Algorithmic Framework for a Hybrid Temperature/Run Time RCCS Field Savings Metric

-------------------------------------Draft 1/30/2015 --------FOR DISCUSSION ONLY-----------------------------------

Background: The U.S. Environmental Protection Agency (EPA) is pursuing development of an ENERGY STAR Residential Climate Control Systems (RCCS) program that will recognize RCCS based on service provider submission of summary field saving data that represents average savings for the RCCS. For a broad outline of the program as currently envisioned, see the June 9, 2014 ENERGY STAR Climate Controls Update Memo. To follow the progress of the program, visit the specification development page here.

In order to make such a program possible, EPA is working with stakeholders to develop metric(s) and methodology that standardizes the calculation of RCCS field savings. This document is one of several drafts in that process. Looking at earlier drafts in this series may make it easier to understand this one.

Note: EPA does not intend for this working document to be used, at this time, to assess RCCS savings claims.

During the stakeholder call regarding metric calculation, there were two questions that arose, which stakeholders agreed were best addressed by trying out various algorithms with actual data.

1. Is it better to correlate the home's HVAC run time (RT) directly with $\Delta T = T_{indoor} - T_{outdoor}$, or with heating degree days and cooling degree days (HDD and CDD) based on ΔT? (This is not the textbook definition of HDD and CDD, but a similar concept based on the difference between indoor and outdoor temperatures for the home.) It was agreed that the method of fitting the data also depends on this: if using HDD and CDD, use a ratio estimation algorithm; if using ΔT, use a linear fit. This question applies to sections 1 through 4 of the algorithm included in the prior draft framework (1/13/2015).
2. For the case where HDD and CDD are used, is it better to derive them from summing readings from shorter periods, or from the average temperatures for the day?

These questions result in several possible algorithms; a version of each is below:

1. Calculate a linear fit of daily total run time vs. daily average ΔT. Somehow decide which data (from the "bottom" of the U) to exclude from the fit.
2. Use ratio estimation to find the relationship between HDD/CDD and run times. Two ways to calculate HDD and CDD: based on ΔT and calculated using the average indoor and outdoor temperatures for the day or based on summing heating and cooling degree hours from hourly ΔT.

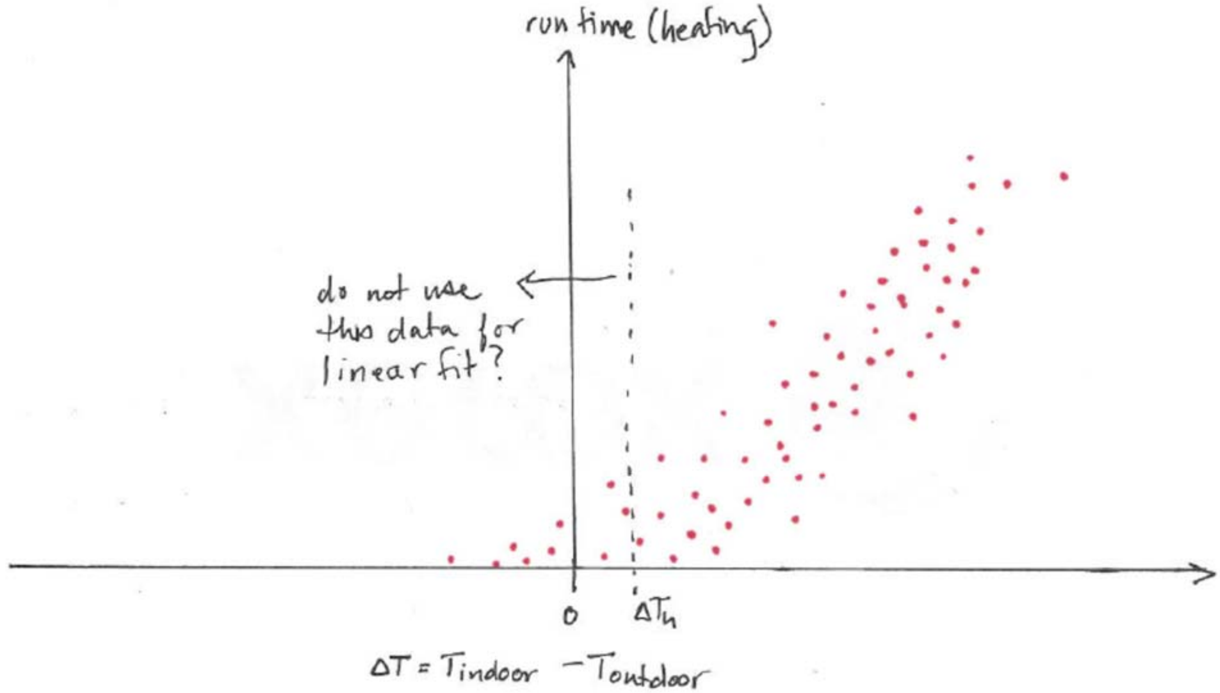
There was also some discussion of how to choose a baseline, but there was clearly not enough time to do all the topics justice. It was agreed that the stakeholder call on 1/30/2015 would concentrate on baselines, and that further discussion of the questions detailed here would take place on 2/13/2015. Stakeholders agreed this time would give them a chance to apply the questions under consideration to testing with representative data. All three versions use our originally proposed baselines of the 90^{th} and 10^{th} percentiles of set temperature. Please interpret them with the idea that this is a placeholder pending further discussion. Several stakeholders have commented on their hesitation about this baseline, and we look forward to robust discussion on the 1/30/2015 call.

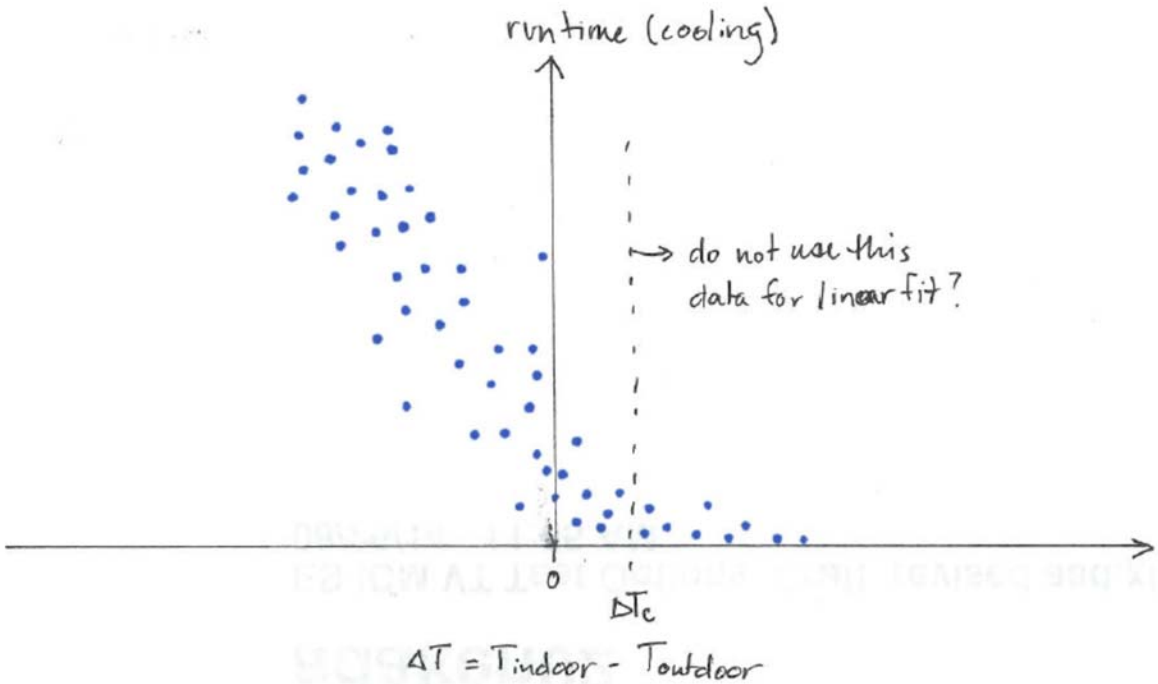
In addition, we note that throughout this algorithm, there are calculations of run time and various temperatures, begging the question of how precisely each need to be calculated. EPA encourages stakeholder feedback on this point, but tentatively recommends that final results be rounded to the nearest minute of run time and the nearest 0.5°F.

VERSION 1: Linear fit of run time vs ∆T

1. Setup
    a. Calculate average outdoor temperature for each day in the year of data, $T_{outdoor}$
    b. Calculate average ∆T = average T_{indoor} – average $T_{outdoor}$ for each day of the year.
    c. Choose a first guess for which data not to include, with low run times, under the assumption that it lies on a nonlinear part of the curve.
    d. Calculate heating system run time for each day in the year of data.
    e. Calculate cooling system run time for each day in the year of data.

Below are plots showing representative data of run time versus ∆T. Note that the data shows a strong linear trend down to low values of run time; once near zero, the tail is rather long. It was suggested that not all data be included in the linear fit of this data, but we do not recall hearing a suggestion of how to choose the point which data is considered useful. The plots show the possibility of using a ∆T cut off; in the algorithm, we propose a run time cut off of 30 minutes per day. Stakeholder feedback on this strategy would be welcome.





2. Calculate daily average ΔT = average daily T_{indoor} – average daily $T_{outdoor}$.

3. To correlate heating system run time with ΔT for that home, perform linear regression of daily average ΔT vs daily total heating system run time. Exclude all data point with run time below 30 minutes per day. The equation for the linear regression is:
   Heating run time = $\alpha_h \bullet \Delta T + \beta_h$

4. To correlate cooling system run time with ΔT for that home, perform linear regression of daily average ΔT vs daily total cooling system run time, excluding points with less than 30 minutes of cooling run time per day. The equation for the linear regression is:
   Cooling run time = $\alpha_c \bullet \Delta T + \beta_c$

5. Algorithm to calculate % heating run time reduction (HS), due to spending more time in higher ΔT parts of the curve in schedules that use more heat:
   a. Find baseline average daily indoor temperature for heating $T_{base\ heat}$ = 90th percentile of user setpoint history for the heating season for the home.
   b. Calculate the baseline average $\Delta T_{base\ heat}$ = daily average $T_{base\ heat}$ – daily average $T_{outdoor}$ for each day.
   c. Calculate the baseline run time $RT_{base\ heat} = \alpha_h \bullet \Delta T_{base\ heat} + \beta_h$ for each day and sum over the year, discarding any negative values of run time per day.
   d. Calculate % reduction in heating run time over the full year:
   $HS = (RT_{base\ heat} - RT_{actual})/RT_{base\ heat}$

6. Algorithm to calculate annual cooling savings is exactly parallel, except that $T_{base\ cool}$ = 10th percentile of indoor temperatures for that home for the entire year.

7. For heat pump systems only, calculate compressor utilization (CU)
   a. Sum up all the time that backup resistance heat was on for the year (t_{res})
   b. Sum up all the time that the compressor was on in heating mode for the year (t_{comp})
   c. Compressor utilization = $t_{comp}/(t_{res} + t_{comp})$

VERSION 2: ratio estimate of run time vs HDD(CDD), with 2 options for how DD are calculated

1. Setup
   a. Choose a first guess for $\Delta T_{b\ heat}$ and for $\Delta T_{b\ cool}$, for instance 0°F for both. These are the indoor vs. outdoor temperatures at which heating/cooling are not needed for the home.
   b. Calculate average indoor and outdoor temperature for each day in the year of data,
      daily average T_{indoor} = sum of hourly indoor temperatures / 24;
      daily average $T_{outdoor}$ = sum of hourly outdoor temperatures / 24
   c. Calculate daily average ΔT for each day in the year.
      daily average ΔT = daily average T_{indoor} – daily average $T_{outdoor}$. Can be positive or negative.
   d. Calculate total heating system run time for each day in the year of data
   e. Calculate total cooling system run time for each day in the year

2. To correlate heating system run time with HDD for that home, iterate the following until the fit is good enough:
   a. Calculate HDD
      i. Option 1: daily HDD = daily average ΔT – current guess of $\Delta T_{b\ heat}$; values less than zero should be entered as 0.
      ii. Option 2: Given hourly ΔT over the day and $\Delta T_{b\ heat}$, for each time hour the heating degree hours $HDH_i = \Delta T_i - \Delta T_{b\ heat}$; count any negative results as zero. HDD = Sum of all HDH divided by 24 hours/day.
   b. Ratio estimation: In theory, on any day when HDD = 0 for that house, there should be no heat on, and for higher HDD, the relationship should be linear. Thus, heating run time = α_h • HDD. To estimate α_h, use sum of all heating run time over the year divided the sum of all HDD over the year.
   c. Mean square error σ_h: for each day, calculate actual heating run time - α_h • HDD; that is, how far from the estimated line the actual heating run time is. Square that difference, then add them all up.
   d. Is the fit good enough, meaning is σ_h small enough? If not, the next guess of $\Delta T_{b\ heat}$ should be based on the last guesses and the trend in σ_h.
3. Record the best guess value of $\Delta T_{b\ heat}$, and the corresponding σ_h.

4. To correlate cooling system run time with CDD for that home, repeat the steps above with cooling equipment data and cooling balance temperature, except that CDD = $\Delta T_{b\ cool} - \Delta T$ for whichever method of calculating degree days you are using.

You now have the following quantities characterizing the home:

For the best model of heating run time proportional to HDD	
$\Delta T_{b\ heat}$	The temperature difference below which no heat is needed
α_h	The proportionality constant between heating run time and HDD
σ_h	The mean square error for the heating model

For the best model of cooling run time proportional to CDD	
$\Delta T_{b\ cool}$	The temperature difference above which no cooling is needed
α_c	The proportionality constant between cooling run time and HDD
σ_c	The mean square error for the cooling model

5. To calculate % heating run time reduction (HS):
   a. Find baseline average daily indoor temperature for heating $T_{base\ heat}$ = 90th percentile of user setpoint history for the heating season.
   b. Calculate the baseline heating degree days HDD_{base}. A presumably higher run time in the baseline condition corresponds to spending more days on the higher HDD part of the line than in the actual case. Note that the model itself does not change. In the baseline indoor temperature condition, $\Delta T_{base\ heat}$ =. For each day,
      i. Option 1: $HDD_{base} = \Delta T_{base\ heat} - \Delta T_{b\ heat} = T_{base\ heat}$ – daily average $T_{outdoor} - \Delta T_{b\ heat}$ for each day. Discard negative values. (Note that $T_{base\ heat}$ is assumed to be constant over the day.)
      ii. Option 2: Given hourly $T_{outdoor}$ over the day, for each hour, $HDH_{base\ i} = = \Delta T_{base\ heat} - \Delta T_{b\ heat} = T_{base\ heat} - T_{outdoor} - \Delta T_{b\ heat}$; discard negative results. HDD_{base} = sum of all HDH divided by 24 hours/day.
   c. The change in run time ΔRT_h is proportional to the change in the change in HDD: $\Delta RT_h = \alpha_h \bullet$ (HDDbase – HDD). $RT_{base\ heat} = \alpha_h \bullet$ HDDbase.
   d. Calculate % reduction in heating run time over the full year:
      $HS = \Delta RT_h / RT_{base\ heat}$

6. Algorithm to calculate annual cooling savings is exactly parallel, except that $CDD = \Delta T_{b\ cool} - \Delta T$ for whichever method of calculating degree days you are using.

7. For heat pump systems only, calculate compressor utilization (CU)
   a. Sum up all the time that backup resistance heat was on for the year (t_{res})
   b. Sum up all the time that the compressor was on in heating mode for the year (t_{comp})
   c. Compressor utilization = $t_{comp}/(t_{res} + t_{comp})$