ENERGY STAR Connected Thermostats

CT Metrics Stakeholder Meeting Slides

February 22, 2020

Attendees

Abigail Daken, EPA
Abhishek Jathar, ICF for EPA
Alan Meier, LBNL
Leo Rainer, LBNL
Nick Turman-Bryant, ICF for EPA
Eric Floehr, Intellovations
Craig Maloney, Intellovations
Michael Blasnik, Google/Nest
Kevin Trinh, Ecobee
Michael Sinclair, Ecobee
Brad Powell, Carrier
Jason Thomas, Carrier
Theresa Gillette, JCI
Rohit Udavant, JCI
Diane Jakobs, Rheem
Carson Burrus, Rheem
Chris Puranen, Rheem
Glen Okita, EcoFactor
John Sartain, Emerson

Eric Ko, Emerson
Albert Chung, Emerson
James Jackson, Emerson
Daniel Stephan, Emerson
Mike Lubliner, Wash State U
Charles Kim, SCE
Michael Fournier, Hydro Quebec
Dan Fredman, VEIC
Robert Weber, BPA
Phillip Kelsven, BPA
Casey Klock, AprilAire
Wade Ferkey, AprilAire
Kristin Heinemeier, Frontier Energy
Ulysses Grundler, Trane
John Hughes, Trane
Mike Caneja, Bosch
Sarathy Palaykar, Bosch
Mike Clapper, UL

Alex Boesenberg, NEMA
Ethan Goldman
Jon Koliner, Apex Analytics
Hassan Shaban, Apex Analytics
Michael Siemann, Resideo
Arnie Meyer, Resideo
Aniruddh Roy, Goodman/Daikin
Jia Tao, Daikin
Dan Baldewicz, Energy Solutions for CA IOUs
Cassidee Kido, Energy Solutions for CA IOUs
Dave Winningham, Lennox
Dan Poplawski, Braeburn
Natasha Reid, Mysa
Peter Gifford, Mysa
Vrushali Mendon, Resource Refocus

Agenda

- Software V2.0
  - Updates and installation
  - Meeting feedback
- February submission preliminary analysis
- Variable capacity metrics – continued from previous meeting
  - Lab test for Communicating Controllers
  - Average Capacity Factor (ACF)
  - System Sizing
- Topics from the floor (**Didn't discuss due to limited time**)
- Connected Thermostat Use Cases (**Didn't discuss due to limited time**)

Software Updates: V2.0

- Updated to version 2.0.0a2
  - pip install --upgrade thermostat==2.0.0a2 or pip install thermostat==2.0.0a2
  - Pins version of numpy to 1.20, as Python versions 3.7+ have an issue with eeweather and will cause all thermostats to be thrown out
  - Ensures that less than 5% of data is missing, will throw out thermostats with more than 5% missing data
  - If you can, we would love it if you would re-run with this version and re-submit data
- Updates to documentation
  - Based on feedback we've worked to make the documentation clearer on the file format and requirements
  - If the documentation is unclear or if you find discrepancies, please email us and let us know

Software Updates: V2.0

- Feedback from the meetings
  - Mitigating remote access
  - Working on solutions to remove the need for remote access to NOAA historical data
  - Long time to get data into the format for the epathermostat software
  - Please let us know if we can help

Discussion: Software Updates

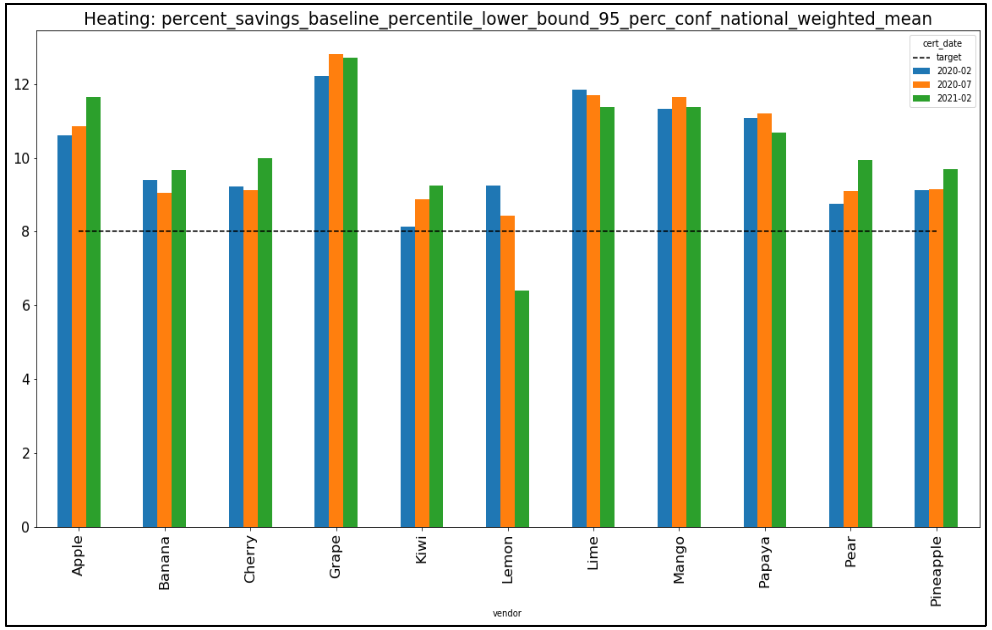
- One yes
- May not be possible without talking to eeweather
- May also align with other efforts by eeweather so we should talk to them (This is for TMY3, not for the other weather stations as we use for our software.)

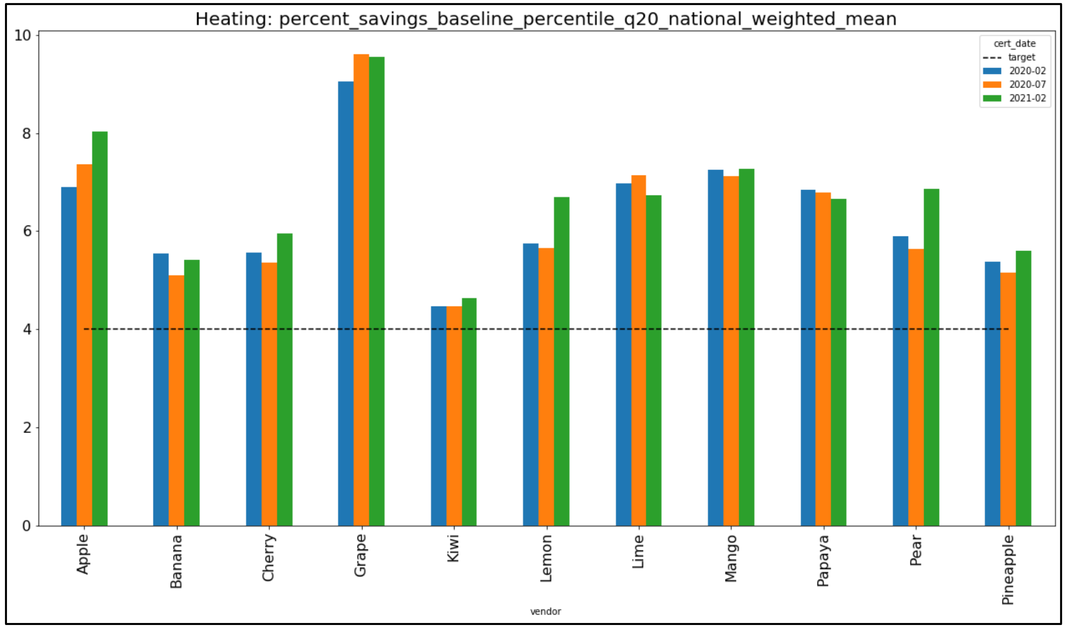
February Submission: Preliminary Analysis

- Reminder: Not withdrawing certification of any vendor for missing the mark with 2020 data due to pandemic and extraordinary circumstances.

- Resubmission Statistics:
  - 14 datasets received for V 1.7.2 single speed equipment
  - 2 datasets received using V 2.0 for same sample as V 1.7.2 and expecting 2 more soon!
  - 2 datasets received for two-stage equipment only sample using V2.0 and expecting 2 more soon!
  - No oversampled sets for RHU2

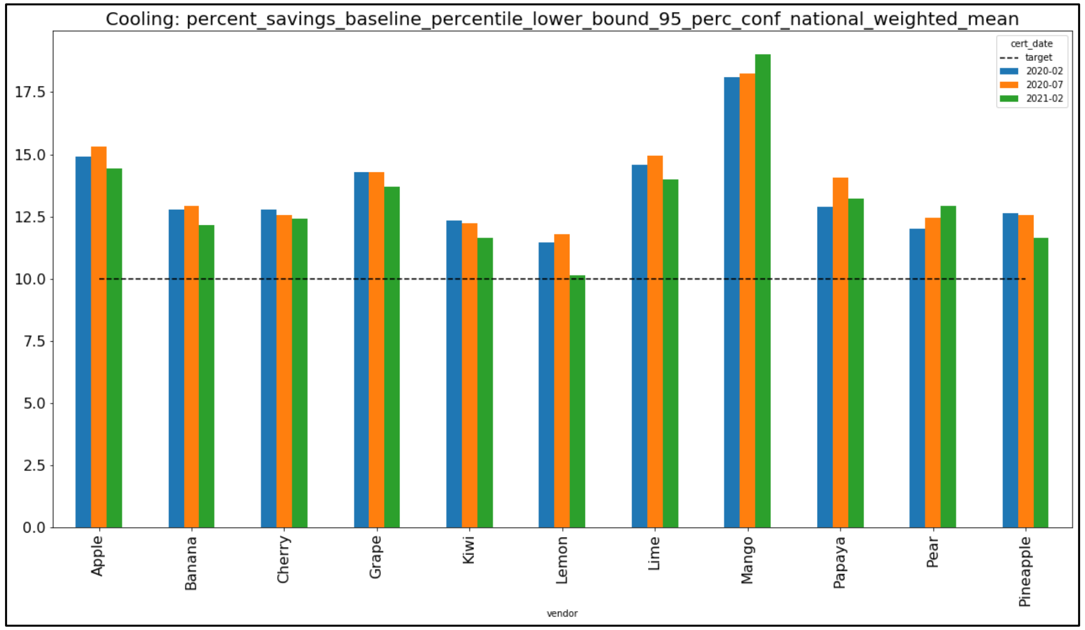
February Submission: Heating Metric

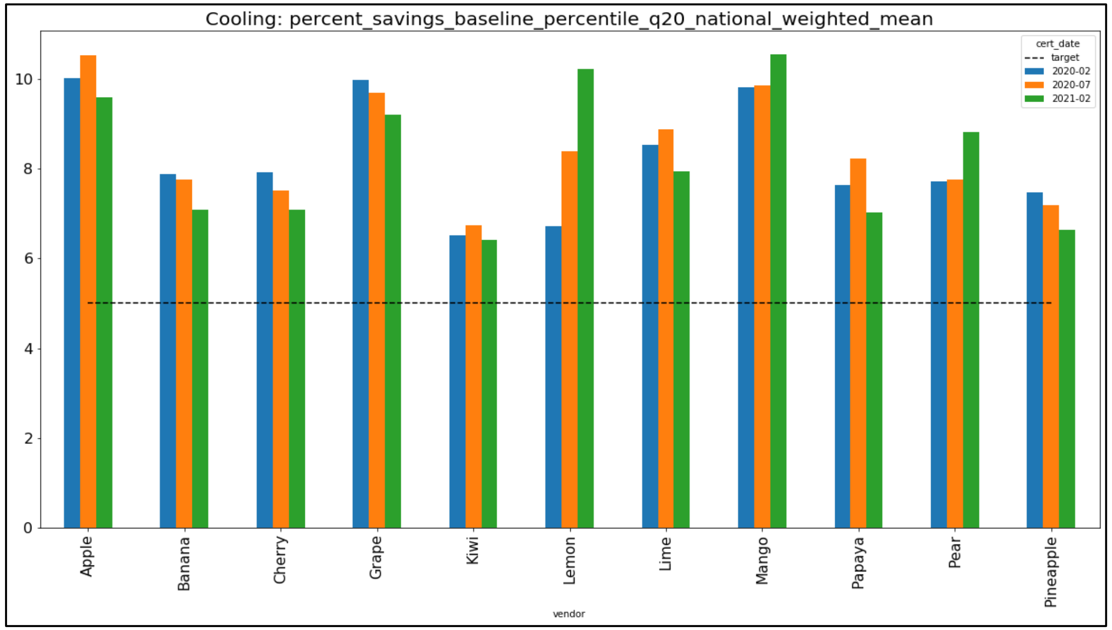




- Included only those vendors who submitted data since Feb 2020 (3 submissions).
- No particular data trends.
- Approximately half vendors witnessed an improvement in their Metric scores
- One vendor missed the metric mark for lower bound 95% confidence interval

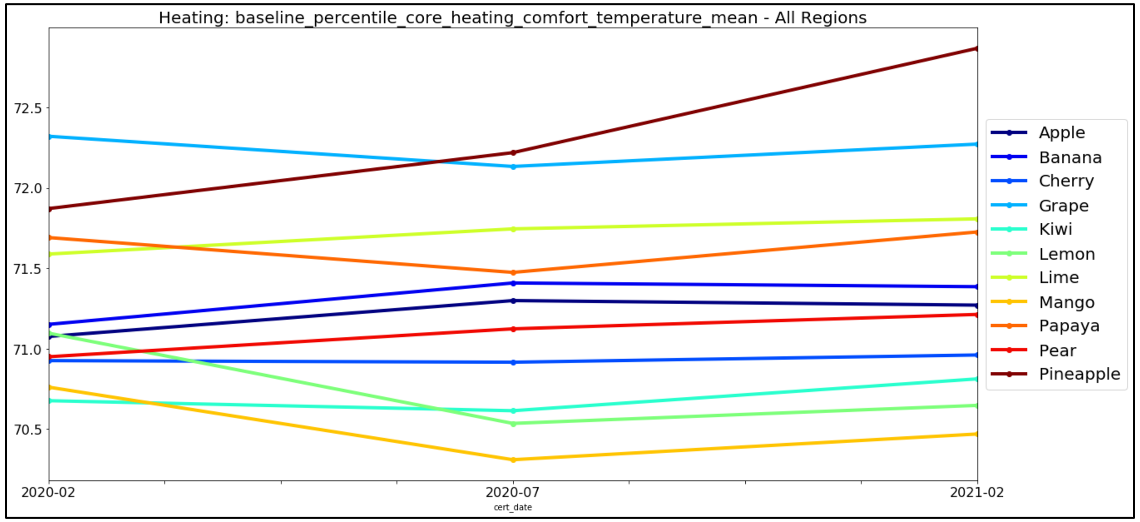
February Submission: Cooling Metric

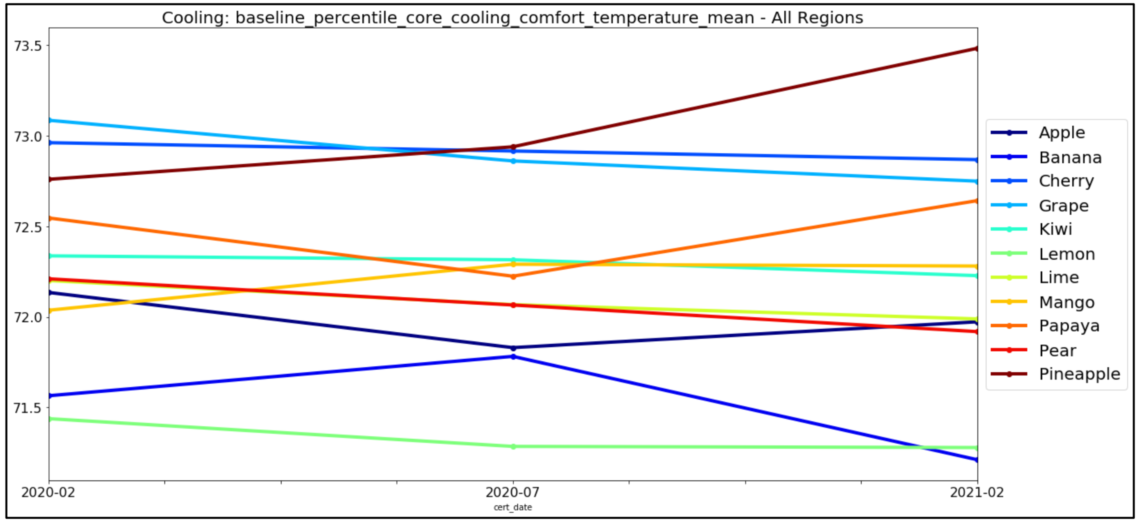




- Included only those vendors who submitted data since Feb 2020 (3 submissions).
- Cooling metric scores declined for majority of the vendors compared to previous years

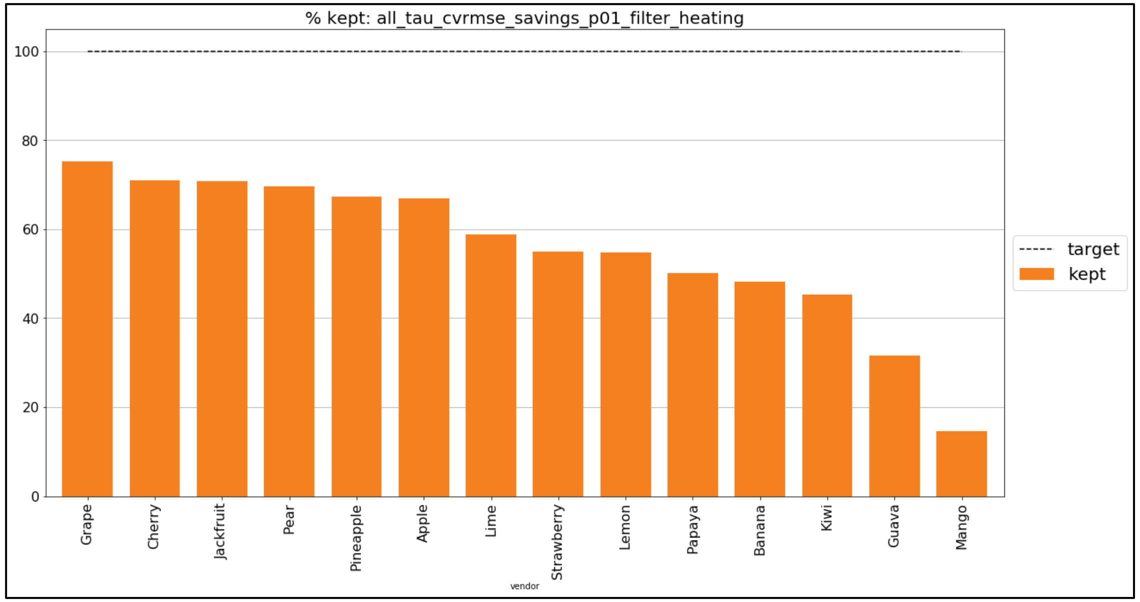
February Submission: Comfort Temperatures

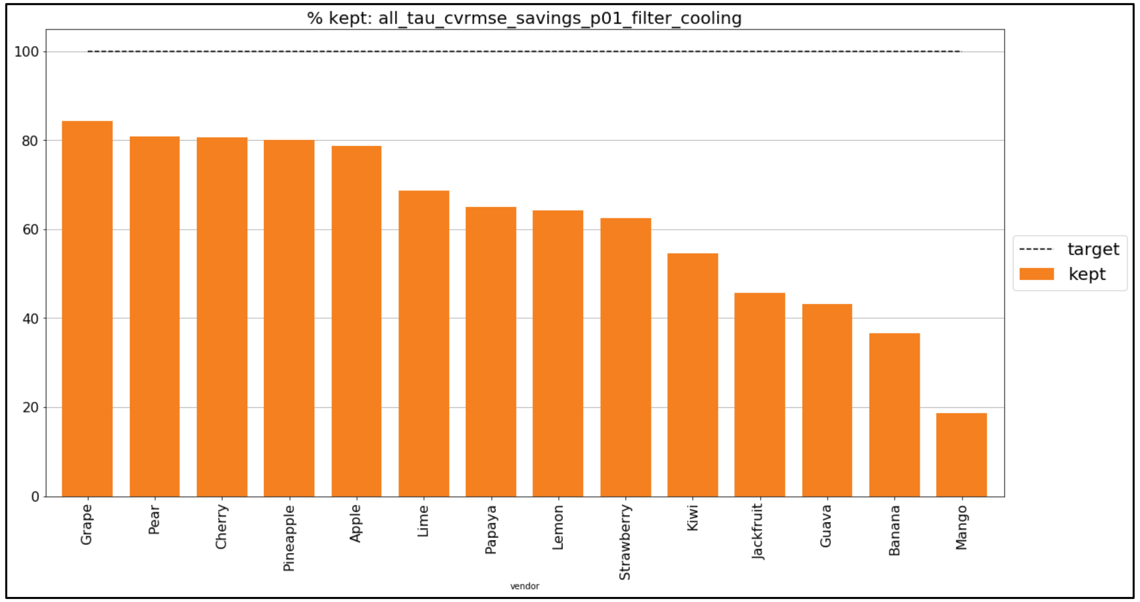




- No significant change in heating comfort temperatures. Marginally up for most cases.
- Cooling comfort temperatures showed marginally downward trend but for 3 vendors.

February Submission: Thermostats filtered





- Results for February 2021 submission only.
- 5 vendors have 50% or more thermostats discarded for heating on filtering.
- 4 vendors have 50% or more thermostats discarded for cooling on filtering.

Discussion: February Resubmission

- COVID-19 impact – why so minimal?:
  - Higher scores might have been from less efficient comfort temperatures
  - In addition, summer was milder, leading to a 10% higher savings per degree of setback
  - Most of the heating happened before the pandemic hit
  - Saw less of an impact on schedules and occupancy than expected, particularly after June 2020. Generally more occupancy during summer.
  - Big occupancy impact during Spring, when relatively little heating or cooling.
- Are there any further details about the type of equipment? All of the equipment in this sample was wired as if it was single stage. Because of this, we would expect most of these to be single stage equipment. Any system wired to control 2-stage heating or cooling is eliminated from the sample.

Discussion: February Resubmission

- We do not currently log why thermostats are eliminated
- Are any of the thermostats using non-internet timestamps, that might be causing data noise? The amount of offset you'd need to cause the error is very large.
- We talk about missing data – could there be a systematic bias produced by throwing out the thermostats with poor connectivity, which may correlate with other properties of the users or the installation.
  - Could relax the 5% missing data crierion if we tested including gappier data.
  - Could also run the metric for thermostats missing more data to see if we get systematically different results.
  - Not clear how missing data would affect the result, other than having little data in the core seasons, which would mean the regression models are poor models for what's actually happening

Discussion: February Resubmission

- Q: are we really able to say much about the actual savings for a product where 80% of the installations were kicked out?
  - Knowing why the thermostat was dropped would be helpful
  - Could figure out the outdoor temp for the times that have no data are – which might provide a way to know whether you’re missing most of a season.
  - Current stats summary file has number of days with insufficient data, total core day run time (unfiltered and filtered samples), outdoor temps filtered and not filtered, and a number of other things that could be useful in examining this.
  - Is there a way for the software to indicate which specific installations were in or out of the sample? Flag in the thermostat file at least for the CVRMSE and tau criteria. Could also do a data dump from the stats file.

Discussion: February Resubmission

- Why do we think tau is a constant? It isn't really, but the model parameter, which comes from a fit to daily average values.
  - The measure of demand is calculated on an hourly basis before being rolled up to daily demand, so some of the hourly variation is captured.
  - How does the algorithm converge? Is psypi's least squares subject to problems with local minima? Because it appears from comparing thermostat results from the EPA software and from other ways of analysis, the fit doesn't seem to be the same. It looks like the EPA software is finding local minima.
  - The problem with tau out of 0 to 25 range is usually tau below zero…often because a fireplace or woodstove is used only on very cold days. Sometimes different user choices in mild and very cold weather can have that.
  - Most tau outside this range are above -5F, which may simply reflect that this is noisy data
- Least squares may not be the issue if we're concerned about local minima

Lab test for controllers for variable capacity systems
Context: Where we've been and how we're getting to this subject

- We've been pursuing a method to evaluate the savings of communicating controls based on field data
- Stakeholders said "good" control of variable capacity systems means
  - Avoid cycling at higher capacity at lower loads
  - Intelligent setback and recovery to avoid use of high-capacity states and (for heat pumps) resistance heat
- Have gotten as far as we can w/o more data – we'll review later in this meeting
- Based on stakeholder interest in December, let's explore lab test – not sure we'll use, but let's brainstorm a bit

Exploration: Lab test for communicating controls

- Coordination with activities for testing variable capacity products (at DOE, CSA, and elsewhere) will be important; may delay development
- Advantages of lab test:
  - Set conditions precisely to exercise capability
  - May be less expensive to develop and/or to run
  - Could be performed on non-connected variable capacity CAC/HP as well
  - Others?
- Disadvantages of lab test:
  - Hard to tell if the lab setup is representative
  - Not a simple test – CAC/HP industry has been struggling with this for years
  - Others?

Discussion: Advantages and disadvantages of lab test

- Advantages & disadvantages
  - Lab costs can actually be quite high, and certainly at the moment are pretty much unavailable either for development of the test or for running it. So much so that we can say in the next several years there would have to be a big payoff.
  - The test conditions will also affect cost sensitively – more test points will better reflect field conditions but also be much more expensive.
  - Could there be a simple methodology that would allow lab testing? The work now on a test method that reflects field performance for CAC/HP goes well beyond what we need for a thermostat. If we were to go this way it would need to be something pretty simple in the short term at least.
- Q: is CSA-EXP07 what folks are referring to?
  - Partly, though there is much work to do in characterizing the repeatability, reproducibility, and representativeness of it, and the cost/benefit hasn't been clearly defined.
  - There are other paths under investigation worldwide.

Discussion: Advantages and disadvantages of lab test

- For context, every piece of residential & commercial CAC/HP will need to be re-rated by 2023, and many will need to be redesigned by 2023 or 2025
- Potential round robin with communicating control being swapped out for non-communicating control with the same type of equipment to test the communicating control
- There is a round robin for reproducibility for CSA-EXP07 going on right now
- Acknowledge that this is important, but the context of lab lack of availability is also important
- ASHRAE 90.2 is looking at measures to meet criteria for residential building reach codes. Proposal to fund something to get into this – how will connected tstats going to work with VRF systems.
- If we are interested in whether the control specifically works well with the variety of buildings, rather than a variety of equipment combinations, it might be better to look to a simulation engine.
  - One yes, one something along these lines could be considered
  - Now we have way more variables, though, because of the complexity of the home as simulated. How would we handle that variety.

Exploration of Lab test for communicating controls: what to test?

- What would we want to test in a lab? What can we test?
  - Recover from setback (or decide not to set back) in a way that saves energy.
  - Validate set points for DOE testing as per the VRF CVP.
  - Avoid cycling at higher capacity at lower loads
  - Others?
- What would the lab test leave un-tested?
  - Interaction between user behavior and control capabilities.
  - What else?

Discussion: What would we want to test in a lab, what can we test?

- Small chamber with a building model temp profile and look at output of the signal it would be sending to equipment, rather than looking at the behavior of the equipment itself.
  - But the signals coming out of the thermostat are proprietary signals, so how would we work with that? In some cases, already providing this to DOE under an NDA of some kind.
  - Sometimes publish an open API, which would be acceptable. Would an open API be able to test what we need?

Discussion: What would we be unable to test in a lab?

- notes

Exploration: Hybrid test for communicating controls

- Another stakeholder suggestion from December meeting sort of in-between lab and field data
- Define a test sequence that units could run through in field installations to evaluate behavior in specific circumstances, without affecting user experience
  - What could be tested this way without affecting user experience?
  - What would the test sequence look like?

Discussion: Hybrid test

- notes

Variable capacity metrics from field data

- Three potential metrics:
  - Average capacity factor
  - Lowest temperature without short cycling
  - Short cycling fraction

- In December we discussed how minimum capacity call (proxy for how low system capacity can go) and climate region informed the estimation of ACF for cooling, and we presented initial estimates of system size based on cooling runtimes

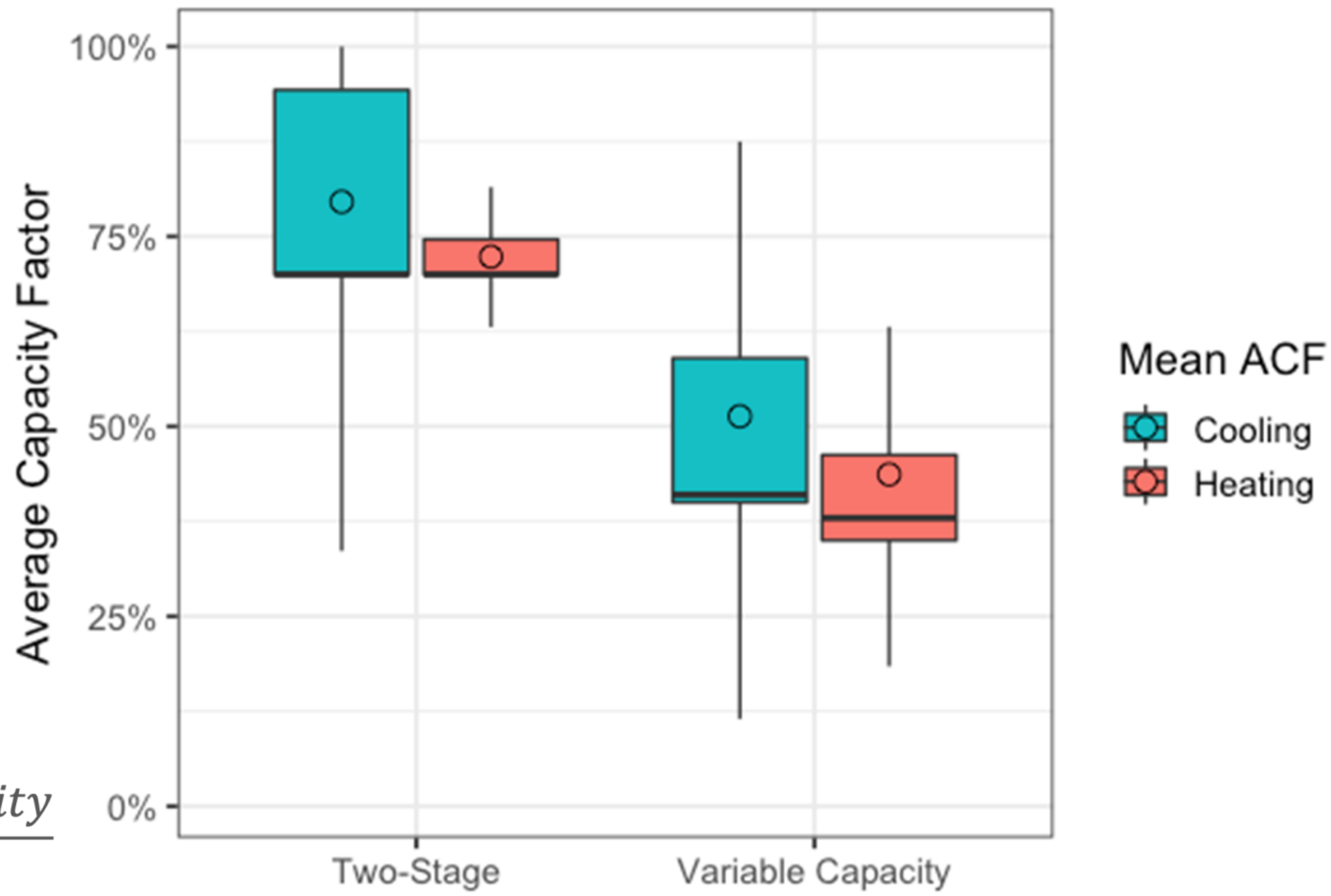
- Today’s call we hope to close the loop by presenting ACF estimates for heating by climate region and relative system sizing estimates based on heating runtimes

- Bear in mind that our heating data are for furnaces only and not heat pumps (may observe different behavior with heat pumps)

Review of Average Capacity Factor

- Variable capacity systems have lower ACFs than two-stage systems for both heating and cooling

- We see much less variability with heating ACFs compared to cooling

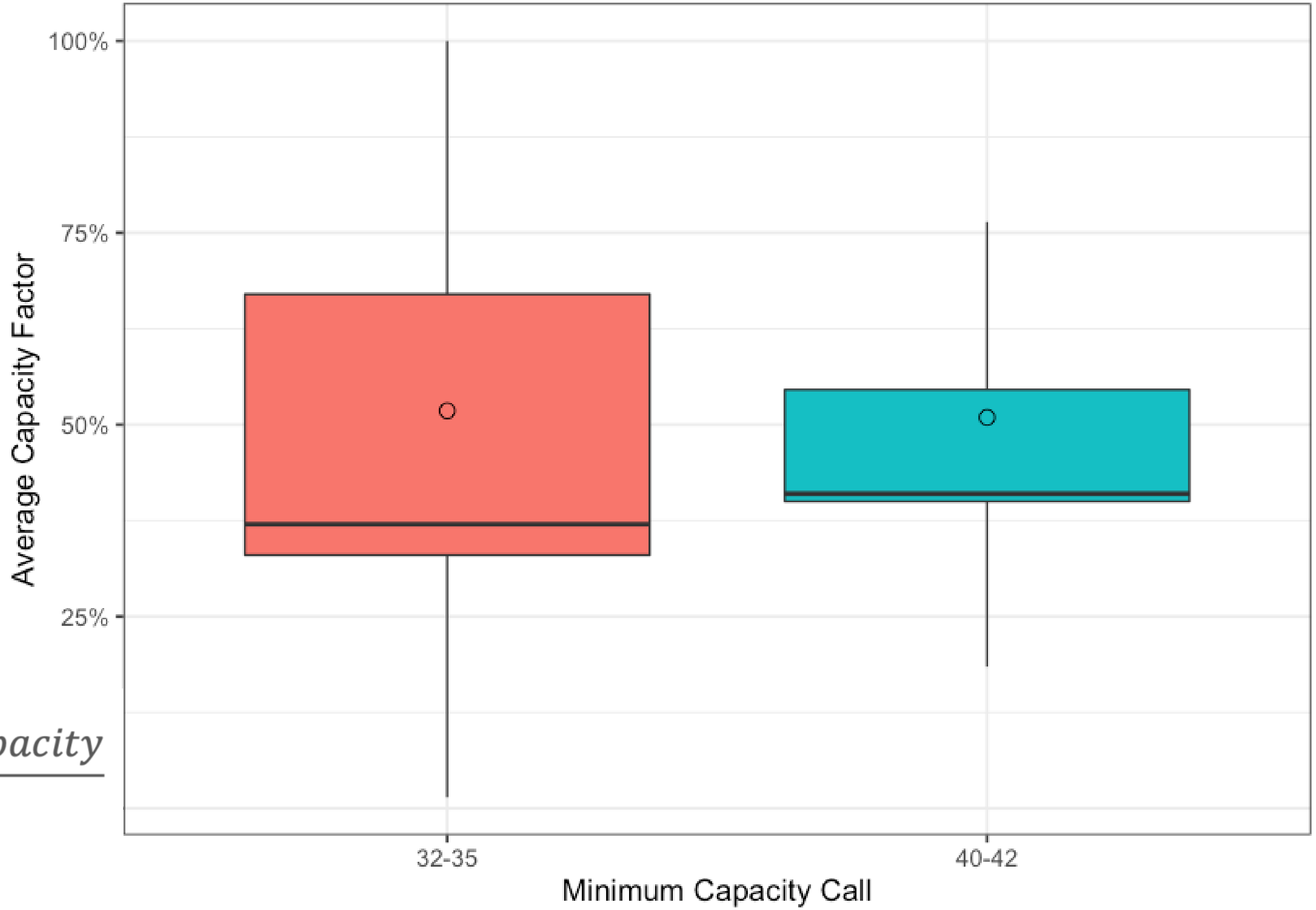


$$ACF = \frac{ERT}{RT} = \frac{\sum_{time} run\ time * relative\ capacity}{\sum_{time} run\ time}$$

* Data are from an anonymous vendor and represent the average capacity factors across 71 two-stage and 91 variable capacity systems with split AC and gas furnace across 58 cities in 5 states for one year

From last time: minimum capacity call does not explain ACF variation

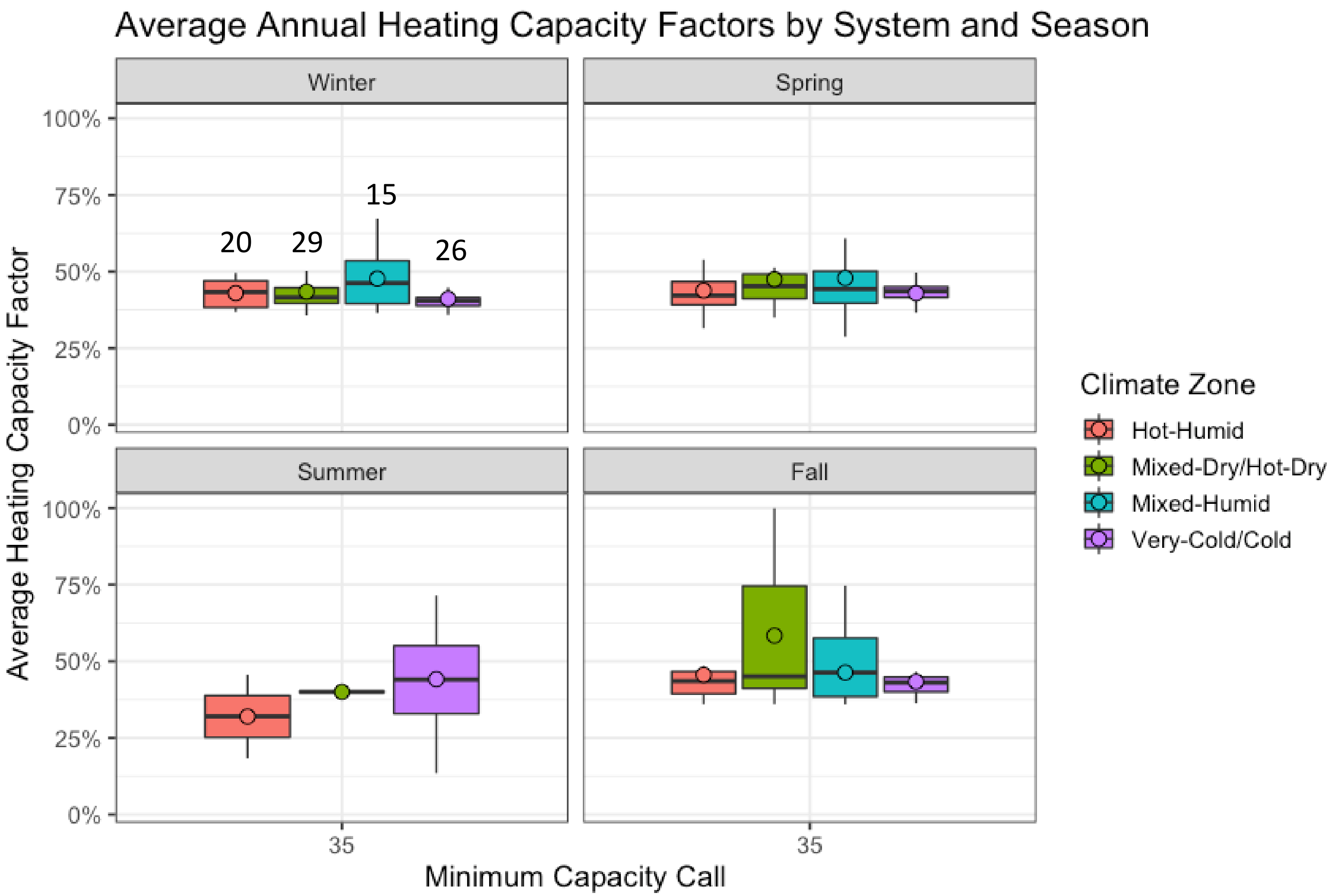
- When looking at the variability in cooling capacity factors across systems for two different minimum capacity calls, we concluded that turndown ratio does not explain much variability in ACF
- Question was raised: are we seeing variation between climates?



$$ACF = \frac{ERT}{RT} = \frac{\sum_{time} run\ time * relative\ capacity}{\sum_{time} run\ time}$$

ACF for heating has less variability than for cooling

- More variability for Mixed-Humid climate zone
- More variability in summer and fall

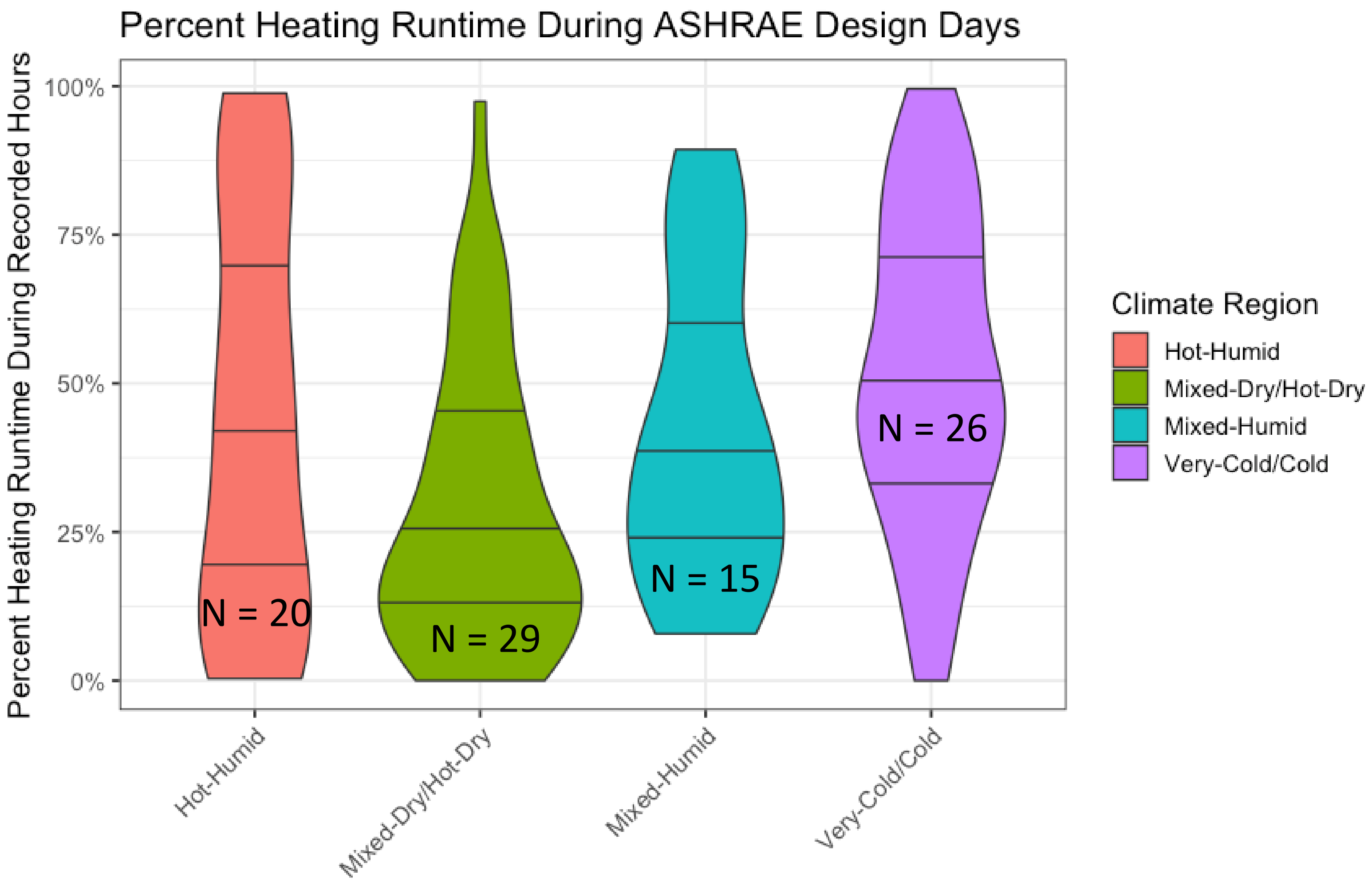


System Sizing: Why and How?

- System sizing affects two of the metrics that stakeholders proposed (outside temperature at which system can't run continuously and starts cycling)
- Controls vendors have little or no influence on system sizing
- If we want to control for sizing, we need to be able to derive it from the data
- In previous meetings we discussed some methods to do so
- Suggested method:
  - Look at run time on "design days", when system should be running near maximum capacity for substantial periods of time
  - Design day defined by the top 1% of temperatures
  - Look at percent runtime for the recorded data for all design days
- Presented cooling in December; now presenting heating analysis

Comparing relative system sizing interesting, but no clear "oversized"

- Distributions across systems are suggestive but do not clearly show a group of "oversized" units



Sizing estimation for heating and cooling (Didn't discuss due to limited time)

- Determine the heating and cooling design temperatures:
  - Use the 1% cooling / 99% heating dry bulb temperature from the appendix of chapter 16 of ASHRAE manual based on closest matching entry
  - Use the 99th / 1st percentile dry bulb temperature measured in the city where the HVAC system is installed
  - Use the 99th / 1st percentile measured outdoor temperature from the HVAC system
- Determine the heating and cooling day selection:
  - Select the days in which the **maximum** temperature measured by the city's weather station crosses the cooling or heating design temperature
  - Select the days in which the HVAC system's **maximum** measured temperature crosses the cooling or heating design temperature
- Deal with missing data:
  - Estimate the percent of the recorded hours that HVAC system is running

Discussion: Variable Capacity Metrics

- Why are so few systems run full out on the coldest days?
  - We would expect 50-60% and no more, because the design day really means it needs to be able to keep up in the middle of the coldest night, not that it needs 100% run time all the hours of the days.
  - Would also need a little over-capacity for recovery from setback or from leaving the door open to bring the groceries in
  - But WAIT! For variable capacity systems, you would expect the capacity to turn down instead, and to run for long times.
- The current situation is that controllers or control algorithm are not tested in the DOE equipment specs.

Topics from the floor (Didn’t discuss due to limited time)

- We’d also like to hear what you are interested in
  - What questions would you like us to explore in these meetings?
  - How do you wish the ENERGY STAR thermostat program was different?

Discussion: Topics from the floor

- notes

Use Cases: Summary of Discussions (Didn't discuss due to limited time)

- Goal was to identify use-cases where current metric would give misleading results or unfairly penalize a vendor
- We identified >10 use-cases and considered:
  - Fraction of CTs in this category
  - Impact on metric, inside temperatures, runtimes
  - Problems in samples
  - Other factors

#	Use Case
1	SF detached home (1 thermostat)
2	Vacation home
3	SF home (>1 thermostat)
4	Multiple thermostats on a single account
5	SF home with multiple temperature sensors
6	Small commercial with own HVAC
7	Apartment with own HVAC
8	Duplex home, multiple thermostats, different accounts, same dwelling
9	Variable capacity heating or cooling
10	2-stage system
11	Dual fuel

High-Level Conclusions

- Limited data available on use-cases
  - Anecdotal or limited metadata
  - Vendors often used creative methods to infer presence of different use-cases
- Most non-standard use-cases:
  1) increased variance but not clear if they biased results
  2) were already filtered out in the software
  3) were a small fraction of the samples
- No significant use-case behaved so differently as to require revisiting metric. (No "show-stoppers")
- More metadata is needed to reliably identify different use-cases

Next Steps

- EPA will not, at this time, make any adjustments to the metric in order to account for different use-cases
- EPA may increase sample size in order to take into account increased variability caused by different use-cases
- EPA will periodically re-visit use-cases to ensure the credibility of the metric
- EPA encourages further investigation of use-cases
  - Vendors are invited to report on internal studies
  - Targeted investigations by researchers, universities, utilities, and other groups