The Elusive Missing Link: Correlating Billing Analysis and Thermostat Data in the Northwest

Jon Koliner, Apex Analytics
Lauren Gage, Apex Analytics
Hassan Shaban, Empower DataWorks
Josh Rushton, Rushton Analytics
Phillip Kelsven, Bonneville Power Administration
Dan Rubado, Energy Trust of Oregon

ABSTRACT

Smart thermostats have become a major component of residential demand-side management programs in the US. As devices that effectively turn existing home heating and cooling systems into connected systems, smart thermostats enable efficient setpoint optimization around user schedules, bring-your-own-device demand response programs, and aggregation into distributed virtual power plants. At present, they can shift demand to times with lower carbon generation. In the future, they can stabilize a renewable-powered grid by shifting load to times of highest capacity. Yet, questions remain about the energy savings of different smart thermostats with differing features.

The Northwest Smart Thermostat study sought to address these persisting questions about thermostat performance by establishing a method to estimate thermostat energy savings using metrics generated from thermostat data only. The study contacted over 50,000 thermostat users in the Northwest to opt in via a web portal, achieving an opt-in rate of 4%, then gathered thermostat metrics from four manufacturers and billing data from four utilities. The opt-in process included a short survey on the timing of major home energy use changes, which showed more changes happened during and after thermostat installation. We joined site-level billing analysis results with performance metrics calculated from each thermostat's data, including ENERGY STAR metrics and new metrics calculated for the study. The study found primary heating fuel savings consistent with other recent studies (≈5% of whole home usage), but found weak or nonexistent correlations between thermostat metrics and site-level energy use changes, testing the inclusion of different primary and secondary metrics in modeling.

Purpose

The primary objective of this study was to develop a method to estimate energy savings for smart thermostats based on thermostat metrics. For the purposes of the study, we defined 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.

Features of smart thermostats on the market have evolved over the past several years and the number of manufacturers has expanded. Today, new and incumbent manufacturers are releasing more economical models with pared down features, while software updates instantaneously modify operating characteristics of entire installed bases. The common evaluation method of pre-post billing analysis requires nearly two years of combined pre- and post-installation data to establish energy savings, and rarely provides a sufficient breadth of insight to differentiate between specific models and feature sets.

To address the evolving market, The Environmental Protection Agency, via their ENERGY STAR® initiative, had established energy savings criteria based on thermostat performance metrics (i.e., thermostat metrics) entirely derived from post-period thermostat data (ENERGY STAR 2022). These metrics included runtime reduction compared to multiple baseline conditions and planned to incorporate resistance heat utilization in the future. The calculation of these metrics is performed by the Connected Thermostat Field Savings Software, run by each manufacturer to protect the anonymity of customers.

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 (RTF 2016) to develop the smart thermostat Research Strategy (RTF 2018) for this study using the ENERGY STAR metrics as a basis. Northwest stakeholders[1] 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 (ENERGY STAR 2016) 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 directly from these thermostat metrics. Therefore, we conducted this study to assess the viability of a key linkage, a correlation between energy savings and thermostat metrics, that would enable future savings certifications.

Recruitment and Data Collection

Opt-In Process

The study attempted to recruit a 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. The study included models from four manufacturers (Nest, ecobee, Resideo, and Emerson), and associated electric and gas consumption data from four utilities (Avista Utilities, Clark Public Utilities, Puget Sound Energy, and Energy Trust of Oregon [representing PacificPower and Portland General Electric]).

---

[1] The Advisory team consisted of 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.

The effort to secure this data required that the study team secure data-sharing agreements from manufacturers and utilities, which ultimately allowed an opt-in process from end users through manufacturer and utility-specific websites. The study team used the opportunity provided by this direct contact requirement to deliver a short survey to these users. 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.

Both utilities and manufacturers emailed customers to request that they opt into data sharing with the study. Between both efforts, 50,000 customers were contacted, with roughly 2,000 opting to share data with the study, for an opt-in rate of 4%. Resideo and ecobee provided data for an additional 2,000 users, enabled by preexisting data sharing agreements from prior pilots and studies.

Table 1. Contacted thermostat users and thermostat users whose data is included in the study.

	Total	Ecobee	Emerson	Nest	Resideo
Contacted/Emailed	50,072	NA[2]	6,099	28,895	15,078
Thermostat Users in Study	3,943	1,641	587	1,177	538[3]

Survey of Changes

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. The final set of questions was a compromise between suspected relevant data and a minimizing burden on thermostat users. Therefore, not all possible energy use changes were included. The survey asked if they conducted any of the following and the timing (i.e., after, during, or before thermostat installation):[4]

- Installation of a new heating and/or cooling equipment (e.g., furnace, central AC, window AC; water heater is not applicable).
- Purchase of an electric vehicle.
- Completion of a major renovation that may affect energy use (e.g., new windows, insulation, remodeling, additions).
- Increase in the number of people living in your home?
- Installation of any other smart home devices (e.g., smart speakers, smart lights, home displays, or home mesh Wi-Fi systems)[5]

The most striking result from this survey is that major changes tended to follow or coincide with thermostat installation, not precede it (40% vs 19%). Although thermostat users most frequently reported not taking a major action, changes were more likely during or after thermostat installation than before. Figure 1 shows this asymmetry in aggregate.

---

[2] ecobee provided telemetry data for 1,641 users with consent gathered via its eco+ optimization opt-in process

[3] Resideo provided thermostat metrics for 500 users in Energy Trust of Oregon Connected Savings programs

[4] To avoid COVID-19 pandemic complications, the survey asked customers to respond for the timeframe between January 2017 and February 2020 (i.e., before the COVID-19 outbreak), but the questions were worded to refer to the time compared to thermostat installation, not a specific set of dates. The opt-in process and websites were deployed between September and December 2020.

[5] Due to timing, this question was only asked for thermostat users from one manufacturer.

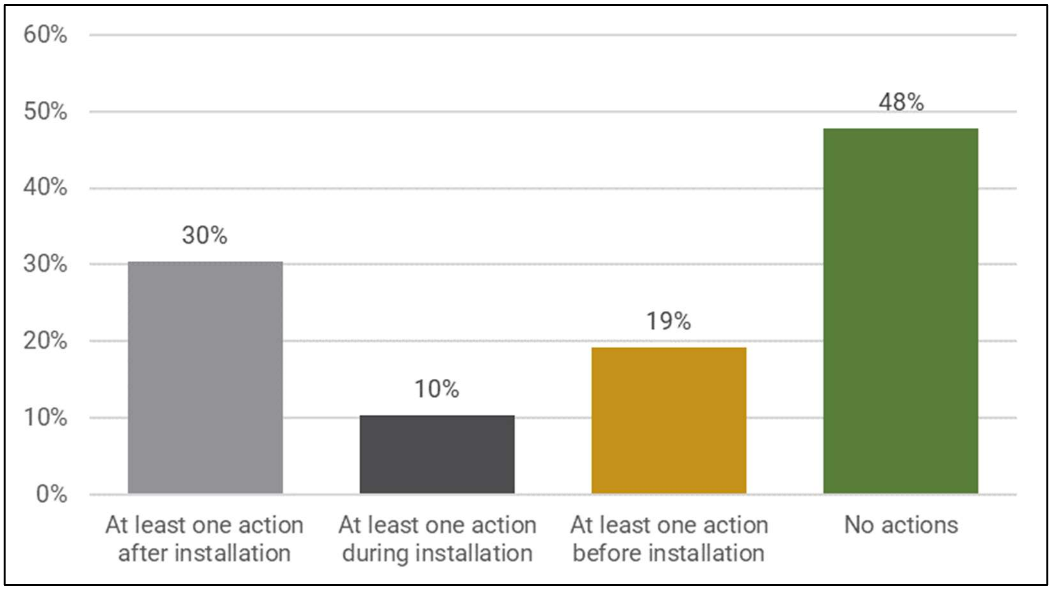

Figure 1. Aggregated survey responses across all questions about major energy use changes relative to the timing of smart thermostat installation. Note that users could select different responses for each type of change, so these do not add to 100%.

When looking at specific actions (Figure 2), this survey showed that thermostat users tended to install new heating and cooling equipment, conduct a major renovation and/or have an occupancy increase during or after thermostat installation. EV purchases were more symmetric in time across all surveyed users (6% vs 5%).

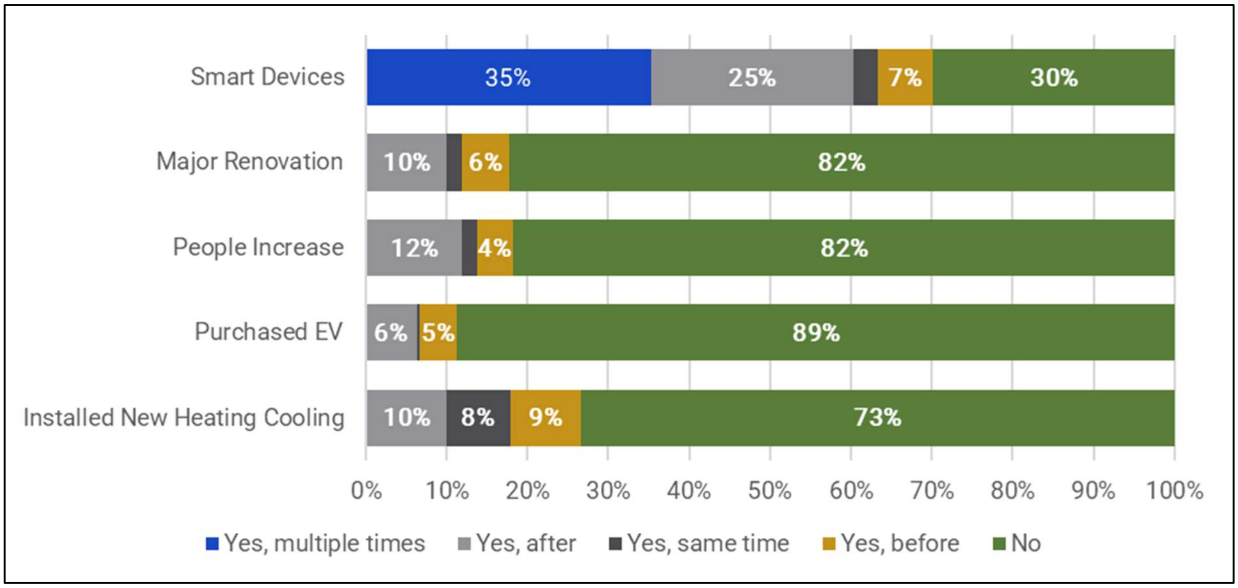

Figure 2. Survey responses across all questions about major energy use changes relative to the timing of smart thermostat installation. The question about smart devices was added for Nest users, and the "multiple times" response was added for that question.

Thermostat Metrics

Given that manufacturers would be running the code to generate the thermostat metrics only once, the Northwest stakeholders requested that the study team include as many descriptive output metrics to the CT Field Savings software as possible, with the goal of accurately representing the Northwest region. In particular, the stakeholders wanted to implement measured baselines from regional studies. Three manufacturers provided anonymous hourly telemetry data sets early in the study period to support the thermostat telemetry analysis and modification of the ENERGY STAR software in this fashion. The data sets were in the same format that the "V2" specification for ENERGY STAR uses as inputs. These data enabled the team to test and create new metrics to assess whether they reflected observable HVAC operation patterns. It also allowed the team to ensure that the input telemetry data from the manufacturers worked correctly with the updated software.

The entire set of outputs from the original ENERGY STAR software were retained in the updated software. These included:

- **Runtime difference from comfort baseline**: How different is actual HVAC runtime from a baseline that assumes the indoor temperature is always maintained at a user-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?

Additionally, the study team developed and tested additional metrics that were included in the software provided to manufacturers. (For additional details, see Appendix 1: Detailed New Metrics[6])

- **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?
- **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?

All new metrics were integrated into a separate version of the ENERGY STAR Connected Thermostat codebase, which was published by the study team for use by the thermostat manufacturers (Shaban 2021).

---

[6] Appendix 1 fully describes these metrics because the code and the calculations may be useful to manufacturers, utilities, and researchers in future analyses.

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. The study did not receive exact thermostat model from manufacturers, and we recommend asking for this information in the future.

Billing Analysis: Site Level and Pooled

Approach

Billing data provided by the utilities[7] was analyzed in two ways for this study: at the site level and within a pooled analysis. As noted above, the primary objective of this study was to develop a method to estimate energy savings for smart thermostats based on thermostat metrics. This required a correlation between the change in site-level energy use and metrics produced from thermostat data. Site-level energy use, normalized for weather, is referred to as Normalized Annual Consumption (NAC), and its change between the pre- and post- period is denoted here as ΔNAC. Secondary study objectives included establishing energy savings from smart thermostat products, which was best performed with pooled regression analysis that incorporated future thermostat installers as a comparison group in the post-period.

For the purposes of this paper, we focus primarily on the site-level analysis and correlations, and direct readers to the Northwest Thermostat Study (Apex 2021) for a full description of the pooled analysis for aggregate savings, including specification, detailed results, and alternate models. The sample gathered for this study was intended to provide a breadth of datapoints for correlation, and not to serve as a representative sample either for utility programs or the general population. The methods and results for the aggregate savings pooled analysis, despite being applied to this convenience sample, align with other more purpose-built studies, which we direct the reader towards (Guidehouse 2020, Guidehouse 2021, Rubado 2020, Kelsven 2016, CPUC 2021).

The baseline offset model[8] 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 model terms are:

$ADC_{p,t}$ = Site-specific average daily energy use for site p in month t
$\alpha_{0,p}$ = Site-specific fixed effect for site p
$HDD_{p,t}$ = Base 60 heating degree days (HDD) experienced by site p in month t
$CDD_{p,t}$ = Base 65 cooling degree days (CDD) experienced by site p in month t

---

[7] To avoid COVID-19 pandemic complications, data were collected for the period January 2017 to February 2020.

[8] The study team developed comparison group adjustments, or "baseline offsets," to apply to the site-level models to account for exogenous changes by non-participants. The team used a simple 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. We added a window2 term to represent the post-period window and isolate the exogenous energy use changes for the comparison thermostat users. This model was run for both natural gas and electricity data.

$window2_t$ = A binary variable indicating whether month t is in the post-period window of analysis

$post_{p,t}$ = A binary variable indicating whether site p has had a smart thermostat installed in month t

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

β_{post} = Coefficient estimating the difference in site-specific energy use due to thermostat installation

The site-level billing analysis followed CalTRACK guidelines (CalTRACK 2019). The team joined the data with NOAA weather data and cleaned it and defined the analysis periods. The CalTRACK methods, as implemented in the open-source library eemeter (Open EE 2020), were then applied to fit variable base degree day models to the billing data. While the variable base process and CALTRACK methods are more involved, the model takes the form:

$$ADC_{p,t} = \mu_p + \beta_{HDD}HDD_{p,t} + \beta_{CDD}CDD_{p,t} + \varepsilon$$

where μ is the mean use for site p, similar to $\alpha_{0,p}$.

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.

The two models were each used to estimate the NAC during the reporting period. We multiplied the model coefficients for each by the total degree days in that period in order to generate comparable estimates for all thermostats in the study, regardless of any missing data. The difference between the baseline and reporting NAC was ΔNAC, or site-level savings, an estimate of savings during the reporting period due to the installation of the smart thermostat. The team applied the adjustment offsets as calculated above to all site-level savings shown in this report.

There was substantial attrition from the sample of thermostat users whose data was released to the study, as shown in Table 2. A primary source of attrition was due to users who opted in or whose data was provided by manufacturers despite being outside a participating utility territory or having installed their thermostat in circumstances where their billing data was not useable (e.g. after only a few months in their home or outside the study window). This was unavoidable due to the study’s design; user data that manufacturers had access to did not guarantee a clean geographic match or the existence of a clean billing data set.

Table 2. Remaining thermostat users after each data joining, cleaning, or filtering step

Attrition Step	Total Thermostat Users
In the Study	3,943
With Thermostat Data	3,367
Match to Billing Data	1,452
At least 1 month of pre- and post- billing data	1,166
At least 9 months of pre- and post- billing data	805
Absolute value of ΔNAC less than 50%	776
More than 30 core heating days	765
Remove top and bottom 0.5% by energy use	762
Survey filters	587

By fuel and system type, the final sample included 497 sites with gas furnaces, natural gas consumption data, and thermostat data. Other cells were much smaller, and the only other group with sufficient power to merit in-depth analysis was the heat pump with electric backup group as shown in Table 3.

Table 3. Thermostat users by System and Fuel Type

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

Results

The pooled analysis found statistically significant savings of approximately 5% for the primary heating fuel of both gas furnace or boiler sites and heat pump sites, Table 4. Despite the convenience sample, this value aligns with other regional and national studies.

Table 4: Pooled Analysis Results

Heating System Type	Fuel	n	Comp. Group n	Post-Installation		
				Average Savings	Standard Error	Percent Savings
Gas Furnace or Boiler	Electricity	550	104	-220 kWh	110 kWh	-2.4%
	Gas	678	133	43 therms	20 therms	5%
Heat Pump w Elec. Backup	Electricity	73	15	670 kWh	402 kWh	4.5%

The study calculated average site-level savings adjusted by comparison group usage changes, which differ somewhat from the pooled analysis results due to two primary factors:

- More stringent inclusion criteria in the site-level analysis due to both the needs of the site-level model and filtering on thermostat metric information
- The inclusion of varying levels of thermostat optimization in the site-level results, which the pooled model could model and isolate by site and month[9]

The aggregated savings from site-level analysis are not reported here to limit confusion – the pooled analysis is the appropriate method to calculate aggregate savings although aggregate savings are not the goal of the study.

Table 5 shows the difference in site level savings estimates from the rest of the data set for those sites with major energy use changes during or after smart thermostat installation. Given the substantial impact of these changes, users who reported these actions were excluded from the report's savings estimates and correlation analysis (below).

Table 5. Difference in site-level savings estimates for sites with major energy use changes during or after smart thermostat installation.

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

The timing of these major changes has implications for quasi-experimental design and future thermostat installers as controls. The use of either matched controls or future installers to extract thermostat energy savings is predicated on the assumption that their behavior during the installers' post-period deviates only by thermostat installation. However, if thermostat installation is a triggering event for other major energy use changes, the installers' energy

---

[9] Note that the inclusion of optimization is intentional – the thermostat metrics will capture this effect, so it is retained in the energy savings estimates for use in correlation analysis

changes due to those events will be indistinguishable from changes due to the installation itself, and therefore introduce unexplained bias into the final estimates.

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. As noted above, 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, and therefore whether they could be used in such a method.

We 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 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, we 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 R2 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 study 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 NMBE is the average difference between the predicted and actual savings divided by the average savings for the out-of-sample group. This metrics indicates 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).

Results

Across all correlation models, across thermostat types and metrics, the study found weak or no correlations. Said another way, the study found that 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 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. 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. For natural gas heated homes like those included in this analysis, the reduction in ΔNAC from thermostat installation should be approximately equal to the ENERGY STAR heating reduction.[10] We classify the model fit as a positive but weak correlation, with insufficient predictive power to be used in future savings estimates. Figure 3 displays the site-level savings and primary metric of heating runtime reduction, with a line of best fit based on the primary metric.

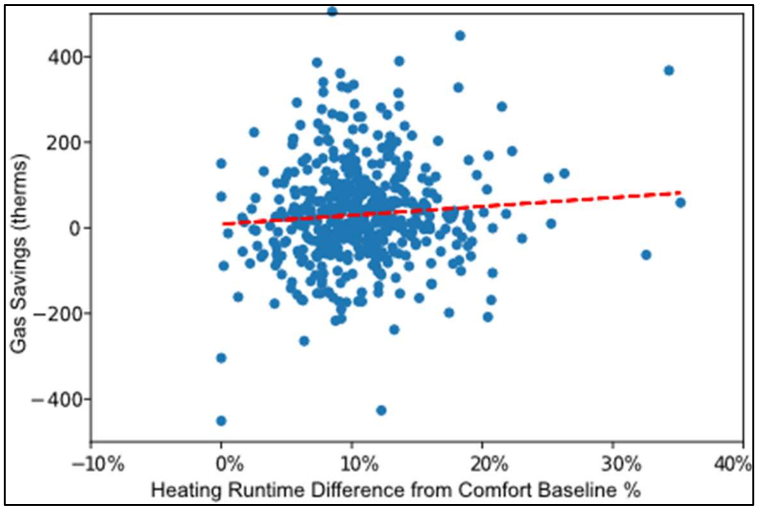

Figure 3. ΔNAC derived from natural gas pre- and post- period data versus Heating Runtime Difference from Comfort Baseline derived from thermostat post-period data alone.

The regional runtime metric performed slightly worse than the comfort runtime metric for gas heated homes, but various combinations of independent variables performed better (R^2, NMBE) with the regional runtime metric than the comfort runtime metric. The regional runtime metric also achieved the best model fit for heat pump homes, but these correlations had an NMBE of -75% to +196%. The correlation between these two metrics shows a moderately strong but surprisingly inconsistent relationship, Figure 4.

---

[10] Note that there is clear correlation ($p << 0.1\%$) between baseline NAC and baseline runtime, but a simple regression retains a substantial amount of unexplained variance ($R^2=0.09$). A given user may have other gas appliances, and NAC is affected by system size and efficiency. Runtime reduction versus ΔNAC would contain variance from system size and efficiency, but would be unaffected by the presence of other gas appliances.

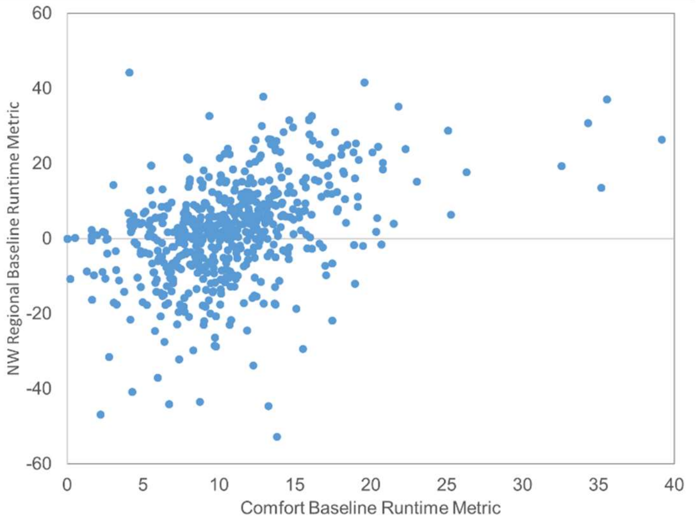

Figure 4. The Northwest Regional Baseline Runtime Metric versus the ENERGY STAR Comfort Baseline Runtime metric, in percent reduction.

The study team tested 28 linear correlation models. Table 6 displays models for each system and fuel type combination:

- 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

In addition to the adjusted R^2 for each model, Table 6 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)[11]. The adjusted R^2 describes the portion of variability in savings that the model captures.

---

[11] 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.

Table 6. 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
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

Summary

The primary objective of this study was to develop a method to estimate energy savings for smart thermostats based on thermostat metrics. Its findings preclude creating such a method from these data sets. However, this study offers a number of new insights into the relationship between user behaviors, thermostat installation, metrics generated from thermostat data, and billing analysis results. The complexity of gathering opt-ins and aligning those users with utility sites resulted in a sample size that precludes definitive conclusions. Nonetheless, some results can be incorporated into future studies and discussions.

First, the study generated several new heat pump performance metrics for use in the Northwest version of the ENERGY STAR software. Demand Normalized Resistance Utilization, Excess Resistance Score, or the Sigmoid fit to Resistance Utilization bins may provide new means of characterizing the degree to which heat pumps reduce reliance on backup heat. The sample size for this study was too small to extract definitive results, so we encourage others to use these metrics in any thermostat-based efforts in the future.

The survey found that smart thermostat installation precedes or coincides with other major energy use changes more than it follows them. Analysis of the energy use data suggests that the confounding impacts of these actions can bias energy savings estimates. We suggest that future studies take the time trend inherent in these choices seriously, and account for these actions either through a disaggregation scheme, a survey, or extensive data on the sampled users.

Finally, the correlation analysis found a weak to nonexistent correlation between metrics generated by the thermostat and site-level energy use changes. Although the sample was small, the estimated correlation was one fifth as strong as it should be for gas heated homes. For the goal of the study, this finding disqualifies thermostat metrics as useable for calculating energy savings for new products. Taken on its face, the finding also suggests that neither a site-specific post-only baseline (ENERGY STAR) nor a sampled population-level estimate for a pre-period baseline (Indoor Temperature Study) can explain the majority of energy savings from these devices.

REFERENCES

Apex (Apex Analytics), CLEAResult, Energy 350, and Empower Dataworks. 2021. “Northwest Smart Thermostat Research Study.” On behalf of NEEA. https://neea.org/resources/northwest-smart-thermostat-research-study

CalTRACK. 2019. “CalTRACK Methods.” Accessed March 2022. http://docs.caltrack.org/en/latest/methods.html

Conduit NW. 2019. “Smart Residential Thermostats Indoor Temperature Baseline Study.” Accessed March 2022. https://conduitnw.org/Pages/RETACProject.aspx?rid=14

CPUC (California Public Utilities Commission). 2021. “Impact Evaluation of Smart Thermostats.” https://pda.energydataweb.com/api/downloads/2508/CPUC%20Group%20A%20Residential%20PY2019_SCT%20Final%20Report_toPDA.pdf

ENERGY STAR. 2016. “ENERGYSTAR® Program Requirements For Connected Thermostat Products.” https://www.energystar.gov/sites/default/files/asset/document/ENERGY%20STAR%20Program%20Requirements%20for%20Connected%20Thermostats%20Version%201.0.pdf

ENERGY STAR. 2022. “Smart Thermostats Key Product Criteria.” Accessed March 2022. https://www.energystar.gov/products/heating_cooling/smart_thermostats/key_product_criteria

Guidehouse. 2020. “ComEd Advanced Thermostat Evaluation.” Prepared for ComEd. https://ilsag.s3.amazonaws.com/ComEd-Adv-Thermostat-Research-Report-Final-2020-11-10.pdf

Guidehouse. 2021. “Residential Wi-Fi and Programmable Thermostat Impacts.” Prepared for the Massachusetts Program Administrators. https://ma-eeac.org/wp-content/uploads/MARES24-Final-Report-2021-09-29.pdf

Kelsven, P., Weber, R., and Urbatsch, E. 2016. “Nest Learning Thermostat Pilot Program Savings Assessment” Prepared for BPA & FPUD. https://www.bpa.gov/EE/Utility/research-archive/Documents/BPA%20-FPUD-Nest-Thermostat-Pilot-Savings-Assessment.pdf

NEEA. 2022. “Northwest End Use Load Research Project Energy Metering Data.” Accessed March 2022. https://neea.org/data/nw-end-use-load-research-project/energy-metering-study-data

Open EE. 2020. “EEweather: Weather station wrangling for EEmeter.” Accessed March 2022. https://eeweather.readthedocs.io/en/latest/

RTF (Regional Technical Forum). 2016. “RTF Research Strategy: Residential Connected Thermostats and Services.” https://nwcouncil.app.box.com/s/v73hd6fq07zuspgs20hi5pybl5wrzci7

RTF (Regional Technical Forum). 2018. “Smart Thermostat Savings Task Force Research Strategy.” https://conduitnw.org/pages/file.aspx?rid=4697

Rubado, D. 2020. “Memo: 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

Shaban, H. 2021. “Connected Thermostat draft field savings metrics module.” https://github.com/hshaban/epathermostat_nw

APPENDIX 1: DETAILED NEW METRICS

The Runtime Difference from Hourly Indoor Temperature Baseline: 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 (Conduit NW 2019) instead of individual comfort temperatures. As with other runtime reduction metrics, it is meant to capture the impact of thermostat control on HVAC runtime and energy consumption through setpoint adjustments and scheduling, and is calculated as follows:

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 α-τ 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.

HVAC and no-HVAC Temperature Change Rates: The average rate of indoor temperature increases and decreases relative to the indoor-outdoor temperature difference when the HVAC systems are either running or not running. When the system is not running, this change rate describes shell characteristics of the building. When the system is running, this change rate describes the sizing of the system relative to the building size. The impact of thermostat control on energy use is expected to depend on both building shell conditions and the ability to displace HVAC load, so these variables may be used as control variables in the energy use correlation analysis. To calculate these metrics:

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.
5. **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.
6. **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.

DNRU Reduction: 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 (NEEA 2022). The metric captures the level of resistance heat use due to thermostat control algorithms within a single score. Better control of resistance heat should correlate with better energy efficiency.

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.

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

Excess Resistance Score: 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 that 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.

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.

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

$$\begin{aligned}
Res_{d,h}^{output} &= \beta_{res} \cdot (RT_{d,h}^{aux} + RT_{d,h}^{emer}) \\
Comp_{d,h}^{output} &= \beta_{comp} \cdot \left(1 - \rho \cdot \left(47 - T_{d,h}^{out}\right)\right) \cdot \left(RT_{d,h}^{comp} + RT_{d,h}^{aux}\right) \\
Comp_{d,h}^{available} &= \beta_{comp} \cdot \left(1 - \rho \cdot \left(47 - T_{d,h}^{out}\right)\right) \cdot \left(60 - RT_{d,h}^{comp} - RT_{d,h}^{aux}\right)
\end{aligned}$$

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 we calculate three potential definitions for the hour-level resistance variable.

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

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

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.

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

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

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

Integral of Sigmoid Resistance Function: DNRU and Excess Resistance Score are complex metrics that mix thermal load calculation with resistance heat runtimes. Fitting the standard ENERGY STAR bins for resistance utilization with a sigmoid function, shown in Figure X, provides three metrics (μ, σ, $\int S$) which can serve as a simpler alternative in the correlation analysis. Resistance utilization by temperature bin has a sigmoid functional form. Fitting such a curve for each thermostat allows 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.

1. Calculate RU_{bin}, by following steps 1-3 under the DNRU metric.
2. Fit a sigmoid (reverse S-shaped) model[12] 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).

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

3. Integrate the sigmoid function over all temperatures between 0 and 60°F to yield a third metric – the sigmoid integral ($\int S$).

---

[12] Note that "erf" refers to the error function: https://mathworld.wolfram.com/Erf.html

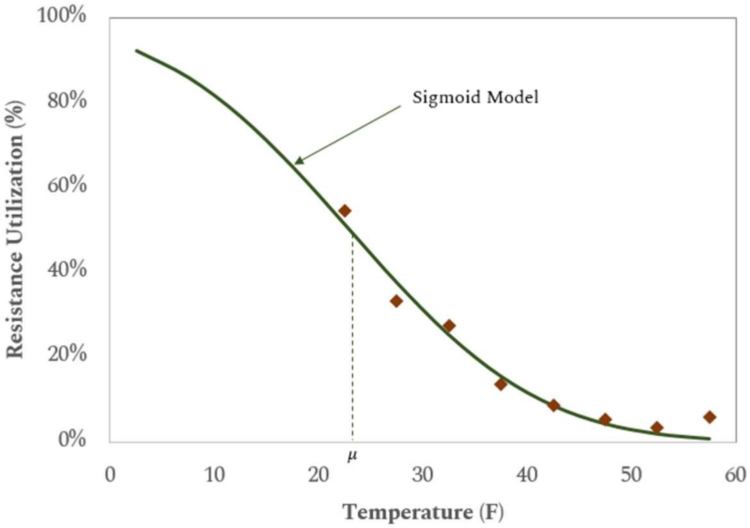

Figure 5. An example of a sigmoid model fit to resistance utilization bins.