EPA ENERGY STAR Connected Thermostats

Stakeholder working meeting

Connected Thermostat Field Savings Metric

10/09/2015

Agenda

- Introduction – anyone new joining the call?

- Software module update

- How to choose a representative sample for metric calculation

- Introduction of possible alternative path to show savings

- Thermostat device testing

<#>

Attendees

- Abigail Daken, EPA
- Doug Frazee, ICF International on behalf of EPA
- Jennifer Kulp, ICF International on behalf of EPA
- Alan Meier, Lawrence Berkeley National Laboratories
- Ethan Goldman, VEIC
- Nick Lange, VEIC
- Michael Blasnik, Nest Labs
- Dave Cassano, Nest Labs
- Raj Shah, Carrier
- Phil Ngo, Impact Labs
- Brent Huchuk, Ecobee
- Michael Siemann, Weatherbug Home

Attendees Continued

- Laurie Sobczak, Comverge
- Alex Boesenberg, NEMA
- Matt Golden, Open EE
- Ed Pike, Energy Solutions on behalf CA IOUs
- Ford Garberson, Ecofactor

<#>

Software Modules Alpha Release

- Update from OEE or Doug on progress in last 2 weeks
  - New version of software to be released this month
  - Will be installable using normal tip install, should take care of dependencies
  - Includes input format updates (e.g. only daily run time, consolidation of heating columns and cooling columns, separate heating and cooling setpoint columns, etc)
  - Documentation will include functional input file examples
  - Output file changes as well
  - Extra regression eliminated
  - Next week: update documentation more fully, tutorials, more input data checking
- If you have an issue, load it to GitHub; OEE will get in touch with you
- If you can't even get the modules running, email Phil Ngo, phil@theimpactlab.co

<#>

Periodic savings info

- Vendors will periodically submit mean run-time reduction and associated statistics of their customer base
  - Some homes must be excluded due to poor data quality (e.g. poor fit, large data gaps, recent subscription, etc.)
  - Can only use data from homes with single speed A/C and furnace, or single speed heat pump, and
  - Reporting of compressor utilization for heat pump homes with aux heat.
- Statistical information to include mean score, standard error of the mean, also possibly decile bins

<#>

Reminder of goals for savings info

- Fair to all partners
- Does not impose an excessive administrative burden on vendors
- Procedure is transparent to EPA, to consumers, and to NGOs
- Technically defensible
- A 3rd party can duplicate the results using the same input data
- Consumers can understand the verification in general terms
- Relies on partner's own data plus publicly-available data sources
- Consistent procedure for updating

<#>

Representative sample option

- Vendors have claimed that the time to process data and compute savings will become a barrier for vendors with many subscribers
- Selection of a representative sample would have to
  - Reflect geographical spread of population of subscribers. Are there statistical measures of geographic spread?
  - Provide statistically meaningful results
  - Include largest useful variety of equipment types
  - Be large enough to average over uncontrolled variables, such as changes in home occupancy or employment
  - Are there other demographic factors important enough that we must include them? (Age, county average income, rate structure)

<#>

"Large enough"

- Relatively easy to make sure sample provides statistically significant results, by requiring standard error below a certain level
- How can we tell if sample is large enough to average out uncontrolled variables?
  - Any factors likely to have an impact and not controlled for (e.g. rate structure), need to try to arrange variation over them
  - For some, geographically clustered samples will have biases (rate structure, possibly occupancy patterns and affluence and age)
  - For these elements, a geographically thinly scattered sample is more likely to average them out
- Simple requirement would be minimum absolute number in the sample.

<#>

Discussion of representative sample

- Choose random sample?
  - Could random selection method be standardized?
  - Systematic sampling: order devices by something natural like date of connection, then pick every x customer. Auditable.
  - Prefer random sampling with same random number generator and a same seed. Use to sort devices (by what?), then take the first x? Each time you choose a sample, you get different devices.
  - Whatever method you use, list of customers would need to be saved.

<#>

Discussion of representative sample

- How to deal with geographic spread?
  - Purely random sample of customers would reflect geographic spread of subscribers
  - Does not reflect geographic spread of U.S. population
  - Customers from various regions get different weight so that vendors' scores are not biased by their regional presence (1F delta T difference makes more difference in mild climate than in extreme climate)
  - Smaller vendors may not have widely spread deployment
  - Require a certain % of sample to be in each of 5 climate zones
  - Require vendors to have results with some statistical significance in each climate zone, calculate score for each, then weight scores by US population in each climate zone. Could even report scores by climate zones.

<#>

Discussion of representative sample

- A (sample in each climate zone) vs B (reflect distribution of subscribers)?
  - Temp and weather correction not significantly more complicated and could be done even if you did not have a lot of customers in each climate zone
  - Look at average delta T and humidity, what fraction of my homes are in each of these bins, and what fraction of the US population are in these bins, and weight
  - Resistance to requiring data in each climate zone
- Quickly getting very complex – can we just for the first stage of this require reporting in different climate zones, or other information, that would allow us to decide if this is a big enough problem to try to address?

<#>

Discussion of representative sample

- "Large enough"
  - Cannot get away from this – biases for each vendor are inevitable
  - Whether the uncontrolled variables are averaged out is entirely expressed in the standard error of the mean
- Does anyone feel there is a need for a minimum number of households or minimum percentage
  - Minimum number that is likely to provide statistical significance can be defined
  - Require that number in each climate zone – if you can't get that in a zone, just report all units
  - Why should a particular sample size be needed at all? Does the sample size matter at all, as long as the sample large enough to have Gaussian distribution
  - Helpful to give some guidance as people start to do this work
  - If you leave the sample size open, what keeps people from gaming, based on what the random # generator happens to give them for various sample sizes?

<#>

Proposal for alternative path for demonstrating savings

- Why?
  - Regional baselines will not be available for initial metric calculations
  - Some CTs drive savings that are not fully captured by the metric
  - Allow providers to demonstrate savings on the order of that required in the spec, through alternate means
- Service providers may submit a study demonstrating savings that meets requirements set out by EPA
- These are questions to inform an initial proposal, which will be released for formal comment
- Process for using the alternative path, and any study requirements, would become part of the CT spec

<#>

Proposal for alternative path for demonstrating savings

- Possible process to use alternative path
    1. Provider submits study design to EPA for approval
    2. If approved, provider conducts study
    3. Provider submits results to EPA
    4. EPA posts study and results on publicly available site
    5. EPA verifies savings comply with statistical certainty of savings as required in the spec (will be added)
    6. If so, EPA provides partner approval documentation that can be submitted to a CB, along with metric data, for initial certification
- Study would be comparison of households with the product, and similar households in which some or all energy saving features of the product are disabled
- Results would be a comparison of run time in control and test group, and would have to have statistical certainty of savings comparable to those required to be demonstrated by metric results

<#>

Alternative path study elements

- What will be provided to control and treatment group, and how does this reflect a realistic baseline case? (e.g. disable algorithms that automatically adjust CT for savings)

- Sample households selection, reflecting similar concerns that we've just discussed for metric samples. Would it be possible to control for a larger number of variables?

- How long will the study be? Cover core heating and cooling only, as we have discussed for the metric. How long do you need to look at household behavior to feel confident you understand how users are reacting to a feature?

- Assessment of results: meets required savings for spec with required level of certainty. Certainty evaluated using statistical methods. For statistically limited samples, may be able to have pre-study period to compare results of sample and test group

- Proposal for semi-annual update/tracking

<#>

Goal is equivalence of alternative path and metric

- Savings demonstrated with the metric and with the alternative path are intended to be largely equivalent
- Both will be required to meet the same level of savings (Y) over homes with "regular" thermostats
  - Metric: if regular thermostats score X, the level set would be X + Y
  - Alternative path: must show at least Y in savings
- Both will be required to have a similar level of certainty
  - Metric: expressed as standard error of the mean in submitted metric savings
  - Alternative path: expressed as level of statistical significance of results, required in spec
- Both require methodological transparency – studies must be available
- Both will have similar requirements for sample selection

<#>

Discussion of alternative path

- Needed sample will be larger by factor of 10 using alternative path
- Will also need to have experiment run for as long as metric calc
- Cannot inform people that they are part of the study because they may behave differently – no disagreement on this point
- General requirement that no additional elements added to service during treatment
- Can existing studies be used if they meet the requirements, even though EPA didn't approve ahead of time? Would reduce burden on vendors.
- Having full features of product not available to those who think they are getting the product for a year would be hard
- Can avoid that by using billing analysis from cooperation with utility
- Does anyone object to allowing the use of billing data if you've got it? (kWh and therms, not $$)
- How do we ensure fair comparison of units with billing data and those with metric data?

<#>

Discussion of alternative path

- What is the baseline? For the alternative path, is the control group the baseline.

- Does the measurement need to go for the full core seasons? For the metric, you need enough data for a good fit – but why do you need that for this kind of study?

- For part of the season you may have more vacations, and also set points change over the course of the season. But aren't you going to have the same effect in the control and test seasons?

- Edge effect? How do households react to change of services? For at least some changes for some services you get a continuum of changes over time. Over time, people over ride or change settings that don't meet their needs (even for a short time) and then leave them in place for the long terms.

- Also over time you could go from a lower efficiency to a higher. Seems more likely to some for these services.

<#>

Droop, Operating Differential & Temperature Accuracy

- EPA proposed leveraging NEMA DC 3-2013 test methodology, but

- Manufacturers have not been able to identify any capable test labs

  - Droop: a lowering of both cut-in and cut-out temperatures due to increase internal heating during high-load conditions. Smaller droop, means more consistent indoor temperatures as outdoor temperatures change.

  - Operating Differential ≤ 2°F equates to acceptable temperature regulation

  - Static Temperature Accuracy ± 0.5°F ensures accuracy of temperature reporting

<#>

Droop, Operating Differential & Temperature Accuracy

- ## Static Temperature Accuracy
  - Testing in a thermal chamber – no specialized equipment
  - But since most thermostats display to the nearest 1°F, is ±0.5°F static temp accuracy achievable? Testable?
  - Will use of reported (rather than displayed) temps facilitate testing to ± 0.5°F?

- ## ≤ 2°F Operating Differential & ≤ 0.5°F Droop
  - NEMA testing is in a specialized test chamber
  - Tight temperature regulation enhances comfort, but does it save energy?
  - Is droop an issue today or is it largely an artifact from mechanical thermostats?
  - In lieu of DC 3, how do manufacturers test?

Discussion of CT device temperature testing

- Static temp accuracy of 0.5 degree?
  - Self-baselining will correct for this anyway, but repeatability is key.
  - Very hard to get this level of complexity because the thermistor is in a very complex environment within the thermostat
  - Drift over time may affect results more, and is a known problem for thermistors, harder to test
  - Test for repeatability instead
  - Are parameters specified for the thermistors? Reaction that this should be a system measurement
  - Air patterns around thermostats in the home, and exactly where it is placed, will make a very significant difference as well.
  - Hard to get a chamber accurate to 0.5 degree anyway

<#>

Discussion of CT device temperature testing

- Does it make sense to have any requirement on temp accuracy
  - 1 F accuracy a lot easier to achieve and to test for
  - Consistency matters much more – less accurate when HVAC running more, during customer interactions, etc. This are ways it is inconsistently wrong.
  - For absolute temperature there will be larger errors (microclimates, placement of the thermostat in the home)

<#>

Discussion of CT device temperature testing

- What kind of internal testing do you do to make sure your customers have a good experience?
  - Control logic and algorithms for differentials, minimum run times, etc.
  - We do some testing in temp controlled boxes, but we do not consider it representative of a real home – broad agreement.
- Which of any of these device measurements would be relevant to the metric or the alternative path?
  - Only care about systematic biases
- Let's say CT has 2F differential, user setpoint is upper end. Means natural bias which will save energy unless user bumps it up. General thought that will be accounted for in the metric or A/B test

<#>

Contact Information

Web site for these notes and all public discussion/comments:
http://www.energystar.gov/products/spec/connected_thermostats_specification_v1_0_pd

Abigail Daken

EPA ENERGY STAR Program

202-343-9375

daken.abigail@epa.gov

Doug Frazee

ICF International

443-333-9267

dfrazee@icfi.com

Phil Ngo

Impact Labs

phil@theimpactlab.co