



# **ENERGY STAR Connected Thermostats**

## **CT Metrics Stakeholder Meeting Slides**

September 13, 2021



## Attendees

Abigail Daken, EPA

Abhishek Jathar, ICF for EPA

Alan Meier, LBNL

Leo Rainer, LBNL

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

Eric Ko, Emerson

Vrinda Gaba, Emerson

Tom Lorenz, Emerson

Daniel Merchant, Emerson

Mike Lubliner, Wash State U

Charles Kim, SCE

Samira Sammoun, Hydro Quebec

Robert Weber, BPA

Phillip Kelsven, BPA

Casey Klock, AprilAire

Wade Ferkey, AprilAire

Kristin Heinemeier, Frontier

Energy

Ulysses Grundler, Trane

John Hughes, Trane

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

Claire Mizolek, Energy Solutions

for CA IOUs

Dave Winningham, Lennox

Dan Poplawski, Braeburn

Natasha Reid, Mysa

Moustafa Elsisy, Mysa

Aidan Girard, Mysa

Sylvain Mayer, Sinopetech

Riana Johnson, Illume advising

Elmer Guest, efficiencyMB



## Agenda

- Software updates
- July 2021 data submission analysis
- Missing data updates and proposed solutions
- Metric topics from discussion guide comments
- Summary of results from disaggregating product families

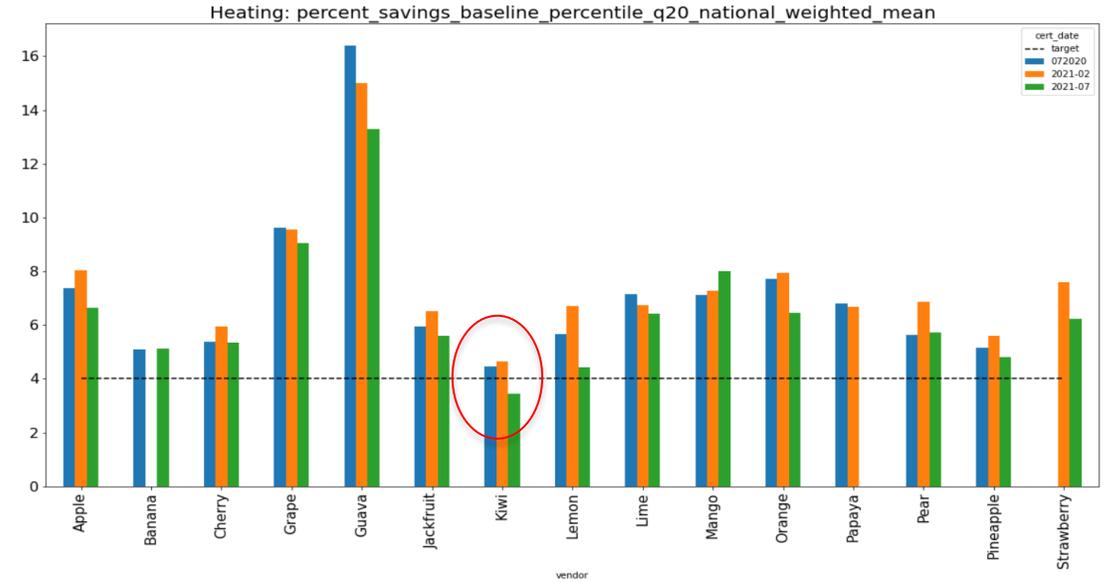
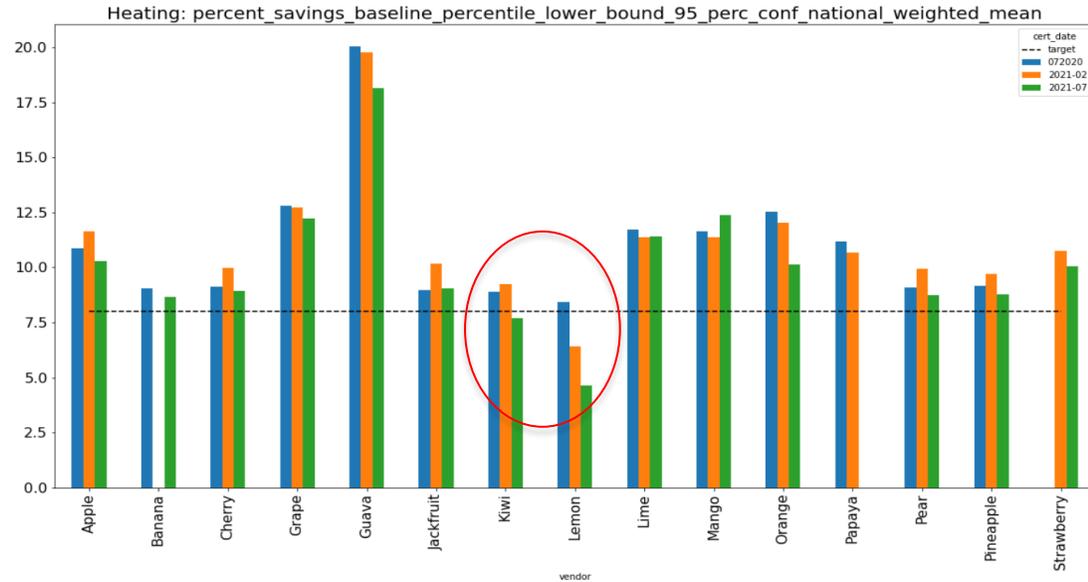


## Software Updates: V2.0

- Speed improvements in the Stats module (multi-processing)
- Removed mid-year and year end-to-end processing
  - (If you used the example scripts you couldn't use these anyway)
  - We're currently processing all data that is presented for the period
- Core-days for heating / cooling are calculated when the thermostat is imported
  - Allows us to explore surfacing poorly-performing thermostats that don't have enough core days earlier in the process
- Added a "common issues" page to the documentation outlining common issues and how to address them
  - [https://epathermostat.readthedocs.io/en/feature-epathermostat\\_2.0/common\\_issues.html](https://epathermostat.readthedocs.io/en/feature-epathermostat_2.0/common_issues.html)
  - This will be a "living document" with issues that folks encounter with the software. Please let us know your pain points and we'll add them to the documentation.



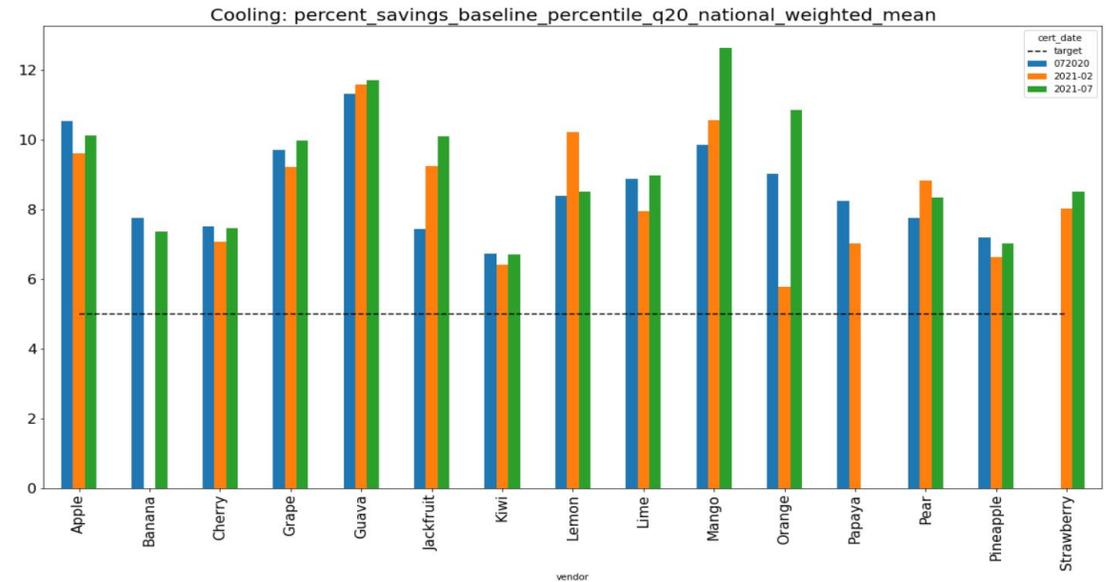
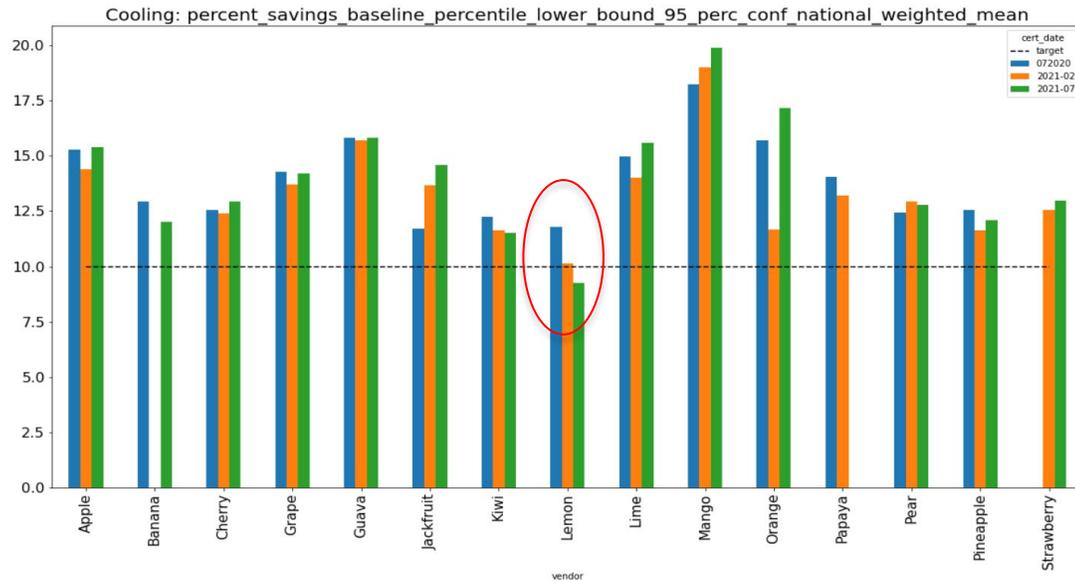
## July 2021 data submission analysis – Heating Savings



- Included past three data submissions.
- Heating metric scores for most vendors fell compared to previous year.
- Two vendors missed the metric mark for lower bound 95% confidence interval.
- One vendor missed the metric mark for q20.



## July 2021 data submission analysis – Cooling Savings



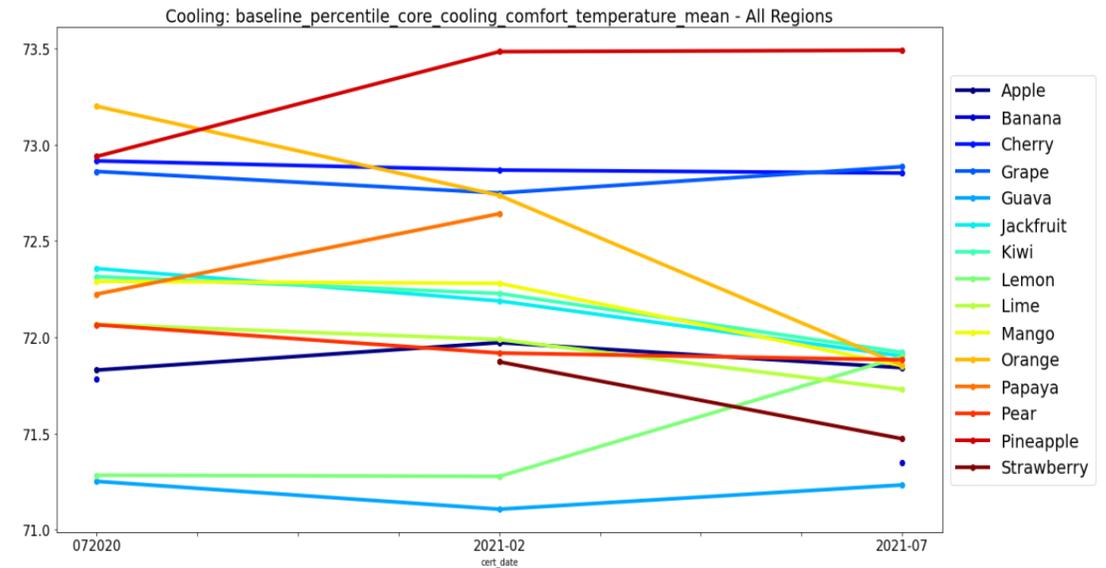
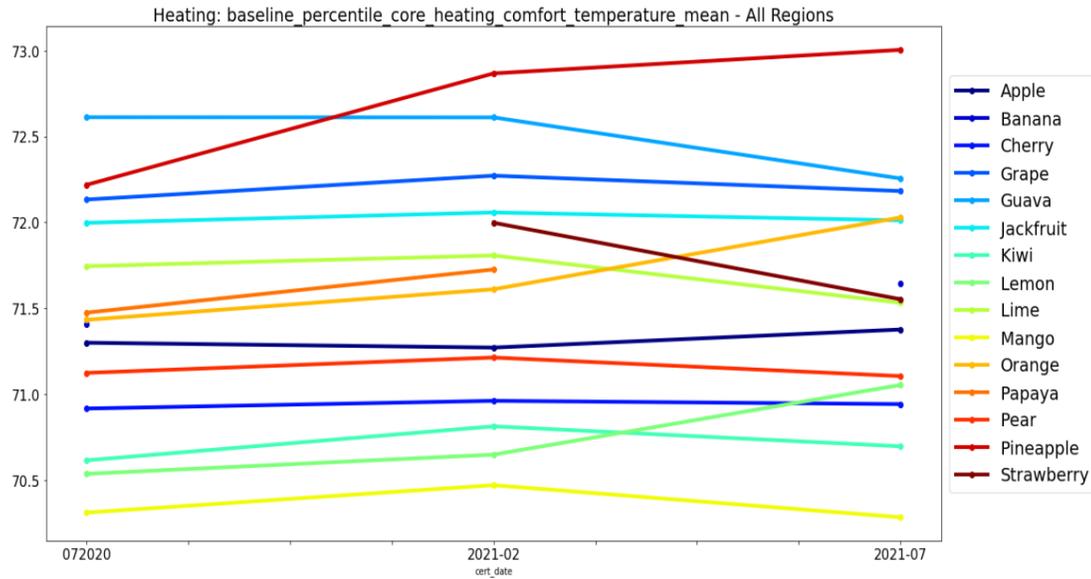
- Included past three data submissions.
- No particular trend with metric scores.
- One vendor missed the metric mark for lower bound 95% confidence interval.



## July 2021 data submission

- Why two vendors missed the metric levels?
  - EPA actively engaged in follow up conversations with both vendors to find the root cause.
  - Vendor 1
    - Bookkeeping problem with filtering out hourly aux heating runtime in the backend caused the vendor to throw out many days of data.
    - Fixing the issue and rerunning the scripts on the same sample showed significant improvement in metric scores.
  - Vendor 2
    - All days present but significant number of thermostats had no data.
    - Missing data was due to keeping files on thermostats that were removed or dropped.
  - How should EPA account for thermostats that go offline for significant periods?

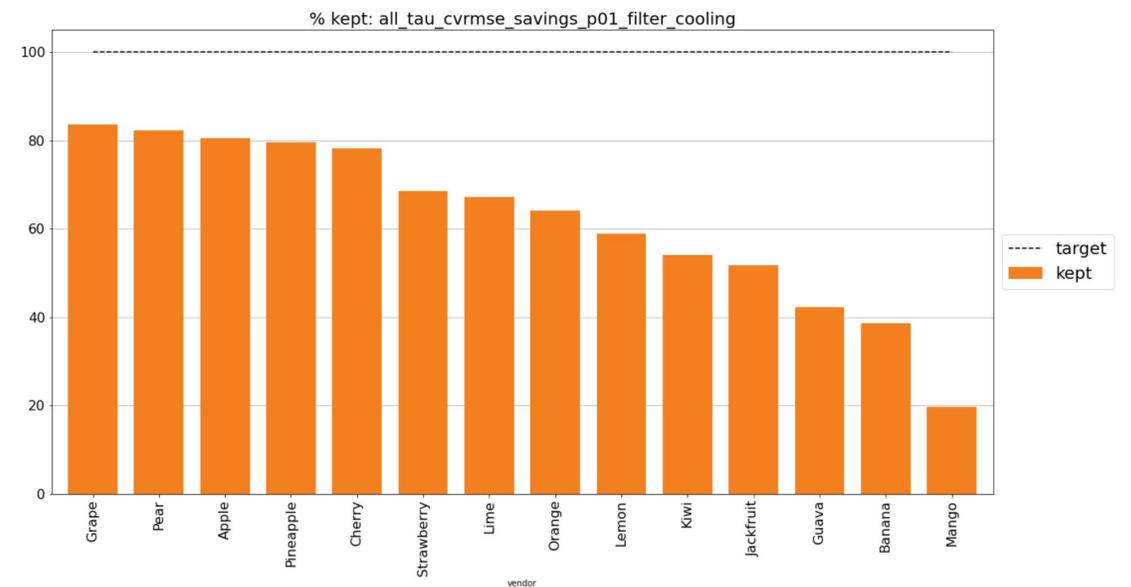
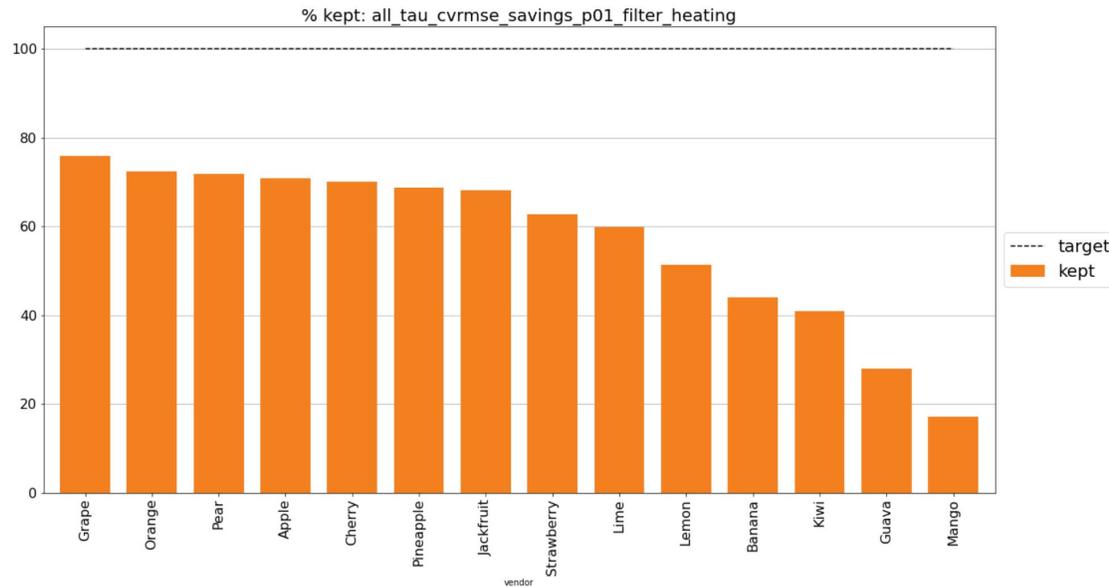
## July 2021 data submission analysis – Comfort Temperature



- No significant trends in heating and cooling comfort temperatures.



## July 2021 data submission analysis – Dropped Thermostats



- 4 vendors have 50% or more thermostats discarded for heating on filtering.
- 3 vendors have 50% or more thermostats discarded for cooling on filtering.



## Missing data updates and proposed solutions

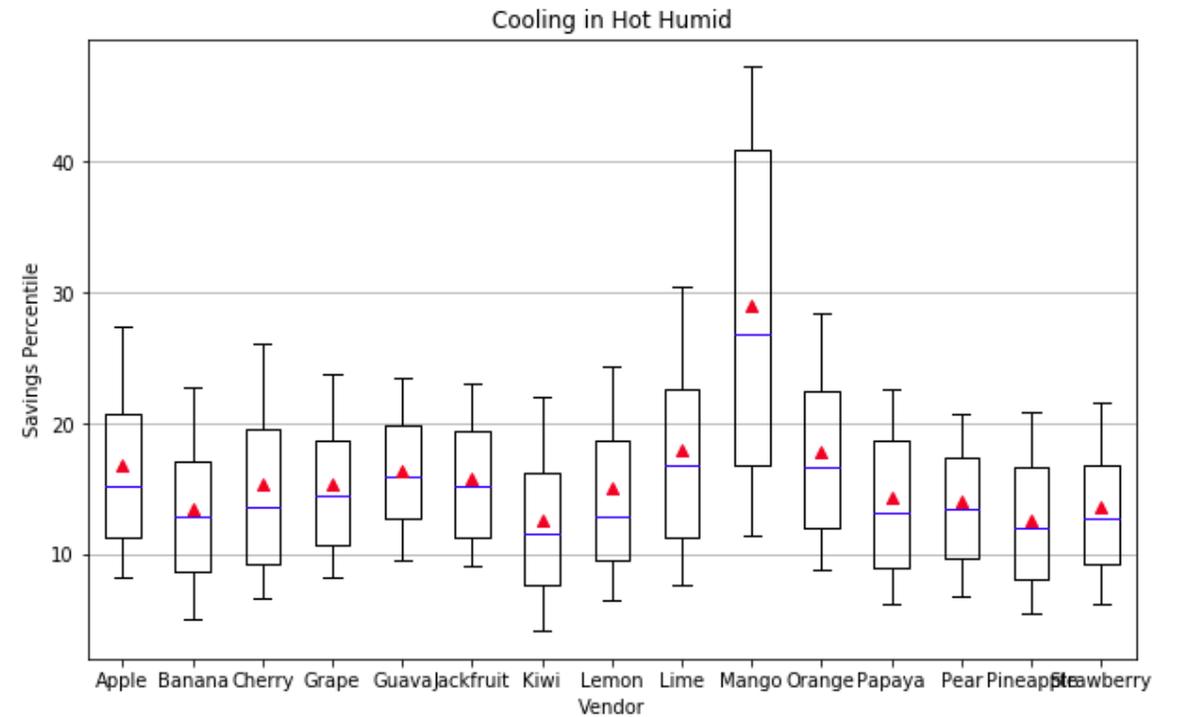
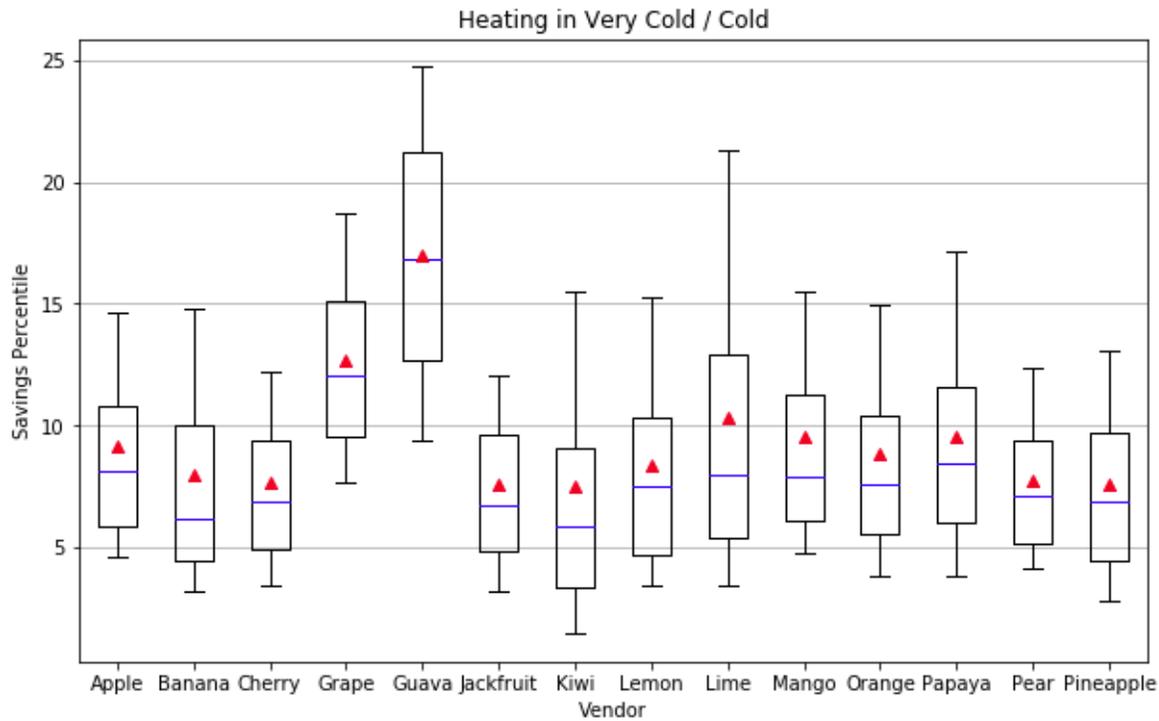
- Goal: ensure that data quality thresholds do not exclude thermostats unnecessarily
  - Missing data threshold in v.2 software: max of two days with > 2 hrs missing
  - Including more thermostats will reduce volatility of sample and Q20 score
  - More installations with somewhat less data may improve reliability and repeatability of results
- Question: how much does missing data impact the metric score?
- Test: remove random days and measure change in savings score
- Proposed solution: require  $x\%$  of *core* days (rather than absolute number of days)

## Missing data updates and proposed solutions

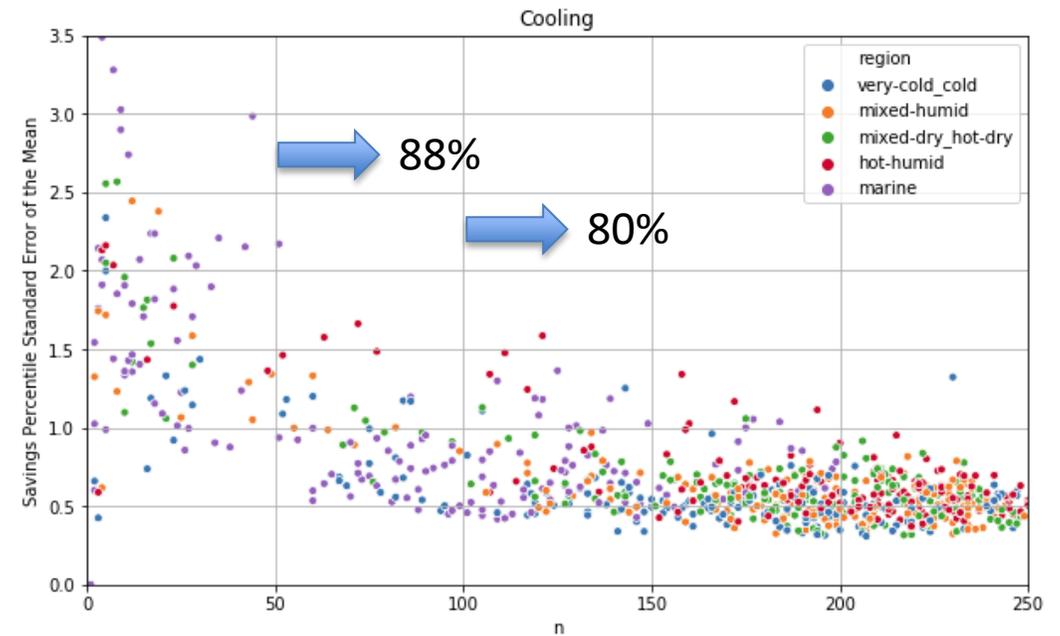
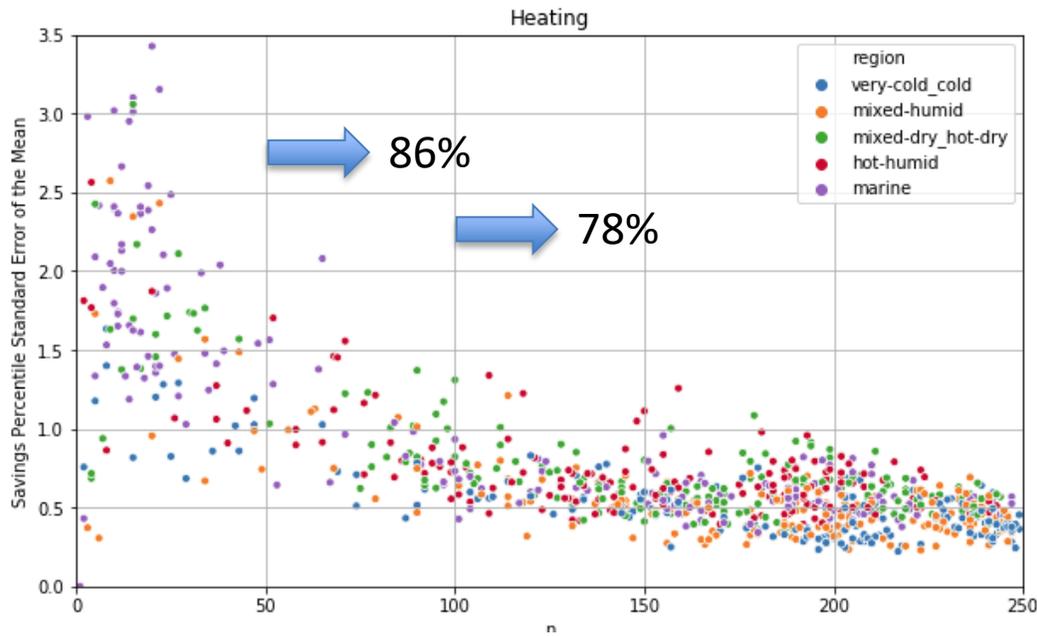
- Next Steps:
  - Working with vendors to run this code on their device data
  - Determine new missing data threshold
  - Implement new missing data test in v.2 epathermostat code
- Open questions:
  - How much variation in the savings metric is worth the gain in sample size?
  - Does random data removal accurately simulate real missing data?
    - What about blocks of missing data? Correlations with run-time?
  - How many thermostats are missing how much data?



## Distribution of Savings Percentile by Vendor for the July 2021 Submission



## Variation in Savings Percentile Standard Error of the Mean by Submission Size (n)



- All vendors for all certification submissions
- SEM > 3.5 and n > 250 trimmed

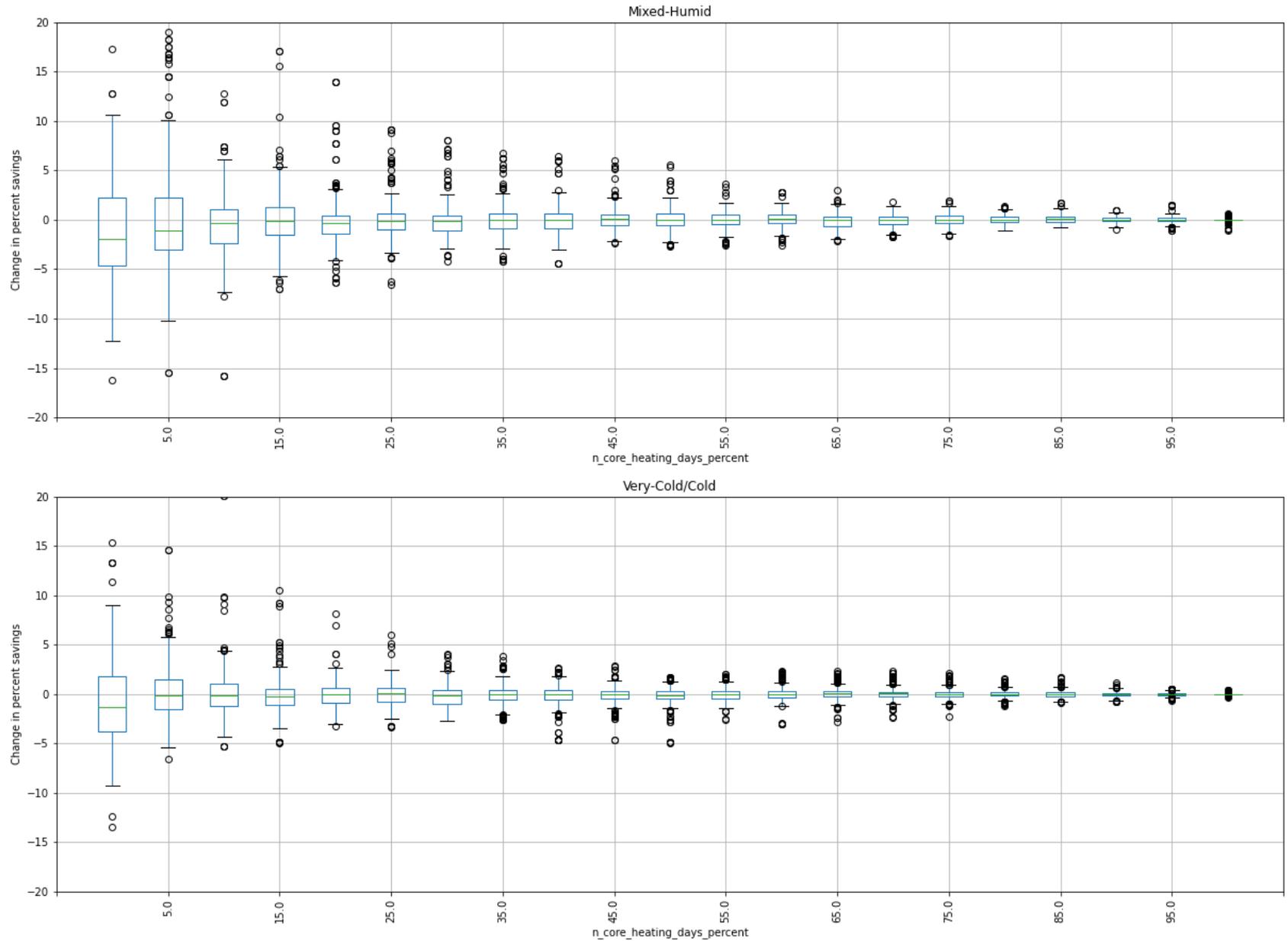


## Simulating missing data

- "Gremlin" code
  - [https://github.com/EPAENERGYSTAR/epathermostat/blob/feature/epa2.0\\_gremlin/scripts/gremlin.py](https://github.com/EPAENERGYSTAR/epathermostat/blob/feature/epa2.0_gremlin/scripts/gremlin.py)
- Sample was from the DYD data set, comprising of 329 thermostats with less than 15 days of missing data.
- The Gremlin code is a naïve data removal tool that randomly removes five (5) days of sample and recalculates the results until there are no more days of sample left.
- The Metrics file for each iteration is stored in a SQLite3 database along with the number of days removed.
- The days that are removed are logged in the program logfile.
- Purpose of the program was to get a broad overview of how the metric is affected by sample removal for ever-increasing numbers of days of missing data.
- Again, this was a naïve approach to indiscriminately remove data. Each thermostat was subjected to its own indiscriminate and random date removal.

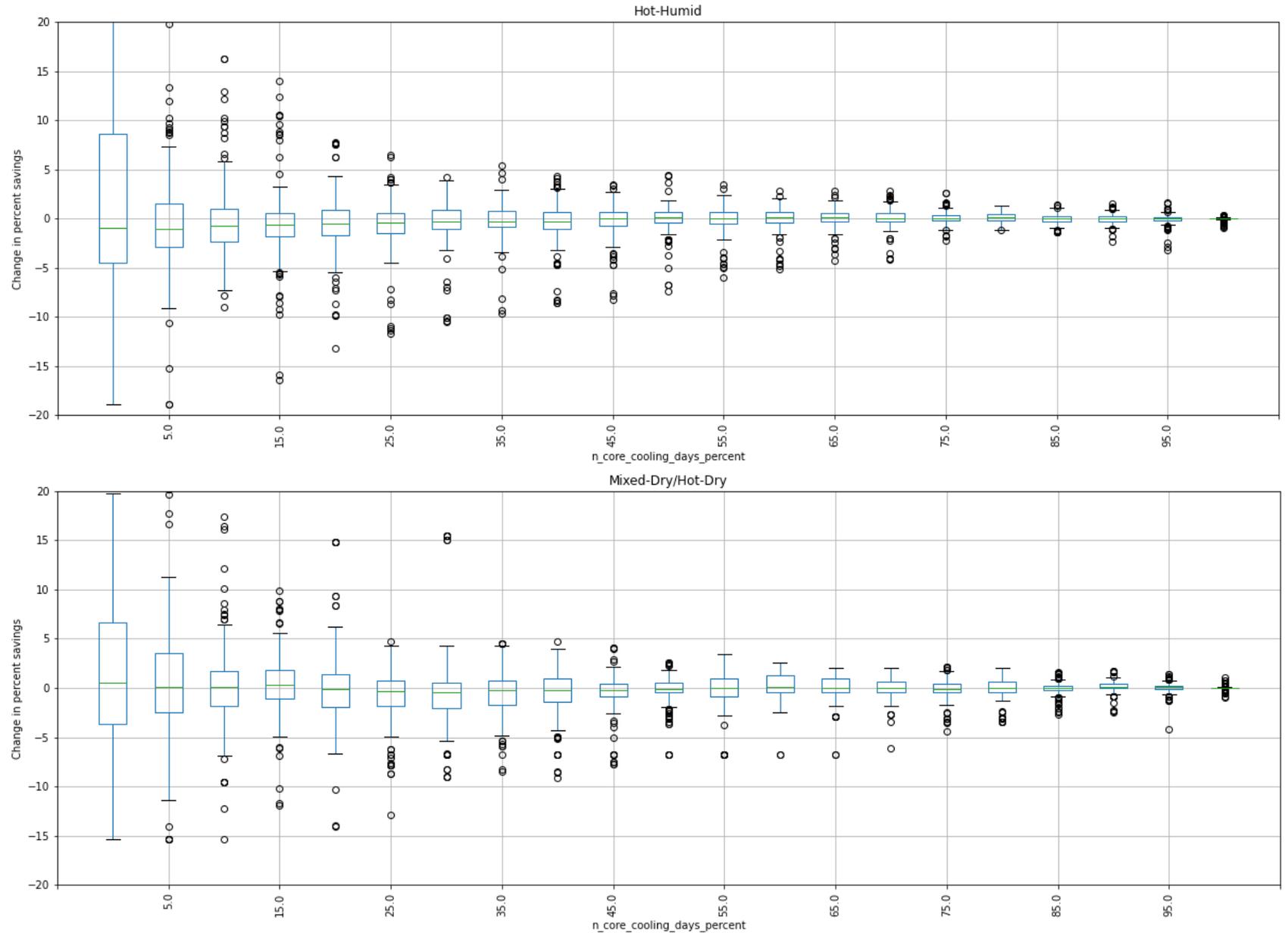


# Heating Savings





# Cooling Savings



## Discussion: Missing data updates and proposed solutions

- Expect to see more impacts of extreme weather events in the next submission. – e.g. large swath of data missing from winter 20-21 in Texas.
- HVAC isn't running during a power outage anyway, so those days wouldn't have information for us anyway. But do we lose internet connectivity with some extreme weather events, even when power is on.
- Back to TX – power was restored before internet, for several days. Though that data might still be anomalous, given recovery of the thermal mass from being very cold.
- Models with more local storage will have less data missing – does that produce bias? Some models will get tossed out more.
- Note that it's a lot harder to do a biased test than unbiased. We could start by throwing out the days with the highest/lowest run time, and would expect to see more effect. Such a test would bracket the effect, at least.
- Why isn't this a post-processing issue? An outage will affect all vendors in an area.
  - EPA has no access to the raw data to compare between vendors.

## Discussion: Missing data updates and proposed solutions

- Distribution of missing data in pool of thermostats – are 90% of thermostats missing fewer than 20 days? If so, our solution is EASY. On the other hand, if a higher proportion of thermostats are missing significant data, we need to work harder
  - Right now, a “missing day” is missing more than 2 total hours of data in the day.
  - Can we develop a standard set of data quality statistics in the software to examine this question?
  - For meter data analysis, looked not at total missing data, but at a measure of whether the range of conditions was covered and/or whether it included a particular range of degree days. E.g. must have something like half of the annual expected degree days. But the essence of suggestion – look at whether data includes most of the range of the conditions at the weather station.
- What about missing days due to missing data from the weather station? One vendor says that it’s not uncommon to find weather files missing a lot of data. (EPA will take a look at this issue.)

## Discussion: Missing data updates and proposed solutions

- NREL doesn't necessarily have cleaned up data available for every station at the time we're doing the analysis.
  - EEweather may need updating with recent changes at NREL and NOAA.
  - [got deep into the weeds of specific weather station here, was not able to follow with the notes] There are bad weather stations (they degrade over time) and good ones (major airports keep theirs up to date).
  - We do check the NOAA quality rating.
  - Curated weather file on the table, but we haven't figured out how much it would cost to do, and whether we can.
  - NREL national solar radiation database is pretty good – there is an API but you need an account. Kevin can share sample Python code. Seems more resilient to weather stations missing data. API call might produce issues, and won't be able to download.
  - Playing around with WeatherSource (free up to an API limit)



## Discussion: Missing data updates and proposed solutions

- Please share more thoughts by email!
- Current interpolation of data gaps not consistent with documentation. This issue has been fixed in V2
- How do we deal with milder climates, which have fewer core days at all? More noise for the same % of the season missing. Possibly different models for why people switch off thermostats – operating as a light switch?
  - A standard based on range of coverage might be more robust against that problem.

## Metric topics from discussion guide comments

- Larger sample sizes supported; mixed feedback on what that should be, mixed feedback on whether a % of population makes sense.
- Question: Would the supplied software handle the selection of a % of thermostats?
  - Answer: no, for that we'd need at least metadata for the entire population of customers. That seems cumbersome and intrusive.
- On the topic of increasing confidence in the ENERGY STAR metric through correlation with meter data, in discussions with DOE. Thanks for the input.
- No new information on metric weighting, but EPA and NRCan are coordinating to try to include Canadian addresses in the analysis. It isn't clear how long this will take.
- We're trying to address issues with tau and with thermostat attrition in time for V2.0 specification.

EPA will be following up with a number of stakeholders in the coming weeks.



## Summary of results from disaggregating product families

- Question: would more granular product families provide more meaningful results?
- Two vendors provided submission files for theoretical sub-divisions of their thermostat population, using characteristics such as:
  - Thermostat model
  - Hardware configuration
  - Software features
  - Sales channel
- Heating and cooling savings for both vendors showed <1% (absolute) difference between different product subsets.
- This variation is not large enough to warrant more stringent rules for product families.



## Discussion: Summary of results from disaggregating product families

- Was there significant difference between the products for those vendors who submitted these data?
  - We don't think we have fully explored this space and may ask for more targeted data at a later date.
- Keep in mind, if we get to a place where we're comfortable sharing the scores the incentives to subsect into smaller product families might be different.