ENERGY STAR Connected Thermostats

Stakeholder Working Meeting

December 17, 2019

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
Jing Li, Carrier
Brian Rigg, JCI
Theresa Gillette, JCI
Kurt Mease, JCI
Diane Jakobs, Rheem
Carson Burrus, Rheem
Chris Puranen, Rheem
Glen Okita, EcoFactor
Brent Huchuk, ecobee
John Sartain, Emerson
James Jackson, 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, IRCO/Trane
Jeff Stewart, IRCO/Trane
Mike Caneja, Bosch
Sarathy Palaykar, Bosch
Brenda Ryan, UL
Mike Clapper, UL
Alex Boesenberg, NEMA
Ethan Goldman, Recurve
Jon Koliner, Apex Analytics
Michael Siemann, Resideo
Aniruddh Roy, Goodman/Daikin

Agenda

- RHU: Further exploration
  - 5F vs 15F bin(s)
  - Regional vs. national
  - Weighting for improved relevance
- NEEA project and coordination with EPA
- Savings metric improvements: what to focus on?
- Software changes heading to 2020 revision
  - Simple: application of 2 stage idea
  - Harder: changes for 2020 – how to manage, what to include

RHU2: Recap

- Previous meetings: shared RHU2 results, discussed need for additional sampling of heat pumps for reasonable results
- Metric proposal: RHU2 in 30F-45F bin, upper 95th confidence limit of the mean ≤ 0.2
  - Questions about national vs. regional
  - Questions about 5F or 15F bin(s)
- Weighting? If there's 0% ER off time in the day, the ER minutes don't count (broken system); if there was heat running only 50% of the time, then you'd get half weighting. If heat running very little, it would be heavily weighted.
- Previously discussed outlier considerations
  - User settings that could cause that to happen? e.g. compressor lockout temp, the way you set the flipped behavior, etc.
  - Comfort setting may also effect this – which contractors would like.
  - Use an outlier threshold – if lots of compressor runtime, no need to include it in the average.

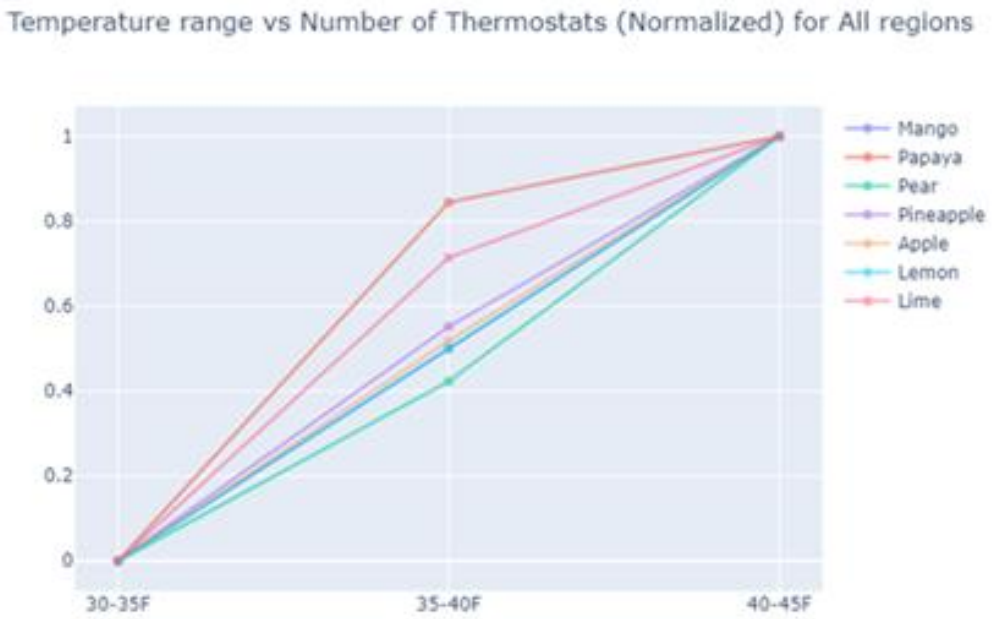
RHU2: National vs. Regional

	All		Mixed Humid		Very Cold		All		Mixed Humid		Very Cold	
	30-40F n	30-40F ub	30-40F n	30-40F ub	30-40F n	30-40F ub	40-45F n	40-45F ub	40-45F n	40-45F ub	40-45F n	40-45F ub
Mango	160	0.086	87	0.090	31	0.110	188	0.1	82	0.074	27	0.125
Papaya	48	0.202	0	nan	0	nan	53	0.164	0	nan	0	nan
Pear	63	0.089	38	0.093	2	0.164	72	0.078	37	0.063	2	0.009
Pineapple	78	0.187	56	0.189	14	0.265	91	0.145	58	0.15	14	0.173
Plum	5	0.799	0	nan	0	nan	5	0.846	0	nan	0	nan
Apple	58	0.138	26	0.125	3	0.215	71	0.115	26	0.081	3	0.259
Grape	7	0.119	0	nan	0	nan	11	0.061	0	nan	0	nan
Lemon	25	0.505	12	0.603	1	0.216	32	0.485	13	0.462	1	0.115
Lime	100	0.160	57	0.129	8	0.562	107	0.147	56	0.131	8	0.431

- Data from July 2019 Submission- (2 sets used V1.5, had no RHU2 data)
- Impact of more samples in Mixed Humid region: low upper 95th confidence mean values

RHU2: Wider temperature bin analysis – All regions

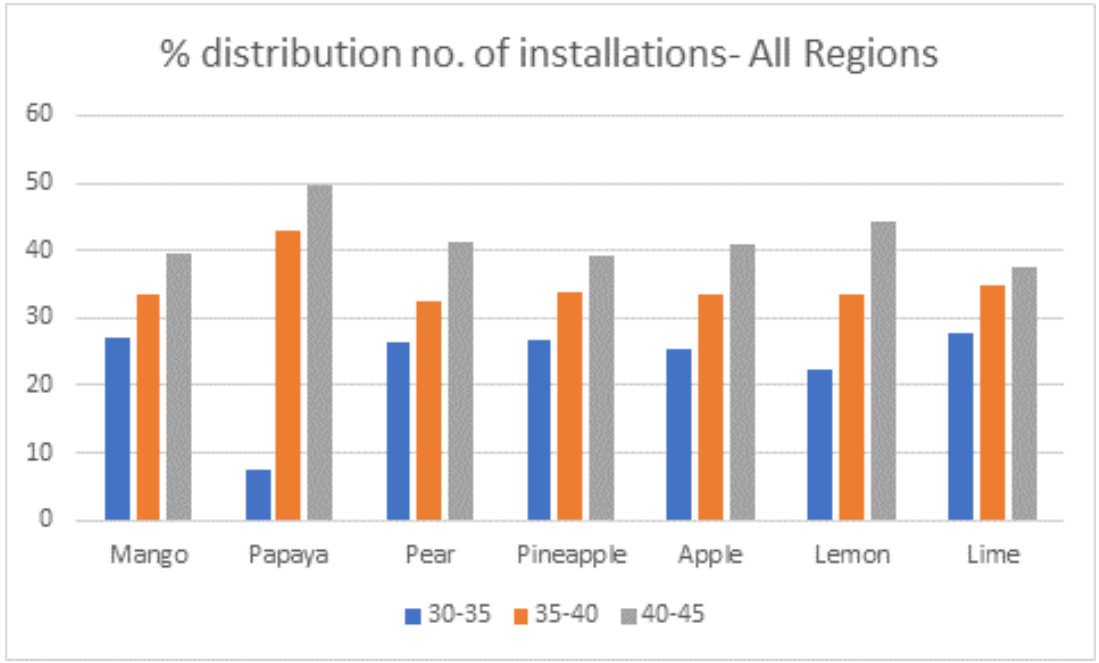
All	30 35F n	30 35F ub	35 40F n	35 40F ub	40 45F n	40 45F ub	30 40F n	30 40F ub
Mango	128	0.105	158	0.081	188	0.1	160	0.086
Papaya	8	0.235	46	0.198	53	0.164	48	0.202
Pear	46	0.135	57	0.075	72	0.078	63	0.089
Pineapple	62	0.191	78	0.178	91	0.145	78	0.187
Apple	44	0.166	58	0.127	71	0.115	58	0.138
Lemon	16	0.548	24	0.511	32	0.485	25	0.505
Lime	79	0.174	99	0.15	107	0.147	100	0.16



- Gradual increase in the number of samples as temperature increases
- Good news: No great difference in pattern between vendors

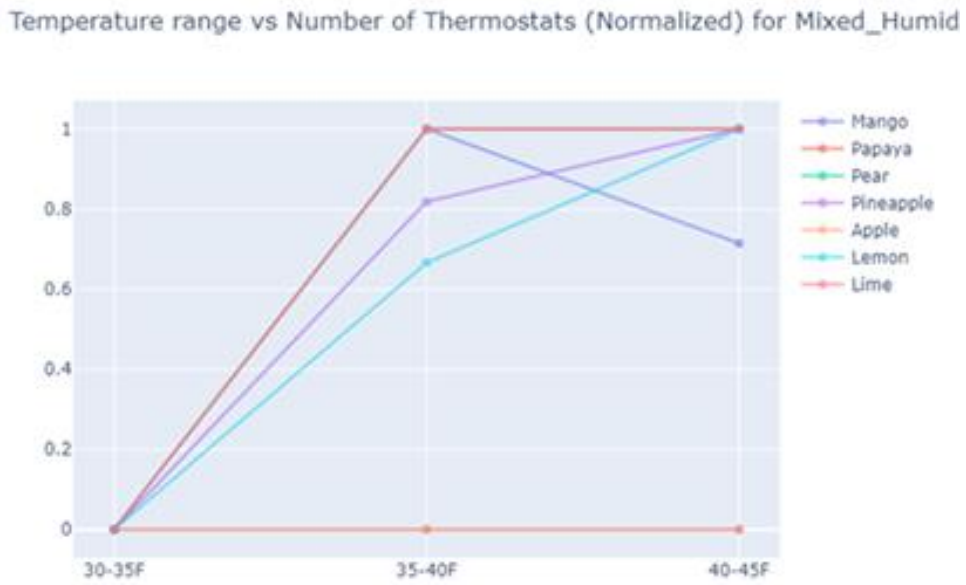
Percentage wise distribution of samples in All regions

All	30 35F		35 40F		40 45F	
	n	%	n	%	n	%
Mango	128	27.00	158	33.33	188	39.66
Papaya	8	7.48	46	42.99	53	49.53
Pear	46	26.29	57	32.57	72	41.14
Pineapple	62	26.84	78	33.77	91	39.39
Apple	44	25.43	58	33.53	71	41.04
Lemon	16	22.22	24	33.33	32	44.44
Lime	79	27.72	99	34.74	107	37.54



RHU2: Wider temperature bin analysis – Mixed Humid region

Mixed Humid	30 35F n	30 35F ub	35 40F n	35 40F ub	40 45F n	40 45F ub	30 40F n	30 40F ub
Mango	72	0.122	86	0.080	82	0.074	87	0.09
Papaya	0	nan	0	nan	0	nan	0	nan
Pear	32	0.126	37	0.084	37	0.063	38	0.093
Pineapple	47	0.210	56	0.188	58	0.150	56	0.189
Apple	23	0.153	26	0.112	26	0.081	26	0.125
Lemon	10	0.607	12	0.604	13	0.462	12	0.603
Lime	53	0.136	56	0.108	56	0.131	57	0.129



- Barring Papaya, nothing that adds to discussion

Discussion: RHU2

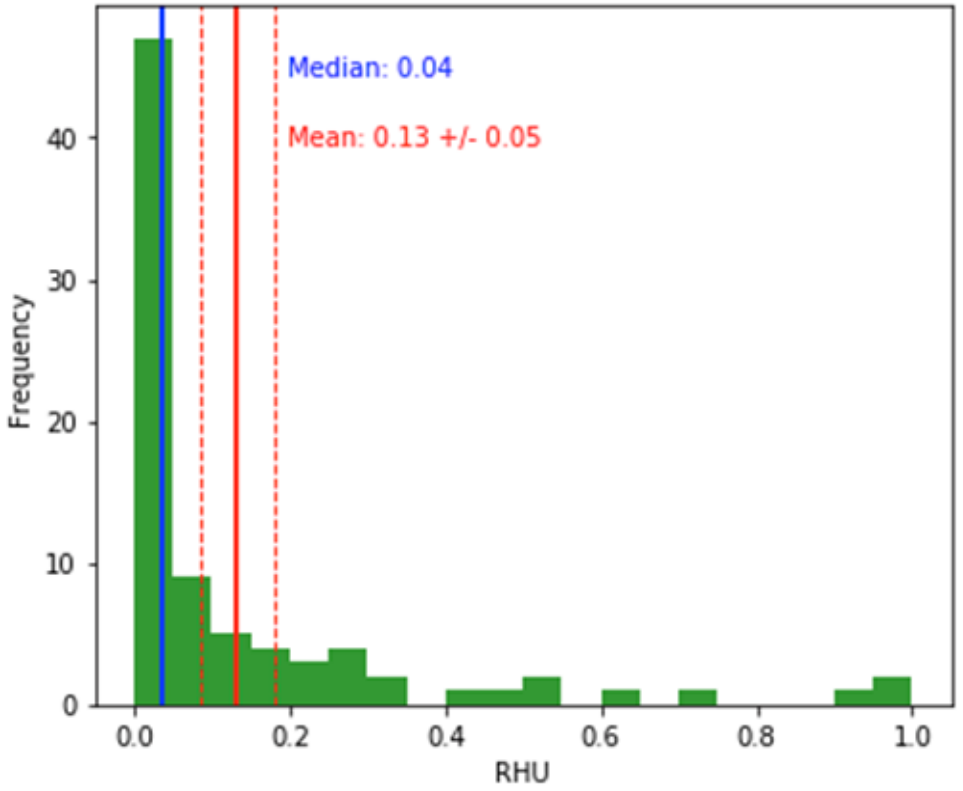
- Reiteration that in a 15F wide bin, the distribution of installations in different parts of that bin, or in different regions within the same bin, may have a significant effect on results. How would we decide if there's a problem?
  - Are the installations from different vendors in a given bin similarly distributed across regions? (Check resubmitted data for this.)
  - Within a 15 F bin, what is the mean temperature installations are experiencing? (Can't tell from resubmission data.)
  - Percent of sample in each of the three 5F bins for each vendor; compare
  - Average outdoor temperatures across all installations in each temperature bin. (Can't tell from resubmission data – not available per bin, just average for vendor's sample.)
- If we use 5F bins, different threshold for each bin?
- Large samples help because assuming all vendors have installations that are in the same population, larger samples will be a more accurate representation.
- Compare one year to another year to see how robust the relationships are – but is that meaningful? Many things may cause them to change. We only have historical data for RHU, not RHU2, and also not a heat pump oversample so numbers are small in some regions.
- Late breaking vendor input: for their installations, RHU variation not large based on regions or temps w/in wide bin

Discussion: RHU2

- As we start to use a bunch of bins, does it make sense to pursue a regression-based metric instead.
  - Data may be sparse
  - Expectation of linearity?
- What triggers aux heat use?
  - Outdoor temperature (combats cold blow)
  - Not reaching temp quickly enough, large difference between indoor and set temp (boost capacity)
  - Compressor failure (emergency heat, not aux)
  - Compressor lockout temp available in some thermostats, will trigger emergency heat use
  - Incorrect wiring? How often do we think this is a problem? Self-install is common for E* thermostats.
- Note that some heat pumps have the ability to use just SOME of the strip heat – requires proprietary control to be in the thermostat's control
- Does the thermostat KNOW if a second stage is resistance heat? At least if it's wired correctly?
- Hypothetically, are vendors roughly able to estimate the output of resistance heat vs. full capacity of the heat pump (e.g. heating capacity at 47F)? Could using this ratio improve the RHU metric?
  - Yes, by run time, assuming that the total heat load of the house is linear with delta-T.

RHU Metric Selection

- Current Metric proposal uses the upper 95th confidence limit of the mean ≤ 0.2
- However, RHU is highly non-normal so the 95th confidence limit may not be a good indicator of performance.
- Better to use a percentile such as 80th or 90th?



Discussion: RHU2 and using a percentile

- Percentiles much less influenced by extreme values, which is an advantage here
- Mean has the advantage that it doesn't throw out information in the 0th to 80th percentile
  - More concerned about outliers on the high side – keep in mind RHU2 already trims the installations with the top 5% of WHU before calculating the mean.
- Not a lot of opinions – feel free to reach out separately

Savings metric improvements

- Now turning away from RHU and back to HS and CS, the heating and cooling savings metrics, representing estimated run time reduction
- Largely agreed on September call (75%) that ENERGY STAR metric, with adjustments, was the best available estimate of smart thermostat savings
- No agreement on what the right improvements are. Ideas? (move to discussion slide)

- Ideas that have come up before:
  - Classify homes based on data (vacation homes), treat differently
  - Subtract baseline metric score based on other data
  - What can we do with the hourly input data we've discussed? (c.f. NEEA project)

Discussion: Savings metric improvements

- SKIPPED TOPIC – ran out of time

Collaboration with NEEA (Apex Analytics)

- Apex Analytics working on NEEA project
- Comparison of ENERGY STAR metric scores to savings derived from billing analysis as per traditional metrics
- Planned part of project is to modify ENERGY STAR metric calculation software and to test it
- EPA and NEEA agree in principle on coordination of effort

Software: 2 Stage systems implementation

- Seeking to include installations wired to control 2-stage heating and maybe cooling.
- Furnace/boiler: efficiency independent of capacity, no issue. Compressor based heat/cool: difference in efficiency will cause systematic nonlinearity. Include these?
- Vendor feedback from previous meetings: too few samples to bother?
  - Low single digit installations wired for control of 2 stage heating (DIY installations)
  - Another vendor: 15% for model that is contractor installed, 7% for DIY
- Proposed implementation: change input file format to have extra columns
  - Heating run time -> equivalent full load run time (ERT)
  - Add columns for stage 1 and stage 2 heating run time
  - If ERT column filled out, use that (assume calculated by vendor using actual relative capacity of stages). Document expectation of how it's calculated?
  - If no ERT data, calculate from stage 1 and stage 2 run times using 0.65 relative capacity
  - Also add installation wiring types for 2-stage furnace/boiler heating

Discussion: 2 stage implementation

- Vendor experimented with various ways of doing this, and found that it didn't make much difference to the scores. The choices of relative capacity affect the slope, but since that drops out of the performance metric, it just doesn't make much difference. Argument for just not worrying about it much. Difference between comfort temp and average temp drives savings, with a consideration of climate.
- In cooling analysis had a slightly bigger impact, but still very small (couple of tenths of % difference on score)

Software: Version 2.0 implementation

- Reminder: we talked about changing input format to be all hourly data, no daily, and this was generally agreed to be a good plan
- Some of the changes we've discussed today also have input file implications
- Hourly data will change calculations somewhat (see next discussion point)
- While we want to test as many V2.0 spec changes as possible with resubmission data, this may be too much to fit within the version 1.x software thread
- As a starting place for discussion EPA proposes:
  - Start Software version 2.0 branch now; not to be used for certification or resubmission data. Use it only for testing
  - Summer 2020 resubmission to include results from both version 1.7.0 and 2.0?
  - Heat pump oversample run, as proposed for V2 Method, used to calculate RHU
  - Development efforts kick off in Q1 2020

Software: Calculating temperature based on hourly timeline

- What do we take in hourly vs daily?
  - Hourly data : Average indoor and outdoor temperature, Auxiliary heat run time, Emergency heat run time, Average conditioned space temperature, Average heating/ cooling set point temperature.
  - Daily data : Heating/ Cooling equipment run time
- Effect on calculation of final metric?
- How would it look different if we used hourly data?
  - ER use is compared to run time in a day
  - Possible to track recovery?
  - Hourly baseline?

Discussion: V2.0 software implementation plan, and hourly data/calcs

- Ran out of time for discussion

Wrap up and Next Steps

- Action Items:

- Next Steps: