ENERGY STAR Connected Thermostats

CT Metrics Stakeholder Meeting Slides

October 26, 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
Joel Jacob, Ecobee
Jing Li, Carrier
Jason Thomas, Carrier
Frank David, Carrier
Theresa Gillette, JCI
Diane Jakobs, Rheem
Glen Okita, EcoFactor
John Sartain, Emerson
Eric Ko, Emerson
James Jackson, Emerson
Phil Jensen, 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
Ulysses Grundler, Trane
Jeff Stewart, Trane
Sarathy Palaykar, Bosch
Brenda Ryan, UL
Mike Clapper, UL
Alex Boesenberg, NEMA
Ethan Goldman
Jon Koliner, Apex Analytics
Michael Siemann, Resideo
Beth Crouchet, 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
Howard Rideout, Mysa
Peter Gifford, Mysa
Vrushali Mendon, Resource Refocus
Kristin Heinemeier, Frontier Energy
Thad Carlson, Tricklestar

Agenda

- Software Updates: V2.0
- Data Quality: How to handle missing hours of runtime?
- Connected Thermostat use cases
- Multi-vendor test set
- Metrics for Variable Capacity Systems
  - Metric updates and underlying assumptions
  - Savings mechanisms
  - Average capacity factor
  - Recovery periods

Software Updates: V2.0

- Added two-stage files for testing
  - Split off single stage and multi-stage data
  - Created tests for multi-stage
  - Cleaned up tests
- Cleaned up code
  - Cleaned up warnings around pandas versions and which versions are accepted
  - Removed dead code like setpoints and other dead code
  - Simplified code
  - Explicitly close file handles in stations
  - Tweaked warning suppression and made it more explicit
  - Removed warnings for expected items (zero division)
  - Updated naming of variables to make their function more clear

How to Handle Missing Data in Thermostat Interval Files

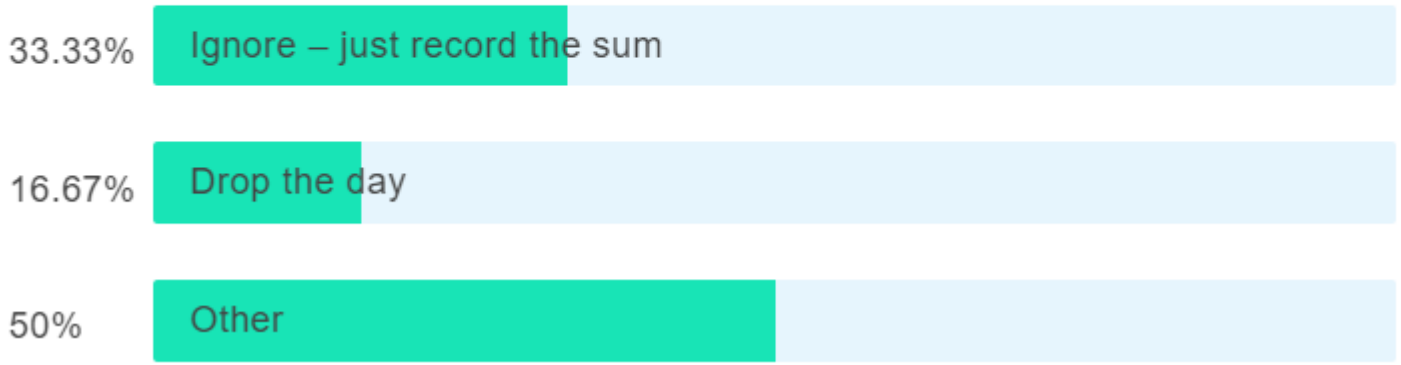
- The version 1 data format currently consists of both hourly and daily data:
  - Temperatures and set points are recorded hourly
  - Runtimes are recorded daily
- The version 2 data format consists of all hourly data.
- Because the metric calculations use daily runtimes, the Version 2 software converts the hourly runtime data to daily totals.
- This requires a decision of what to do if there are any missing runtime data during a day:
  - Ignore the missing values and just record the sum
  - Fill the missing hours using interpolation (up to some maximum gap)
  - Drop the day from the calculations
- To keep consistency with Version 1 it would be best to use rules for missing hourly data that are as close as possible to what vendors are currently implementing when creating the version 1 interval data.
- So what are vendors currently doing?

Discussion: Missing data

- One vendor: Less than 0.001 of days are missing an hour inside a day there is otherwise data for
- Another: Similar, very small issue
- Poll:

1 of 8. How do you currently deal with missing runtime data when calculating the daily value?

Multiple choice with single answer



Connected Thermostat Use-Cases

ENERGYSTAR's goal is to maintain a level playing field for vendors. To this end, it will consider various strategies to prevent distorted results.

CTs are being used in situations beyond the simple single-family home. These alternative use-cases can influence the vendor's calculation of the metric in two ways:

- Certain use-cases will result in misleading metrics
- A vendor will have a distribution of use-cases that differ from other vendors

Does the metric accurately capture CT performance in common use-cases?

How are CTs being used?

- Building types (single-family, apartments, vacation homes)
- Ownership scenarios (one per home, multiple)
- Other (home and customer different)

Does the existence of these configurations suggest alternative sampling procedures?

Which inputs will generate a misleading metric calculation?

- Unrealistic comfort temperatures?
- Unusual indoor or outdoor temperatures?
- Unusual relationship between temperature and runtimes?
- Other?

#	Use Case	Does Current ES metric make sense? (Y/N/ Maybe)	Notes/ Explanation/ Drawbacks	Fraction of CTs in this category	Type of problem (Sample or metric calculation)

#	Use Case	Does Current ES metric make sense? (T_{90}, runtime, ΔT)	Notes/ Explanation/ Drawbacks	Fraction of CTs in this category	Problem in Sample or Metric?
1	SF detached home (1 tstat)		Our base case, single or dual-speed, unspecified auxiliary heating source,		
2	Vacation home				
3	SF home (>1 tstat)				
4	Multiple thermostats on a single account		(like a motel? Dorm?)		
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		Variation on the Apartment idea above		
9	Variable capacity heating or cooling		We’re investigating a metric for effectiveness of variable capacity		
10	2-stage system				
11	Dual fuel		They are currently excluded		

Discussion: Connected thermostat use cases

- Signs of vacation homes vendors use to estimate: >25% of days non-occupied; multiple thermostats with one account
  - Some spread, but most chose 6-20%. Overall estimate about 10% (order of magnitude)
  - Does the current metric make sense? Is weather accurate – no one said no (but in another meeting we heard from another vendor who said yes)
  - Clearly a different primary schedule than other homes
  - If the home is occupied very rarely, may affect comfort temperature. If not, home gets very high score. But, how would they have managed the temperature w/o a smart thermostat, but they might have used a higher unoccupied temp either by accident or deliberately. If comfort temp is in setback state, the score will be negative (extreme cases might get kicked out)
  - Comfort temperature all the time baseline is less realistic
  - Would difference in vacation home population size distort scores?
  - .

Discussion: Connected thermostat use cases

- vacation homes
  - If customers turn systems off when away (not where pipes will freeze), how does that impact the metric? Not included in core heating/cooling days, so indoor temps not part of comfort temp determination.
  - Impact of indoor temperature of other days may be outsized if recovering from deep setbacks; would look like noise or systematic distortion in our model – if noisy enough thermostat not included in statistics.
- Homes with more than one thermostat
  - Homes with multiple thermostats and multiple systems are relatively common
  - Homes may also have a single thermostat controlling a single system with zone dampers (may have external temperature sensors in additional zones)
  - 3rd party vendors can't tell if 2 thermostats are controlling a single HVAC with zone dampers

Discussion: Connected thermostat use cases

- Homes with more than one thermostat
  - Home with 2 systems and 2 thermostats: Large homes with separate wings operate fairly separately.
  - In 2 story homes, upstairs thermostat carries more of the cooling load and downstairs more of the heating load. The thermostat carrying less of the load would probably still have similar core days but may be noisier or otherwise be less likely to be included in the statistics.
  - Vendors that can tell which homes have more than one thermostat and system do indeed see this effect – one typically looks less linear in a given season
  - This is common, and there is a significant variation between vendors
  - So, is it a problem? Not clear. Baseboard thermostat example, with a home energy meter as well. Whole home load shed about 50% of what you'd expect from just the controlled system – this is a takeback effect. Relevant where there's more than one type of heating and they aren't controlled by a coordinated system.
  - May also affect the tau estimate, because there's a heat gain that isn't controlled by the thermostat.

Discussion: Connected thermostat use cases

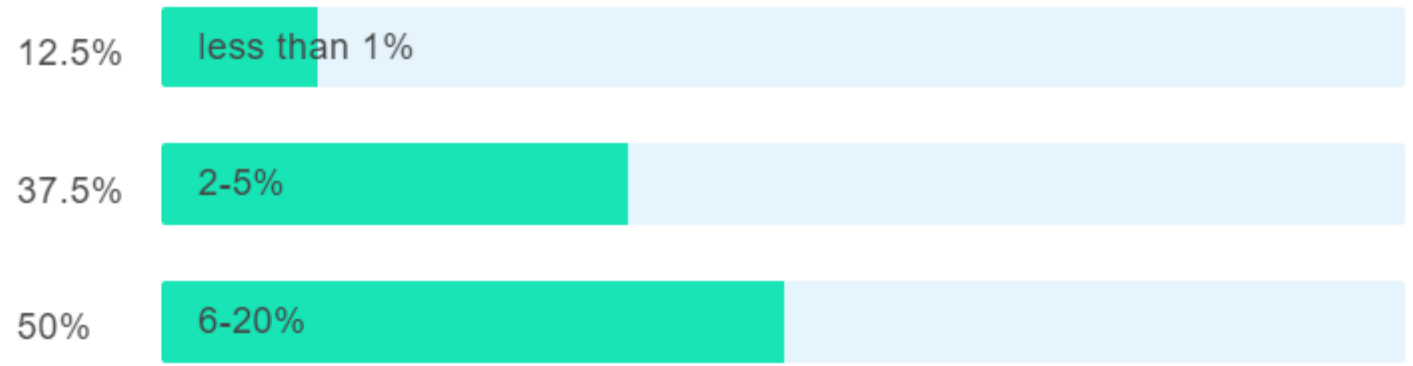
- Poll results:

2 of 8. What fraction of your company's thermostats fall in Single Family detached home (1 tstat)?

Multiple choice with single answer

3 of 8. What fraction of your company's thermostats fall in Vacation homes?

Multiple choice with single answer

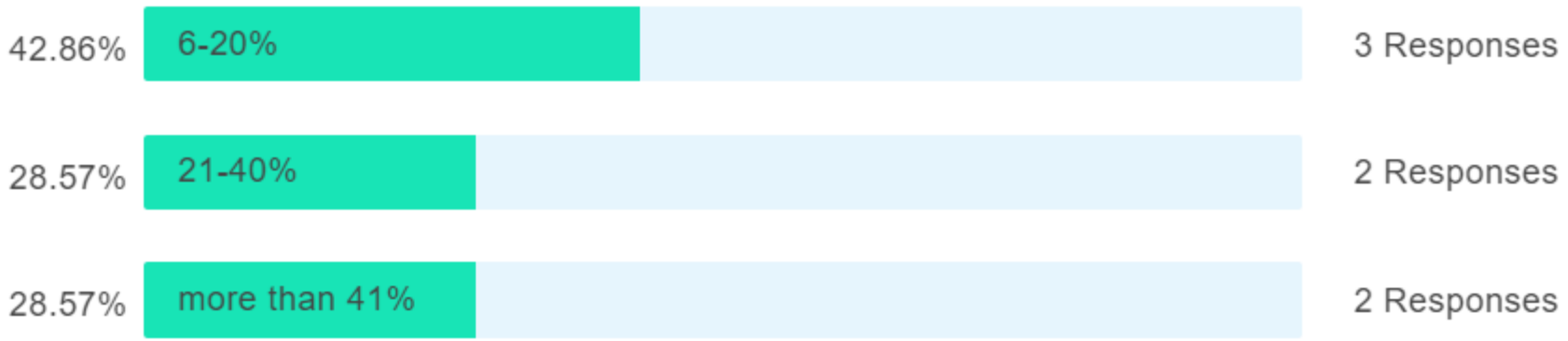


Discussion: Connected thermostat use cases

- Poll results:

4 of 8. What fraction of your company's thermostats fall in Single Family home (more than1 tstat)?

Multiple choice with single answer



The Multivendor Test Set

ENERGYSTAR would like to create a test set containing data from multiple thermostats to test and debug new versions of the software. The objective is to speed releases of future versions, reduce errors encountered by vendors, and fairly treat all vendors.

- Participation is voluntary
- The test set will contain thermostat interval data and metadata used to calculate the metric in V2.0 format
- EPA proposes these underlying principles:
  - No PII; vendors will submit anonymized data
  - Vendor anonymity will be preserved by requiring a minimum number of participants
  - LBNL will create and maintain the test set
  - The test set will be accessible by only EPA and designated contractors
- Should there be other principles?

Other Considerations

- Number of files per vendor
- Oversampling of specific technologies
- Mechanics of submission
- Updates
- Sunsetting / Retraction
- Security
- Other?

Next Steps in Multivendor Test Set

- Should LBL draft a common agreement?
- Decide on test set size and characteristics
- Define storage mechanics and access
- Schedule driven by V2.0 revision schedule – would like to test software before February resubmission

Discussion: Multivendor database

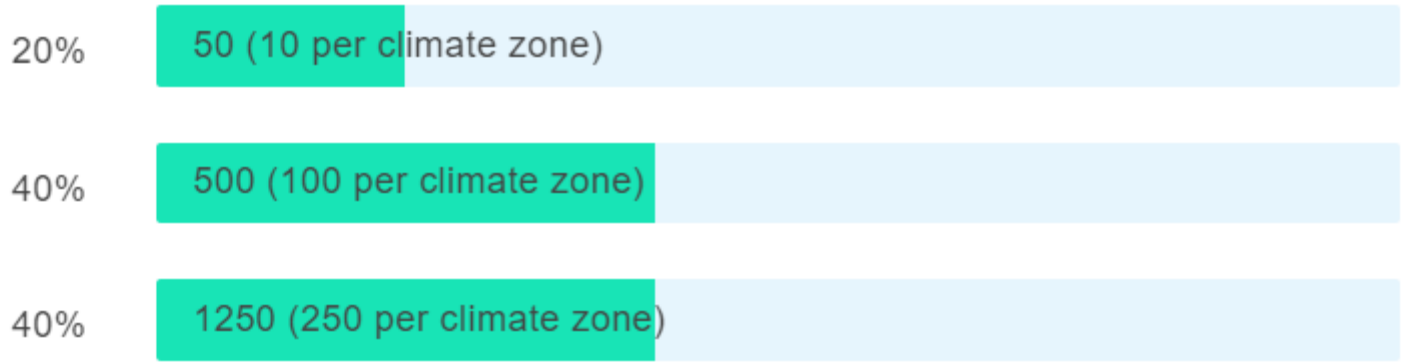
- Might be more comfortable talking one on one, and could then bring someone from legal
- Would it help to bring a strawman to such a meeting? This info here might be enough
- General comment: would you want data sets to be representative? Or a slanted sample? Or what? Can we limit sets based on reidentification risk (e.g. how many homes with heat pumps in a given zip code? The more metadata you need, the larger this problem it is. Can we gang zipcodes together?)
  - Not critical to have representative sample, though we do need a sample from each climate zone
  - The point here is to exercise the software fully, so more important to have a variety
  - The only metadata we need are what's in the 2.0 files: HVAC type and zip code
- Any elements that aren't important?
  - Regular updates more of a problem
  - Might want expiration date = sunsetting
- Don't use 2020 data!

Discussion: Multivendor database

- Poll Results:

5 of 8. How many files would you be willing to provide for the test set?

Multiple choice with single answer



Update on progress toward a variable capacity system metric

- Workshops on communicating controls for variable capacity HVAC were held on April 3rd and April 6th
- Our team has been exploring new metrics, when they would be applied, and baselines for performance comparison
- Scope of conversation hasn't changed:
  - Zoning not currently being considered
  - Focused on the communicating controller's ability to take advantage of the opportunities for efficiency that variable capacity systems make available, not the HVAC system itself

Underlying assumptions

- Variable capacity systems can save energy compared to a single- or two-stage system because:
  - Matching HVAC operation to the load at reduced capacity where the system operates more efficiently
  - Avoid energy associated with repeated start-ups by running at lower capacities for longer periods
- Controls can affect how well systems take advantage of these capabilities, and can
  - Avoid using high capacity to recover from scheduled setback by starting recovery earlier

What makes a savings mechanisms a good fit for ENERGY STAR?

- Affected by control strategies and algorithms, and in particular the control strategy to match load, efficiently ramp, learn building and occupancy effects, and control equipment in a way that optimizes efficiency and consumer comfort
- Able to be examined with either prescriptive requirements (better to avoid) or with data as we've talked about it
- Can identify a reasonable baseline to compare data to
- Some products on the market are significantly better at it than others
- Actual savings are significant in each home and/or apply to every home so that the total savings are significant

Savings mechanisms and potential metrics identified during the workshop

- Savings mechanisms identified:
  - Running at lower capacity states *when possible*
  - Avoiding short cycling
  - Set back and set up
  - Fan control

- Potential metrics:
  - Average capacity factor
  - Temperature at which system starts running uninterrupted for long periods of time
  - When cycling, unit comes on at a lower capacity
  - Short-cycling fraction

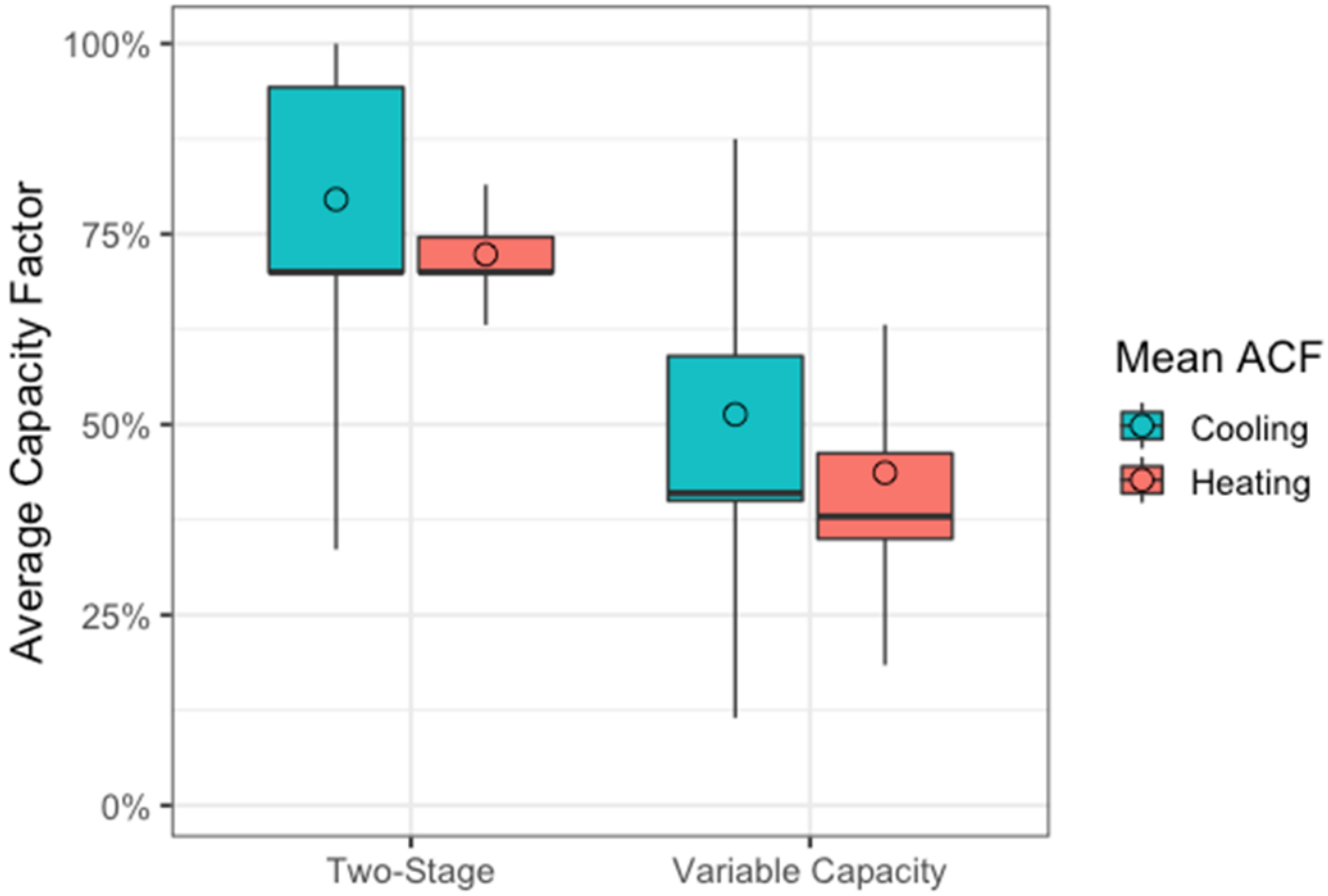
Average Capacity Factor (ACF)

- Discussion from previous meetings:
  - Depending on the type of equipment, there is a small or medium sized efficiency advantage from using a lower capacity for a longer time
  - Also comfort and equipment longevity advantages
- Metrics to use field data to distinguish units that do a better job
- As with RHU, start with a basic calculation, and then look for where the results are meaningful

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

Average Capacity Factor Summary Statistics Across All Systems

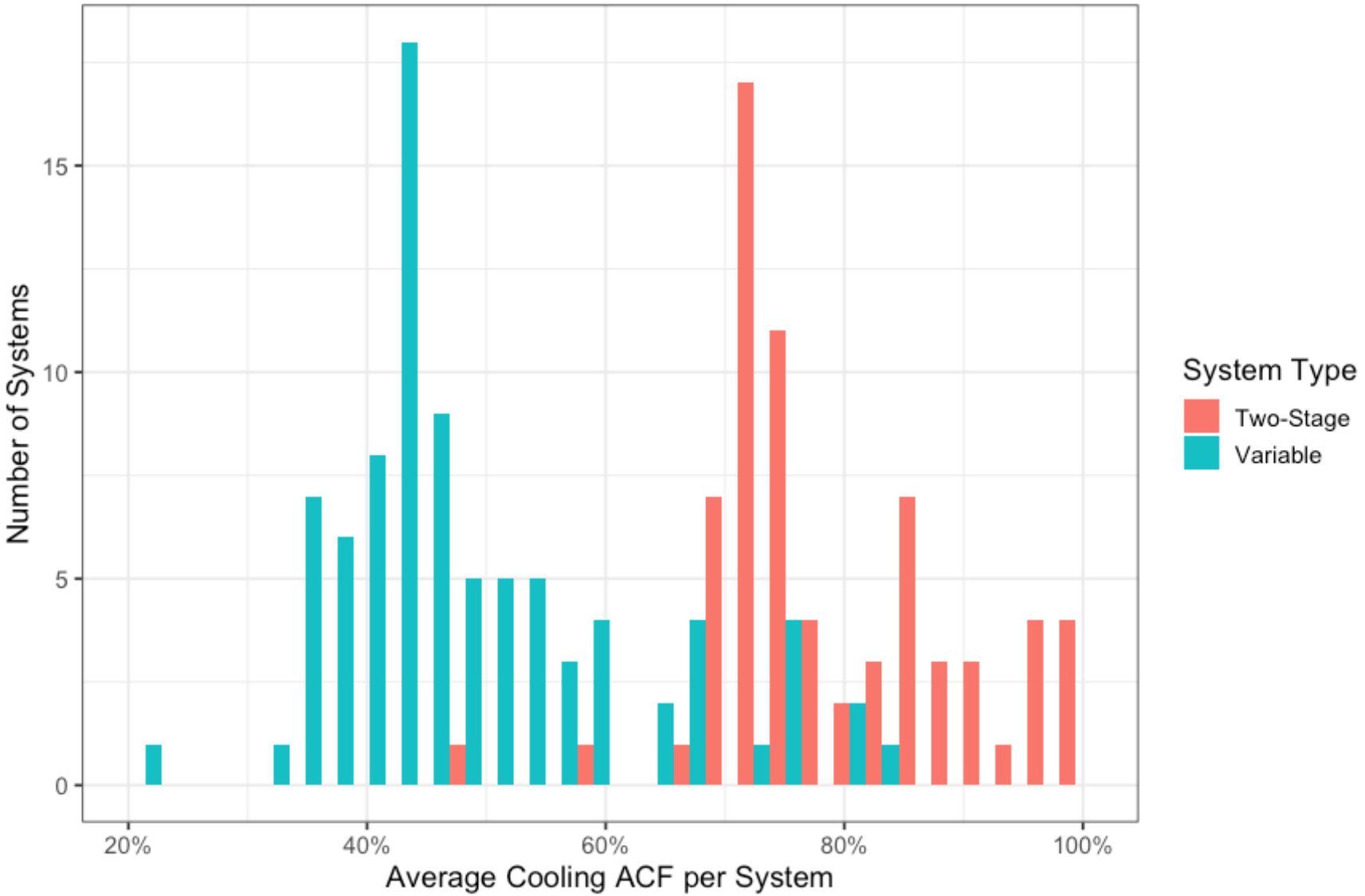
- As expected, variable capacity systems have lower ACFs than two-stage systems for both heating and cooling
- This comparison is to explore the metric; eventually we will compare ACFs of variable capacity systems to see if it can differentiate efficient operation



* 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

Average Cooling Capacity Factors by System

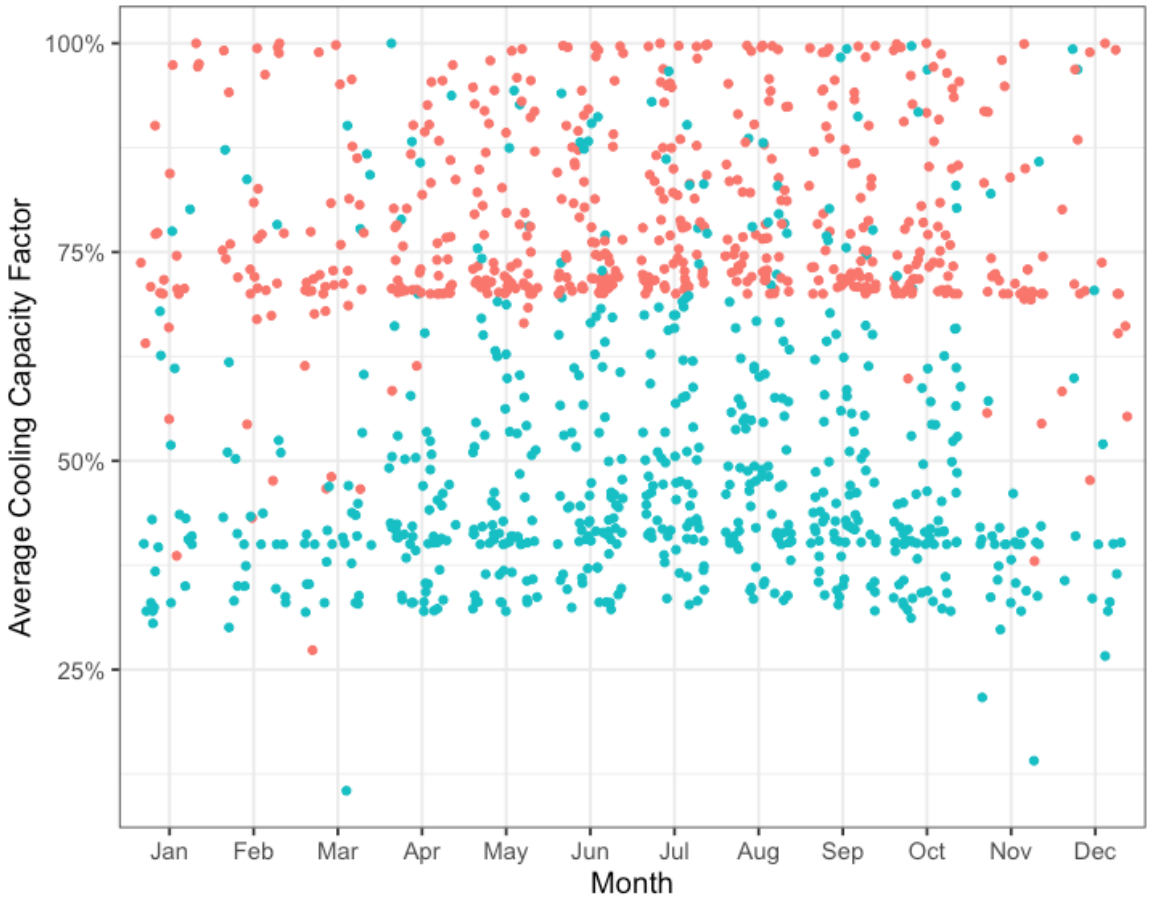
- Clear dispersion between two-stage and variable capacity systems
- No obvious ”line in the sand” to determine which variable capacity units are operating efficiently vs. inefficiently



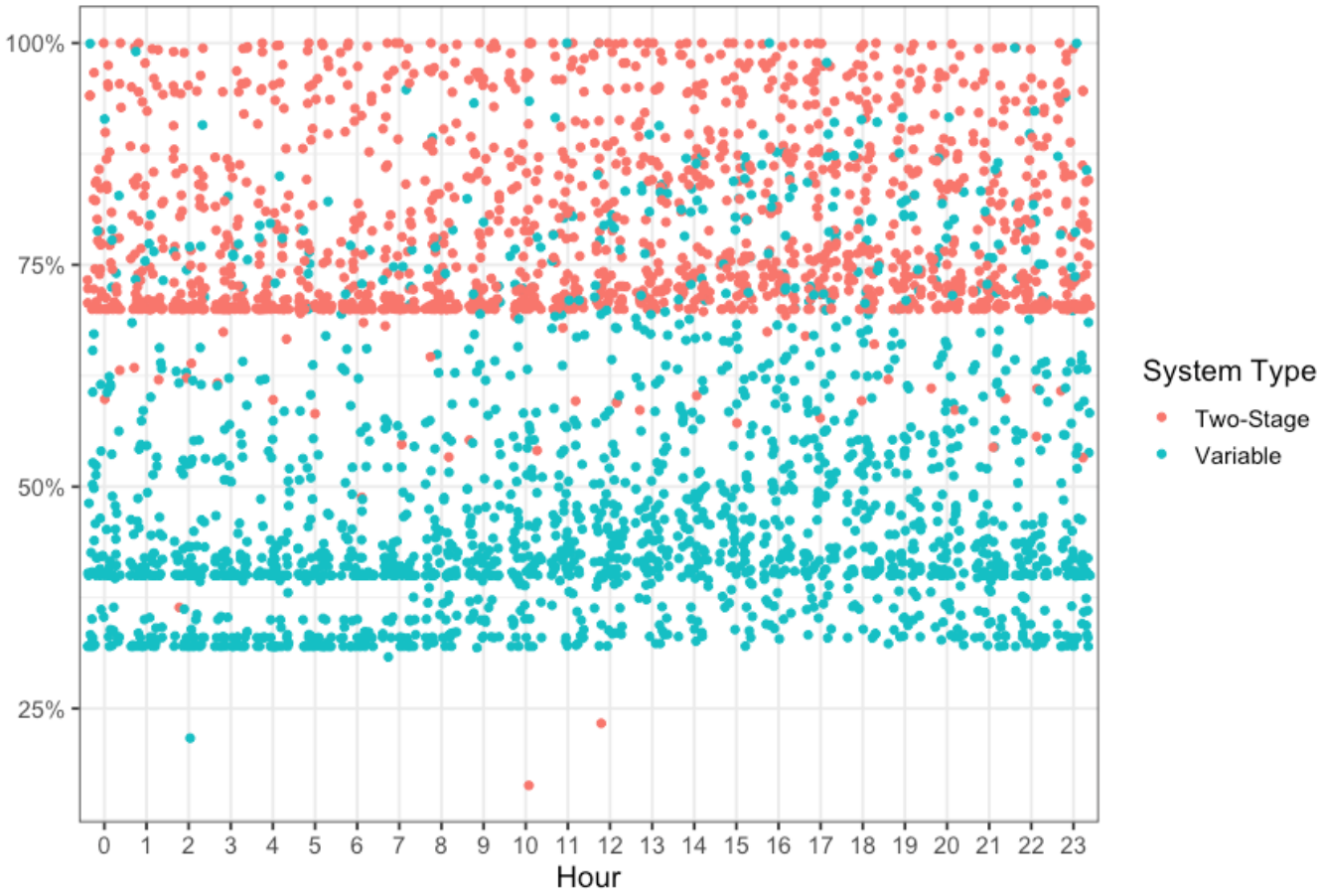
Calculating Average Cooling Capacity Factor by Month and Hour

- We see a slight lift in average capacity factors for variable capacity systems during the summer months and the afternoon hours of the day

Average Cooling Capacity Factor by Month

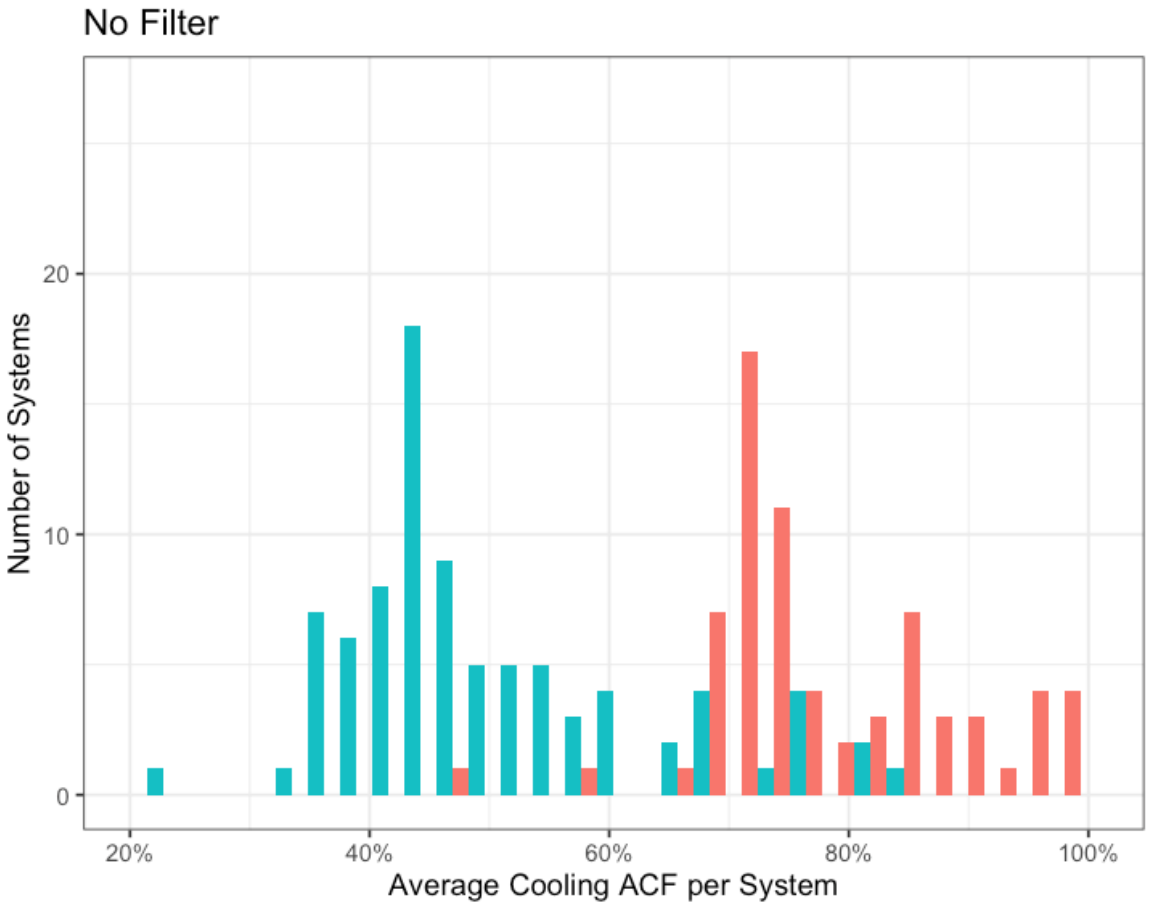


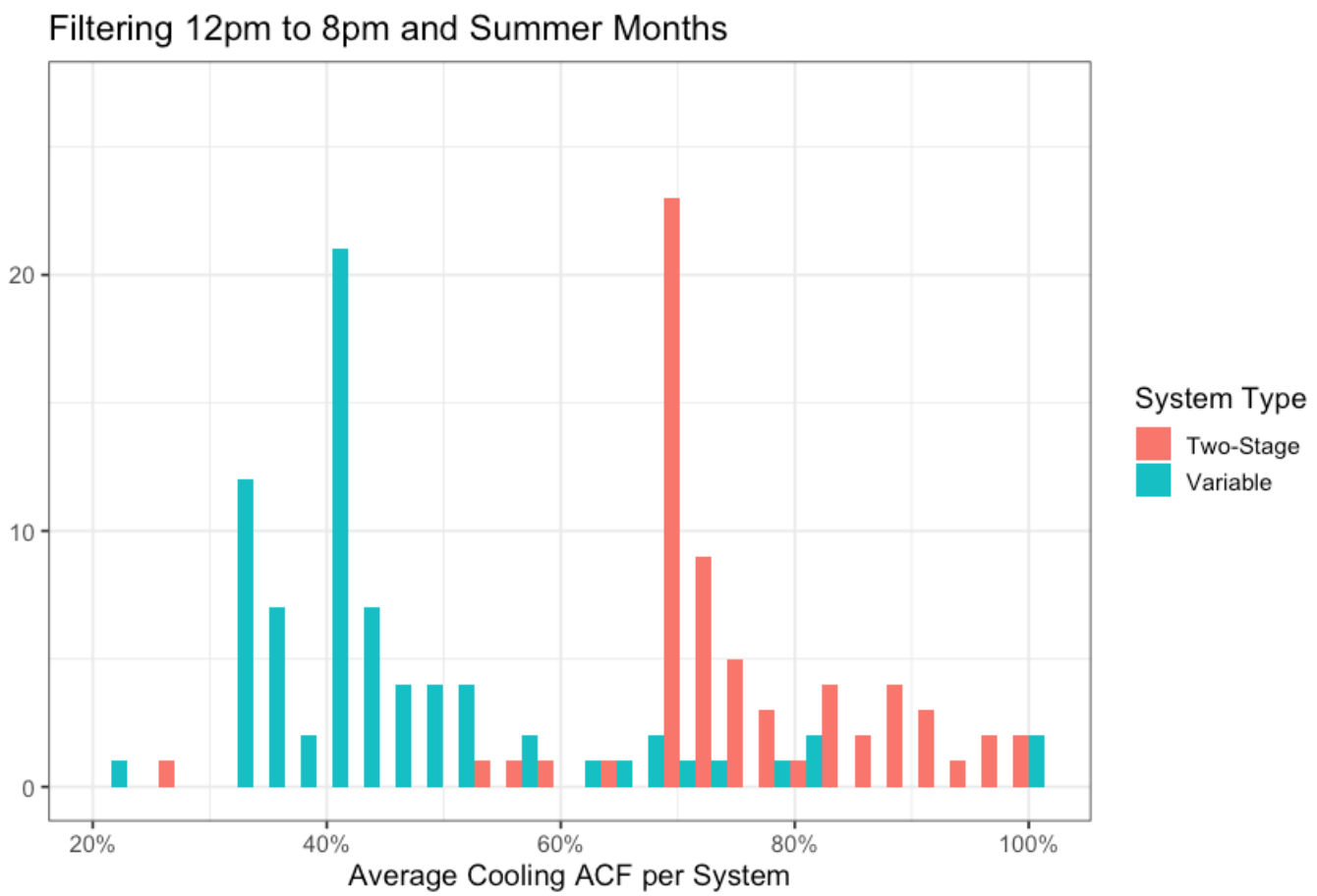
Average Cooling Capacity Factor by Hour



Filtering Out 12pm to 8pm and Summer Months

- However, when we filter out periods with higher average cooling capacity factors there is only a slight shift to the left for variable capacity systems





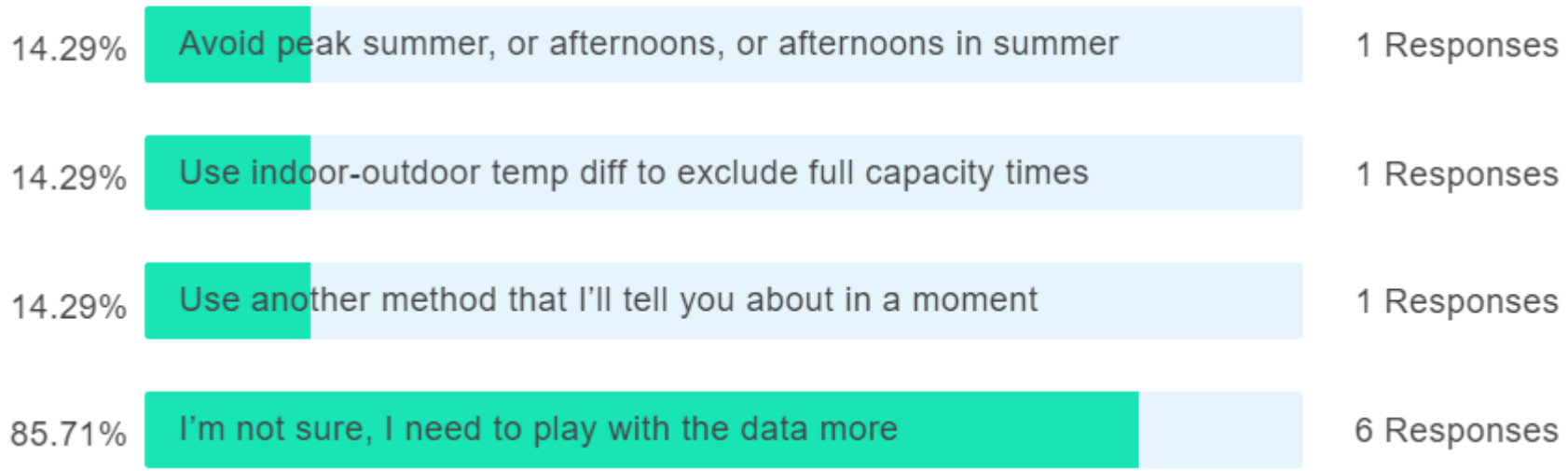
Poll

- What time periods or conditions do you think should be filtered out of an Average Cooling Capacity Factor calculation?
  - Avoid peak summer, or afternoons, or afternoons in summer
  - Use indoor-outdoor temperature difference to exclude times the system needs full capacity
  - Use differences in indoor temperature and set temperature to exclude recovery periods
  - Use another method that I'll tell you about in a moment
  - I'm not sure, I need to play with the data more

Poll

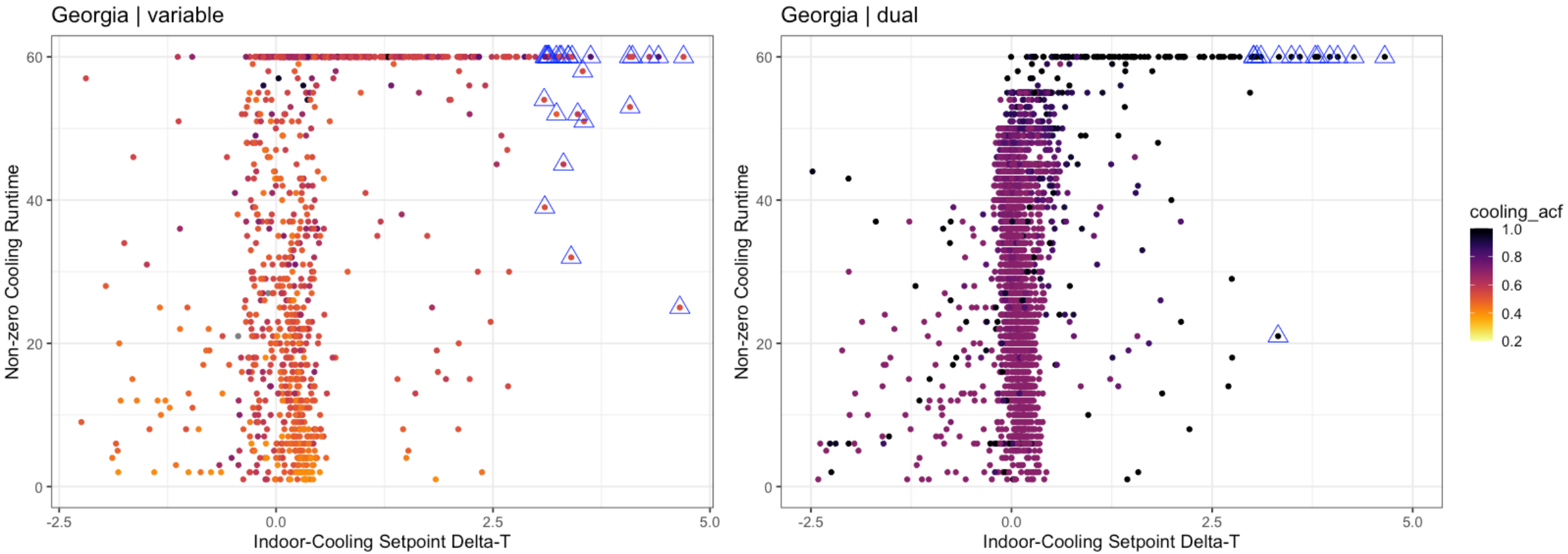
6 of 8. What time periods or conditions do you think should be filtered out of an Average Cooling Capacity Factor calculation?

Multiple choice with multiple answers



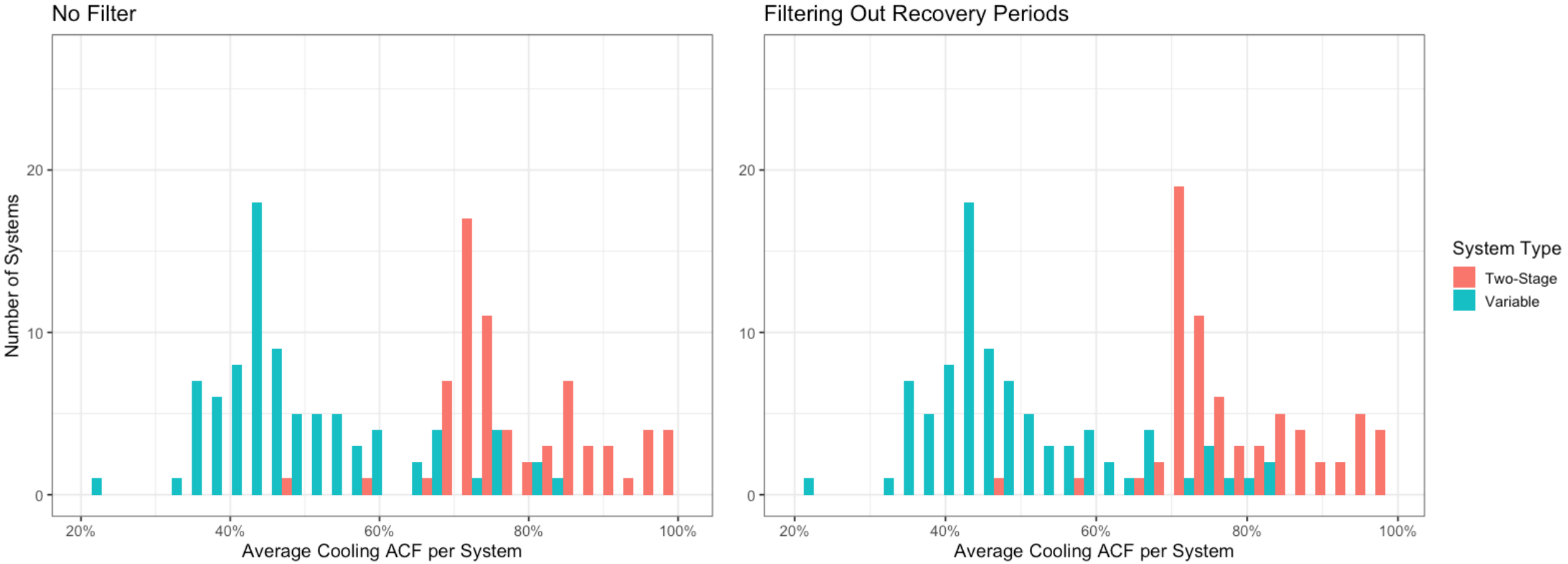
Identifying Recovery Periods

- The charts below depict cooling runtime as a function of the cooling setpoint and the indoor temperature difference for two example systems.
- The points highlighted by the blue triangles represent potential recovery periods when the difference between the indoor and setpoint temperatures were greater than 3 degrees.



Filtering Out Recovery Periods

- When we filter out hours where there is at least a 3-degree difference between the average indoor setpoint and the average indoor temperature (recovery period) there is not much of a difference



Poll

- What difference between indoor temperature and indoor cooling setpoint should be used to define a recovery hour?
  - 2 degrees
  - 3 degrees
  - 4 degrees
  - 5 degrees
  - I'm not sure

Poll

7 of 8. What difference between indoor temperature and indoor cooling setpoint should be used to define a recovery?

Multiple choice with single answer

Questions for Discussion

- Under what circumstances or conditions should the ACF metric be applied?
  - We shared examples based on time of year, time of day, and recovery periods, but are there other ways to filter the data?
- Is the ACF metric a good-enough metric to continue exploring?

Poll

- How promising do you think the Average Capacity Factor is for gauging the performance of variable capacity HVAC systems?
  - Not promising
  - There's promise but it still needs a lot of work
  - It seems promising

*This poll was skipped due to limited time

Questions for Discussion

- Does the group have any thoughts or concerns about the other metrics that are being explored, and is it possible to record the timestamps for each capacity call that would be necessary for these metrics?
  - Temperature at which system starts running uninterrupted for long periods of time
  - When cycling, unit comes on at a lower capacity
  - Short-cycling fraction
- Note that the data we are now exploring doesn't include time stamps, just classifies calls for cooling and heating into which hour they occurred, so we can't tell the order of the calls. We would need more detailed data to explore these metrics.

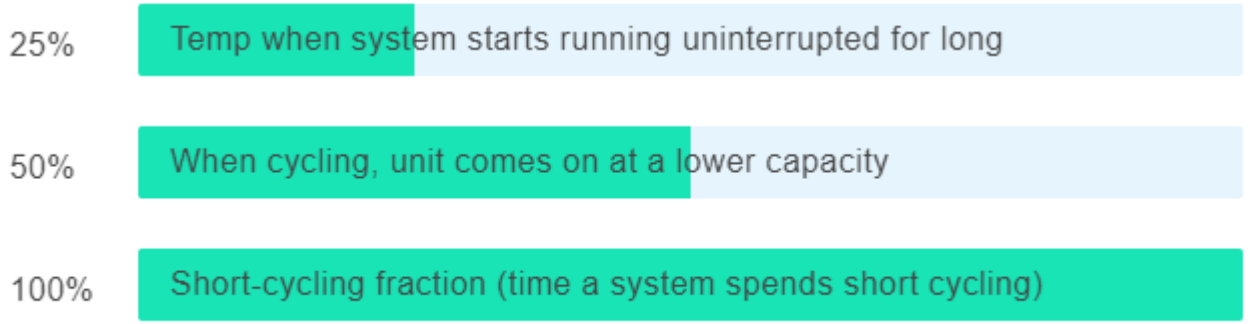
Poll

- Which of the other metrics seem the most promising to you?
  - Temperature at which system starts running uninterrupted for long periods of time
  - When cycling, unit comes on at a lower capacity
  - Short-cycling fraction (fraction of time a system spends short cycling)

Poll

8 of 8. Which of the other metrics seem the most promising to you?

Multiple choice with multiple answers



Discussion: Variable capacity systems

- Why should you screen out any periods? The thermostat has an effect on how well or badly recovery is done to save energy.
  - What about very hot/very cold times? Maybe.
  - Second vote to keep recovery periods in – that's when the control has an opportunity to be smart
  - Pushback – there is short cycling that may be under control of the thermostat. (Data from field suggests that some brands do better at avoiding short-cycling than others)
  - Pushback on that – note that location of the thermostat can produce short cycling
- ID'ing recovery periods:
  - Look at distribution of data and cut it at some quantile.
- What about looking at ACF for specifically long run time hours?
- If we are to look at the alternative metrics we need time stamped data
  - Could also count cycle starts per interval
- For temp at which system starts to run uninterrupted might be able to find in the distribution even without time stamps – plot run times against outside temp or demand of some kind