November 7, 2016

To: ConnectedThermostats@energystar.gov

From: Nest Labs
contact: Michael Blasnik mblasnik@nestlabs.com

RE: comments on EPA Energy Star Connected Thermostats "Draft 2 Method to Demonstrate Fields Savings Rev. Sep-2016"

Thank you for the opportunity to provide these comments on the draft document. Nest's comments are listed below in order of importance, with some later comments perhaps already having been addressed via stakeholder conversations since the document was initially released.

1) HVAC runtime model selection

The metric calculation has been employing three different methods for statistically estimating how HVAC runtime depends on indoor and outdoor temperatures. Two of the methods are based on an approach similar to degree-days and the third is based on degree-hours. Modeling results for a thermostat are considered sufficiently reliable to include in the performance metric based on the quality of the fit and the estimated "temperature float" -- requiring a CV(RMSE) less than 0.6 and a temperature float between 0°F and 25°F.

Nest ran the metric software on more than 2,000 thermostats and found that the hourly model provided the best fit to the data. The hourly model had more thermostat results classified as reliable -- especially in the heating season (about 3% more) and in the hot dry climate zone (about 30% more). This finding makes sense because diurnal temperature swings are largest in the hot dry region and an hourly analysis should better reflect the impacts of cold nights with warm days.

In addition to having more results classified as reliable, the hourly model also produced more reasonable estimates of the temperature float caused by solar and internal gains. The daily delta T model estimated a temperature float of about 8.5°F in the heating season and 12°F in the cooling season -- a surprisingly large difference between seasons. The hourly model estimated similar values of temperature float -- between 9°F and 10°F -- in both seasons. These values are more consistent with expectations than the daily model results.

Because the hourly model provides more results that pass reliability screens and also more reasonable values of the temperature float, Nest believes that the hourly model should be the approach used in the final metric.

2) Required Sample Sizes

On page 3 starting on line 108, it says "*Noting that the software tool will exclude certain CTs from the assessment of mean savings, ensure that the output file includes at least 100 installed CT Products in each of the five EIA climate zones for both heating and cooling. If not, repeat the procedure from step 4) 2f with additional CT Products in the data set.*"

This approach of requiring a specific number of thermostats with usable results and then allowing each vendor to select whatever sample sizes they want makes the procedure less consistent and repeatable and more open to potential gaming. Nest recommends that the standard require a specific sample size per climate region on the input file for running the metric. Based on our analysis thus far, Nest suggests the fixed sample size per region be set at least at 200 thermostats. If an allowance needs to be made for products with a smaller installed base, then the standard could simply require sampling all eligible units available (i.e. analyze a census of units in any climate region that doesn't meet the requirement).

3) Calculation of Confidence Interval on National Average Savings

The draft standard is based on the lower bound of the 95% confidence interval on the national average savings surpassing the required threshold values. But this lower bound is being calculated as the weighted average of the lower bounds for each climate region's average savings. That calculation will result in too wide a confidence interval because the sample sizes for each region must be smaller than the overall national sample and therefore the standard errors will be larger (the standard error is inversely proportional to the square root of the sample size). Larger standard errors mean wider confidence intervals. In a simple example, if the savings in each region were 10% and the standard deviation of savings were also 10% and each region had a sample of 100 homes, then the lower bounds on the 95% confidence intervals would be 8% [10% - 1.96 * 10%/sqrt(100) = 8.04%) in each region and so the weighted average would be 8.0%. But the actual lower bound of the 95% confidence on the national weighted average savings would be 9.1% [10% - 1.96 * 10%/sqrt(500) = 9.12%].

4) Selecting Random Samples

The description of how the random samples will be selected from the list of thermostats appears to have some ambiguity. The specific wording (p.2 lines 58-59) is "*sort each climate zone metadata file by the unique thermostat_id assigned to each thermostat*".* It does not appear that the thermostat_id is unambiguously defined. A vendor could potentially assign thermostat_id values in order to select a specific sample. The thermostat_id should be clearly defined (e.g., the thermostat serial number or a direct mapping of sorted serial_numbers to new ID numbers).

5) Screening of results - trimming tails after other screens

The most recent version of the metric code applies the trimming of the top and bottom 1% of the savings values to the entire unscreened sample and not to the subset that passed the basic model quality screens. It is more common to only apply trimming of outlier values after the basic model screening has been applied. It appears that EPA agrees with this sequential approach to the trimming and so Nest would just like re-iterate our support for that approach.

6) Sample Frame Attrition Report

On page 2, starting on line 44, the document lists five reasons why a thermostat should be removed from the list of thermostats to be sampled -- a null or invalid ID, an excluded HVAC system type, a bad zip code, or not having data during the target date range. The results from this process should be summarized and reported as part of the compliance procedure so that unusual removal rates can potentially be explored.