# ENERGY STAR® Connected Thermostat Products

# Draft 2 Method to Demonstrate Field Savings
# Rev. Sep-2016

## 1) OVERVIEW

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

**Note:** This Draft 2 Method is based on the Draft 1, 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) TEST CONDUCT

Field savings of CT Products shall be developed by ENERGY STAR CT Partners and will be assessed for product certification as well as for periodic reporting as detailed in the Partner Agreement.

1. Install and configure the ***most recent release*** of the ENERGY STAR CT Field Savings Software into a Python environment using the open-source code and instructions available at:
   - Documentation: http://thermostat.readthedocs.org/en/latest/
   - Source code: https://github.com/impactlab/thermostat

**Note:** EPA has added language requiring use of the most recent released software to assess CT savings.

2. From the set of all instances of a fielded CT Product in the US, generate a randomly selected data set via the following procedure: (***** throughout this section, see Sample Code in Appendix A)
   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
      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 aux and/or emergency heat
         2: Single stage heat pump without aux or emergency heat
         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
      vi. date2 is the last date for which interval data was reported

b. from the above metadata file, generate a 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
   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 EIA Building America Climate Zone to Zipcode Database. These climate zones are:

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

d. using Python Natsort (V 5.0.1), sort each climate zone metadata file by the unique thermostat_id assigned to each thermostat.*****

e. using the Python Numpy (V 1.10.4) random number generator, set a seed (supplied by EPA) for each climate zone manually with numpy.random.seed (number) and record these numbers.*****

**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 EIA climate zone, generate metadata files with the desired quantity of CTs (250 recommended) using the Numpy function numpy.random.choice, using replace=False to prevent sampling duplicates. If an EIA climate zone has too few thermostats to sample, include all thermostats for that zone.*

g. using the above metadata files, generate:
- a single metadata file, and
- a separate interval data file for each included unique thermostat_id,

that follow the requisite file format and content requirements in http://thermostat.readthedocs.io/en/latest/tutorial.html#input-data.*

**Note:** EPA has added a detailed, reproducible procedure to select a random sample, using the Python libraries Natsort, numpy.random.seed, and numpy.random.choice. Particular care is used in specifying a seed, so that the random selection created by running the script can be reproduced exactly if needed. This, combined with data retention requirements in section 4)8, significantly increases auditability of random sampling. Python sample code is included in Annex 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) and:

b. includes CT Products in each of the five Energy Information Agency (EIA) climate zones: Very Cold/Cold, Hot Dry/Mixed Dry, Hot Humid, Mixed Humid, and Marine.

c. adheres to the relevant reporting period criteria.

d. includes only CT Products that control one of the following types of HVAC equipment:
- Single stage heat pump with or without aux and/or emergency heat
- Single stage non heat pump with or without single-stage central air conditioning
- Single stage central air conditioning without central heating

**Note:** EPA has added language that ensures that 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.

4. Process the data set using the ENERGY STAR CT Field Savings software tool. 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 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 sufficiently certainty.
5. Noting that the software tool will exclude certain CTs from the assessment of mean savings, ensure that the output file includes at least **100** installed CT Products in each of the five EIA climate zones for both heating and cooling.[1] If not, repeat the procedure from step 4) 2f with additional CT Products in the data set.

**Note:** As the savings software employs selective filtering that eliminates certain CTs from being used to assess mean savings, EPA has proposed that mean savings be calculated using at least 100 CTs per climate zone. This criteria strives to ensure statistical significance while reducing potential barriers to new CT products and/or CT service providers with lower market penetration. EPA welcomes feedback on this strategy.

The outputs from the software in this draft reflect the current outputs. The outputs of the released V1.0 of the software tool may be adjusted, depending on what savings criteria are included in the final specification.

6. If the results do not indicate sufficient statistical certainty, the procedure may be repeated from step 4) 2f with a larger number of homes in the sample to decrease the uncertainty.
7. Submit the CT Field Savings software tool output file
   a. to the Certification Body for initial product certification, or
   b. to EPA for ongoing reporting.
8. Retain the following data for a period not less than five years from the associated file submission:
   c. raw data files for each thermostat_id in the CT data set used to assess savings
   d. all metadata files created in step 4) 2
   e. all random seeds with a reference to the metadata files for which each seed was used
   f. CT data set files, generated in step 4) 2 g, and used with the ENERGY STAR CT Field Savings Software to assess reported savings
   g. All output files submitted to EPA or to certification bodies.

[1] The version 1.0 ENERGY STAR CT Field Savings software tool will employ data filtering which may result in a reduction in the number of CT Products used to assess regional and National savings.

**Note:** EPA expects that, in accordance with stakeholder discussions, the requirement for demonstrated field savings in the specification will include some specific statistical certainty. For instance, the requirement may be written as minimum savings as demonstrated by the lower bound of the 95% confidence interval. This Method specifically allows re-running the procedure with a larger sample in order to improve the certainty of the results.

While Draft 1 required submission of specified outputs, this draft requires the software output file itself to be submitted.

Finally, in order to support auditability, this draft adds a minimum retention period of five years for all source data and input files to the ENERGY STAR CT Field Savings software tool. Note that while EPA does not plan to audit savings submissions associated with the Version 1.0 ENERGY STAR CT specification, several utility stakeholders have indicated that auditability is important to their adoption of the metric and of the specification. Thus, auditability supports broader adoption of the specification. In addition, it is in line with EPA's commitment to credible savings for ENERGY STAR products.

While EPA strives for a robust repeatable methodology for demonstrating savings, it is necessary to balance those goals with the need to minimize the burden on ENERGY STAR CT partners. With these goals in mind, EPA encourages stakeholder to provide specific feedback to inform random sample selection, as well as on the minimum sample size criteria and on data retention criteria.

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

```
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:
    pickle.dump(prng_state_exact,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,200, replace=False )
np.random.seed(102)
SampEIAHotHumid = np.random.choice(EIAHotHumid,200, replace=False )
np.random.seed(103)
SampEIAMixedHumid = np.random.choice(EIAMixedHumid,200, replace=False )
np.random.seed(104)
SampEIAHDMD = np.random.choice(EIAHDMD,200, replace=False )
np.random.seed(105)
SampEIAMarine = np.random.choice(EIAMarine,200, 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
```

```
210     results = SampEIAColdVCold
211     outfile =
212     np.append(results,[SampEIAHotHumid,SampEIAMixedHumid,SampEIAHDMD,SampEIAMarine])
213
214     #Save Sample items to file
215     np.savetxt('PRNG.csv',SortedEIASample, delimiter=",")
216     np.savetxt('PRNGvector.csv',outfile, delimiter=",")
217
218
```

## APPENDIX B: DESCRIPTION OF SAVINGS METHODOLOGY & ALGORITHMS USED IN THE SOFTWARE

**Note:** One stakeholder commented that it would be useful to have a text description of the algorithms used in the field savings software, since it is not included in the software documentation. In response, EPA has developed a Discussion Document that details the v0.3.4 software implementation, which currently includes (3) candidate methods for modeling the unique relationship between thermal load and HVAC run time in a home. EPA intends to select one of these methods for use in the Version 1.0 ENERGY STAR Connected Thermostat (CT) Program. Once the method is selected, EPA will add a detailed description herein. In the meantime, please refer to the Discussion Document.