Energy Trust of Oregon

MEMO

Date: February 7, 2020
To: Board of Directors
From: Dan Rubado, Evaluation Project Manager
Subject: Summary of Recurve Residential Smart Thermostat Impact Analysis

Energy Trust used an impact analysis tool built by Recurve Analytics to evaluate gas and electricity savings from smart thermostats installed in single-family homes with forced air heating systems in Oregon from 2015 to 2017. The tool used monthly utility billing data to conduct pre/post analyses of whole home energy usage. Energy usage data are weather-normalized using typical meteorological year data. Normalized annual energy usage in the year immediately preceding the installation is compared with that of the year immediately following installation. The change in normalized annual energy usage is then evaluated against changes in energy usage during the same time period in two comparison groups—a site-level matched non-participant comparison group and a group of homes that installed smart thermostats in later years (future participants). These calculations provide two estimates of the average annual energy savings resulting from the measure given typical weather conditions.

We restricted the analysis to smart thermostat purchases where no other efficiency measures were installed in the home during the analysis period. This was to isolate the energy impact of smart thermostats, although these homes may not be totally representative of the larger population of homes installing smart thermostats. Several standard data screens were also applied to remove atypical homes. As shown in the Recurve snapshot reports that follow this memo, energy savings were small but statistically significant in most scenarios with gas-heated homes. Savings in electrically-heated homes appeared to be negligible, although the sample size was small and the precision was poor, making this finding uncertain. We focused our analysis on homes located across the state to maximize sample sizes.

Heating zones are geographic areas defined by the Regional Technical Forum, based on the number of heating degree-days during a typical winter. Heating zone 1 represents areas of the state with relatively mild winters, such as Western Oregon. Heating zones 2 and 3 represent areas of the state with cold winters, like the mountains and Central and Eastern Oregon. Most of our analyses spanned across heating zones to preserve sample size and because Energy Trust's smart thermostat measures are not stratified by heating zone.

From 2015 to 2017, overall average gas savings in gas-heated homes ranged from 24 to 32 therms per year (+/- 7 therms) or 3-4% of baseline gas usage. There were 924 treatment homes analyzed, which had an average annual baseline gas usage of 713 therms. They were distributed across Western and Central Oregon but concentrated in the Portland metro area. For heating zone 1, during the same time period, average gas savings in gas-heated homes ranged 26 to 32 therms per year (+/- 7) or 4-5% of baseline gas usage. There were 878 treatment homes analyzed in heating zone 1, which had an average annual baseline gas usage of 704 therms. These homes were concentrated in the Portland metro area with some distribution across heating zone 1 in

Western Oregon. Heating zone 1 results were nearly identical to the overall results because 95% of homes in the treatment group were located in heating zone 1. We were unable to quantify savings in heating zone 2 due to the small number of homes available for analysis. For both the overall and heating zone 1 analyses, the matched and future comparison groups provided good representations of the baseline gas usage in the treatment group and a reasonable point of comparison as similar homes that did not install smart thermostats. The large sample sizes, relatively good precision and close matches between groups give us high confidence in these gas savings results.

We analyzed gas savings for each year individually (2015, 2016, and 2017) to see if there were any changes in savings occurring over time. We did not see a coherent time trend, but savings appeared to be much lower than average in 2015 and somewhat higher than average in 2016. However, these differences could easily be explained by variability in gas usage, lower sample sizes and lower precision. Results for 2016 and 2017 had larger sample sizes and were more robust than the 2015 results. They also more closely aligned with the overall gas savings estimate.

We were interested to see if there might be differences in gas savings between the two primary thermostat manufacturers supported through Energy Trust's programs: Nest and ecobee. However, the power of this analysis was limited by the uneven split between Nest and ecobee purchases. From 2015 to 2017, Nest thermostats accounted for 84% of installations in treatment homes and ecobee thermostats made up the remaining 16%. For Nest thermostats across heating zones and installation years, average gas savings in gas-heated homes ranged from 21 to 29 therms per year (+/- 8) or 3-4% of baseline gas usage. There were 775 Nest treatment homes analyzed, which had an average annual baseline gas usage of 713 therms. These results were very similar to the overall gas savings. For ecobee thermostats, average gas savings in gas-heated homes ranged from 36 to 45 therms (+/- 16) or 5-6% of baseline gas usage. There were 146 ecobee treatment homes analyzed, which had an average annual baseline gas usage of 723 therms. Although the ecobee savings results are notably higher than results for Nest, they are based on a much smaller sample size with lower precision. It is unclear whether these results will persist with a larger sample of homes.

We also analyzed electric savings for gas-heated homes, which result from reduced furnace fan runtime and summer cooling savings. Across all heating zones from 2015 to 2017, average electric savings ranged from 178 to 225 kWh per year (+/- 90 kWh) or 2-3% of baseline electricity usage. There were 450 homes available for this analysis with average annual baseline electricity usage of 8,675 kWh. The magnitude of these savings is relatively small, but statistically significant and moderately precise. The comparison groups provided fair matches to the treatment homes based on energy consumption and a decent point of comparison as similar homes that did not install smart thermostats. Thus, we have moderate confidence in the direction and magnitude of the electric savings even though the precision is somewhat lower than for the gas results.

Overall electric usage in electrically-heated homes across heating zones from 2015 to 2017 increased slightly after the installation of a smart thermostat, with average savings ranging from -72 to -317 kWh per year (+/- 428) or 1-3% of baseline electricity usage. There were only 77 treatment homes available for this analysis with average annual baseline electricity usage of 12,563 kWh. While these results indicate smart thermostats had an insignificant but slightly negative impact on energy use in electrically-heated homes, the sample size was very small given the expected level of savings. There was also higher variability in electricity usage in electrically-heated homes compared to gas-heated homes, resulting in very poor precision. The matched comparison group provided a good match on electricity consumption and geographic distribution; however, the future participant group provided a relatively poor match. This may indicate the future participant group provided a somewhat skewed point of comparison. In addition, the baseline energy usage of the treatment group was surprisingly low for homes with electric heat, indicating that these homes may be more energy efficient than average or do not exclusively heat with electricity. In either case, the savings estimate presented for this group may not be representative of the savings we would expect in a typical electrically-heated home. While these results are not encouraging for smart thermostats in electrically-heated homes, they are far from definitive.

In the table below, we summarize the results of the various smart thermostat analysis scenarios we looked at. Results are provided for kWh and therm savings for gas- and electrically-heated homes that installed thermostats from 2015 to 2017. For most analyses, we combined the two heating zones to preserve sample sizes. We present the midpoint savings estimate of the two comparison group methodologies (matched non-participants and future participants).

Table 1: Smart thermostat energy savings analysis summary of results

Fuel Analyzed	Heating Fuel	Heating Zone	Make	Years	N*	Baseline Energy Usage	Average Savings†	Absolute Precision†	Percent Savings†	Conf. Level
Therms	Gas	All	All	2015-2017	924	713	28	+/- 7	4%	High
Therms	Gas	1	All	2015-2017	878	704	29	+/- 7	4%	High
Therms	Gas	All	Nest	2015-2017	775	713	25	+/- 8	4%	Moderate
Therms	Gas	All	ecobee	2015-2017	146	723	40	+/- 16	6%	Moderate
Therms	Gas	All	All	2015	111	699	6	+/- 20	1%	Low
Therms	Gas	All	All	2016	374	705	40	+/- 10	6%	Moderate
Therms	Gas	All	All	2017	438	727	19	+/- 10	3%	Moderate
kWh	Gas	All	All	2015-2017	450	8,675	202	+/- 90	2%	Moderate
kWh	Electricity	All	All	2015-2017	77	12,563	-194	+/- 428	-2%	Low

* N is the final treatment group sample size in the analysis.

† The average savings, absolute precision and percent savings values represent the midpoint estimates between the two comparison group methodologies used.

These results confirm that smart thermostats continue to save a small percentage of energy in gas-heated Oregon homes beyond the pilot period. They also provide an early warning that electricity savings in electrically-heated homes may be lower than expected, although it is too soon to say for sure. Energy Trust will use the results from this and other Recurve analyses to update savings assumptions used in our standard residential measures when updates are made.

Impact Evaluation Report

Gas Impact of Thermostats in Program Year 2015, 2016, 2017

Result Summary

Measure: Thermostats	Program Year: 2015, 2016, 2017		Fuel: Gas	*Last Consumption Data Update:* October 1, 2019 *Last Participation Data Update:* October 1, 2019 *CalTRACK Version:* 2.0
Meter Data Filters:	DNAC: <75%	DNAC Percentile: None	Annual Consumption Percentile: Remove Top and Bottom 0.5%	
Model Filters:	Period Length: 11 Months or Longer	R-Squared: >0.5	CV(RMSE): < 1	
Metadata Filters:	Cooling Zone(s): All		Heating Fuel: Gas	
Thermostat Name: All / Heat Pump Baseline: All	Heating Zone(s): All		Multi Measure Filter: Single Measure Only	
Heat Pump Manufacturer: All / Heat Pump Adv. Controls or Commissioning: All				

924 Treatment Meters	17 +/- 7 Therms Average Normal Year Pre-Post Difference in Consumption per Participant	2 +/- 1 % Percent Normal Year Pre-Post Difference in Consumption per Participant	713 Mean Baseline Consumption (Gas)	53% Realization Rate
4,578 Site-level Matched Meters	24 +/- 7 Therms Average Savings Relative to Site-level Matched Comparison Group	3 +/- 1% Percent Savings Relative to Site-level Matched Comparison Group	704 Mean Baseline Consumption (Gas)	75% Realization Rate
10.5k Future Participant Meters	32 +/- 7 Therms Average Savings Relative to Future Participant Group	4 +/- 1% Savings Relative to Future Participant Group	707 Mean Baseline Consumption (Gas)	99% Realization Rate

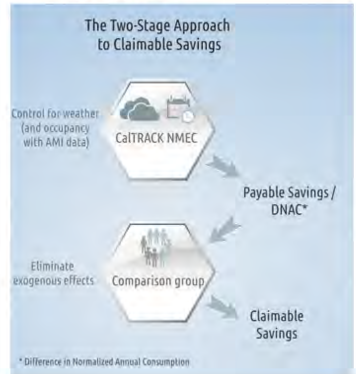
1. Introduction

This report contains the results of applying the two-stage approach (informed by the DOE's uniform methods chapter on whole building analysis) for calculating claimable savings to the selected portfolio of energy efficiency projects (see Figure). This approach begins with identification of two comparison groups for the treatment sample: (a) a site-level matched comparison group and (b) a future participant group. These groups are described below along with summary statistics (site locations, sample size, baseline consumption and baseline load disaggregation).

The CalTRACK methods are then applied to arrive at site-level savings, normalized for weather, and reflective of energy consumption changes for customers at the meter. Using a difference of differences for the treatment group with each comparison group accounts for population-level consumption changes (e.g. economic changes, rate changes, natural energy efficiency adoption etc.). The methods contained within this report are the outcome of a recent peer-reviewed study completed by Energy Trust of Oregon and Open Energy Efficiency (see "Methodology" section for more details).

The report includes the following sections:
Result Summary - Includes the overall portfolio results
Section 1. Introduction - Overview of report and the different groups included in the analysis
Section 2. Data Preparation - Data cleaning and sample attrition
Section 3. Modeling Results - CalTRACK model outputs and Difference in Normalized Annual Consumption (DNAC) results
Section 4. Methodology - Description of methods used in this report

Two-Stage Approach

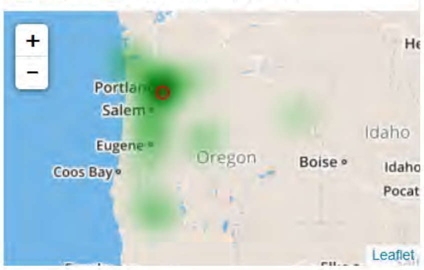


Treatment Group

The treatment group consists of sites that participated in the specified energy efficiency projects in the specified program year. Only sites that installed single measures are included in the treatment group. And this group includes the subset of sites that had sufficient data quality for modeling.

Treatment Site Locations

Projects · Treatment Group Centroid



38.3 miles

80% of projects lie within this distance from treatment group centroid

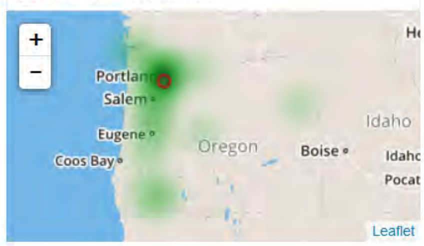
923	713
Meters	Mean Baseline Consumption (Gas)

Site-level Matched Comparison Group

This group includes comparison group sites that were matched at the site-level to treatment group sites. Each treatment group site is matched to five comparison group sites from the same zipcode, but only the sites with sufficient data quality were included in the group. Matching was performed using monthly consumption in the baseline period as detailed in the Methodology section.

Site-level Matched Site Locations

Comparison Group Centroid · Projects · Treatment Group Centroid



0.1 miles

Distance between treatment and comparison group centroids

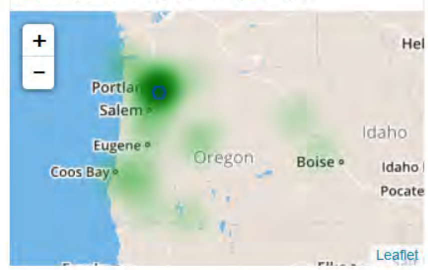
4,578	704
Meters	Mean Baseline Consumption (Gas)

Future Participant Group

The pool of sites that was used to create this group was composed of sites that installed the same measure in the year following the specified program year. The final sites were selected by stratified sampling using deciles of annual energy consumption.

Future Participant Site Locations

Comparison Group Centroid · Projects

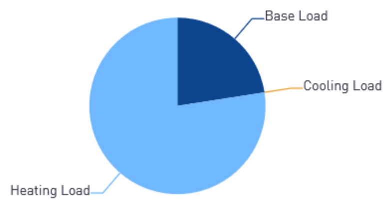


0.8 miles

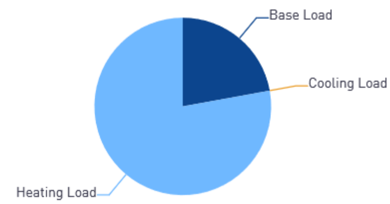
Distance between treatment and future participant group centroids

10.5k	707
Meters	Mean Baseline Consumption (Gas)

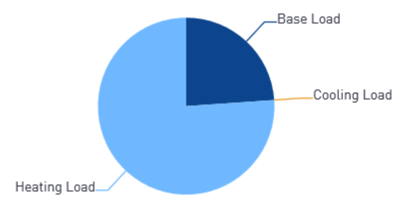
Load Disaggregation



Load Disaggregation



Load Disaggregation



2. Data Preparation

Consumption data preparation and cleaning followed best practices defined in the CalTRACK 2.0 billing methods. Some key aspects of the data cleaning process are highlighted here; please see the resources section for links to more detailed documentation. The initial and final sample sizes are shown below along with the percent of the treatment population that is represented by the sample. The sample attrition table shows the impact of each filtering criterion on sample size.

5,347	924	17%
Meters in Treatment Population	Final Sample Size	Percent of Treatment Population Represented by Sample

Sample Attrition Table

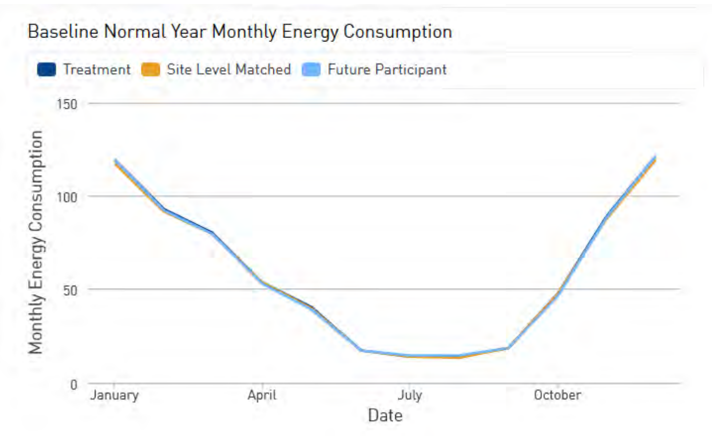
Filter	Selected Filter Value (if applicable)	Number of Dropped Meters	Sample Size after Applying Filter
Measure: Meters associated with a particular measure in program participation data. **Year:** Program year. **Fuel:** Type of metered fuel.	Measure: Thermostats -- Year: 2015, 2016, 2017 -- Fuel: Gas	--	5,347
Meters with valid consumption data in baseline and/or reporting periods.	--	148	5,199
MultiMeasure_Filter: Meters with single/multiple measure installations in baseline and/or reporting periods.	Multi Measure Filter: Single Measure Only	3,830	1,369
HeatingFuel: Meters with a valid heating fuel that corresponds to the selected filter value.	Heating Fuel: Gas	128	1,241
HeatingZone, CoolingZone: Meters in selected heating and/or cooling climate zones.	Heating Zone: All -- Cooling Zone: All	0	1,241
Other measure-specific filters.	--	0	1,241
PeriodLength_Threshold: Meters meeting a threshold number of months of valid consumption data.	Period Length: 11 Months or Longer	196	1,045
Meters with at least 5 site-level matched meters from the comparison group pool.	--	88	957
DNAC_Threshold: Meters with normalized change in annual energy consumption under a specified threshold.	DNAC: <75%	21	936

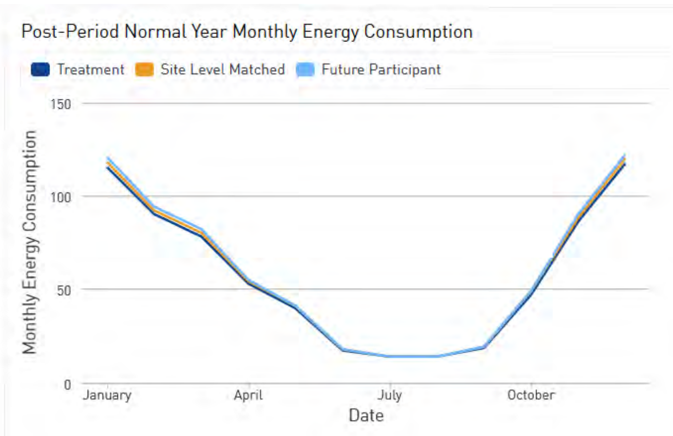
DNACPercentile_Threshold: Meters within specified percentile bands of normalized change in annual consumption.	DNAC Percentile: None	0	937
ConsumptionPercentile_Threshold: Meters within specified percentile bounds of annual energy consumption.	Annual Consumption Percentile: Remove Top and Bottom 0.5%	4	932
R2_Threshold: Meters with valid model R-squared for the baseline and reporting periods that meet a specified threshold. Models may have invalid R-squared due to data issues.	R-Squared: >0.5	9	923
CVRMSE_Threshold: Meters with valid model CV(RMSE) for the baseline and reporting periods that meet a specified threshold.	CV(RMSE): < 1	0	922

3. Modeling Results

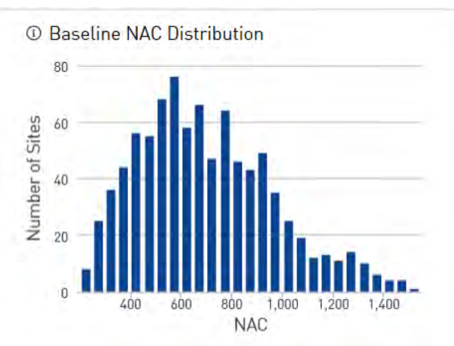
This section includes summaries of the Difference in Normalized Annual Consumption (DNAC) results for the treatment and comparison groups. The time series of monthly energy consumption illustrates the similarities and/or differences in energy consumption for the different groups in the baseline and reporting periods.

Below, you will find a breakdown of the DNAC results by group, showing the histograms of DNAC as well as the mean value expressed in raw units and as a percent of baseline annual consumption. Finally, the distribution of model types in the baseline and reporting periods are also provided as an additional layer of analysis.

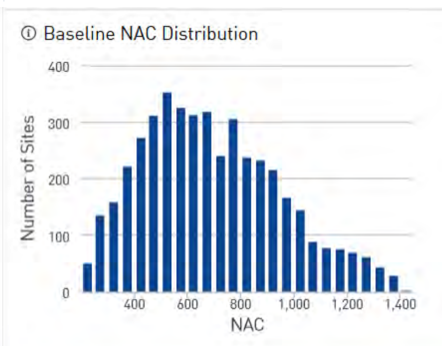




Treatment Group



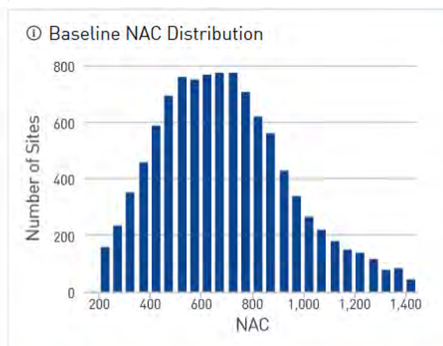
Site-level Matched Comparison Group



0.172

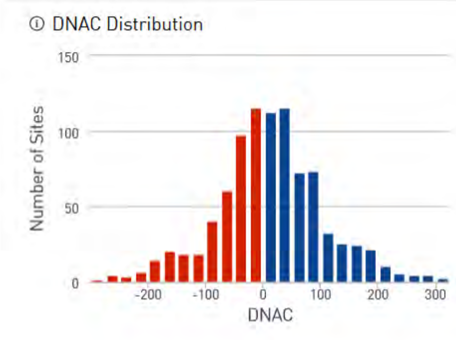
Annual Consumption p-value

Future Participant Group



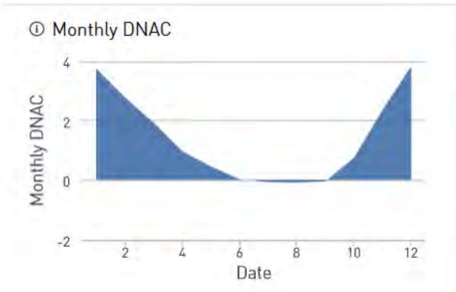
0.11

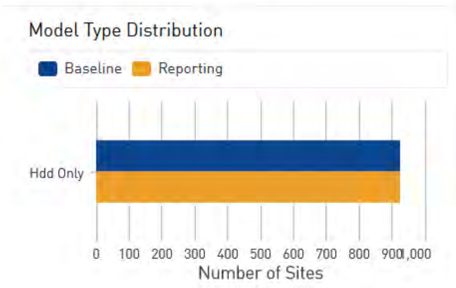
Annual Consumption p-value



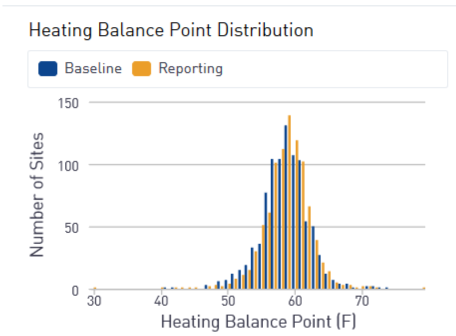
17 +/- 7 Therms
Average Difference in Normalized Annual Consumption per Participant

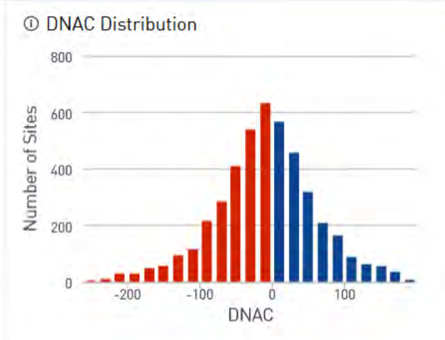
3 +/- 1 %
Difference in Normalized Annual Consumption as a Percent of Baseline





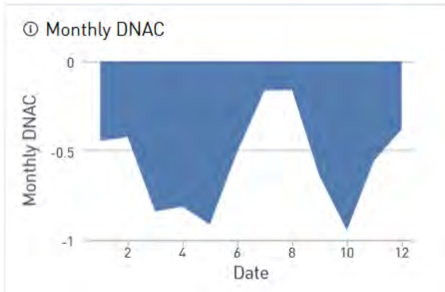
Treatment Group

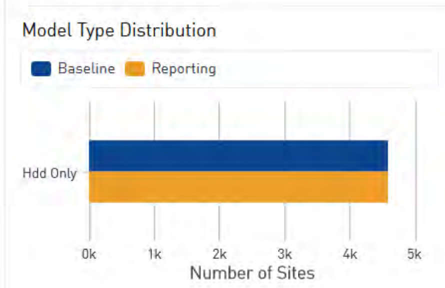




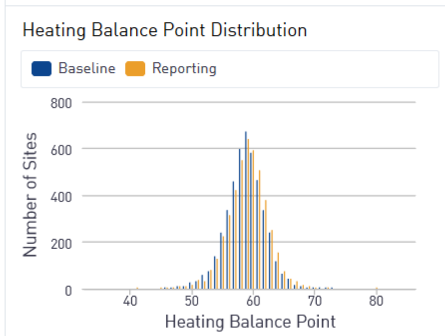
-7 +/- 2 Therms
Average Difference in Normalized Annual Consumption per Participant

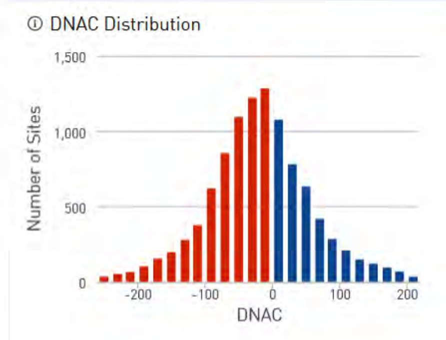
-1 +/- 0 %
Difference in Normalized Annual Consumption as a Percent of Baseline





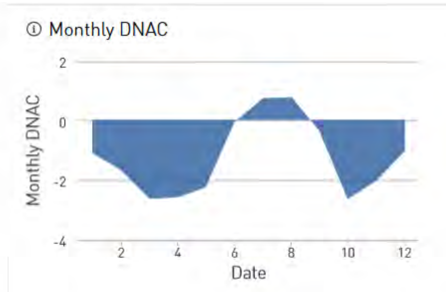
Site-level Matched Comparison Group

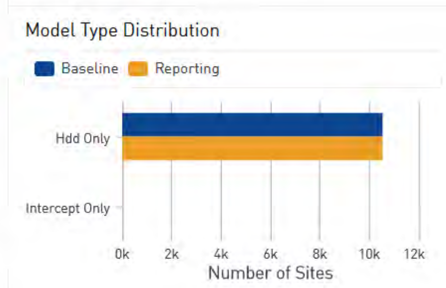




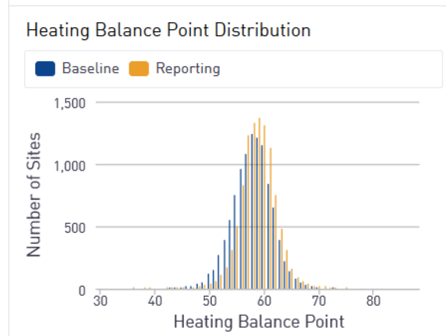
-15 +/- 2 Therms
Average Difference in Normalized Annual Consumption per Participant

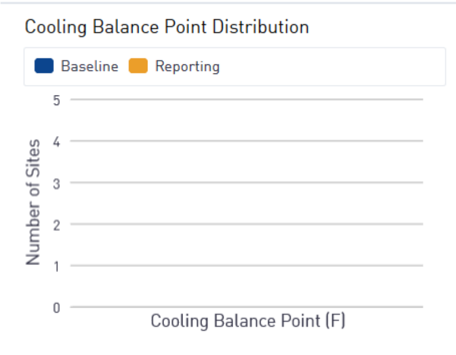
-2 +/- 0 %
Difference in Normalized Annual Consumption as a Percent of Baseline

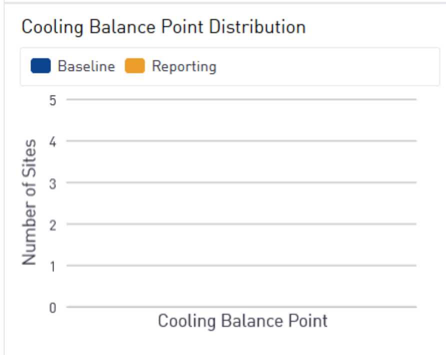


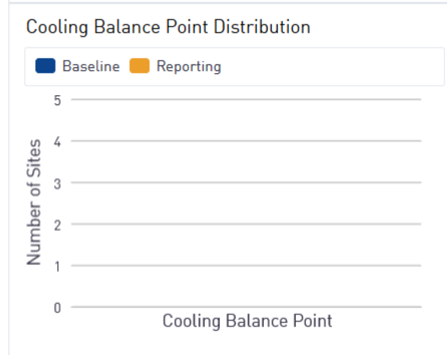


Future Participant Group









4. Methodology

CalTRACK and Comparison Group Methods
Documentation: docs.caltrack.org
Code: https://github.com/energy-market-methods/caltrack

Data Preparation

Baseline period: Since the predicted baseline may be unstable with different baseline period lengths, which may, in turn, affect calculated savings, the consensus of the CalTRACK 2.0 working group was to set the maximum baseline period at 12 months, since the year leading to the energy efficiency intervention is the most indicative of recent energy use trends and prolonging the baseline period increases the chance of other unmeasured factors affecting the baseline. In addition, CalTRACK uses a minimum 12-month baseline by default.

Blackout period: The blackout period refers to the time period between the end of the baseline period and the beginning of the reporting period. In this analysis, it is specified to coincide with the project installation time period, meaning that the billing period that contains the project installation date is dropped from the analysis.

Analysis periods: Different portions of the analysis used different time periods of consumption data, therefore, it is useful to clearly define these time periods and where they were used. Consider a project with an installation date on a particular day d in a particular month m in a particular program year y. The year before the program year is labelled as y-1, the year prior to that as y-2 and so on, while the years following the program year are labelled y+1, y+2 etc. In all cases, the billing period that contains the project installation was dropped from the analysis. Other sections of the analysis use the following time periods:
- *Treatment and site-level matched groups:* Baseline period includes the 12 months preceding the installation billing period. Reporting period includes the 12 months following the installation billing period.
- *Future participant group:* Baseline period is the calendar year preceding the program year (Year *y-1*). Reporting period is the program year itself (Year *y*).
- Site-level consumption matching was performed using the 12 months of data immediately prior to the project installation date.
- Equivalence tests were performed using data from the previous calendar year (y-1).

Modeling

Weather Normalization: Weather normalization of billing data in CalTRACK follows certain model foundations in literature (PRISM, ASHRAE Guideline 14, IPMVP Option C and the Uniform Methods Project for Whole Home Building Analysis). Building energy use is modeled as a combination of base load, heating load, and cooling load. Heating load and cooling load are assumed to have a linear relationship with heating and cooling demand, as approximated by heating and cooling degree days, beyond particular heating and cooling balance points. A number of candidate OLS models are fit to the consumption data using different combinations of heating and cooling balance points (ranging from 30 to 90 F) and different sets of independent variables. The model with the highest adjusted R-squared that contains strictly positive coefficients is selected as the final model and used to calculate normalized energy usage.

Model Types: CalTRACK specifies a linear relationship between energy use and temperature as reflected in the building consumption profile. In the most generic case, a model would include an intercept term, a heating balance point and heating slope coefficient, and a cooling balance point and a cooling slope coefficient. Depending on the fuel a building uses for heating or cooling or its consumption patterns, models with a single temperature coefficient and balance point (i.e., heating or cooling) may be more appropriate.

Difference in Normalized Annual Consumption (DNAC): The DNAC is calculated by using two CalTRACK regression models in conjunction with Typical Meteorological Year (TMY3) weather data, as follows:
- Two models are fit to the consumption data - one model for the baseline (pre-intervention) period and one for the reporting (post-intervention) period.
- Long-term heating and cooling degree days based on TMY3 data are substituted in both regression equations to calculate the Normalized Annual Consumption (NAC) for each period. TMY3 data is maintained by NREL and includes weather averages for 1020 locations in the US between 1991-2005.
- DNAC is determined by subtracting the two NACs (DNAC = Baseline NAC - Reporting NAC).

Disaggregation: Disaggregated loads are calculated from the different components of the statistical model fit. The weather sensitive components (heating and cooling load) are calculated by multiplying the relevant model coefficients (beta_hdd or beta_cdd) by the total degree days in a normal weather year (total HDD or CDD). For each site, the total HDD or CDD can be calculated using that site's estimated degree day balance points (also an output of the model) and the temperature for its closest weather station. The base load is estimated by multiplying the intercept of the statistical model by the number of days (365 for a full year).

Savings calculation: Savings are calculated by subtracting the DNAC for either comparison group from the DNAC for the treatment group.

Savings Uncertainty: Uncertainty presented in this analysis is calculated using the ASHRAE Guideline 14 formulation for aggregating the prediction uncertainty of point estimates in a time series. It is calculated at a 90% confidence level. The total uncertainty at the site-level is calculated using the sum of squares of the baseline and reporting models. Other aggregate uncertainty values (e.g. for a portfolio or for a difference-in-differences estimate) are also aggregated using the square root of the sum of squares.

Comparison Group Generation

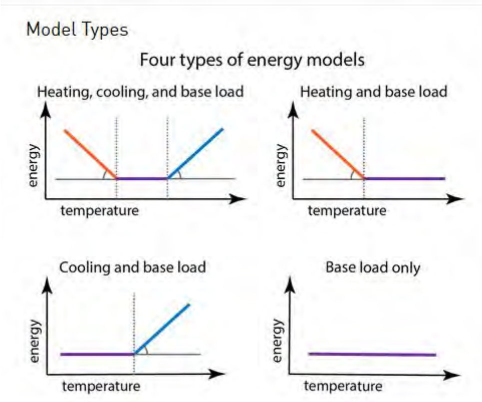
Site-level Matching: In monthly consumption matching, a comparison group is constructed by selecting n matches (n=5 in this analysis) from the comparison group pool with the shortest distance d to the treatment group customer under consideration. The pool is limited to non-participants within the same zipcode as the treatment group customer. The distance d is, in essence, a way to reduce 12 monthly consumption differences between any two customers to one metric (see Figure). In the present analysis, we selected (without replacement) five nearest neighbors for each treatment site based on the Euclidean distance of monthly consumption.

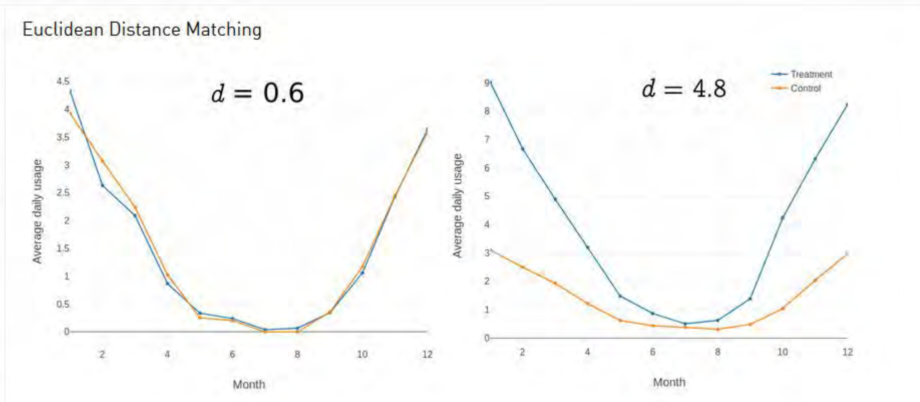
Future Participant Groups: Comparison groups comprising future participants are considered to be representative of participants in most aspects (observable and non-observable). For example, future participants are known to be eligible to receive the measure, and for some measures, they may have the same baseline equipment as the participants. Future participants have the same propensity to participate in the program as participants, thus reducing or eliminating self-selection bias, something that is otherwise difficult to control for in a quasi-experimental study. More comprehensive data is typically collected for future participants, allowing for potentially better matching and more insightful analysis. From a practical perspective, future participant groups may be difficult to construct for all measures, unless a program has been running for multiple years and is considered stable with sufficient data collection over the analysis period. Sample sizes for the comparison group may also be constrained if using future participants.

Stratified sampling is applied to future participant groups to attempt to replicate the distributions of the underlying variable (annual consumption) in the comparison group. Annual consumption of all treatment sites is first split into deciles, then a random sample is selected from within each corresponding bin in the comparison group pool of future participants.

Geographical screen: For the site-level matched group, only sites within the same zipcode as the treatment site were considered as potential comparison group matches.

Sampling method: In all cases where sampling was required from the comparison group, sampling was performed without replacement.





Impact Evaluation Report

Gas Impact of Thermostats in Program Year 2015, 2016, 2017

Result Summary

Measure: Thermostats	Program Year: 2017, 2015, 2016		Fuel: Gas	*Last Consumption Data Update:* October 1, 2019 *Last Participation Data Update:* October 1, 2019 *CalTRACK Version:* 2.0
Meter Data Filters:	DNAC: <75%	DNAC Percentile: None	Annual Consumption Percentile: Remove Top and Bottom 0.5%	
Model Filters:	Period Length: 11 Months or Longer	R-Squared: >0.5	CV(RMSE): < 1	
Metadata Filters:	Cooling Zone(s): All		Heating Fuel: Gas	
Thermostat Name: All / Heat Pump Baseline: All	Heating Zone(s): 1 - Hdd <= 6000		Multi Measure Filter: Single Measure Only	
Heat Pump Manufacturer: All / Heat Pump Adv. Controls or Commissioning: All				

878 Treatment Meters	17 +/- 7 Therms Average Normal Year Pre-Post Difference in Consumption per Participant	2 +/- 1 % Percent Normal Year Pre-Post Difference in Consumption per Participant	704 Mean Baseline Consumption (Gas)	53% Realization Rate
4,356 Site-level Matched Meters	26 +/- 7 Therms Average Savings Relative to Site-level Matched Comparison Group	4 +/- 1% Percent Savings Relative to Site-level Matched Comparison Group	694 Mean Baseline Consumption (Gas)	81% Realization Rate
10.1k Future Participant Meters	32 +/- 7 Therms Average Savings Relative to Future Participant Group	5 +/- 1% Savings Relative to Future Participant Group	701 Mean Baseline Consumption (Gas)	102% Realization Rate

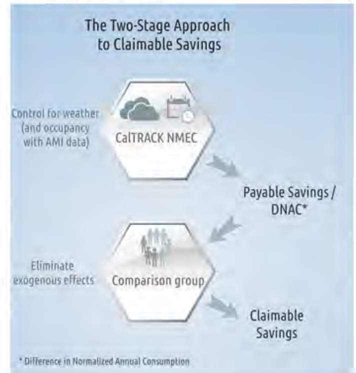
1. Introduction

This report contains the results of applying the two-stage approach (informed by the DOE's uniform methods chapter on whole building analysis) for calculating claimable savings to the selected portfolio of energy efficiency projects (see Figure). This approach begins with identification of two comparison groups for the treatment sample: (a) a site-level matched comparison group and (b) a future participant group. These groups are described below along with summary statistics (site locations, sample size, baseline consumption and baseline load disaggregation).

The CalTRACK methods are then applied to arrive at site-level savings, normalized for weather, and reflective of energy consumption changes for customers at the meter. Using a difference of differences for the treatment group with each comparison group accounts for population-level consumption changes (e.g. economic changes, rate changes, natural energy efficiency adoption etc.). The methods contained within this report are the outcome of a recent peer-reviewed study completed by Energy Trust of Oregon and Open Energy Efficiency (see "Methodology" section for more details).

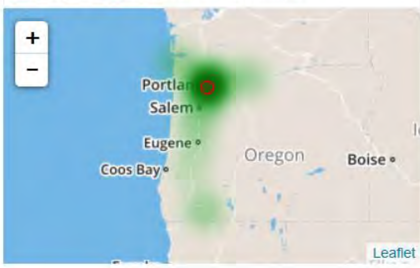
The report includes the following sections:
Result Summary - Includes the overall portfolio results
Section 1. Introduction - Overview of report and the different groups included in the analysis
Section 2. Data Preparation - Data cleaning and sample attrition
Section 3. Modeling Results - CalTRACK model outputs and Difference in Normalized Annual Consumption (DNAC) results
Section 4. Methodology - Description of methods used in this report

Two-Stage Approach



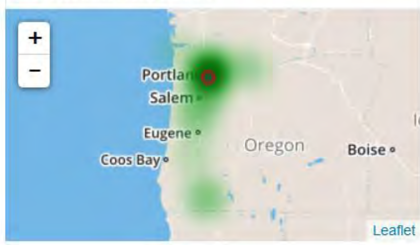
Treatment Site Locations

Projects   Treatment Group Centroid



Site-level Matched Site Locations

Comparison Group Centroid   Projects
Treatment Group Centroid



Future Participant Site Locations

Comparison Group Centroid   Projects
Treatment Group Centroid

28.6 miles
80% of projects lie within this distance from treatment group centroid

878
Meters

704
Mean Baseline Consumption (Gas)

0.2 miles
Distance between treatment and comparison group centroids

4,356
Meters

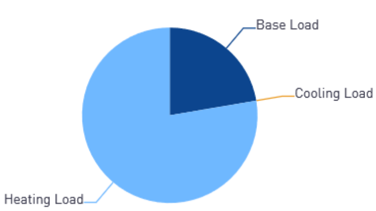
694
Mean Baseline Consumption (Gas)

0.9 miles
Distance between treatment and future participant group centroids

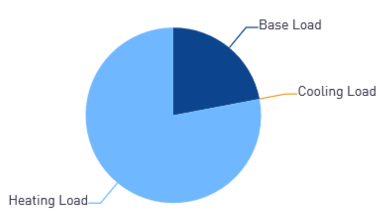
10.1k
Meters

701
Mean Baseline Consumption (Gas)

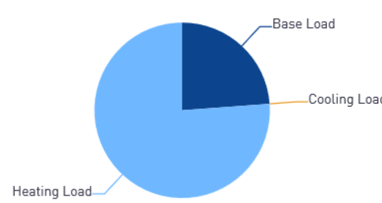
Load Disaggregation



Load Disaggregation



Load Disaggregation



2. Data Preparation

Consumption data preparation and cleaning followed best practices defined in the CalTRACK 2.0 billing methods. Some key aspects of the data cleaning process are highlighted here; please see the resources section for links to more detailed documentation. The initial and final sample sizes are shown below along with the percent of the treatment population that is represented by the sample. The sample attrition table shows the impact of each filtering criterion on sample size.

5,347	878	16%
Meters in Treatment Population	Final Sample Size	Percent of Treatment Population Represented by Sample

Sample Attrition Table

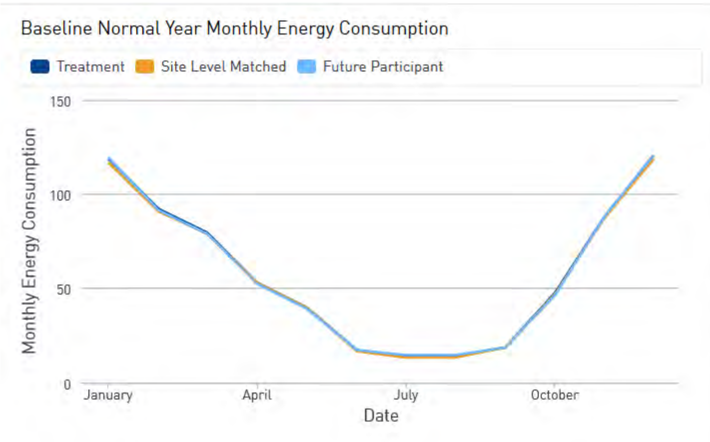
Filter	Selected Filter Value (if applicable)	Number of Dropped Meters	Sample Size after Applying Filter
Measure: Meters associated with a particular measure in program participation data. **Year:** Program year. **Fuel:** Type of metered fuel.	Measure: Thermostats -- Year: 2015, 2016, 2017 -- Fuel: Gas	--	5,347
Meters with valid consumption data in baseline and/or reporting periods.	--	148	5,199
MultiMeasure_Filter: Meters with single/multiple measure installations in baseline and/or reporting periods.	Multi Measure Filter: Single Measure Only	3,830	1,369
HeatingFuel: Meters with a valid heating fuel that corresponds to the selected filter value.	Heating Fuel: Gas	128	1,241
HeatingZone, CoolingZone: Meters in selected heating and/or cooling climate zones.	Heating Zone: 1 - Hdd <= 6000 -- Cooling Zone: All	59	1,182
Other measure-specific filters.	--	0	1,182
PeriodLength_Threshold: Meters meeting a threshold number of months of valid consumption data.	Period Length: 11 Months or Longer	183	999
Meters with at least 5 site-level matched meters from the comparison group pool.	--	87	912
DNAC_Threshold: Meters with normalized change in annual energy consumption under a specified threshold.	DNAC: <75%	20	892

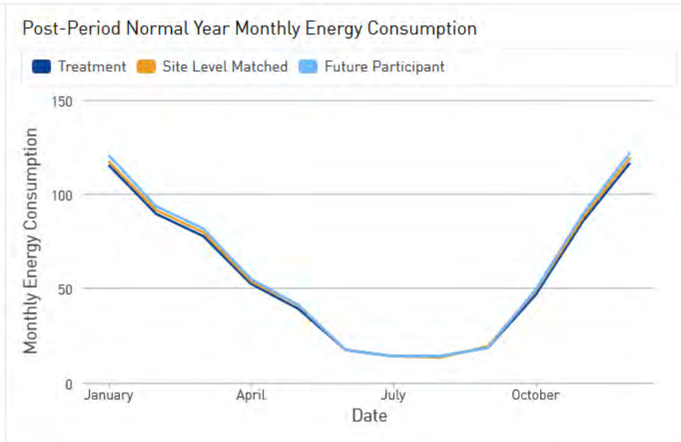
DNACPercentile_Threshold: Meters within specified percentile bands of normalized change in annual consumption.	DNAC Percentile: None	0	892
ConsumptionPercentile_Threshold: Meters within specified percentile bounds of annual energy consumption.	Annual Consumption Percentile: Remove Top and Bottom 0.5%	4	888
R2_Threshold: Meters with valid model R-squared for the baseline and reporting periods that meet a specified threshold. Models may have invalid R-squared due to data issues.	R-Squared: >0.5	10	878
CVRMSE_Threshold: Meters with valid model CV(RMSE) for the baseline and reporting periods that meet a specified threshold.	CV(RMSE): < 1	0	878

3. Modeling Results

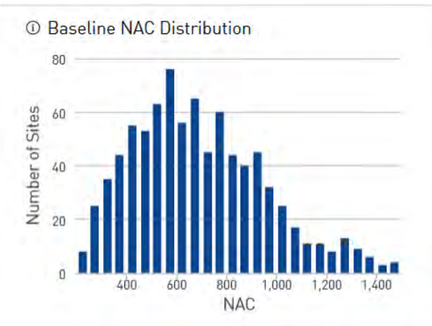
This section includes summaries of the Difference in Normalized Annual Consumption (DNAC) results for the treatment and comparison groups. The time series of monthly energy consumption illustrates the similarities and/or differences in energy consumption for the different groups in the baseline and reporting periods.

Below, you will find a breakdown of the DNAC results by group, showing the histograms of DNAC as well as the mean value expressed in raw units and as a percent of baseline annual consumption. Finally, the distribution of model types in the baseline and reporting periods are also provided as an additional layer of analysis.

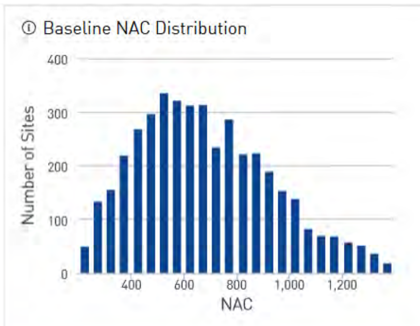




Treatment Group



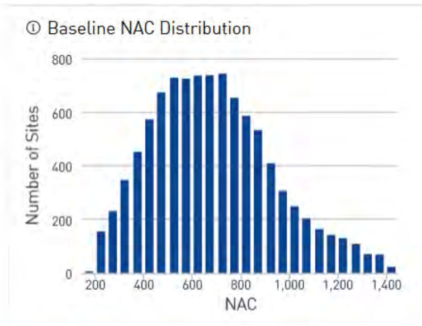
Site-level Matched Comparison Group



0.148

Annual Consumption p-value

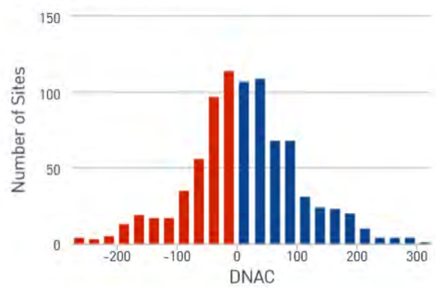
Future Participant Group



0.116

Annual Consumption p-value

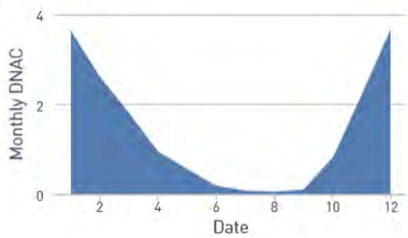
ⓘ DNAC Distribution



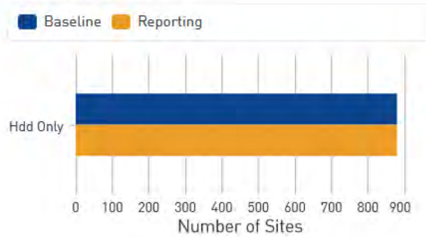
17 +/- 7 Therms
Average Difference in Normalized Annual Consumption per Participant

2 +/- 1 %
Difference in Normalized Annual Consumption as a Percent of Baseline

ⓘ Monthly DNAC

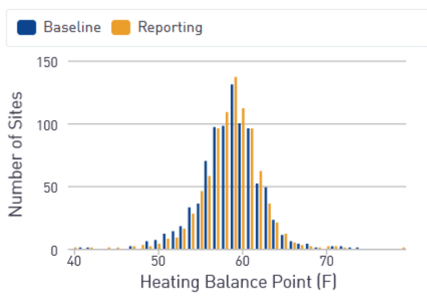


Model Type Distribution

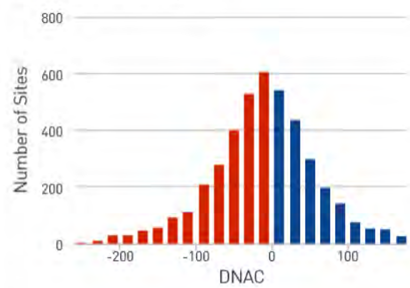


Treatment Group

Heating Balance Point Distribution



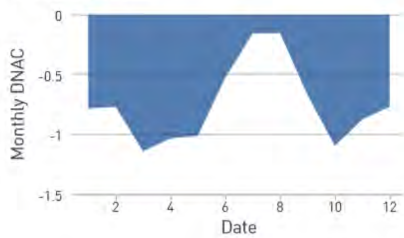
ⓘ DNAC Distribution



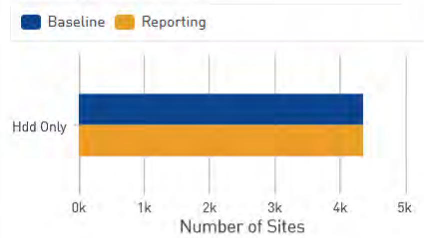
-9 +/- 2 Therms
Average Difference in Normalized Annual Consumption per Participant

-1 +/- 0 %
Difference in Normalized Annual Consumption as a Percent of Baseline

ⓘ Monthly DNAC

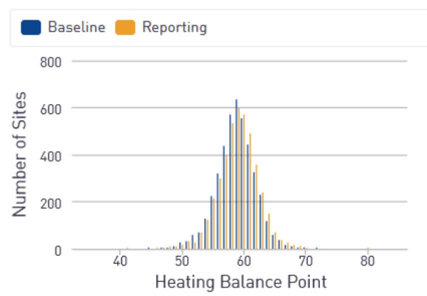


Model Type Distribution

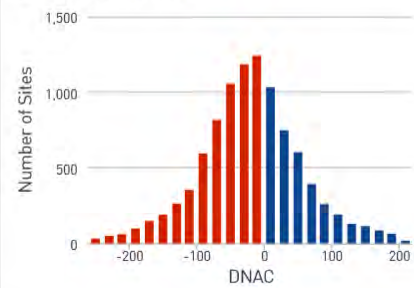


Site-level Matched Comparison Group

Heating Balance Point Distribution



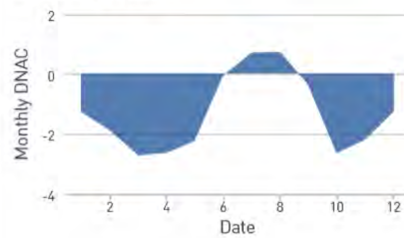
ⓘ DNAC Distribution



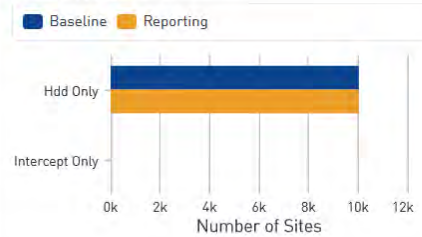
-16 +/- 2 Therms
Average Difference in Normalized Annual Consumption per Participant

-2 +/- 0 %
Difference in Normalized Annual Consumption as a Percent of Baseline

ⓘ Monthly DNAC

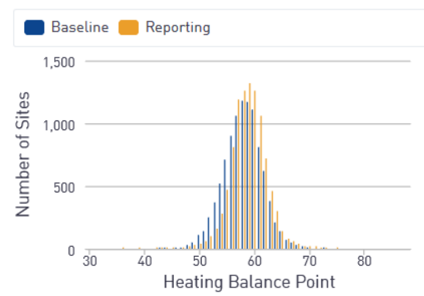


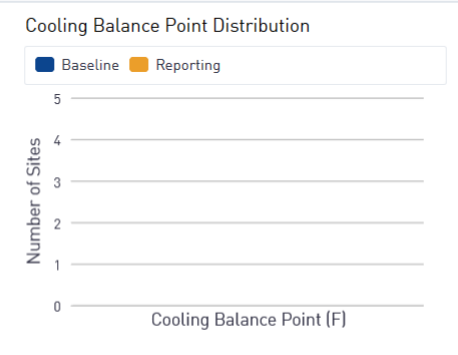
Model Type Distribution

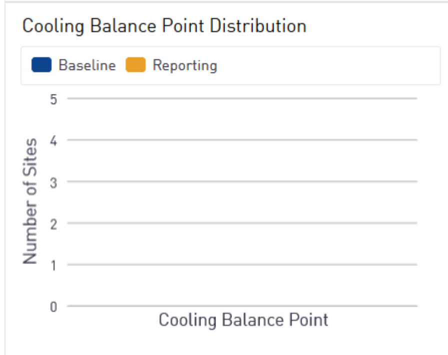


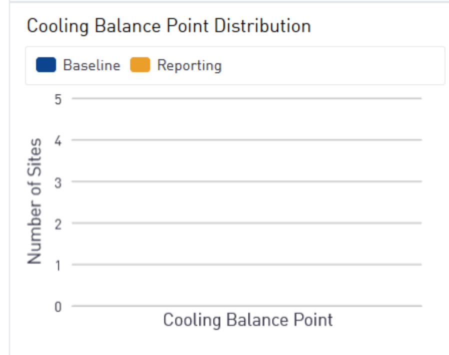
Future Participant Group

Heating Balance Point Distribution









4. Methodology

CalTRACK and Comparison Group Methods
Documentation: docs.caltrack.org
Code: https://github.com/energy-market-methods/caltrack

Data Preparation

Baseline period: Since the predicted baseline may be unstable with different baseline period lengths, which may, in turn, affect calculated savings, the consensus of the CalTRACK 2.0 working group was to set the maximum baseline period at 12 months, since the year leading to the energy efficiency intervention is the most indicative of recent energy use trends and prolonging the baseline period increases the chance of other unmeasured factors affecting the baseline. In addition, CalTRACK uses a minimum 12-month baseline by default.

Blackout period: The blackout period refers to the time period between the end of the baseline period and the beginning of the reporting period. In this analysis, it is specified to coincide with the project installation time period, meaning that the billing period that contains the project installation date is dropped from the analysis.

Analysis periods: Different portions of the analysis used different time periods of consumption data, therefore, it is useful to clearly define these time periods and where they were used. Consider a project with an installation date on a particular day d in a particular month m in a particular program year y. The year before the program year is labelled as y-1, the year prior to that as y-2 and so on, while the years following the program year are labelled y+1, y+2 etc. In all cases, the billing period that contains the project installation was dropped from the analysis. Other sections of the analysis use the following time periods:
- *Treatment and site-level matched groups:* Baseline period includes the 12 months preceding the installation billing period. Reporting period includes the 12 months following the installation billing period.
- *Future participant group:* Baseline period is the calendar year preceding the program year (Year *y-1*). Reporting period is the program year itself (Year *y*).
- Site-level consumption matching was performed using the 12 months of data immediately prior to the project installation date.
- Equivalence tests were performed using data from the previous calendar year (y-1).

Modeling

Weather Normalization: Weather normalization of billing data in CalTRACK follows certain model foundations in literature (PRISM, ASHRAE Guideline 14, IPMVP Option C and the Uniform Methods Project for Whole Home Building Analysis). Building energy use is modeled as a combination of base load, heating load, and cooling load. Heating load and cooling load are assumed to have a linear relationship with heating and cooling demand, as approximated by heating and cooling degree days, beyond particular heating and cooling balance points. A number of candidate OLS models are fit to the consumption data using different combinations of heating and cooling balance points (ranging from 30 to 90 F) and different sets of independent variables. The model with the highest adjusted R-squared that contains strictly positive coefficients is selected as the final model and used to calculate normalized energy usage.

Model Types: CalTRACK specifies a linear relationship between energy use and temperature as reflected in the building consumption profile. In the most generic case, a model would include an intercept term, a heating balance point and heating slope coefficient, and a cooling balance point and a cooling slope coefficient. Depending on the fuel a building uses for heating or cooling or its consumption patterns, models with a single temperature coefficient and balance point (i.e., heating or cooling) may be more appropriate.

Difference in Normalized Annual Consumption (DNAC): The DNAC is calculated by using two CalTRACK regression models in conjunction with Typical Meteorological Year (TMY3) weather data, as follows:
- Two models are fit to the consumption data - one model for the baseline (pre-intervention) period and one for the reporting (post-intervention) period.
- Long-term heating and cooling degree days based on TMY3 data are substituted in both regression equations to calculate the Normalized Annual Consumption (NAC) for each period. TMY3 data is maintained by NREL and includes weather averages for 1020 locations in the US between 1991-2005.
- DNAC is determined by subtracting the two NACs (DNAC = Baseline NAC - Reporting NAC).

Disaggregation: Disaggregated loads are calculated from the different components of the statistical model fit. The weather sensitive components (heating and cooling load) are calculated by multiplying the relevant model coefficients (beta_hdd or beta_cdd) by the total degree days in a normal weather year (total HDD or CDD). For each site, the total HDD or CDD can be calculated using that site's estimated degree day balance points (also an output of the model) and the temperature for its closest weather station. The base load is estimated by multiplying the intercept of the statistical model by the number of days (365 for a full year).

Savings calculation: Savings are calculated by subtracting the DNAC for either comparison group from the DNAC for the treatment group.

Savings Uncertainty: Uncertainty presented in this analysis is calculated using the ASHRAE Guideline 14 formulation for aggregating the prediction uncertainty of point estimates in a time series. It is calculated at a 90% confidence level. The total uncertainty at the site-level is calculated using the sum of squares of the baseline and reporting models. Other aggregate uncertainty values (e.g. for a portfolio or for a difference-in-differences estimate) are also aggregated using the square root of the sum of squares.

Comparison Group Generation

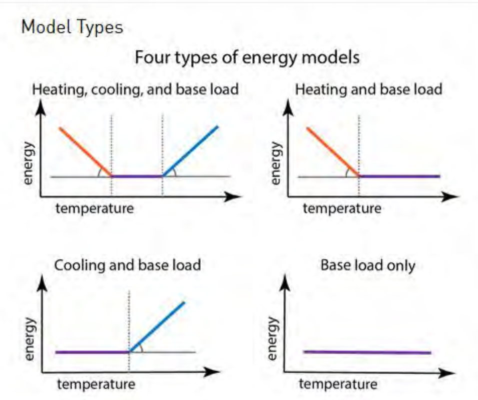
Site-level Matching: In monthly consumption matching, a comparison group is constructed by selecting n matches (n=5 in this analysis) from the comparison group pool with the shortest distance d to the treatment group customer under consideration. The pool is limited to non-participants within the same zipcode as the treatment group customer. The distance d is, in essence, a way to reduce 12 monthly consumption differences between any two customers to one metric (see Figure). In the present analysis, we selected (without replacement) five nearest neighbors for each treatment site based on the Euclidean distance of monthly consumption.

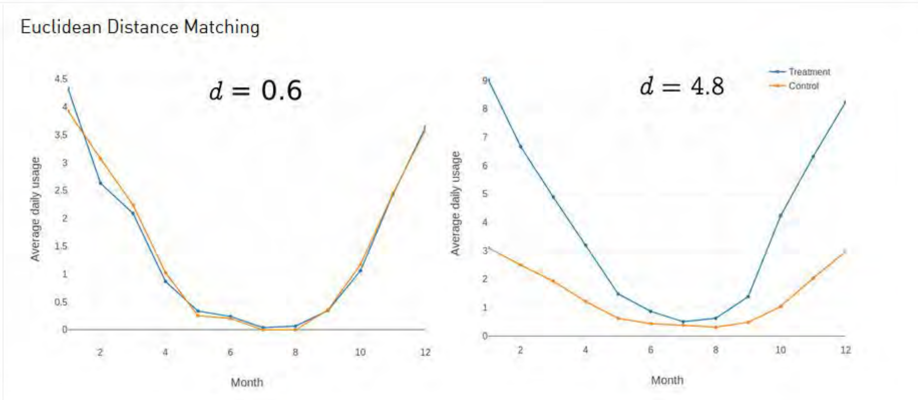
Future Participant Groups: Comparison groups comprising future participants are considered to be representative of participants in most aspects (observable and non-observable). For example, future participants are known to be eligible to receive the measure, and for some measures, they may have the same baseline equipment as the participants. Future participants have the same propensity to participate in the program as participants, thus reducing or eliminating self-selection bias, something that is otherwise difficult to control for in a quasi-experimental study. More comprehensive data is typically collected for future participants, allowing for potentially better matching and more insightful analysis. From a practical perspective, future participant groups may be difficult to construct for all measures, unless a program has been running for multiple years and is considered stable with sufficient data collection over the analysis period. Sample sizes for the comparison group may also be constrained if using future participants.

Stratified sampling is applied to future participant groups to attempt to replicate the distributions of the underlying variable (annual consumption) in the comparison group. Annual consumption of all treatment sites is first split into deciles, then a random sample is selected from within each corresponding bin in the comparison group pool of future participants.

Geographical screen: For the site-level matched group, only sites within the same zipcode as the treatment site were considered as potential comparison group matches.

Sampling method: In all cases where sampling was required from the comparison group, sampling was performed without replacement.





Impact Evaluation Report

Gas Impact of Thermostats in Program Year 2015

Result Summary

Measure: Thermostats	Program Year: 2015		Fuel: Gas	*Last Consumption Data Update:* October 1, 2019 *Last Participation Data Update:* October 1, 2019 *CalTRACK Version:* 2.0
Meter Data Filters:	DNAC: <75%	DNAC Percentile: None	Annual Consumption Percentile: Remove Top and Bottom 0.5%	
Model Filters:	Period Length: 11 Months or Longer	R-Squared: >0.5	CV(RMSE): < 1	
Metadata Filters:	Cooling Zone(s): All		Heating Fuel: Gas	
Thermostat Name: All / Heat Pump Baseline: All	Heating Zone(s): All		Multi Measure Filter: Single Measure Only	
Heat Pump Manufacturer: All / Heat Pump Adv. Controls or Commissioning: All				

111 Treatment Meters	-29 +/- 20 Therms Average Normal Year Pre-Post Difference in Consumption per Participant	-4 +/- 3 % Percent Normal Year Pre-Post Difference in Consumption per Participant	699 Mean Baseline Consumption (Gas)	-91% Realization Rate
554 Site-level Matched Meters	11 +/- 21 Therms Average Savings Relative to Site-level Matched Comparison Group	2 +/- 3% Percent Savings Relative to Site-level Matched Comparison Group	693 Mean Baseline Consumption (Gas)	35% Realization Rate
3,133 Future Participant Meters	-5 +/- 20 Therms Average Savings Relative to Future Participant Group	-1 +/- 3% Savings Relative to Future Participant Group	690 Mean Baseline Consumption (Gas)	-18% Realization Rate

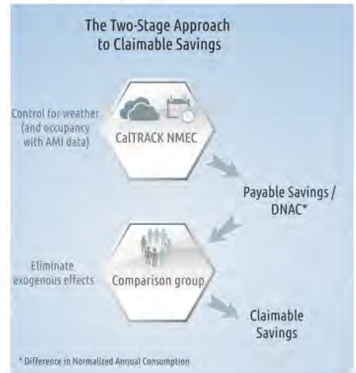
1. Introduction

This report contains the results of applying the two-stage approach (informed by the DOE's uniform methods chapter on whole building analysis) for calculating claimable savings to the selected portfolio of energy efficiency projects (see Figure). This approach begins with identification of two comparison groups for the treatment sample: (a) a site-level matched comparison group and (b) a future participant group. These groups are described below along with summary statistics (site locations, sample size, baseline consumption and baseline load disaggregation).

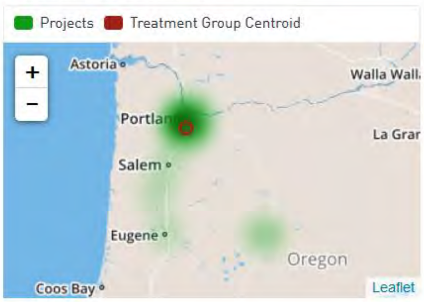
The CalTRACK methods are then applied to arrive at site-level savings, normalized for weather, and reflective of energy consumption changes for customers at the meter. Using a difference of differences for the treatment group with each comparison group accounts for population-level consumption changes (e.g. economic changes, rate changes, natural energy efficiency adoption etc.). The methods contained within this report are the outcome of a recent peer-reviewed study completed by Energy Trust of Oregon and Open Energy Efficiency (see "Methodology" section for more details).

The report includes the following sections:
Result Summary - Includes the overall portfolio results
Section 1. Introduction - Overview of report and the different groups included in the analysis
Section 2. Data Preparation - Data cleaning and sample attrition
Section 3. Modeling Results - CalTRACK model outputs and Difference in Normalized Annual Consumption (DNAC) results
Section 4. Methodology - Description of methods used in this report

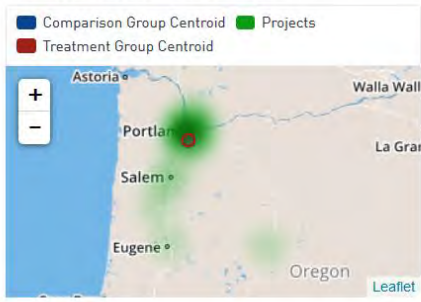
Two-Stage Approach



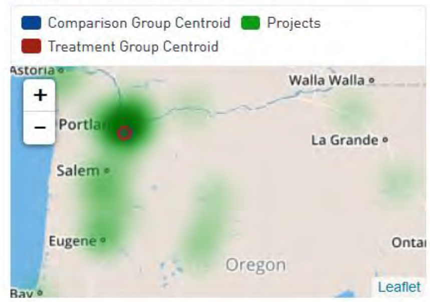
Treatment Site Locations



Site-level Matched Site Locations



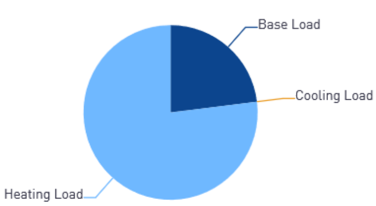
Future Participant Site Locations



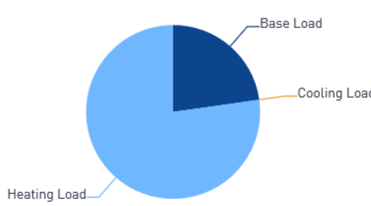
21.5 miles	0.1 miles	1.7 miles
80% of projects lie within this distance from treatment group centroid	Distance between treatment and comparison group centroids	Distance between treatment and future participant group centroids

111	699	554	693	3,135	690
Meters	Mean Baseline Consumption (Gas)	Meters	Mean Baseline Consumption (Gas)	Meters	Mean Baseline Consumption (Gas)

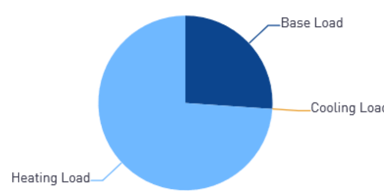
Load Disaggregation



Load Disaggregation



Load Disaggregation



2. Data Preparation

Consumption data preparation and cleaning followed best practices defined in the CalTRACK 2.0 billing methods. Some key aspects of the data cleaning process are highlighted here; please see the resources section for links to more detailed documentation. The initial and final sample sizes are shown below along with the percent of the treatment population that is represented by the sample. The sample attrition table shows the impact of each filtering criterion on sample size.

488	111	23%
Meters in Treatment Population	Final Sample Size	Percent of Treatment Population Represented by Sample

Sample Attrition Table

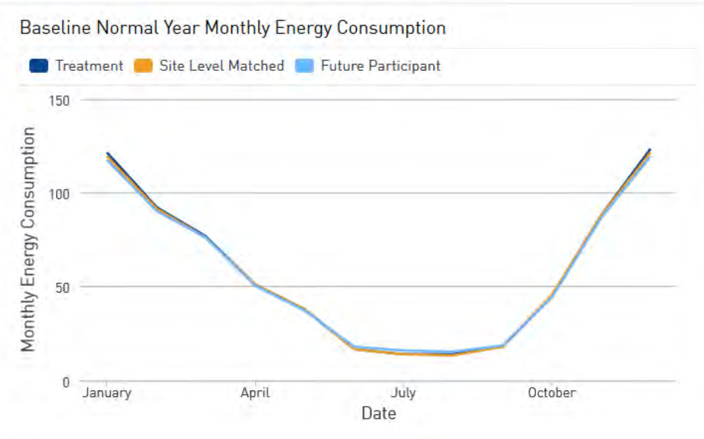
Filter	Selected Filter Value (if applicable)	Number of Dropped Meters	Sample Size after Applying Filter
Measure: Meters associated with a particular measure in program participation data. **Year:** Program year. **Fuel:** Type of metered fuel.	Measure: Thermostats -- Year: 2015 -- Fuel: Gas	--	488
Meters with valid consumption data in baseline and/or reporting periods.	--	15	473
MultiMeasure_Filter: Meters with single/multiple measure installations in baseline and/or reporting periods.	Multi Measure Filter: Single Measure Only	334	139
HeatingFuel: Meters with a valid heating fuel that corresponds to the selected filter value.	Heating Fuel: Gas	8	131
HeatingZone, CoolingZone: Meters in selected heating and/or cooling climate zones.	Heating Zone: All -- Cooling Zone: All	0	131
Other measure-specific filters.	--	0	131
PeriodLength_Threshold: Meters meeting a threshold number of months of valid consumption data.	Period Length: 11 Months or Longer	14	117
Meters with at least 5 site-level matched meters from the comparison group pool.	--	3	114
DNAC_Threshold: Meters with normalized change in annual energy consumption under a specified threshold.	DNAC: <75%	2	112

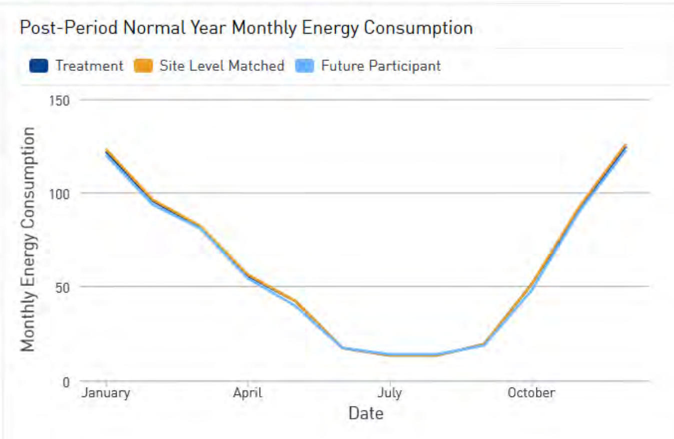
DNACPercentile_Threshold: Meters within specified percentile bands of normalized change in annual consumption.	DNAC Percentile: All	0	112
ConsumptionPercentile_Threshold: Meters within specified percentile bounds of annual energy consumption.	Annual Consumption Percentile: Remove Top and Bottom 0.5%	0	112
R2_Threshold: Meters with valid model R-squared for the baseline and reporting periods that meet a specified threshold. Models may have invalid R-squared due to data issues.	R-Squared: >0.5	1	111
CVRMSE_Threshold: Meters with valid model CV(RMSE) for the baseline and reporting periods that meet a specified threshold.	CV(RMSE): < 1	0	111

3. Modeling Results

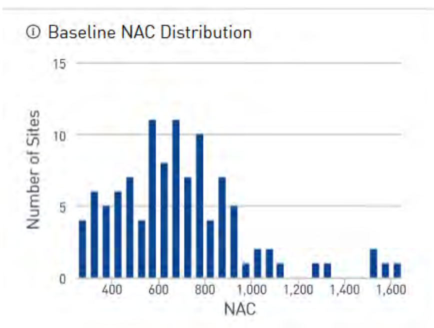
This section includes summaries of the Difference in Normalized Annual Consumption (DNAC) results for the treatment and comparison groups. The time series of monthly energy consumption illustrates the similarities and/or differences in energy consumption for the different groups in the baseline and reporting periods.

Below, you will find a breakdown of the DNAC results by group, showing the histograms of DNAC as well as the mean value expressed in raw units and as a percent of baseline annual consumption. Finally, the distribution of model types in the baseline and reporting periods are also provided as an additional layer of analysis.

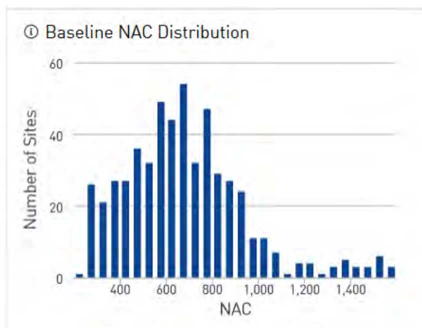




Treatment Group



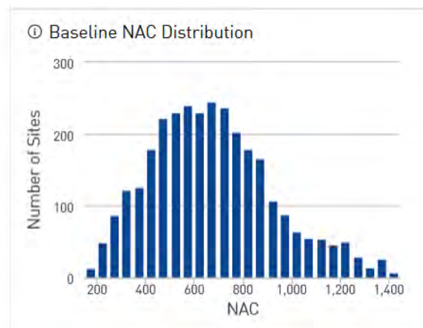
Site-level Matched Comparison Group



0.266

Annual Consumption p-value

Future Participant Group

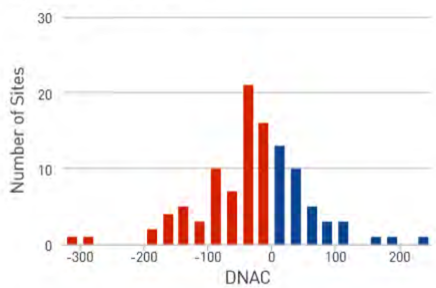


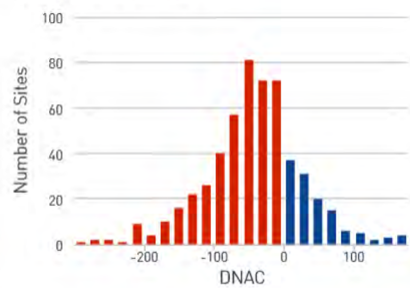
0.171

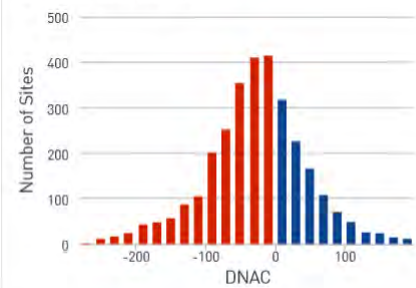
Annual Consumption p-value

Treatment Group	Site-level Matched Comparison Group	Future Participant Group
-29 +/- 20 Therms — Average Difference in Normalized Annual Consumption per Participant	-40 +/- 8 Therms — Average Difference in Normalized Annual Consumption per Participant	-23 +/- 3 Therms — Average Difference in Normalized Annual Consumption per Participant
-4 +/- 3 % — Difference in Normalized Annual Consumption as a Percent of Baseline	-6 +/- 1 % — Difference in Normalized Annual Consumption as a Percent of Baseline	-3 +/- 0 % — Difference in Normalized Annual Consumption as a Percent of Baseline

DNAC Distribution







-29 +/- 20 Therms

Average Difference in Normalized Annual Consumption per Participant

-4 +/- 3 %

Difference in Normalized Annual Consumption as a Percent of Baseline

-40 +/- 8 Therms

Average Difference in Normalized Annual Consumption per Participant

-6 +/- 1 %

Difference in Normalized Annual Consumption as a Percent of Baseline

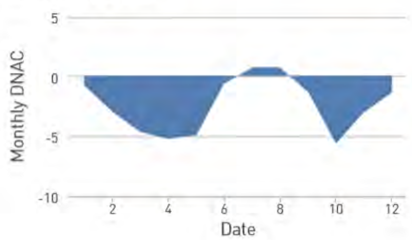
-23 +/- 3 Therms

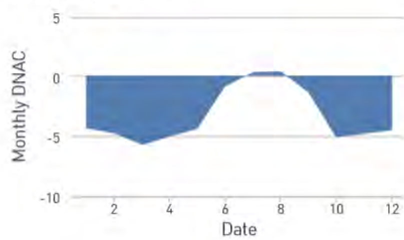
Average Difference in Normalized Annual Consumption per Participant

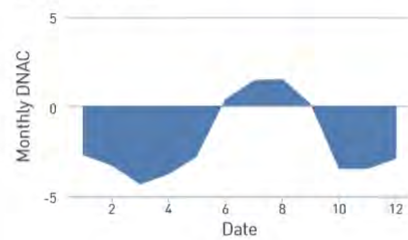
-3 +/- 0 %

Difference in Normalized Annual Consumption as a Percent of Baseline

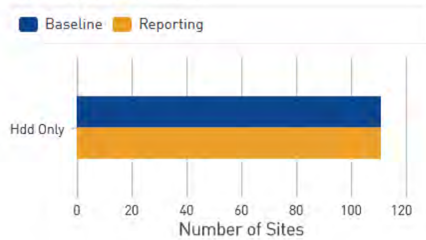
Monthly DNAC

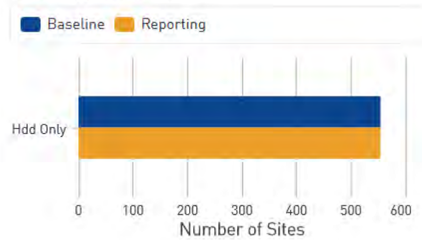


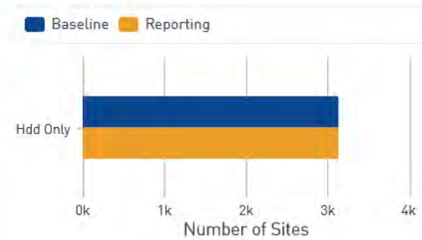




Model Type Distribution





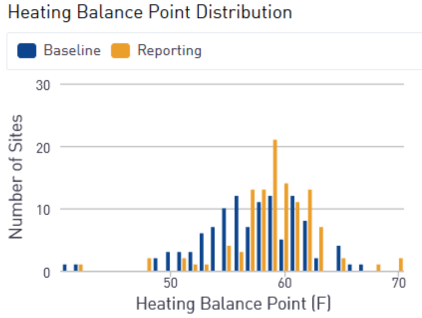


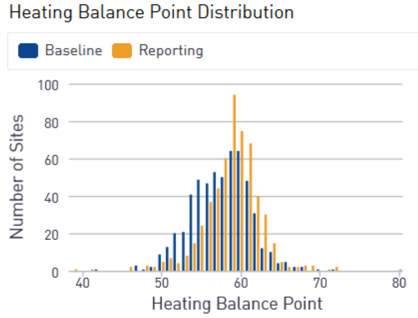
Treatment Group

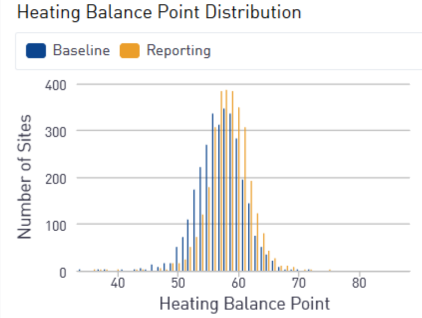
Site-level Matched Comparison Group

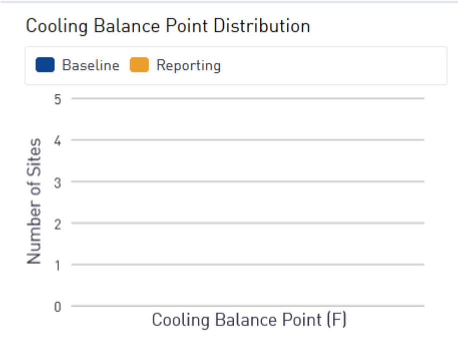
Future Participant Group

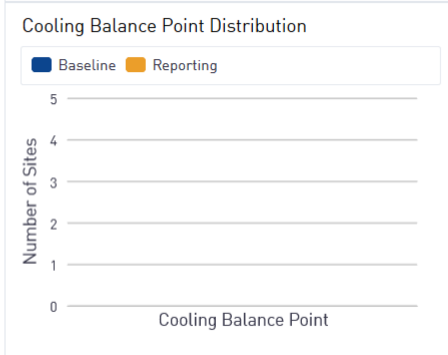
Heating Balance Point Distribution

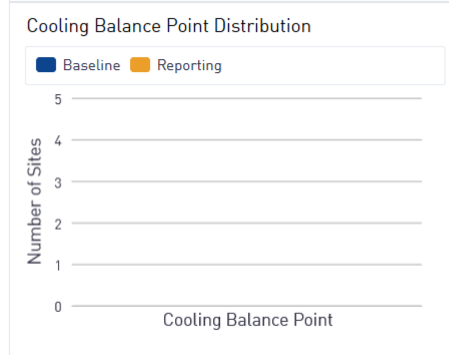












4. Methodology

CalTRACK and Comparison Group Methods
Documentation: docs.caltrack.org
Code: https://github.com/energy-market-methods/caltrack

Data Preparation

Baseline period: Since the predicted baseline may be unstable with different baseline period lengths, which may, in turn, affect calculated savings, the consensus of the CalTRACK 2.0 working group was to set the maximum baseline period at 12 months, since the year leading to the energy efficiency intervention is the most indicative of recent energy use trends and prolonging the baseline period increases the chance of other unmeasured factors affecting the baseline. In addition, CalTRACK uses a minimum 12-month baseline by default.

Blackout period: The blackout period refers to the time period between the end of the baseline period and the beginning of the reporting period. In this analysis, it is specified to coincide with the project installation time period, meaning that the billing period that contains the project installation date is dropped from the analysis.

Analysis periods: Different portions of the analysis used different time periods of consumption data, therefore, it is useful to clearly define these time periods and where they were used. Consider a project with an installation date on a particular day d in a particular month m in a particular program year y. The year before the program year is labelled as y-1, the year prior to that as y-2 and so on, while the years following the program year are labelled y+1, y+2 etc. In all cases, the billing period that contains the project installation was dropped from the analysis. Other sections of the analysis use the following time periods:
- *Treatment and site-level matched groups:* Baseline period includes the 12 months preceding the installation billing period. Reporting period includes the 12 months following the installation billing period.
- *Future participant group:* Baseline period is the calendar year preceding the program year (Year *y-1*). Reporting period is the program year itself (Year *y*).
- Site-level consumption matching was performed using the 12 months of data immediately prior to the project installation date.
- Equivalence tests were performed using data from the previous calendar year (y-1).

Modeling

Weather Normalization: Weather normalization of billing data in CalTRACK follows certain model foundations in literature (PRISM, ASHRAE Guideline 14, IPMVP Option C and the Uniform Methods Project for Whole Home Building Analysis). Building energy use is modeled as a combination of base load, heating load, and cooling load. Heating load and cooling load are assumed to have a linear relationship with heating and cooling demand, as approximated by heating and cooling degree days, beyond particular heating and cooling balance points. A number of candidate OLS models are fit to the consumption data using different combinations of heating and cooling balance points (ranging from 30 to 90 F) and different sets of independent variables. The model with the highest adjusted R-squared that contains strictly positive coefficients is selected as the final model and used to calculate normalized energy usage.

Model Types: CalTRACK specifies a linear relationship between energy use and temperature as reflected in the building consumption profile. In the most generic case, a model would include an intercept term, a heating balance point and heating slope coefficient, and a cooling balance point and a cooling slope coefficient. Depending on the fuel a building uses for heating or cooling or its consumption patterns, models with a single temperature coefficient and balance point (i.e., heating or cooling) may be more appropriate.

Difference in Normalized Annual Consumption (DNAC): The DNAC is calculated by using two CalTRACK regression models in conjunction with Typical Meteorological Year (TMY3) weather data, as follows:
- Two models are fit to the consumption data - one model for the baseline (pre-intervention) period and one for the reporting (post-intervention) period.
- Long-term heating and cooling degree days based on TMY3 data are substituted in both regression equations to calculate the Normalized Annual Consumption (NAC) for each period. TMY3 data is maintained by NREL and includes weather averages for 1020 locations in the US between 1991-2005.
- DNAC is determined by subtracting the two NACs (DNAC = Baseline NAC - Reporting NAC).

Disaggregation: Disaggregated loads are calculated from the different components of the statistical model fit. The weather sensitive components (heating and cooling load) are calculated by multiplying the relevant model coefficients (beta_hdd or beta_cdd) by the total degree days in a normal weather year (total HDD or CDD). For each site, the total HDD or CDD can be calculated using that site's estimated degree day balance points (also an output of the model) and the temperature for its closest weather station. The base load is estimated by multiplying the intercept of the statistical model by the number of days (365 for a full year).

Savings calculation: Savings are calculated by subtracting the DNAC for either comparison group from the DNAC for the treatment group.

Savings Uncertainty: Uncertainty presented in this analysis is calculated using the ASHRAE Guideline 14 formulation for aggregating the prediction uncertainty of point estimates in a time series. It is calculated at a 90% confidence level. The total uncertainty at the site-level is calculated using the sum of squares of the baseline and reporting models. Other aggregate uncertainty values (e.g. for a portfolio or for a difference-in-differences estimate) are also aggregated using the square root of the sum of squares.

Comparison Group Generation

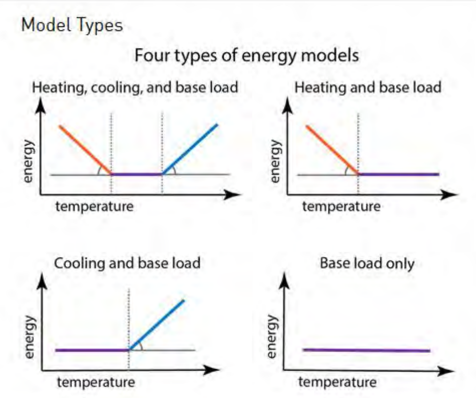
Site-level Matching: In monthly consumption matching, a comparison group is constructed by selecting n matches (n=5 in this analysis) from the comparison group pool with the shortest distance d to the treatment group customer under consideration. The pool is limited to non-participants within the same zipcode as the treatment group customer. The distance d is, in essence, a way to reduce 12 monthly consumption differences between any two customers to one metric (see Figure). In the present analysis, we selected (without replacement) five nearest neighbors for each treatment site based on the Euclidean distance of monthly consumption.

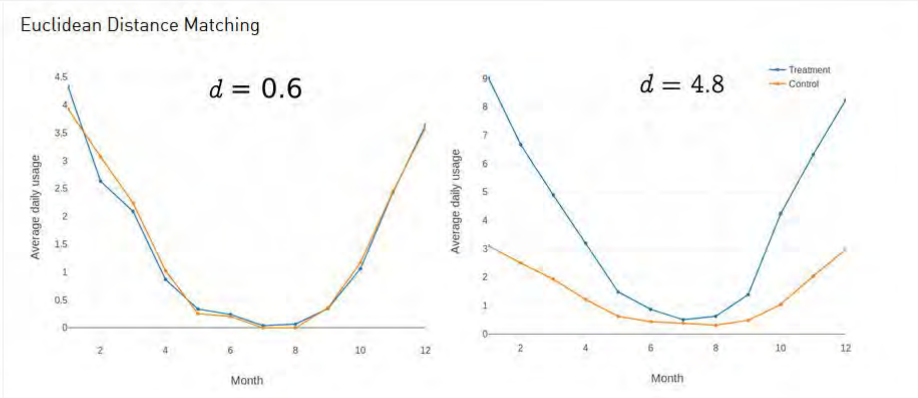
Future Participant Groups: Comparison groups comprising future participants are considered to be representative of participants in most aspects (observable and non-observable). For example, future participants are known to be eligible to receive the measure, and for some measures, they may have the same baseline equipment as the participants. Future participants have the same propensity to participate in the program as participants, thus reducing or eliminating self-selection bias, something that is otherwise difficult to control for in a quasi-experimental study. More comprehensive data is typically collected for future participants, allowing for potentially better matching and more insightful analysis. From a practical perspective, future participant groups may be difficult to construct for all measures, unless a program has been running for multiple years and is considered stable with sufficient data collection over the analysis period. Sample sizes for the comparison group may also be constrained if using future participants.

Stratified sampling is applied to future participant groups to attempt to replicate the distributions of the underlying variable (annual consumption) in the comparison group. Annual consumption of all treatment sites is first split into deciles, then a random sample is selected from within each corresponding bin in the comparison group pool of future participants.

Geographical screen: For the site-level matched group, only sites within the same zipcode as the treatment site were considered as potential comparison group matches.

Sampling method: In all cases where sampling was required from the comparison group, sampling was performed without replacement.





Impact Evaluation Report

Gas Impact of Thermostats in Program Year 2016

Result Summary

Measure: Thermostats		Program Year: 2016	Fuel: Gas	*Last Consumption Data Update:* October 1, 2019 *Last Participation Data Update:* October 1, 2019 *CalTRACK Version:* 2.0
Meter Data Filters:	DNAC: <75%	DNAC Percentile: None	Annual Consumption Percentile: Remove Top and Bottom 0.5%	
Model Filters:	Period Length: 11 Months or Longer	R-Squared: >0.5	CV(RMSE): < 1	
Metadata Filters:	Cooling Zone(s): All		Heating Fuel: Gas	
Thermostat Name: All	Heat Pump Baseline: All	Heating Zone(s): All	Multi Measure Filter: Single Measure Only	
Heat Pump Manufacturer: All	Heat Pump Adv. Controls or Commissioning: All			

374 Treatment Meters	20 +/- 10 Therms Average Normal Year Pre-Post Difference in Consumption per Participant	3 +/- 1 % Percent Normal Year Pre-Post Difference in Consumption per Participant	705 Mean Baseline Consumption (Gas)	64% Realization Rate
1,867 Site-level Matched Meters	25 +/- 11 Therms Average Savings Relative to Site-level Matched Comparison Group	3 +/- 1% Percent Savings Relative to Site-level Matched Comparison Group	696 Mean Baseline Consumption (Gas)	80% Realization Rate
3,921 Future Participant Meters	55 +/- 10 Therms Average Savings Relative to Future Participant Group	8 +/- 1% Savings Relative to Future Participant Group	695 Mean Baseline Consumption (Gas)	175% Realization Rate

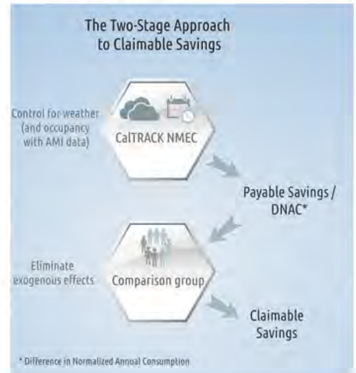
1. Introduction

This report contains the results of applying the two-stage approach (informed by the DOE's uniform methods chapter on whole building analysis) for calculating claimable savings to the selected portfolio of energy efficiency projects (see Figure). This approach begins with identification of two comparison groups for the treatment sample: (a) a site-level matched comparison group and (b) a future participant group. These groups are described below along with summary statistics (site locations, sample size, baseline consumption and baseline load disaggregation).

The CalTRACK methods are then applied to arrive at site-level savings, normalized for weather, and reflective of energy consumption changes for customers at the meter. Using a difference of differences for the treatment group with each comparison group accounts for population-level consumption changes (e.g. economic changes, rate changes, natural energy efficiency adoption etc.). The methods contained within this report are the outcome of a recent peer-reviewed study completed by Energy Trust of Oregon and Open Energy Efficiency (see "Methodology" section for more details).

The report includes the following sections:
Result Summary - Includes the overall portfolio results
Section 1. Introduction - Overview of report and the different groups included in the analysis
Section 2. Data Preparation - Data cleaning and sample attrition
Section 3. Modeling Results - CalTRACK model outputs and Difference in Normalized Annual Consumption (DNAC) results
Section 4. Methodology - Description of methods used in this report

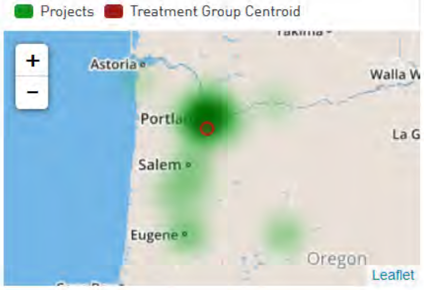
Two-Stage Approach



Treatment Group
The treatment group consists of sites that participated in the specified energy efficiency projects in the specified program year. Only sites that installed single measures are included in the treatment group. And this group includes the subset of sites that had sufficient data quality for modeling.

Treatment Site Locations

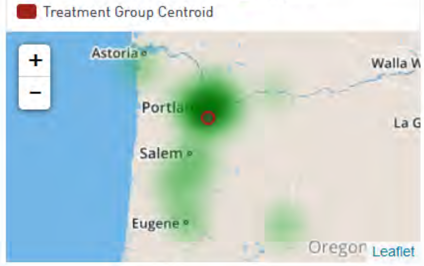
Projects Treatment Group Centroid



Site-level Matched Comparison Group
This group includes comparison group sites that were matched at the site-level to treatment group sites. Each treatment group site is matched to five comparison group sites from the same zipcode, but only the sites with sufficient data quality were included in the group. Matching was performed using monthly consumption in the baseline period as detailed in the Methodology section.

Site-level Matched Site Locations

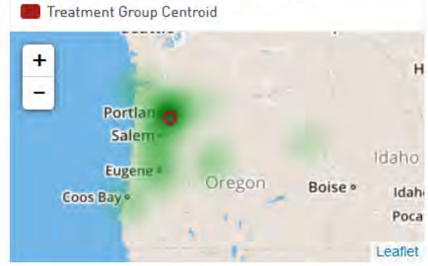
Comparison Group Centroid Projects
Treatment Group Centroid



Future Participant Group
The pool of sites that was used to create this group was composed of sites that installed the same measure in the year following the specified program year. The final sites were selected by stratified sampling using deciles of annual energy consumption.

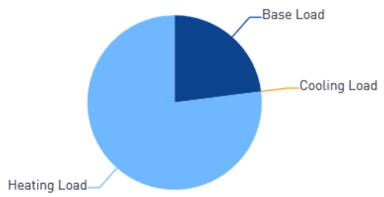
Future Participant Site Locations

Comparison Group Centroid Projects
Treatment Group Centroid

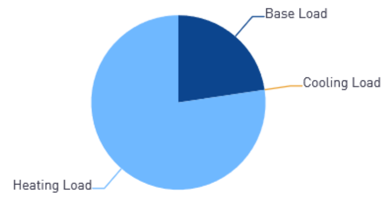


Treatment Group		Site-level Matched Comparison Group		Future Participant Group	
38.3 miles 80% of projects lie within this distance from treatment group centroid		0.1 miles Distance between treatment and comparison group centroids		1.3 miles Distance between treatment and future participant group centroids	
374 Meters	705 Mean Baseline Consumption (Gas)	1,867 Meters	696 Mean Baseline Consumption (Gas)	3,921 Meters	695 Mean Baseline Consumption (Gas)

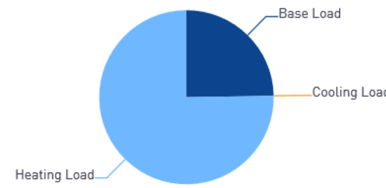
Load Disaggregation



Load Disaggregation



Load Disaggregation



2. Data Preparation

Consumption data preparation and cleaning followed best practices defined in the CalTRACK 2.0 billing methods. Some key aspects of the data cleaning process are highlighted here; please see the resources section for links to more detailed documentation. The initial and final sample sizes are shown below along with the percent of the treatment population that is represented by the sample. The sample attrition table shows the impact of each filtering criterion on sample size.

1,894	374	20%
Meters in Treatment Population	Final Sample Size	Percent of Treatment Population Represented by Sample

Sample Attrition Table

Filter	Selected Filter Value (if applicable)	Number of Dropped Meters	Sample Size after Applying Filter
Measure: Meters associated with a particular measure in program participation data. **Year:** Program year. **Fuel:** Type of metered fuel.	Measure: Thermostats -- Year: 2016 -- Fuel: Gas	--	1,894
Meters with valid consumption data in baseline and/or reporting periods.	--	46	1,848
MultiMeasure_Filter: Meters with single/multiple measure installations in baseline and/or reporting periods.	Multi Measure Filter: Single Measure Only	1,298	550
HeatingFuel: Meters with a valid heating fuel that corresponds to the selected filter value.	Heating Fuel: Gas	67	483
HeatingZone, CoolingZone: Meters in selected heating and/or cooling climate zones.	Heating Zone: All -- Cooling Zone: All	0	483
Other measure-specific filters.	--	0	483
PeriodLength_Threshold: Meters meeting a threshold number of months of valid consumption data.	Period Length: 11 Months or Longer	80	401
Meters with at least 5 site-level matched meters from the comparison group pool.	--	16	387
DNAC_Threshold: Meters with normalized change in annual energy consumption under a specified threshold.	DNAC: <75%	8	380

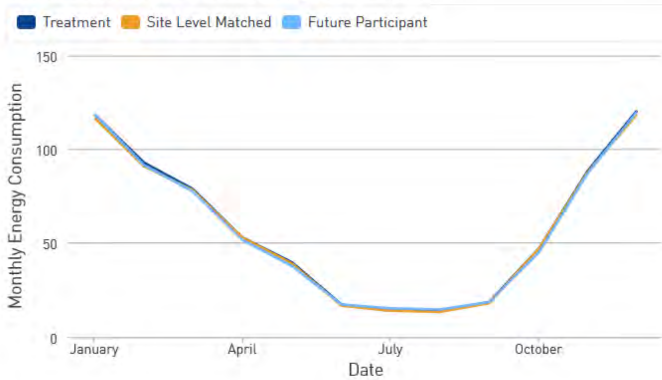
DNACPercentile_Threshold: Meters within specified percentile bands of normalized change in annual consumption.	DNAC Percentile: None	-2	379
ConsumptionPercentile_Threshold: Meters within specified percentile bounds of annual energy consumption.	Annual Consumption Percentile: Remove Top and Bottom 0.5%	1	379
R2_Threshold: Meters with valid model R-squared for the baseline and reporting periods that meet a specified threshold. Models may have invalid R-squared due to data issues.	R-Squared: >0.5	6	375
CVRMSE_Threshold: Meters with valid model CV(RMSE) for the baseline and reporting periods that meet a specified threshold.	CV(RMSE): < 1	0	374

3. Modeling Results

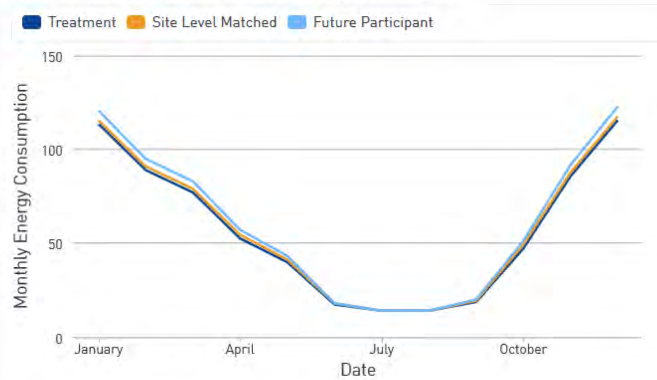
This section includes summaries of the Difference in Normalized Annual Consumption (DNAC) results for the treatment and comparison groups. The time series of monthly energy consumption illustrates the similarities and/or differences in energy consumption for the different groups in the baseline and reporting periods.

Below, you will find a breakdown of the DNAC results by group, showing the histograms of DNAC as well as the mean value expressed in raw units and as a percent of baseline annual consumption. Finally, the distribution of model types in the baseline and reporting periods are also provided as an additional layer of analysis.

Baseline Normal Year Monthly Energy Consumption

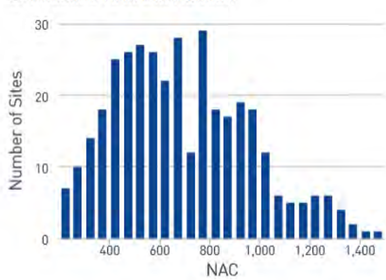


Post-Period Normal Year Monthly Energy Consumption



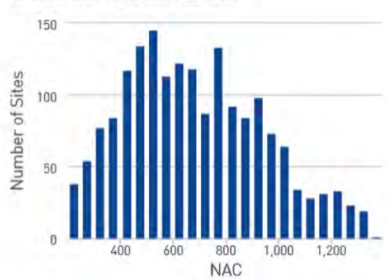
Treatment Group

ⓘ Baseline NAC Distribution



Site-level Matched Comparison Group

ⓘ Baseline NAC Distribution



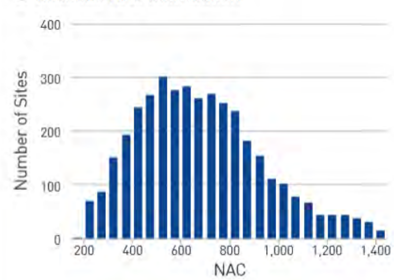
ⓘ

0.153

Annual Consumption p-value

Future Participant Group

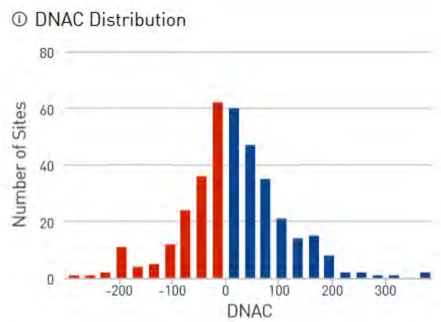
ⓘ Baseline NAC Distribution



ⓘ

0.172

Annual Consumption p-value

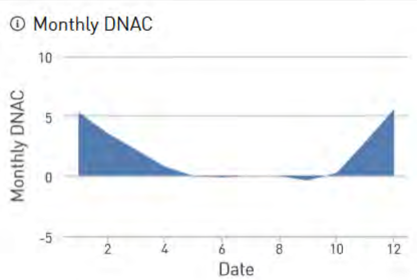


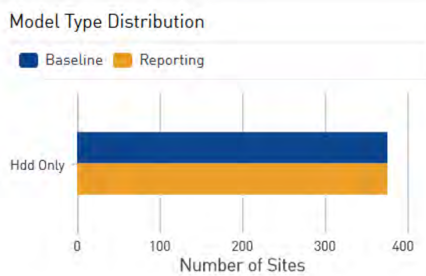
20 +/- 10 Therms

Average Difference in Normalized Annual Consumption per Participant

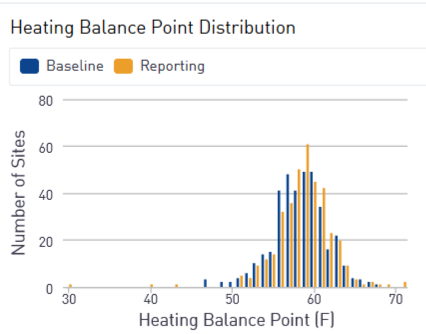
3 +/- 1 %

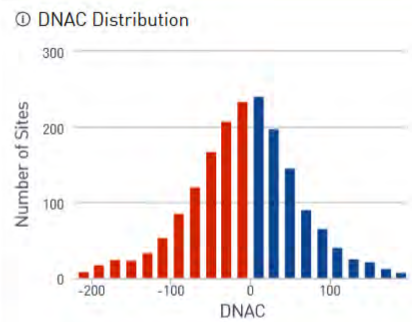
Difference in Normalized Annual Consumption as a Percent of Baseline





Treatment Group



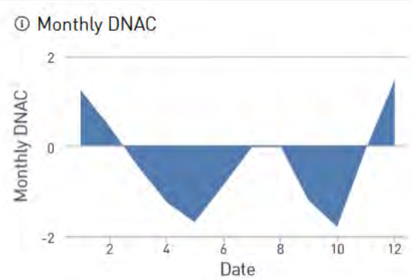


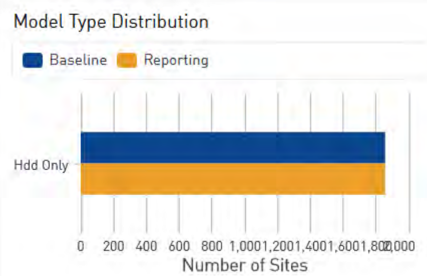
-5 +/- 4 Therms

Average Difference in Normalized Annual Consumption per Participant

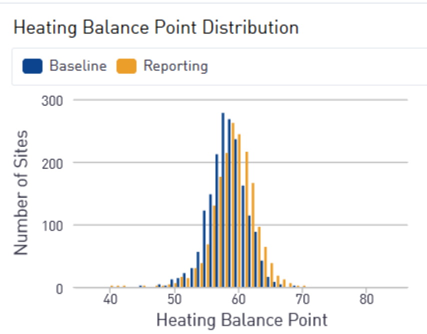
-1 +/- 1 %

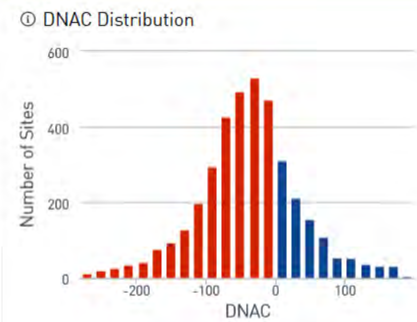
Difference in Normalized Annual Consumption as a Percent of Baseline





Site-level Matched Comparison Group



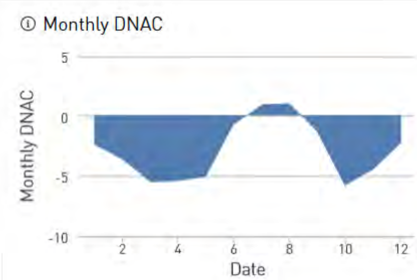


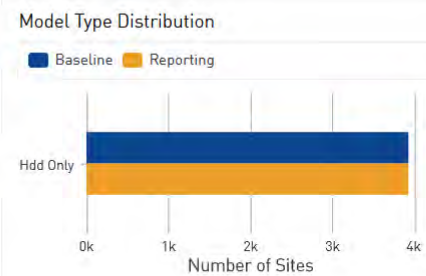
-34 +/- 3 Therms

Average Difference in Normalized Annual Consumption per Participant

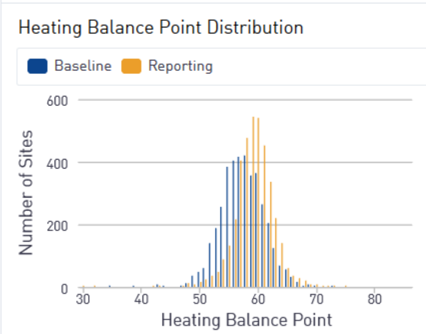
-5 +/- 0 %

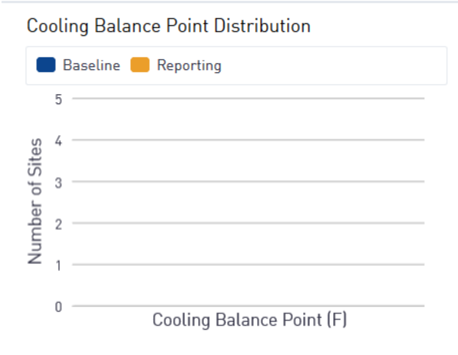
Difference in Normalized Annual Consumption as a Percent of Baseline

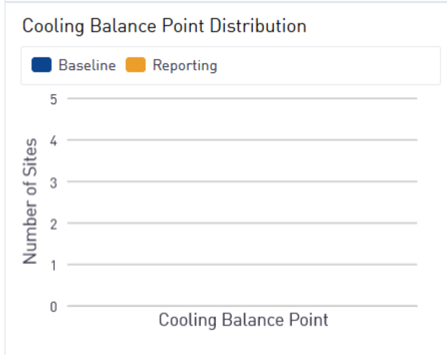


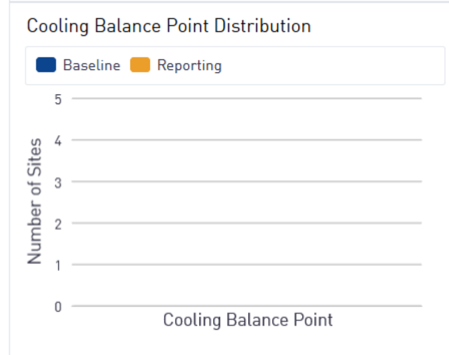


Future Participant Group









4. Methodology

CalTRACK and Comparison Group Methods
Documentation: docs.caltrack.org
Code: https://github.com/energy-market-methods/caltrack

Data Preparation

Baseline period: Since the predicted baseline may be unstable with different baseline period lengths, which may, in turn, affect calculated savings, the consensus of the CalTRACK 2.0 working group was to set the maximum baseline period at 12 months, since the year leading to the energy efficiency intervention is the most indicative of recent energy use trends and prolonging the baseline period increases the chance of other unmeasured factors affecting the baseline. In addition, CalTRACK uses a minimum 12-month baseline by default.

Blackout period: The blackout period refers to the time period between the end of the baseline period and the beginning of the reporting period. In this analysis, it is specified to coincide with the project installation time period, meaning that the billing period that contains the project installation date is dropped from the analysis.

Analysis periods: Different portions of the analysis used different time periods of consumption data, therefore, it is useful to clearly define these time periods and where they were used. Consider a project with an installation date on a particular day d in a particular month m in a particular program year y. The year before the program year is labelled as y-1, the year prior to that as y-2 and so on, while the years following the program year are labelled y+1, y+2 etc. In all cases, the billing period that contains the project installation was dropped from the analysis. Other sections of the analysis use the following time periods:
- *Treatment and site-level matched groups:* Baseline period includes the 12 months preceding the installation billing period. Reporting period includes the 12 months following the installation billing period.
- *Future participant group:* Baseline period is the calendar year preceding the program year (Year *y-1*). Reporting period is the program year itself (Year *y*).
- Site-level consumption matching was performed using the 12 months of data immediately prior to the project installation date.
- Equivalence tests were performed using data from the previous calendar year (y-1).

Modeling

Weather Normalization: Weather normalization of billing data in CalTRACK follows certain model foundations in literature (PRISM, ASHRAE Guideline 14, IPMVP Option C and the Uniform Methods Project for Whole Home Building Analysis). Building energy use is modeled as a combination of base load, heating load, and cooling load. Heating load and cooling load are assumed to have a linear relationship with heating and cooling demand, as approximated by heating and cooling degree days, beyond particular heating and cooling balance points. A number of candidate OLS models are fit to the consumption data using different combinations of heating and cooling balance points (ranging from 30 to 90 F) and different sets of independent variables. The model with the highest adjusted R-squared that contains strictly positive coefficients is selected as the final model and used to calculate normalized energy usage.

Model Types: CalTRACK specifies a linear relationship between energy use and temperature as reflected in the building consumption profile. In the most generic case, a model would include an intercept term, a heating balance point and heating slope coefficient, and a cooling balance point and a cooling slope coefficient. Depending on the fuel a building uses for heating or cooling or its consumption patterns, models with a single temperature coefficient and balance point (i.e., heating or cooling) may be more appropriate.

Difference in Normalized Annual Consumption (DNAC): The DNAC is calculated by using two CalTRACK regression models in conjunction with Typical Meteorological Year (TMY3) weather data, as follows:
- Two models are fit to the consumption data - one model for the baseline (pre-intervention) period and one for the reporting (post-intervention) period.
- Long-term heating and cooling degree days based on TMY3 data are substituted in both regression equations to calculate the Normalized Annual Consumption (NAC) for each period. TMY3 data is maintained by NREL and includes weather averages for 1020 locations in the US between 1991-2005.
- DNAC is determined by subtracting the two NACs (DNAC = Baseline NAC - Reporting NAC).

Disaggregation: Disaggregated loads are calculated from the different components of the statistical model fit. The weather sensitive components (heating and cooling load) are calculated by multiplying the relevant model coefficients (beta_hdd or beta_cdd) by the total degree days in a normal weather year (total HDD or CDD). For each site, the total HDD or CDD can be calculated using that site's estimated degree day balance points (also an output of the model) and the temperature for its closest weather station. The base load is estimated by multiplying the intercept of the statistical model by the number of days (365 for a full year).

Savings calculation: Savings are calculated by subtracting the DNAC for either comparison group from the DNAC for the treatment group.

Savings Uncertainty: Uncertainty presented in this analysis is calculated using the ASHRAE Guideline 14 formulation for aggregating the prediction uncertainty of point estimates in a time series. It is calculated at a 90% confidence level. The total uncertainty at the site-level is calculated using the sum of squares of the baseline and reporting models. Other aggregate uncertainty values (e.g. for a portfolio or for a difference-in-differences estimate) are also aggregated using the square root of the sum of squares.

Comparison Group Generation

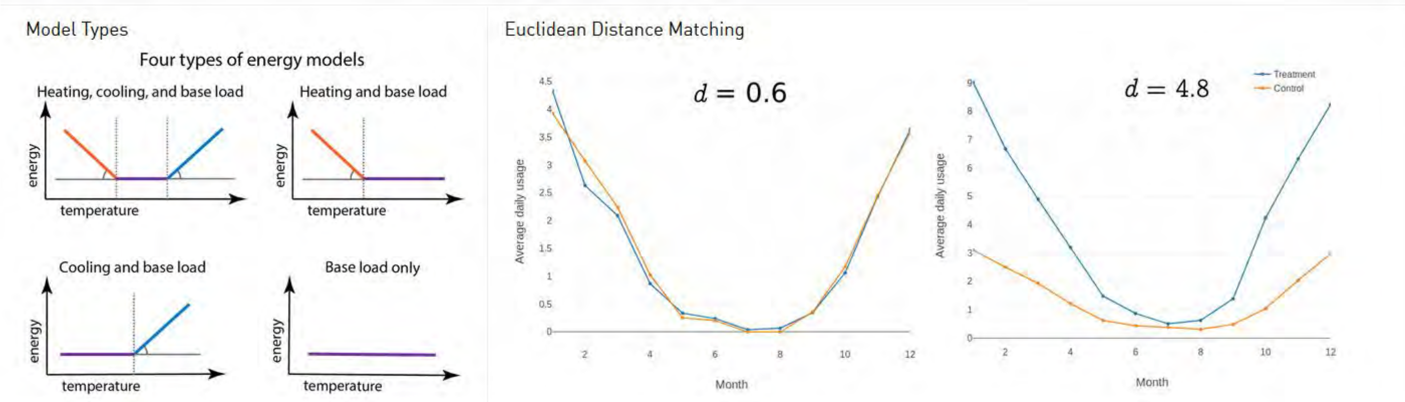
Site-level Matching: In monthly consumption matching, a comparison group is constructed by selecting n matches (n=5 in this analysis) from the comparison group pool with the shortest distance d to the treatment group customer under consideration. The pool is limited to non-participants within the same zipcode as the treatment group customer. The distance d is, in essence, a way to reduce 12 monthly consumption differences between any two customers to one metric (see Figure). In the present analysis, we selected (without replacement) five nearest neighbors for each treatment site based on the Euclidean distance of monthly consumption.

Future Participant Groups: Comparison groups comprising future participants are considered to be representative of participants in most aspects (observable and non-observable). For example, future participants are known to be eligible to receive the measure, and for some measures, they may have the same baseline equipment as the participants. Future participants have the same propensity to participate in the program as participants, thus reducing or eliminating self-selection bias, something that is otherwise difficult to control for in a quasi-experimental study. More comprehensive data is typically collected for future participants, allowing for potentially better matching and more insightful analysis. From a practical perspective, future participant groups may be difficult to construct for all measures, unless a program has been running for multiple years and is considered stable with sufficient data collection over the analysis period. Sample sizes for the comparison group may also be constrained if using future participants.

Stratified sampling is applied to future participant groups to attempt to replicate the distributions of the underlying variable (annual consumption) in the comparison group. Annual consumption of all treatment sites is first split into deciles, then a random sample is selected from within each corresponding bin in the comparison group pool of future participants.

Geographical screen: For the site-level matched group, only sites within the same zipcode as the treatment site were considered as potential comparison group matches.

Sampling method: In all cases where sampling was required from the comparison group, sampling was performed without replacement.



Impact Evaluation Report

Gas Impact of Thermostats in Program Year 2017

Result Summary

Measure: Thermostats		Program Year: 2017	Fuel: Gas	*Last Consumption Data Update:* October 1, 2019 *Last Participation Data Update:* October 1, 2019 *CalTRACK Version:* 2.0
Meter Data Filters:	DNAC: <75%	DNAC Percentile: None	Annual Consumption Percentile: Remove Top and Bottom 0.5%	
Model Filters:	Period Length: 11 Months or Longer	R-Squared: >0.5	CV(RMSE): < 1	
Metadata Filters:	Cooling Zone(s): All		Heating Fuel: Gas	
Thermostat Name: All	Heat Pump Baseline: All	Heating Zone(s): All	Multi Measure Filter: Single Measure Only	
Heat Pump Manufacturer: All	Heat Pump Adv. Controls or Commissioning: All			

438 Treatment Meters	25 +/- 10 Therms Average Normal Year Pre-Post Difference in Consumption per Participant	4 +/- 1 % Percent Normal Year Pre-Post Difference in Consumption per Participant	727 Mean Baseline Consumption (Gas)	77% Realization Rate
2,163 Site-level Matched Meters	27 +/- 10 Therms Average Savings Relative to Site-level Matched Comparison Group	3 +/- 1% Percent Savings Relative to Site-level Matched Comparison Group	717 Mean Baseline Consumption (Gas)	78% Realization Rate
3,485 Future Participant Meters	11 +/- 10 Therms Average Savings Relative to Future Participant Group	1 +/- 1% Savings Relative to Future Participant Group	738 Mean Baseline Consumption (Gas)	32% Realization Rate

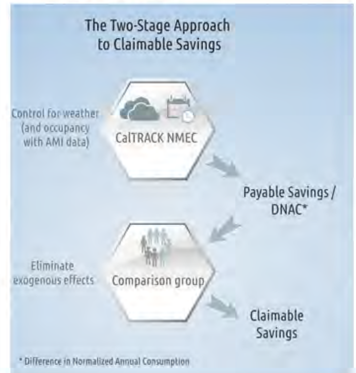
1. Introduction

This report contains the results of applying the two-stage approach (informed by the DOE's uniform methods chapter on whole building analysis) for calculating claimable savings to the selected portfolio of energy efficiency projects (see Figure). This approach begins with identification of two comparison groups for the treatment sample: (a) a site-level matched comparison group and (b) a future participant group. These groups are described below along with summary statistics (site locations, sample size, baseline consumption and baseline load disaggregation).

The CalTRACK methods are then applied to arrive at site-level savings, normalized for weather, and reflective of energy consumption changes for customers at the meter. Using a difference of differences for the treatment group with each comparison group accounts for population-level consumption changes (e.g. economic changes, rate changes, natural energy efficiency adoption etc.). The methods contained within this report are the outcome of a recent peer-reviewed study completed by Energy Trust of Oregon and Open Energy Efficiency (see "Methodology" section for more details).

The report includes the following sections:
Result Summary - Includes the overall portfolio results
Section 1. Introduction - Overview of report and the different groups included in the analysis
Section 2. Data Preparation - Data cleaning and sample attrition
Section 3. Modeling Results - CalTRACK model outputs and Difference in Normalized Annual Consumption (DNAC) results
Section 4. Methodology - Description of methods used in this report

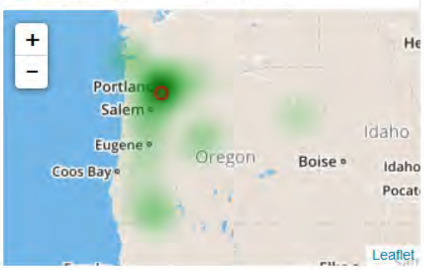
Two-Stage Approach



Treatment Group
The treatment group consists of sites that participated in the specified energy efficiency projects in the specified program year. Only sites that installed single measures are included in the treatment group. And this group includes the subset of sites that had sufficient data quality for modeling.

Treatment Site Locations

Projects | Treatment Group Centroid



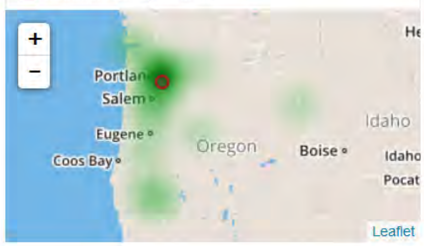
52.2 miles
80% of projects lie within this distance from treatment group centroid

438	727
Meters	Mean Baseline Consumption (Gas)

Site-level Matched Comparison Group
This group includes comparison group sites that were matched at the site-level to treatment group sites. Each treatment group site is matched to five comparison group sites from the same zipcode, but only the sites with sufficient data quality were included in the group. Matching was performed using monthly consumption in the baseline period as detailed in the Methodology section.

Site-level Matched Site Locations

Comparison Group Centroid | Projects | Treatment Group Centroid



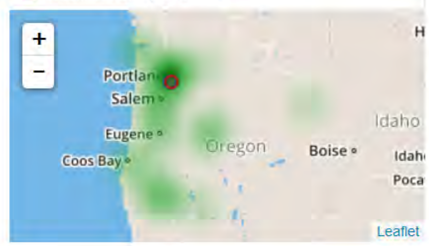
0.7 miles
Distance between treatment and comparison group centroids

2,163	717
Meters	Mean Baseline Consumption (Gas)

Future Participant Group
The pool of sites that was used to create this group was composed of sites that installed the same measure in the year following the specified program year. The final sites were selected by stratified sampling using deciles of annual energy consumption.

Future Participant Site Locations

Comparison Group Centroid | Projects | Treatment Group Centroid

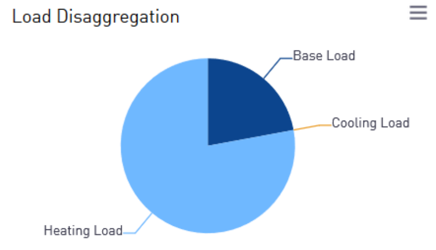


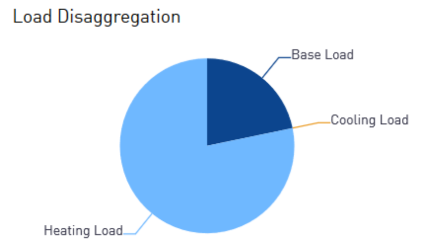
2.3 miles
Distance between treatment and future participant group centroids

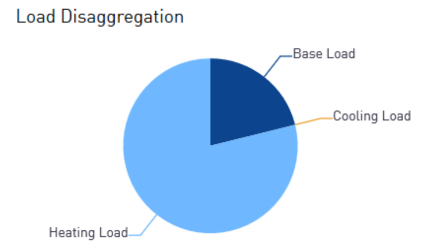
3,485	738
Meters	Mean Baseline Consumption (Gas)

Report Date: December 6, 2019

RECURVE







2. Data Preparation

Consumption data preparation and cleaning followed best practices defined in the CalTRACK 2.0 billing methods. Some key aspects of the data cleaning process are highlighted here; please see the resources section for links to more detailed documentation. The initial and final sample sizes are shown below along with the percent of the treatment population that is represented by the sample. The sample attrition table shows the impact of each filtering criterion on sample size.

2,965	438	15%
Meters in Treatment Population	Final Sample Size	Percent of Treatment Population Represented by Sample

Sample Attrition Table

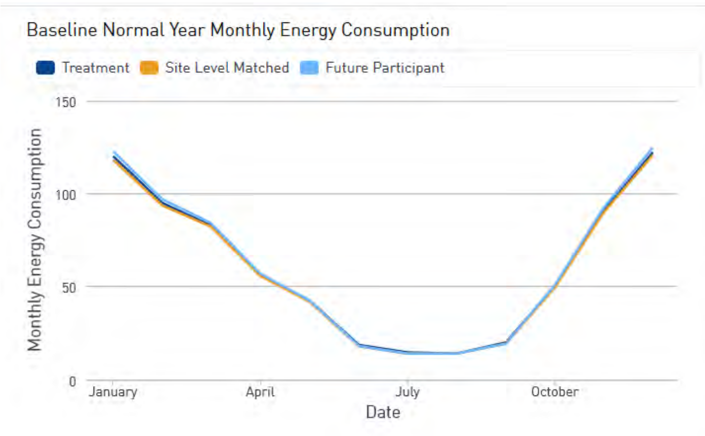
Filter	Selected Filter Value (if applicable)	Number of Dropped Meters	Sample Size after Applying Filter
Measure: Meters associated with a particular measure in program participation data. **Year:** Program year. **Fuel:** Type of metered fuel.	Measure: Thermostats -- Year: 2017 -- Fuel: Gas	--	2,965
Meters with valid consumption data in baseline and/or reporting periods.	--	87	2,878
MultiMeasure_Filter: Meters with single/multiple measure installations in baseline and/or reporting periods.	Multi Measure Filter: Single Measure Only	2,198	680
HeatingFuel: Meters with a valid heating fuel that corresponds to the selected filter value.	Heating Fuel: Gas	53	627
HeatingZone, CoolingZone: Meters in selected heating and/or cooling climate zones.	Heating Zone: All -- Cooling Zone: All	0	627
Other measure-specific filters.	--	0	627
PeriodLength_Threshold: Meters meeting a threshold number of months of valid consumption data.	Period Length: 11 Months or Longer	100	527
Meters with at least 5 site-level matched meters from the comparison group pool.	--	70	457
DNAC_Threshold: Meters with normalized change in annual energy consumption under a specified threshold.	DNAC: <75%	12	445

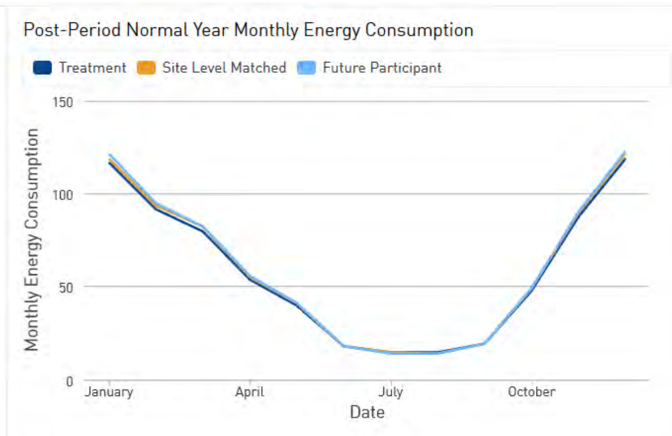
DNACPercentile_Threshold: Meters within specified percentile bands of normalized change in annual consumption.	DNAC Percentile: All	0	445
ConsumptionPercentile_Threshold: Meters within specified percentile bounds of annual energy consumption.	Annual Consumption Percentile: Remove Top and Bottom 0.5%	2	443
R2_Threshold: Meters with valid model R-squared for the baseline and reporting periods that meet a specified threshold. Models may have invalid R-squared due to data issues.	R-Squared: >0.5	5	438
CVRMSE_Threshold: Meters with valid model CV(RMSE) for the baseline and reporting periods that meet a specified threshold.	CV(RMSE): < 1	0	438

3. Modeling Results

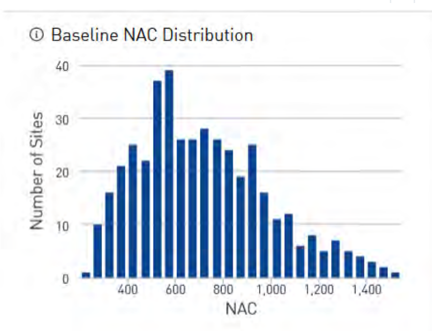
This section includes summaries of the Difference in Normalized Annual Consumption (DNAC) results for the treatment and comparison groups. The time series of monthly energy consumption illustrates the similarities and/or differences in energy consumption for the different groups in the baseline and reporting periods.

Below, you will find a breakdown of the DNAC results by group, showing the histograms of DNAC as well as the mean value expressed in raw units and as a percent of baseline annual consumption. Finally, the distribution of model types in the baseline and reporting periods are also provided as an additional layer of analysis.

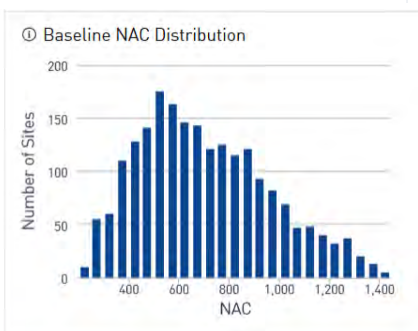




Treatment Group



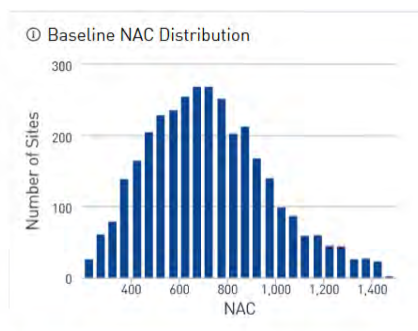
Site-level Matched Comparison Group



0.176

Annual Consumption p-value

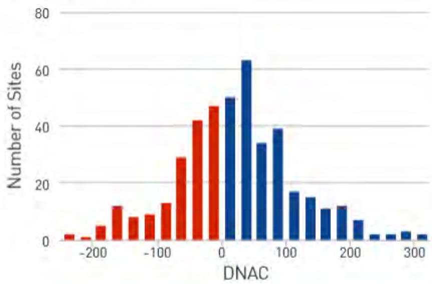
Future Participant Group



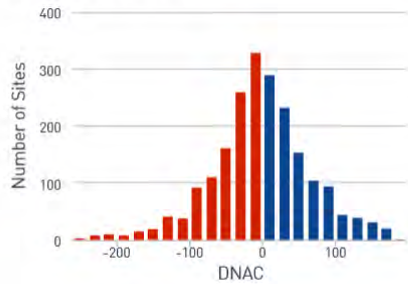
0.196

Annual Consumption p-value

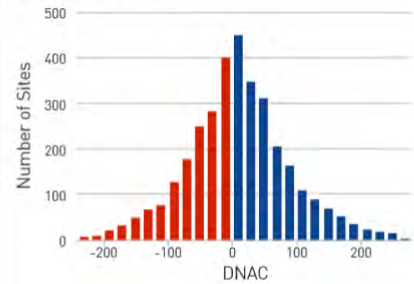
ⓘ DNAC Distribution



ⓘ DNAC Distribution



ⓘ DNAC Distribution



25 +/- 10 Therms
Average Difference in Normalized Annual Consumption per Participant

4 +/- 1 %
Difference in Normalized Annual Consumption as a Percent of Baseline

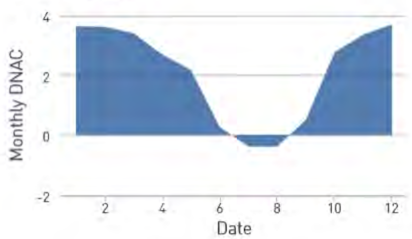
0 +/- 3 Therms
Average Difference in Normalized Annual Consumption per Participant

0 +/- 0 %
Difference in Normalized Annual Consumption as a Percent of Baseline

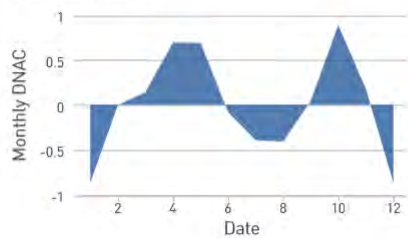
15 +/- 4 Therms
Average Difference in Normalized Annual Consumption per Participant

2 +/- 0 %
Difference in Normalized Annual Consumption as a Percent of Baseline

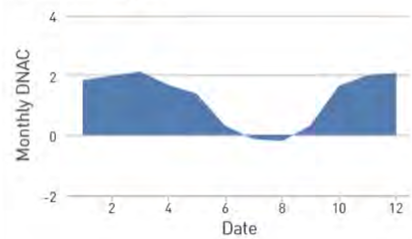
ⓘ Monthly DNAC



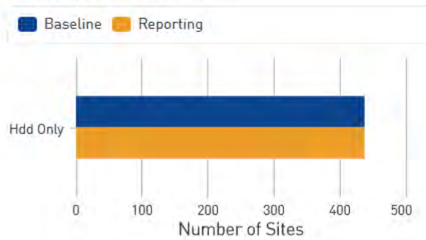
ⓘ Monthly DNAC



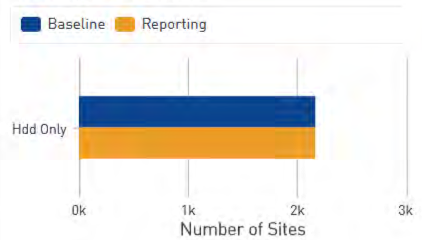
ⓘ Monthly DNAC



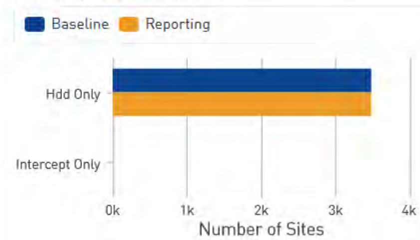
Model Type Distribution



Model Type Distribution

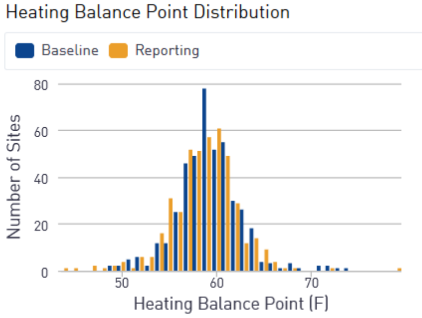


Model Type Distribution



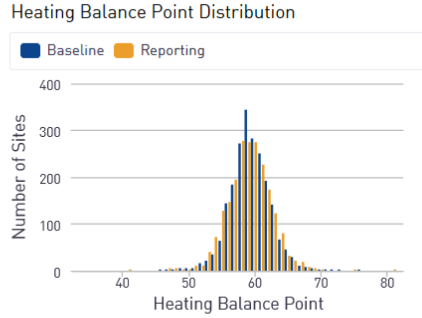
Treatment Group

Heating Balance Point Distribution



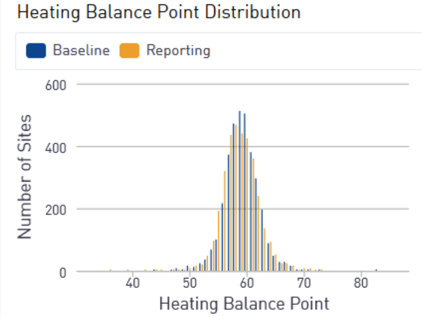
Site-level Matched Comparison Group

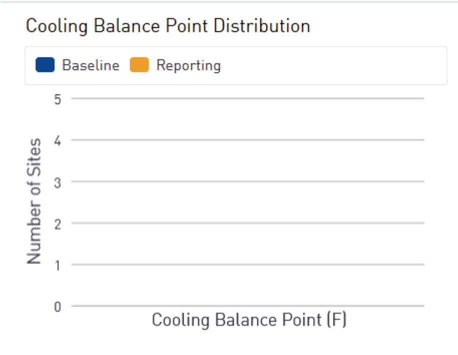
Heating Balance Point Distribution

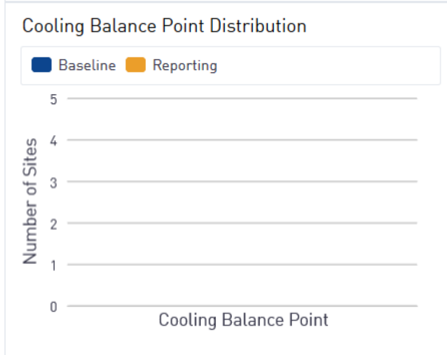


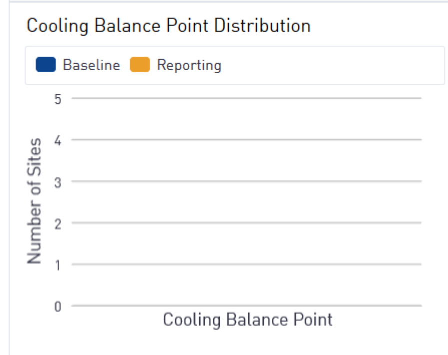
Future Participant Group

Heating Balance Point Distribution









4. Methodology

CalTRACK and Comparison Group Methods
Documentation: docs.caltrack.org
Code: https://github.com/energy-market-methods/caltrack

Data Preparation

Baseline period: Since the predicted baseline may be unstable with different baseline period lengths, which may, in turn, affect calculated savings, the consensus of the CalTRACK 2.0 working group was to set the maximum baseline period at 12 months, since the year leading to the energy efficiency intervention is the most indicative of recent energy use trends and prolonging the baseline period increases the chance of other unmeasured factors affecting the baseline. In addition, CalTRACK uses a minimum 12-month baseline by default.

Blackout period: The blackout period refers to the time period between the end of the baseline period and the beginning of the reporting period. In this analysis, it is specified to coincide with the project installation time period, meaning that the billing period that contains the project installation date is dropped from the analysis.

Analysis periods: Different portions of the analysis used different time periods of consumption data, therefore, it is useful to clearly define these time periods and where they were used. Consider a project with an installation date on a particular day d in a particular month m in a particular program year y. The year before the program year is labelled as y-1, the year prior to that as y-2 and so on, while the years following the program year are labelled y+1, y+2 etc. In all cases, the billing period that contains the project installation was dropped from the analysis. Other sections of the analysis use the following time periods:
- *Treatment and site-level matched groups:* Baseline period includes the 12 months preceding the installation billing period. Reporting period includes the 12 months following the installation billing period.
- *Future participant group:* Baseline period is the calendar year preceding the program year (Year *y-1*). Reporting period is the program year itself (Year *y*).
- Site-level consumption matching was performed using the 12 months of data immediately prior to the project installation date.
- Equivalence tests were performed using data from the previous calendar year (y-1).

Modeling

Weather Normalization: Weather normalization of billing data in CalTRACK follows certain model foundations in literature (PRISM, ASHRAE Guideline 14, IPMVP Option C and the Uniform Methods Project for Whole Home Building Analysis). Building energy use is modeled as a combination of base load, heating load, and cooling load. Heating load and cooling load are assumed to have a linear relationship with heating and cooling demand, as approximated by heating and cooling degree days, beyond particular heating and cooling balance points. A number of candidate OLS models are fit to the consumption data using different combinations of heating and cooling balance points (ranging from 30 to 90 F) and different sets of independent variables. The model with the highest adjusted R-squared that contains strictly positive coefficients is selected as the final model and used to calculate normalized energy usage.

Model Types: CalTRACK specifies a linear relationship between energy use and temperature as reflected in the building consumption profile. In the most generic case, a model would include an intercept term, a heating balance point and heating slope coefficient, and a cooling balance point and a cooling slope coefficient. Depending on the fuel a building uses for heating or cooling or its consumption patterns, models with a single temperature coefficient and balance point (i.e., heating or cooling) may be more appropriate.

Difference in Normalized Annual Consumption (DNAC): The DNAC is calculated by using two CalTRACK regression models in conjunction with Typical Meteorological Year (TMY3) weather data, as follows:
- Two models are fit to the consumption data - one model for the baseline (pre-intervention) period and one for the reporting (post-intervention) period.
- Long-term heating and cooling degree days based on TMY3 data are substituted in both regression equations to calculate the Normalized Annual Consumption (NAC) for each period. TMY3 data is maintained by NREL and includes weather averages for 1020 locations in the US between 1991-2005.
- DNAC is determined by subtracting the two NACs (DNAC = Baseline NAC - Reporting NAC).

Disaggregation: Disaggregated loads are calculated from the different components of the statistical model fit. The weather sensitive components (heating and cooling load) are calculated by multiplying the relevant model coefficients (beta_hdd or beta_cdd) by the total degree days in a normal weather year (total HDD or CDD). For each site, the total HDD or CDD can be calculated using that site's estimated degree day balance points (also an output of the model) and the temperature for its closest weather station. The base load is estimated by multiplying the intercept of the statistical model by the number of days (365 for a full year).

Savings calculation: Savings are calculated by subtracting the DNAC for either comparison group from the DNAC for the treatment group.

Savings Uncertainty: Uncertainty presented in this analysis is calculated using the ASHRAE Guideline 14 formulation for aggregating the prediction uncertainty of point estimates in a time series. It is calculated at a 90% confidence level. The total uncertainty at the site-level is calculated using the sum of squares of the baseline and reporting models. Other aggregate uncertainty values (e.g. for a portfolio or for a difference-in-differences estimate) are also aggregated using the square root of the sum of squares.

Comparison Group Generation

Site-level Matching: In monthly consumption matching, a comparison group is constructed by selecting n matches (n=5 in this analysis) from the comparison group pool with the shortest distance d to the treatment group customer under consideration. The pool is limited to non-participants within the same zipcode as the treatment group customer. The distance d is, in essence, a way to reduce 12 monthly consumption differences between any two customers to one metric (see Figure). In the present analysis, we selected (without replacement) five nearest neighbors for each treatment site based on the Euclidean distance of monthly consumption.

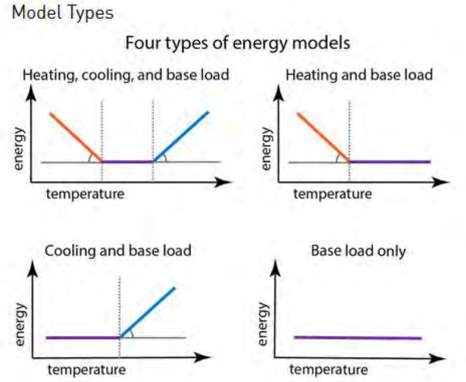
Future Participant Groups: Comparison groups comprising future participants are considered to be representative of participants in most aspects (observable and non-observable). For example, future participants are known to be eligible to receive the measure, and for some measures, they may have the same baseline equipment as the participants. Future participants have the same propensity to participate in the program as participants, thus reducing or eliminating self-selection bias, something that is otherwise difficult to control for in a quasi-experimental study. More comprehensive data is typically collected for future participants, allowing for potentially better matching and more insightful analysis. From a practical perspective, future participant groups may be difficult to construct for all measures, unless a program has been running for multiple years and is considered stable with sufficient data collection over the analysis period. Sample sizes for the comparison group may also be constrained if using future participants.

Stratified sampling is applied to future participant groups to attempt to replicate the distributions of the underlying variable (annual consumption) in the comparison group. Annual consumption of all treatment sites is first split into deciles, then a random sample is selected from within each corresponding bin in the comparison group pool of future participants.

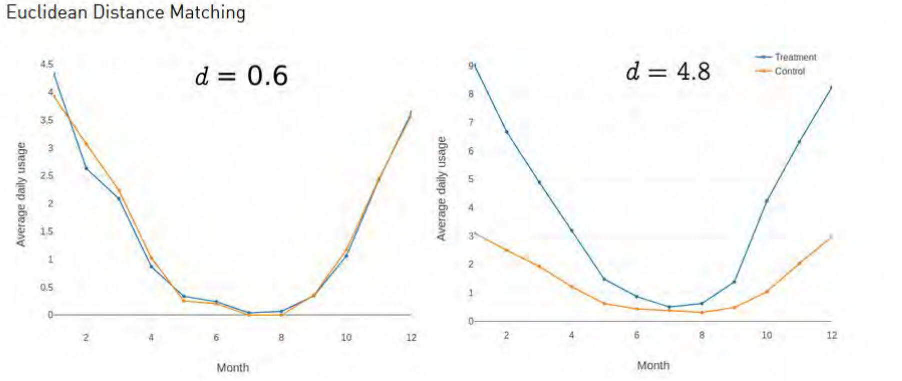
Geographical screen: For the site-level matched group, only sites within the same zipcode as the treatment site were considered as potential comparison group matches.

Sampling method: In all cases where sampling was required from the comparison group, sampling was performed without replacement.

Model Types



Euclidean Distance Matching



Impact Evaluation Report

Gas Impact of Thermostats in Program Year 2015, 2016, 2017

Result Summary

Measure: Thermostats	Program Year: 2015, 2016, 2017		Fuel: Gas
Meter Data Filters:	DNAC: <75%	DNAC Percentile: None	Annual Consumption Percentile: Remove Top and Bottom 0.5%
Model Filters:	Period Length: 11 Months or Longer	R-Squared: >0.5	CVIRMSEI: < 1
Metadata Filters:	Cooling Zone(s): All		Heating Fuel: Gas
Thermostat Name: Nest, Nest / Heat Pump Baseline: All	Heating Zone(s): All		Multi Measure Filter: Single Measure Only
Heat Pump Manufacturer: All / Heat Pump Adv. Controls or Commissioning: All			

Last Consumption Data Update: October 1, 2019

Last Participation Data Update: October 1, 2019

CalTRACK Version: 2.0

775 Treatment Meters	14 +/- 7 Therms Average Normal Year Pre-Post Difference in Consumption per Participant	2 +/- 1 % Percent Normal Year Pre-Post Difference in Consumption per Participant	713 Mean Baseline Consumption (Gas)	44% Realization Rate
3,847 Site-level Matched Meters	21 +/- 8 Therms Average Savings Relative to Site-level Matched Comparison Group	3 +/- 1% Percent Savings Relative to Site-level Matched Comparison Group	703 Mean Baseline Consumption (Gas)	70% Realization Rate
10.5k Future Participant Meters	29 +/- 7 Therms Average Savings Relative to Future Participant Group	4 +/- 1% Savings Relative to Future Participant Group	708 Mean Baseline Consumption (Gas)	94% Realization Rate

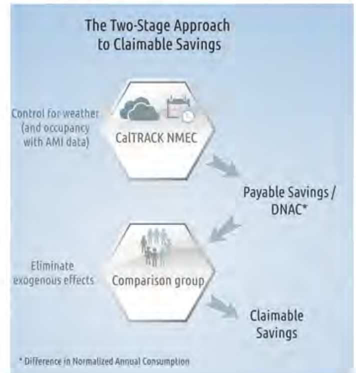
1. Introduction

This report contains the results of applying the two-stage approach (informed by the DOE's uniform methods chapter on whole building analysis) for calculating claimable savings to the selected portfolio of energy efficiency projects (see Figure). This approach begins with identification of two comparison groups for the treatment sample: (a) a site-level matched comparison group and (b) a future participant group. These groups are described below along with summary statistics (site locations, sample size, baseline consumption and baseline load disaggregation).

The CalTRACK methods are then applied to arrive at site-level savings, normalized for weather, and reflective of energy consumption changes for customers at the meter. Using a difference of differences for the treatment group with each comparison group accounts for population-level consumption changes (e.g. economic changes, rate changes, natural energy efficiency adoption etc.). The methods contained within this report are the outcome of a recent peer-reviewed study completed by Energy Trust of Oregon and Open Energy Efficiency (see "Methodology" section for more details).

The report includes the following sections:
Result Summary - Includes the overall portfolio results
Section 1. Introduction - Overview of report and the different groups included in the analysis
Section 2. Data Preparation - Data cleaning and sample attrition
Section 3. Modeling Results - CalTRACK model outputs and Difference in Normalized Annual Consumption (DNAC) results
Section 4. Methodology - Description of methods used in this report

Two-Stage Approach

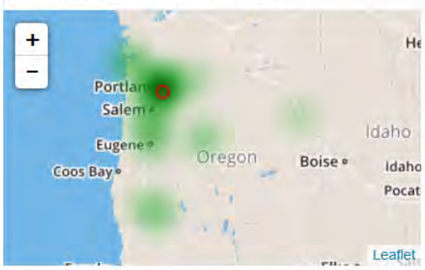


Treatment Group

The treatment group consists of sites that participated in the specified energy efficiency projects in the specified program year. Only sites that installed single measures are included in the treatment group. And this group includes the subset of sites that had sufficient data quality for modeling.

Treatment Site Locations

Projects | Treatment Group Centroid



38.2 miles

80% of projects lie within this distance from treatment group centroid

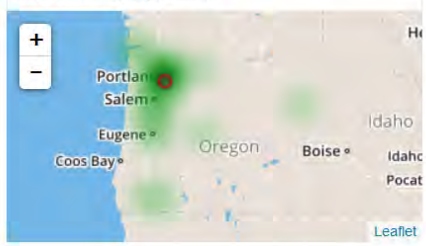
774	713
Meters	Mean Baseline Consumption (Gas)

Site-level Matched Comparison Group

This group includes comparison group sites that were matched at the site-level to treatment group sites. Each treatment group site is matched to five comparison group sites from the same zipcode, but only the sites with sufficient data quality were included in the group. Matching was performed using monthly consumption in the baseline period as detailed in the Methodology section.

Site-level Matched Site Locations

Comparison Group Centroid | Projects | Treatment Group Centroid



0.2 miles

Distance between treatment and comparison group centroids

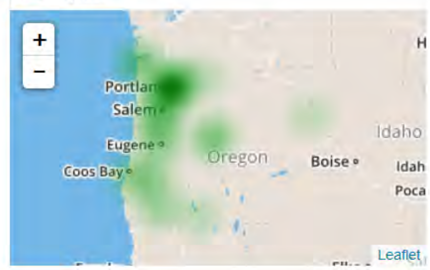
3,847	703
Meters	Mean Baseline Consumption (Gas)

Future Participant Group

The pool of sites that was used to create this group was composed of sites that installed the same measure in the year following the specified program year. The final sites were selected by stratified sampling using deciles of annual energy consumption.

Future Participant Site Locations

Projects

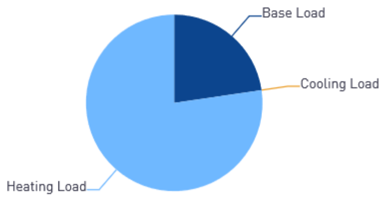


1.0 miles

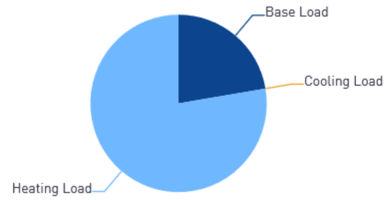
Distance between treatment and future participant group centroids

10.5k	708
Meters	Mean Baseline Consumption (Gas)

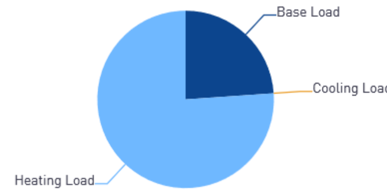
Load Disaggregation



Load Disaggregation



Load Disaggregation



2. Data Preparation

Consumption data preparation and cleaning followed best practices defined in the CalTRACK 2.0 billing methods. Some key aspects of the data cleaning process are highlighted here; please see the resources section for links to more detailed documentation. The initial and final sample sizes are shown below along with the percent of the treatment population that is represented by the sample. The sample attrition table shows the impact of each filtering criterion on sample size.

5,347	775	15%
Meters in Treatment Population	Final Sample Size	Percent of Treatment Population Represented by Sample

Sample Attrition Table

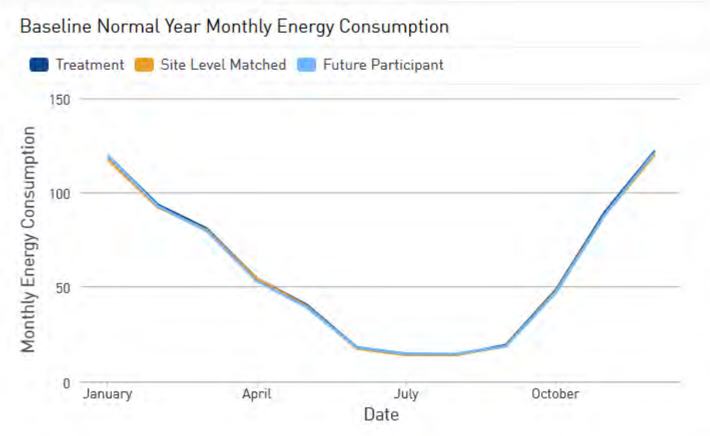
Filter	Selected Filter Value (if applicable)	Number of Dropped Meters	Sample Size after Applying Filter
Measure: Meters associated with a particular measure in program participation data. **Year:** Program year. **Fuel:** Type of metered fuel.	Measure: Thermostats -- Year: 2015, 2016, 2017 -- Fuel: Gas	--	5,347
Meters with valid consumption data in baseline and/or reporting periods.	--	148	5,199
MultiMeasure_Filter: Meters with single/multiple measure installations in baseline and/or reporting periods.	Multi Measure Filter: Single Measure Only	3,830	1,369
HeatingFuel: Meters with a valid heating fuel that corresponds to the selected filter value.	Heating Fuel: Gas	128	1,241
HeatingZone, CoolingZone: Meters in selected heating and/or cooling climate zones.	Heating Zone: All -- Cooling Zone: All	0	1,241
Other measure-specific filters.	--	230	1,011
PeriodLength_Threshold: Meters meeting a threshold number of months of valid consumption data.	Period Length: 11 Months or Longer	138	872
Meters with at least 5 site-level matched meters from the comparison group pool.	--	68	804
DNAC_Threshold: Meters with normalized change in annual energy consumption under a specified threshold.	DNAC: <75%	17	787

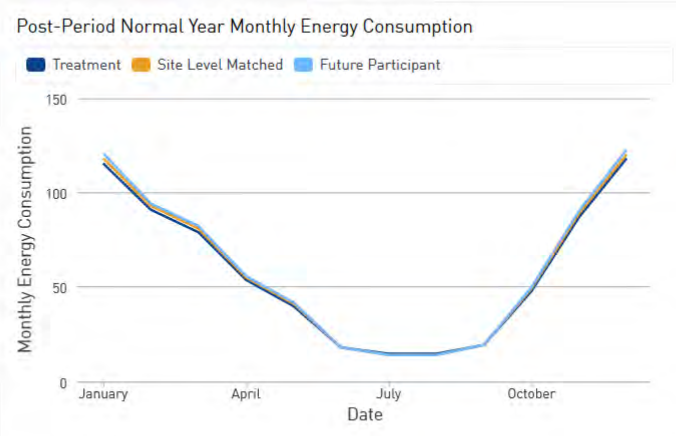
DNACPercentile_Threshold: Meters within specified percentile bands of normalized change in annual consumption.	DNAC Percentile: None	1	787
ConsumptionPercentile_Threshold: Meters within specified percentile bounds of annual energy consumption.	Annual Consumption Percentile: Remove Top and Bottom 0.5%	1	783
R2_Threshold: Meters with valid model R-squared for the baseline and reporting periods that meet a specified threshold. Models may have invalid R-squared due to data issues.	R-Squared: >0.5	9	774
CVRMSE_Threshold: Meters with valid model CV(RMSE) for the baseline and reporting periods that meet a specified threshold.	CV(RMSE): < 1	0	775

3. Modeling Results

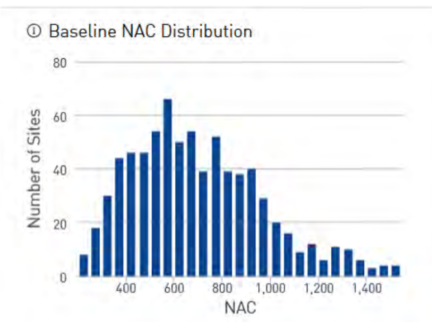
This section includes summaries of the Difference in Normalized Annual Consumption (DNAC) results for the treatment and comparison groups. The time series of monthly energy consumption illustrates the similarities and/or differences in energy consumption for the different groups in the baseline and reporting periods.

Below, you will find a breakdown of the DNAC results by group, showing the histograms of DNAC as well as the mean value expressed in raw units and as a percent of baseline annual consumption. Finally, the distribution of model types in the baseline and reporting periods are also provided as an additional layer of analysis.

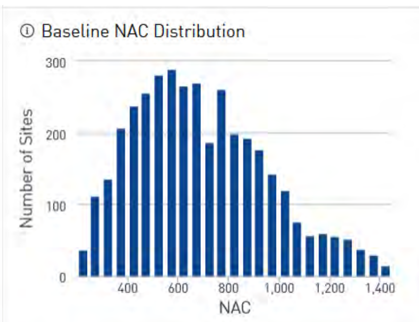




Treatment Group



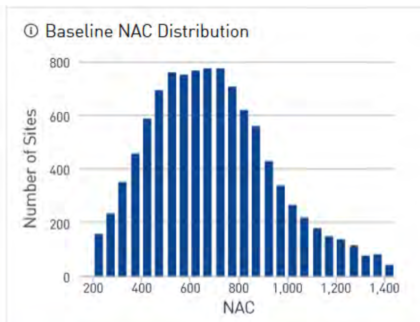
Site-level Matched Comparison Group



0.169

Annual Consumption p-value

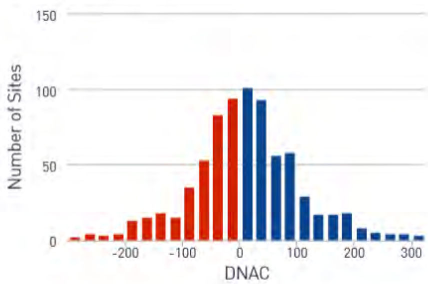
Future Participant Group



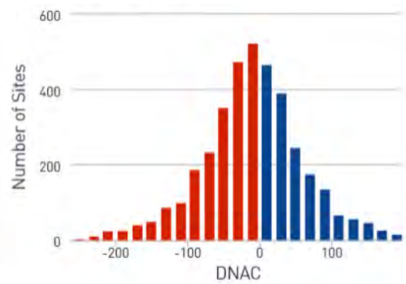
0.115

Annual Consumption p-value

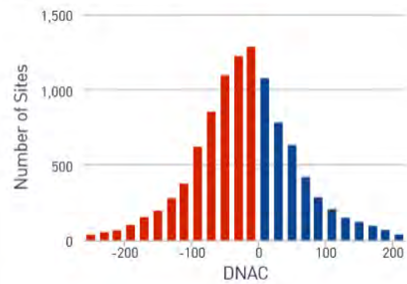
ⓘ DNAC Distribution



ⓘ DNAC Distribution



ⓘ DNAC Distribution



14 +/- 7 Therms

Average Difference in Normalized Annual Consumption per Participant

2 +/- 1 %

Difference in Normalized Annual Consumption as a Percent of Baseline

-7 +/- 3 Therms

Average Difference in Normalized Annual Consumption per Participant

-1 +/- 0 %

Difference in Normalized Annual Consumption as a Percent of Baseline

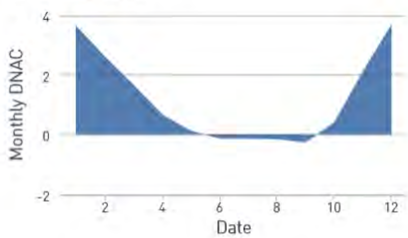
-15 +/- 2 Therms

Average Difference in Normalized Annual Consumption per Participant

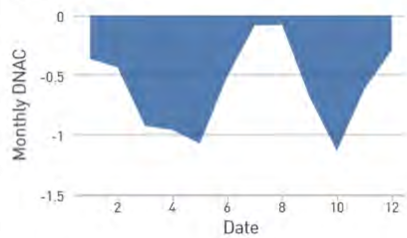
-2 +/- 0 %

Difference in Normalized Annual Consumption as a Percent of Baseline

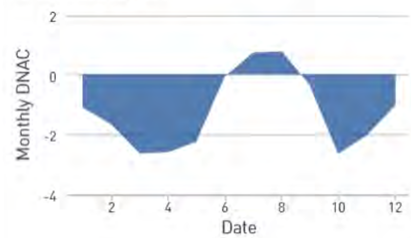
ⓘ Monthly DNAC



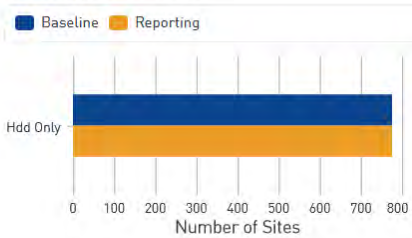
ⓘ Monthly DNAC



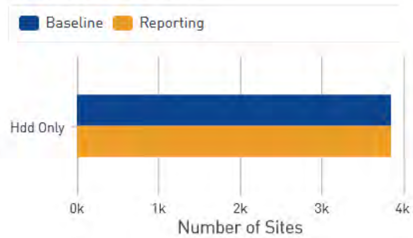
ⓘ Monthly DNAC



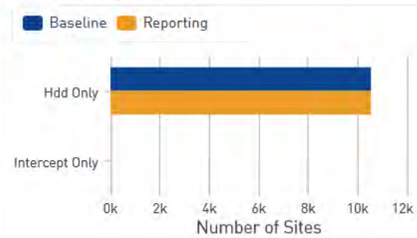
Model Type Distribution



Model Type Distribution



Model Type Distribution

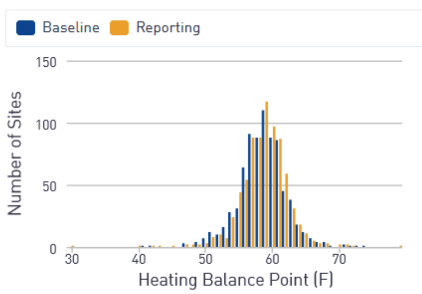


Treatment Group

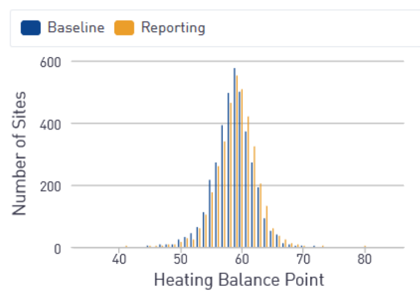
Site-level Matched Comparison Group

Future Participant Group

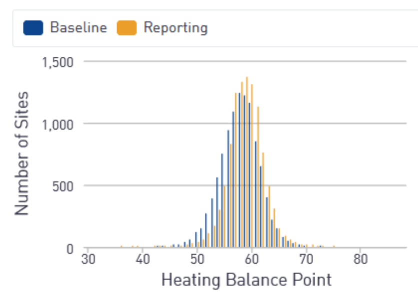
Heating Balance Point Distribution

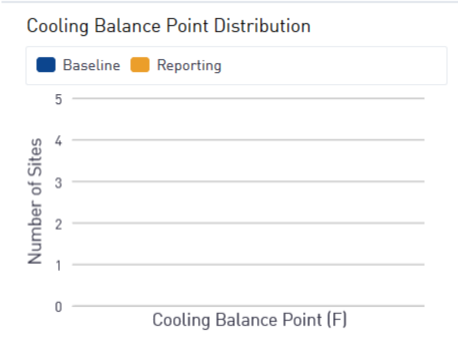


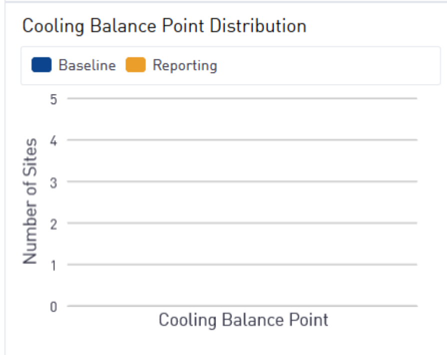
Heating Balance Point Distribution

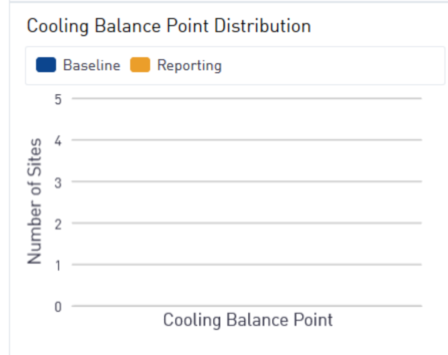


Heating Balance Point Distribution









4. Methodology

CalTRACK and Comparison Group Methods
Documentation: docs.caltrack.org
Code: https://github.com/energy-market-methods/caltrack

Data Preparation

Baseline period: Since the predicted baseline may be unstable with different baseline period lengths, which may, in turn, affect calculated savings, the consensus of the CalTRACK 2.0 working group was to set the maximum baseline period at 12 months, since the year leading to the energy efficiency intervention is the most indicative of recent energy use trends and prolonging the baseline period increases the chance of other unmeasured factors affecting the baseline. In addition, CalTRACK uses a minimum 12-month baseline by default.

Blackout period: The blackout period refers to the time period between the end of the baseline period and the beginning of the reporting period. In this analysis, it is specified to coincide with the project installation time period, meaning that the billing period that contains the project installation date is dropped from the analysis.

Analysis periods: Different portions of the analysis used different time periods of consumption data, therefore, it is useful to clearly define these time periods and where they were used. Consider a project with an installation date on a particular day d in a particular month m in a particular program year y. The year before the program year is labelled as y-1, the year prior to that as y-2 and so on, while the years following the program year are labelled y+1, y+2 etc. In all cases, the billing period that contains the project installation was dropped from the analysis. Other sections of the analysis use the following time periods:
- *Treatment and site-level matched groups:* Baseline period includes the 12 months preceding the installation billing period. Reporting period includes the 12 months following the installation billing period.
- *Future participant group:* Baseline period is the calendar year preceding the program year (Year *y-1*). Reporting period is the program year itself (Year *y*).
- Site-level consumption matching was performed using the 12 months of data immediately prior to the project installation date.
- Equivalence tests were performed using data from the previous calendar year (y-1).

Modeling

Weather Normalization: Weather normalization of billing data in CalTRACK follows certain model foundations in literature (PRISM, ASHRAE Guideline 14, IPMVP Option C and the Uniform Methods Project for Whole Home Building Analysis). Building energy use is modeled as a combination of base load, heating load, and cooling load. Heating load and cooling load are assumed to have a linear relationship with heating and cooling demand, as approximated by heating and cooling degree days, beyond particular heating and cooling balance points. A number of candidate OLS models are fit to the consumption data using different combinations of heating and cooling balance points (ranging from 30 to 90 F) and different sets of independent variables. The model with the highest adjusted R-squared that contains strictly positive coefficients is selected as the final model and used to calculate normalized energy usage.

Model Types: CalTRACK specifies a linear relationship between energy use and temperature as reflected in the building consumption profile. In the most generic case, a model would include an intercept term, a heating balance point and heating slope coefficient, and a cooling balance point and a cooling slope coefficient. Depending on the fuel a building uses for heating or cooling or its consumption patterns, models with a single temperature coefficient and balance point (i.e., heating or cooling) may be more appropriate.

Difference in Normalized Annual Consumption (DNAC): The DNAC is calculated by using two CalTRACK regression models in conjunction with Typical Meteorological Year (TMY3) weather data, as follows:
- Two models are fit to the consumption data - one model for the baseline (pre-intervention) period and one for the reporting (post-intervention) period.
- Long-term heating and cooling degree days based on TMY3 data are substituted in both regression equations to calculate the Normalized Annual Consumption (NAC) for each period. TMY3 data is maintained by NREL and includes weather averages for 1020 locations in the US between 1991-2005.
- DNAC is determined by subtracting the two NACs (DNAC = Baseline NAC - Reporting NAC).

Disaggregation: Disaggregated loads are calculated from the different components of the statistical model fit. The weather sensitive components (heating and cooling load) are calculated by multiplying the relevant model coefficients (beta_hdd or beta_cdd) by the total degree days in a normal weather year (total HDD or CDD). For each site, the total HDD or CDD can be calculated using that site's estimated degree day balance points (also an output of the model) and the temperature for its closest weather station. The base load is estimated by multiplying the intercept of the statistical model by the number of days (365 for a full year).

Savings calculation: Savings are calculated by subtracting the DNAC for either comparison group from the DNAC for the treatment group.

Savings Uncertainty: Uncertainty presented in this analysis is calculated using the ASHRAE Guideline 14 formulation for aggregating the prediction uncertainty of point estimates in a time series. It is calculated at a 90% confidence level. The total uncertainty at the site-level is calculated using the sum of squares of the baseline and reporting models. Other aggregate uncertainty values (e.g. for a portfolio or for a difference-in-differences estimate) are also aggregated using the square root of the sum of squares.

Comparison Group Generation

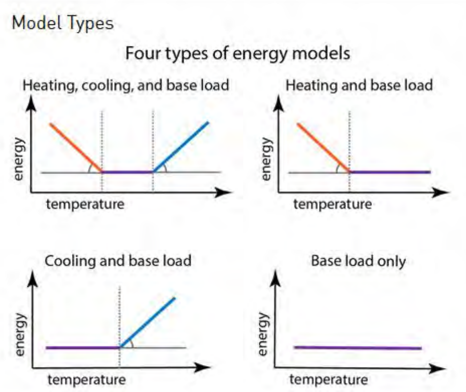
Site-level Matching: In monthly consumption matching, a comparison group is constructed by selecting n matches (n=5 in this analysis) from the comparison group pool with the shortest distance d to the treatment group customer under consideration. The pool is limited to non-participants within the same zipcode as the treatment group customer. The distance d is, in essence, a way to reduce 12 monthly consumption differences between any two customers to one metric (see Figure). In the present analysis, we selected (without replacement) five nearest neighbors for each treatment site based on the Euclidean distance of monthly consumption.

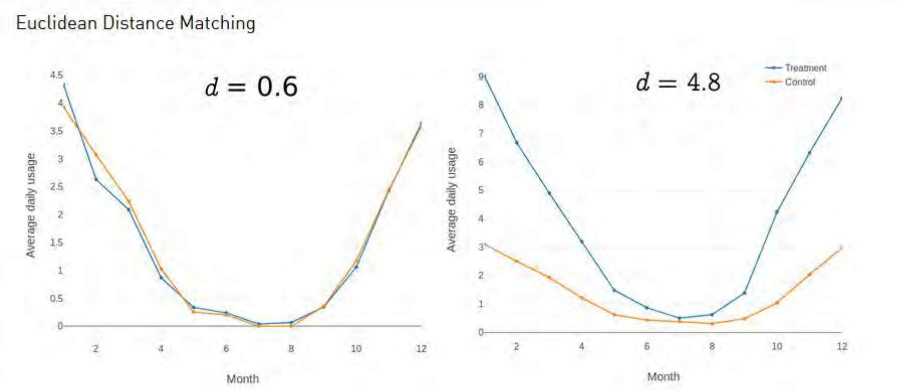
Future Participant Groups: Comparison groups comprising future participants are considered to be representative of participants in most aspects (observable and non-observable). For example, future participants are known to be eligible to receive the measure, and for some measures, they may have the same baseline equipment as the participants. Future participants have the same propensity to participate in the program as participants, thus reducing or eliminating self-selection bias, something that is otherwise difficult to control for in a quasi-experimental study. More comprehensive data is typically collected for future participants, allowing for potentially better matching and more insightful analysis. From a practical perspective, future participant groups may be difficult to construct for all measures, unless a program has been running for multiple years and is considered stable with sufficient data collection over the analysis period. Sample sizes for the comparison group may also be constrained if using future participants.

Stratified sampling is applied to future participant groups to attempt to replicate the distributions of the underlying variable (annual consumption) in the comparison group. Annual consumption of all treatment sites is first split into deciles, then a random sample is selected from within each corresponding bin in the comparison group pool of future participants.

Geographical screen: For the site-level matched group, only sites within the same zipcode as the treatment site were considered as potential comparison group matches.

Sampling method: In all cases where sampling was required from the comparison group, sampling was performed without replacement.





Impact Evaluation Report

Gas Impact of Thermostats in Program Year 2015, 2016, 2017

Result Summary

Measure: Thermostats	Program Year: 2015, 2016, 2017		Fuel: Gas	*Last Consumption Data Update:* October 1, 2019 *Last Participation Data Update:* October 1, 2019 *CalTRACK Version:* 2.0
Meter Data Filters:	DNAC: <75%	DNAC Percentile: None	Annual Consumption Percentile: Remove Top and Bottom 0.5%	
Model Filters:	Period Length: 11 Months or Longer	R-Squared: >0.5	CV(RMSE): < 1	
Metadata Filters:	Cooling Zone(s): All		Heating Fuel: Gas	

Thermostat Name: Ecobee, Ecobee, Ecobee, Ecobee	Heat Pump Baseline: All	Heating Zone(s): All	Multi Measure Filter: Single Measure Only
Heat Pump Manufacturer: All	Heat Pump Adv. Controls or Commissioning: All		

146 Treatment Meters	31 +/- 15 Therms Average Normal Year Pre-Post Difference in Consumption per Participant	4 +/- 2 % Percent Normal Year Pre-Post Difference in Consumption per Participant	723 Mean Baseline Consumption (Gas)	95% Realization Rate
725 Site-level Matched Meters	36 +/- 16 Therms Average Savings Relative to Site-level Matched Comparison Group	5 +/- 2% Percent Savings Relative to Site-level Matched Comparison Group	709 Mean Baseline Consumption (Gas)	111% Realization Rate
10.5k Future Participant Meters	45 +/- 15 Therms Average Savings Relative to Future Participant Group	6 +/- 2% Savings Relative to Future Participant Group	707 Mean Baseline Consumption (Gas)	141% Realization Rate

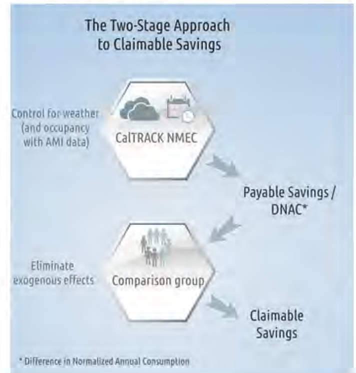
1. Introduction

This report contains the results of applying the two-stage approach (informed by the DOE's uniform methods chapter on whole building analysis) for calculating claimable savings to the selected portfolio of energy efficiency projects (see Figure). This approach begins with identification of two comparison groups for the treatment sample: (a) a site-level matched comparison group and (b) a future participant group. These groups are described below along with summary statistics (site locations, sample size, baseline consumption and baseline load disaggregation).

The CalTRACK methods are then applied to arrive at site-level savings, normalized for weather, and reflective of energy consumption changes for customers at the meter. Using a difference of differences for the treatment group with each comparison group accounts for population-level consumption changes (e.g. economic changes, rate changes, natural energy efficiency adoption etc.). The methods contained within this report are the outcome of a recent peer-reviewed study completed by Energy Trust of Oregon and Open Energy Efficiency (see "Methodology" section for more details).

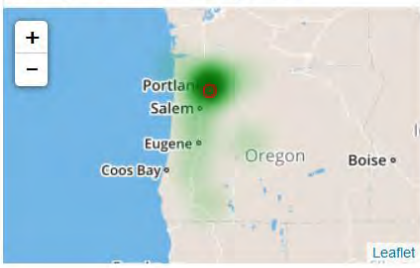
The report includes the following sections:
Result Summary - Includes the overall portfolio results
Section 1. Introduction - Overview of report and the different groups included in the analysis
Section 2. Data Preparation - Data cleaning and sample attrition
Section 3. Modeling Results - CalTRACK model outputs and Difference in Normalized Annual Consumption (DNAC) results
Section 4. Methodology - Description of methods used in this report

Two-Stage Approach



Treatment Site Locations

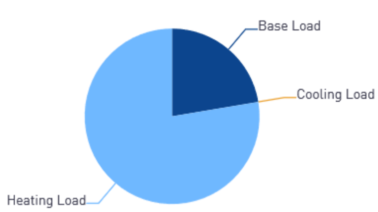
Projects | Treatment Group Centroid



39.6 miles
80% of projects lie within this distance from treatment group centroid

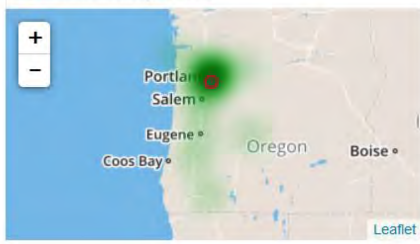
146	723
Meters	Mean Baseline Consumption (Gas)

Load Disaggregation



Site-level Matched Site Locations

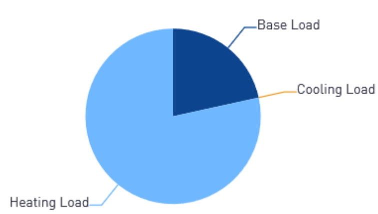
Comparison Group Centroid | Projects | Treatment Group Centroid



0.6 miles
Distance between treatment and comparison group centroids

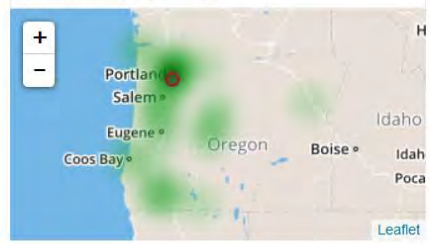
725	709
Meters	Mean Baseline Consumption (Gas)

Load Disaggregation



Future Participant Site Locations

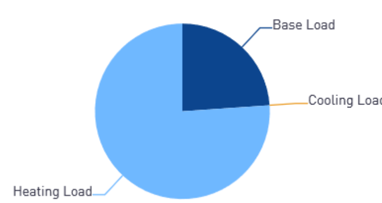
Comparison Group Centroid | Projects | Treatment Group Centroid



1.9 miles
Distance between treatment and future participant group centroids

10.5k	707
Meters	Mean Baseline Consumption (Gas)

Load Disaggregation



2. Data Preparation

Consumption data preparation and cleaning followed best practices defined in the CalTRACK 2.0 billing methods. Some key aspects of the data cleaning process are highlighted here; please see the resources section for links to more detailed documentation. The initial and final sample sizes are shown below along with the percent of the treatment population that is represented by the sample. The sample attrition table shows the impact of each filtering criterion on sample size.

5,347	146	2.7%
Meters in Treatment Population	Final Sample Size	Percent of Treatment Population Represented by Sample

Sample Attrition Table

Filter	Selected Filter Value (if applicable)	Number of Dropped Meters	Sample Size after Applying Filter
Measure: Meters associated with a particular measure in program participation data. **Year:** Program year. **Fuel:** Type of metered fuel.	Measure: Thermostats -- Year: 2015, 2016, 2017 -- Fuel: Gas	--	5,347
Meters with valid consumption data in baseline and/or reporting periods.	--	148	5,199
MultiMeasure_Filter: Meters with single/multiple measure installations in baseline and/or reporting periods.	Multi Measure Filter: Single Measure Only	3,830	1,369
HeatingFuel: Meters with a valid heating fuel that corresponds to the selected filter value.	Heating Fuel: Gas	128	1,241
HeatingZone, CoolingZone: Meters in selected heating and/or cooling climate zones.	Heating Zone: All -- Cooling Zone: All	0	1,241
Other measure-specific filters.	--	1,034	207
PeriodLength_Threshold: Meters meeting a threshold number of months of valid consumption data.	Period Length: 11 Months or Longer	37	170
Meters with at least 5 site-level matched meters from the comparison group pool.	--	19	151
DNAC_Threshold: Meters with normalized change in annual energy consumption under a specified threshold.	DNAC: <75%	4	147

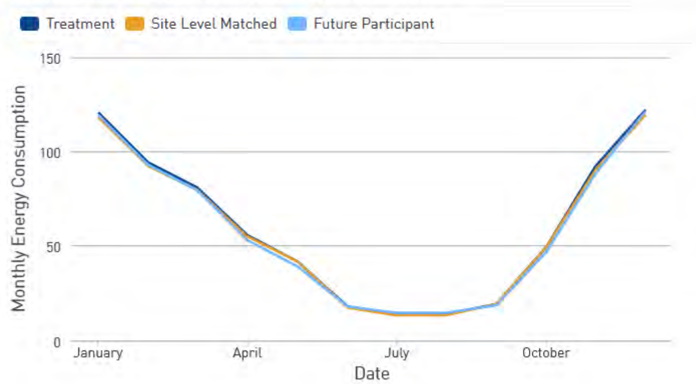
DNACPercentile_Threshold: Meters within specified percentile bands of normalized change in annual consumption.	DNAC Percentile: None	0	147
ConsumptionPercentile_Threshold: Meters within specified percentile bounds of annual energy consumption.	Annual Consumption Percentile: Remove Top and Bottom 0.5%	0	147
R2_Threshold: Meters with valid model R-squared for the baseline and reporting periods that meet a specified threshold. Models may have invalid R-squared due to data issues.	R-Squared: >0.5	1	146
CVRMSE_Threshold: Meters with valid model CV(RMSE) for the baseline and reporting periods that meet a specified threshold.	CV(RMSE): < 1	0	146

3. Modeling Results

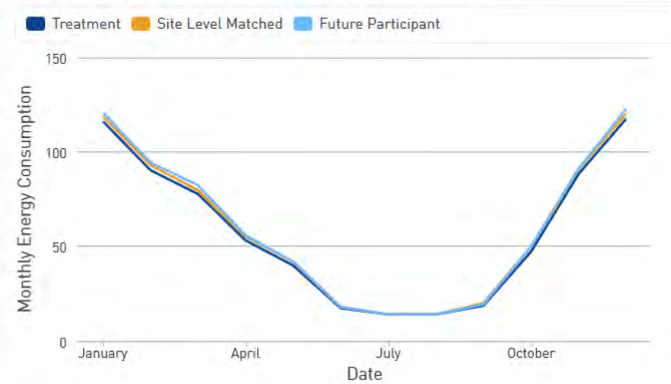
This section includes summaries of the Difference in Normalized Annual Consumption (DNAC) results for the treatment and comparison groups. The time series of monthly energy consumption illustrates the similarities and/or differences in energy consumption for the different groups in the baseline and reporting periods.

Below, you will find a breakdown of the DNAC results by group, showing the histograms of DNAC as well as the mean value expressed in raw units and as a percent of baseline annual consumption. Finally, the distribution of model types in the baseline and reporting periods are also provided as an additional layer of analysis.

Baseline Normal Year Monthly Energy Consumption

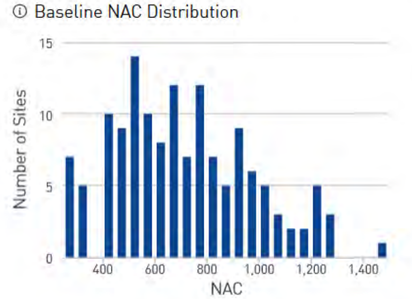


Post-Period Normal Year Monthly Energy Consumption



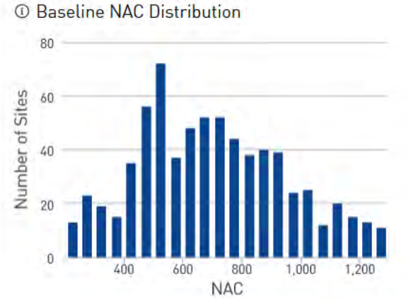
Treatment Group

ⓘ Baseline NAC Distribution



Site-level Matched Comparison Group

ⓘ Baseline NAC Distribution

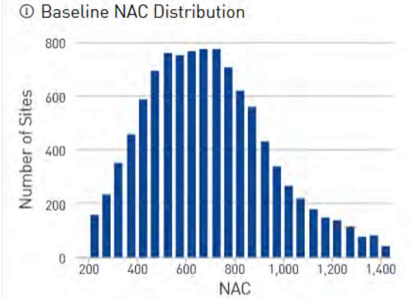


ⓘ 0.267

Annual Consumption p-value

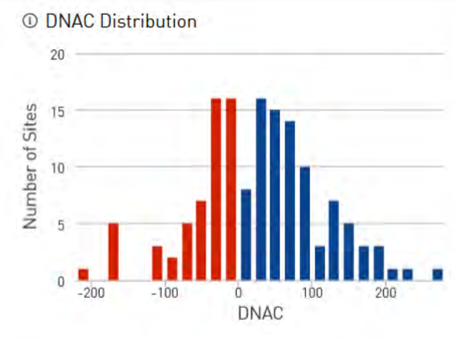
Future Participant Group

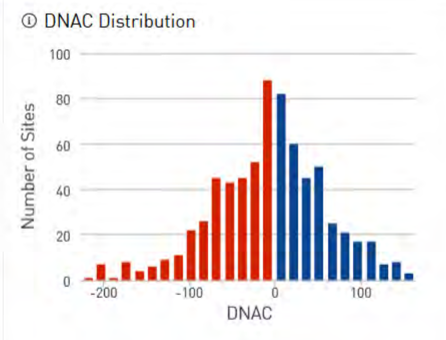
ⓘ Baseline NAC Distribution

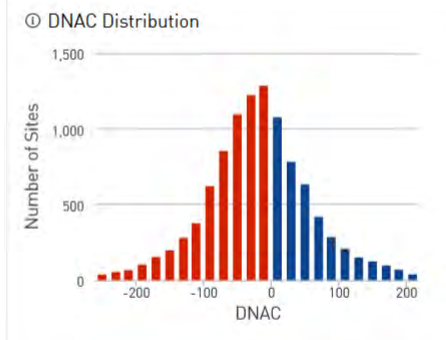


ⓘ 0.114

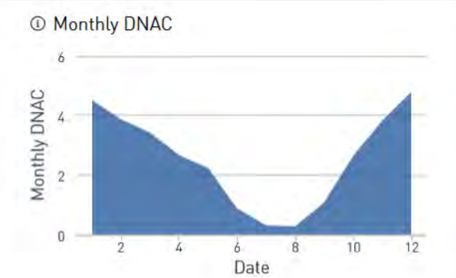
Annual Consumption p-value

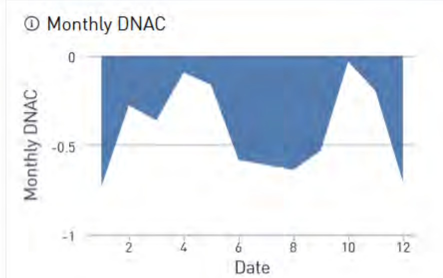


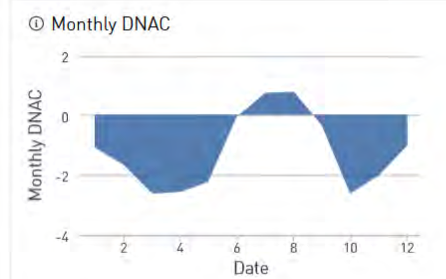


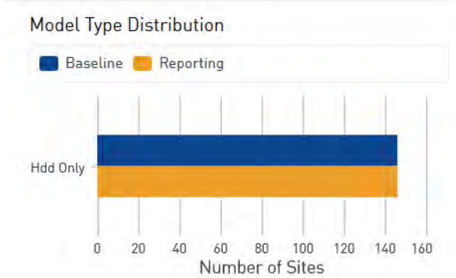


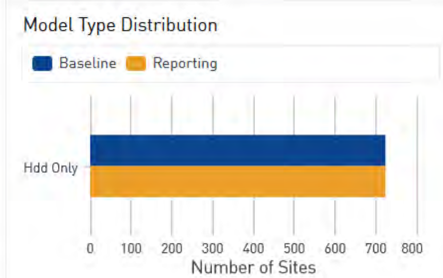
31 +/- 15 Therms	4 +/- 2 %	-5 +/- 6 Therms	-1 +/- 1 %	-15 +/- 2 Therms	-2 +/- 0 %
Average Difference in Normalized Annual Consumption per Participant	Difference in Normalized Annual Consumption as a Percent of Baseline	Average Difference in Normalized Annual Consumption per Participant	Difference in Normalized Annual Consumption as a Percent of Baseline	Average Difference in Normalized Annual Consumption per Participant	Difference in Normalized Annual Consumption as a Percent of Baseline

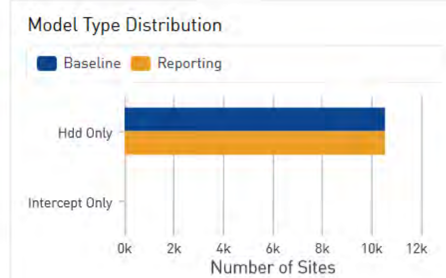




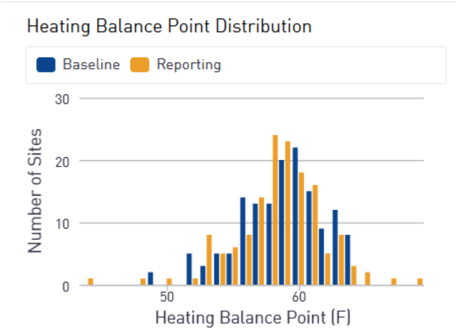


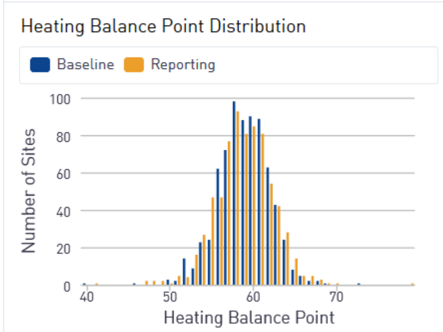


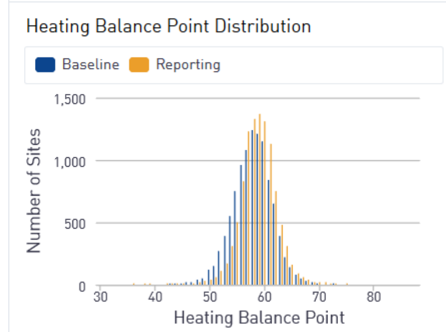


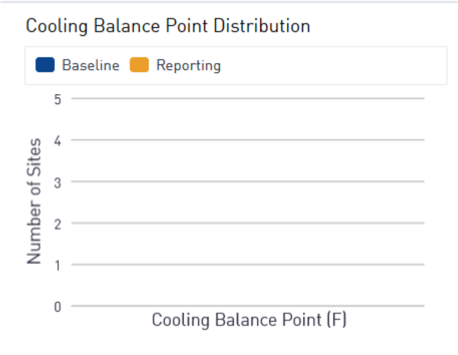


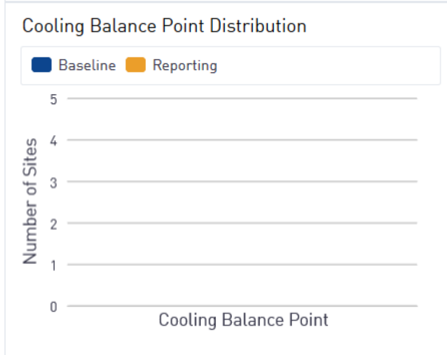
Treatment Group	Site-level Matched Comparison Group	Future Participant Group

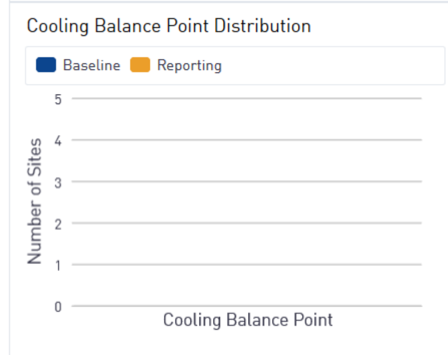












4. Methodology

CalTRACK and Comparison Group Methods
Documentation: docs.caltrack.org
Code: https://github.com/energy-market-methods/caltrack

Data Preparation

Baseline period: Since the predicted baseline may be unstable with different baseline period lengths, which may, in turn, affect calculated savings, the consensus of the CalTRACK 2.0 working group was to set the maximum baseline period at 12 months, since the year leading to the energy efficiency intervention is the most indicative of recent energy use trends and prolonging the baseline period increases the chance of other unmeasured factors affecting the baseline. In addition, CalTRACK uses a minimum 12-month baseline by default.

Blackout period: The blackout period refers to the time period between the end of the baseline period and the beginning of the reporting period. In this analysis, it is specified to coincide with the project installation time period, meaning that the billing period that contains the project installation date is dropped from the analysis.

Analysis periods: Different portions of the analysis used different time periods of consumption data, therefore, it is useful to clearly define these time periods and where they were used. Consider a project with an installation date on a particular day d in a particular month m in a particular program year y. The year before the program year is labelled as y-1, the year prior to that as y-2 and so on, while the years following the program year are labelled y+1, y+2 etc. In all cases, the billing period that contains the project installation was dropped from the analysis. Other sections of the analysis use the following time periods:
- *Treatment and site-level matched groups:* Baseline period includes the 12 months preceding the installation billing period. Reporting period includes the 12 months following the installation billing period.
- *Future participant group:* Baseline period is the calendar year preceding the program year (Year *y-1*). Reporting period is the program year itself (Year *y*).
- Site-level consumption matching was performed using the 12 months of data immediately prior to the project installation date.
- Equivalence tests were performed using data from the previous calendar year (y-1).

Modeling

Weather Normalization: Weather normalization of billing data in CalTRACK follows certain model foundations in literature (PRISM, ASHRAE Guideline 14, IPMVP Option C and the Uniform Methods Project for Whole Home Building Analysis). Building energy use is modeled as a combination of base load, heating load, and cooling load. Heating load and cooling load are assumed to have a linear relationship with heating and cooling demand, as approximated by heating and cooling degree days, beyond particular heating and cooling balance points. A number of candidate OLS models are fit to the consumption data using different combinations of heating and cooling balance points (ranging from 30 to 90 F) and different sets of independent variables. The model with the highest adjusted R-squared that contains strictly positive coefficients is selected as the final model and used to calculate normalized energy usage.

Model Types: CalTRACK specifies a linear relationship between energy use and temperature as reflected in the building consumption profile. In the most generic case, a model would include an intercept term, a heating balance point and heating slope coefficient, and a cooling balance point and a cooling slope coefficient. Depending on the fuel a building uses for heating or cooling or its consumption patterns, models with a single temperature coefficient and balance point (i.e., heating or cooling) may be more appropriate.

Difference in Normalized Annual Consumption (DNAC): The DNAC is calculated by using two CalTRACK regression models in conjunction with Typical Meteorological Year (TMY3) weather data, as follows:
- Two models are fit to the consumption data - one model for the baseline (pre-intervention) period and one for the reporting (post-intervention) period.
- Long-term heating and cooling degree days based on TMY3 data are substituted in both regression equations to calculate the Normalized Annual Consumption (NAC) for each period. TMY3 data is maintained by NREL and includes weather averages for 1020 locations in the US between 1991-2005.
- DNAC is determined by subtracting the two NACs (DNAC = Baseline NAC - Reporting NAC).

Disaggregation: Disaggregated loads are calculated from the different components of the statistical model fit. The weather sensitive components (heating and cooling load) are calculated by multiplying the relevant model coefficients (beta_hdd or beta_cdd) by the total degree days in a normal weather year (total HDD or CDD). For each site, the total HDD or CDD can be calculated using that site's estimated degree day balance points (also an output of the model) and the temperature for its closest weather station. The base load is estimated by multiplying the intercept of the statistical model by the number of days (365 for a full year).

Savings calculation: Savings are calculated by subtracting the DNAC for either comparison group from the DNAC for the treatment group.

Savings Uncertainty: Uncertainty presented in this analysis is calculated using the ASHRAE Guideline 14 formulation for aggregating the prediction uncertainty of point estimates in a time series. It is calculated at a 90% confidence level. The total uncertainty at the site-level is calculated using the sum of squares of the baseline and reporting models. Other aggregate uncertainty values (e.g. for a portfolio or for a difference-in-differences estimate) are also aggregated using the square root of the sum of squares.

Comparison Group Generation

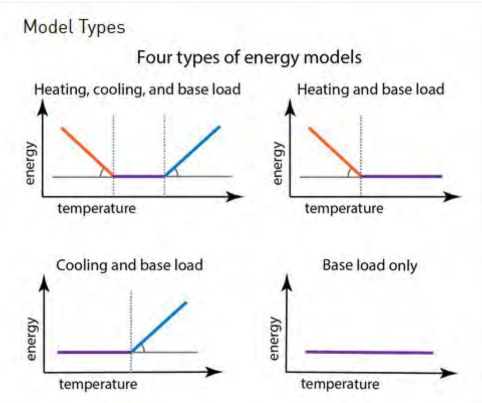
Site-level Matching: In monthly consumption matching, a comparison group is constructed by selecting n matches (n=5 in this analysis) from the comparison group pool with the shortest distance d to the treatment group customer under consideration. The pool is limited to non-participants within the same zipcode as the treatment group customer. The distance d is, in essence, a way to reduce 12 monthly consumption differences between any two customers to one metric (see Figure). In the present analysis, we selected (without replacement) five nearest neighbors for each treatment site based on the Euclidean distance of monthly consumption.

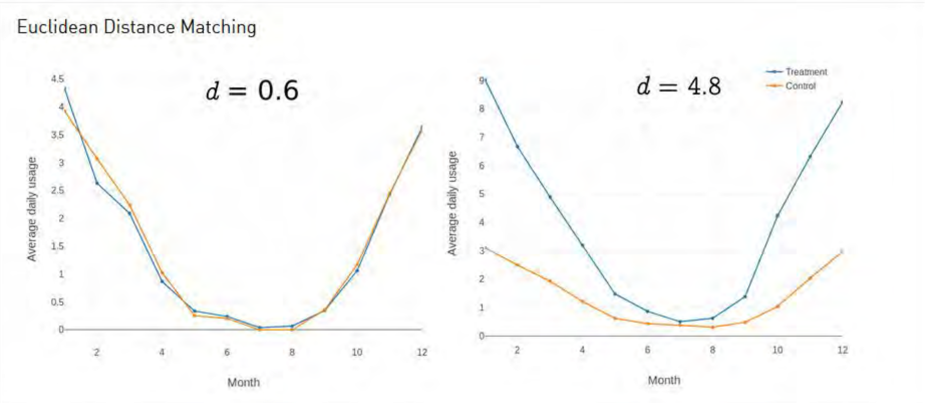
Future Participant Groups: Comparison groups comprising future participants are considered to be representative of participants in most aspects (observable and non-observable). For example, future participants are known to be eligible to receive the measure, and for some measures, they may have the same baseline equipment as the participants. Future participants have the same propensity to participate in the program as participants, thus reducing or eliminating self-selection bias, something that is otherwise difficult to control for in a quasi-experimental study. More comprehensive data is typically collected for future participants, allowing for potentially better matching and more insightful analysis. From a practical perspective, future participant groups may be difficult to construct for all measures, unless a program has been running for multiple years and is considered stable with sufficient data collection over the analysis period. Sample sizes for the comparison group may also be constrained if using future participants.

Stratified sampling is applied to future participant groups to attempt to replicate the distributions of the underlying variable (annual consumption) in the comparison group. Annual consumption of all treatment sites is first split into deciles, then a random sample is selected from within each corresponding bin in the comparison group pool of future participants.

Geographical screen: For the site-level matched group, only sites within the same zipcode as the treatment site were considered as potential comparison group matches.

Sampling method: In all cases where sampling was required from the comparison group, sampling was performed without replacement.





Impact Evaluation Report

Electricity Impact of Thermostats in Program Year 2015, 2016, 2017

Result Summary

Measure: Thermostats	Program Year: 2015, 2016, 2017		Fuel: Electricity	*Last Consumption Data Update:* October 1, 2019 *Last Participation Data Update:* October 1, 2019 *CalTRACK Version:* 2.0
Meter Data Filters:	DNAC: <75%	DNAC Percentile: None	Annual Consumption Percentile: Remove Top and Bottom 0.5%	
Model Filters:	Period Length: 11 Months or Longer	R-Squared: >0.5	CV(RMSE): < 1	
Metadata Filters:	Cooling Zone(s): All		Heating Fuel: Gas	
Thermostat Name: All; Heat Pump Baseline: All	Heating Zone(s): All		Multi Measure Filter: Single Measure Only	
Heat Pump Manufacturer: All; Heat Pump Adv. Controls or Commissioning: All				

450 Treatment Meters	283 +/- 84 kWh Average Normal Year Pre-Post Difference in Consumption per Participant	3 +/- 1 % Percent Normal Year Pre-Post Difference in Consumption per Participant	8,675 Mean Baseline Consumption (Electricity)	1,951% Realization Rate
2,200 Site-level Matched Meters	178 +/- 91 kWh Average Savings Relative to Site-level Matched Comparison Group	2 +/- 1% Percent Savings Relative to Site-level Matched Comparison Group	8,585 Mean Baseline Consumption (Electricity)	1,262% Realization Rate
6,256 Future Participant Meters	225 +/- 88 kWh Average Savings Relative to Future Participant Group	3 +/- 1% Savings Relative to Future Participant Group	8,489 Mean Baseline Consumption (Electricity)	1,552% Realization Rate

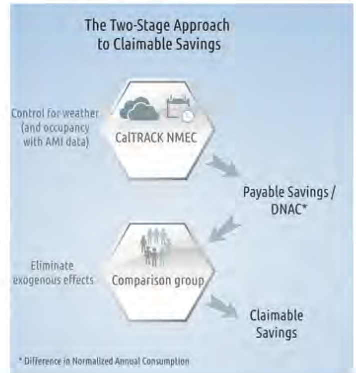
1. Introduction

This report contains the results of applying the two-stage approach (informed by the DOE's uniform methods chapter on whole building analysis) for calculating claimable savings to the selected portfolio of energy efficiency projects (see Figure). This approach begins with identification of two comparison groups for the treatment sample: (a) a site-level matched comparison group and (b) a future participant group. These groups are described below along with summary statistics (site locations, sample size, baseline consumption and baseline load disaggregation).

The CalTRACK methods are then applied to arrive at site-level savings, normalized for weather, and reflective of energy consumption changes for customers at the meter. Using a difference of differences for the treatment group with each comparison group accounts for population-level consumption changes (e.g. economic changes, rate changes, natural energy efficiency adoption etc.). The methods contained within this report are the outcome of a recent peer-reviewed study completed by Energy Trust of Oregon and Open Energy Efficiency (see "Methodology" section for more details).

The report includes the following sections:
Result Summary - Includes the overall portfolio results
Section 1. Introduction - Overview of report and the different groups included in the analysis
Section 2. Data Preparation - Data cleaning and sample attrition
Section 3. Modeling Results - CalTRACK model outputs and Difference in Normalized Annual Consumption (DNAC) results
Section 4. Methodology - Description of methods used in this report

Two-Stage Approach



Treatment Group

The treatment group consists of sites that participated in the specified energy efficiency projects in the specified program year. Only sites that installed single measures are included in the treatment group. And this group includes the subset of sites that had sufficient data quality for modeling.

Treatment Site Locations

Projects Treatment Group Centroid

22.8 miles

80% of projects lie within this distance from treatment group centroid

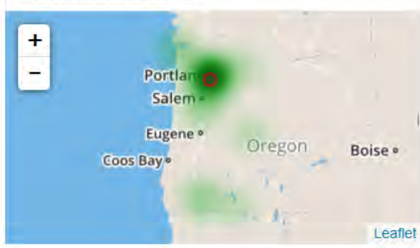
450	8,675
Meters	Mean Baseline Consumption (Electricity)

Site-level Matched Comparison Group

This group includes comparison group sites that were matched at the site-level to treatment group sites. Each treatment group site is matched to five comparison group sites from the same zipcode, but only the sites with sufficient data quality were included in the group. Matching was performed using monthly consumption in the baseline period as detailed in the Methodology section.

Site-level Matched Site Locations

Comparison Group Centroid Projects
Treatment Group Centroid



0.6 miles

Distance between treatment and comparison group centroids

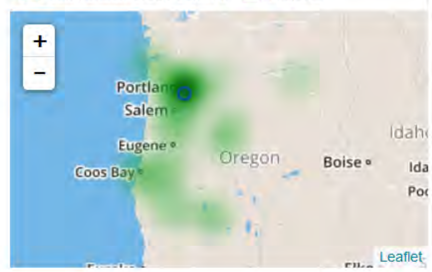
2,200	8,585
Meters	Mean Baseline Consumption (Electricity)

Future Participant Group

The pool of sites that was used to create this group was composed of sites that installed the same measure in the year following the specified program year. The final sites were selected by stratified sampling using deciles of annual energy consumption.

Future Participant Site Locations

Comparison Group Centroid Projects

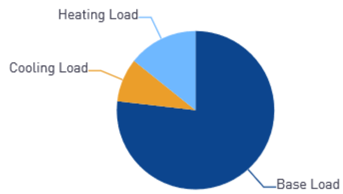


4.0 miles

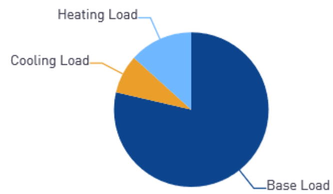
Distance between treatment and future participant group centroids

6,256	8,489
Meters	Mean Baseline Consumption (Electricity)

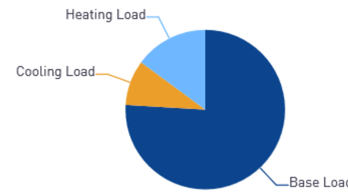
Load Disaggregation



Load Disaggregation



Load Disaggregation



2. Data Preparation

Consumption data preparation and cleaning followed best practices defined in the CalTRACK 2.0 billing methods. Some key aspects of the data cleaning process are highlighted here; please see the resources section for links to more detailed documentation. The initial and final sample sizes are shown below along with the percent of the treatment population that is represented by the sample. The sample attrition table shows the impact of each filtering criterion on sample size.

4,911	450	9.2%
Meters in Treatment Population	Final Sample Size	Percent of Treatment Population Represented by Sample

Sample Attrition Table

Filter	Selected Filter Value (if applicable)	Number of Dropped Meters	Sample Size after Applying Filter
Measure: Meters associated with a particular measure in program participation data. **Year:** Program year. **Fuel:** Type of metered fuel.	Measure: Thermostats -- Year: 2015, 2016, 2017 -- Fuel: Electricity	--	4,911
Meters with valid consumption data in baseline and/or reporting periods.	--	88	4,823
MultiMeasure_Filter: Meters with single/multiple measure installations in baseline and/or reporting periods.	Multi Measure Filter: Single Measure Only	3,559	1,264
HeatingFuel: Meters with a valid heating fuel that corresponds to the selected filter value.	Heating Fuel: Gas	210	1,054
HeatingZone, CoolingZone: Meters in selected heating and/or cooling climate zones.	Heating Zone: All -- Cooling Zone: All	0	1,054
Other measure-specific filters.	--	0	1,054
PeriodLength_Threshold: Meters meeting a threshold number of months of valid consumption data.	Period Length: 11 Months or Longer	393	661
Meters with at least 5 site-level matched meters from the comparison group pool.	--	107	554
DNAC_Threshold: Meters with normalized change in annual energy consumption under a specified threshold.	DNAC: <75%	9	548

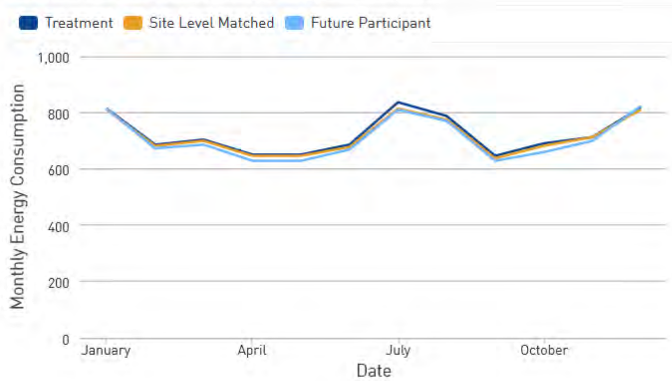
DNACPercentile_Threshold: Meters within specified percentile bands of normalized change in annual consumption.	DNAC Percentile: None	0	548
ConsumptionPercentile_Threshold: Meters within specified percentile bounds of annual energy consumption.	Annual Consumption Percentile: Remove Top and Bottom 0.5%	2	546
R2_Threshold: Meters with valid model R-squared for the baseline and reporting periods that meet a specified threshold. Models may have invalid R-squared due to data issues.	R-Squared: >0.5	96	450
CVRMSE_Threshold: Meters with valid model CV(RMSE) for the baseline and reporting periods that meet a specified threshold.	CV(RMSE): < 1	0	450

3. Modeling Results

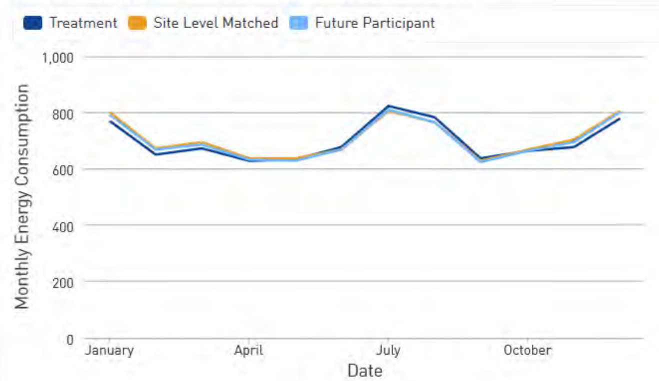
This section includes summaries of the Difference in Normalized Annual Consumption (DNAC) results for the treatment and comparison groups. The time series of monthly energy consumption illustrates the similarities and/or differences in energy consumption for the different groups in the baseline and reporting periods.

Below, you will find a breakdown of the DNAC results by group, showing the histograms of DNAC as well as the mean value expressed in raw units and as a percent of baseline annual consumption. Finally, the distribution of model types in the baseline and reporting periods are also provided as an additional layer of analysis.

Baseline Normal Year Monthly Energy Consumption

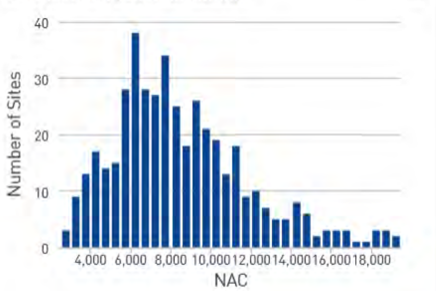


Post-Period Normal Year Monthly Energy Consumption



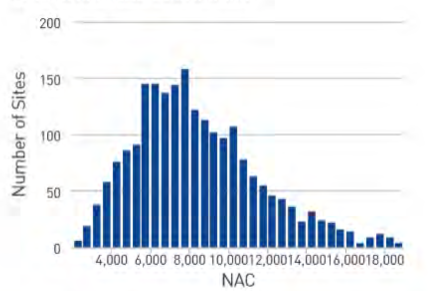
Treatment Group

Baseline NAC Distribution



Site-level Matched Comparison Group

Baseline NAC Distribution

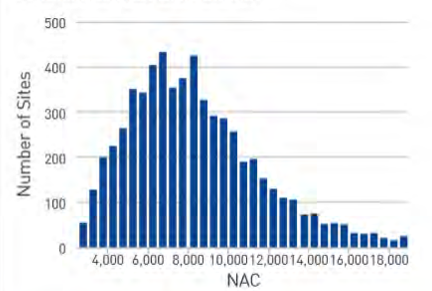


0.178

Annual Consumption p-value

Future Participant Group

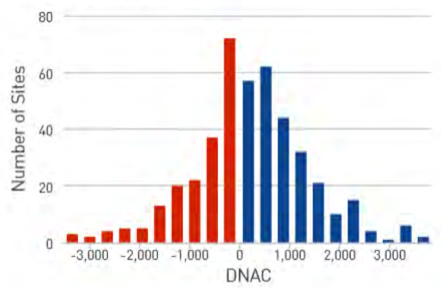
Baseline NAC Distribution



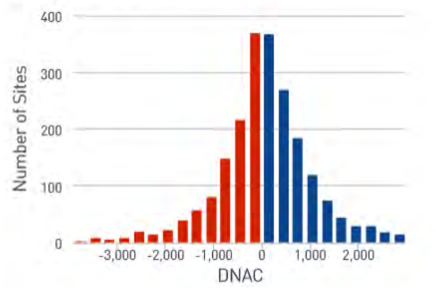
0.156

Annual Consumption p-value

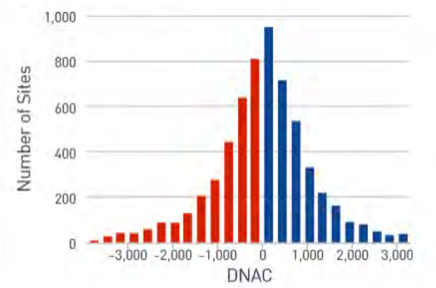
DNAC Distribution



DNAC Distribution



DNAC Distribution



283 +/- 84 kWh
Average Difference in Normalized Annual Consumption per Participant

4 +/- 1 %
Difference in Normalized Annual Consumption as a Percent of Baseline

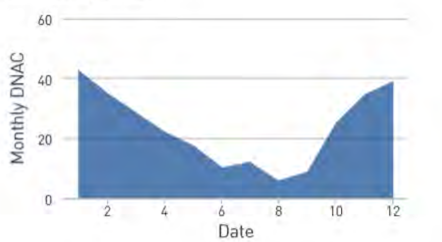
104 +/- 35 kWh
Average Difference in Normalized Annual Consumption per Participant

1 +/- 0 %
Difference in Normalized Annual Consumption as a Percent of Baseline

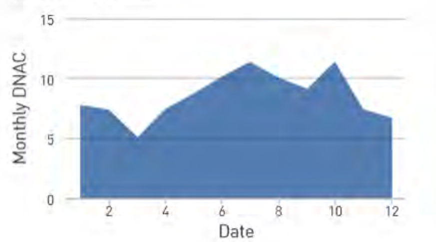
58 +/- 24 kWh
Average Difference in Normalized Annual Consumption per Participant

1 +/- 0 %
Difference in Normalized Annual Consumption as a Percent of Baseline

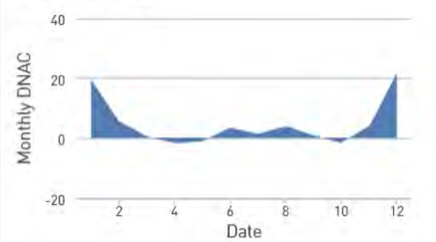
Monthly DNAC



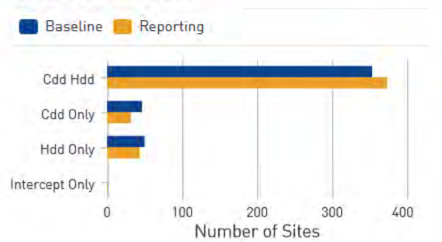
Monthly DNAC



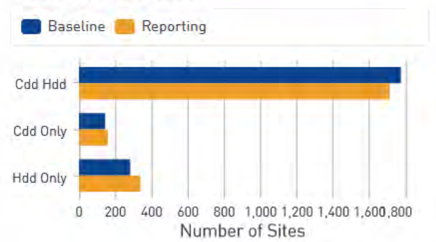
Monthly DNAC



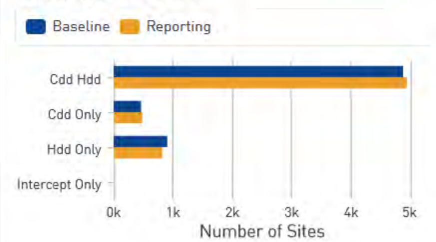
Model Type Distribution



Model Type Distribution



Model Type Distribution

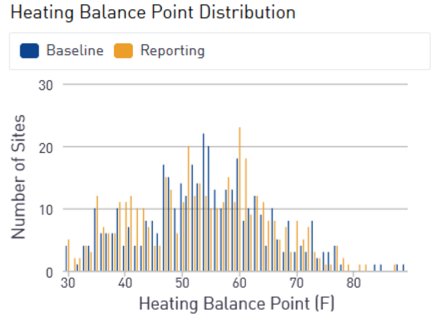


Treatment Group

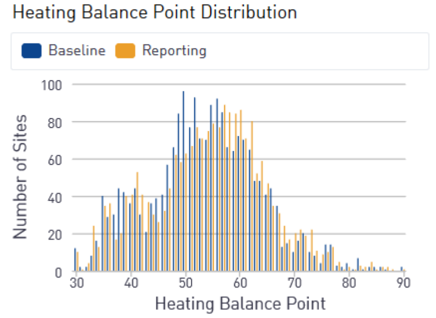
Site-level Matched Comparison Group

Future Participant Group

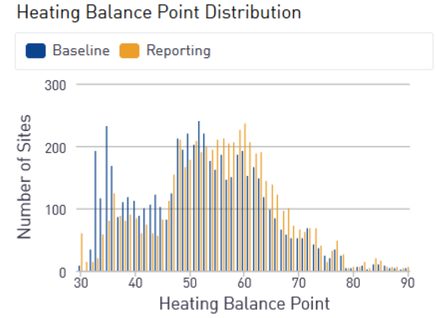
Heating Balance Point Distribution

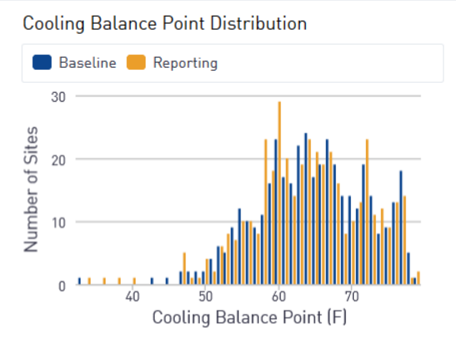


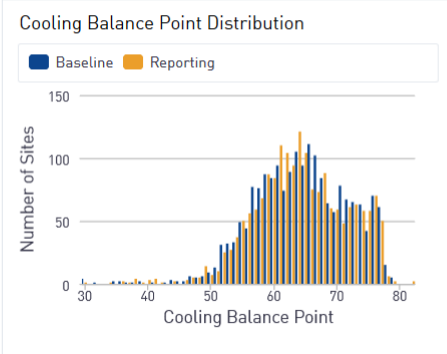
Heating Balance Point Distribution

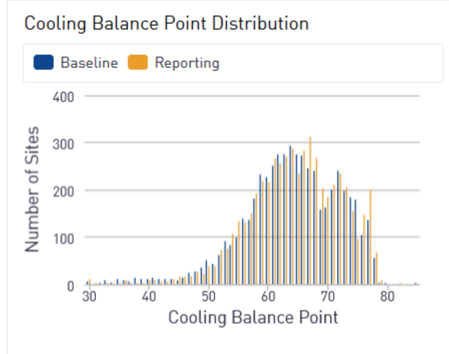


Heating Balance Point Distribution









4. Methodology

CalTRACK and Comparison Group Methods
Documentation: docs.caltrack.org
Code: https://github.com/energy-market-methods/caltrack

Data Preparation

Baseline period: Since the predicted baseline may be unstable with different baseline period lengths, which may, in turn, affect calculated savings, the consensus of the CalTRACK 2.0 working group was to set the maximum baseline period at 12 months, since the year leading to the energy efficiency intervention is the most indicative of recent energy use trends and prolonging the baseline period increases the chance of other unmeasured factors affecting the baseline. In addition, CalTRACK uses a minimum 12-month baseline by default.

Blackout period: The blackout period refers to the time period between the end of the baseline period and the beginning of the reporting period. In this analysis, it is specified to coincide with the project installation time period, meaning that the billing period that contains the project installation date is dropped from the analysis.

Analysis periods: Different portions of the analysis used different time periods of consumption data, therefore, it is useful to clearly define these time periods and where they were used. Consider a project with an installation date on a particular day d in a particular month m in a particular program year y. The year before the program year is labelled as y-1, the year prior to that as y-2 and so on, while the years following the program year are labelled y+1, y+2 etc. In all cases, the billing period that contains the project installation was dropped from the analysis. Other sections of the analysis use the following time periods:
- *Treatment and site-level matched groups:* Baseline period includes the 12 months preceding the installation billing period. Reporting period includes the 12 months following the installation billing period.
- *Future participant group:* Baseline period is the calendar year preceding the program year (Year *y-1*). Reporting period is the program year itself (Year *y*).
- Site-level consumption matching was performed using the 12 months of data immediately prior to the project installation date.
- Equivalence tests were performed using data from the previous calendar year (y-1).

Modeling

Weather Normalization: Weather normalization of billing data in CalTRACK follows certain model foundations in literature (PRISM, ASHRAE Guideline 14, IPMVP Option C and the Uniform Methods Project for Whole Home Building Analysis). Building energy use is modeled as a combination of base load, heating load, and cooling load. Heating load and cooling load are assumed to have a linear relationship with heating and cooling demand, as approximated by heating and cooling degree days, beyond particular heating and cooling balance points. A number of candidate OLS models are fit to the consumption data using different combinations of heating and cooling balance points (ranging from 30 to 90 F) and different sets of independent variables. The model with the highest adjusted R-squared that contains strictly positive coefficients is selected as the final model and used to calculate normalized energy usage.

Model Types: CalTRACK specifies a linear relationship between energy use and temperature as reflected in the building consumption profile. In the most generic case, a model would include an intercept term, a heating balance point and heating slope coefficient, and a cooling balance point and a cooling slope coefficient. Depending on the fuel a building uses for heating or cooling or its consumption patterns, models with a single temperature coefficient and balance point (i.e., heating or cooling) may be more appropriate.

Difference in Normalized Annual Consumption (DNAC): The DNAC is calculated by using two CalTRACK regression models in conjunction with Typical Meteorological Year (TMY3) weather data, as follows:
- Two models are fit to the consumption data - one model for the baseline (pre-intervention) period and one for the reporting (post-intervention) period.
- Long-term heating and cooling degree days based on TMY3 data are substituted in both regression equations to calculate the Normalized Annual Consumption (NAC) for each period. TMY3 data is maintained by NREL and includes weather averages for 1020 locations in the US between 1991-2005.
- DNAC is determined by subtracting the two NACs (DNAC = Baseline NAC - Reporting NAC).

Disaggregation: Disaggregated loads are calculated from the different components of the statistical model fit. The weather sensitive components (heating and cooling load) are calculated by multiplying the relevant model coefficients (beta_hdd or beta_cdd) by the total degree days in a normal weather year (total HDD or CDD). For each site, the total HDD or CDD can be calculated using that site's estimated degree day balance points (also an output of the model) and the temperature for its closest weather station. The base load is estimated by multiplying the intercept of the statistical model by the number of days (365 for a full year).

Savings calculation: Savings are calculated by subtracting the DNAC for either comparison group from the DNAC for the treatment group.

Savings Uncertainty: Uncertainty presented in this analysis is calculated using the ASHRAE Guideline 14 formulation for aggregating the prediction uncertainty of point estimates in a time series. It is calculated at a 90% confidence level. The total uncertainty at the site-level is calculated using the sum of squares of the baseline and reporting models. Other aggregate uncertainty values (e.g. for a portfolio or for a difference-in-differences estimate) are also aggregated using the square root of the sum of squares.

Comparison Group Generation

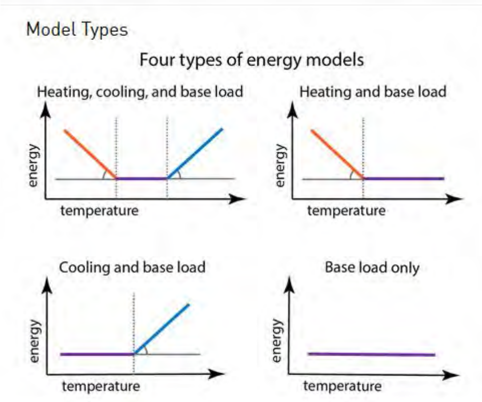
Site-level Matching: In monthly consumption matching, a comparison group is constructed by selecting n matches (n=5 in this analysis) from the comparison group pool with the shortest distance d to the treatment group customer under consideration. The pool is limited to non-participants within the same zipcode as the treatment group customer. The distance d is, in essence, a way to reduce 12 monthly consumption differences between any two customers to one metric (see Figure). In the present analysis, we selected (without replacement) five nearest neighbors for each treatment site based on the Euclidean distance of monthly consumption.

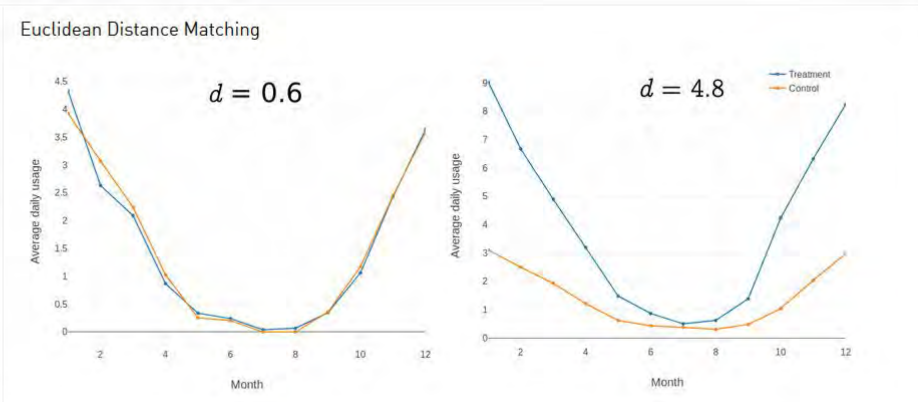
Future Participant Groups: Comparison groups comprising future participants are considered to be representative of participants in most aspects (observable and non-observable). For example, future participants are known to be eligible to receive the measure, and for some measures, they may have the same baseline equipment as the participants. Future participants have the same propensity to participate in the program as participants, thus reducing or eliminating self-selection bias, something that is otherwise difficult to control for in a quasi-experimental study. More comprehensive data is typically collected for future participants, allowing for potentially better matching and more insightful analysis. From a practical perspective, future participant groups may be difficult to construct for all measures, unless a program has been running for multiple years and is considered stable with sufficient data collection over the analysis period. Sample sizes for the comparison group may also be constrained if using future participants.

Stratified sampling is applied to future participant groups to attempt to replicate the distributions of the underlying variable (annual consumption) in the comparison group. Annual consumption of all treatment sites is first split into deciles, then a random sample is selected from within each corresponding bin in the comparison group pool of future participants.

Geographical screen: For the site-level matched group, only sites within the same zipcode as the treatment site were considered as potential comparison group matches.

Sampling method: In all cases where sampling was required from the comparison group, sampling was performed without replacement.





Impact Evaluation Report

Electricity Impact of Thermostats in Program Year 2015, 2016, 2017

Result Summary

Measure: Thermostats	Program Year: 2015, 2016, 2017	Fuel: Electricity

Last Consumption Data Update: October 1, 2019

Last Participation Data Update: October 1, 2019

CalTRACK Version: 2.0

Meter Data Filters:	DNAC: <75%	DNAC Percentile: None	Annual Consumption Percentile: Remove Top and Bottom 0.5%
Model Filters:	Period Length: 11 Months or Longer	R-Squared: >0.5	CV(RMSE): < 1
Metadata Filters:	Cooling Zone(s): All		Heating Fuel: Electricity

Thermostat Name: All	Heat Pump Baseline: All	Heating Zone(s): All	Multi Measure Filter: Single Measure Only
Heat Pump Manufacturer: All	Heat Pump Adv. Controls or Commissioning: All		

77 Treatment Meters	-197 +/- 407 kWh Average Normal Year Pre-Post Difference in Consumption per Participant	-2 +/- 3 % Percent Normal Year Pre-Post Difference in Consumption per Participant	12,563 Mean Baseline Consumption (Electricity)	-61% Realization Rate
384 Site-level Matched Meters	-317 +/- 441 kWh Average Savings Relative to Site-level Matched Comparison Group	-3 +/- 4% Percent Savings Relative to Site-level Matched Comparison Group	12,507 Mean Baseline Consumption (Electricity)	-99% Realization Rate
2,007 Future Participant Meters	-72 +/- 414 kWh Average Savings Relative to Future Participant Group	-1 +/- 3% Savings Relative to Future Participant Group	13,684 Mean Baseline Consumption (Electricity)	-23% Realization Rate

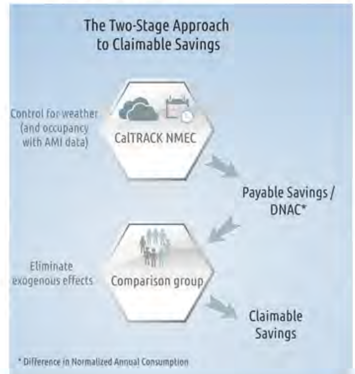
1. Introduction

This report contains the results of applying the two-stage approach (informed by the DOE's uniform methods chapter on whole building analysis) for calculating claimable savings to the selected portfolio of energy efficiency projects (see Figure). This approach begins with identification of two comparison groups for the treatment sample: (a) a site-level matched comparison group and (b) a future participant group. These groups are described below along with summary statistics (site locations, sample size, baseline consumption and baseline load disaggregation).

The CalTRACK methods are then applied to arrive at site-level savings, normalized for weather, and reflective of energy consumption changes for customers at the meter. Using a difference of differences for the treatment group with each comparison group accounts for population-level consumption changes (e.g. economic changes, rate changes, natural energy efficiency adoption etc.). The methods contained within this report are the outcome of a recent peer-reviewed study completed by Energy Trust of Oregon and Open Energy Efficiency (see "Methodology" section for more details).

The report includes the following sections:
Result Summary - Includes the overall portfolio results
Section 1. Introduction - Overview of report and the different groups included in the analysis
Section 2. Data Preparation - Data cleaning and sample attrition
Section 3. Modeling Results - CalTRACK model outputs and Difference in Normalized Annual Consumption (DNAC) results
Section 4. Methodology - Description of methods used in this report

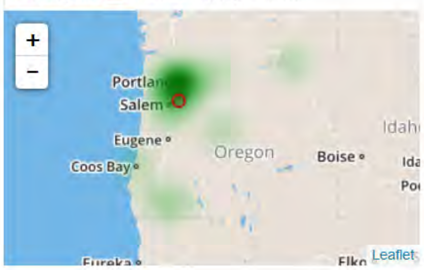
Two-Stage Approach



Treatment Group
The treatment group consists of sites that participated in the specified energy efficiency projects in the specified program year. Only sites that installed single measures are included in the treatment group. And this group includes the subset of sites that had sufficient data quality for modeling.

Treatment Site Locations

Projects | Treatment Group Centroid



180.6 miles

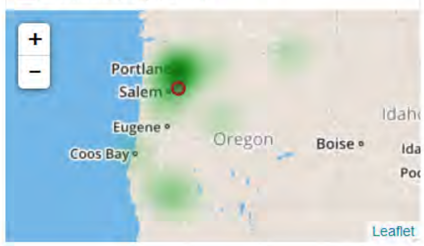
80% of projects lie within this distance from treatment group centroid

77	12,563
Meters	Mean Baseline Consumption (Electricity)

Site-level Matched Comparison Group
This group includes comparison group sites that were matched at the site-level to treatment group sites. Each treatment group site is matched to five comparison group sites from the same zipcode, but only the sites with sufficient data quality were included in the group. Matching was performed using monthly consumption in the baseline period as detailed in the Methodology section.

Site-level Matched Site Locations

Comparison Group Centroid | Projects | Treatment Group Centroid



2.0 miles

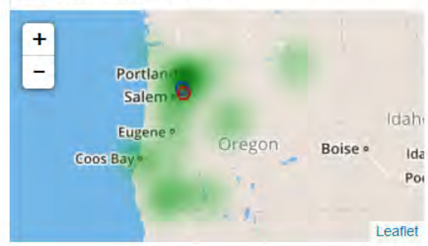
Distance between treatment and comparison group centroids

384	12,507
Meters	Mean Baseline Consumption (Electricity)

Future Participant Group
The pool of sites that was used to create this group was composed of sites that installed the same measure in the year following the specified program year. The final sites were selected by stratified sampling using deciles of annual energy consumption.

Future Participant Site Locations

Comparison Group Centroid | Projects | Treatment Group Centroid

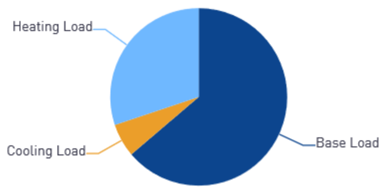


9.6 miles

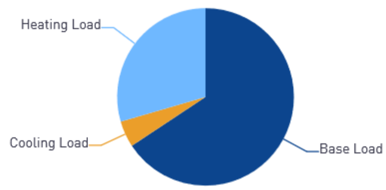
Distance between treatment and future participant group centroids

2,007	13,684
Meters	Mean Baseline Consumption (Electricity)

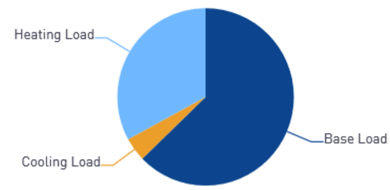
Load Disaggregation



Load Disaggregation



Load Disaggregation



2. Data Preparation

Consumption data preparation and cleaning followed best practices defined in the CalTRACK 2.0 billing methods. Some key aspects of the data cleaning process are highlighted here; please see the resources section for links to more detailed documentation. The initial and final sample sizes are shown below along with the percent of the treatment population that is represented by the sample. The sample attrition table shows the impact of each filtering criterion on sample size.

4,911	77	1.6%
Meters in Treatment Population	Final Sample Size	Percent of Treatment Population Represented by Sample

Sample Attrition Table

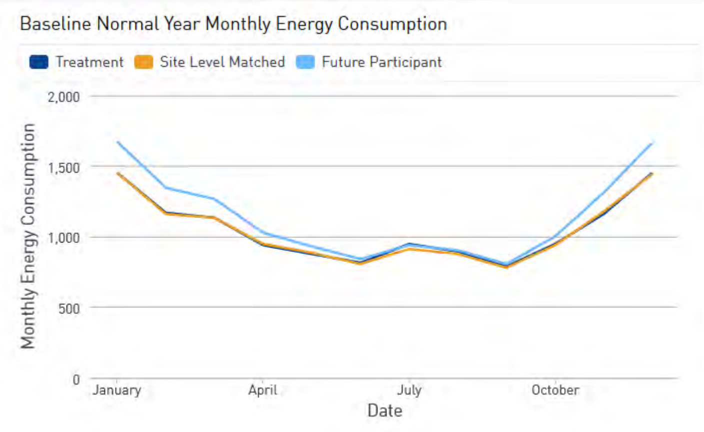
Filter	Selected Filter Value (if applicable)	Number of Dropped Meters	Sample Size after Applying Filter
Measure: Meters associated with a particular measure in program participation data. **Year:** Program year. **Fuel:** Type of metered fuel.	Measure: Thermostats -- Year: 2015, 2016, 2017 -- Fuel: Electricity	--	4,911
Meters with valid consumption data in baseline and/or reporting periods.	--	88	4,823
MultiMeasure_Filter: Meters with single/multiple measure installations in baseline and/or reporting periods.	Multi Measure Filter: Single Measure Only	3,559	1,264
HeatingFuel: Meters with a valid heating fuel that corresponds to the selected filter value.	Heating Fuel: Electricity	1,095	169
HeatingZone, CoolingZone: Meters in selected heating and/or cooling climate zones.	Heating Zone: All -- Cooling Zone: All	0	169
Other measure-specific filters.	--	0	169
PeriodLength_Threshold: Meters meeting a threshold number of months of valid consumption data.	Period Length: 11 Months or Longer	72	97
Meters with at least 5 site-level matched meters from the comparison group pool.	--	9	88
DNAC_Threshold: Meters with normalized change in annual energy consumption under a specified threshold.	DNAC: <75%	3	85

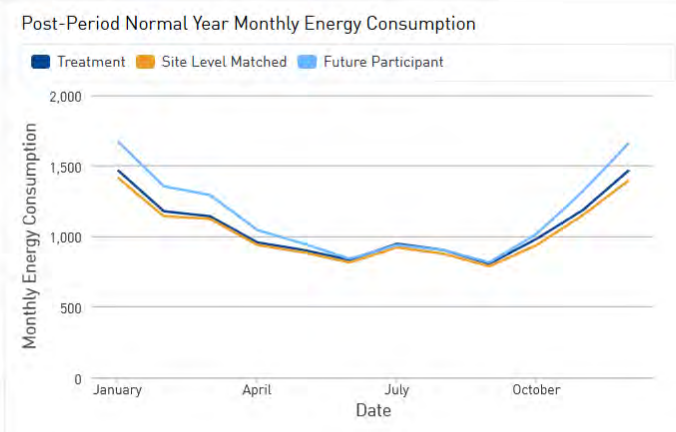
DNACPercentile_Threshold: Meters within specified percentile bands of normalized change in annual consumption.	DNAC Percentile: None	0	85
ConsumptionPercentile_Threshold: Meters within specified percentile bounds of annual energy consumption.	Annual Consumption Percentile: Remove Top and Bottom 0.5%	0	85
R2_Threshold: Meters with valid model R-squared for the baseline and reporting periods that meet a specified threshold. Models may have invalid R-squared due to data issues.	R-Squared: >0.5	8	77
CVRMSE_Threshold: Meters with valid model CV(RMSE) for the baseline and reporting periods that meet a specified threshold.	CV(RMSE): < 1	0	77

3. Modeling Results

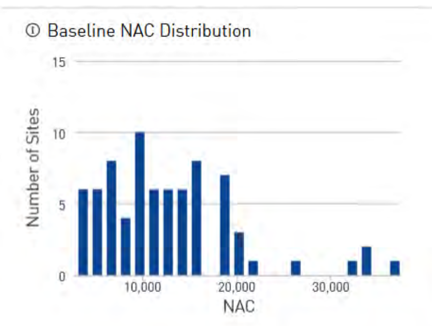
This section includes summaries of the Difference in Normalized Annual Consumption (DNAC) results for the treatment and comparison groups. The time series of monthly energy consumption illustrates the similarities and/or differences in energy consumption for the different groups in the baseline and reporting periods.

Below, you will find a breakdown of the DNAC results by group, showing the histograms of DNAC as well as the mean value expressed in raw units and as a percent of baseline annual consumption. Finally, the distribution of model types in the baseline and reporting periods are also provided as an additional layer of analysis.

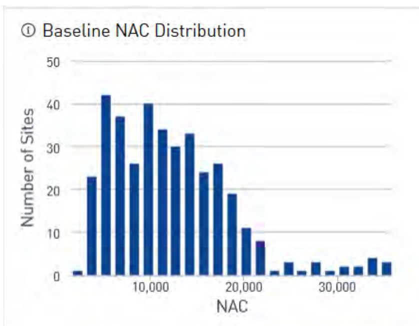




Treatment Group



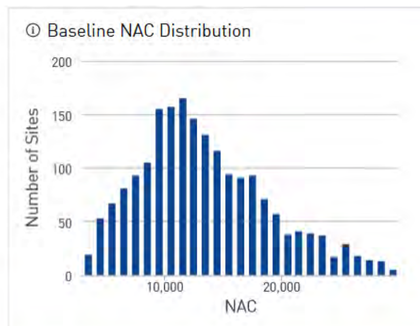
Site-level Matched Comparison Group



0.285

Annual Consumption p-value

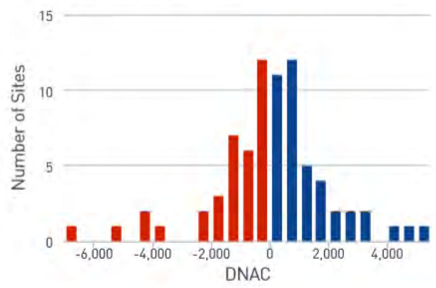
Future Participant Group



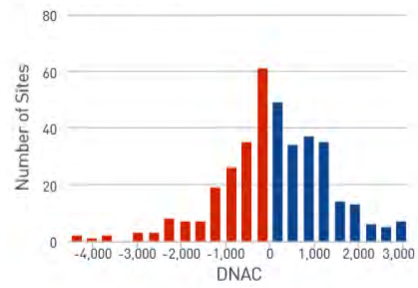
0.237

Annual Consumption p-value

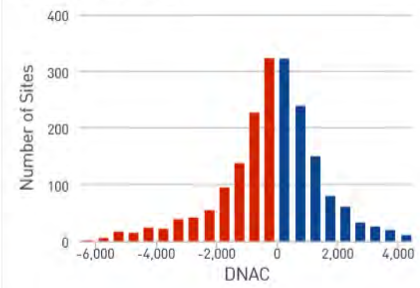
DNAC Distribution



DNAC Distribution

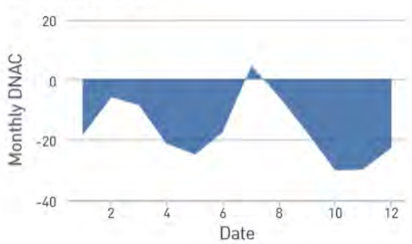


DNAC Distribution

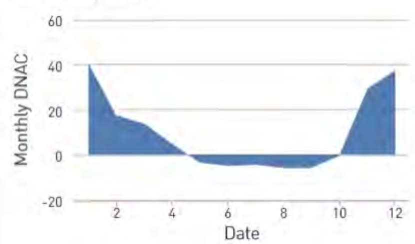


-197 +/- 407 kWh	-1 +/- 3 %	119 +/- 169 kWh	1 +/- 1 %	-125 +/- 74 kWh	-1 +/- 1 %
Average Difference in Normalized Annual Consumption per Participant	Difference in Normalized Annual Consumption as a Percent of Baseline	Average Difference in Normalized Annual Consumption per Participant	Difference in Normalized Annual Consumption as a Percent of Baseline	Average Difference in Normalized Annual Consumption per Participant	Difference in Normalized Annual Consumption as a Percent of Baseline

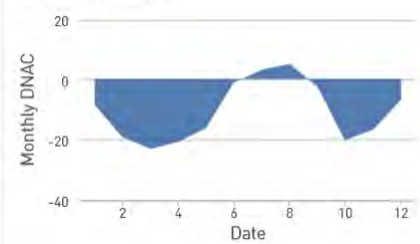
Monthly DNAC



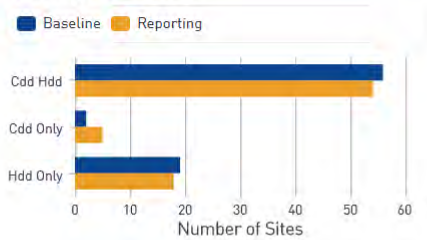
Monthly DNAC



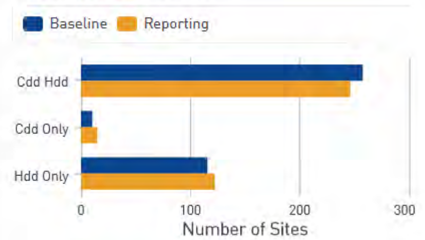
Monthly DNAC



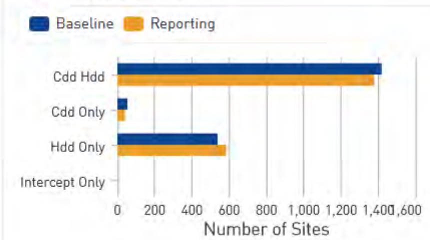
Model Type Distribution



Model Type Distribution

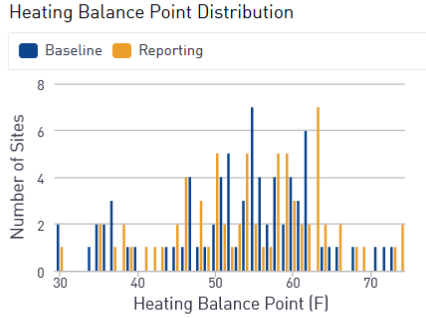


Model Type Distribution



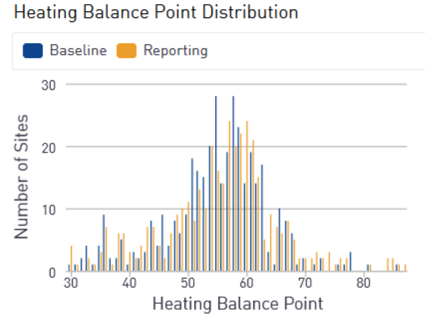
Treatment Group

Heating Balance Point Distribution



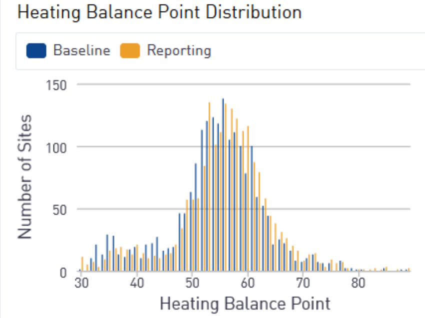
Site-level Matched Comparison Group

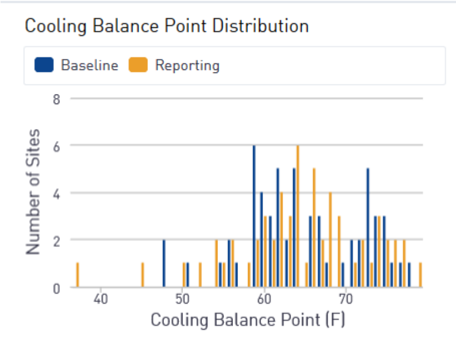
Heating Balance Point Distribution

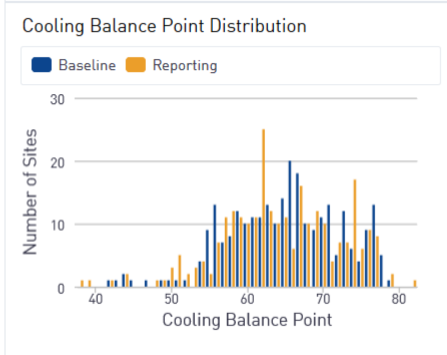


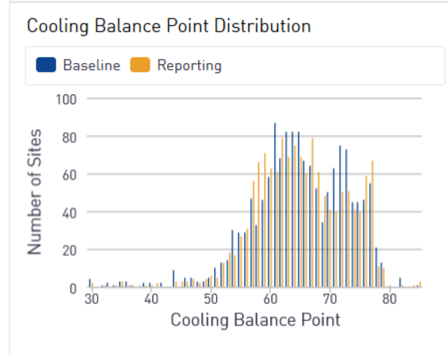
Future Participant Group

Heating Balance Point Distribution









4. Methodology

CalTRACK and Comparison Group Methods
Documentation: docs.caltrack.org
Code: https://github.com/energy-market-methods/caltrack

Data Preparation

Baseline period: Since the predicted baseline may be unstable with different baseline period lengths, which may, in turn, affect calculated savings, the consensus of the CalTRACK 2.0 working group was to set the maximum baseline period at 12 months, since the year leading to the energy efficiency intervention is the most indicative of recent energy use trends and prolonging the baseline period increases the chance of other unmeasured factors affecting the baseline. In addition, CalTRACK uses a minimum 12-month baseline by default.

Blackout period: The blackout period refers to the time period between the end of the baseline period and the beginning of the reporting period. In this analysis, it is specified to coincide with the project installation time period, meaning that the billing period that contains the project installation date is dropped from the analysis.

Analysis periods: Different portions of the analysis used different time periods of consumption data, therefore, it is useful to clearly define these time periods and where they were used. Consider a project with an installation date on a particular day d in a particular month m in a particular program year y. The year before the program year is labelled as y-1, the year prior to that as y-2 and so on, while the years following the program year are labelled y+1, y+2 etc. In all cases, the billing period that contains the project installation was dropped from the analysis. Other sections of the analysis use the following time periods:
- *Treatment and site-level matched groups:* Baseline period includes the 12 months preceding the installation billing period. Reporting period includes the 12 months following the installation billing period.
- *Future participant group:* Baseline period is the calendar year preceding the program year (Year *y-1*). Reporting period is the program year itself (Year *y*).
- Site-level consumption matching was performed using the 12 months of data immediately prior to the project installation date.
- Equivalence tests were performed using data from the previous calendar year (y-1).

Modeling

Weather Normalization: Weather normalization of billing data in CalTRACK follows certain model foundations in literature (PRISM, ASHRAE Guideline 14, IPMVP Option C and the Uniform Methods Project for Whole Home Building Analysis). Building energy use is modeled as a combination of base load, heating load, and cooling load. Heating load and cooling load are assumed to have a linear relationship with heating and cooling demand, as approximated by heating and cooling degree days, beyond particular heating and cooling balance points. A number of candidate OLS models are fit to the consumption data using different combinations of heating and cooling balance points (ranging from 30 to 90 F) and different sets of independent variables. The model with the highest adjusted R-squared that contains strictly positive coefficients is selected as the final model and used to calculate normalized energy usage.

Model Types: CalTRACK specifies a linear relationship between energy use and temperature as reflected in the building consumption profile. In the most generic case, a model would include an intercept term, a heating balance point and heating slope coefficient, and a cooling balance point and a cooling slope coefficient. Depending on the fuel a building uses for heating or cooling or its consumption patterns, models with a single temperature coefficient and balance point (i.e., heating or cooling) may be more appropriate.

Difference in Normalized Annual Consumption (DNAC): The DNAC is calculated by using two CalTRACK regression models in conjunction with Typical Meteorological Year (TMY3) weather data, as follows:
- Two models are fit to the consumption data - one model for the baseline (pre-intervention) period and one for the reporting (post-intervention) period.
- Long-term heating and cooling degree days based on TMY3 data are substituted in both regression equations to calculate the Normalized Annual Consumption (NAC) for each period. TMY3 data is maintained by NREL and includes weather averages for 1020 locations in the US between 1991-2005.
- DNAC is determined by subtracting the two NACs (DNAC = Baseline NAC - Reporting NAC).

Disaggregation: Disaggregated loads are calculated from the different components of the statistical model fit. The weather sensitive components (heating and cooling load) are calculated by multiplying the relevant model coefficients (beta_hdd or beta_cdd) by the total degree days in a normal weather year (total HDD or CDD). For each site, the total HDD or CDD can be calculated using that site's estimated degree day balance points (also an output of the model) and the temperature for its closest weather station. The base load is estimated by multiplying the intercept of the statistical model by the number of days (365 for a full year).

Savings calculation: Savings are calculated by subtracting the DNAC for either comparison group from the DNAC for the treatment group.

Savings Uncertainty: Uncertainty presented in this analysis is calculated using the ASHRAE Guideline 14 formulation for aggregating the prediction uncertainty of point estimates in a time series. It is calculated at a 90% confidence level. The total uncertainty at the site-level is calculated using the sum of squares of the baseline and reporting models. Other aggregate uncertainty values (e.g. for a portfolio or for a difference-in-differences estimate) are also aggregated using the square root of the sum of squares.

Comparison Group Generation

Site-level Matching: In monthly consumption matching, a comparison group is constructed by selecting n matches (n=5 in this analysis) from the comparison group pool with the shortest distance d to the treatment group customer under consideration. The pool is limited to non-participants within the same zipcode as the treatment group customer. The distance d is, in essence, a way to reduce 12 monthly consumption differences between any two customers to one metric (see Figure). In the present analysis, we selected (without replacement) five nearest neighbors for each treatment site based on the Euclidean distance of monthly consumption.

Future Participant Groups: Comparison groups comprising future participants are considered to be representative of participants in most aspects (observable and non-observable). For example, future participants are known to be eligible to receive the measure, and for some measures, they may have the same baseline equipment as the participants. Future participants have the same propensity to participate in the program as participants, thus reducing or eliminating self-selection bias, something that is otherwise difficult to control for in a quasi-experimental study. More comprehensive data is typically collected for future participants, allowing for potentially better matching and more insightful analysis. From a practical perspective, future participant groups may be difficult to construct for all measures, unless a program has been running for multiple years and is considered stable with sufficient data collection over the analysis period. Sample sizes for the comparison group may also be constrained if using future participants.

Stratified sampling is applied to future participant groups to attempt to replicate the distributions of the underlying variable (annual consumption) in the comparison group. Annual consumption of all treatment sites is first split into deciles, then a random sample is selected from within each corresponding bin in the comparison group pool of future participants.

Geographical screen: For the site-level matched group, only sites within the same zipcode as the treatment site were considered as potential comparison group matches.

Sampling method: In all cases where sampling was required from the comparison group, sampling was performed without replacement.