



**BSR/ASHRAE Addendum d to
ANSI/ASHRAE Standard 140-2023**

First Public Review Draft

Proposed Addendum d to Standard 140- 2023, Method of Test for Evaluating Building Performance Simulation Software

**First Public Review Draft (April 2026)
(Draft shows Proposed Changes to Current Standard)**

This draft has been recommended for public review by the responsible project committee. To submit a comment on this proposed standard, go to the ASHRAE website at www.ashrae.org/standards-research--technology/public-review-drafts and access the online comment database. The draft is subject to modification until it is approved for publication by the Board of Directors and ANSI. Until this time, the current edition of the standard (as modified by any published addenda on the ASHRAE website) remains in effect. The current edition of any standard may be purchased from the ASHRAE Online Store at www.ashrae.org/bookstore or by calling 404-636-8400 or 1-800-727-4723 (for orders in the U.S. or Canada).

The appearance of any technical data or editorial material in this public review document does not constitute endorsement, warranty, or guaranty by ASHRAE of any product, service, process, procedure, or design, and ASHRAE expressly disclaims such.

© 2026 ASHRAE. This draft is covered under ASHRAE copyright. Permission to reproduce or redistribute all or any part of this document must be obtained from the ASHRAE Manager of Standards, 180 Technology Parkway, Peachtree Corners, GA 30092. Phone: 404-636-8400, Ext. 1125. Fax: 404-321-5478. E-mail: standards.section@ashrae.org.

ASHRAE, 180 Technology Parkway, Peachtree Corners GA 30092

(This foreword is not part of this standard. It is merely informative and does not contain requirements necessary for conformance to the standard. It has not been processed according to the ANSI requirements for a standard and may contain material that has not been subject to public review or a consensus process. Unresolved objectors on informative material are not offered the right to appeal at ASHRAE or ANSI.)

Foreword to Addendum D

This addendum adds acceptance criteria for the building thermal fabric empirical validation tests of Section 13. The criteria apply to reported hourly heater energy consumption time-series results for the best-steady-state period for a given test case, included in the standard output reports for Section 13.

Normative Annex A3 contains statistical (CVRMSE) bounds for specific test cases that will be applied to software results to determine acceptance. The minimum number of range cases within the test group to pass has been updated to reflect including more cases in the acceptance criteria.

Informative Annex B12 has been updated to include the procedure for developing the acceptance criteria ranges for the Section 13 time-series results.

[Note to Reviewers: This addendum makes proposed changes to the current standard. These changes are indicated in the text by underlining (for additions) and ~~striketrough~~ (for deletions) except where the reviewer instructions specifically describe some other means of showing the changes. Additional material is provided for context only and is not open for comment except as it relates to the proposed changes.]

Addendum d to Standard 140-2023

Add the following new term to Section 3.1.

[...]

Coefficient of Variation of the Root Mean Square Error of the Simulation ($CVRMSE_{sim}$): the variation of simulation results from a measured mean value during a given best-steady-state time period, formulated as:

$$CVRMSE_{sim} = \sqrt{\frac{1}{n} \sum_{i=1}^n (S_i - \bar{M})^2 / \bar{M}}$$

where:

n = number of data points, S_i = individual simulation results (e.g., hourly) during the designated time period, and \bar{M} = mean of the measurements.

Add the following new terms to Section 3.2.

[...]

$CVRMSE_{sim}$: Coefficient of Variation of the Root Mean Square Error of the Simulation

[...]

\bar{M} : Mean of the measurements during a given best-steady-state time period. See *Coefficient of Variation of the Root Mean Square Error of the Simulation ($CVRMSE_{sim}$)*.

[...]

$Q_{hr.meas.mean}$: Mean of the measured heating energy during a given best-steady-state time period, Wh/h. See *\bar{M} and Coefficient of Variation of the Root Mean Square Error of the Simulation ($CVRMSE_{sim}$)*.

Revise Annex A3 as follows.

**NORMATIVE ANNEX A3
SOFTWARE ACCEPTANCE CRITERIA**

This annex provides the acceptance criteria for the software. In order to comply, the software must demonstrate that a sufficient number of software results (see Section A3.2) are within the ranges specified in Section A3.1 for each set of test cases.

Software includes the calculation engine and the corresponding user interface. The testing of software only meets the requirements of this annex for that software and cannot be used as a proxy for documenting compliance of another software that uses the same calculation engine but a different user interface.

A3.1 Class I Software Ranges. Tables A3-1 through A3-10~~9~~ give the acceptance criteria ranges for the test groups. Ranges are given for selected individual test-case results (e.g., Case 600); differences between the results of two test cases, i.e., result sensitivities (e.g., Case 610 – Case 600); and daily hourly time series results (e.g., February 1 Case 600).

For single-value results (see Tables A3-1, A3-4, A3-7 through A3-9), a software result is within a given range when the software result is greater than or equal to the lower limit and less than or equal to the upper limit shown for the given range in the tables. (**Informative Note:** The setting of these ranges is based on the calculations shown in Informative Annex B12, Section B12.1.1.1, “Acceptance Criteria Bounds Calculations,” “Equations for Single-Value Results.”)

For the time series ranges bounds for software-to-software comparative tests, comparison metrics (Pearson correlation coefficient [for curve shape] and Root Mean Square Error (RMSE) [for value scaling]) for the software results are calculated versus the representative curve results (see Tables A3-2 and A3-5), and then these metrics are compared to the ranges (see Tables A3-3 and A3-6). A software results set is within a given range if both the calculated Pearson correlation coefficient is above the limit in the table and the calculated RMSE value is below the limit in the table. (**Informative Note:** The setting of these ranges is based on the calculations shown in Informative Annex B12, Section B12.1.1.2, “Acceptance Criteria Bounds Calculations,” “Equations for Time Series Results for Software-to-Software Comparative Tests.”)

For the time-series range bounds for empirical validation tests, statistical bounds are defined by the measured data uncertainty, which includes the coefficient of variation of the root mean squared error (CVRMSE) of the measured data for a given test case and the instrument uncertainty. A software results set is within a given range if its calculated $CVRMSE_{sim}$ value is at or below the $CVRMSE_{sim}$ Upper Limit for the best-steady-state time period of a given test case shown in Table A3-10. (**Informative Note:** $CVRMSE_{sim}$ Upper Limit values are developed from the equations shown in Informative Annex B12, Section B12.1.1.3, “Acceptance Criteria Bounds Calculations”, “Equations for Time Series Results for Empirical Validation Tests”).

For comparison of $CVRMSE_{sim}$ versus the $CVRMSE_{sim}$ Upper Limit for a given test case, $CVRMSE_{sim}$ shall be determined by applying the following as listed in Table A3-10:

- The reference measurement (\bar{M}) as mean measured heating energy for the best-steady-state period of a given test case ($Q_{hr, meas, mean}$)
- The hourly simulated heating energy (Q_{hr}) for the best-steady-state time period for a given test case per the listed test specification sections, taken from simulation results entered in the listed cell ranges of the normative output template (Std140_ET_Output.xlsx).

The criteria are organized by test suite as follows.

- Tables A3-1 through A3-6; ~~are related to the~~ building thermal envelope and fabric load tests of Section 7
- Table A3-7; ~~is related to the~~ space-cooling equipment performance tests of Sections 9.2.3 and 9.2.4
- Table A3-8; ~~is related to the~~ space-heating equipment performance tests of Section 10
- Table A3-9; ~~is related to the~~ air-side HVAC equipment performance tests of Section 11
- Table A3-10: building thermal fabric empirical validation tests of Section 13.

[...]

Table A3-10 Building Thermal Fabric Empirical Validation—Time Series Ranges

Range Case	$CVRMSE_{sim}^a$ Upper Limit	\bar{M}^b as $Q_{hr.meas.mean}$ (Wh/h)	Best-Steady-State Time Period		
			Section Reference	Std140 ET Output.xlsx Cells	
				Tab	Cell Ranges
ET110A1	0.0099	634.75	13.3.2.4.1	ET110A1	K370:K387
ET110B1	0.0109	640.40	13.3.3.4.1	ET110B1	K370:K387
ET100A1	0.0074	736.61	13.3.4.3.1	ET100A1	K190:K244
ET100B1	0.0096	737.59	13.3.5.3.1	ET100B1	K190:K244
ET100A3	0.0186	736.61	13.3.6.4.1	ET100A3	K190:K244
ET100B3	0.0241	737.59	13.3.7.4.1	ET100B3	K190:K244

a. As defined in Section 3. (**Informative Note:** Application of this metric is elaborated in Informative Annex B12, Section B12.1.1.3. A utility for calculating $CVRMSE_{sim}$ for a tested software is provided at <https://data.ashrae.org/standard140/analysis.html>.)

b. As defined in Section 3.2. (**Informative Note:** \bar{M} is the measurement basis [i.e., measured mean heating energy ($Q_{hr.meas.mean}$)] for determining $CVRMSE_{sim}$ as described in Informative Annex B12, Section B12.1.1.3. Simulation results are not expected to exactly match \bar{M} .)

Table A3-110 Acceptance Range Pass Criteria

Test Group	Tables of Ranges	Number of Range Cases in Test Group	Minimum Number of Range Cases within the Test Group to Pass
Thermal Fabric Low Mass	A3-1, A3-3	34	30
Thermal Fabric High Mass	A3-4, A3-6	28	25
<u>Thermal Fabric Empirical Validation</u>	<u>A3-10</u>	<u>6</u>	<u>5</u>
Cooling Equipment	A3-7	14	12
Heating Equipment	A3-8	6	5
Air-side Equipment	A3-9	54	48

A3.2 Number of Results within Acceptance Range to Pass a Test Group. Table A3-110 gives the number of range cases for which a software’s results must be within the acceptable ranges to pass a test group. Software shall pass a test group when the count of results computed for each range case, from the corresponding tables of ranges, meets or exceeds the minimum count required in Table A3-110.

Informative Note: See Informative Annex B12, Section B12.4 for a description of the procedures used to determine the numbers in the table.

A3.3 Submission of Results. Test results shall be provided in the normative output forms of Normative Annex A2. Submittals shall also include a complete set of reports, as described in Normative Annex A2, Attachment A2.9, with all report blocks completed.

Revise Annex B12 as follows.

INFORMATIVE ANNEX B12 DEVELOPMENT OF ACCEPTANCE CRITERIA

B12.1 Acceptance Criteria Bounds Calculations. The results from a reference set of software programs or empirically measured data where available (see Section B12.1.2), also called “reference results” here, are used to calculate the upper and lower bound limits for the acceptance criteria. Where a reference set of software are applied (see Sections B12.1.1.1, B12.1.1.2, and B12.1.2), they are also called “reference results” here. Applying empirically measured data for setting bound limits is described in Section B12.1.1.3.

B12.1.1 Equations for Setting Bound Limits. There are two different types of results included in the acceptance criteria: single-value results (e.g., annual totals) and time series results (e.g., hourly zone loads or heating energy for a specific time period day). There are different calculation methodologies for each of these types of results.

[...]

B12.1.1.2 Equations for Time Series Results for Software-to-Software Comparative Tests.

[...]

B12.1.1.3 Equations for Time-Series Results for Empirical Validation Tests.

B12.1.1.3.1 Considerations for an accuracy metric based on measured data

- The dependent parameter of measured data and simulation models are compared at intended steady-state, so for the purpose of comparison, a single underlying physical steady-state value is sought for comparing models to measurements.
- Measurement is subject to instrument error as well as noise from external disturbances, control system response, etc. These noise elements should be uncorrelated to the steady-state value and are minor relative to it, thus, for the purpose of comparing models to measurements, the best estimate of the underlying physics is the mean of the dependent parameter of the measured data during its steady-state period.
- Simulation software is intended to model the same steady-state conditions as the measurements, but with some simplifications (e.g., different control responses). Hence, the best reference for simulation is the best estimate of the underlying physics, i.e., the mean of the measured data (see item b above) and *not* the hourly measured data points that contain multiple noise elements.

B12.1.1.3.2 Characterizing Measured Data

During a designated steady-state time period, the measured data have minor variations from the mean caused by instrument uncertainty as well as external disturbances, control response, etc. During this time period, the measurement quality is characterized by estimating the total “mean square error to the mean” (MSE_{meas}), which is of the same form as a standard deviation calculation. Additionally, the effects of instrument uncertainty are included. This leads to the following equation:

$$MSE_{meas} = \frac{1}{n-1} \sum_{i=1}^n (M_i - \bar{M})^2 + u_{instr}^2$$

where:

- n = number of data points, M_i = individual measurements (e.g., hourly), \bar{M} = mean of the measurements, and u_{instr}^2 = instrument uncertainty variance
- the denominator is $n - 1$ rather than n because the M_i are used to estimate \bar{M} , which reduces the total degrees-of-freedom by one.

B12.1.1.3.3 Characterizing Simulation Data

The simulations also have variation from the mean, with MSE_{sim} as:

$$MSE_{sim} = \frac{1}{n} \sum_{i=1}^n (S_i - \bar{M})^2$$

where:

- S_i = individual simulation results (e.g., hourly)

- MSE_{sim} has denominator n , because \bar{M} is independent of the S_i values.

B12.1.1.3.4 Defining a Metric to Characterize Accuracy of Simulations Versus Measurements

Based on the above characterizations of measurement and simulation data, a single metric to characterize the simulation accuracy is applied:

$$MSE_{sim} \leq k_e^2 \cdot MSE_{meas}$$

where k_e is a scaling factor related to the desired confidence interval (e.g., $k_e = 2, 3,$ or 5 are, respectively, 95%, 99.7%, or 99.9994% confidence intervals).

Root Mean Square Error (RMSE) has typical physical units, so model results RMSE ($RMSE_{sim}$) versus allowed RMSE ($RMSE_{allowed}$) are bounded as:

$$RMSE_{sim} = \sqrt{MSE_{sim}} \leq RMSE_{allowed} = k_e \sqrt{MSE_{meas}}$$

Note: It can be shown that $|\bar{S} - \bar{M}| \leq \sqrt{MSE_{sim}}$,^{C-15} thus a bound on $RMSE_{sim}$ is a bound on $\bar{S} - \bar{M}$.

The Coefficient of Variation of Root Mean Square Error (CVRMSE) normalizes RMSE by some reference to get a dimensionless fraction. For steady-state test cases, the natural reference is \bar{M} , so allowed $CVRMSE_{sim}$ is bounded as:

$$CVRMSE_{sim} = \frac{RMSE_{sim}}{\bar{M}} \leq CVRMSE_{allowed} = k_e \frac{\sqrt{MSE_{meas}}}{\bar{M}}$$

Then $CVRMSE_{allowed}$ is a maximum bound as a dimensionless value, which is applied for developing the upper limits shown in Table A3-10 (see Annex A3, Section A3.1), as follows:

- For the most constrained cases, where constant combined convective and radiative interior surface heat transfer coefficients (based on empirical measurements) are specified:
 - $k_e = 2$
 - This applies to Cases ET110A1, ET110B1, ET100A1, and ET100B1
- For the less constrained cases where convective and radiative interior surface heat transfer algorithms are selected by the user:
 - $k_e = 5$
 - This applies to Cases ET100A3 and ET100B3.

B12.1.2 Reference Software for Setting Bound Limits for Software-to-Software Comparative Tests.

For software-to-software comparative tests, the reference set of software for setting the bound limits is selected based on the following criteria:

- A single set of results per calculation engine is included (multiple user interfaces for the same calculation engine are not included).
- Results are submitted by the calculation engine copyright holder(s) or by a modeler designated by the calculation engine copyright holder(s), applying a user interface specified by the calculation engine copyright holder(s) if appropriate.
- The set of selected software should represent a diversity of modeling methods, such as control-based and load-based.
- The calculation engine should be actively, reasonably, and widely used and actively maintained.
- Software satisfies the following:
 - Able to run at a time step of an hour or less
 - Able to include hourly variations in occupancy, lighting power, miscellaneous equipment power, thermostat set points, and HVAC system operation
 - Include thermal mass effects
- If a software did not participate in the simulation trials for the initial development of a test suite, then prior to the submittal, its results are reviewed and undergo iterative rounds of review and revisions as determined during the submission review process. The reviewer comments and submitter responses are documented from each round, with a logical basis for corrections to programs or inputs included in the documentation.

This review process only applies to the initial submittal of a software and not subsequent submittals of that software for version updates.

The results from a reference software

- include justification via Normative Annex A2, Attachment A2.9, Item G (“Report Block for Anomalous Results”) if they fall outside of the established bounds and
- pass the minimum number of tests indicated for each test group in Table A3-110 for which a given reference software provided results.

The results from a reference software may be updated for each new public release of the calculation engine. For future continuous maintenance revisions of Standard 140, SSPC 140 may

- include any additional software in the acceptance criteria-bounds reference set that apply and meet the criteria for inclusion;
- remove any software from the reference set that no longer meet the criteria for inclusion; and
- update the acceptance criteria with the most recent results from the reference set of software.

The current reference set of software used to determine the acceptance criteria are (in alphabetical order): CSE, DeST, DOE-2.2, EnergyPlus, ESP-r, IDA ICE, IES-VE, **Modelica Buildings Library**, TAS, and TRNSYS. Not all software are applied for all test groups.

B12.2 Acceptance Criteria Results Comparison. ~~Given that the upper and lower acceptance bound limits for the test cases are determined from the reference set of software, the results from other tested software can be compared against those the acceptance bound limits of Normative Annex A3. A software result is designated as an outlier for a given acceptance-criteria range case if that result is less than the lower limit or greater than the upper outside of the provided bound limits.~~

~~Tolerances will be used~~ are applied for comparisons to bounds.

- a. A relative tolerance of 1% (i.e., numbers within $\leq 1\%$ of each other are considered equivalent)
- b. An absolute tolerance of $1E-4$ (i.e., numbers within $1E-4$ of each other are considered equivalent even if that is more than 1%)

B12.3 Acceptance Criteria Tests Included. Not all test suites and test cases included in Standard 140 are appropriate to be used in acceptance criteria. The following criteria are applied in determining which test results are used in the acceptance criteria:

- For empirical validation tests (which have measurement-based criteria), all test cases are included.
 - The more-constrained cases have more restrictive acceptance bounds based on 95%-confidence measurement uncertainty.
 - The less-constrained cases have expanded acceptance bounds based on the ability of software that participated in the simulation trials for initial development of the test suite (that ran the less-constrained cases) to comply with the bounds, where such software also comply with the acceptance bounds of the more constrained cases.
- For software-to-software comparative tests, A_a test group has enough of the reference set of software that can adequately perform all of the tests within the test group to produce statistically meaningful acceptance ranges. (Beginning with Standard 140-2023, which first introduced the acceptance criteria, there are at least five reference-set software that can adequately perform the tests within each test group.)
- The tests are grouped to provide separate acceptance criteria for different sections of the standard.
- When available, acceptance-criteria tests include outputs such as annual total load, energy use, or other annual summations based on realistic, time-varying conditions.
- Differences between the results of two tests (i.e., result sensitivities) are preferred.
- A base case for each test suite or portion of the test suite should be included.

Based on these criteria, the following ~~five~~ test groups were selected for the acceptance criteria (selected outputs within each test group are also described, and Test Group 5 has separate selected outputs specific to its test cases):

1. Thermal Fabric Low Mass Test Group Range Cases
 - Annual heating load and annual sensible cooling load
 - Case 600
 - Case 610 – Case 600
 - Case 620 – Case 600
 - Case 630 – Case 620
 - Case 640 – Case 600

First Public Review Draft

Case 650 – Case 600 (cooling only)

Case 660 – Case 600

Case 670 – Case 600

Case 680 – Case 600

Case 685 – Case 600

Case 695 – Case 685

Daily hourly loads

Case 600

Case 640

Case 660

Case 670

Case 680

Case 685

Case 695

2. Thermal Fabric High Mass Test Group Range Cases

Annual heating load and annual sensible cooling load

Case 900

Case 900 – Case 910

Case 920 – Case 900

Case 930 – Case 920

Case 940 – Case 900

Case 950 – Case 900 (cooling only)

Case 960 – Case 900

Case 980 – Case 900

Case 985 – Case 900

Case 995 – Case 985

Daily hourly loads

Case 900

Case 940

Case 980

Case 985

Case 995

3. Cooling Equipment Test Group Range Cases

Annual total cooling energy consumption

CE300

CE310 – CE300

CE320 – CE300

CE330 – CE320

CE330 – CE300

CE340 – CE300

CE330 – CE340

CE350 – CE300

CE360 – CE300

CE400 – CE300

CE410 – CE300

CE420 – CE300

CE430 – CE300

CE440 – CE300

4. Heating Equipment Test Group Range Cases

Furnace input energy and cumulative fan energy

HE210

HE220 – HE210

HE230 – HE220

5. Air-side Equipment Test Group Range Cases

Heating-coil load

AE201

Preheat-coil load

AE301

AE401

AE301 – AE401

Cooling-coil total load and cooling-coil sensible load

AE205

AE245 – AE205

AE206

AE226 – AE206

AE305

AE305 – AE405

AE345 – AE305

AE345 – AE445

AE306

AE306 – AE406

AE326 – AE306

AE326 – AE426

AE405

AE445 – AE405

AE406

AE426 – AE406

Reheat-coil loads—zone 1 and zone 2

AE301

AE401

AE301 – AE401

AE305

AE345 – AE305

AE306

AE326 – AE306

AE405

AE406

6. Thermal Fabric Empirical Validation Test Group Range Cases

Hourly heating energy for designated time periods

ET110A1

ET110B1

ET100A1

ET100B1

ET100A3

ET100B3

B12.4 Acceptance Criteria Overall Passing. Because there may be legitimate physical modeling reasons (see Informative Annex B11, Section B11.1.4) why a software's results would fall outside of the acceptance range for an individual test or difference between two tests, it is not reasonable to expect a software to pass all acceptance criteria for every test in a test group.

To determine the number of range cases for which a software's results need to be within the acceptance ranges to pass a test group, the number of range cases in the test group is multiplied by 0.9, and the resulting number is rounded down to the nearest integer. For example, for the thermal fabric low mass test group, there are 34 range cases; $34 \times 0.9 = 30.6$; rounding down yields 30 range cases.

Add the following reference to Informative Annex C2.

[...]

C-15. G. Golub and C. Van Loan. (1989). "Matrix Computations." Second Edition. Johns Hopkins University Press. See Equation 2.2.