November 16, 2021

REPORT #E21-324

Northwest Smart Thermostat Research Study

Prepared For NEEA:
Tamara Anderson, Sr. Program Coordinator

Prepared by:
Lauren Miller Gage, Principal
Jon Koliner, PhD

Apex Analytics, LLC
apexanalyticsllc.com

and

Hassan Shaban, Principal, Energy & Data

Empower Dataworks
216 Ridgemont Dr. East
Wenatchee, WA 98802

Northwest Energy Efficiency Alliance
PHONE
503-688-5400
EMAIL
info@neea.org

Table of Contents

Acknowledgements i
Executive Summary 1
  Key Findings 3
  Future Research Considerations 7
1. Background and Goals 9
  1.1 Background 9
  1.2 Study Objectives 10
  1.3 Other Relevant Thermostat Studies 11
2. Study Methodology 12
  2.1 Data Collection 12
  2.2 Thermostat Telemetry Analysis 15
  2.3 Billing Analysis Approach 17
  2.4 Correlation Analysis Approach 21
3. Sample Characteristics and Survey Results 22
  3.1 Thermostat User Opt-Ins 22
  3.2 Survey Results 23
  3.3 Merged Sample of Thermostat and Billing Data 25
  3.4 Site-Level Sample Data 27
  3.5 Pooled Analysis Data Set 29
4. Energy Savings Analysis Results 29
  4.1 Thermostat Telemetry Analysis 29
  4.2 Billing Analysis 35
  4.3 Correlation of Metrics with Savings 43
5. Key Findings and Future Research Considerations 49
Appendix 1: Telemetry Analysis Memo Addendum 51
  New Core Thermostat Metrics 51
  Additional Test Metrics 53
Appendix 2: Billing Analysis Details 60
Appendix 3: Correlation Analysis Details 60

Acknowledgements

This study is a collaboration between NEEA, Avista Power, the Bonneville Power Administration, Chelan County PUD, Clark Public Utilities, Energy Trust of Oregon, Idaho Power, Northwest Power & Conservation Council, Puget Sound Energy, Seattle City Light, Snohomish County PUD, and Tacoma Power.

The research was led by Apex Analytics, in partnership with Empower Dataworks, CLEAResult, and Energy 350 (the Apex team).

This study would not have been possible without the collaboration with the thermostat manufacturers who voluntarily provided data to this study (Nest/Google, ecobee, Emerson and Resideo); the organizations that provided billing data for this study (Avista Power, Clark Public Utilities, Energy Trust of Oregon, Puget Sound Energy); the codebase and assistance from ENERGY STAR® staff; the guidance of the study's Advisory Team; and NEEA staff's dedication, support, and management.

Executive Summary

In 2018, a group of Northwest utilities and energy efficiency organizations began planning a regional research project for smart thermostats[1]. Smart thermostat models were coming on to the market quickly, with new and various features. Yet, using the common evaluation method of pre-post billing analysis of participants required one to two years to capture energy savings in both seasons. Concurrently, the Environment Protection Agency, via their ENERGY STAR® Connected Thermostat initiative, had established energy savings criteria based on thermostat performance metrics (i.e., thermostat metrics) such as runtime reduction compared to a specified baseline condition.

The Northwest group (i.e., NEEA, Avista Power, the Bonneville Power Administration, Chelan County PUD, Clark Public Utilities, Energy Trust of Oregon, Idaho Power, Northwest Power & Conservation Council, Puget Sound Energy, Seattle City Light, Snohomish County PUD, and Tacoma Power; collectively, the Advisory Team) envisioned that there could be an improved process to enable Northwest utilities to quickly screen new products for inclusion in Qualified Products Lists (QPLs) and estimate energy savings without repeated one-off evaluations. The envisioned method was to align with the ENERGY STAR process and data requirements, using data from regional studies as available, particularly for baseline usage. With a newly developed method, the Northwest would be able to more efficiently calculate expected energy savings from these thermostat metrics.

Therefore, the **primary objective of this study was to develop a method to estimate energy savings for smart thermostats based on thermostat performance metrics** (i.e., hourly thermostat-gathered operational data such as hourly runtimes and indoor temperatures). To do this, the Apex team (Apex Analytics, LLC; CLEAResult; and Energy350) attempted to establish if a statistically significant relationship between energy savings from billing data analysis and thermostat metrics exists and was stable across thermostat characteristics. Secondary objectives included determining regional baseline thermostat behavior and determining energy savings from smart thermostats with sufficient reliability while controlling for other changes through a comparison group.

The major steps in the project included:

1) **Enroll Manufacturers and Opt-In Thermostat Users**. The project recruited manufacturers to collaborate with the study, run the updated software, and provide thermostat metrics for thermostat users. To provide thermostat-

---

[1] Per Regional Research Strategy, smart thermostats are defined as programmable, internet-connected devices that incorporate occupancy sensing; adaptive control to optimize performance based on user behavior and weather conditions; and control for electric heat pumps, gas forced air furnaces, electric forced air furnaces, and central A/C systems.

specific data, several manufacturers needed signed participation agreements from thermostat users. To obtain these legal agreements, the study created an opt-in process with nine separate, customized websites. The Apex team contacted over 50,000 Northwest residential customers, and 1,800 opted in to share their data and completed a survey of their concurrent energy-using activities (e.g., heating system upgrades, purchase of an electric vehicle). Manufacturers also provided data for 2,100 additional customers under existing agreements associated with setpoint optimization programs.

2) **Perform Telemetry Analysis and Software Modification**. To collect thermostat performance metrics relevant for the Northwest, the Apex team analyzed a set of anonymous thermostat data to modify the ENERGY STAR Thermostat Field Savings software. This step resulted in a modified version of the ENERGY STAR software to calculate new metrics for the study, including runtime difference using a regional indoor temperature baseline, temperature change rates with and without HVAC, excess resistance score, and additional resistance heat metrics for heat pumps.

3) **Collect Utility and Manufacturer Data.** Based on the enrolled thermostat users, the Apex team collected billing data from four regional utilities and thermostat metrics from four manufacturers. The resulting merged data set included approximately 1,000 thermostats for the final analysis dataset covering January 2017–February 2020. These thermostat users included participants and non-participants in utility thermostat programs who primarily used gas furnaces to heat their homes and mainly lived in Northwest Heating Zone 1[2]. Although the Apex team attempted to develop a large and representative sample of thermostat users in the Northwest, the complexity of data collection limited the sample size and characteristics. Thus, this study’s sample is likely not representative of the region.

4) **Conduct Billing Analysis**. To estimate energy savings, the Apex team conducted a pooled regression analysis of billing data using future thermostat users as a comparison group and a site-level normalized annual whole-home billing analysis.

---

[2] Climate Zones | Regional Technical Forum (nwcouncil.org). Heating Zone 1 is a mild climate of the western portions of Oregon and Washington, with dense population centers such as Portland and Seattle.

5) **Correlate Metrics and Savings.** Finally, the Apex team conducted a correlation analysis using the site-level energy savings and all thermostat performance metrics.[3]

Key Findings

Using the data collected and analyzed as described above, our study had three key findings, as described below.

Smart Thermostat Installation Resulted in Statistically Significant Energy Savings

The Apex team conducted a pooled analysis of energy savings using future participants as a comparison group[4] as shown in Table ES-1. The Apex team found statistically significant gas savings (43 therms, 5% of total gas use) for sites with gas furnaces and statistically significant electric savings (670 kWh, 4.5%) for sites with heat pumps and electric backup. The Apex team found a moderate increase in electric use (220 kWh, 2.4%) for homes with gas furnaces, potentially representing an increase in either furnace fan usage or cooling usage. For all three groups, the Apex team estimated and removed the effect of post-installation setpoint optimization[5] during portions of the post-period. Due to sample characteristics, these results should not be interpreted as representative of the region.

---

[3] The correlation analysis does not require that the sample represents the population and benefited from inclusion of multiple thermostat models with a breadth of thermostat characteristics. The correlation holds even for customers receiving optimization, because optimization should impact both thermostat metrics and energy savings.

[4] Installations dated from January 2018 to March 2020, with 2017 serving as pre-period and post-treatment analysis window of March 2019 to February 2020. An alternate pooled analysis post-period analysis window is presented in Appendix 2.

[5] Post-installation setpoint optimization is when either a manufacturer or a third-party provider remotely accesses the thermostat (usually with user consent) and nudges it to a more efficient schedule.

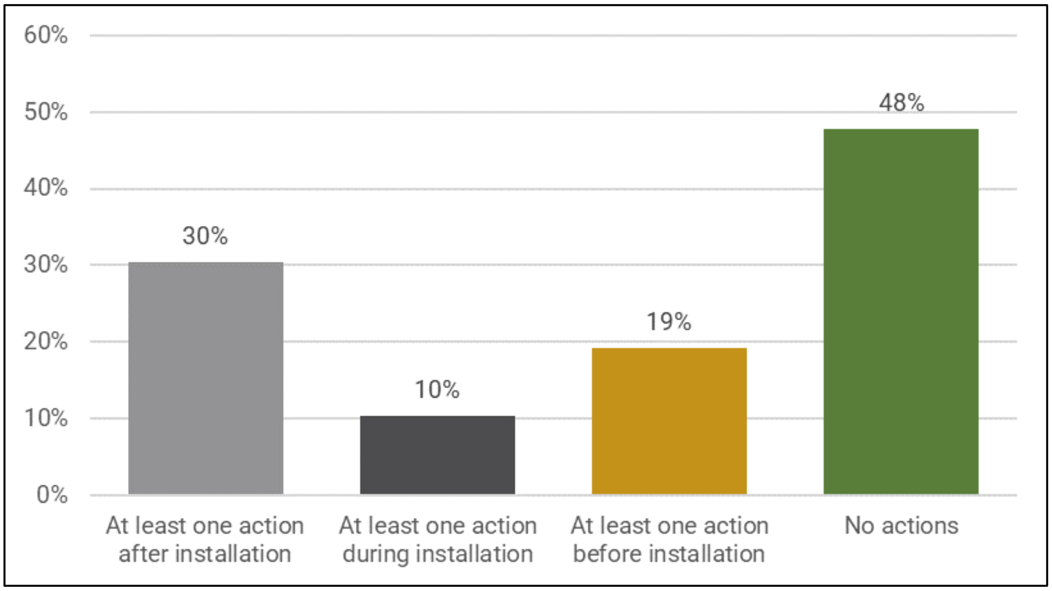
Table ES-1. Pooled Analysis Results

Heading System Type	Fuel	n	Post-Installation		
			Average Savings	Std Error of Avg Savings	Savings as % of Whole-Home Usage
Gas Furnace or Boiler	Gas	678	43 therms	20 therms	5%
	Electricity	550	-220 kWh	110 kWh	-2.4%
Heat Pump with Electric Backup	Electricity	73	670 kWh	402 kWh	4.5%

Major Home and Life Changes Occurring in a Similar Timeframe to Thermostats Impact Energy Savings Substantially

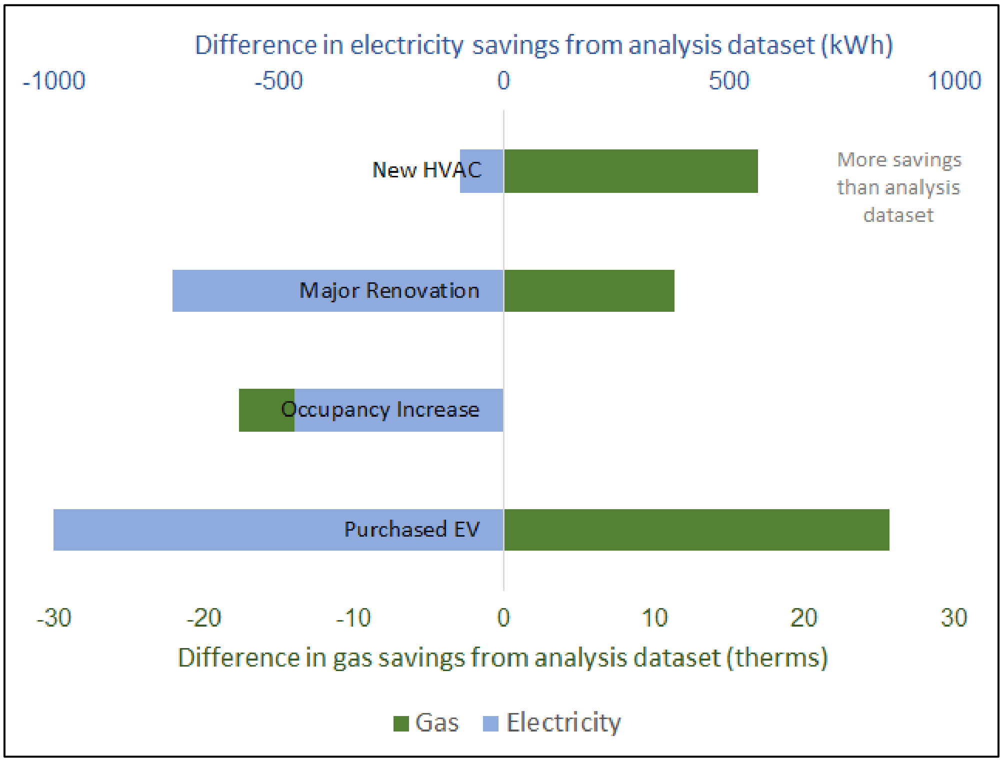
During the opt-in process, this study surveyed thermostat users and found that residents tend to conduct other major energy-changing behaviors after, during, or before installing thermostats (e.g., purchase of an electric vehicle, home renovations, HVAC system changes or occupancy changes), as shown in Figure ES-1.

Figure ES-1. Major Energy Use Changing Actions by Thermostat Users (All Activities Combined)



Using a site-level analysis, the study calculated the approximate energy savings impact of these activities (reported as during or after thermostat installation) compared to thermostat users who did not report engaging in these activities. The total sample in each of these groups is not sufficiently large to make statistically significant conclusions (see Section 4.2 for more information). However, the Apex team found that thermostat users with a new HVAC system had approximately 17 therms more in gas savings and 100 kWh less in electric savings than the main analysis, while additional occupants in the home erased or reversed all detectable gas and electric savings. Both electric vehicle purchases and major renovations increased electricity use substantially (700 to 1,100 kWh), although both groups saved additional therms compared to the main dataset (see Figure ES-2).

Figure ES-2. Energy-Use Changes due to Concurrent or Subsequent Activities (Individual Activities)

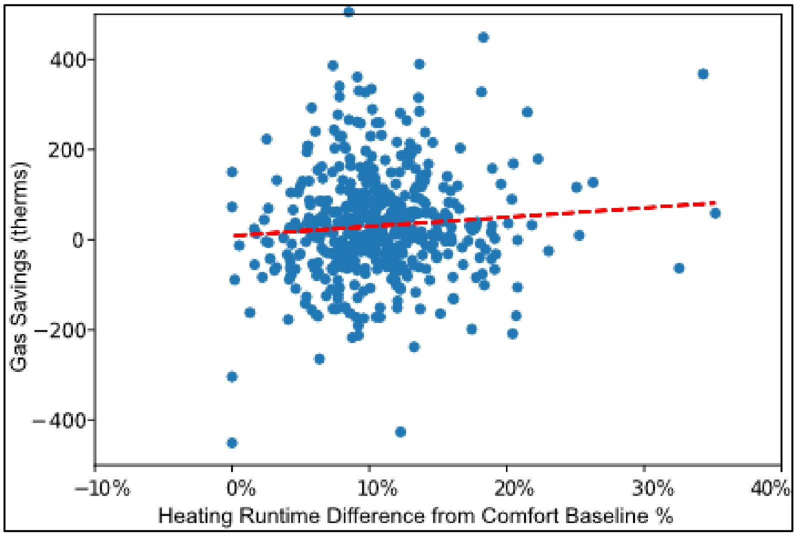


Based on this finding, the Apex team removed thermostat users who reported energy-changing behaviors at the same time or after smart thermostat installation[6]. Energy-changing behaviors, if not controlled for, could bias billing analysis of consumption. The billing analysis and correlation analysis in this report represent those customers who did not report these major changes in the participant survey questions.

Energy Savings Were Insufficiently Correlated with Thermostat Metrics to Establish a Method.

Using the standard ENERGY STAR metrics, the new thermostat metrics, and site-level savings, the Apex team conducted numerous correlation analyses of various thermostat metrics with site-level savings and found very weak or no correlation. This lack of strong correlation persisted across thermostat metrics, including new metrics with regional baselines, and was not dependent on thermostat models. For example, for gas furnaces, the raw correlation between thermostat metrics and savings is very weak[7] as shown in Figure ES-3 below. The correlation for heat pumps was statistically insignificant.

Figure ES-3. Site-Level Gas Savings (Gas Furnaces and Boilers) versus Runtime Reduction with a Comfort Baseline, with a Line of Best Fit



---

[6] Due to concerns regarding survey fatigue that could have negatively impacted thermostat user willingness to share data with the study, the Apex team worked with the Advisory Team to constrain the set of questions to the highest priority questions. Therefore, the study was not able to collect or analyze all desired energy activity changes. There may be other exogenous changes that could affect savings estimates.

[7] Out-of-sample bias error for savings (95% confidence interval): -38% to +99%.

In summary, none of the models attempting to correlate thermostat metrics with site-level savings generated more than a weak correlation. Although models for the primary heating fuel of gas furnaces (i.e., natural gas) were suggestive of underlying relationships that are roughly consistent with estimated savings (Figure ES-1), the regression coefficients had wide uncertainty bands. The resulting correlations were not strong enough to function as bases for estimating savings for a QPL or to distinguish thermostat differences. Therefore, the Apex team could not establish a method to use thermostat-derived metrics to estimate these energy savings with sufficient reliability for use by Northwest utilities.

There are two likely causes of weak or non-existent correlation:

- The variation in site-level normalized annual consumption (ΔNAC) is large and often unrelated to smart thermostat installation (i.e., it is related to occupant behavior and other end uses). If this inherent variation is larger than the impact of the thermostat, the explanatory power of the model is necessarily limited.
- The runtime reduction metrics do not incorporate information about the site-level pre-period baselines. Therefore, it is likely that using inequivalent baseline calculations—a pre-post method for energy use and a post-only method for runtime reduction metrics—is a driver of weak correlation.

The Apex team expects that, while a larger study could reduce uncertainty in the correlations, the two factors listed above would limit the ability of program administrators and other organizations to use thermostat-sourced metrics to reliably predict energy savings for given groups of thermostats.

Future Research Considerations

The Apex team's future research considerations are as follows:

- **Controlling for Major Life Changes.** It is important to control for major life changes to establish accurate energy savings for smart thermostats. As shown above, major life changes, if not controlled for, could influence energy savings using billing analysis. The Apex team recommends future research control for these factors using surveys or other methods to detect major energy-use changes.
- **Integrating Baseline Information.** For comparing thermostat metrics and site-level energy savings, information on the true baseline (prior

thermostat type, setpoints, and/or monitored indoor temperatures) is likely necessary. Relying on post-installation data alone will require samples of thermostats with sufficiently differentiated metric values, either by improving these metrics with additional information or by vastly increasing the size and therefore differentiation power of the study.

- **Future Design Considerations.** If organizations want to correlate thermostat metrics and energy savings in the future, entities will need legal and technical infrastructure in place with each thermostat manufacturer and potentially with customers. Future designs could gain customer and manufacturer agreements on the program front-end.
- **Thermostat Metrics Opportunities.** There is potential to use some of the additional thermostat metrics related to building shell and HVAC performance for behavioral messaging or HVAC diagnostics by energy efficiency programs. There is also potential to conduct additional research with the existing anonymous data set of ΔNAC and thermostat metrics, which is included with this report as **Appendix 4.**

1. Background and Goals

1.1 Background

In 2018, a group of Northwest utilities and energy efficiency organizations began planning a regional research project for smart thermostats. This group refined and expanded upon the Northwest Regional Technical Forum's 2016 Connected Thermostat Research Strategy[8] to develop the smart thermostat Research Strategy[9] (Research Strategy) for this study. Using this Research Strategy as a basis, this group competitively solicited a contractor to conduct this work on behalf of the Northwest region.

This project is a collaboration between NEEA, Avista Power, the Bonneville Power Administration (BPA), Chelan County PUD, Clark Public Utilities, Energy Trust of Oregon, Idaho Power, Northwest Power & Conservation Council, Puget Sound Energy, Seattle City Light, Snohomish County PUD, and Tacoma Power (collectively, the Advisory Team). NEEA managed this research project on behalf of the funding organizations.

The Research Strategy defines smart thermostats as programmable, internet-connected devices that incorporate the following features:

- Occupancy sensing (e.g., proximity, geo-fencing, or other techniques to determine occupancy).
- Adaptive control to optimize performance based on user behavior and/or weather conditions.
- Control of electric heat pumps, gas forced air furnaces, electric forced air furnaces, and central A/C systems.

The regional research group sought to determine a rigorous method to establish energy savings for these thermostats. They recognized that smart thermostat models were coming on the market quickly with new features. Yet, using the common evaluation method of pre-post billing analysis requires nearly two years of combined pre- and post-installation data to establish energy savings. Also, basing the savings determination on features (e.g., on-board occupancy sensing) was difficult because manufacturers deploy proprietary control algorithms to generate energy savings. Therefore, the group was interested in finding a new way to establish energy savings.

---

[8] https://nwcouncil.app.box.com/s/v73hd6fq07zuspgs20hi5pybl5wrzci7

[9] https://conduitnw.org/pages/file.aspx?rid=4697

Concurrently, the Environmental Protection Agency, via their ENERGY STAR® initiative, had established energy savings criteria based on thermostat performance metrics (i.e., thermostat metrics) such as runtime reduction compared to a specified baseline condition, and had plans to incorporate resistance heat utilization in the future.[10] Therefore, the group envisioned that there could be an improved process to enable Northwest utilities to quickly screen new products for inclusion in Qualified Products Lists (QPLs) and estimate energy savings without repeated one-off evaluations. The envisioned method was to align with ENERGY STAR Connected Thermostat process and data requirements, using regional data as available, particularly for baseline usage. Data from new and existing thermostat models would be periodically collected from random samples of field installs to calculate up-to-date thermostat metrics. With a newly developed method, the Northwest would then be able to calculate expected energy savings from these thermostat metrics.

1.2 Study Objectives

As defined in the Research Strategy, the **primary objective of this study was to develop a method to estimate energy savings for smart thermostats based on thermostat metrics**. To do this, the Apex team (Apex Analytics, LLC; CLEAResult; and Energy350) attempted sought to establish if a statistically significant relationship between savings and thermostat metrics exists and is stable across thermostat models.

Additional objectives include:

- Determine regional baseline thermostat behavior through the proxy measurement of indoor temperature profiles in the Northwest. (See Section 4.1 for results.)
- Determine energy savings from specific smart thermostat products and applications with sufficient reliability, controlling for confounding variables and exogenous effects through a comparison group and additional site-level data. (See Section 4.2 for results.)
- Establish thermostat metrics based on smart thermostat-reported data that correlate to energy savings and provide a rationale for how savings

---

[10] Smart Thermostats Key Product Criteria | Products | ENERGY STAR includes Connected Thermostat Energy Savings Criteria based on thermostat metrics such as runtime reduction and resistance heat utilization. https://www.energystar.gov/products/heating_cooling/smart_thermostats/key_product_criteria

are achieved, both through setpoint changes that reduce system runtime and the reduction of reliance on backup heat in heat pumps. (See Section 4.3 for results.)

- Create a variable speed heat pump smart thermostat inventory. (See NEEA website for results.[11])

1.3 Other Relevant Thermostat Studies

The Apex team performed a brief literature review of public studies to assess current state of savings estimates for smart thermostat installation. Several recent smart thermostat studies have estimated smart thermostat energy savings and are used to compare to this study's pooled results in Section 4.2. Most of the studies have found that installing smart thermostats results in both gas and electric energy savings.

- Energy Trust of Oregon and Recurve conducted a thermostat study in the Northwest between 2015 and 2017 using Nest and ecobee thermostats on single-family homes using forced-air heating systems in Oregon[12]. They found that the overall average gas savings in gas-heated homes was 3% to 4% of whole-home gas usage. The study also analyzed electric savings from gas heated homes and found that they saved 2% to 3% of baseline electricity usage.
- California Public Utilities Commission (CPUC) conducted an impact evaluation of direct install smart thermostats in the residential sector for program year 2019[13]. Direct install participants typically have multiple measures installed at once, so the study employed a method to disaggregate whole-home savings into measure-level savings. CPUC found a -1% whole-home gas savings rate and a 2% whole-home electric savings for single-family homes. Note that this study sample is weighted towards Southern California, which has a low demand for winter heating.

---

[11] https://neea.org/resources/variable-speed-heat-pump-smart-thermostat-findings

[12] Recurve; Summary of Recurve Residential Smart Thermostat Impact Analysis. https://www.energytrust.org/wp-content/uploads/2020/02/Recurve-Smart-Thermostat-Impact-Analysis-Reports-2015-2017.pdf

[13] California Public Utilities Commission; Impact Evaluation of Smart Thermostats. https://pda.energydataweb.com/api/downloads/2508/CPUC%20Group%20A%20Residential%20PY2019_SCT%20Final%20Report_toPDA.pdf

- ComEd and Ameren Illinois ran an advanced thermostat study to find the impact of residential advanced thermostats on cooling season electric consumption. In the study, they used econometric analysis and an adjusted ENERGY STAR analysis to estimate savings[14]. The econometric analysis found a 3% to 6% whole-home electric energy savings and a 10% to 16% peak demand reduction savings depending on the method of the regression. The ENERGY STAR analysis found 3% to 29% cooling runtime savings, and the study noted the inconsistency with billing analysis.
- BPA and Franklin Public Utility District ran a Nest Learning thermostat pilot program on single family home with heat pumps in the Northwest[15]. They utilized a pooled fixed effects model, a normalized annual consumption (NAC) model, and an ECAM+ model. All models found a 4% savings of total kWh consumption, and the NAC model found a 12% savings of estimated heating and cooling load.

2. Study Methodology

This section describes the study data collection and the methodology for each analysis involved.

2.1 Data Collection

2.1.1 Manufacturer and Utility Recruiting

To accomplish the primary objective, NEEA and the Advisory Team wanted to ensure that the study had a sufficiently large, diverse sample of thermostat users in the Northwest with matched billing data and thermostat metrics. This requirement meant that the study needed to recruit, collect, and join data from multiple sources. Key steps in this process included:

- **Manufacturer Recruiting.** The study's goal was to recruit and include four manufacturers in the study. The recruiting process required approximately 18 months of communications with manufacturers regarding study

---

[14] Guidehouse; ComEd Advanced Thermostat Evaluation. https://ilsag.s3.amazonaws.com/ComEd-Adv-Thermostat-Research-Report-Final-2020-11-10.pdf

[15] Bonneville Power Administration & Franklin Public Utility District. https://www.bpa.gov/EE/Utility/research-archive/Documents/BPA%20-FPUD-Nest-Thermostat-Pilot-Savings-Assessment.pdf

methodology, data requests, legal agreements, and other specifics to support manufacturer participation. Ultimately, the study recruited and collected data from four manufacturers who contributed greatly to this study including NEST/Google, ecobee, Emerson, and Resideo. Several of these manufacturers were unable to release thermostat metrics on individual customers without customer consent, which required that the Apex team establish opt-in processes for each manufacturer's customer base (see Section 2.1.2).

- **Utility Recruiting.** The study's goal was to include three to four utilities with regional weather variations that could match thermostat users to billing data and provide this information to the study. Four of the funding utilities committed to providing data for the study: Avista Utilities, Clark Public Utilities, Puget Sound Energy, and Energy Trust of Oregon (representing Pacific Power and Portland General Electric). Similar to manufacturers, several of these utilities needed customer consent to release billing data (see Section 2.1.2).

2.1.2 Thermostat User Opt-In Process

To address the needs of utilities and manufacturers to have signed customer consent before sharing individual customers' billing and thermostat metrics, the study designed an opt-in process for thermostat users. This opt-in process included:

- **Manufacturer-Specific Requirements**. Several manufacturers had specific requirements for the participation agreement language and/or approach (e.g. users needed to validate through an OAuth digital authentication process). Additionally, some manufacturers were able to send recruiting emails to thermostat users, while others needed utilities to conduct recruiting. Finally, two manufacturers provided data for customers who had existing signed agreements through optimization pilots.
- **Utility-Specific Requirements.** Utilities also had specific needs in terms of participation agreement approach and agreement terms to gain customer consent. For those thermostat manufacturers that needed utilities to recruit customers, each utility had a different approach to recruitment (e.g., emails, mail).
- **Nine Websites.** The various requirements by manufacturer and utility combination led to the development of nine different websites for the opt-in of thermostat users.

- **Survey of Thermostat Users.** To understand if there were major energy-use changes in the household at a similar time or after thermostat installation, the opt-in process included the following short survey of thermostat users. Note that the questions presented were a compromise between suspected relevant data and a minimal impediment to opt-in, so not all possible complicating actions are represented. Users were asked if between January 2017 and February 2020 (before the COVID-19 outbreak) they conducted any of the following and the timing (i.e., after, during, or before thermostat installation).
    - *Did you install new heating and/or cooling equipment (e.g., furnace, central AC, window AC; water heater is not applicable)?*
    - *Did you purchase an electric vehicle?*
    - *Did you complete a major renovation that may affect your energy use (e.g., new windows, insulation, remodeling, additions)?*
    - *Was there an increase in the number of people living in your home?*
    - ***Nest users only:** Did you install any other smart home devices (e.g., smart speakers, smart lights, home displays, or home mesh wifi systems)? [Additional option of selecting "multiple times" for installation time]*
- **Additional Opt-In Data.** Depending on the thermostat manufacturer and utility, participants were required to input additional data at opt-in in order to enable matching of billing data, such as serial number, address, or email.

To encourage opt-ins, the study offered a donation to a charity of the thermostat users' choice, including Red Cross of America, Alliance to Save Energy, and Feeding America. The opt-in process and websites were deployed between September and December 2020.

2.1.3 Utility and Manufacturer Data Collection

After thermostat users opted into the study, utilities found those participants in their system and provided unique identifiers (e.g., study ID or email) and monthly billing data (date, consumption, therms/kWh) for the period January 2017 to February 2020[16].

The study requested additional information from utilities such as heating system type, housing information, and program participation information, but these data were not provided for the majority of thermostat users.

---

[16] This date was selected to avoid the impacts of the COVID-19 pandemic.

Manufacturers provided anonymous data sets early in the study period to support the thermostat telemetry analysis and modification of the ENERGY STAR software (see Section 3.2 below). For the correlation analysis, manufacturers provided performance metrics based on telemetry data for the study participants, as well as other pieces of information such as heating system type, thermostat installation date or date range, presence of on-board occupancy sensing, and the timing of any setpoint optimization after installation.

2.1.4 Other Data Sets

The Apex team also acquired and processed other data sets for inclusion in this report.

- **Regional Weather Data.** Used in calculations of thermal demand (indoor-outdoor temperature difference): https://eeweather.readthedocs.io/en/latest/
- **Indoor Temperature Monitoring Study (Cadmus + RTF).** Used to develop regional hourly temperature baselines by climate zone and heating system type: https://conduitnw.org/Pages/RETACProject.aspx?rid=14
- **Residential Building Stock Assessment (RBSA) Metering Study (NEEA).** Used to develop a regional baseline for demand-normalized resistance utilization (DNRU): https://neea.org/data/nw-end-use-load-research-project

2.2 Thermostat Telemetry Analysis

Modification of ENERGY STAR Software and Anonymous Data Analysis

To support the primary objective of developing a method to estimate energy savings based on thermostat metrics, the Research Strategy identified that this study should analyze on the existing ENERGY STAR software metrics and new, regional-specific metrics.

Thermostat manufacturers collect hourly time series of indoor and outdoor temperatures and runtimes for the different heating and cooling stages (i.e., telemetry data). These raw data are exported into a formatted file to be the primary input for the ENERGY STAR software. Both the original ENERGY STAR software and the ENERGY STAR software modified for this study generate a set of summary metrics from the raw data. These are the thermostat metrics referenced throughout this report. The thermostat metrics are summary results per thermostat, with separate metrics for the cooling and heating season.

The Apex team received hourly anonymous telemetry data from three manufacturers to understand the hourly data available and their relationship with existing ENERGY STAR metrics. This data enabled the team to test and create new metrics to assess whether they reflected observable patterns in the data. It also allowed the Apex team to ensure that the input telemetry data from the manufacturers worked correctly with the software.

A preliminary set of metrics was recommended in the smart thermostat Research Strategy developed by the smart thermostat Advisory Team. These were fully described, with small modifications, in the Telemetry Analysis Memo Addendum (**Appendix 1: Telemetry Analysis Memo Addendum**). The Apex team developed several additional metrics after investigating anonymous thermostat data from, and receiving input from the Advisory Team. These new metrics are described in the Data Analysis Implementation Attachment (**Appendix 1: Telemetry Analysis Memo Addendum**). All new metrics were integrated into a separate version of the ENERGY STAR Connected Thermostat codebase, which was published[17] by the Apex team for use by the thermostat manufacturers.

The list of thermostat metrics used in the analysis includes:

- Metrics in original ENERGY STAR software
    - Runtime difference from comfort baseline: How different is actual HVAC runtime from a baseline that assumes the indoor temperature is always maintained at a customer-specific empirically derived comfort temperature?
    - Model fit metrics that diagnose the ENERGY STAR thermal demand model, including thermal demand model coefficients, model fit, and the number of core heating and cooling days.
    - Resistance Heat utilization in different temperature bins: For heat pumps, how much does the electric resistance heating kick in the colder it gets?
- Metrics developed for this study (See **Appendix 1: Telemetry Analysis Memo Addendum**)
    - Runtime difference from the Northwest hourly indoor temperature baseline: How different is actual HVAC runtime from the runtime needed to maintain a regional average baseline of hourly indoor temperature?
    - HVAC and no-HVAC temperature change rates: How quickly does the home lose heat in winter or gain heat in summer?

---

[17] https://pypi.org/project/thermostat-nw/

  - Integral of sigmoid resistance function: How much resistance heat is used by a heat pump in the 0–60°F temperature range?
  - Excess resistance score: How much of the home's thermal load is met by resistance heat when the heat pump compressor could have met it?
  - DNRU reduction: For heat pumps, how different is the demand-normalized resistance heat utilization compared to a regional baseline?
  - Linear model between thermal demand and HVAC runtime (model fit metrics were used as data quality filters): Is indoor temperature driven by the HVAC system? Or are there other drivers?

Resulting Thermostat Metrics Analysis

Three manufacturers successfully used the modified software to generate thermostat metrics for customers in the opt-in dataset, and the Apex team reviewed the resulting dataset. The Apex team generated the metrics for the fourth manufacturer. The Apex team also requested and received supplemental data that identified sensor-based or geofencing-based occupancy sensing (ecobee and Resideo), thermostat installation date (all manufacturers), and participation dates in optimization programs (ecobee and Resideo). These supplemental data points are not typically provided in the ENERGY STAR process.

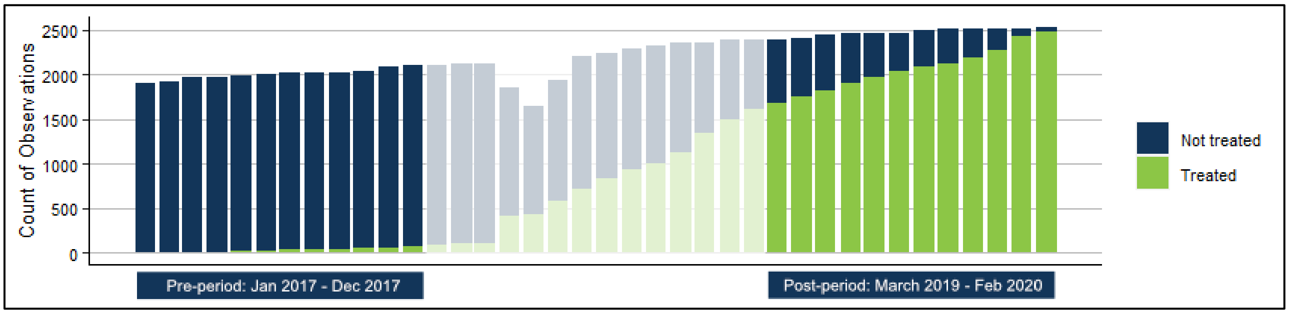
The metrics and additional data received from the manufacturers were then merged with energy consumption provided by the utilities. Depending on the opt-in method, the joining was performed by participant ID or email address. When the merging process was completed, the Apex team had a master dataset of thermostat metrics and energy consumption, which could then be used for subsequent stages of the analysis.

2.3 Billing Analysis Approach

Billing data provided by the utilities was analyzed in two ways for this study: within a pooled analysis and at the site level. The primary study objective requires a correlation between site-level change in NAC (ΔNAC, or site-level savings) and metrics produced from thermostat data. Secondary study objectives included establishing energy savings from smart thermostat products, which was best performed with pooled regression analysis that incorporates future thermostat installers as a comparison group in the post-period.

The available data spanned three years, with installations occurring across the second two years, but only two years were needed for analysis. For both analyses, the Apex team used the latest post-period window possible to maximize the data available for correlation. Therefore, the 12-month timespan used for the post-period is the "post-period window" (March 2019–February 2020). The analysis window for both pooled and site-level models is shown in Figure 1.

Figure 1. Counts of Customers with and without Thermostats Installed, with the Window Excluded from Analysis Greyed Out



Pooled Analysis

Pooled analysis produces aggregate energy-use change estimates and baseline offsets (see the following section) to account for other exogenous energy-use changes during the same time frame using a comparison group of future thermostat installers. The Apex team aligned the formatting of the four utilities' data, cleaned, and calendarized (i.e., aligned with calendar months) if necessary (see Section 3 for sample data attrition details). The data were calendarized by calculating average daily consumption for each billing period, assigning it to all represented days, and then grouping and summing these values by month. Calendarization was required to use a lagged dependent variable (LDV) model.

The resulting billing data set after cleaning was used in pooled and site-level analysis. The Apex team used an LDV model, where pre-period consumption in the corresponding calendar month is included as an independent variable to calculate energy-use changes. Unlike a fixed effects model, the LDV estimates post-period baseline use as a proportion of pre-period use, the mitigating bias introduced by autocorrelation within the time series data. The team then joined the energy-use data from the pre-period window (Jan–Dec 2017) to the post-period data (Mar 2019–Feb 2020) by month-of-year. Then, the following model specification was used:

$$ADC_{p,t} = \beta_0 + \beta_{pre} \cdot preADC_{p,t} + \beta_{mo} \cdot month_t + \beta_{premo} \cdot preADC_{p,t} \cdot month_t + \beta_{post} \cdot post_{p,t} + \beta_{opt} \cdot opt_{p,t} + \varepsilon$$

Where:

p = An index corresponding to a single unique site

t = An index corresponding to a calendar month in the post-period

$ADC_{p,t}$ = The post-period average daily consumption (ADC) for site p in calendar month t

$preADC_{p,t}$ = The pre-period average daily consumption for site p in calendar month t

$month_t$ = A dummy variable for calendar month t

$post_{p,t}$ = A binary variable indicating whether site p has the smart thermostat installed in calendar month t of the post-period

$opt_{p,t}$ = A binary variable indicating whether the smart thermostat at site p is receiving optimization in calendar month t of the post-period

β_0 = The aggregate intercept coefficient estimate

β_{pre} = Coefficient estimating the aggregate impact of pre-period average daily consumption on post-period average daily consumption

β_{mo} = The monthly intercept coefficient estimate

β_{premo} = Coefficient estimating the impact, by calendar month, of pre-period average daily consumption on post-period average daily consumption

β_{post} = Coefficient estimating the impact of smart thermostat installation on average daily consumption in the post-period

β_{opt} = Coefficient estimating the impact of smart thermostat optimization on average daily consumption in the post-optimization period

ε = The error term

The Apex team ran one model per system type and fuel source for a total of eight models and tested the inclusion of occupancy sensing and climate zone as terms in separate models. Each model included the additional term as a standalone and interacted with the *post* variable to isolate differences among the groups before installation and differences in energy change after installation.

Although the preferred model for billing savings analysis was as specified above, the Apex team also wanted to develop comparison group adjustments, or “baseline offsets,” to apply to the site-level models (see next section) to account for exogenous changes by non-participants. Therefore, the Apex team used a simpler weather-normalized fixed effects model for adjusting to the site-level model, excluding post-installation data for residents who had thermostats installed for the entire post-period. The Apex team added a *window2* term to represent the post-period window and isolate the exogenous energy use changes for the comparison thermostat users.

The model formula is shown below:

$$ADC_{p,t} = \alpha_{0,p} + \beta_{HDD}HDD_{p,t} + \beta_{CDD}CDD_{p,t} + \beta_{w2}window2_t + \beta_{post}post_{p,t} + \varepsilon$$

Where the new terms are:

$\alpha_{0,p}$	= Site-specific fixed effect for site *p*
$HDD_{p,t}$	= The heating degree days (HDD) experienced by site *p* in month *t*
$CDD_{p,t}$	= The cooling degree days (CDD) experienced by site *p* in month *t*
$window2_t$	= A binary variable indicating whether month *t* is in the post-period window of analysis
β_{HDD}	= Coefficient estimating the impact per HDD on ADC
β_{CDD}	= Coefficient estimating the impact per CDD on ADC
β_{w2}	= Coefficient estimating the difference in site-specific energy use, on average, during the second (post-period) window of analysis

Site-Level Analysis

The site-level billing analysis followed CalTRACK guidelines[18]. The Apex team conducted the following steps to analyze each participant in the master dataset.

The Apex team began with a data set with the same cleaned and calendarized billing data used in billing analysis (See Section 3.3). Then, the zip code for each participant was matched to the closest NOAA weather station to download hourly weather data for the analysis period (2017–2020) and typical meteorological year (TMY3) data. In most cases, the activation date for the smart thermostats was obtained from the manufacturers or inferred from the telemetry data when unavailable.

For consistency with the analysis period in the thermostat telemetry data, the baseline and reporting periods for the site-level billing analysis were set as follows:

- The start of the baseline period was set as January 1, 2017.

---

[18] http://docs.caltrack.org/en/latest/

- The end of the baseline period was set as the earlier of December 31, 2017, or the thermostat activation date.
- The start of the reporting (post-install) period was set as the later of the thermostat activation date or March 1, 2019.
- The end of the reporting period was set at 12 months after the start of the reporting period.

The CalTRACK methods, as implemented in the open-source library eemeter[19], were then applied to fit variable base degree day models to the billing data. Two models were fit—one in the baseline period and a second in the reporting period. The outputs of this process included the model parameters and model fit metrics, particularly R-squared (R^2) and CVRMSE[20].

The two models were used to estimate the NAC during the reporting period in the absence and presence of the smart thermostat by multiplying the model coefficients with the total degree days in those periods. The difference between the baseline and reporting NAC was the ΔNAC, or site-level savings, an estimate of savings during the reporting period due to the installation of the smart thermostat. Refer to the CalTRACK documentation[21] for a full description of this procedure. The Apex team applied the adjustment offsets from the pooled analysis (see Section 4.2) to all site-level savings shown in this report.

2.4 Correlation Analysis Approach

The final step in the analysis was the correlation analysis, which attempted to correlate site-level ΔNAC with thermostat metrics and other data. The primary study goal was to develop a method to estimate energy savings for smart thermostats based on thermostat metrics. The correlation analysis tested whether any metrics calculated using thermostat telemetry data correlate with energy savings.

To answer these questions, the Apex team tested six to eight different models for ΔNAC for each heating system type, each with at least one primary variable and between zero and five secondary variables. The primary variables included standard ENERGY STAR metrics and the new metrics to capture the major mechanisms for

---

[19] https://eemeter.readthedocs.io/

[20] Coefficient of Variation of the Root Mean Squared Error. What is a CVRMSE value? | Support Article (recurve.com)

[21] http://docs.caltrack.org/en/latest/

potential energy savings (e.g., runtime reduction from regional baseline, avoided excess resistance utilization), in both the heating and cooling seasons. The secondary variables included several different metrics to characterize resistance heat utilization, and home and HVAC system characterization metrics.

For each model, the Apex team conducted the same process to extract model coefficients estimate prediction uncertainty. The prediction uncertainty (bias and variance) addresses the central goal of the study by evaluating how far off a future prediction for savings could be based on thermostat metrics. The steps were as follows:

- Perform a linear regression between adjusted ΔNAC and a set of primary and secondary variables.
- Capture all model coefficients and standard errors, and model fit metric such as R^2 and CVRMSE.
- Perform ten-fold cross-validation to determine normalized mean bias error (NMBE, the average percent difference between actual and predicted ΔNAC for out-of-sample sites). For cross-validation, the Apex team fit a model on a 70% random sample of sites to calculate the model coefficients and predicted savings for the other 30% of out-of-sample sites, then repeated ten times. The mean and standard deviation of the NMBE tell us the expected bias in our predictions.

The Apex team expanded the analysis to include residual analysis and correlation residuals with other independent variables for selected models.

3. Sample Characteristics and Survey Results

This section outlines the sample characteristics and survey results from thermostat user opt-ins.

3.1 Thermostat User Opt-Ins

The Apex team, in partnership with manufacturers and utilities, contacted over 50,000 Northwest residential customers to participate in the study, as shown in Table 1. This led to over 3,900 thermostat users in the study via an opt-in process or through manufacturers providing data directly under prior customer agreements.

Table 1. Thermostat Users in the Study

	Total	Ecobee	Emerson	Nest	Resideo
Contacted/Emailed	**50,072**	NA*	6,099	28,895	15,078
Thermostat Users in Study	**3,943**	1,641	587	1,177	538**

* Ecobee's eco+ participants did not require opt-in due to existing participation agreements that allowed data sharing.

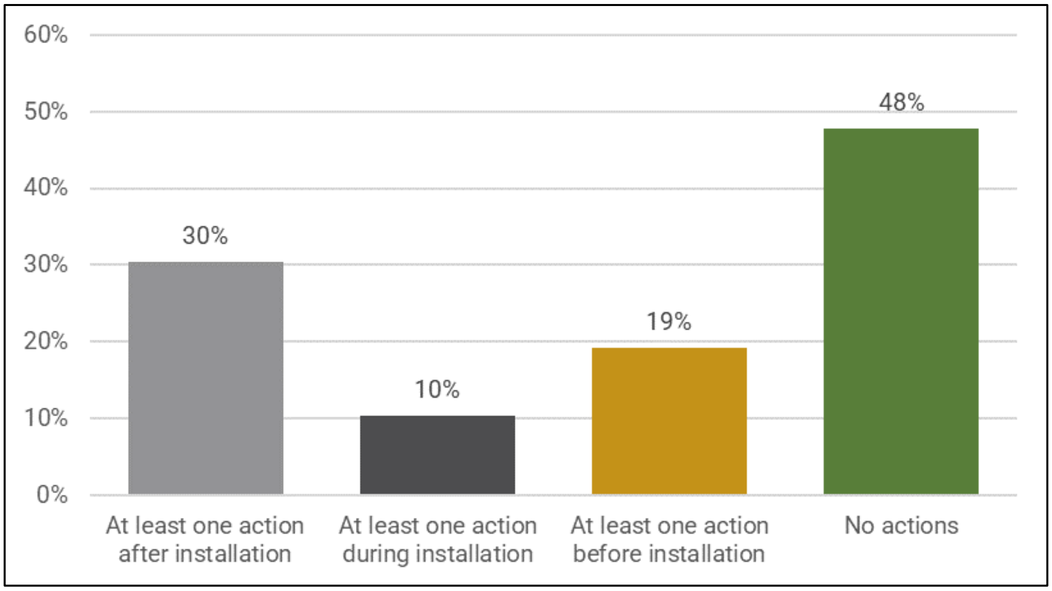
** Although approximately 15,000 Resideo customers were recruited to opt in to the study, only 38 thermostat users agreed to opt in. To support the study and sample, Energy Trust of Oregon provided data for an additional 500 participants and comparison group customers for the Resideo ConnectedSavings pilot.

Although the Apex team attempted to develop a representative sample of thermostat users in the Northwest, this study's sample is likely not representative due to the complexity of data collection that limited the sample size and characteristics. The primary intent of this study was to establish a correlation between thermostat metrics and energy savings (Section 4.3). This analysis does not require that the sample represent the population under the assumption that the metrics and savings are correlated and benefited from the inclusion of multiple thermostat models with a breadth of thermostat characteristics. The correlation holds even for customers receiving optimization because optimization should impact both thermostat metrics and energy savings. Outside of this correlation (see Section 4.2), the gathered data was a "convenience sample" (i.e., not representative), and results should be treated as suggestive rather than definitive.

3.2 Survey Results

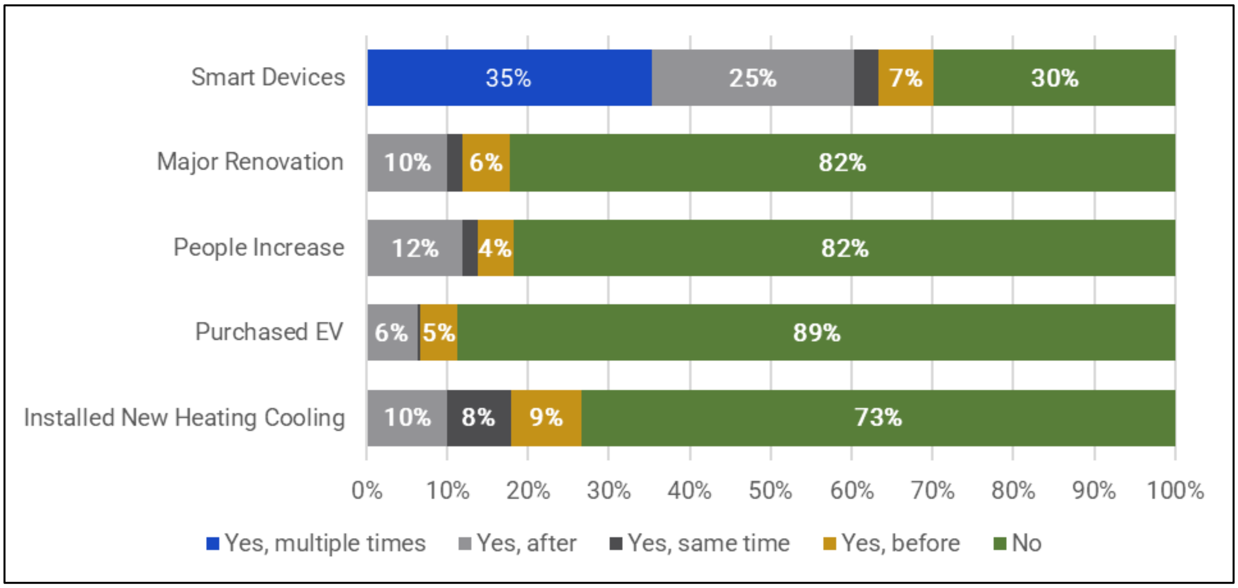
Of the nearly 4,000 thermostat users in the study, approximately 1,850 completed a survey of major energy use change behaviors as part of the opt-in process. Surveys of opt-in customers found that they tend to install their smart thermostats prior to other energy-use affecting actions. Figure 2 shows that a higher proportion of the residents who made major changes took at least one action after or during smart thermostat installation than before. Figure 3 breaks the survey results down by the various activities. It shows that the asymmetry in time (i.e., more actions after/during than before) holds true across most of the activities.

Figure 2. Overall Timing of Major Changes Relative to Smart Thermostat Installation (Electric Vehicles, Renovations, HVAC System, Occupancy)*



* Sums to more than 100% because of multiple response survey questions.

Figure 3. Individual Energy Use Changes and Timing Relative to Smart Thermostat Installation



See Section 4.2 for an analysis of the site-level impacts on energy use from each of these changes.

3.3 Merged Sample of Thermostat and Billing Data

As described in Section 2.1.3, once the thermostat users were in the study, the Apex team collected data from manufacturers and utilities. As shown in

Table 2, thermostat manufacturers were able to provide thermostat data for over 85% of the thermostat users (3,367 of 3,943) in the study using serial numbers, email addresses, or direct recruiting links. Reasons for data loss at this stage included lack of availability of thermostat data during the analysis period or inability to match the provided identifiers within the manufacturers' customer database.

Next, utilities matched approximately 43% of the thermostat users (1,452 of 3,367) with billing data. Most of the attrition at this stage was in the ecobee dataset due to the inability to match customer emails to the emails on file with the participating utilities.

Finally, to be included in the merged sample, the study required at least one month of baseline (pre-installation) and reporting (post-installation) data, which was approximately 80% of those thermostat users with billing data (1,166 of 1,452). This step also included several minor data cleaning filters. They were:

- Drop records with a net metered billing code.
- Drop records with missing or negative data, or 0 kWh reads.
- Drop records with excessively high energy use (greater than 20 kW or 120 kBTU/hr continuous operation).
- After calendarization, drop records with less than 15 days of data in the month.
- Drop records more than 3 standard deviations above or below the average use for the month.
- (For Energy Trust of Oregon only) Drop records where bill duration was less than 15 days or greater than 45 days, and usage differed by greater than 30% from the prior month, assuming either a missed bill or the median bill duration (30 days).[22]

---

[22] Energy Trust of Oregon data was provided with bill dates but not bill durations. Therefore, bill durations had to be calculated from prior bill dates. When there was a gap of greater than one standard bill duration, this logic had to be applied.

Table 2. Attrition of Thermostats from Data Received to Merged Sample

Attrition Step	Total Thermostat Users	Ecobee	Emerson	Nest	Resideo
In the Study	**3,943**	1,641	587	1,177	538
With Thermostat Data	**3,367**	1,641	95	1,106	525
Match to Billing Data	**1,452**	247	61	747	397
Merged Sample	**1,166**	194	58	576	338

Table 3 shows the characteristics of the merged sample by utility and thermostat manufacturer. The majority of thermostats in the study are in the Energy Trust of Oregon territory, followed by Puget Sound Energy.

Table 3. Merged Sample Sites by Utility and Manufacturer

	Avista	Clark	Energy Trust of Oregon	PSE
Ecobee	0	0	123	66
Emerson	16	6	39	0
Nest	80	54	360	113
Resideo	2	1	315	1

Note: The sum of sites is higher than Table 2 as some participants had accounts from multiple utilities.

After the merged sample data cleaning, the Apex team applied additional filters on data quality to arrive at the site-level analysis dataset and the pooled analysis dataset, described in Sections 3.4 and 3.5 below. These filters were similar but slightly different for the two analyses. The pooled analysis aimed to assess the aggregate energy savings, which required different filters than for the goal of correlation after calculating site-level savings. In brief, the two main differences were as follows:

- The pooled analysis included a comparison group, so post-period data sufficiency checks refer to the second window of analysis, not the period of time after installation. Sites with no thermostat installation were important to retain because their data in the second window, without thermostat installation, constituted the counterfactual for thermostat installers.

- Thermostat model fits, and other data quality filters are important to the site-level analysis but irrelevant to the pooled analysis.

3.4 Site-Level Sample Data

The data quality filters on the site-level sample focused on obtaining the best data for correlation analysis. That effort required filters on both thermostat data quality and billing data quality. The data for each site had to meet the following criteria:

- At least 9 months of data in the baseline and reporting periods.
- Absolute values of site-level savings less than 50% of baseline usage.
- At least 30 core heating days in the calculation of telemetry metrics.
- Not in the top and bottom 0.5% of customers by annual energy use.
- Survey filters: Resident had not purchased an electric vehicle (for electricity meters only), updated their HVAC system, performed a major renovation, or had increased occupancy in their home, during or after the thermostat installation. Thermostats without associated survey responses were left in the sample.

Table 4**Error! Reference source not found.** shows the attrition across each of these steps. In the site-level analysis dataset, about 10% of the sites have a thermostat connected to a heat pump, while 90% have a thermostat connected to a furnace or boiler. Table 5 shows the final sample by system type and fuel.

Table 4. Attrition from Site-Level Analysis Data Quality Filters, by Manufacturer

Attrition Step	Total	Ecobee	Emerson	Nest	Resideo
Merged sample	**1,166**	194	58	576	338
At least 9 months of pre- and post-billing data	**805**	128	58	404	215
Absolute value of ΔNAC less than 50%	**776**	127	58	392	199
More than 30 core heating days	**765**	125	58	386	196
Remove top and bottom 0.5% by energy use	**762**	125	58	383	196
Survey filters	**587**	125	37	236	189

In the site-level analysis dataset, about 10% of the sites have a thermostat connected to a heat pump, while 90% have a thermostat connected to a furnace or boiler, as shown in Table 5.

Table 5. Site-Level Analysis Sample by System Type and Analyzed Fuel Consumption Data

Heating Type	Fuel	n
Gas Furnace	Gas	497
	Electricity	381
Heat Pump with Electric Backup	Gas	13
	Electricity	43
Electric Furnace	Gas	2
	Electricity	15
Heat Pump without Electric Backup	Gas	1
	Electricity	2

In terms of heating zones (HZ), 13% of the sites are located in HZ 2/3 (cold/very cold), while 87% are located in HZ 1 (marine). The Apex team had hoped to acquire a larger sample of heat pumps and cold-climate sites. The relative size of the installed base in Puget Sound and Energy Trust of Oregon territory dictated the concentration of HZ results.

3.5 Pooled Analysis Data Set

The pooled analysis data set used slightly different data quality filters than the site-level analysis data set. Table 6 shows the attrition due to these data quality filters, by system type. Table 7 shows the final counts of sites by manufacturer and system type.

Table 6. Attrition from Pooled Analysis Data Quality Filters, by Manufacturer

Attrition Step*	Nest	Ecobee	Resideo	Emerson
Merged sample	732	245	391	60
Activation date before the pre-period	732	221	391	60
Survey filters	413	221	380	37
At least 10 months billing data in each window	306	167	300	35

* Note that this sample includes comparison group customers who had no post-period billing data after thermostat installation, and therefore starts with a larger initial n.

Table 7. Final Numbers by Manufacturer Type and Analyzed Fuel Consumption Data

Heating Type	Fuel	Nest	Ecobee	Resideo	Emerson
Gas Furnace or Boiler	Electricity	207	114	202	27
	Gas	267	139	242	30
Heat Pump with Electric Backup	Electricity	21	14	36	2
	Gas	6	10	6	1
Electric Furnace or Boiler	Electricity	10	5	10	0
Heat Pump with Non-Electric Backup	Electricity	2	1	2	0

4. Energy Savings Analysis Results

4.1 Thermostat Telemetry Analysis

Modification of ENERGY STAR Software and Anonymous Data Analysis

Using the method described in Section 2.2, the Apex team analyzed hourly, anonymous thermostat telemetry data from 512 thermostats, including 275 with heat pumps, to test the planned additional metrics for inclusion in the ENERGY STAR Connected Thermostat Field Savings software. Secondarily, the Apex team performed an exploratory analysis of this data to determine whether other additions could potentially capture mechanisms for savings effectively.

To support the primary goal of correlating metrics with site-level energy changes, the Apex team aimed to create the most descriptive metrics possible. The more effective a given metric is at characterizing thermostat operation characteristics (e.g., setpoints, variation in indoor temperature, strip heat use), the better it will represent the thermostat side of the correlation. Therefore, the Apex team assessed hourly telemetry data and resulting summary metrics to identify whether they described real states of operation and varied among sites. The remainder of this section provides an overview of key metrics and their relationship with hourly telemetry data.

The ENERGY STAR method fits a linear model between average daily equipment runtime and thermal demand (indoor-outdoor temperature difference) to produce a modeled runtime. Thermostat metrics are then calculated from this runtime model.

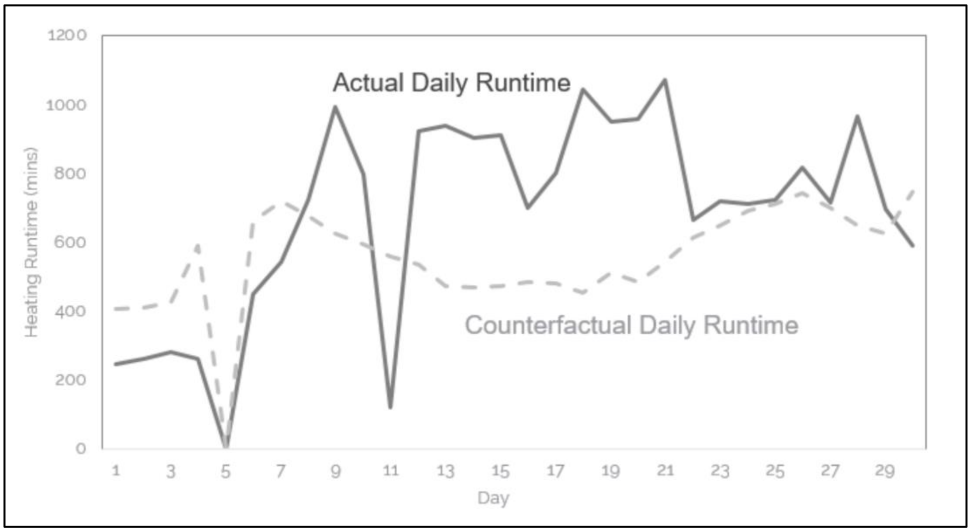
For each site and each HVAC season (heating and cooling), the ENERGY STAR software fits two basic parameters for the runtime metric. The first is a temperature offset representing the indoor-outdoor Delta-T, at which active space conditioning comes online (heating- or cooling-degree hours start being counted when Delta-T exceeds this value). The second is a loading constant that relates heating- or cooling-degree-hours to HVAC runtime. Runtime is then modeled as the product of degree-hours and the loading constant. The primary runtime reduction metrics are generated by calculating the percent difference between actual runtime and a modeled counterfactual based on a stipulated indoor temperature profile. In this percent-difference formulation, the same loading constant appears in the numerator and the denominator, so by cancellation, what we call "runtime reduction" is equivalent to the percent reduction in estimated heating- or cooling-degree-hours. Because of this, it can be useful (and more direct) to think of the runtime reduction metric as a set-back metric.

In the base ENERGY STAR formulation, this counterfactual is a constant "comfort temperature" equal to the 90th percentile most comfortable (cooler in the cooling season, warmer in the heating season). In this study, the Apex team uses regional baseline indoor temperatures calculated from the RBSA study and segmented by heating and cooling climate zone. The associated metrics are:

- Runtime Reduction with Comfort Temperature Baseline
- Runtime Reduction with Regional Baseline

Figure 4Figure 4. shows an example of the actual and counterfactual daily runtime for a single thermostat generated via this method.

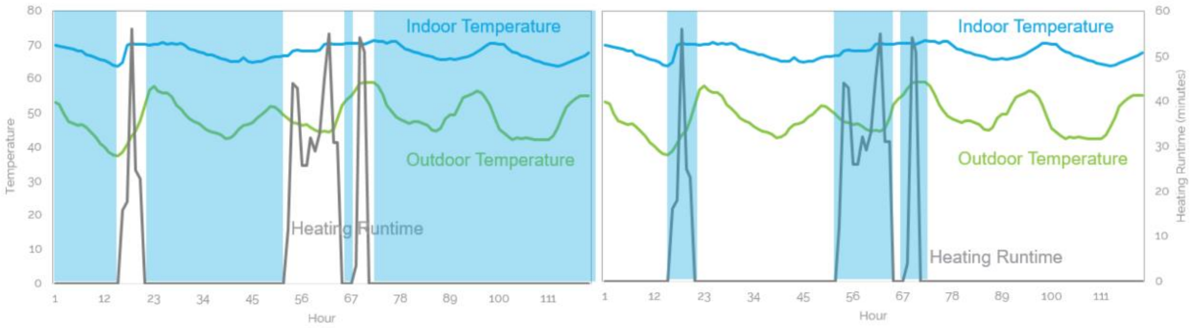
Figure 4. Daily Runtime Based on Comfort Temperature



The thermostat telemetry data also contains useful information about the rate of heat loss and gain within the residence. The rate of indoor temperature change towards the outdoor temperature when no HVAC is running may indicate the quality of the building shell. The rate of heating or cooling when HVAC is running may indicate undersizing or oversizing of HVAC systems. As secondary metrics in the correlation analysis, the Apex team believed these terms might be useful and added them to the software. The terms are defined as follows, and examples of calculation periods are shown in Figure 5. The associated metrics are:

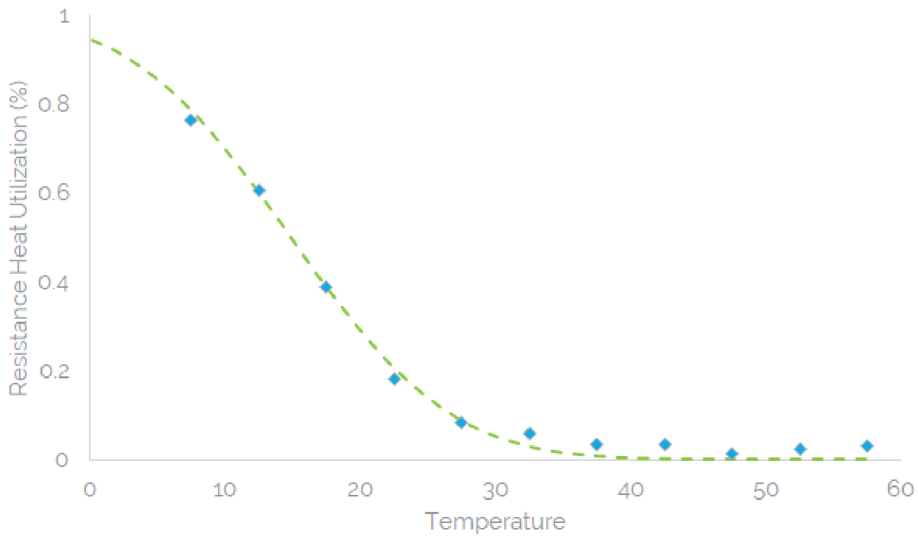
- **Heat loss constant (heating) / heat gain constant (cooling):** Average temperature change rate relative to indoor-outdoor temperature difference, when HVAC runs less than 5 minutes per hour (this 5-minute threshold was selected to account for hours with very low HVAC usage and the gradient metric is relatively insensitive to it).
- **HVAC constant:** Average temperature change rate relative to indoor-outdoor temperature difference, when HVAC runs over 15 minutes per hour.

Figure 5. Indoor and Outdoor Temperature versus Heating Runtime, with the Periods of Time Used for Calculating the Heat Loss Constant and HVAC Constant Highlighted



The Apex team added calculations for several additional metrics related to heat pump resistance utilization, including DNRU, Excess Resistance Score, and a sigmoid integral. The base of these metrics comes from the ENERGY STAR-defined resistance utilization metric. All three additional heat pump metrics are described in **Appendix 1: Telemetry Analysis Memo Addendum**. For example, the sigmoid calculation is a compact representation of resistance heat utilization (RHU) across all temperature bins. A sigmoid function is fit to RHU and integrated between 0°F and 60°F. The result is a single value that corresponds to how much resistance heat the thermostat has called for within a range of temperatures. A fitted sigmoid to RHU is shown in Figure 6.

Figure 6. Sigmoid Fitted to Binned Resistance Heat Utilization



Resulting Thermostat Metrics Analysis

After the thermostat metrics were generated by manufacturers and transferred to the Apex team, the team reviewed each metric and assessed the correlation among metrics to gain a better understanding of which metrics might provide duplicative or new information. Figure 7 and Figure 8 show a selection of the output metrics compared against each other. The distribution of each metric is found in the diagonals for gas and electric meters, respectively. A few key findings on individual metrics include:

- The regional baseline runtime metric is centered around zero, indicating that the indoor temperatures in the analysis sample are close to the regional averages used to develop the baseline.
- The comfort baseline runtime metric is centered around 10%. By its nature, it is assuming an inefficient baseline temperature, so it is always positive and therefore may not necessarily reflect actual savings or runtime reductions.
- Heat loss and HVAC heating constants have skewed distributions: homes in the tails of these distributions likely have weatherization and HVAC optimization opportunities, respectively.

In terms of correlations among metrics (non-diagonals), notable findings include:

- The runtime reduction using an individual comfort temperature baseline showed a weak correlation with runtime reduction using a Northwest regional baseline, suggesting that they measure different effects. However, ENERGY STAR also includes its regional baseline for runtime reduction, which correlates strongly with the Northwest baseline.
- There appears to be a weak positive correlation between the heat loss constant and the runtime metrics and a weak negative correlation between the HVAC constants and the runtime metrics, indicating that some of the information about building and HVAC performance may be captured through these metrics.

Figure 7. Gas Furnace Metrics (n=497)

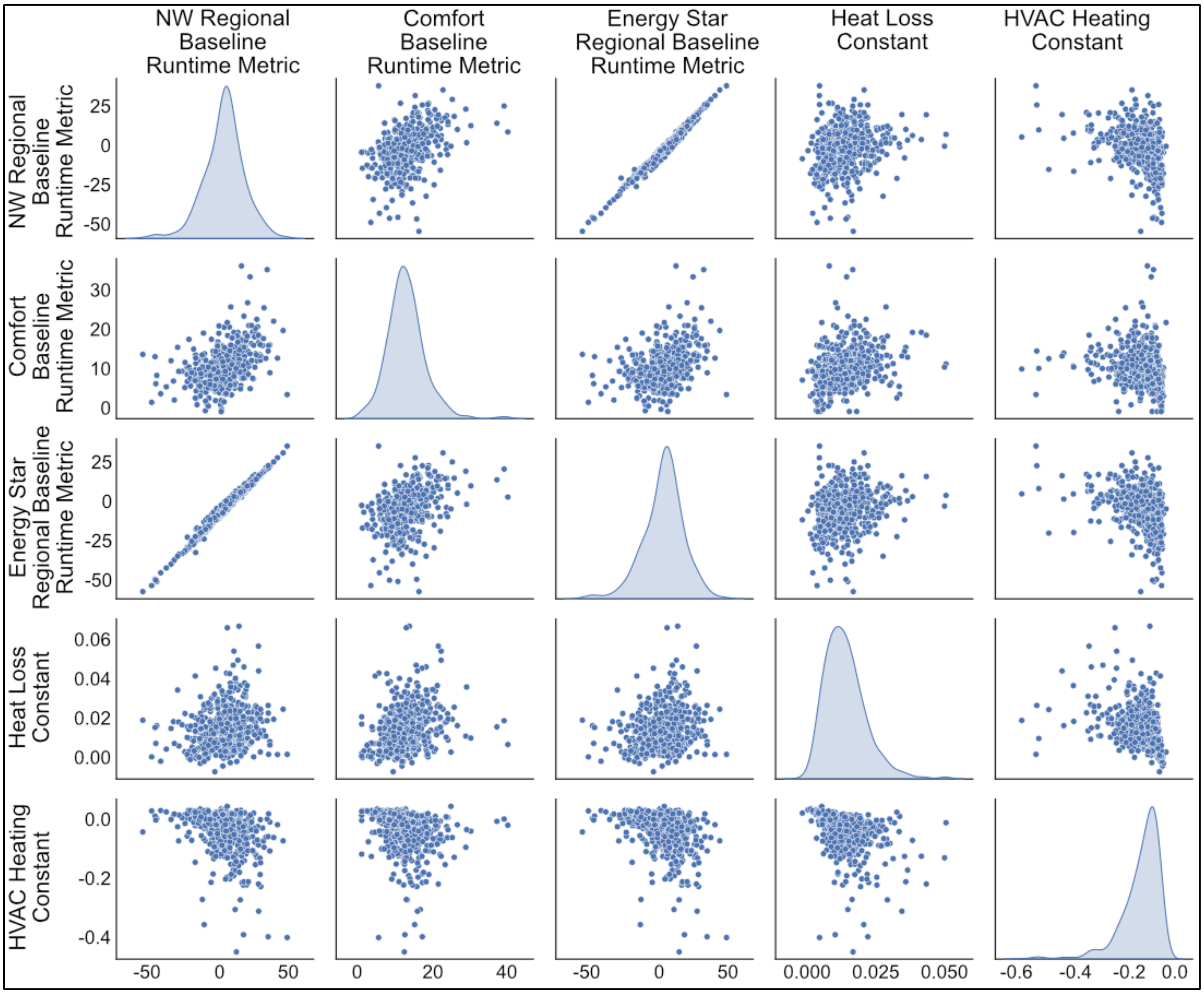


Figure 8. Heat Pump (n=43) and Electric Furnace Metrics (n=15)

4.2 Billing Analysis

Pooled Analysis

After cleaning the consumption data, the Apex team conducted the pooled analysis as described in Section 2.3. The Apex team modeled all eight possible combinations of fuel type and system type, but only three of the combinations had a sufficient

(>50) number of sites: gas data for sites with gas furnaces, electric data for sites with gas furnaces, and electric data for sites with heat pumps and electric backup.

The modeled whole-home energy savings due to smart thermostat installation are shown in Table 8. As noted previously, the sample in this analysis is not necessarily representative of the region or typical thermostat users, so results of the pooled analysis should be considered suggestive (i.e., containing unknown levels of bias) and not definitive estimates of energy savings. The Apex team found statistically significant gas savings (43 therms, 5%) for sites with gas furnaces and statistically significant electric savings (670 kWh, 4.5%) for sites with heat pumps and electric backup.

The team found a moderate increase in electric use (2.4%, 220 kWh) for homes with gas furnaces, potentially representing an increase in either furnace fan usage or cooling usage. A model interacting the *post* variable with month suggests that increased use for this group was concentrated in summer months, suggesting an increase in cooling usage, but given the relatively low precision on the yearly estimate, the monthly estimates should be interpreted with caution.

For all three groups, the team isolated the effect of optimization during portions of the post-period. Its impact enhanced savings by 2% to 6% in addition to savings from smart thermostat installation. Only 27% of thermostats (185 gas furnace, 15 heat pump) received optimization during the analysis period. Table 9 provides pre- and post-consumption data; other detailed modeling results, including the gas furnace electric use model with monthly estimates for impacts, can be found in

Appendix 2: Billing Analysis Details.

Table 8. Pooled Analysis Results

Heating System Type	Fuel	n	Comp. Group n	Post-Installation			
				Average Savings	Std Error of Avg Savings	Percent Savings	Optimization Monthly Avg. Effect*
Gas Furnace or Boiler	Electricity	550	104	-220 kWh	110 kWh	-2.4%	2.8%*
	Gas	678	133	43 therms	20 therms	5%	6.3%*
Heat Pump with Electric Backup	Electricity	73	15	670 kWh	402 kWh	4.5%	1.9%*
	Gas	23	-	-34 therms	54 therms	-5.1%	-28.8%
Electric Furnace or Boiler	Electricity	25	-	760 kWh	789 kWh	5%	-2.3%

Heating System Type	Fuel	n	Comp. Group n	Post-Installation			
				Average Savings	Std Error of Avg Savings	Percent Savings	Optimization Monthly Avg. Effect*
Heat Pump with Non-Electric Backup	Electricity	5	-	1477 kWh	1257 kWh	10%	NA

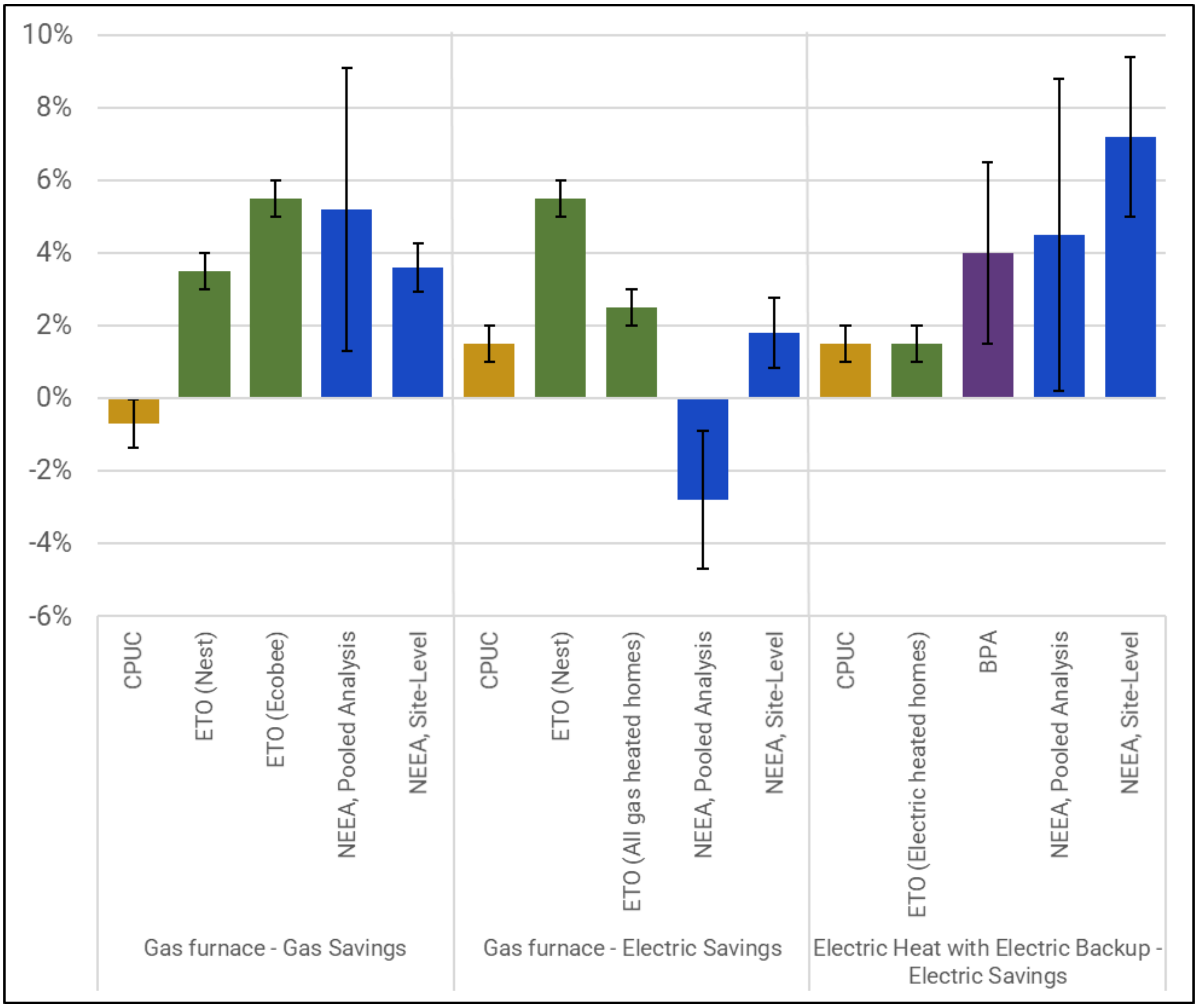
* Optimization savings do not reflect a yearly estimate of savings. They are an estimate of the monthly average effect during optimization for affected users, with users having differing numbers of months of optimization.

Table 9. Pre- and Post-Consumption Usage

Heating System Type	Fuel	Pre-Period Average Daily Consumption	Post-Period Average Daily Consumption
Gas Furnace or Boiler	Electricity (kWh)	25.09	23.73
	Gas (therms)	2.36	2.2
Heat Pump with Electric Backup	Electricity (kWh)	40.62	36.66
	Gas (therms)	1.82	1.48
Electric Furnace or Boiler	Electricity (kWh)	41.59	35.88
Heat Pump with Non-Electric Backup	Electricity (kWh)	40.29	35.84

The detected gas savings align with the Energy Trust of Oregon studies by Apex Analytics in 2015 and by Recurve in 2015 and 2017, as described in Section 1.3. Although this study's primary goal was to correlate site-level energy savings with thermostat metrics, the fact that aggregate energy savings are similar to other studies (see Figure 9) provides a measure of confidence that the sample sourced from the opt-in process is not biased due to the data collection strategy approach. The findings for heat pumps from the pooled analysis show a much higher point estimate than other studies, but error bars are nearly as wide as the point estimate itself.

Figure 9. Comparison of Pooled Analysis Results with Other Studies



In addition to comparing with other studies, the Apex team tested several variations of this base model, which are reported below. Detailed outputs can be found in

Appendix 2: Billing Analysis Details.

Occupancy Sensing: The Apex team tested the impact of including an occupancy sensing term and interacting it with the post period term. This term was a dummy variable that was true for thermostats with a built-in onboard or functioning external occupancy sensor, not for thermostats that rely on geofencing. For the gas model, sites with occupancy sensing were predicted to use more energy prior to installation (178 more therms per year) and save more energy after installing an occupancy sensing thermostat (112 therms per year). However, the estimated savings for thermostats without this feature dropped to zero. The opposite was true for the electric groups (predicted energy use and savings also went down for occupancy sensing thermostats). Interpretation of these results is difficult—prior studies have found that direct occupancy sensing thermostats save more energy, but introducing these additional terms to the model may consume too many degrees of freedom, resulting in overfitting. Given the limited sample sizes, we do report this as a main study finding. Table 10 shows the number of sites with occupancy sensing versus the number of sites that received optimization at some point.

Table 10. Counts of Thermostats with Occupancy Sensing Features, and Who Received Optimization (Gas Model for Gas Furnaces)

	No Optimization	Optimization
No Occupancy	196	57
Occupancy	297	128

Climate Zone: None of the tested models showed significant impacts or changes to post-installation estimates by climate zone. However, only 64 customers were outside of HZ1.

Time Window: As described in Section 2.3, the modeling windows were selected to align with the site-level analysis, which maximized available sites with post-period data by setting its post-period as late as possible. The Apex team also tested setting the post-period window to December 2018–November 2019 to create a larger available comparison group. Doing so narrowed the confidence intervals slightly on the *post* term but increased the confidence intervals on the *opt* term because the number of time points with optimization dropped. The savings values in the new window did not change directionally, although gas savings increased from 43 to 65 therms, and electric heat pump savings decreased from 670 kWh to 563 kWh. Therefore, the Apex team's results are based on the post-period window that is consistent with the site-level analysis (March 2019–February 2020).

Baseline Offsets for Site-Level Analysis

To establish adjustments for the site-level models, the Apex team implemented a simple fixed-effects model with HDD and CDD terms to calculate baseline offsets for use in site-level billing analysis. These baseline offsets do not impact the correlation analysis because they shift the estimated savings for all sites. However, they adjust for bias due to changes in the counterfactual energy usage across the pre- and post-periods, providing a more realistic number for average savings. The team reports the baseline offsets applied to savings for each fuel and system type in Table 11.

Table 11. Baseline Offsets from Pooled Analysis for Site-Level Analysis

System Type	Fuel Type	N for Calculation	Adjustment to Savings
Gas furnace or boiler	Gas	312	+12 therms
Gas furnace or boiler	Electric	219	-379 kWh
Heat pump with electric backup	Electric	52	N/A*

*Sample size too small to calculate baseline offset

Site-Level Billing Results

The Apex team calculated ΔNAC (i.e., site-level savings) for all sites in the site-level analysis dataset, and then applied the two baseline offsets to gas and electric savings for the gas furnace or boiler group to arrive at adjusted site-level savings estimates. Because these calculations are at the site level, the analysis returns a distribution site-level savings values—one for each site.

The distributions of site-level savings estimates in the three main groups are shown in Figure 10, Figure 11, and Figure 12. Notable findings include:

- Gas site-level savings for gas-heated homes have a wide distribution ranging from −200 to +250 therms of savings.
- Most electricity ΔNAC for gas-heated homes fall in the −1000 to +1000 kWh range, while electricity ΔNAC from heat pumps comes from a much smaller sample and has a relatively wider distribution.

Figure 10. Gas Savings (Site Level) for Gas Furnaces and Boilers - Savings by Percent (left) and Therms (right)

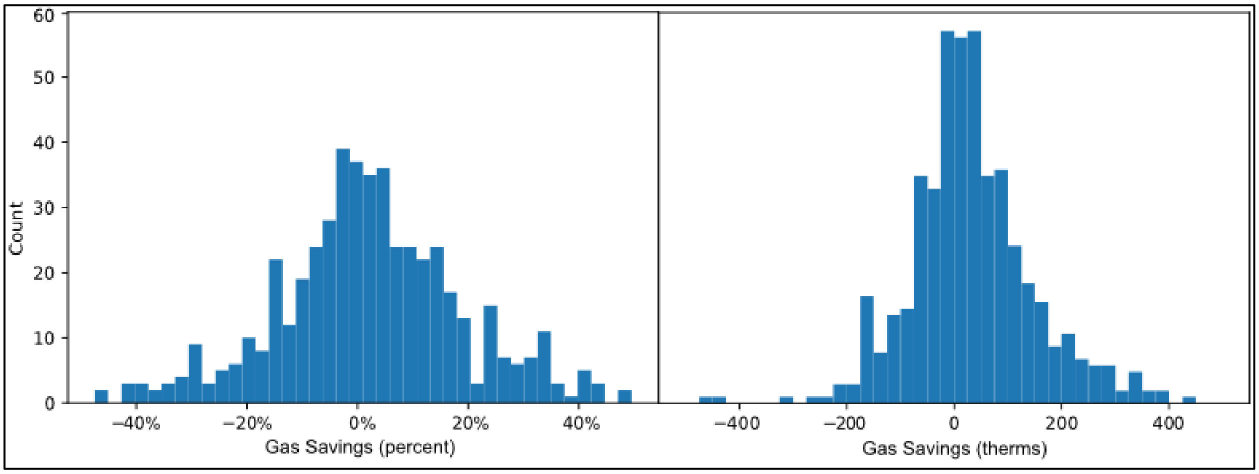


Figure 11. Electric Savings (Site-Level) for Gas Furnaces and Boilers, Savings by Percent (left) and kWh (right)

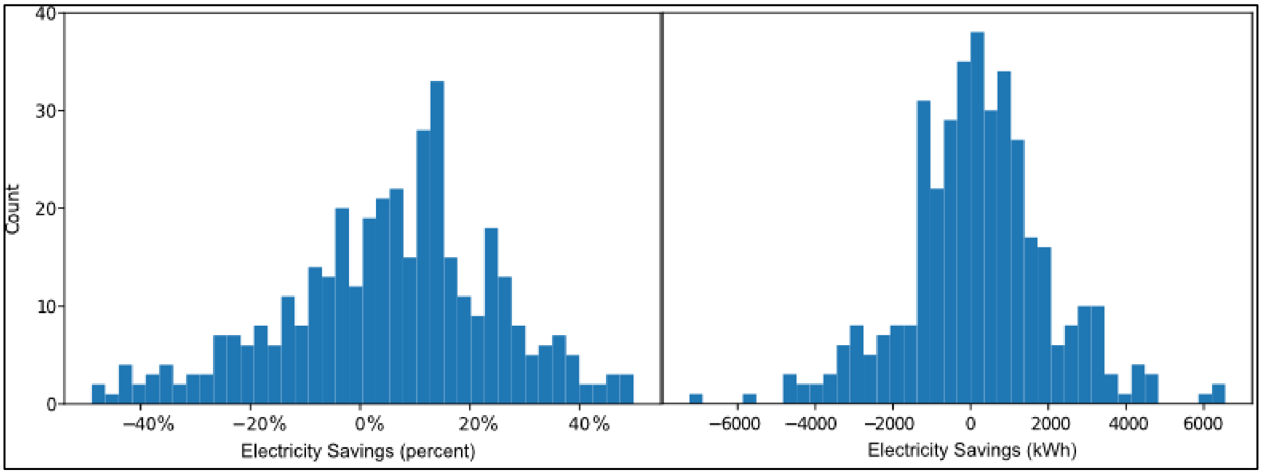
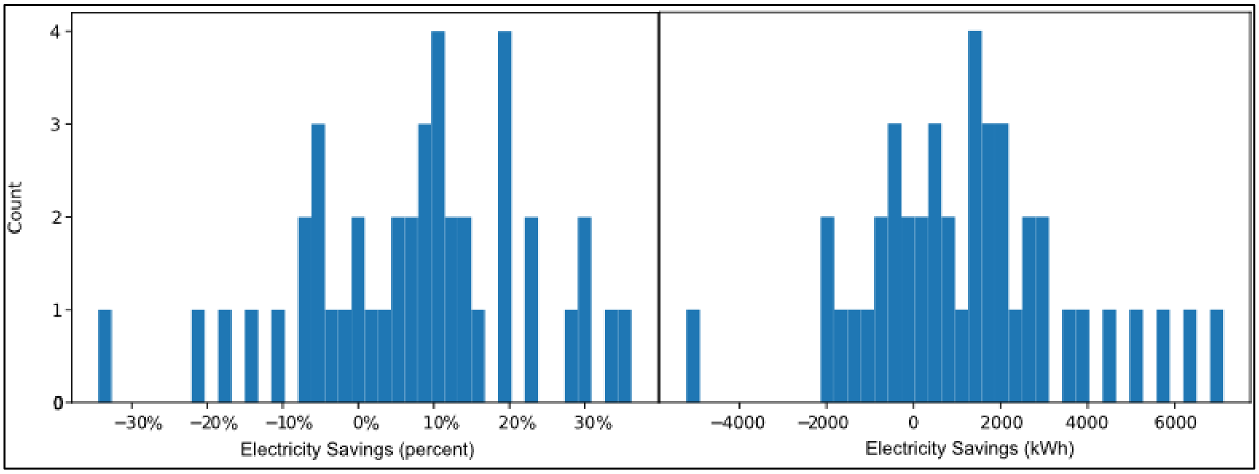


Figure 12. Electric Savings (Site-Level) for Heat Pumps with Electric Backup by Percent (left) and kWh (right)



Although there is a relatively wide distribution evident in the histograms, this distribution comes from site-level exogenous changes in energy use. The baseline offsets derived from the pooled analysis were modest and consistently applied and therefore did not influence the distribution or the correlations.

Table 12 shows the average adjusted ΔNAC estimates from site-level analysis. Note that these differ from the pooled analysis for two reasons:

- They include optimization (which creates real changes to thermostat runtimes and therefore should be kept for correlation analysis).
- They have additional filters applied that exclude sites based on thermostat data criteria.

Table 12. Site-Level Savings Results

Heating System Type	Fuel	n	Average ΔNAC	Std Error of Average Savings
Gas Furnace or Boiler	Electricity	381	187	96
	Gas	497	31	6
Heat Pump with Electric Backup	Electricity	43	1291	364
	Gas	13	37	36
Electric Furnace or Boiler	Electricity	15	721	458
	Gas	2	-7	22
Heat Pump with Non-Electric Backup	Electricity	2	-1340	1353
	Gas	1	158	-

The major energy-use changes reported in the opt-in surveys had a detectable impact on energy use in the post-period, as measured at the site level. Figure 13 shows the difference in site-level estimated ΔNAC for each group with major energy use changes in the post-period, compared to the sites in the main data set with no reported changes in the post-period.

The total sample in each of these groups is not sufficiently large to make statistically significant conclusions. However, the Apex team found that thermostat users with a new HVAC system had approximately 17 therms more in gas savings and 100 kWh less in electric savings than the main analysis, while the addition of occupants in the home erased or reversed all detectable gas and electric savings (Table 13). Both electric vehicle purchases, and major renovations increased electricity use substantially (700 to 1,100 kWh), although both groups saved additional therms compared to the main dataset. Although this is a small sample of the main data set, the magnitude and direction of these differences were substantial enough to merit removing them from the analysis. Sites that underwent any of these changes during or after installation were removed from the energy use and correlation analyses.

Figure 13. Major Energy-Use Changes and Site-Level Savings Differences

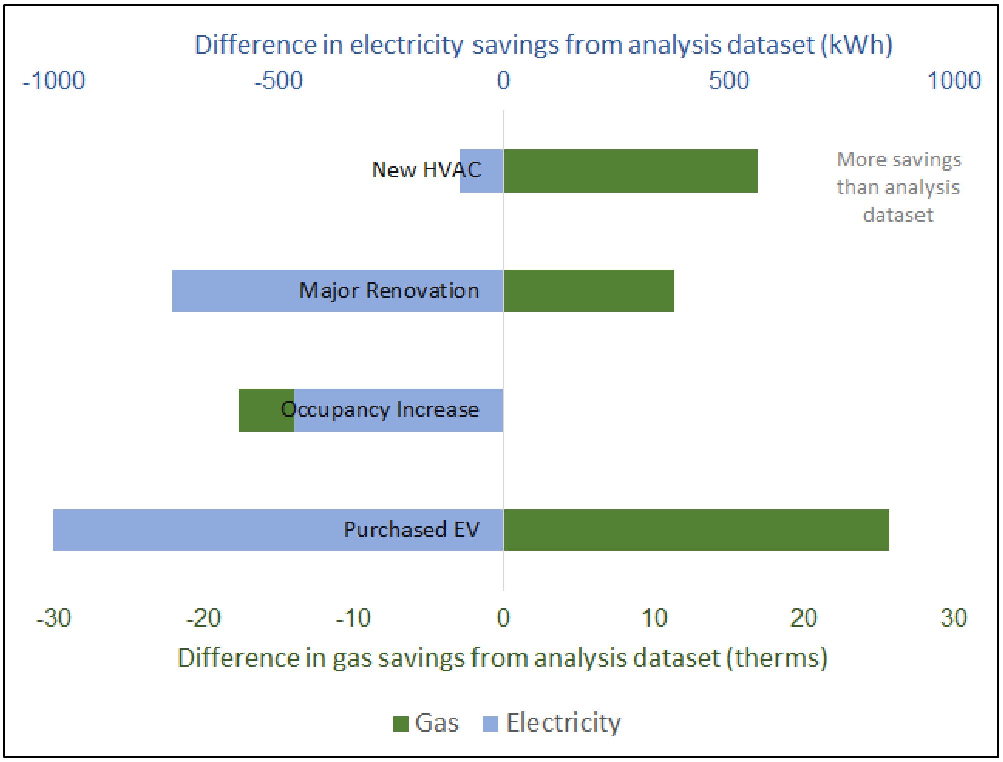


Table 13. Major Energy-Use Changes and Site-Level Savings Differences

	n Gas (% of total)	n Electricity (% of total)	Difference in Savings from Analysis Dataset			
			Therms	kWh	Therms %	kWh %
Electric Vehicle Purchase	38 (4%)	23 (3%)	26	-1074	3%	-11%
Occupancy Increase	59 (6%)	45 (5%)	-18	-466	-2%	-5%
Major Renovation	53 (5%)	38 (5%)	11	-735	1%	-8%
New HVAC	73 (7%)	55 (7%)	17	-97	2%	-1%

4.3 Correlation of Metrics with Savings

The Apex team tested 28 linear correlation models as described in Section 2.4. For each system and fuel type combination, the Apex team reports three model types in Table 14.

- Model A: Model that closely aligns with the expected ENERGY STAR software
- Model B: Minimal model with only the relevant primary metric(s)
- Model C: "Best" model(s) that includes additional secondary metrics to improve the fit

For additional tested models and additional details on the models presented below, see **Appendix 3: Correlation Analysis Details.**

In addition to the adjusted R^2 for each model, the table also includes the 5th and 95th percentiles of NMBE for one hundred out-of-sample cross-validation runs. Seventy percent of the sample (in-sample group) is used to fit a model and predict the mean for the other thirty percent (out-of-sample group). The NMBE is the average difference between the predicted and actual savings divided by the average savings for the out-of-sample group. These would indicate the bounds of bias error if the model were used to predict savings for a similar number of out-of-sample thermostats (such as new models providing thermostat metrics only)[23]. The adjusted R^2 describes the portion of variability in savings that the model captures.

Table 14. Table of Correlation Results

System Type	Fuel	n	Model Type	Variables in Model	Bias 5th %	Bias 95th %	R^2 adj.
Gas Furnace or Boiler	Electricity	322	A	savings ~ regional_runtime_metric_heating + regional_runtime_metric_cooling	-517%	983%	0.00
Gas Furnace or Boiler	Electricity	322	B	savings ~ comfort_runtime_metric_heating + comfort_runtime_metric_cooling	-283%	699%	0.00
Gas Furnace or Boiler	Electricity	314	C	savings ~ regional_runtime_metric_heating + regional_runtime_metric_cooling + heat_loss_constant_heating + heat_gain_constant_cooling + hvac_constant_heating + hvac_constant_cooling + weekly_temperature_variance_heating	-333%	366%	0.00
Gas Furnace or Boiler	Gas	497	A	savings ~ regional_runtime_metric_heating	-46%	117%	0.03
Gas Furnace or Boiler	Gas	497	B	savings ~ comfort_runtime_metric_heating	-39%	105%	0.00
Gas Furnace or Boiler	Gas	497	C	savings ~ comfort_runtime_metric_heating + heat_loss_constant_heating + hvac_constant_heating	-38%	99%	0.01

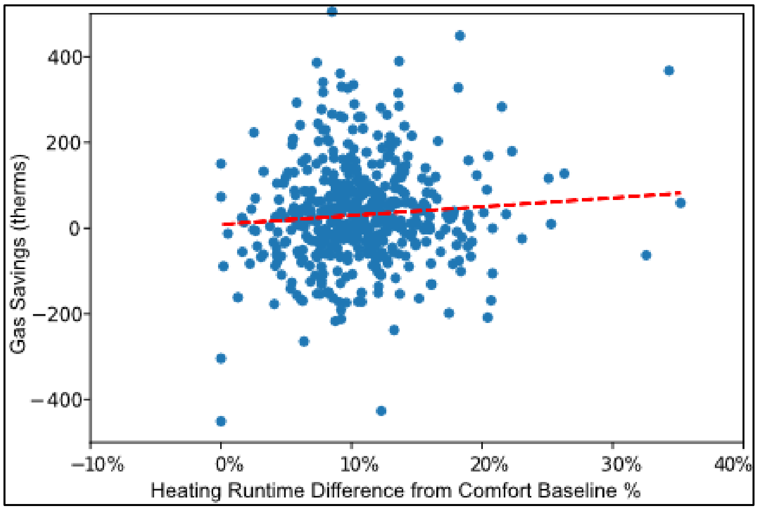
---

[23] Acceptable NMBE bounds can be understood as similar to a 95% confidence interval width, where ±10% indicates that 95% of predictions will fall within 10% of the true mean.

System Type	Fuel	n	Model Type	Variables in Model	Bias 5th %	Bias 95th %	R^2 adj.
Heat Pump w Electric Backup	Electricity	39	A	savings ~ regional_runtime_metric_heating + regional_runtime_metric_cooling + excess_resistance_score	-86%	274%	0.13
Heat Pump w Electric Backup	Electricity	39	B	savings ~ comfort_runtime_metric_heating + comfort_runtime_metric_cooling + excess_resistance_score	-75%	204%	0.12
Heat Pump w Electric Backup	Electricity	39	C	savings ~ regional_runtime_metric_heating + regional_runtime_metric_cooling + dnru_reduction	-79%	237%	0.04
Heat Pump w Electric Backup	Electricity	39	C	savings ~ regional_runtime_metric_heating + regional_runtime_metric_cooling + sigmoid_integral	-75%	196%	-0.05

The best correlation model for gas usage from the gas furnace or boiler group used the comfort runtime metric as a primary metric and included secondary metrics for the heat loss constant and the HVAC constant during the heating season (Model C above). The resulting correlation factor was 0.2% ΔNAC per 1% ENERGY STAR heating reduction, with an out-of-sample bias of -38% to +99%. This results in a range of 0.12% to 0.40% ΔNAC per 1% ENERGY STAR heating reduction. The Apex team classifies this as a positive but weak correlation. Figure 14 displays the site-level savings and primary metric of heating runtime reduction, with a line of best fit based only on the primary metric.

Figure 14. Site-Level Natural Gas Savings for Gas Furnace and Boiler Group versus Runtime Reduction Using the Regional Baseline, with a Line of Best Fit



The best correlation model for electric use from the gas furnace or boiler group used regional runtime metrics for heating and cooling as primary metrics and included secondary metrics for heat loss/gain, HVAC constant for heating and cooling, and weekly temperature variance. The resulting correlation factor was 0.07% ΔNAC per 1% ENERGY STAR heating reduction, with an out-of-sample bias of -333% to +336%. This results in a range of -0.16% to 0.31% ΔNAC per 1% ENERGY STAR heating reduction. The Apex team classifies this result as no correlation. Figure 15 displays the site-level savings and primary metric of heating runtime reduction, with a line of best fit based only on the primary metric.

Figure 16 displays the site-level savings and primary metric of cooling runtime reduction, with a line of best fit based only on the primary metric.

Figure 15. Site-Level Electricity Savings for Gas Furnace and Boiler Group versus Heating Runtime Reduction Using the Regional Baseline, with a Line of Best Fit

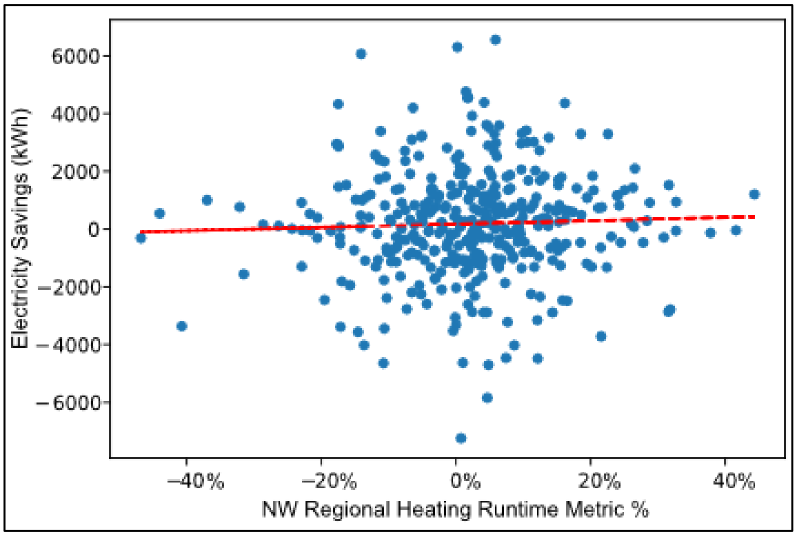
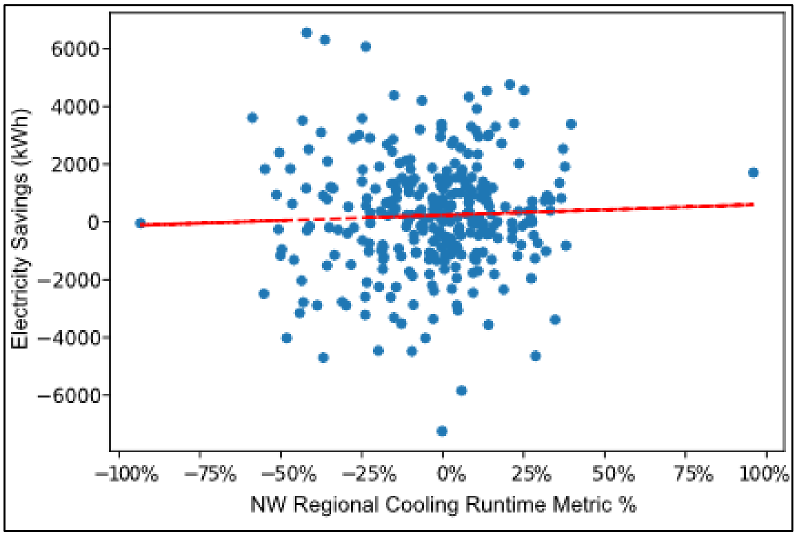


Figure 16. Site-Level Electricity Savings for Gas Furnace and Boiler Group versus Cooling Runtime Reduction Using the Regional Baseline, with a Line of Best Fit



The best correlation model for electric use from the heat pump with electric backup group used the regional runtime metrics for heating and cooling as primary metrics and included secondary metrics for the excess resistance score. The resulting correlation factor was -0.2% ΔNAC per 1% ENERGY STAR heating reduction, with an out-of-sample bias of -86% to 274%. This results in a range of -0.75% to -0.03% ΔNAC per 1% ENERGY STAR heating reduction. The Apex team classifies this result as no correlation or a weak correlation in a non-intuitive direction (increased compressor runtime could result in resistance heat savings and overall energy savings), with the caveat that the number of sites in this analysis is too low to generate a conclusive result. Given that the out-of-sample bias was lower than for the electric use model with the gas furnace or boiler group, a larger sample might have generated a tighter correlation. Also, note that the excess resistance score and the sigmoid integral performed about equally, and any future research including one should also include the other. Figure 17 depicts the site-level savings and primary metric of heating runtime reduction, with a line of best fit based only on the primary metric.

Figure 18 displays the site-level savings and primary metric of cooling runtime reduction, with a line of best fit based only on the primary metric.

Figure 17. Site-Level Electricity Savings for Heat Pump with Electric Backup Group versus Heating Runtime Reduction Using the Regional Baseline, with a Line of Best Fit

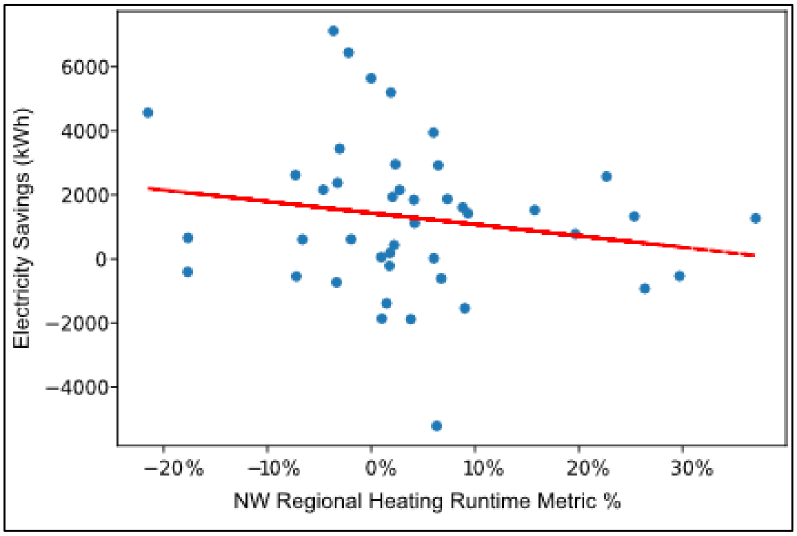
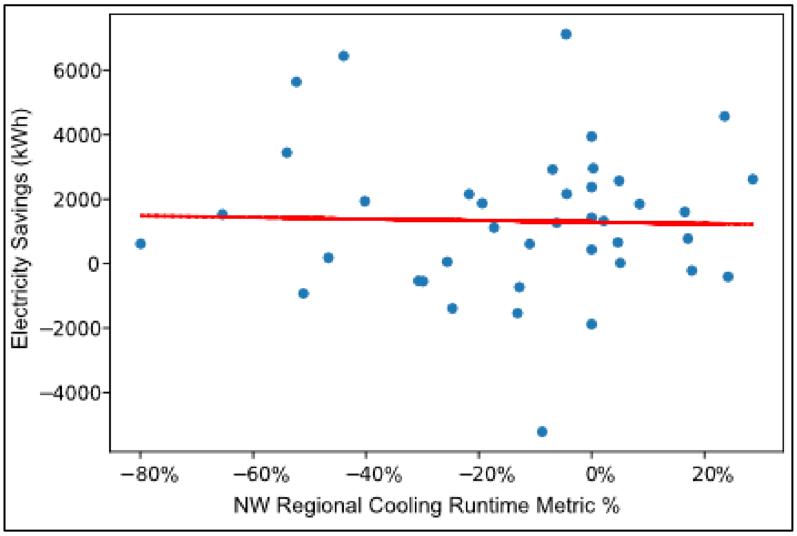


Figure 18. Site-Level Electricity Savings for Heat Pump with Electric Backup Group versus Cooling Runtime Reduction Using the Regional Baseline, with a Line of Best Fit



In summary, none of the models attempting to correlate metrics generated from thermostat metrics with site-level savings generated more than a weak correlation. Models for the primary heating fuel of gas furnaces (gas) and heat pumps (electricity) were suggestive of an underlying correlation but were not strong enough to function as bases for estimating savings for a QPL or distinguish one brand of thermostat from another. In practical terms, the out-of-sample bias error for the best model (i.e., gas savings) suggests that a thermostat model with a true savings of 32 therms could be assigned a savings value of 20 to 64 therms, a large range for a product that will be submitted to cost-effectiveness testing and differentiated from other competing products.

Therefore, the Apex team could not establish a method to use thermostat-derived metrics to estimate these energy savings with sufficient reliability for use by Northwest utilities. There are two likely causes of weak or non-existent correlation:

- The variation in ΔNAC is large and often unrelated to smart thermostat installation.
- Because the runtime reduction metrics cannot be conclusively adjusted for site-level pre-period baselines, it is likely that a large portion of the variation in this metric is tied to pre-existing behaviors while a smaller portion is because of the thermostat installation.

The Apex team expects that, while a larger study could improve the out-of-sample bias, these two factors would limit the ability of program administrators and other organizations to use thermostat-sourced metrics to predict energy savings for given groups of thermostats. A pre-period site-level baseline for indoor temperatures or thermostat runtimes could substantially improve all the generated thermostat metrics, but the Apex team recognizes that the lack of such a baseline was the impetus for developing the ENERGY STAR comfort temperature methodology.

5. Key Findings and Future Research Considerations

Finding: Smart thermostat installation resulted in statistically significant energy savings, reducing energy use by about 5% of primary heating fuel for gas furnaces and heat pumps. This result, and its consistency with other studies, suggests that smart thermostats continue to save energy. Although the primary goal of this study was not to assess aggregate energy savings, it nonetheless confirms that they exist.

Finding: Major home and life changes occurring in a similar timeframe to thermostat installation impact energy savings substantially. Compared to the analysis dataset, groups where major energy-use changes occurred differed in ΔNAC values by up to 25 therms and 1000 kWh, on the order of the impact of thermostat installation. In the Apex team's sample, these changes tended to occur after smart thermostat installation, implying that both a quasi-experimental and future installer comparison group would not generate a comparable counterfactual without adjustment.

> **Future Research Consideration:** Future studies should use either surveys like those conducted for this study, or another disaggregation method to account for major energy-use changes or these changes will introduce bias into the final results.

Finding: Energy savings were insufficiently correlated with thermostat metrics to establish a method of estimating savings for qualifying thermostats into QPLs using thermostat metrics only. The most promising model, with approximately 500 thermostats controlling gas furnaces or boilers and several secondary metrics in the model, found only a weak correlation between runtime reduction and savings.

> **Future Research Consideration:** Either a very large sample (i.e., 10,000 sites) or a different method of conducting this research is required to definitively assess the validity of thermostat metrics in predicting energy savings. A larger sample would help reduce the prediction uncertainty, but reducing bias and variability in the sample requires information about the true baseline, such as prior thermostat type and setpoints. This type of research would inherently require legal and technical infrastructure in place with customers and/or manufacturers before thermostat installation.

Future Research Consideration: There is potential to use some of the additional thermostat metrics for behavioral messaging or HVAC diagnostics by energy efficiency programs. Specifically, the metrics related to heat gain or loss can help identify issues with the building shell (for example, leaky homes, poor insulation), while metrics related to HVAC and resistance heat performance can help diagnose potential HVAC commissioning or sizing issues.

Future Research Consideration: Navigating legal and technical requirements with thermostat manufacturers is a time-intensive process. Sufficient lead time on any study or ongoing engagement should be built in to allow adequate time for these discussions.

Appendix 1: Telemetry Analysis Memo Addendum

The metrics proposed in this memo are *in addition to* the existing EPA thermostat metrics. The Apex team has also incorporated feedback from the Regional Technical Forum.

New Core Thermostat Metrics

Runtime Reduction with Regional Temperature Baselines

Description

An estimate of percent HVAC runtime reduction after installing a smart thermostat. The reduction is estimated relative to the HVAC runtime required to maintain an indoor temperature profile that is typical for similar HVAC systems in the same climate zone. This calculation is different than the typical ENERGY STAR runtime reduction metric because it uses regional baselines from the BPA Smart Residential Thermostats Indoor Temperature Baseline Study[24] instead of individual comfort temperatures.[25]

Why is the Apex team calculating this metric?

This metric is meant to capture the impact of thermostat control on HVAC runtime and energy consumption through setpoint adjustments and scheduling.

How is this metric calculated?

1. Identify the relevant baseline hourly temperature time series for a thermostat based on its climate zone and HVAC type.

2. Merge the outdoor temperature at the thermostat location in the post-period to the baseline temperature using the hour of year as a key.

3. Use the thermostat's τ coefficient from its $\alpha - \tau$ model (calculated in accordance with EPA's methodology) to estimate baseline HVAC runtime

$$RT_{base,d} = \alpha \times \frac{1}{24} \sum_{h=1}^{24} [\Delta T_{d,h(base)} - \tau]_{+}$$

4. Calculate various outputs by comparing $RT_{base,d}$ and RT_{actual} (the actual heating/cooling equipment runtime), including percent runtime reduction ($RT_{reduction}$) and absolute runtime reduction.

Demand-Normalized Resistance Utilization Reduction

Description

An estimate of the impact of thermostat control on resistance heat utilization. Thermostats that rely more on the heat pump compressor than resistance heat in the heating season will utilize less energy. Resistance heat utilization is compared to the average expected resistance heat utilization from the 2011 RBSA Metering dataset[26].

Why is the Apex team calculating this metric?

The metric captures the level of resistance heat use due to thermostat control algorithms. Better control of resistance heat should correlate with better energy efficiency.

How is this metric calculated?

1. Build an hourly time series of compressor and resistance heat runtime and outdoor temperature.

2. Calculate the outdoor temperature bin for each hour. The bin endpoints are specified in the EPA smart thermostat spec.

3. Calculate the average resistance utilization (RU) within each temperature bin.

$$RU_{bin} = \sum_{OT \in bin} \frac{(aux\ RT_{OT} + emergency\ RT_{OT})}{(aux\ RT_{OT} + compressor\ RT_{OT})}$$

4. Calculate a time series of thermal demand.

$$TD_{d,h} = [\Delta T_{d,h} - \tau]_{+}$$

5. Assign a resistance utilization value to each hour based on its outdoor temperature. Calculate the weighted average resistance utilization, using thermal demand as the weights.

$$DNRU = \frac{\sum_{d=1}^{D} \sum_{h=1}^{24} RU_{bin(d,h)} \times TD_{d,h}}{\sum_{d=1}^{D} \sum_{h=1}^{24} TD_{d,h}}$$

6. Follow steps 1 through 4 for a baseline heat pump runtime time series.

7. Calculate the demand-normalized resistance utilization (DNRU) reduction by merging the baseline resistance utilization to the actual resistance utilization time series on the hour of year.

$$DNRU_{reduction} = \frac{\sum_{d=1}^{D}\sum_{h=1}^{24}[RU_{hoy(d,h),base} - RU_{bin(d,h)}] \times TD_{d,h}}{\sum_{d=1}^{D}\sum_{h=1}^{24}TD_{d,h}}$$

Baseline Calculations

1. For the proposed runtime reduction metric and the DNRU, indoor temperature baselines are calculated by averaging indoor temperatures from the RBSA Metering dataset, grouped by climate zone and HVAC system type. The existing EPA metric uses "baseline" indoor temperatures (90th percentile core-season indoor temperatures) that are site-specific but do not clearly relate to pre-thermostat operating conditions.

2. Resistance heat utilization baseline is calculated by averaging resistance heat and heat pump compressor energy use across the RBSA Metering dataset. The energy use is converted to runtime estimates, assuming 2.5 kW and 10 kW as the heat pump and auxiliary heat capacities, respectively.

3. The excess resistance metric is a performance score; it is not a savings metric relative to a defined baseline.

Additional Test Metrics

In addition to the two new core metrics that were outlined in the Northwest smart thermostat Research Strategy, the Apex team has identified several other metrics that could prove beneficial in understanding or strengthening the correlation between the thermostat metrics and thermostat energy savings.

Delta-T Runtime Regression

Description

These are coefficients and model fit parameters from an OLS regression between runtime and Delta-T (indoor-outdoor temperature difference).

Why is the Apex team calculating this metric?

The regression is meant to provide additional perspective on the strength of the relationship between runtime and Delta-T. Poor correlation is a red flag for possible secondary heat sources which can create problems for the billing data correlation analysis.

How is this metric calculated?

1. Calculate average Delta-T and total runtime for core heating (or cooling) days.

2. Fit regression based on system type.

- Gas furnace/boiler, air conditioner

$$\Delta T_d = \beta_0 + \beta_{runtime} \cdot RT_d$$

- Heat pump

$$\Delta T_d = \beta_0 + \beta_{comp} \cdot (1 - 0.012 * (47 - T_d^{outdoor}))\, RT_d^{comp} + \beta_{res} \cdot (RT_d^{aux} + RT_d^{emer})$$

In the β_{comp} term, the factor $[1 - \rho \cdot (47 - T_{d,h}^{out})]$ is meant to reflect the fact that compressor capacity decreases with colder outdoor air temperatures. In this formulation, temperature differences are taken relative to 47°F only because nameplate capacity refers to that temperature. The parameter ρ captures the rate at which compressor capacity diminishes as outdoor temperatures drop. In systems with efficient ducts, it might be common to see ρ values around 0.012, meaning that capacity decreases by 1.2% for every 1°F decrease in outdoor air temperature. For systems with significant uninsulated ductwork in unconditioned spaces, ρ could be 2 or 3 times greater. For this study, the Apex team uses a constant $\rho = 0.012$.

3. Record the regression outputs: Model coefficients and their standard errors, R-squared, CVRMSE.

Excess Resistance Score

Description

This metric quantifies resistance usage that could have been met by available (unused) compressor capacity in the same hour or in a nearby hour. It is normalized to estimate the fraction of total thermal output (compressor + resistance) supplied with resistance heat but could have been supplied with compressor heat. As with all metrics, the final formulation should be informed by exploratory analysis with the anonymized data (some prominent decision points are noted below).

A value of zero indicates that resistance is only called when the compressor is fully utilized and cannot meet the load. Values greater than zero indicate some amount of resistance usage that could have been met with unused resistance capacity. Because resistance and compressor typically run simultaneously for stage two heating calls, it is impossible to get one value unless a system runs entirely in (resistance-only) fault mode.

The metric is built up in a way that is specific to single-speed heat pumps with electric resistance backup heat. The approach may be adaptable to two-speed and variable-speed systems if desired. It may also be adaptable to dual-fuel systems, but this would only be appropriate if the thermostat manages change-over controls.

Why is the Apex team calculating this metric?

The DNRU metric (above) and the resistance utilization sigmoid parameters (below) speak to overall resistance usage, reflecting a wide range of factors. Some highly influential factors (especially heat pump sizing) are outside of the thermostat's control. This excess resistance metric focuses on the portion of resistance usage that can be mitigated by thermostat controls for a given home and heating system.

How is this metric calculated?

1. Use model coefficients from the Delta-T runtime regression β_{comp}, β_{res} that capture relative[27] magnitude of compressor and resistance output rates. Define thermal output variables based on runtime data and the fitted model parameters.

$$Res_{d,h}^{output} \quad = \quad \beta_{res} \cdot \left(RT_{d,\,h}^{aux} + RT_{d,\,h}^{emer}\right)$$

$$Comp_{d,h}^{output} \quad = \quad \beta_{comp} \cdot \left(1 - \rho \cdot \left(47 - T_{d,h}^{out}\right)\right) \cdot \left(RT_{d,\,h}^{comp} + RT_{d,\,h}^{aux}\right)$$

$$Comp_{d,h}^{available} \quad = \quad \beta_{comp} \cdot \left(1 - \rho \cdot \left(47 - T_{d,h}^{out}\right)\right) \cdot \left(60 - RT_{d,\,h}^{comp} - RT_{d,\,h}^{aux}\right)$$

2. Define the hour-level excess resistance variable. This step is complicated because resistance usage can sometimes be avoided through pre-heating, which may involve compressor usage from the previous hour or even earlier. In defining this variable, there are risks of under-counting (by ignoring usable compressor capacity from nearby hours) and double-counting (by counting a given amount of unused compressor capacity as available to displace resistance usage in two separate hours). Because of this, the Apex team calculates three potential definitions for the hour-level resistance variable.

$$Res_{d,h}^{excess,1} \quad = \quad \text{minimum}\left(Res_{d,h}^{output},\ Comp_{d,h}^{available}\right)$$

$$Res_{d,h}^{excess,2} \quad = \quad \text{minimum}\left(Res_{d,h}^{output} + Res_{d,h-1}^{output},\ Comp_{d,h}^{available} + Comp_{d,h-1}^{available}\right)/2$$

$$Res_{d,h}^{excess,3} = \text{minimum}\left(Res_{d,h}^{output} + Res_{d,h-1}^{output} + Res_{d,h-2}^{output},\ Comp_{d,h}^{available} + Comp_{d,h-1}^{available} + Comp_{d,h-2}^{available}\right)/3$$

The first definition counts resistance as "excess" if it could have been met with unused compressor capacity from the same hour. The second definition looks at a rolling two-hour window, comparing resistance usage in each window to unused compressor capacity in that same window and dividing by two because of systematic double-counting. And the third definition uses a three-hour rolling window.

3. Define metric for overall excess resistance as a fraction of total thermal output.

$$Res^{excess,1} = \frac{\sum_{d,\,h} Res_{d,h}^{excess,1}}{\sum_{d,\,h}\left(Res_{d,h}^{output} + Comp_{d,h}^{output}\right)}$$

$$Res^{excess,2} = \frac{\sum_{d,\,h} Res_{d,h}^{excess,2}}{\sum_{d,\,h}\left(Res_{d,h}^{output} + Comp_{d,h}^{output}\right)}$$

$$Res^{excess,3} = \frac{\sum_{d,\,h} Res_{d,h}^{excess,3}}{\sum_{d,\,h}\left(Res_{d,h}^{output} + Comp_{d,h}^{output}\right)}$$

Analysis Period

The analysis period for the thermostat telemetry data is typically defined and can extend over multiple years. At the same time, operational changes are not uncommon, and many sites exhibit a clear change in operation at some point over longer time horizons.

The Apex team has included a procedure to calculate a separate set of metrics in each year in addition to using the entire dataset. For cooling, the years are split on January 1 and for heating, they are split on July 1 to get contiguous cooling and heating seasons, respectively. So, for every thermostat, in both heating and cooling seasons, there will be one set of all metrics for the full analysis period and one set for each "seasonal year" (Jan 1–Dec 31 for cooling, Jul 1–Jun 30 for heating).

For each thermostat, the Apex team will investigate the best metric set for the correlation analysis using the model fit parameters captured.

No-HVAC Temperature Constants

Description

The average rate of indoor temperature increases and decreases relative to the indoor-outdoor temperature difference when the HVAC systems are not actively heating or cooling. Four separate metrics are calculated: a heat gain constant and a heat loss constant during cooling and heating seasons.

Why is the Apex team calculating this metric?

The variable will allow the team to control for different building shell conditions and serve as a proxy for building stock. A higher value for these temperature constants indicates rapidly changing indoor temperatures caused by poor insulation or air sealing, among other factors. The impact of thermostat control on energy use is expected to depend on building shell conditions, so these variables may be used as control variables in the energy use correlation.

How is this metric calculated?

1. Create an hourly time series with heating/cooling runtime and indoor/outdoor temperatures.

2. Calculate the hourly temperature change rate. This is the difference between indoor temperature (IT) in the current hour minus the previous hour divided by the difference between indoor and outdoor (OT) temperature difference in the current hour.

$$TG_{d,h} = \frac{(IT_{d,h} - IT_{d,h-1})}{(OT_{d,h} - IT_{d,h})}$$

3. Heat gain constant: calculate the average temperature change rate for the hours when the outdoor temperature exceeds the indoor temperature, and the heating and cooling runtimes are under 5 minutes.

4. Heat loss constant: calculate the average temperature change rate for the hours when the indoor temperature exceeds the outdoor temperature, and the heating and cooling runtimes are under 5 minutes.

HVAC Temperature Constant

Description

The average rate of indoor temperature increases or decreases relative to the indoor-outdoor temperature difference during times when the HVAC systems are actively heating or cooling. Two separate metrics are calculated during cooling and heating seasons.

Why is the Apex team calculating this metric?

The variable should allow the team to control for different levels of HVAC performance. A higher value for these constants indicates a home that is rapidly heated or cooled by its HVAC system (potentially due to an oversized system, for example), while a lower value indicates a home with a slower temperature response (undersized system). The impact of thermostat control on energy use is expected to depend on HVAC response, so these variables may be used as control variables in the energy use correlation.

How is this metric calculated?

1. Calculate the hourly temperature change rate as explained previously.

2. In the heating season: Calculate the average temperature change rate for the hours when the indoor temperature exceeds the outdoor temperature, and the heating runtime is over 15 minutes.

3. In the cooling season: Calculate the average temperature change rate for the hours when the outdoor temperature exceeds the indoor temperature, and the cooling runtime is over 15 minutes.

Indoor Temperature Variance

Description

A quantification of indoor temperature variation during heating and cooling seasons as a proxy for the use of thermostat features.

Why is the Apex team calculating this metric?

This simple metric that should be highly correlated with the runtime reduction metric calculated by the standard ENERGY STAR software. Its inclusion will allow the team to explain findings related to the base ENERGY STAR software metrics, without assumptions about each individual’s baseline. If indoor temperature variance is the only meaningful portion of the runtime reduction calculation with a comfort temperature, it will correlate just as well with energy savings.

The EPA specification does not include any scheduling information, so it cannot quantify occupant behavior and thermostat use preferences. This metric is intended to test whether the team can control for occupant behavior on thermostat savings. Higher indoor temperature variance should correspond with higher use of thermostat scheduling and setbacks, whereas a customer that uses a smart thermostat as a manual thermostat with a constant temperature hold should experience minimal variance in the indoor temperature (and correspondingly minimal energy savings).

How is this metric calculated?

1. Overall temperature variance: The standard deviation of indoor temperature separately calculated across all hours in the heating and cooling seasons.

2. Weekly temperature variance: The standard deviation of indoor temperature in a typical week, separately calculated across the heating and cooling seasons. This is done by first grouping hourly temperature values by the hour of week.

Sigmoid Model Parameters for Resistance Heat

Description

Three additional metrics that can be used individually or in combination to reflect the reduction in resistance heat utilization caused by a smart thermostat.

Why is the Apex team calculating this metric?

It is not yet known whether DNRU will sufficiently capture the relationship between resistance heat utilization and energy savings. These metrics are intended to provide additional parsimonious metrics that can serve as an alternative in the correlation analysis. Resistance utilization by temperature bin has a sigmoid functional form, and fitting such a curve by thermostat will allow the resistance utilization behavior to be described in either two metrics to describe the full behavior across bins or one totaled metric to provide a single input for correlation analysis.

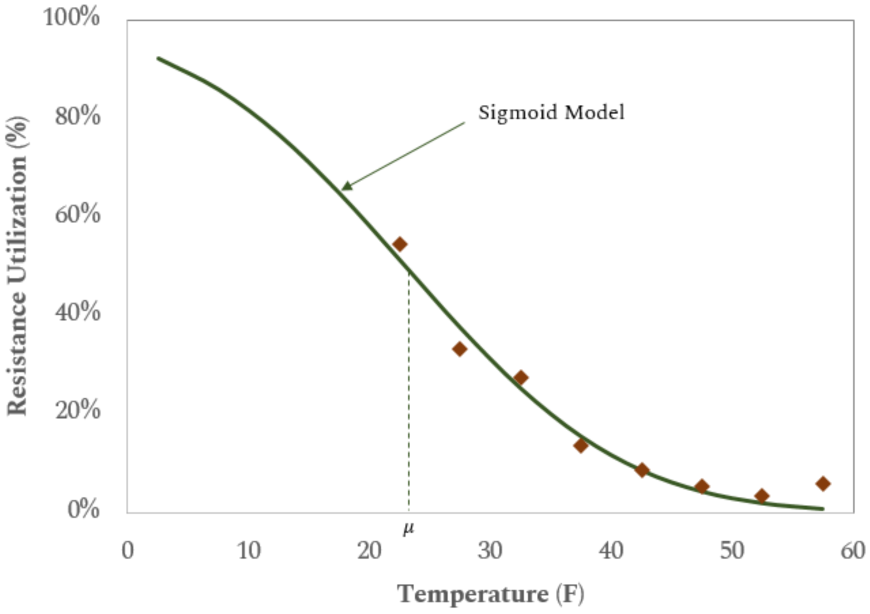
How is this metric calculated?

1. Calculate RU_{bin}, by following steps 1-3 under the *Demand Normalized Resistance Utilization* metric.

2. Fit a sigmoid (reverse S-shaped) model using the temperature bin midpoints (TBM_{bin}) as the independent variable and the resistance heat utilization as the dependent variable. This model will yield two parameters, μ (the average temperature below which resistance heat is used more than 50% of the time) and σ (the temperature delta required to go from 33% resistance utilization to 67% resistance utilization).

$$RU_{bin} = \frac{1}{2} \times \left(1 - \mathrm{erf}\left(\frac{TBM_{bin} - \mu}{\sigma \times \sqrt{2}}\right)\right)$$

3. Integrate the sigmoid function over all temperatures between 0 and 60°F to yield a third metric – the sigmoid integral.



Appendix 2: Billing Analysis Details

Appendix 3: Correlation Analysis Details

.