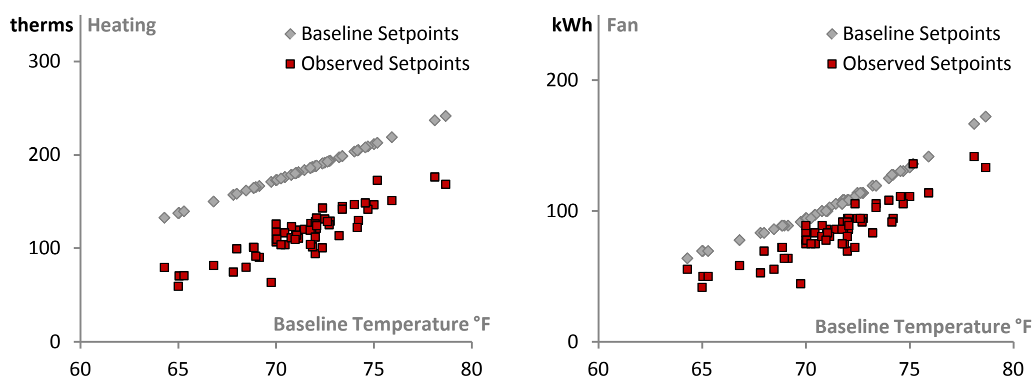


Figure 12: Simulated heating season energy vs. baseline setpoint temperature.

By comparing the user-specific baseline case with the observed setpoints case, shown in Figure 13, we can calculate the expected energy consumption and reduction for each user for the given simulation scenario, as in Figure 14. Heating plus fan energy reduction, relative to the baseline, ranged from 0-51% (mean 19% or 2.9 GJ). Cooling plus fan energy reduction ranged from 4-47% (mean 18% or 0.7 GJ).

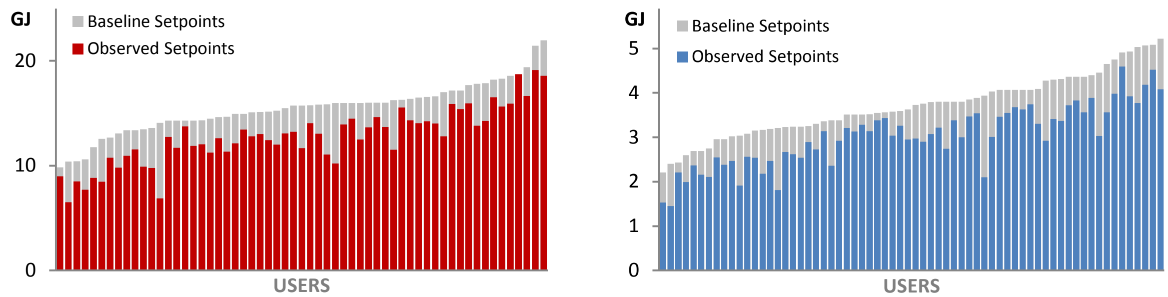


Figure 13: Simulated heating and cooling energy (incl. fan) by user. LEFT=heating season, RIGHT=cooling season.

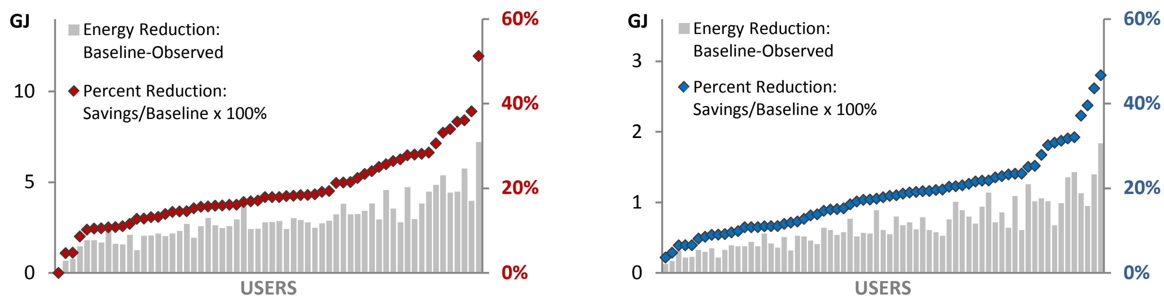


Figure 14: Simulated heating and cooling energy reductions (incl. fan) by user. LEFT=heating season, RIGHT=cooling season.

The range of baseline site energy consumption was 9.9-22.0 GJ (mean 15.5 GJ) for heating and 2.2-5.2 GJ (mean 3.7 GJ) for cooling. Some households whose baseline consumption was lower than average were still able to achieve significant reductions in energy consumption through setbacks, possibly due to extended periods of vacancy. A large enough sample of users must be selected so that such periods do not distort the energy reduction potential predicted by the simulations.

5 Conclusion

Prior attempts to assess and compare thermostat energy performance have proven elusive, largely because of limitations with previous assessment methods. In particular, variable occupant behavior and indoor temperature preferences continue to make comparisons difficult because thermostat performance depends deeply on time-varying human interaction.

Two-way communicating thermostats offer new opportunities for data-driven assessment techniques that can help improve understanding of thermostat usage and field performance. By separating the largely behavioral attributes (setpoint adjustment) from the largely non-behavioral attributes (HVAC control strategies and FDD) and assessing their energy impacts independently with appropriate methods, we have developed a framework that can be used to more completely and precisely quantify the energy performance of deployed thermostats.

In this framework, user-specific setpoint histories are used to inform prototypical building energy simulations that can isolate individual users' energy performance under typical conditions. The energy impact of non-behavioral features may be determined using more traditional techniques, such as controlled field experiments, and may be supplemented by data from installed thermostats to identify the how frequently such features are enabled in homes. To illustrate the feasibility of the data-driven modeling approach, we have presented an example based on data from Nest thermostat users.

The next step in this development effort is to codify the assessment procedure, with input from industry stakeholders and additional research. Applying user data from a larger and more diverse sample of connected thermostats will be crucial for advancing this process and improving understanding of how thermostats are being used and for developing a truly meaningful metric for assessing real-world thermostat effectiveness.

6 References

Andersen, Rune Vinther, Jorn Toftum, Kluas Andersen, Bjarne Olesen. 2009. Survey of Occupant Behaviour and Control of Indoor Environments in Danish Dwellings. Energy and Buildings. 41(2009)11-6.

CADMUS. 2012a. Massachusetts Multifamily Impact Analysis. The Cadmus Group. Final Report Prepared for The Electric and Gas Program Administrators of Massachusetts. Jul.

CADMUS. 2012b. Wi-Fi Programmable Thermostat Pilot Program Evaluation. The Cadmus Group. Final Report Prepared for The Electric and Gas Program Administrators of Massachusetts. September.

Conant, A. and J. Proctor. 2008. "Field Tests of Specially Selected Air Conditioners for Hot Dry Climates". *Proc. ACEEE Summer Study on Energy Efficiency in Buildings*. Aug.

DOE. 2013. EnergyPlus Energy Simulation Software v8.1.0. U.S. Department of Energy. http://apps1.eere.energy.gov/buildings/energyplus/.

EIA. 2009 Residential Energy Consumption Survey. Space Heating in U.S. Homes. Table HC6.6. U.S. Energy Information Administration. http://www.eia.gov/consumption/residential/data/2009.

EIA. 2009 Residential Energy Consumption Survey. Air Conditioning in U.S. Homes. Table HC7.6. U.S. Energy Information Administration. http://www.eia.gov/consumption/residential/data/2009.

Gao, Ge and Kamin Whitehouse. 2009. The Self-Programming Thermostat: Optimizing Setback Schedules based on Home Occupancy Patterns. BuildSys'09.

Guo, Wei and Darin W. Nutter. 2010. Setback and Setup Temperature Analysis for a Classic Double-Corridor Classroom Building. Energy and Buildings. 42(2010)189-97.

Gupta, Manu, Stephen S. Intille, and Kent Larson. 2009. Adding GPS-Control to Traditional Thermostats: An Exploration of Potential Energy Savings and Design Challenges. MIT House_n.

Henderson, H.I., D.B. Shirey, and R.A. Raustad. 2007. "Closing the Gap: Getting Full Performance from Residential Central Air Conditioner". Report to the National Association of State Energy Officials (NASEO) and the New York State Energy Research and Development Authority (NYSERDA). April.

Ingersoll, John and Joe Huang. 1985. Heating Energy Use Management in Residential Buildings by Temperature Control. Energy and Buildings. 8(1985)27-35.

LBNL. Hopkins, Lekov, Lutz, and Gu. 2011. Simulating a Nationally Representative Housing Sample Using EnergyPlus. Lawrence Berkeley National Laboratory. 2011.

Lu, Jaikang, Tamim Sookoor, Vijay Srinivasan, Ge Gao, Brian Holben, John Stankovic, Eric Field, Kamin Whitehouse. 2010. The Smart Thermostat: Using Occupancy Sensors to Save Energy in Homes. SenSys '10. Zurich Switzerland. Nov.

Lopes, John S. and Patrick Agnew. 2010. FPL Residential Thermostat Load Control Pilot Project Evaluation. ACEEE Summer Study on Energy Efficiency in Buildings.

Malnick, Todd, Nate Wilairat, Jennifer Holmes, and Lisa Perry. 2012. Destined to Disappoint: Programmable Thermostat Savings Only as Good as the Assumptions about Their Operating Characteristics. ACEEE Summer Study on Energy Efficiency in Buildings.

Meier, Alan, Cecilia Aragon, Becky Hurwitz, Dhawal Mujundar, Daniel Perry, Therese Peffer, and Marco Pritoni. 2010. How People Actually Use Thermostats. ACEEE Summer Study on Energy Efficiency in Buildings.

Meier, Alan, Cecilia Aragon, Therese Peffer, Daniel Perry, Marco Pritoni. 2011. Usability of Residential Thermostats: Preliminary Investigations. Building and Environment. 46(2011)1891-8.

Michaud, Normand, Lori Medgal, Pierre Baillargeon, Carl Acocella. 2009. Billing Analysis & Environment that Re-Sets Savings for Programmable Thermostats in New Homes. Energy Program Evaluation Conference, Portland. 597-606.

Moon, Jin Woo and Seung-Hoon Han. Thermostat Strategies Impact on Energy Consumption in Residential Buildings. Energy and Buildings. 43(2011)338-46.

Nevius, Monica J. and Scott Pigg. 2000. Programmable Thermostats That Go Berzerk? Taking a social perspective on space heating in Wisconsin. Consumer behavior and non-energy effects. 8.233.

NREL. 2013a. BEopt v2.1.0. National Renewable Energy Laboratory. http://beopt.nrel.gov/.

NREL. Roberts, David and Kerylyn Lay. 2013b. Variability in Measured Space Temperatures in 60 Homes. National Renewable Energy Laboratory.

NREL. Hendron, Robert and Cheryn Engebrecht. 2010. "Building America Research Benchmark Definition." NREL/TP-550-47246. National Renewable Energy Laboratory. http://www.nrel.gov/docs/fy10osti/47246.pdf. Jan.

Pilati, David. 1976. Residential Energy Savings Through Modified Control of Space-Conditioning Equipment. Energy. 1(1976) 233-9. Mar.

Palani, Manivannan, Dennis O'Neal, and Jeff Haberl. 1992. The Effect of Reduced Evaporator Air Flow on the Performance of a Residential Air Conditioner. *Proceedings of the Eighth Symposium on Improving Building Systems in Hot and Humid Climates.*

Peffer, Therese, Marco Pritoni, Alan Meier, Cecilia Aragon, and Daniel Perry. 2011. How people use thermostats in homes: A review. Building and Environment. 46(2011) 2529-41.

Proctor, J., A. Conant, and A. Hairrell. 2010. "Proportional Time Delay Relay for Air Conditioner Latent Capacity Recovery". Proctor Engineering Group. November.

RLW. 2007. Validating the Impact of Programmable Thermostats. RLW Analytics. Final Report to Gas Networks. Jan.

Rossi, T. 2010. "Residential Air Conditioning Fault Detection and Diagnostics (FDD) and Protocols to Support Efficient Operation." Building America Program. Residential Buildings Integration Meeting - July 20-22, 2010.

Roth, K., D. Westphalen, and J. Brodrick. 2006. "Residential Central AC Fault Detection & Diagnostics". *ASHRAE Journal*. May. pp. 96-76.

Sachs, Olga, Verena Tiefenbeck, Caroline Duvia, Angela Qin, Kate Cheney, Craig Akers, and Kurt Roth. 2012. Field Evaluation of Programmable Thermostats. Final Report to the U.S. Department of Energy Building America Program. Dec.

Shirley, D.B. 2008. "AC System Equipment Specification, Installation, and Operational Options for Improved Indoor Humidity Control". Florida Solar Energy Center Report, FSEC-PF-443-08.

Sookoor, Tamim, Brian Holben, Kamin Whitehouse. 2011. Feasability of retrofitting centralized HVAC systems for room-level zoning. Sustainable Computing: Informatics and Systems.

Surles, William and Gregor P. Henze. 2012. Evaluation of Automatic Price Based Thermostat Control for Peak Energy Reduction under Residential Time-of-Use Utility Tariffs. Energy and Buildings. 49(2012)99-108.

Urban, Bryan, Diana Elliott, and Olga Sachs. 2012. Towards Better Modeling of Residential Thermostats. *Proceedings of the IBPSA SimBuild Conference.* Madison, WI.

Urban, Bryan and Carla Gomez. 2013. A Case for Thermostat User Models. *Proceedings of the IBPSA International BuildingSim Conference.* Chambery, France.

Urban, Bryan, Carla Gomez, Chris Geibe, and Kurt Roth. 2013. "Improving Programmable Thermostats with Less Aggressive Default Setbacks." *Behavior Energy and Climate Change Conference.* Nov.

Woolley, Jonathan and Therese Peffer. 2012. Occupancy sensing adaptive thermostat controls: A market review and observations from multiple field installations in university residence halls. ACEEE Summer Study on Energy Efficiency in Buildings.

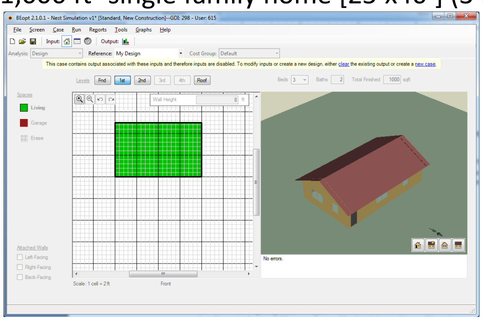
WSU. 2010. Electric Heat Lockout on Heat Pumps. Washington State University Extension Program.

Appendix A: User Data and Simulation Processing Notes

1. Data Format.
   CSV file format.
   Row 1: column headers: Timestamp, User IDs
   Column 1 = timestamp values (e.g., YYYY-MM-DD hh:00:00 + timezone)
   Columns 2:N+1 = observed setpoint temperatures

2. Data Processing.
   If HVAC system is available during the full hour:
   Time-weighted hourly averaged setpoint temperatures.
   When the HVAC system is disabled (heating or cooling mode switched off):
   Substitute the lowest typical acceptable settings (e.g., 50°F for heating and 90°F for cooling).

3. Simulation input file summary (default values from NREL's BEopt simulation tool)
   a. 1,000 ft^2 single family home [25'x40'] (3 beds, 2 baths [used to define hot water usage])



   b. Typical new construction building characteristics:
      i. Wood stud walls; R-13 insulation; light vinyl siding; 1/2" drywall interior
      ii. Unfinished attic; R-30 fiberglass insulation; Asphalt shingle roof
      iii. Uninsulated concrete slab foundation
      iv. 15% window by area; Double pane windows, medium gain, low-e
      v. Air leakage: 7 air changes per hour @ 50 Pa; exhaust ventilation at 40cfm
      vi. Major appliances: refrigerator; gas range; dishwasher; clothes washer/dryer
      vii. Other electric loads: 1927 kWh/yr
      viii. Benchmark lighting: 1022 kWh/yr
      ix. Gas water heater: 0.59 energy factor; 40 gal tank, located in conditioned space
      x. HVAC
         1. Central AC: SEER 13, single stage, 0.07 cycling fraction
         2. Gas furnace: 78% AFUE
         3. Ducts: 15% leakage, R-8 insulation, located in unfinished attic
   c. Weather file:
      EPW file for Arlington, VA, typical weather year

Appendix B: Simulation Results

Table 2: Energy simulation results for heating and cooling seasons.

		Heating Season Energy (GJ)											
		Baseline Setpoints			Observed Setpoints			Difference			% Difference		
USER	Baseline Temp °F	Heat	Fan	Heat +Fan	Heat	Fan	Heat +Fan	Heat	Fan	Heat +Fan	Heat	Fan	Heat +Fan
1	71.4	15.1	0.4	15.5	12.7	0.3	13.0	2.4	0.1	2.5	16%	18%	16%
2	71.1	14.9	0.4	15.2	11.7	0.3	12.0	3.2	0.1	3.2	21%	22%	21%
3	70.0	14.0	0.3	14.3	12.4	0.3	12.7	1.5	0.0	1.6	11%	12%	11%
4	75.0	18.1	0.5	18.6	15.5	0.4	15.9	2.6	0.1	2.7	15%	17%	15%
5	71.8	15.4	0.4	15.8	13.4	0.3	13.7	2.0	0.1	2.1	13%	15%	13%
6	72.8	16.2	0.4	16.6	13.6	0.3	14.0	2.6	0.1	2.6	16%	17%	16%
7	74.2	17.4	0.5	17.9	13.7	0.3	14.1	3.7	0.1	3.8	21%	26%	21%
8	78.7	21.3	0.6	22.0	17.8	0.5	18.3	3.5	0.1	3.7	17%	23%	17%
9	72.0	15.6	0.4	16.0	10.0	0.3	10.2	5.6	0.1	5.8	36%	36%	36%
10	75.2	18.3	0.5	18.7	18.3	0.5	18.7	0.0	0.0	0.0	0%	0%	0%
11	72.1	15.6	0.4	16.0	12.8	0.3	13.1	2.8	0.1	2.9	18%	21%	18%
12	71.0	14.8	0.4	15.1	12.6	0.3	12.9	2.1	0.1	2.2	14%	14%	14%
13	67.8	12.3	0.3	12.6	7.9	0.2	8.1	4.4	0.1	4.5	36%	37%	36%
14	65.0	10.2	0.3	10.4	7.5	0.2	7.6	2.7	0.1	2.8	27%	28%	27%
15	72.0	15.6	0.4	16.0	13.6	0.3	13.9	2.0	0.1	2.1	13%	13%	13%
16	70.8	14.6	0.4	15.0	13.0	0.3	13.3	1.6	0.0	1.6	11%	11%	11%
17	72.5	16.0	0.4	16.4	13.9	0.3	14.2	2.1	0.1	2.2	13%	17%	13%
18	64.3	9.6	0.2	9.9	8.4	0.2	8.6	1.2	0.0	1.3	13%	13%	13%
19	68.0	12.4	0.3	12.7	10.5	0.3	10.8	1.9	0.1	1.9	15%	17%	15%
20	74.0	17.3	0.5	17.7	15.5	0.4	15.9	1.8	0.1	1.8	10%	13%	10%
21	71.9	15.5	0.4	15.9	10.7	0.3	11.0	4.7	0.1	4.9	31%	31%	31%
22	74.7	17.8	0.5	18.3	15.0	0.4	15.4	2.9	0.1	3.0	16%	19%	16%
23	78.1	20.8	0.6	21.4	18.6	0.5	19.1	2.2	0.1	2.3	11%	15%	11%
24	69.1	13.3	0.3	13.6	9.6	0.2	9.8	3.7	0.1	3.8	28%	28%	28%
25	69.0	13.2	0.3	13.5	9.7	0.2	9.9	3.5	0.1	3.6	26%	28%	26%
26	68.8	13.1	0.3	13.4	10.7	0.3	11.0	2.4	0.1	2.4	18%	19%	18%
27	74.1	17.4	0.5	17.8	12.9	0.3	13.3	4.5	0.1	4.6	26%	28%	26%
28	72.1	15.6	0.4	16.0	14.0	0.3	14.3	1.6	0.1	1.7	10%	13%	11%
29	69.7	13.8	0.3	14.1	6.7	0.2	6.9	7.1	0.2	7.2	51%	52%	51%
30	70.0	14.0	0.3	14.3	11.3	0.3	11.6	2.7	0.1	2.7	19%	21%	19%
31	72.0	15.6	0.4	16.0	13.3	0.3	13.7	2.3	0.1	2.3	15%	15%	15%
32	70.0	14.0	0.3	14.3	13.3	0.3	13.6	0.7	0.0	0.7	5%	6%	5%
33	65.0	10.2	0.3	10.4	6.3	0.2	6.4	3.9	0.1	4.0	38%	40%	38%
34	70.4	14.3	0.4	14.7	12.3	0.3	12.6	2.0	0.1	2.0	14%	14%	14%
35	73.4	16.7	0.4	17.2	15.3	0.4	15.7	1.4	0.1	1.5	9%	12%	9%
36	65.3	10.4	0.3	10.6	7.5	0.2	7.7	2.9	0.1	3.0	28%	28%	28%
37	72.4	15.9	0.4	16.3	10.6	0.3	10.9	5.3	0.1	5.4	33%	35%	33%
38	68.5	12.8	0.3	13.1	8.4	0.2	8.6	4.3	0.1	4.4	34%	35%	34%
39	75.9	18.9	0.5	19.4	16.0	0.4	16.4	3.0	0.1	3.1	16%	20%	16%
40	72.4	15.9	0.4	16.3	15.1	0.4	15.5	0.8	0.0	0.8	5%	5%	5%
41	72.7	16.2	0.4	16.6	13.2	0.3	13.6	2.9	0.1	3.0	18%	20%	18%
42	71.8	15.4	0.4	15.8	12.6	0.3	12.9	2.8	0.1	2.9	18%	21%	18%
43	66.8	11.5	0.3	11.8	8.6	0.2	8.8	2.9	0.1	3.0	25%	25%	25%
44	72.6	16.1	0.4	16.5	13.6	0.3	13.9	2.5	0.1	2.6	16%	17%	16%
45	70.5	14.3	0.4	14.7	11.0	0.3	11.2	3.4	0.1	3.4	23%	23%	23%
46	70.8	14.6	0.4	14.9	11.8	0.3	12.0	2.8	0.1	2.9	19%	19%	19%
47	71.0	14.8	0.4	15.2	12.1	0.3	12.4	2.7	0.1	2.8	18%	17%	18%
48	73.2	16.6	0.4	17.0	12.0	0.3	12.3	4.6	0.1	4.7	28%	30%	28%
49	68.9	13.1	0.3	13.4	10.7	0.3	10.9	2.4	0.1	2.5	19%	19%	19%
50	74.6	17.7	0.5	18.2	15.7	0.4	16.1	2.0	0.1	2.1	12%	15%	12%
51	71.7	15.4	0.4	15.7	12.6	0.3	12.9	2.8	0.1	2.8	18%	18%	18%
52	73.4	16.7	0.4	17.2	15.0	0.4	15.4	1.7	0.1	1.8	10%	14%	10%
53	71.0	14.7	0.4	15.1	11.6	0.3	11.9	3.2	0.1	3.2	21%	22%	21%
54	72.0	15.6	0.4	16.0	11.8	0.3	12.1	3.7	0.1	3.8	24%	26%	24%
55	72.1	15.6	0.4	16.0	13.1	0.3	13.4	2.5	0.1	2.6	16%	18%	16%
56	70.0	14.0	0.3	14.3	11.6	0.3	11.9	2.4	0.1	2.4	17%	18%	17%
57	70.0	14.0	0.3	14.3	11.6	0.3	11.9	2.4	0.1	2.5	17%	18%	17%
58	70.2	14.2	0.3	14.5	11.0	0.3	11.2	3.2	0.1	3.3	22%	21%	22%
59	71.8	15.4	0.4	15.8	11.0	0.3	11.3	4.4	0.1	4.5	28%	29%	28%
Median	71.8	15.4	0.4	15.8	12.4	0.3	12.7	2.7	0.1	2.8	18%	19%	18%
Mean	71.3	15.1	0.4	15.5	12.2	0.3	12.5	2.8	0.1	2.9	19%	21%	19%
S.D.	2.9	2.3	0.1	2.4	2.7	0.1	2.8	1.2	0.0	1.3	9%	9%	9%
Max	78.7	21.3	0.6	22.0	18.6	0.5	19.1	7.1	0.2	7.2	51%	52%	51%
Min	64.3	9.6	0.2	9.9	6.3	0.2	6.4	0.0	0.0	0.0	0%	0%	0%

Cooling Season Energy (GJ)

USER	Baseline Temp. °F	Baseline Setpoints Cool	Baseline Setpoints Fan	Baseline Setpoints Cool +Fan	Observed Setpoints Cool	Observed Setpoints Fan	Observed Setpoints Cool +Fan	Difference Cool	Difference Fan	Difference Cool +Fan	% Difference Cool	% Difference Fan	% Difference Cool +Fan
1	73.0	2.7	0.8	3.5	2.5	0.7	3.2	0.2	0.1	0.3	8%	9%	9%
2	72.2	2.9	0.8	3.7	2.3	0.7	3.0	0.6	0.2	0.8	20%	20%	20%
3	67.0	4.0	1.2	5.2	3.2	0.9	4.1	0.9	0.3	1.1	21%	24%	22%
4	72.8	2.8	0.8	3.6	2.4	0.7	3.0	0.4	0.1	0.5	15%	16%	15%
5	72.0	2.9	0.9	3.8	2.5	0.7	3.2	0.4	0.1	0.6	15%	16%	15%
6	70.2	3.3	1.0	4.3	2.6	0.8	3.4	0.7	0.2	0.9	22%	22%	22%
7	72.0	2.9	0.9	3.8	2.4	0.7	3.1	0.6	0.2	0.7	19%	20%	19%
8	71.7	3.0	0.9	3.9	2.8	0.8	3.5	0.3	0.1	0.4	9%	10%	9%
9	76.3	2.0	0.6	2.6	1.5	0.5	2.0	0.5	0.1	0.6	24%	22%	23%
10	74.2	2.5	0.7	3.2	1.7	0.5	2.2	0.8	0.2	1.0	31%	32%	31%
11	73.7	2.6	0.7	3.3	2.1	0.6	2.7	0.4	0.1	0.6	17%	18%	17%
12	68.6	3.7	1.1	4.8	3.1	0.9	4.0	0.6	0.2	0.8	16%	17%	16%
13	70.0	3.4	1.0	4.4	2.9	0.8	3.7	0.5	0.2	0.6	15%	15%	15%
14	67.6	3.9	1.2	5.0	2.9	0.9	3.8	1.0	0.3	1.3	25%	26%	25%
15	70.3	3.3	1.0	4.3	2.3	0.7	2.9	1.0	0.3	1.4	31%	33%	32%
16	74.0	2.5	0.7	3.2	2.0	0.6	2.6	0.5	0.1	0.6	19%	19%	19%
17	72.8	2.8	0.8	3.6	2.7	0.8	3.4	0.1	0.0	0.1	4%	4%	4%
18	71.8	3.0	0.9	3.9	2.7	0.8	3.5	0.3	0.1	0.4	10%	10%	10%
19	74.0	2.5	0.7	3.2	2.0	0.6	2.5	0.5	0.2	0.7	22%	22%	22%
20	72.0	2.9	0.9	3.8	2.1	0.6	2.7	0.8	0.3	1.1	28%	29%	28%
21	71.0	3.2	0.9	4.1	2.7	0.8	3.5	0.5	0.1	0.6	15%	15%	15%
22	74.1	2.5	0.7	3.2	1.4	0.4	1.8	1.1	0.3	1.4	43%	44%	44%
23	72.0	2.9	0.9	3.8	2.6	0.8	3.4	0.3	0.1	0.4	11%	12%	11%
24	68.0	3.8	1.1	4.9	3.0	0.9	3.9	0.8	0.2	1.0	20%	21%	20%
25	71.0	3.2	0.9	4.1	2.6	0.7	3.3	0.6	0.2	0.8	19%	20%	19%
26	74.2	2.5	0.7	3.2	1.9	0.6	2.5	0.6	0.2	0.7	22%	24%	23%
27	70.2	3.3	1.0	4.3	2.6	0.8	3.4	0.7	0.2	0.9	20%	21%	21%
28	76.0	2.1	0.6	2.7	1.8	0.5	2.4	0.3	0.1	0.3	12%	12%	12%
29	73.0	2.7	0.8	3.5	2.4	0.7	3.1	0.3	0.1	0.4	11%	11%	11%
30	71.5	3.1	0.9	3.9	1.6	0.5	2.1	1.4	0.4	1.8	47%	47%	47%
31	69.9	3.4	1.0	4.4	3.0	0.9	3.9	0.4	0.1	0.5	11%	12%	12%
32	67.5	3.9	1.2	5.1	3.2	0.9	4.2	0.7	0.2	0.9	17%	19%	18%
33	70.0	3.4	1.0	4.4	3.0	0.9	3.8	0.4	0.1	0.5	12%	13%	12%
34	72.8	2.8	0.8	3.6	2.6	0.8	3.4	0.1	0.0	0.2	5%	5%	5%
35	67.5	3.9	1.2	5.1	3.5	1.0	4.5	0.4	0.1	0.6	11%	12%	11%
36	69.7	3.5	1.0	4.5	2.4	0.7	3.0	1.1	0.3	1.4	32%	33%	32%
37	72.7	2.8	0.8	3.6	2.5	0.7	3.3	0.3	0.1	0.3	9%	10%	9%
38	73.4	2.6	0.8	3.4	1.8	0.5	2.4	0.8	0.2	1.0	31%	29%	30%
39	71.0	3.2	0.9	4.1	2.8	0.8	3.6	0.4	0.1	0.5	13%	13%	13%
40	75.0	2.3	0.7	3.0	2.0	0.6	2.6	0.3	0.1	0.4	14%	14%	14%
41	73.0	2.7	0.8	3.5	2.5	0.7	3.3	0.2	0.1	0.2	7%	6%	7%
42	74.6	2.4	0.7	3.1	2.0	0.6	2.6	0.4	0.1	0.5	17%	17%	17%
43	71.0	3.2	0.9	4.1	2.9	0.8	3.7	0.3	0.1	0.4	10%	10%	10%
44	76.0	2.1	0.6	2.7	1.7	0.5	2.2	0.4	0.1	0.5	20%	20%	20%
45	74.8	2.3	0.7	3.0	1.9	0.6	2.5	0.4	0.1	0.6	18%	19%	18%
46	75.8	2.1	0.6	2.8	1.6	0.5	2.1	0.5	0.2	0.6	23%	24%	23%
47	73.5	2.6	0.8	3.4	2.3	0.7	2.9	0.4	0.1	0.5	13%	14%	14%
48	77.0	1.9	0.5	2.4	1.7	0.5	2.2	0.2	0.1	0.2	9%	9%	9%
49	74.7	2.4	0.7	3.0	1.5	0.4	1.9	0.9	0.3	1.1	37%	38%	37%
50	71.0	3.2	0.9	4.1	2.8	0.8	3.6	0.3	0.1	0.4	11%	11%	11%
51	74.0	2.5	0.7	3.2	2.1	0.6	2.7	0.4	0.1	0.6	17%	18%	17%
52	71.2	3.1	0.9	4.0	2.3	0.7	3.0	0.8	0.2	1.0	25%	26%	25%
53	74.3	2.4	0.7	3.2	2.0	0.6	2.5	0.5	0.1	0.6	19%	20%	19%
54	72.6	2.8	0.8	3.6	2.3	0.7	3.0	0.5	0.2	0.7	19%	20%	19%
55	77.1	1.9	0.5	2.4	1.1	0.3	1.5	0.7	0.2	1.0	39%	41%	40%
56	72.0	2.9	0.9	3.8	2.3	0.7	3.0	0.6	0.2	0.8	21%	22%	21%
57	70.0	3.4	1.0	4.4	2.8	0.8	3.6	0.6	0.2	0.8	18%	19%	18%
58	74.0	2.5	0.7	3.3	2.2	0.7	2.9	0.3	0.1	0.4	11%	11%	11%
59	69.0	3.6	1.1	4.7	2.8	0.8	3.6	0.8	0.3	1.1	23%	24%	23%
60	73.6	2.6	0.8	3.4	2.4	0.7	3.1	0.2	0.1	0.2	6%	8%	7%
61	68.1	3.8	1.1	4.9	3.5	1.1	4.6	0.3	0.1	0.3	7%	6%	7%
62	75.0	2.3	0.7	3.0	1.9	0.5	2.4	0.5	0.1	0.6	20%	20%	20%
63	77.9	1.7	0.5	2.2	1.2	0.3	1.5	0.5	0.2	0.7	31%	31%	31%
64	72.1	2.9	0.9	3.8	2.3	0.7	2.9	0.7	0.2	0.9	23%	24%	23%
65	73.0	2.7	0.8	3.5	2.4	0.7	3.1	0.3	0.1	0.4	11%	11%	11%
66	71.0	3.2	0.9	4.1	2.9	0.8	3.7	0.3	0.1	0.3	8%	9%	8%
Median	72.4	2.9	0.8	3.7	2.4	0.7	3.1	0.5	0.1	0.6	18%	19%	18%
Mean	72.3	2.9	0.8	3.7	2.4	0.7	3.0	0.5	0.2	0.7	18%	19%	18%
S.D.	2.5	0.5	0.2	0.7	0.5	0.2	0.7	0.3	0.1	0.3	9%	9%	9%
Max	77.9	4.0	1.2	5.2	3.5	1.1	4.6	1.4	0.4	1.8	47%	47%	47%
Min	67.0	1.7	0.5	2.2	1.1	0.3	1.5	0.1	0.0	0.1	4%	4%	4%

Appendix C: User Setpoint History Data

User setpoint histories during two weeks for the heating and cooling seasons. Horizontal gray line indicates each user's 10th or 90th percentile setpoint for heating and cooling seasons, respectively.

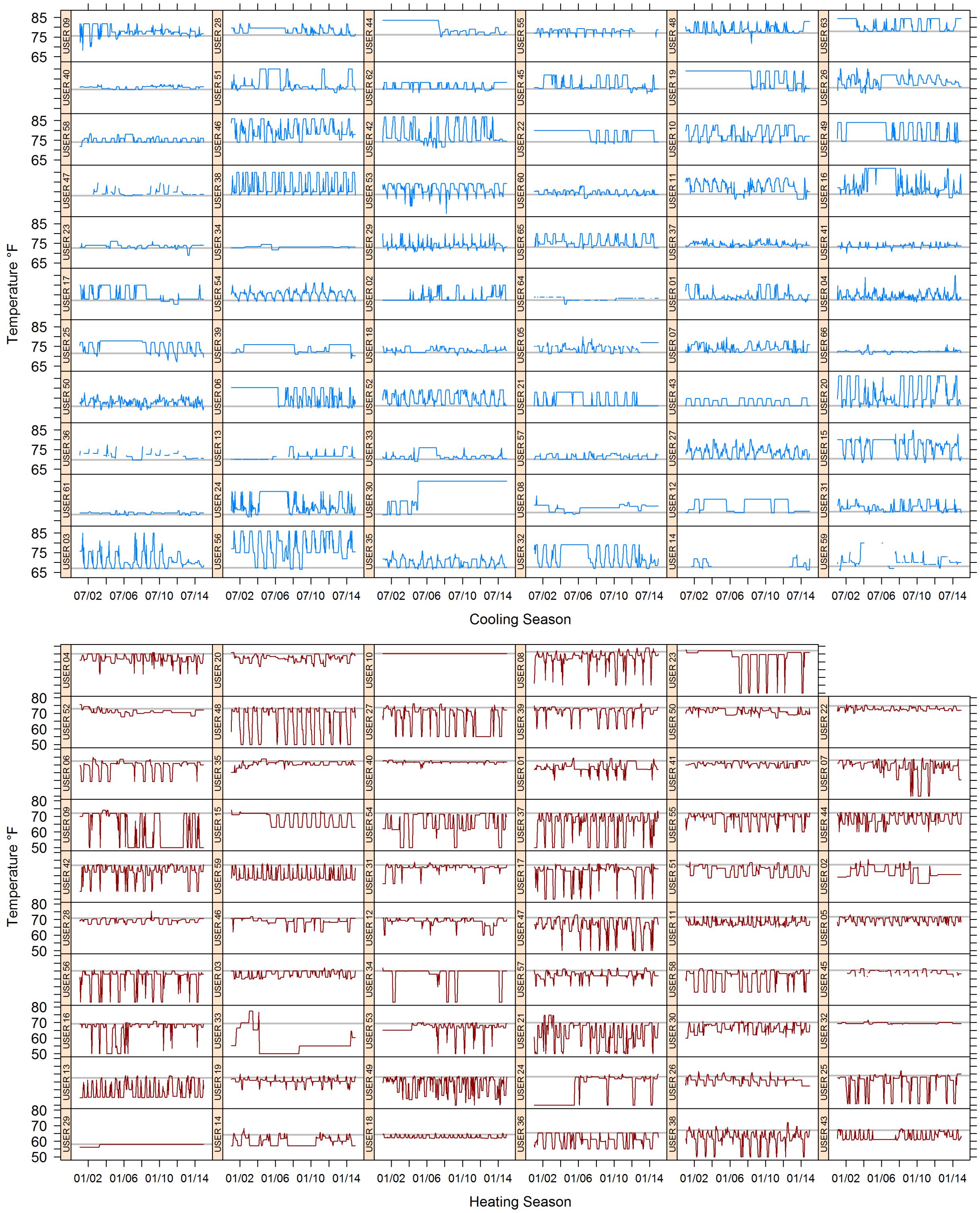


Figure 15: User setpoint histories for two weeks during the heating and cooling season.