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 95\textsuperscript{th} confidence limit of the mean $\leq 0.2$
  - 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

<table>
<thead>
<tr>
<th>Fruit</th>
<th>All (n)</th>
<th>Mixed Humid (ub)</th>
<th>Very Cold (ub)</th>
<th>All (n)</th>
<th>Mixed Humid (ub)</th>
<th>Very Cold (ub)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mango</td>
<td>160</td>
<td>0.086</td>
<td>31</td>
<td>188</td>
<td>0.1</td>
<td>82</td>
</tr>
<tr>
<td>Papaya</td>
<td>48</td>
<td>0.202</td>
<td>2</td>
<td>53</td>
<td>0.164</td>
<td>0</td>
</tr>
<tr>
<td>Pear</td>
<td>63</td>
<td>0.089</td>
<td>2</td>
<td>72</td>
<td>0.078</td>
<td>37</td>
</tr>
<tr>
<td>Pineapple</td>
<td>78</td>
<td>0.187</td>
<td>14</td>
<td>91</td>
<td>0.145</td>
<td>58</td>
</tr>
<tr>
<td>Plum</td>
<td>5</td>
<td>0.799</td>
<td>0</td>
<td>5</td>
<td>0.846</td>
<td>0</td>
</tr>
<tr>
<td>Apple</td>
<td>58</td>
<td>0.138</td>
<td>3</td>
<td>71</td>
<td>0.115</td>
<td>26</td>
</tr>
<tr>
<td>Grape</td>
<td>7</td>
<td>0.119</td>
<td>0</td>
<td>11</td>
<td>0.061</td>
<td>0</td>
</tr>
<tr>
<td>Lemon</td>
<td>25</td>
<td>0.505</td>
<td>1</td>
<td>32</td>
<td>0.485</td>
<td>13</td>
</tr>
<tr>
<td>Lime</td>
<td>100</td>
<td>0.160</td>
<td>8</td>
<td>107</td>
<td>0.147</td>
<td>56</td>
</tr>
</tbody>
</table>

- 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**

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

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

![Temperature range vs Number of Thermostats (Normalized) for All regions](image-url)
### Percentage wise distribution of samples in All regions

<table>
<thead>
<tr>
<th></th>
<th>30 35F</th>
<th>35 40F</th>
<th>40 45F</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>n</strong></td>
<td><strong>%</strong></td>
<td><strong>n</strong></td>
<td><strong>%</strong></td>
</tr>
<tr>
<td>Mango</td>
<td>128</td>
<td>27.00</td>
<td>158</td>
</tr>
<tr>
<td>Papaya</td>
<td>8</td>
<td>7.48</td>
<td>46</td>
</tr>
<tr>
<td>Pear</td>
<td>46</td>
<td>26.29</td>
<td>57</td>
</tr>
<tr>
<td>Pineapple</td>
<td>62</td>
<td>26.84</td>
<td>78</td>
</tr>
<tr>
<td>Apple</td>
<td>44</td>
<td>25.43</td>
<td>58</td>
</tr>
<tr>
<td>Lemon</td>
<td>16</td>
<td>22.22</td>
<td>24</td>
</tr>
<tr>
<td>Lime</td>
<td>79</td>
<td>27.72</td>
<td>99</td>
</tr>
</tbody>
</table>

% distribution no. of installations: All Regions
**RHU2: Wider temperature bin analysis – Mixed Humid region**

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

- 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 0\text{th} to 80\text{th} 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: