# BSR/RESNET 1201-201x PDS-01

# Standard Method of Test for the Evaluation of Model Calibration Methods

6/08/2016

## Foreword (Informative)

This standard presents an analytical method of test for evaluating model calibration techniques. The basic concepts for this method are:

- A building energy simulation program is used to generate pre- and post-retrofit energy consumption data as a synthetic surrogate for data that could be measured in a real building (typically obtained from utility bills).
- The same simulation program that generates the surrogate measured consumption data is used in conjunction with a calibration technique to produce calibrated and non-calibrated energy consumption data.
- The calibration technique is evaluated by comparing the calibrated and non-calibrated consumption data to the surrogate measured consumption data along with other metrics.

The test method is useful in several ways including a) testing a single calibration method to see how well it works under a variety of test conditions, b) testing several calibration methods to determine under what test conditions each is best, c) investigating how much, and what kind, of informational content is needed in the synthetic data to achieve good calibrations with different calibration methods (eg. monthly vs daily vs hourly data and availability of different types of submetered or disaggregated data), d) testing with various amounts and kinds of "noise" in the synthetic data, and e) diagnostic testing.

The test method allows users of the standard to construct their own test cases and specifications. However, to avoid the work of creating new test cases, a set of tests and specifications is available from several studies known as the National Renewable Energy Laboratory (NREL) BESTEST-EX reports (Judkoff et. al. 2010, 2011a, 2011b. Information on how to use the BESTEST-EX test specifications is given in informative Appendix A.

A more complete explanation of the test method, its metrics, and its uses is given in informative Appendix B.

## Standard Method of Test for the Evaluation of Building Energy Analysis Model Calibration Methods (Normative)

### Purpose

This standard specifies a method of test for evaluating calibration methods that are used to reconcile building energy models with measured energy consumption data.

## Scope

This standard test procedure applies to calibration methods used with computer programs that predict the energy performance of buildings.

## Normative Definitions

(Informative note: Definitions in the Informative Definitions section in Appendix B shall not be used in the normative parts of this standard. The Informative Definitions section contains explanations, examples, and synonyms to afford the reader understanding and insight into the normative definitions and procedures.)

(Informative note: Italics designate a term that is defined in the Normative Definitions section.)

Absolute input: an input related to the retrofit that replaces a *base case model* input (see *Relative input*).

Approximate input: An input that has been determined to be uncertain and sensitive.

Approximate Input Range: Defined range of input value uncertainty for a given *approximate input*.

Automated Calibration Technique: A calibration technique that would not be helped by *non-permissible data*.

Base-Case Model: The model of the building before any retrofits are applied.

Calibrated inputs: Inputs that have been determined by fitting to synthetic energy use output data generated with the *explicit model*.

Calibrated model: The simulation model that contains the *calibrated inputs*.

Calibrated results: Output from the *calibrated model*.

Calibration Method: A technique or procedure that attempts to improve energy-related predictions by utilizing existing energy-related building performance data.

Calibrator: The individual human, team, or *automated calibration technique* that will perform the calibration.

Energy performance: The outputs or results, generated during execution of the tests, that quantify the energy-related performance of the test case buildings.

Explicit Input: An input value selected from within a defined range of uncertainty (see *approximate input range*).

Explicit Model: The simulation model that contains the *explicit inputs*.

Explicit results: Output from the *explicit model*.

Model: For purposes of this document, a model is 1) that part of a building energy simulation tool that contains the inputs and is used during execution of a simulation run; and/or 2) the mathematical and computer code representation inside a building energy simulation tool of a physical phenomenon.

Nominal Input: The input value that would be assumed if no calibration were performed.

Nominal Model: The simulation model that contains the *nominal inputs*.

Nominal Results: Output from the *nominal model*.

Non-permissible data: Data that shall not be known or used by the *calibrator* and includes all data which has not been defined as *permissible data*.

Permissible data: Data types and frequencies which have been defined at the beginning of the test procedure by *the test designer* as allowed to be known and used by the *calibrator*.

Post-retrofit model: *The base case model* that has been revised to include individual or combined packages of retrofit measures.

Pre-retrofit model: The *base-case model*.

Reference Results: The outputs or results from the *explicit model*.

Relative input: an input related to the retrofit that adjusts a *base case model* input (see *Absolute input*).

Savings: (pre-retrofit energy performance) – (post-retrofit energy performance)

Synthetic Energy Performance data: *Energy performance* data generated with the *explicit model*.

Synthetic Utility Bill Data: *Base-case energy performance* data generated with the *explicit model*.

Test Designer: Individual or team that designs the test and specifies which data shall be *permissible*, the *test metrics*, and units that shall be used to evaluate the test results.

Test Metrics: An evaluation basis for the test results that shall be defined by the *test designer* and shall at least include: a) a "goodness of fit" comparison between the *synthetic* and *calibrated* pre-retrofit building *energy performance data* (for *calibration methods* that produce a pre-retrofit *calibrated model*), b) a comparison between *explicit* and *calibrated* input values (for *calibration methods* that calibrate inputs), c) a comparison between the *explicit* and *calibrated* post-retrofit *savings* (applicable for all *calibration methods*), and d) a comparison between the *explicit* and *calibrated* post-retrofit building *energy performance data* (for *calibration methods* that produce a post-retrofit *calibrated model*). The types, frequencies, and mathematical relationships of the *Test Metrics* shall also be defined by the *Test Designer*.

Tool: The building energy simulation computer program that contains the model.

## **Procedures** (Normative)

**Rule 1)** The *calibrator* shall only know and use *permissible data*.

**Rule 2)** The simulation *tool* and modeling algorithms used for generating *nominal results, explicit results,* and *calibrated results* shall be the same.

**Rule 3)** The simulation *tool* and modeling algorithms used for generating *pre-retrofit* and *post-retrofit model* results shall be the same.

**Rule 4)**

- The inputs in the *nominal pre-retrofit model* shall be the same as the inputs for a *nominal post-retrofit model*, except where the input values are changed by the retrofit.
- The inputs in the *explicit pre-retrofit model* shall be the same as the inputs for an *explicit post-retrofit model*, except where the input values are changed by the retrofit.
- The inputs in the *calibrated pre-retrofit model* shall be the same as the inputs for a *calibrated post-retrofit model,* except where the input values are changed by the retrofit.

**Rule 5)** The simulation *tool* used for testing a calibration technique shall be the same simulation *tool* that is used for generating, *nominal results, explicit results*, and *calibrated results*.

**Rule 6)** Where there are options in a simulation program for modeling a specific energy- or temperature-related phenomenon, consistent modeling methods shall be used for all cases.

**Rule 7)** *Test metrics* shall at least include a) *calibrated results* versus *explicit results* for energy savings, b) *calibrated input* values versus *explicit input* values (for calibration techniques that calibrate inputs), c) *calibrated results* versus *explicit results* for *base-case energy performance* (for calibration techniques that calibrate the *base-case* model), and d) a comparison between the explicit and calibrated post-retrofit building energy performance data (for calibration methods that produce a post-retrofit calibrated model).

**Rule 8)** The *test designer, calibrator* and all who perform the tests shall adhere to rules 1 through 8.

**Step 1) The *test designer* shall define and document the nature, purpose, and scope of the calibration testing to be performed, including a) the *permissible data*, and b) the *test metrics*.**

**Step 2) The *test designer* shall create a *base-case* specification and *model* with all the required *nominal inputs* for the simulation *tool* to be used to model the *pre-retrofit base-case* building.**

**Step 3) The *test designer* shall use the *model* created in Step 2 to generate *nominal results* for *pre-retrofit energy performance*.**

**Step 4) The *test designer* shall create specifications and *models* with all the required *relative* or *absolute nominal inputs* for the retrofit measures to be applied, singly or in combination, to the *base-case* building.**

**Step 5) The *test designer* shall use the *model(s*) from Step 4 to generate *nominal post-retrofit energy performance* results and *savings.***

**Step 6) The *test designer* shall introduce input uncertainty into the *pre-retrofit* test specification:**

a. Using the *nominal model* from Step 2, perform sensitivity tests on inputs with potentially high uncertainties to determine their relative effects on *nominal model outputs*; select inputs that have both substantial uncertainties and substantial effects on outputs as *approximate inputs*.

b. Specify an uncertainty range (*approximate input range*) for each *approximate input*.

c. Select *explicit inputs* from the *approximate input ranges*.

**Step 7) The *test designer* shall perform simulations using the *explicit inputs* to generate *pre-retrofit base-case synthetic utility bill data.***

**Step 8) The *test designer* shall perform simulations to generate *reference results* and *savings* for the *post-retrofit* building:**

a. This shall be accomplished by adjusting appropriate *base case explicit inputs* with either *relative* or *absolute input* values, as specified for each retrofit case and combination of cases.

**Step 9) The *calibrator* shall perform the calibration and generate *base-case* and *calibrated savings results* using the calibration technique being tested and the associated building energy simulation *tool*.**

a. Predict energy savings via one of the following:

   i. Calibrate the *nominal base-case model inputs* using the *synthetic utility bills* and the uncertainty ranges (described in Steps 6 and 7 above), then apply the specified retrofit cases to the calibrated model.

   ii. Apply the specified retrofit to the uncalibrated *base-case model* and then calibrate or correct energy savings predictions using the *synthetic utility bills* (without adjustment to *base-case* model inputs).

   iii. Other calibration methods. ***Informative Note:*** This test method makes no recommendation about how to perform calibrations. Any calibration method that seeks to improve energy savings predictions through use of pre-retrofit building energy performance data may be tested via this method.

**Step 10) Evaluate the calibration technique*:***

a. Compare the “goodness of fit” between the *synthetic* and *calibrated* pre-retrofit building *energy performance data* (for *calibration methods* that produce a *pre-retrofit calibrated model*).

b. Compare the *explicit* and *calibrated* input values (for *calibration methods* that calibrate inputs).

c. Compare the *explicit* and *calibrated* post-retrofit savings (applicable for all *calibration methods*).

d. Compare the *explicit* and *calibrated* post-retrofit building *energy performance data* (for *calibration methods* that produce a post-retrofit *calibrated model*).

e. Informative Note: It may also be useful to compare *calibrated results* and *explicit reference results* to the *nominal*, uncalibrated results. Such a comparison can be used to determine the benefit of calibration.

f. Informative Note: The *test designer* may specify additional *test metrics* for evaluating the calibration technique.

## Appendix A: Example Test Specification (Informative)

The "pure" calibration test method may be applied using aspects of the BESTEST-EX test specification (Judkoff et al 2010, 2011a) as described in this Appendix. BESTEST-EX details a method in which several simulation programs were used as reference programs to generate average synthetic utility bill data and savings data (Judkoff et. al., 2011a, 52414). Such an approach tests both the simulation program and the associated calibration technique together. The test method described in this document is different in that any given building simulation tool can be used as its own reference in conjunction with a model calibration technique to test the calibration technique. BESTEST-EX introduced this concept and named it the "pure" calibration test method. Such an approach is a "pure" test of the calibration technique alone. The reader is advised to substitute the concept of a simulation tool generating its own synthetic utility bill data for the multiple reference simulation tool approach whenever it is discussed in BESTEST-EX. Furthermore, there are many items in the BESTEST-EX specification that the test taker may choose to disregard because they are difficult to model in their simulation tool. If it is easier to use the modeling approaches in your simulation tool instead of attempting to follow every detail of the BESTEST-EX specification, then use the modeling approaches in your tool (e.g., using your typical space conditioning model rather than the idealized space conditioning system as specified in BESTEST-EX). Be aware that if the exact BESTEST-EX specifications are not used, then comparison to the BESTEST-EX field trial results is not recommended. A good example of a study that employs the pure calibration test method is [Robertson et. al., 2013]. This study used a combination of data from the BESTEST-EX test specification and specification data created by the authors which is an allowable practice in this method of test.

Even though this is an informative section, the mandatory steps are repeated below with information about how BESTEST-EX can be used to assist in the creation of test specifications.

**Step 1) The *test designer* shall define and document the nature, purpose, and scope of the calibration testing to be performed, including a) the *permissible data*, and b) the *test metrics*.**

**Step 2) The *test designer* shall create a *base case* specification and model with all the required *nominal inputs* for the simulation tool to be used to model the *pre-retrofit base-case* building.**

- *Specifications: BESTEST-EX Section 1.2.1 can be used to guide the specification of the base-case building.*

**Step 3) The *test designer* shall use the model created in Step 2 to generate nominal pre-retrofit energy use output results.**

**Step 4) The *test designer* shall create specifications and models with all the required *relative* or *absolute* nominal inputs for the retrofit measures to be applied, singly or in combination, to the *base case* building model.**

- *Specifications: BESTEST-EX Sections 1.2.2 (physics retrofit cases) and 1.3.2 (calibration retrofit cases) can be used to guide the specification of retrofit measures.*

**Step 5) The *test designer* shall use the model(s) from Step 4 to generate *nominal post-retrofit* energy use output results and savings.**

**Step 6) The *test designer* shall introduce input uncertainty into the pre-retrofit test specification:**

a. Using the *nominal model* from Step 2, perform sensitivity tests on inputs with potentially high uncertainties to determine their relative effects on *nominal model outputs*; select inputs that have both substantial uncertainties and substantial effects on outputs as *approximate inputs*.
   - *Specifications: Uncertain and influential parameters are identified throughout BESTEST-EX Section 1.2.1. These are the inputs that have "min" and "max" values in additional to "nom" values. The user can identify alternative or additional approximate inputs.*

b. Specify an uncertainty range (*approximate input range*) for each *approximate input*.
   - *Specifications: Uncertainty ranges are specified for each approximate input in Section 1.2.1. The "min", "max", and "nom" values specify a triangular probability distribution that approximates the uncertainty for each input, as described in BESTEST-EX Appendix F.*

c. Select *explicit inputs* from the *approximate input ranges*.
   - *Specifications: Multiple calibration scenarios were generated for BESTEST-EX. Each calibration scenario has its own set of randomly selected explicit inputs. The randomly selected values for these scenarios can be found in Table 15 of [Judkoff et. al., 2011b]. If the test taker has followed the test procedures in BESTEST-EX, then it may be possible to use the values. If the test taker has deviated from the test specification (e.g., changes in uncertainty ranges, influential inputs, buildings specifications), then they need to generate new calibration scenarios by randomly selecting sets of explicit input values specific to their test specification (an approach similar to that described in BESTEST-EX Appendix F could be used).*

**Step 7) The *test designer* shall perform simulations using the *explicit inputs* to generate *pre-retrofit base-case synthetic utility bill data.***

- *Reminder: The synthetic utility bill data generated with your own tool and model should be used instead of the reference synthetic utility bill data in BESTEST-EX.*

**Step 8) The *test designer* shall perform simulations to generate *explicit post-retrofit reference energy use output results* and savings:**

a. This shall be accomplished by adjusting appropriate *base case explicit inputs* with either *relative* or *absolute input* values, as specified for each retrofit case and combination of cases.

**Step 9) The *calibrator* shall perform the calibration and generate *base case* and *calibrated savings* results using the *calibration technique* being tested and the associated building energy simulation tool.**

a. Predict energy savings via one of the following:
   i. Calibrate the *nominal base-case model* inputs using the *synthetic utility bills* and the uncertainty ranges (described in Steps 6 and 7 above), then apply the specified *retrofit cases* to the calibrated model.
   ii. Apply the specified retrofit to the uncalibrated base case model and then calibrate or correct energy savings predictions using the *synthetic utility bills* (without adjustment to *base-case* model inputs).
   iii. Other calibration methods. ***Informative Note:*** This test method makes no recommendation about how to perform calibrations. Any calibration method that seeks to improve energy savings predictions through use of pre-retrofit building energy performance data may be tested via this method.

**Step 10) Evaluate the calibration technique:**

a. Compare the "goodness of fit" between the *synthetic* and *calibrated* pre-retrofit building *energy performance data* (for *calibration methods* that produce a *pre-retrofit calibrated model*).
b. Compare the *explicit* and *calibrated* input values (for *calibration methods* that calibrate inputs).
c. Compare the *explicit* and *calibrated* post-retrofit *savings* (applicable for all *calibration methods*)
d. Compare the *explicit* and *calibrated* post-retrofit building *energy performance data* (for *calibration methods* that produce a post-retrofit *calibrated model*).
e. Informative Note: It may also be useful to compare *calibrated results* and *explicit reference results* to the *nominal*, uncalibrated results. Such a comparison can be used to determine the benefit of calibration.
f. Informative Note: The *test designer* may specify additional *test metrics* for evaluating the calibration technique.

- *Specifications: BESTEST-EX Appendix G describes how results can be evaluated to compare calibration methods and to assess the benefit of calibration. For the pure calibration test method, all results used in the comparison must be generated by the simulation tool that is used for the testing. Additionally, [Robertson et. al., 2013] demonstrates how goodness of fit, input agreement, and savings prediction accuracy can be evaluated using the pure calibration approach.*

## Appendix B: Testing Model Calibration Techniques Using Synthetic Data and the "Pure" Test Method Path (Informative)

This section summarizes a method for testing model calibration procedures. Calibration is commonly used in conjunction with energy retrofit audit models (Judkoff et. al., 2010, 2011a, 2011b; Reddy et. al., 2006; Robertson et. al., 2013). This test method was initially developed by NREL for testing calibration procedures used with residential retrofit audit software, however, the method is applicable in a commercial building context. Other terms frequently used to describe model calibration include model tuning, model true-up, and model reconciliation.

Typically, residential and commercial model calibration has been implemented using monthly energy data collected from utility bills for an existing building that is about to receive an energy retrofit. Sometimes sub-metered, disaggregated, or higher frequency data is also available. An audit is conducted to gather information about the building needed to assemble an input file for a building energy simulation tool. A calibration technique is used to reconcile model predictions with the utility data, and then the "calibrated model" is used to predict energy savings and energy cost savings from a variety of retrofit measures and combinations thereof. Many variations on this approach exist, including some where the savings predictions are subjected to calibration instead of, or along with the model inputs.

While it is logical to use the actual performance data of the building to tune the model, it is not certain that this will result in a model that better predicts post-retrofit energy savings. When calibrating a large number of inputs to a limited number of outputs (mathematically this is called an underdetermined or over-parameterized problem), there can be many combinations of input parameters that will result in a close match to the utility bill data, so a close match is not in itself proof of a good calibration. The lower the frequency of the building performance data, or the lower the informational content of that data, the lower the probability that the calibration actually improves the model and associated energy savings predictions. Therefore any method to test calibration techniques should use at least the following three figures of merit: a) the accuracy of the savings prediction, b) how closely the calibrated input parameter values match the actual parameter values (when these can be known), and c) the goodness of fit between the modeled and measured data. A limiting factor in attempting to empirically validate calibration techniques is the lack of high quality monthly energy data for at least a year pre- and post-retrofit (higher frequency data and sub-metered data are better), good pre- and post-retrofit building characteristics data, local pre-and post-retrofit weather data, and the dates of the retrofit installations. Until a sufficient amount of such empirical data is available to researchers, an alternative test method can be used in which a building energy simulation tool is used to generate its own synthetic utility bill energy use data, post retrofit energy use data, and energy savings data. The synthetic data may be considered as a surrogate for actual data. Ideally, empirical data would be available for "bottom line" validation, while the analytical test method described here is used for diagnosis and improvement of calibration techniques.

This method of test refers to BESTEST-EX in Appendix A (below) to provide example test specifications that can be used in lieu of creating new test specifications (Judkoff et. al., 2010, 47427). BESTEST-EX details a method in which several simulation programs were used as reference programs to generate average synthetic utility bill data and savings data (Judkoff et. al., 2011a, 52414). Such an approach tests both the simulation program and the associated calibration technique together. The test method described in this document is different in that any given building simulation tool can be used as its own reference in conjunction with a model calibration technique to test the calibration technique. BESTEST-EX introduced this concept and named it the "pure" calibration test method. Here, we further develop this method. In the pure method a simulation tool generates its' own synthetic utility bill data. Such an approach is a "pure" test of the calibration technique alone. The reader is advised to substitute the concept of a simulation tool generating its' own synthetic utility bill data for the multiple reference simulation tool approach whenever it is discussed in BESTEST-EX. Furthermore, there are items in the BESTEST-EX specification that the test taker may choose to disregard because they are difficult to model in their simulation tool. If it is easier to use the modeling approaches in a given simulation tool instead of attempting to follow every detail of the BESTEST-EX specification, then use the modeling approaches in the tool (e.g., using the tool's typical space conditioning model rather than the idealized space conditioning system as specified in BESTEST-EX). Be aware that if the exact BESTEST-EX specifications are not used, then comparison to the BESTEST-EX field trial results is not recommended. A good example of a study that employs the pure calibration test method is [Robertson et. al., 2013]. This study used a combination of data from the BESTEST-EX test specification and specification data created by the authors which is an allowable practice in this method of test.

The "pure" method for testing calibration techniques follows the general procedures outlined below immediately after the Definitions section.

## Informative Definitions

Note: Definitions of terms in this section are explanatory with examples to help the reader gain understanding and insight regarding this method of test. Informative Definitions shall not be used in the normative parts of this standard. Synonymous terms, are given to help in understanding the concepts. Normative definitions used in the normative sections are expressed in mandatory language.
Note: *Italics* in this section designate that a word or phrase is defined in this section.

Absolute input: an input related to the retrofit that replaces a *base case model* input. For example, the new Solar Heat Gain Coefficient (SHGC) for a replacement window. (see *Relative Input*)

Approximate input: An input that has been determined to be uncertain and sensitive. Such inputs are good candidates to test model input calibration techniques, and are also frequently relevant to applied retrofit measures.

Approximate Input Range: Defined range of input value uncertainty for a given *approximate input*.

Automated Calibration Procedure: A calibration procedure that requires no human judgment; i.e., a calibration procedure that would not be helped by a human knowing the test case explicit inputs or any other *non-permissible* data.

Automated Calibration Technique: same as *automated calibration procedure*.

Base-Case Model: The model of the building before any retrofits are applied.

Calibrated inputs: Inputs that have been determined by fitting to synthetic utility bill data (typically gas and electric consumption output). The synthetic utility bill data is generated with the *explicit model*.

Calibrated model: The simulation model that contains the *calibrated inputs*.

Calibrated results: Output from the *calibrated model*.

Calibration Method: A technique or procedure that attempts to improve energy-related predictions by utilizing existing energy-related building performance data.

Calibrator: The human or *automated calibration technique* that will perform the calibration.

Energy performance: A general term for the outputs or results generated during execution of the tests in this test method. These outputs quantify the energy-related performance of the test case building(s).

Explicit Input: An input value selected from within a defined range of uncertainty (see *approximate input range).*

Explicit Model: The simulation model that contains the *explicit inputs* and which is used to generate the synthetic energy performance data (typically gas and electric consumption output such as would appear on a utility bill, hence the term *synthetic utility bill data*).

Explicit results: Output from the *explicit model*.

Model: Has a variety of meanings. For purposes of this document two meanings are important. 1) A model is that part of a building energy simulation tool that contains the inputs and is used during execution of a simulation run; 2) a model is the mathematical and computer code representation inside a building energy simulation tool of a physical phenomenon.

Nominal Input: The input value that would be assumed if no calibration were performed, such as the audit-recorded value, the software-defaulted value, or values obtained using credible sources such as the ASHRAE Handbook of Fundamentals. For certain kinds of tests, nominal inputs could be defined that are erroneous (very far from typical values) mimicking human error (e.g., typographical or bias errors). In BESTEST-EX nominal inputs are sometimes referred to as physics inputs, although there are a few instances in BESTEST-EX where the physics inputs and nominal inputs are different.

Nominal Model: The simulation model that contains the *nominal inputs*. In BESTEST-EX nominal models are sometimes referred to as physics models or physics test cases.

Nominal Results: Output from the *nominal model*. In BESTEST-EX nominal results are sometimes referred to as physics results or physics test case results.

Non-permissible data: Non-permissible data should not be known or used by the *calibrator* and includes all data which has not been defined as *permissible data*.

Permissible data: Data types and frequencies which have been defined at the beginning of the test procedure by *the test designer* as allowed to be known and used by the *calibrator*.

Physics Tests: A term used in BESTEST-EX to describe tests of the correctness of the building energy simulation tool. Physics tests use *nominal inputs*.

Post-retrofit model(s): *The base case model* that has been revised to include individual or combined packages of retrofit measures.

Pre-retrofit model: The *base-case model* (the model before any retrofit measures have been applied).

“Pure” Method for testing calibration techniques: A method in which the building energy simulation tool to be used in conjunction with a *calibration technique*, is used to generate the *synthetic utility bill data* and all other simulation results needed to perform the test.

Reference Model: The *explicit model*.

Reference Results: The outputs or results from the *explicit model*.

Relative input: An input related to the retrofit that adjusts a base case input. For example, add R-6 to the existing roof insulation. (see *Absolute input*)

Synthetic Energy Performance data: Base case and retrofit energy output data generated with the *explicit model*. This data is a surrogate for actual energy use data measured in a real building for purposes of this test method.

Synthetic Utility Bill Data: Utility Bill Data generated using the *explicit model*. Synthetic utility bill data can be at any frequency, level of sub-metering, or degree of disaggregation.

Test Designer: Individual or team that designs the test and specifies which data shall be “permissible”, the “test metrics”, and units that shall be used to evaluate the test results.

Test Metrics: An evaluation basis for the test results that shall be defined by the *test designer* and shall at least include: a) a “goodness of fit” comparison between the synthetic and calibrated pre-retrofit building energy performance data (for calibration methods that produce a pre-retrofit calibrated model), b) a comparison between explicit and calibrated input values (for calibration methods that calibrate inputs), c) a comparison between the explicit and calibrated post-retrofit savings (applicable for all calibration methods), and d) a comparison between the explicit and calibrated post-retrofit building energy performance data (for calibration methods that produce a post-retrofit calibrated model). The types, frequencies, and mathematical relationships of the *test metrics* shall also be defined by the t*est designer*.

Tool: The building energy simulation computer program that contains the model.

## **Outline and Explanation of Procedures** (Informative)

The test designer defines the scope of the calibration testing. The purpose is to determine at the beginning of the test process the exact nature of the test, what information is permitted to be known and used by a human or automated calibrator, and what metrics shall be used to evaluate the test results. For example, if the scope of the calibration testing is intended to mimic a common audit and calibration scenario (e.g., the approach used in BESTEST-EX), then the only permissible data for the calibrator would be a) synthetic whole-building monthly gas and electric energy use data (no disaggregated data allowed for this example), b) nominal inputs representing the input data that would be collected by an auditor or modeler, and c) input uncertainty ranges for those inputs that are to be calibrated. The metrics for this example would typically be a) calibrated monthly and annual energy savings versus explicit monthly and annual energy savings, b) calibrated input values versus explicit input values (for methods that calibrate inputs), and c) calibrated base-case monthly and annual energy use data versus explicit base-case monthly and annual energy use data. If the goal of the testing is to determine the benefit of using hourly synthetic smart meter data, or submetered or disaggregated data, (e.g., the approach used in Robertson et. al.) then the test designer shall define precisely what type and frequency of data shall be permitted to be known and used by the calibrator and what types, frequencies, and mathematical relationships of results data shall be used as *test metrics*.

Create specifications for a pre-retrofit base case test building(s) defining the values for nominal input parameters that a building simulation model would need. This does not have to be a real existing building, but the specification should be representative of the types of buildings for which the calibration technique(s) will be used.

Use the nominal model to generate nominal pre-retrofit energy use output results. The nominal model is sometimes referred to as the physics model in BESTEST-EX because BESTEST-EX included a series of building physics tests. The nominal model can also be used by the test designer to identify sensitive parameters.

Create specifications for individual and packages of retrofit measures to be applied to the pre-retrofit test building(s). Nominal input values for the retrofit measures should be expressed as “relative” or “absolute” values depending on the kind of retrofit. For example, adding R-15 of insulation to existing insulation in the attic would be a “relative” retrofit parameter value, whereas a window replacement would consist of several “absolute” retrofit parameter values. This distinction becomes important when applying retrofit measures to the calibrated base case because key base case parameter values are considered uncertain.

Use the nominal retrofit models to generate post-retrofit energy use output results and savings. Nominal retrofit models are sometimes referred to as physics retrofit models in BESTEST-EX.

Introduce input uncertainty into the pre-retrofit test specification. This represents the uncertainty associated with collecting audit survey data and developing inputs from that data.

- Use the nominal pre-retrofit model to perform sensitivity tests on inputs with potentially high uncertainties to determine their relative effects on outputs; select inputs that have both substantial uncertainties and effects on outputs as *approximate inputs*.
- Specify an uncertainty range (*approximate input range*) for each *approximate input*.
- Select *explicit inputs* from within the *approximate input ranges*. It is necessary that those who will be performing the calibrations, *the calibrators,* are blind to the explicit inputs. It may be useful to choose combinations of explicit inputs that yield high, medium, and low pre-retrofit energy use. One way to select *explicit inputs* is randomly within the *approximate input ranges*. Another would be to select *explicit inputs* as the most-probably “correct” input, for example when the *nominal inputs* are designed to test a calibration process for correcting errors such as typographical or bias in the model input.

Perform simulations using the *explicit inputs* to generate the pre-retrofit base case synthetic utility bill data. This is typically monthly electric and gas consumption data, but the method can be used to generate and test against higher frequency or lower frequency synthetic building energy performance data. Also, end-use data at varying levels of disaggregation can be used, mimicking the availability of sub-metered data.

Perform simulations to generate explicit post-retrofit energy results and savings results. Starting with the appropriate explicit pre-retrofit base-case model, adjust appropriate base case inputs for each retrofit case and combinations of cases.

Develop tested program and calibration technique results. The *calibrator* may only know *permissible data*. Typically, this would mean that the calibrator could know the nominal model and the approximate input ranges, but could not know the explicit inputs or the explicit retrofit outputs and savings results until the test is completed.

Predict energy savings via one of the following:

- Calibrate the base-case model inputs using the synthetic utility bills (described in 6 above), then apply the specified retrofit cases to the calibrated model.
- Apply the specified retrofits to the uncalibrated base-case model and then calibrate or correct energy savings predictions using the synthetic utility bills (without adjustment to base-case model inputs), e.g., as (calibrated savings) = (predicted savings) × (base case actual bills)/(base case predicted bills).
- Other calibration methods. This test method makes no recommendation about how to perform calibrations. Any calibration method that seeks to improve energy savings predictions through use of pre-retrofit building energy performance data may be tested via this method.

Use, at least, the following comparisons to evaluate the adequacy of the tested calibration technique:

- Compare the goodness of fit between the synthetic and calibrated pre-retrofit building energy performance data(for calibration methods that produce a pre-retrofit calibrated model).
- Compare the explicit and calibrated post-retrofit building energy performance data (for those calibration techniques that produce a post-retrofit calibrated model).
- Compare the savings predictions from the tested simulation tool and associated calibration technique, versus the savings predictions from the same simulation tool run with explicit inputs (applies to all calibration techniques).
- Compare explicit and calibrated input values (for calibration methods that calibrate inputs).

All of these comparisons are important for assessing the accuracy of the calibration method. A large disagreement in any one of them indicates the presence of compensating errors or some other error. Not all calibration methods will allow all of the above comparisons, however, all calibration methods will allow comparison of the savings predictions from the tested simulation tool and any associated calibration techniques, to the savings predictions from the same tool run with the explicit inputs. If generated, the calibrated model outputs and savings may also be compared to the nominal outputs and savings (the "uncalibrated" results) to gauge the benefit of calibration (Judkoff et. al., 2010, 47427, Appendix G).

The method for testing model calibration techniques described above is a "pure" calibration test in that the synthetic utility billing data is generated with the tested program, and the program accuracy related to building physics modeling is not tested. The pure calibration test requires a) automated calibration where no human judgment is necessary (i.e., a calibration procedure that would not be helped by a human knowing "non-permissible data"), or b) that the modeler running the calibration test does not know "non-permissible data". In the most common cases non-permissible data would include a) the explicit inputs used to develop the synthetic utility bills, b) the results from the explicit pre-retrofit model(s) except for those results defined as permissible such as synthetic utility bill data, and c) the results from the post-retrofit model(s) and the savings from the retrofits until the test is complete.

This method facilitates "self-testing" of a calibration technique, and is useful in several ways, including: a) testing a single calibration method, b) testing several calibration methods to determine under what test conditions each is best, and c) investigating how much, and what kind, of informational content is needed in the synthetic calibration data to achieve good calibrations with different calibration techniques (eg. monthly vs daily vs hourly data and availability of different types of submetered or disaggregated data). The pure calibration test, however, may not be practical for a certification test that must be administered by a third party organization and where an honor system is not deemed appropriate. A method to facilitate third party testing which assures that the person performing the test does not know the explicit inputs, has also been developed (Judkoff et al. [2011a, 2011b]) and is referred to as the "reference program method." The main difference between the two test methods is that for the reference program method several (preferably at least three) reference programs are used to generate the synthetic utility bills, and to create the reference energy savings data. The bills and the savings are taken as the average of the reference program results. The reference program method is both a test of the calibration technique, and a test of how closely the physics models in the tested program match the physics models in the reference programs. Example acceptance criteria may be used to facilitate the comparison of energy savings predictions (Judkoff et al 2011a). Figure 1 illustrates the overall conceptual approach to testing model calibration techniques and illustrates both the "reference program" and "pure" methods.

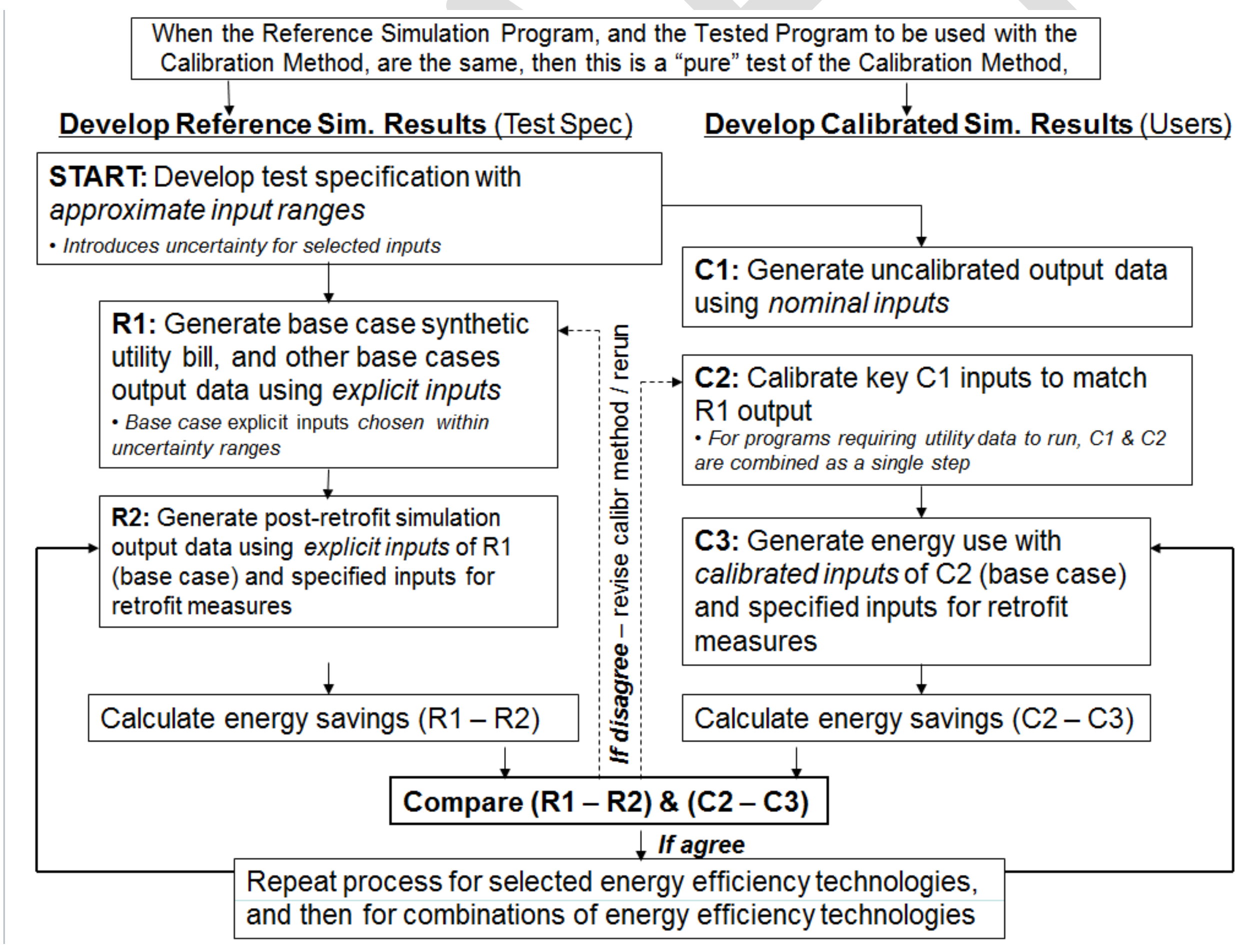


**Fig. 1 Calibration Cases Conceptual Flow (Judkoff et al. 2011a)**

## Appendix C: References (Informative)


Judkoff, R.; Polly, B.; Bianchi, M.; Neymark, J. 2010. *Building Energy Simulation Test for Existing Homes (BESTEST-EX).* NREL/TP-550-47427. Golden, CO, USA: National Renewable Energy Laboratory. http://www.nrel.gov/docs/fy10osti/47427.pdf

Judkoff, R., Neymark, J., Polly, B., Bianchi, M. 2011a. "The Building Energy Simulation Test For Existing Homes (BESTEST-EX) Methodology." *Proceedings Building Simulation 2011a,* International Building Performance Simulation Association. Also see pre-printed version, NREL CP-5500-51655. National Renewable Energy Laboratory, Golden, CO. http://www.nrel.gov/docs/fy12osti/51655.pdf.

Judkoff, R., Polly, B., Bianchi, M., Neymark, J., Kennedy, M. 2011b. *Building Energy Simulation Test For Existing Homes (BESTEST-EX): Instructions for Implementing the Test Procedure, Calibration Test Reference Results, and Example Acceptance-Range Criteria*. NREL TP-5500-52414. National Renewable Energy Laboratory, Golden, CO. http://www.nrel.gov/docs/fy11osti/52414.pdf.

Reddy, T.A., I. Maor, C. Panjapornporn, and J. Sun. 2006. Procedures for reconciling computer-calculated results with measured energy use, RP-1051 final research report. Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.

Robertson, J.; Polly, B.; Collis, J. (2013). Evaluation of Automated Model Calibration Techniques for Residential Building Energy Simulation. 91 pp.; NREL Report No. TP-5500-60127.