ENERGY STAR® Connected Thermostat Products

Final Draft Method to Demonstrate Field Savings
Rev. Nov-2016

1) OVERVIEW

This method shall be used to demonstrate field savings for ENERGY STAR Connected Thermostat (CT) Products.

Note: This Final Draft Method is based on the Draft 2, as modified in response to stakeholder comments and based on continuing stakeholder discussions.

2) APPLICABILITY

This ENERGY STAR Method is applicable to Connected Thermostat Products as defined in the ENERGY STAR Eligibility Criteria for Connected Thermostat Products.

3) DEFINITIONS

Unless otherwise specified, all terms used in this document are consistent with the definitions contained in the ENERGY STAR Eligibility Criteria for Connected Thermostat Products.

4) DEMONSTRATING FIELD SAVINGS

Field savings of CT Products shall be assessed for product certification as well as for periodic reporting as detailed in the Partner Commitments section of the ENERGY STAR Program Requirements for Connected Thermostat Products.

1. Install and configure the ***most recent release*** of the ENERGY STAR CT Field Savings Software using the open source code and instructions available at:
   - Documentation: http://thermostat.readthedocs.org/en/latest/
   - Source code: https://github.com/impactlab/thermostat
2. From the set of all instances of a fielded CT Product Family in the US, generate a randomly selected data set via the following procedure:
   a. Generate a metadata file that includes all instances of a fielded CT Product in the US, with the following information:
      i. included data: “thermostat_id”, “equipment_type”, “zipcode”, “date1”, “date2”;
      ii. thermostat_id is a unique alphanumeric string;

      Important Note: thermostat_id shall be a preexisting unique identifier associated with the CT Device (e.g. MAC address or serial number) or directly mapped from it using consistent rules (e.g. to remove non alphanumeric characters or to eliminate duplicates).

      iii. equipment_type:
         0: Other – e.g. multi-zone multi-stage, modulating. Note: module will not output savings data for this type
         1: Single stage heat pump with electric resistance aux and/or emergency heat (i.e. strip heat)
         2: Single stage heat pump without additional and/or supplemental heating sources
         3: Single stage non heat pump with single-stage central air conditioning
         4: Single stage non heat pump without central air conditioning
         5: Single stage central air conditioning without central heating;

iv. zipcode is the US ZIP code where the thermostat is installed;
v. date1 is the first date for which interval data was reported; and
vi. date2 is the last date for which interval data was reported.

b. from the above metadata file, generate a new metadata file that eliminates CTs according to the following rules:
   i. thermostat_id is null or invalid;
   ii. equipment_type = 0, or changes during the reporting period;
   iii. zipcode is null, invalid, or changes during the reporting period;
   iv. date1 is after the reporting period start date; and
   v. date2 is before the reporting period stop date.

c. from the above metadata file, split CTs into 5 climate zone metadata files, using the Energy Information Administration (EIA) Building America Climate Zone to ZIP code Database. These climate zones are:

   - Cold/Very Cold
   - Hot Dry/Mixed Dry
   - Hot Humid
   - Mixed Humid
   - Marine

d. using Python Natsort, sort each climate zone metadata file by the unique thermostat_id assigned to each thermostat.*

e. using the Python Numpy random number generator, set a seed (supplied by EPA) for each climate zone manually with numpy.random.seed (number) and record these numbers.[1]

   Important Note: Seeds must be set manually and documented by Partner in order for sample to be reproducible and/or auditable. Additionally, using numpy.random.seed() without specifying a seed (empty parenthesis) is unable to be reconstructed, thus *should not* be used when submitting data.

f. for each of the five EIA climate zones, generate metadata files with 250 CTs using the Numpy function numpy.random.choice, using replace=False to prevent sampling duplicates. If the Partner has fewer than 250 thermostats in an EIA climate zone from which to sample to sample, include all thermostats for that climate zone.[1]

g. using the above metadata files, generate files that follow the requisite file format and content requirements in http://thermostat.readthedocs.io/en/latest/tutorial.html#input-data.[1]

   - a single metadata file that includes all CTs from the above step; and
   - a separate interval data file for each included unique thermostat_id.

Note: EPA received extensive feedback that the flexibility for the manufacturer to define the thermostat IDs and select sample size without limit presents opportunities for gaming and makes the sample selection not reproducible. In addition, a required sample size only on the output may require running the software several times, given the likelihood that some CTs will be filtered out. EPA agrees with this and proposes a specific sample size on the input to the EPA tools. EPA has also added language requiring the thermostat ID be associated with the CT Device.

EPA is concerned that requiring specific versions of Numpy and Natsort may create dependency conflicts. Rather than specify versions, EPA is investigating how the versions used might be captured and retained in case review is needed later. Intent in specifying versions was to guard against changes that would impact sample selection, possibly without EPA's or partners' knowledge. We note it is unlikely this situation would arise, as these are mature utility functions in broad use.

[1] see Sample Code in Appendix A

3. Verify that this data set:
    a. consists of one metadata file and *n* interval data files (one interval data file for each CT in the sample);
    b. includes CT Products in each of the five EIA climate zones: Very Cold/Cold, Hot Dry/Mixed Dry, Hot Humid, Mixed Humid, and Marine;
    c. adheres to the relevant reporting period criteria as detailed in the Partner Commitments section of the ENERGY STAR Program Requirements for Connected Thermostat Products; and
    d. includes only CT Products that control one of the following types of HVAC equipment types 1 thru 5:

Note: EPA has added language that ensures CTs that are not in service for the applicable reporting period are removed prior to random sample selection. Reporting period criteria is included in the ENERGY STAR Connected Thermostat specification. EPA has clarified here and elsewhere that heat pumps with electric resistance auxiliary and/or emergency heat are included, but dual-fuel systems, including both a heat pump and a central oil- or gas-fired furnace, are not. EPA is aware that some CT vendors are not able to distinguish these systems by the way the CT is wired, and states this expectation knowing CT providers will meet it to the best of their ability.

EPA has also clarified the equipment types list to meet the expectation of the software tools that all CTs included for analysis have single stage cooling or no cooling, and/or single stage heating, or no heating.

4. Process the data set using the ENERGY STAR CT Field Savings Software. The output includes:
    - Mean cooling and heating savings scores (% run time reduction) with associated standard error of the mean, and decile bins, in each of five EIA climate zones;
    - The lower 95% confidence interval of the mean cooling and heating savings scores (% run time reduction) with decile bins in each of the five EIA climate zones;
    - Weighted national average cooling and heating savings scores (% run time reduction) with associated standard error of the mean;
    - Weighted national average of the lower 95% confidence interval of the mean cooling and heating savings scores (% run time reduction);
    - Mean resistance heat utilization for heat pumps with backup electric resistance heating, in 5°F outdoor temperature bins from 60°F to 0°F across all climate zones; and
    - Estimate of statistical power in each climate zone and for national savings, in order to guide partners toward a data set that provides sufficient certainty.
5. Submit the CT Field Savings Software output file
    a. to the Certification Body for initial product certification, or
    b. to EPA for ongoing reporting.

Note: This Final Draft specifies a fixed sample size of 250 CTs for each of the five EIA climate zones, but allows CT service providers that have fewer CTs in one or more climate zones to evaluate savings using all CTs (<250) in one or more climate zones.

The outputs from the ENERGY STAR CT Field Savings Software in this Final Draft reflect the outputs expected in the Version 1.0 release of the software, aligning with the statistics required by the specification.

Some commenters suggested setting up a system whereby all outputs of running the EPA software are automatically forwarded to EPA at the same time they are sent to the CT service provider. EPA understands this would require that the software run on a secure server to which providers upload data. EPA is interested in this idea, but is unable to guarantee data privacy at this time. In addition, there is no evidence of the problems it seeks to solve. We look forward to further conversations with stakeholders about such a system as a future improvement.

6. Retain the following data for a period not less than five years from the associated file submission:
    a. raw data files for each thermostat_id in the CT data set used to assess savings;
    b. all metadata files created in step 4.2. all random seeds with a reference to the metadata files; for which each seed was used
    c. CT data set files, generated in step 4.2.g., and used with the ENERGY STAR CT Field Savings Software to assess reported savings; and
    d. All output files submitted to EPA or to certification bodies.

APPENDIX A: CODE FOR SORTING AND REPRODUCABLE RANDOM SAMPLING OF THERMOSTATS

```python
import numpy as np
import pickle
#Package Notes, Natsort 5.0.1
from natsort import natsorted

#DEBUG: state caching
prng_state_exact = np.random.get_state()

#DEBUG: Save state via pickle
with open('prng_state.pickle','wb') as f:


# #DEBUG: Code to load old state and set PRNG to that state
# with open('prng_state.pickle','rb') as f:
#     reload_state = pickle.load(f)
# np.random.set_state(reload_state)

#Load sample target data; Note if 0:n index replaced with data/thermostat id's,
will sample unique id's instead of indicies
EIAColdVCold = np.arange(500)
EIAHotHumid = np.arange(500)
EIAMixedHumid = np.arange(500)
EIAHDMD = np.arange(500)
EIAMarine = np.arange(500)

#Natural Sort Imported data by value
SampEIAColdVCold = natsorted(SampEIAColdVCold)
SampEIAHotHumid = natsorted(SampEIAHotHumid)
SampEIAMixedHumid = natsorted(SampEIAMixedHumid)
SampEIAHDMD = natsorted(SampEIAHDMD)
SampEIAMarine = natsorted(SampEIAMarine)

#Sample target data, applying 1 seed per climate zone
np.random.seed(101)
SampEIAColdVCold = np.random.choice(EIAColdVCold,250, replace=False )
np.random.seed(102)
SampEIAHotHumid = np.random.choice(EIAHotHumid,250, replace=False )
np.random.seed(103)
SampEIAMixedHumid = np.random.choice(EIAMixedHumid,250, replace=False )
np.random.seed(104)
SampEIAHDMD = np.random.choice(EIAHDMD,250, replace=False )
np.random.seed(105)
SampEIAMarine = np.random.choice(EIAMarine,250, replace=False )

#Sort Sampled data by value
SampEIAColdVCold = np.sort(SampEIAColdVCold)
SampEIAHotHumid = np.sort(SampEIAHotHumid)
SampEIAMixedHumid = np.sort(SampEIAMixedHumid)
SampEIAHDMD = np.sort(SampEIAHDMD)
SampEIAMarine = np.sort(SampEIAMarine)

#Create matrix for all samples, Matrix format best for indicies
SortedEIASample =
np.vstack((SampEIAColdVCold,SampEIAHotHumid,SampEIAMixedHumid,SampEIAHDMD,SampEIAM
arine))

#Create long format output, best for vector of thermostat id's
```

```
results = SampEIAColdVCold
outfile =
np.append(results,[SampEIAHotHumid,SampEIAMixedHumid,SampEIAHDMD,SampEIAMarine])

#Save Sample items to file
np.savetxt('PRNG.csv',SortedEIASample, delimiter=",")
```

APPENDIX B: DESCRIPTION OF SAVINGS METHODOLOGY & ALGORITHMS USED IN THE ENERGY STAR CT FIELD SAVINGS SOFTWARE

Note: Based on the results of the data call using the Beta software release 0.4 and on further analysis and conversation with stakeholders, EPA was able to settle on a method to calculate the metric and several details of the calculations. This Appendix now reflects these decisions and describes the content and use of the ***forthcoming V1.0 release***.

The software consists of documentation plus two modules which assess CT savings expressed as run time reduction relative to the baseline.

- *Thermostat Module*

  The thermostat module separately assesses HVAC heating and cooling savings attributed to a CT.

  1. Input files consist of:
     - a metadata file that contains the unique ID, controlled HVAC type, ZIP code and interval data file name for each CT in the data set; and
     - an interval data file for each CT in the data set.
  2. The module will not assess savings for CT data files that are missing too much data, or if corresponding outdoor temperature data is not available. Indoor and outdoor temperature data gaps of up to 2 hours are interpolated. Days with gaps longer than 2 hours are excluded from the analysis. If more than 5% of days are missing HVAC run time data, the CT is excluded.
  3. The thermostat module output is a.csv file that includes a separate row for each CT in the data set. Columns in the output file include:
     a. Unique CT ID;
     b. Controlled HVAC type;
     c. ZIP code in which the thermostat is installed;
     d. Heating savings;
     e. Cooling savings;
     f. Resistance Heating Utilization; and
     g. Associated statistical data.

- *Statistics Module*

  1. A thermostat module output file comprises the input to the statistics module.
  2. The statistics module output is a .csv file that includes:
     a. Regional heating and cooling savings as average savings of multiple instances of a CT model in each of the five EIA climate zones. The output file includes the climate zone, decile bins, means, and standard errors of the mean for all numerical outputs in the individual thermostat output files generated by the thermostat module. It also includes the lower 95% confidence limit for heating and cooling savings;
     b. National heating and cooling savings assessed as a weighted average of regional savings. Weighting is by the proportion of national heating and cooling energy, respectively, used in each of the five EIA climate zones. The output file also includes the weighted average decile bins, weighted average standard error of the mean, and the lower 95% confidence limit of the weighted average; and
     c. Resistance heat utilization (%) in 5°F bins from 0°F to 60°F (applicable to heating with heat pumps only)

Source code and detailed documentation are available at the following locations:

Documentation: http://thermostat.readthedocs.org/en/latest/

Source code: https://github.com/impactlab/thermostat

OVERVIEW OF SOFTWARE USAGE

Step 1 – The thermostat module is used to assess savings for n CTs that comprise a sample set.

1. *Assess Savings for CT1:*
    a. *Develop Thermal/HVAC Model* – the software constructs a model of the relationship between heating and cooling HVAC run time, outside temperature, and temperature choices for the CT;
    b. *Determine Baseline Comfort Temperatures* – the software parses the CT indoor temperature history to determine occupants' preferred temperatures for heating and cooling;
    c. *Assess Baseline Run Times* – the software uses the thermal/HVAC model to assess baseline heating and cooling run times for the CT; i.e. what run times would have been under 24/7 use of baseline comfort temperatures;
    d. *Assess CT Savings* – the software generates an output file that includes CT savings; expressed as percent run time reduction relative to the baseline. For heat pumps, the software also outputs Resistance Heat Utilization (RHU), the proportion of total heating run time that includes auxiliary and emergency heating.
2. *Repeat for each of the remaining CTs in the sample set.* Once evaluation of the sample set is complete, the output file will include *n* rows of data, one row for each of the *n* CTs in the sample set.

Step 2 – The statistics module is used to assess aggregate national and regional savings

1. *Regional CT savings* – With a thermostat module output file serving as the input, the statistics module outputs average regional HVAC savings in the five EIA[2] climate zones, along with associated statistics.
2. *National CT savings* – the software also outputs national heating and cooling savings as a weighted average of regional savings, along with associated statistics. Regional savings are weighted by the proportion of national heating or cooling energy consumed in each of the five EIA climate zones.
3. *Resistance heating utilization (RHU)* – for heat pumps, the statistics module also outputs average regional and national RHU and associated statistics.

ASSESSMENT OF SAVINGS FOR ONE THERMOSTAT

Common Terms and Calculations

- *Core Days*
    - Core heating days – days where daily heating run time ≥ 30 minutes with no cooling, excluding days that exceed the missing data thresholds, as previously noted.
    - Core cooling days – days where daily cooling ≥ 30 minutes with no heating, excluding days that exceed the missing data thresholds, as previously noted.

[2] https://www.eia.gov/consumption/residential/reports/images/climatezone_eere-lg.jpg

Note: Only primary heating source run times are assessed to determine core heating days. That is, auxiliary and emergency heat run times associated with heat pumps are not included.

- *Baseline Assessment* – occupant's preferred comfort temperatures for heating and cooling[3]:
  - Using data reported by the CT for core heating days, determine the preferred comfort temperature for heating:

    $T\ base\ heat\ (°F)$ = *90th percentile of indoor temperature history for core heating days*. Not a Number (NaN) values are ignored.

  - Using data reported by the CT for core cooling days, determine the preferred comfort temperature for cooling:

    $T\ base\ cool\ (°F)$ = *10th percentile of indoor temperature history for core cooling days*. NaN values are ignored.

Assess Heating Savings

- Develop the CT/home's unique thermal/HVAC heating run time model (hourly HTD model)
  - o Calculate daily Heating Thermal Demand ($daily\ HTD_d$) for each core heating day in the interval data file:
    - Calculate the average hourly indoor minus outdoor temperature difference ($hourly\ \Delta T_{d.n}$) for each core heating day:
      - Using the ZIP code included in the metadata file, determine the closest NOAA weather station that is in the same climate zone as the CT and look up hourly outdoor temperatures:
        $hourly\ outdoor\ T_{d.n}\ (°F)$, where
        d is the core cooling day ($d = 001, 002, 003 \ldots x$),
        n is the hour; ($n = 01, 02, 03, \ldots 24$)
      - $hourly\ \Delta T_{d.n}\ (°F) = hourly\ indoor\ T_{d.n} - hourly\ outdoor\ T_{d.n}$, where
        d is the core heating day ($d = 001, 002, 003 \ldots x$), and
        n is the hour; ($n = 01, 02, 03, \ldots 24$).
    - Starting with an assumed value of zero for Tau ($\tau_h = 0$), calculate the daily Heating Thermal Demand ($daily\ HTD_d$), as follows:

      $daily\ HTD_d(°F) = \frac{\sum_{n=1}^{24}[hourly\ \Delta T_{d.n} - \tau_h]_+}{24}$, where
      d is the core heating day ($d = 001, 002, 003 \ldots x$),
      n is the hour; ($n = 01, 02, 03, \ldots 24$),
      τ_h is the ΔT associated with $HTD = 0$, reflecting that homes with no heat running tend to be warmer than the outdoors, and
      $[\ ]_+$ indicates that the term is zero if its value would be negative
  - o For the set of all core heating days in the CT interval data file, use ratio estimation to calculate α_h, the homes responsiveness to heating, which should be positive

    $\alpha_h(^{minutes}/_{°F}) = RT\ actual\ heat / \sum_{d=1}^{x} daily\ HTD_d$, where
    $RT\ actual\ heat$ is the sum of heating run times for all core heating days in the CT interval data file.
  - o For the set of all core heating days in the CT interval data file, optimize τ_h that results in minimization of the sum of squares of the difference between daily run times reported by

[3] Methodology described in paragraph 3.1.2 of "A DATA-DRIVEN FRAMEWORK FOR COMPARING RESIDENTIAL THERMOSTAT ENERGY PERFORMANCE," Bryan Urban and Kurt Roth, Fraunhofer USA, July 2014.

the CT, and calculated daily heating run times. Next recalculate α_h (iaw the above step) and record the model's parameters $(\tau_h\ ,\ \alpha_h)$[4]:

$\sum_{d=1}^{x}(actual\ RT\ heat_d - daily\ RT\ heat_d)^2$ is minimized, where

$daily\ RT\ heat_d = \alpha_h * daily\ HTD_d$

$actual\ RT\ heat_d$ is the total daily heating run time reported by the CT for that core heating day

d is the core heating day $(d = 001, 002, 003 \ldots x)$

Note that α_h characterizes the response of the home to heating and τ_h is the difference between inside and outside temperatures when heating run time = 0

- Calculate the cumulative baseline run time for the collection of all core heating days in the CT interval data file (i.e. what would have occurred over the set of all core heating days in the reporting period, had the home been held constant at the preferred heating comfort temperature).
  - o Calculate the baseline daily Heating Thermal Demand $(daily\ HTD\ base_d)$for each core heating day in the CT interval data file
    - ▪ Calculate the difference between the occupant's preferred comfort temperature for heating and the average outside temperature for each hour of each core heating day $(hourly\ \Delta T\ base\ heat_{d.n})$:

      $hourly\ \Delta T\ base\ heat_{d.n}\ (°F) = T\ base\ heat - hourly\ outdoor\ T_{d.n}$, where

      d is the core heating day $(d = 001, 002, 003 \ldots x)$,

      n is the hour $(n = 01, 02, 03, \ldots 24)$,

      $T\ base\ heat$ is the occupant's preferred comfort temperature for heating, and

      $hourly\ outdoor\ T_{d.n}$ is the average outdoor temperature for that hour reported by the nearest NOAA weather station in the same climate zone as the CT.
    - ▪ Calculate baseline daily Heating Thermal Demand $(daily\ HTD\ base_d)$

      $daily\ HTD\ base_d = \frac{\sum_{n=1}^{24}[hourly\ \Delta T\ base\ heat_{d.n} - \tau_h]_+}{24}$, where

      d is the core heating day $(d = 001, 002, 003 \ldots x)$,

      n is the hour $(n = 01, 02, 03, \ldots 24)$,

      τ_h, determined above, is a constant that is part of CT/home's thermal/HVAC heating run time model, and

      $[\ ]_+$ indicates that the term is zero if its value would be negative.
  - o Calculate baseline run time as the sum of daily baseline run times for the set of core heating days in the CT interval data file

    $RT\ base\ heat\ (minutes) = \sum_{d=1}^{x} \alpha_h * daily\ HTD\ base_d\ (°F)$, where

    α_h , determined above, is a constant that is part of CT/home's thermal/HVAC heating run time model

    $daily\ HTD\ base_d$ is the baseline daily Heating Thermal Demand

    d is the core heating day $(d = 001, 002, 003 \ldots x)$
- CT heating savings is the percent heating run time reduction, for the reporting period in the CT input file.

  $HS\ (\%\ RT\ reduction) = 100 * \frac{(RT\ base\ heat - RT\ actual\ heat)}{RT\ base\ heat}$, where

[4] Thermostat module calls Scipy Function leastsq, which uses Levenberg-Marquardt as implemented in MINPACK: LMDIF 1980

$RT\ base\ heat$ is the sum of baseline heating run times for all core heating days in the CT interval data file, and

$RT\ actual\ heat$ is the sum of heating run times for all core heating days in the CT interval data file.

Assess Cooling Savings

- Develop the CT/home's unique thermal/HVAC cooling run time model (hourly CTD method)
  - o Calculate daily Cooling Thermal Demand $(daily\ CTD_d)$ for each core cooling day in the interval data file:
    - ▪ Calculate the average hourly indoor minus outdoor temperature difference $(hourly\ \Delta T_{d.n})$ for each core heating day:
      - Using the ZIP code included in the metadata file, determine the closest NOAA weather station that is in the same climate zone as the CT and look up hourly outdoor temperatures:
        $hourly\ outdoor\ T_{d.n}$, where
        d is the core cooling day $(d = 001, 002, 003 \dots x)$,
        n is the hour; $(n = 01, 02, 03, \dots 24)$
      - $hourly\ \Delta T_n\ (°F) = \ hourly\ indoor\ T_{d.n} - \ hourly\ outdoor\ T_{d.n}$, where
        d is the core heating day $(d = 001, 002, 003 \dots x)$, and
        n is the hour; $(n = 01, 02, 03, \dots 24)$
    - ▪ Starting with an assumed value of zero for Tau $(\tau_c = 0)$, calculate the daily Cooling Thermal Demand $(daily\ CTD_d)$, as follows:
      $daily\ CTD_d(°F) = \frac{\sum_{n=1}^{24}[\tau_c - hourly\ \Delta T_{d.n}]_+}{24}$, where
      d is the core cooling day $(d = 001, 002, 003 \dots x)$,
      n is the hour; $(n = 01, 02, 03, \dots 24)$
      τ_c is the ΔT associated with $CTD = 0$ (zero cooling run time), and
      $[\ ]_+$ indicates that the term is zero if its value would be negative
  - o For the set of all core cooling days in the CT interval data file, use ratio estimation to calculate α_c, the homes responsiveness to cooling, which should be positive
    $\alpha_c(^{minutes}/_{°F}) = {}^{RT\ actual\ cool}/_{\sum_{d=1}^{x} daily\ CTD}$, where
    $RT\ actual\ cool$ is the sum of cooling run times for all core cooling days in the CT interval data file.
  - o For the set of all core cooling days in the CT interval data file, optimize τ_c that results in minimization of the sum of squares of the difference between daily run times reported by the CT, and calculated daily cooling run times. Next recalculate α_c (iaw the above step) and record the model's parameters $(\tau_c\ ,\ \alpha_c)$[5]:
    $\sum_{d=1}^{x}(actual\ RT\ cool_d - daily\ RT\ cool_d)^2$ is minimized, where
    $daily\ RT\ cool_d = \ \alpha_c * daily\ CTD_d$
    $actual\ RT\ cool_d$ is the total daily cooling run time reported by the CT for that core cooling day
    d is the core cooling day $(d = 001, 002, 003 \dots x)$

    Note that α_c characterizes the response of the home to cooling and τ_c is the difference between inside and outside temperatures when cooling run time = 0

---

[5] Thermostat module calls Scipy Function leastsq, which uses Levenberg-Marquardt as implemented in MINPACK: LMDIF 1980

- Calculate the cumulative baseline run time for the collection of all core cooling days in the CT interval data file (i.e. what would have occurred over the set of all core cooling days in the reporting period, had the home been held constant at the preferred cooling comfort temperature).
  - o Calculate the baseline daily Cooling Thermal Demand ($daily\ CTD\ base_d$) for each core cooling day in the CT interval data file
    - ▪ Calculate the difference between the occupant's preferred comfort temperature for cooling and the average outside temperature for each hour of each core cooling day ($hourly\ \Delta T\ base\ cool_{d.n}$):

      $hourly\ \Delta T\ base_n\ (°F) = T\ base\ cool - \ hourly\ outdoor\ T_n$, where

      d is the core heating day ($d = 001, 002, 003 \dots x$),

      n is the hour ($n = 01, 02, 03, \dots 24$),

      $T\ base\ cool$ is the occupant's preferred comfort temperature for cooling, and

      $hourly\ outdoor\ T_n$ is the average outdoor temperature for that hour reported by the nearest NOAA weather station in the same climate zone as the CT.
    - ▪ Calculate baseline daily Cooling Thermal Demand ($daily\ CTD\ base_d$)

      $daily\ CTD\ base_d = \frac{\sum_{n=1}^{24}[\tau_c - hourly\ \Delta T\ base\ cool_{d.n}]_+}{24}$, where

      d is the core cooling day ($d = 001, 002, 003 \dots x$),

      n is the hour ($n = 01, 02, 03, \dots 24$),

      τ_c , determined above, is a constant that is part of CT/home's thermal/HVAC cooling run time model, and

      $[\ ]_+$ indicates that the term is zero if its value would be negative.
  - o Calculate baseline run time as the sum of daily baseline run times for the set of core cooling days in the CT interval data file

    $RT\ base\ heat\ (minutes) = \sum_{d=1}^{x} \alpha_c * daily\ CTD\ base_d\ (°F)$, where

    α_c , determined above, is a constant that is part of CT/home's thermal/HVAC cooling run time model

    $daily\ CTD\ base_d$ is the baseline daily Cooling Thermal Demand

    d is the core cooling day ($d = 001, 002, 003 \dots x$)
  - o CT cooling savings is the percent cooling run time reduction for the reporting period in the CT input file.

    $CS(\%\ RT\ reduction) = 100 * \frac{(RT\ base\ cool - RT\ actual\ cool)}{RT\ base\ cool}$, where

    $RT\ base\ cool$ is the sum of baseline cooling run times for all core cooling days in the CT interval data file, as detailed for each method, and

    $RT\ actual\ cool$ is the sum of cooling run times for all core cooling days in the CT interval data file.

Calculate RHU

- o Resistance Heating Utilization (RHU) – for heat pump systems only, calculate RHU in 12 (daily average) outdoor temperature bins (0≤T<5°F, 5≤T<10°F,…, 55≤T≤60°F). For example $RHU_{0\text{-}5F}$ is calculated as follows:

$$RHU_{0-5F} = \frac{(t_{emerg_{0-5F}} + \ t_{aux_{0-5F}})}{(t_{emerg_{0-5F}} + \ t_{comp_{0-5F}})}$$

where,

$t_{emerg_{0-5F}}$ = total emergency resistance heating run time in the interval data file that occurs on core heating days where 0°F ≤ average daily outdoor temperature < 5°F, t_{emerg} is in lieu of compressor operation, e.g. use of electric resistance strip heat under fault conditions or when the outside temperature is very low,

$t_{aux_{0-5F}}$ = total annual auxiliary resistance heating run time in the interval data file that occurs on core heating days where 0°F ≤ average daily outdoor temperature < 5°F, t_{aux} is supplemental to compressor operation, e.g. use of electric resistance strip heat to increase heating capacity, and

$t_{comp_{0-5F}}$ = total compressor heating run time in the interval data file that occurs on core heating days where 0°F ≤ average daily outdoor temperature < 5°F. A proportion of t_{comp} may occur when auxiliary resistance heating also occurs.

ASSESSMENT OF AGGREGATE NATIONAL AND REGIONAL SAVINGS

After the thermostat module generates an output file for the sample set, the statistics module leverages these files to assess average savings, expressed as percent run time reduction.

- Assess average heating savings, cooling savings, and RU for each of the five EIA climate zones, for example

  CT Savings $_{\text{heat Mixed-Humid}}$ = ∑ CT$_i$ Savings $_{\text{heat Mixed-Humid}}$ / *n*, where

  i = 1 thru n, and

  n = the number of CTs in the EIA Mixed-Humid climate region

- Nationally – from regional savings, national savings are calculated as weighted averages for heating and cooling, see below for details on Climate Zone weightings. This group of nationally weighted savings estimates include the mean, standard error of the mean, decile bins (q10-q90) and the Lower Bound of the 95th Percentile Confidence Interval (LB95).
  The Nationally Weighted LB95 is calculated via the following formula:

$$SE_{National} = \sqrt{\sum_{CZ=1}^{5} (W_{CZ} * SE_{CZ})^2}$$

  Where SE_{CZ} are the standard errors in each climate zone, calculated as the standard deviation in the climate zone divided by the square root of the number of CTs analyzed in that climate zone, W_{CZ} are the climate zone weightings, and $SE_{National}$ is the standard error of the weighted mean.

  The Nationally Weighted LB95 is calculated from the $SE_{National}$ via the following:

$$LB95_{National} = Mean_{National} - 1.96 * SE_{National}$$

  Key performance metrics, Nationally Weighted LB95 and 20th decile bin q20 are specifically labelled in the output file for ease of reference.

- Weighting for heating and cooling is by the proportion of national heating or cooling energy used in each of the five EIA climate zones, as follows:

Heating	
Climate Region	%BTU vs National
Very Cold/Cold	54.9%
Mixed-Humid	31.2%
Mixed-Dry/Hot-Dry	5.4%
Hot-Humid	4.9%
Marine	3.6%

Cooling	
Climate Region	%BTU vs National
Very Cold/Cold	9.6%
Mixed-Humid	34.0%
Mixed-Dry/Hot-Dry	14.4%
Hot-Humid	42.0%
Marine	Not included

- All other National outputs, including National RHU, are calculated as an unweighted average over all CTs in the sample set.