Algorithmic Framework for a Hybrid Temperature/Run Time Connected Thermostat Field Savings Metrics

--------------------------------------Draft 7/21/2015 --------FOR DISCUSSION ONLY-----------------------------------

Background: The U.S. Environmental Protection Agency (EPA) is pursuing development of an ENERGY STAR Connected Thermostat (CT) program that will recognize CTs based on service provider submission of summary field saving data that represents average savings for the CT. For a broad outline of the program as currently envisioned, see the June 9, 2014 ENERGY STAR Climate Controls Update Memo and the June 17, 2015 Draft Connected Thermostats specification and accompanying cover memo. To follow the progress of the program, visit the specification development page.

In order to make such a program possible, EPA has been working with stakeholders to develop one or more metric(s) and methodology that standardizes the calculation of CT field savings. This document is one of several drafts in that process. This version documents the metrics used in the Alpha software modules released in June 2015.

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

During stakeholder calls 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.
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 total daily HVAC run time vs. daily average ΔT. Exclude a portion of data (from the "bottom" of the U) from the fit.
2. Use ratio estimation to find the relationship between HDD/CDD and HVAC run times. Two ways of calculating HDD and CDD are included:
    a. based on ΔT and calculated using the average daily indoor and outdoor temperatures, or
    b. based on the average of heating and cooling degree hours for the day using 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. All three versions use our originally proposed baselines of the 90^{th} and 10^{th} percentiles of set point history. Please interpret them with the idea that this is a placeholder pending further discussion.

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 second of run time and the nearest 0.5°F.

Input CT data requirements

- Zip code
- average hourly indoor temperatures
- average hourly set temperatures
- For single stage heat pumps
  - hourly run time of heat pump in heating mode
  - hourly run time of heat pump in cooling mode
  - hourly run time of electric resistance aux. heating (if applicable)
  - hourly run time of electric resistance emerg. heating (if applicable)
- For single stage non-heat-pumps
  - hourly run time of single stage non-heat pump heating equipment
  - hourly run time of single stage cooling equipment
- No other system types are supported

Setup – for Version 1 & Version 2

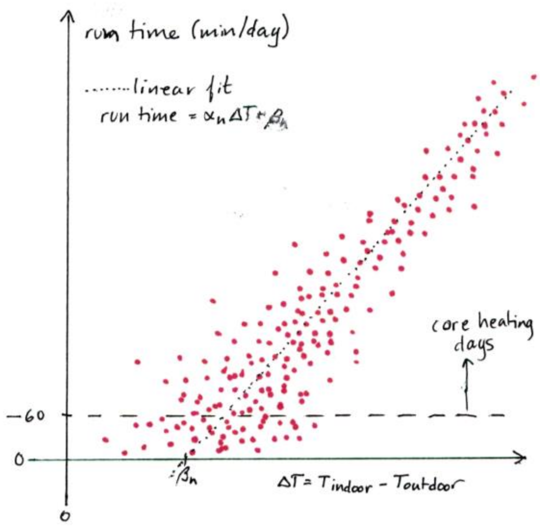
1. Parse data reported by the CT into:
   a. Core heating days – days where daily heating run time ≥ 1 h with no cooling.
   b. Core cooling days – days where daily cooling ≥ 1 h with no heating.
2. Record average daily indoor temperature, T_{indoor} = sum of hourly indoor temperatures / 24, for each:
   a. core heating day, and
   b. core cooling day
3. Using data from the closest NOAA weather station, record the hourly outdoor temperature, and the daily average $T_{outdoor}$ = sum of hourly outdoor temperatures / 24, for each:
   a. core heating day, and
   b. core cooling day
4. Record average daily ΔT = average daily T_{indoor} – average daily $T_{outdoor}$ for each:
   a. core heating day, and
   b. core day

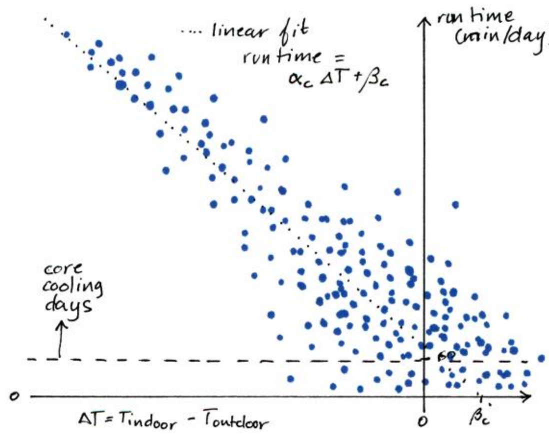
Note that ΔT values in core heating days will be positive (inside is warmer than outside) and ΔT values in core cooling days may be negative or positive, because heat generated indoors and solar gain mean that it will sometimes be necessary for the cooling system to run when inside and outside are the same temperature ($\Delta T = 0$) or it is warmer inside than outside ($\Delta T > 0$).

5. Record total daily HVAC run time, reported by the CT, for each:
   a. core heating day, and
   b. core cooling day

VERSION 1: Linear fit of run time vs ∆T

Below are plots showing daily run time versus ∆T. Note that while 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; this version of the metrics only include core heating and cooling days using a run time cut off of 1 hour per day.





1. For the core heating and cooling days, model the home's linear relationship between HVAC run time and ∆T

   a. For core heating days, record heating run time vs ∆T for that home, by performing linear regression of daily average ∆T vs total daily heating run time. The equation for the linear regression is:

      Heating run time = $\alpha_h \bullet \Delta T + \beta_h$, where

      β_h is the ∆T associated with zero run time (expected to be positive), and

      α_h is the slope of the line (expected to be positive).

   b. For core cooling days, record cooling system run time vs ∆T for that home, by performing linear regression of daily average ∆T vs total daily cooling run time. The equation for the linear regression is:

      Cooling run time = $\alpha_c \bullet \Delta T + \beta_c$, where

      β_c is the ∆T associated with zero run time (expected to be positive), and

      α_c is the slope of the line (expected to be negative).

2. Calculate the core heating baseline run time (what run time would have been had the home been kept at preferred comfort temperatures 24/7)

   a. Using data reported by the CT for core heating days, record the preferred comfort set point for heating:

      $T_{base\ heat}$ = 90th percentile of user set point (target temperature) history for core heating days.

   b. Record the baseline average $\Delta T_{base\ heat} = T_{base\ heat}$ – daily average $T_{outdoor}$ for each core heating day.

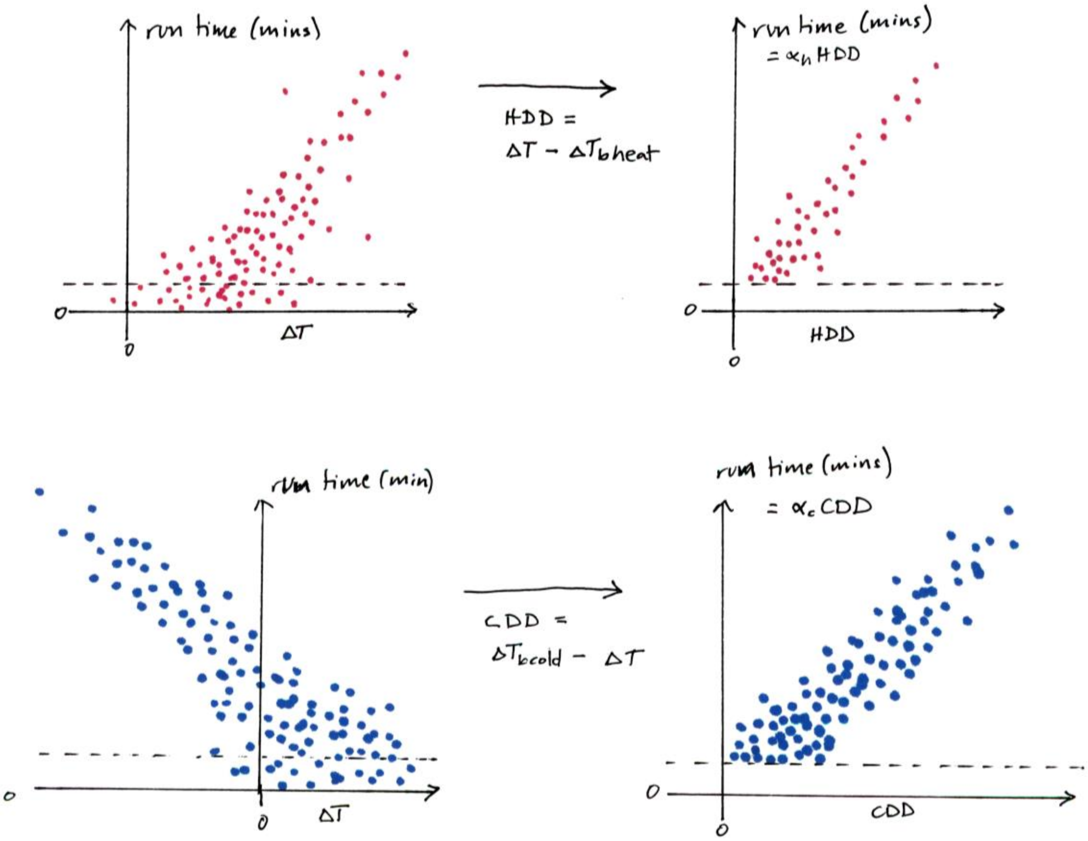
c. Record the baseline run time $RT_{heat\ base} = \alpha_h \bullet \Delta T_{base\ heat} + \beta_h$ for each core heating day and sum, discarding any negative values of run time per day.

3. Calculate % reduction in heating run time for core heating:
   $HS = (RT_{base\ heat} - RT_{actual}) / RT_{base\ heat}$

4. Similarly calculate % reduction in cooling run time for the core cooling, except that $T_{base\ cool}$ = 10^{th} percentile of user set point history for core cooling days.

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

Considering the plots shown in version 1, it is notable that there is a "balance ΔT" at which no heating or cooling is needed – perhaps a range of them, depending on the preferences of residents in the home and the equipment present. This method correlates run time required with deviation from these balance temperatures, called $\Delta T_{b\ heat}$ and $\Delta T_{b\ cool}$. Note that the balance temperatures are defined as those at which no heating (or cooling) is needed, so that the run time should be exactly zero when there is no deviation from the balance temperature. This is the same principle used to calculate heating degree days and cooling degree days, though those are calculated using outdoor temperatures rather than ΔT. We are using ΔT here in order for the algorithm to be able to reflect the indoor temperature preferences of the particular household. Savings come from having a different ΔT, and therefore HDD (CDD) relative to the base case.



To develop the data of HDD (CDD) vs run time, a value for $\Delta T_{b\ heat}$ ($\Delta T_{b\ cool}$) must be established. This is accomplished by making successive guesses for these values, then finding the slope and calculating the mean square error. When the mean square error is minimized, the best value of $\Delta T_{b\ heat}$ ($\Delta T_{b\ cool}$) is recorded.

Because we have defined HDD (CDD) such that a zero value equates to 0 run time, we know for the fit that the intercept will go through zero – in other words, run time will be directly proportional to HDD (CDD) and the "fit" we need to find is the proportionality constant. This allows us to use a ratio estimation method rather than a linear fit. The estimation of the ratio is simply the sum of all run times divided by the sum of all HDDs (CDDs) for core heating (cooling) days. Using this method has the advantage that all points have equal influence on the derived slope; in a linear fit, those nearest the intercept have outsize influence. This is especially problematic because there will be more noise in the lower run time data points. The goodness of fit for the final values is simply the mean square error, just as for the linear fit of method 1.

Because negative values of HDD (CDD) are discarded due to the meaning of the measure, it may make a difference whether hourly temperatures are averaged and daily HDD (CDD) is calculated, or whether HDD (CDD) is calculated hourly and then averaged.

1. To correlate core heating run time with HDD for that home, iterate the following until the fit is good enough:
    a. Make a first guess for $\Delta T_{b\ heat}$ and for $\Delta T_{b\ cool}$, for instance 0°F for both. These are the indoor minus outdoor temperatures at which heating/cooling are not needed for the home. Both are expected to be positive.
    b. Calculate HDD
        i. Option 1: Heating Degree Day method
           Daily HDD = daily average ΔT – current guess of $\Delta T_{b\ heat}$; values less than zero should be entered as 0.
        ii. Option 2: Heating Degree Hour method
           For each hour of a given day, calculate the Heating Degree Hour, count any negative results as zero. $HDH_i = \Delta T_i - \Delta T_{b\ heat}$, i = 1, 2, 3,…,24.
           Record HDD = Σ $HDH_i/24$
    c. 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 = $\alpha_h \bullet$ HDD, reflecting direct proportionality. To estimate α_h, use sum of all heating run time, divided the sum of all HDD over the core heating days. α_h is expected to be positive.
    d. Mean square error σ_h: for each day, calculate actual heating run time - $\alpha_h \bullet$ HDD; that is, how far from the estimated line the actual heating run time is. Square that difference, then add them all up.
    e. 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. The code implements this using a least squares optimization from SciPy (scipy.optimize.leastsq), a commonly used open-source library for scientific computation in Python.
2. Record the final value of $\Delta T_{b\ heat}$, α_h, and the corresponding σ_h.

3. To correlate core 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. Note that in this case α_c is expected to be positive, as the CDD values increase for higher cooling demand.

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

4. Calculate % reduction in heating run time for core heating days:

   a. Using data reported by the CT, record the preferred comfort set point for heating: $T_{base\ heat}$ = 90th percentile of user set point (target temperature) history for core heating days.

   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.

      i. Option 1: Daily average – Record HDD_{base} for each core heating day: $HDD_{base} = \Delta T_{base\ heat} - \Delta T_{b\ heat} = T_{base\ heat}$ – daily average $T_{outdoor} - \Delta T_{b\ heat}$. Discard negative values. (Note that $T_{base\ heat}$ is assumed to be constant over the day.)

      ii. Option 2: Hourly average – For each hour of a core heating day, calculate the baseline Heating Degree Hour, $HDH_{base\ i} = \Delta T_{base\ heat} - \Delta T_{b\ heat} = T_{base\ heat} - T_{outdoor} - \Delta T_{b\ heat}$,
      i = 1-24; discard negative results.
      Record HDD_{base} for each core heating day:
      $HDD_{base} = \Sigma\ HDH_{base\ i} / 24$.

   c. The change in run time ΔRT_h is proportional to the change in the change in HDD: $\Delta RT_h = \alpha_h \bullet (HDD_{base} - HDD)$, expected to be positive, as the baseline condition is expected to require more heating.
   $RT_{base\ heat} = \alpha_h \bullet HDD_{base}$

   d. Record % reduction in heating run time for core heating days:
   $HS = \Delta RT_h / RT_{base\ heat}$

5. Similarly calculate and record % reduction in cooling run time for core cooling days, using CDD = $\Delta T_{b\ cool} - \Delta T_{base\ cool}$ for whichever method of calculating degree days you are using.

Heat Pump Systems Only: Resistance Heating Utilization

For heat pump systems only, calculate Resistance heat Utilization (RU):

1. Using data reported by the CT, calculate the following in twelve outdoor temperature bins ($0 \le T < 5°F$, $5 \le T < 10°F$,…, $55 \le T \le 60°F$):
    a. t_{emerg} = total annual emergency resistance heating run time
    b. t_{aux} = total annual auxiliary resistance heating run time
    c. t_{comp} = total compressor heating run time

    Record RU for each bin:

    $RU = (t_{emerg} + t_{aux})/(t_{emerg} + t_{comp})$