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

--------------------------------------Draft 1/12/2015 --------FOR DISCUSSION ONLY------------------------------------

The purpose of this algorithm is to calculate

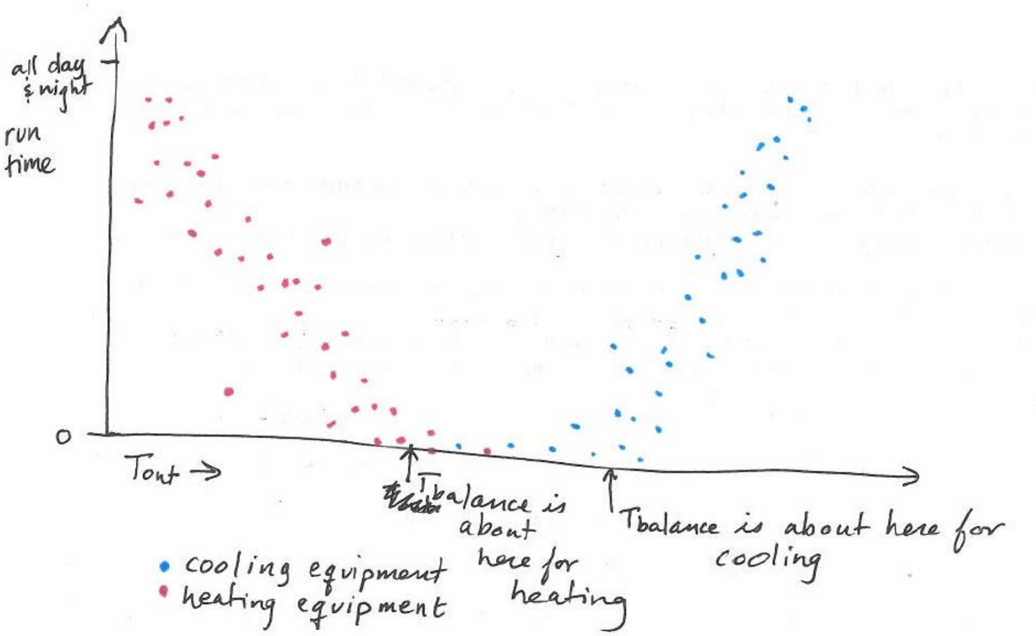
% run time reduction = $\Delta RT/RT_{baseline}$

$\Delta RT = RT_{baseline} - RT_{actual}$
RT_{actual} = sum of run time (for heating or for cooling) for the year
$RT_{baseline}$ = modeled run time in for a presumed temperature setting choice without the climate control

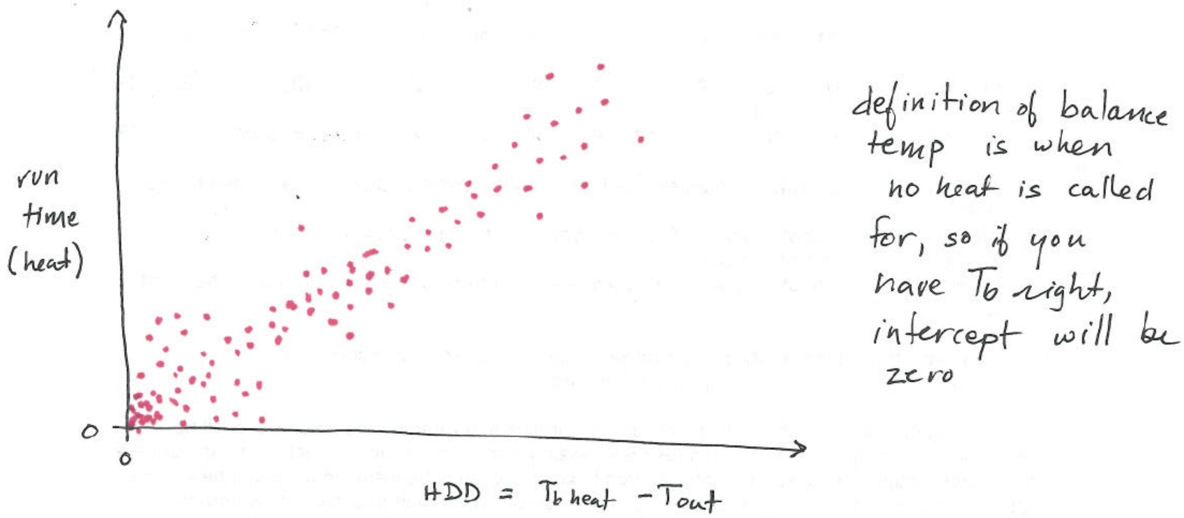
Calculate separately for heating and cooling, and only use data from single stage heating and cooling systems, including single stage heat pumps.
HS = % heating run time reduction
CS = % cooling run time reduction

The following quantities will be used and derived:
$T_{b\ heat}$ = Balance temperature for heating (for a given home), outdoor ambient temperature at which the heat does not run
$T_{b\ cool}$ = Balance temperature for cooling (for a given home), outdoor ambient temperature at which the cooling equipment does not run
α_h = linear regression coefficient for heating; heating run time/HDD
α_c = linear regression coefficient for cooling; cooling run time/CDD
T_{out} = average outside temperature for the given day

To help understand these concepts, this plot illustrates what plots of daily heating and cooling run times vs. daily average temperature might look like. There is a suggestion of a linear relationship.



Once you transform to heating degree days (HDD), which can't be less than zero, you might see something like this:



All days where no heat is called for, including all summer, are at (0,0), so you don't have to worry about defining heating season and cooling season.

For each home with a valid data set, correlate heating degree days (HDD) with heating run time:

1. Setup
    a. Choose a first guess for $T_{b\ heat}$, for instance 70 °F.
    b. Choose a first guess for $T_{b\ cool}$, for instance 70 °F.
    c. Calculate average outdoor temperature for each day in the year of data, T_{out}
    d. Calculate heating system run time for each day in the year of data

2. To correlate heating system run time with HDD for that home, iterate the following until intercept is close enough to zero and/or fit is good enough:
    a. Calculate HDD = $T_{b\ heat} - T_{out}$ for each day; values less than zero should be entered as 0.
    b. Perform linear regression: daily HDD vs daily total heating system run time. In theory, on any day when HDD = 0 for that house, there should be no heat on, so that the intercept will be zero. In fact, that will only be true if you have the right guess for the balance temperature (and if the model actual works for that home, of course). The equation for the linear regression is:

       Heating run time = $\alpha_h \bullet$ HDD + r
    c. Choose the next guess to $T_{b\ heat}$, possibly based on r and on the last guesses.
3. Record the best guess value of $T_{b\ heat}$

4. To correlate cooling system run time with CDD for that home, repeat the steps above with cooling equipment data and cooling balance temperature. Reverse all differences, e.g. CDD = $T_{out} - T_{b\ cool}$ for each day.

5. Algorithm to calculate % heating run time reduction (HS):
    a. Data inputs for each day: average daily outside temperature (T_{out}), actual average daily indoor temperature (T_{heat}), baseline average daily indoor temperature for heating ($T_{base\ heat}$).
        i. T_{heat}: arithmetic average of the indoor temperature over the day
        ii. T_{out}: arithmetic average of the outside temperature throughout the day
        iii. $T_{base\ heat}$: 90th percentile of indoor temperatures for that home for the entire year
    b. Calculate the heating degree days:
        i. HDD_{actual} for each day: current guess average daily actual temperature
        ii. $HDD_{baseline}$ for each day: Estimate the difference between $HDD_{baseline}$ and HDD_{actual} as the difference between the average actual indoor temp T_{actual} and the baseline indoor temp $T_{base\ heat}$.
    c. Calculate the run times:
        i. ΔRT = sum over all days of (α_h • 1 day •($T_{base\ heat} - T_{heat}$))
        ii. RT_{check} = sum over all days of (α_h • HDD_{actual})
        iii. RT_{actual} = sum over all time periods of heating equipment run time
    d. Quality cross-check: Compare RT_{check} to RT_{actual} which should be closely correlated; if not, there is a problem with the data and/or the method. Calculate $Q = RT_{actual} - RT_{check}$.
    e. Calculate % reduction in heating run time over the full year:
       $HS = \Delta RT/RT_{actual}$

6. Algorithm to calculate annual cooling savings is exactly parallel, though temperature differences are reversed, e.g. ΔRT = sum over all days of (α_c • ($T_{cool} - T_{base\ cool}$))

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 = $t_{comp}/(t_{res} + t_{comp})$

<u>To calculate the reported metrics:</u>

A. Across the entire population of installations that have valid data sets, for compressor utilization CU, heating savings HS, cooling savings CS and quality cross-check Q, compute
    a. Arithmetic mean
    b. Standard deviation
    c. 90th percentile
    d. 10th percentile
B. In consideration that there are a limited number of stakeholders that will be able to submit this data, EPA notes that it is acceptable for stakeholders to submit analyses of subsets of their population (e.g. by location or how long they've been enrolled, etc.) in order to mask results.
C. Also report statistics indicating the geographic spread of the installations and the percent of data sets that were deemed invalid, along with an outline of reasons sets were found to be invalid.

<u>Commentary</u>

Correlation with HDD/CDD instead of ΔT (2,3): We realize that HDD and CDD fail to capture the dynamics of set up and set back schedules, and are interested in suggested improvements. The great advantage is that it gives an excellent method for calculating the balance temperature and understanding where the "bottom" of the V is. There are various well-understood methods for iterating to find the right T_b, and we have to reason to think the choice is critical at this point. In terms of how good is "good enough", we note that we should be guided in part by the precision of the thermostats' thermistors.

Baseline temperature (5a iii): The current algorithm uses the 90th and 10th percentile indoor temperatures as constant indoor points for the baseline. 10% of the indoor temperatures detected over the year would be below the 10th percentile temperature, and that would be the comfort temperature for cooling. 90% of indoor temperatures detected over the year would be below the 90th percentile temperature, and that would be the comfort temperature for heating.
This method fails to capture variation over the course of a season, or pre-existing set up and set back behavior. A possible improvement is to use regional daily average indoor temperatures. This may provide better correlation to meter savings. However, reliable data for this is not broadly available and it would need to be developed. Thus, given that we are trying to keep this simple, we start where we are and see how good the correlation to savings is.

Baseline HDD (5b): HDD compares actual outdoor temperature to the balance temperature, which also represents an outdoor temp. The balance temperature depends on the temperature preferences of the people in the house (i.e. the set temperatures), but not in an obvious way, because it also depends on the level and type of activity in the house. We propose to use the difference between the average actual indoor temp and the baseline temperature as the HDD difference.

Heat pumps (5c iii): For heat pumps, we propose the heating run time should be the sum of compressor run time and backup heat run time, since we are also asking for a calculation of compressor utilization separately.

Capturing dynamics (5): The run time on a given day may (in some climates and seasons particularly) depend a lot on the distribution of outdoor and indoor temps throughout the day, not just the daily average. Is there a simple way to capture this? Using statistics for variation of ΔT perhaps?

Compressor utilization (8): Those with more experience can tell us if this is the most useful measure. 1 if resistance heat is never used, 0 if it is the only source of heat used.