
SPECIAL NOTE

This American National Standard (ANS) is a national voluntary consensus standard developed under the auspices of the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). *Consensus* is defined by the American National Standards Institute (ANSI), of which ASHRAE is a member and which has approved this standard as an ANS, as "substantial agreement reached by directly and materially affected interest categories. This signifies the concurrence of more than a simple majority, but not necessarily unanimity. Consensus requires that all views and objections be considered, and that an effort be made toward their resolution." Compliance with this standard is voluntary until and unless a legal jurisdiction makes compliance mandatory through legislation.

ASHRAE obtains consensus through participation of its national and international members, associated societies, and public review.

ASHRAE Standards are prepared by a Project Committee appointed specifically for the purpose of writing the Standard. The Project Committee Chair and Vice-Chair must be members of ASHRAE; while other committee members may or may not be ASHRAE members, all must be technically qualified in the subject area of the Standard. Every effort is made to balance the concerned interests on all Project Committees.

The Manager of Standards of ASHRAE should be contacted for:
a. interpretation of the contents of this Standard,
b. participation in the next review of the Standard,
c. offering constructive criticism for improving the Standard, or

CONTENTS

ANSI/ASHRAE Standard 140-2011, Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs

CONTENTS *(Continued)*

CONTENTS *(Continued)*

NOTE

Approved addenda, errata, or interpretations for this standard can be downloaded free of charge from the ASHRAE Web site at www.ashrae.org/technology.

(This foreword is not part of the 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

This standard method of test (SMOT) can be used for identifying and diagnosing predictive differences from whole-building energy simulation software that may possibly be caused by algorithmic differences, modeling limitations, input differences, or coding errors. These tests are part of an overall validation methodology described in Informative Annex B23. The procedures test software over a broad range of parametric interactions and for a number of different output types, thus minimizing the concealment of algorithmic differences by compensating errors. Different building energy simulation programs, representing different degrees of modeling complexity, can be tested. However, some of the tests may be incompatible with some building energy simulation programs.

The tests are a subset of all the possible tests that could occur. A large amount of effort has gone into establishing a sequence of tests that examines many of the thermal models relevant to simulating the energy performance of a building and its mechanical equipment. However, because building energy simulation software operates in an immense parameter space, it is not practical to test every combination of parameters over every possible range of function.

The tests consist of a series of carefully described test case building plans and mechanical equipment specifications. Output values for the cases are compared and used in conjunction with diagnostic logic to determine the sources of predictive differences.

The test cases are divided into separate test classes to satisfy different levels of software modeling detail. Such classification allows more convenient citation of specific sections of Standard 140 by other codes and standards, and certifying and accrediting agencies, as appropriate. The Class I test cases (Section 5) are detailed diagnostic tests intended for simulation software capable of hourly or sub-hourly simulation time steps. The Class II (Section 7) test cases may be used for all types of building load calculation methods, regardless of time-step granularity, and are often favored by those needing to test simplified software for residential buildings. The Class I (Section 5) test cases are designed for more detailed diagnosis of simulation models than the Class II (Section 7) test cases.

Class I Test Procedures (Section 5)

The set of Class I tests included herein consist of

- *software-to-software comparative tests that focus on building thermal envelope and fabric loads and mechanical equipment performance and*
- *analytical verification tests (comparison of software to analytical or quasi-analytical solutions) that focus on mechanical equipment performance.*

In addition to comparative and analytical verification tests, the overall methodology for model validation and testing described in Informative Annex B23, 2009 ASHRAE Handbook—Fundamentals[1] *(see Chapter 19) and elsewhere*[2] *includes empirical validation testing, where tested software models are validated to within the uncertainty of measured data. Such tests will be considered for Standard 140, and additional research on this topic is recommended, as discussed in Informative Annex B23.*

The current set of Class 1 test cases were initially developed by the National Renewable Energy Laboratory (NREL) with the International Energy Agency (IEA)[3,4,5] *and by Natural Resources Canada, also in collaboration with IEA*[6]*.*

For the building thermal envelope and fabric load cases of Section 5.2, the "basic" cases (Sections 5.2.1 and 5.2.2) test the ability of the programs to model such combined effects as thermal mass, direct solar gain windows, window-shading devices, internally generated heat, infiltration, sunspaces, and deadband and setback thermostat control. The "in-depth" cases (Section 5.2.3) facilitate diagnosis by allowing excitation of specific heat transfer mechanisms. The space-cooling equipment cases of Section 5.3 test the ability of programs to model the performance of unitary space-cooling equipment using manufacturer design data presented as empirically derived performance maps. In the steady-state analytical verification cases of Sections 5.3.1 and 5.3.2, which utilize a typical range of performance data, the following parameters are varied: sensible internal gains, latent internal gains, zone thermostat setpoint (entering dry-bulb temperature), and outdoor dry-bulb temperature. Parametric variations isolate the effects of the parameters singly and in various combinations and isolate the influence of part-loading of equipment, varying sensible heat ratio, "dry" coil (no latent load) versus "wet" coil (with dehumidification) operation, and operation at typical Air-Conditioning, Heating, and Refrigeration Institute (AHRI) rating conditions. Quasi-analytical solution results are presented for the test cases in this section. The comparative test cases of Sections 5.3.3 and 5.3.4 utilize an expanded range of performance data, an outdoor air mixing system, and hourly varying weather data and internal gains. These cases cannot be solved analytically. In these cases, the following parameters are varied: sensible internal gains, latent internal gains, infiltration rate, outdoor air fraction, thermostat setpoints, and economizer control settings. Through analysis of results, the influence of part-loading of equipment, outdoor dry-bulb (ODB) temperature sensitivity, and "dry" coil (no latent load) versus "wet" coil (with dehumidification) operation can also be isolated. These cases help to scale the significance of simulation result disagreements in a realistic context, which is less obvious in the steady-state cases of Sections 5.3.1 and 5.3.2. The space-heating equipment cases of Section 5.4 test the ability of programs to model the performance of residential fuel-fired furnaces. These tests are divided into two tiers. The Tier 1 cases (Sections 5.4.1 and 5.4.2) employ simplified boundary conditions and test the *basic functionality of furnace models. More realistic boundary conditions are used in the Tier 2 cases (Section 5.4.3), where specific aspects of furnace models are examined. The full set of space-heating test cases is designed to test the implementation of specific algorithms for modeling the following aspects of furnace performance: furnace steady-state efficiency, furnace part-load ratio, furnace fuel consumption, circulating fan operation, and draft fan operation. These cases also test the effects of thermostat setback and undersized capacity.*

Class II Test Procedures (Section 7)

The Class II (Section 7) test cases were adapted from HERS BESTEST, developed by the National Renewable Energy Laboratory[7]*. This set of test cases formally codifies the Tier 1 and Tier 2 tests for certification of residential energy performance analysis tools, as described in the* 2006 Mortgage Industry National Home Energy Rating Systems Standards[8].

The Section 7 test cases are divided into Tier 1 and Tier 2 tests. The Tier 1 base building plan (Section 7.2.1) is a single-story house with 1539 ft^2 of floor area, with one conditioned zone (the main floor), an unconditioned attic, a raised floor exposed to air, and typical glazing and insulation. Additional Tier 1 cases (Section 7.2.2) test the ability of software to model building envelope loads in the base-case configuration with the following variations: infiltration; wall and ceiling R-values; glazing physical properties, area, and orientation; shading by a south overhang; internal loads; exterior surface color; energy inefficient building; raised floor exposed to air; uninsulated and insulated slabs-on-grade; and uninsulated and insulated basements. The Tier 2 tests (Section 7.2.3) consist of the following additional elements related to passive solar design: variation in mass, glazing orientation, east and west shading, glazing area, and south overhang. The Section 7 test cases were developed in a more realistic residential context and have a more complex base building construction than the Section 5 test cases (which have more idealized and simplified construction for enhancement of diagnostic capability). To help avoid user input errors for the Section 7 test cases, the input for the test cases is simple, while remaining as close as possible to "typical" residential constructions and thermal and physical properties. Typical building descriptions and physical properties published by sources such as the National Association of Home Builders, the U.S. Department of Energy, American Society of Heating, Refrigerating and Air Conditioning Engineers, and the National Fenestration Rating Council are used for the Section 7 test cases.

Comparing Tested Results

The tests have a variety of uses, including

a. *comparing the predictions from other building energy programs to the example results provided in Informative Annexes B8 and B16 for Class I tests, Informative Annex B20 for Class II tests, and/or to other results that were generated using this SMOT;*
b. *checking a program against a previous version of itself after internal code modifications to ensure that only the intended changes actually resulted;*
c. *checking a program against itself after a single algorithmic change to understand the sensitivity between algorithms; and*
d. *diagnosing the algorithmic sources and other sources of prediction differences (diagnostic logic flow diagrams are included in Informative Annex B9).*

Regarding the comparative test results of Annex B8, selected parts of Annex B16, and Annex B20, the building energy simulation computer programs used to generate these results have been subjected to a number of analytical verification, empirical validation, and comparative testing studies. However, there is no such thing as a completely validated building energy simulation computer program. All building models are simplifications of reality. The philosophy here is to generate a range of results from several programs that are generally accepted as representing the state of the art in whole-building energy simulation programs. To the extent possible, input errors or differences have been eliminated from the presented results. Thus, for a given case, the range of differences between comparative test results presented in Informative Annexes B8, B16, and B20 represents legitimate algorithmic differences among these computer programs. For any given case, a tested program may fall outside this range without necessarily being incorrect. However, it is worthwhile to investigate the sources of substantial differences, as the collective experience of the authors of this standard is that such differences often indicate problems with the software or its usage, including, but not limited to

- *user input error, where the user misinterpreted or incorrectly entered one or more program inputs;*
- *a problem with a particular algorithm in the program; or*
- *one or more program algorithms used outside their intended range.*

Also, for any given case, a program that yields values in the middle of the range established by the comparative test example results should not be perceived as better or worse than a program that yields values at the borders of the range.

Informative (non-mandatory) Annex B22 provides an example procedure for establishing acceptance range criteria to assess annual or seasonal heating and cooling load results for software undergoing the Class II tests contained in Section 7. Inclusion of this example is intended to be illustrative only and does not imply in any way that results from software tests are required by Standard 140 to be within any specific limits. However, certifying or accrediting agencies using Section 7 may wish to adopt procedures for developing acceptance-range criteria for tested software. Informative Annex B22 presents an example range setting methodology that may be useful for these purposes.

Importance of Analytical and Quasi-Analytical Solution Results

Analytical verification test results for the Class I HVAC equipment performance tests include both quasi-analytical solutions and simulation results in selected sections of Informative Annex B16. In general, it is difficult to develop *worthwhile test cases that can be solved analytically or quasi-analytically, but such solutions are extremely useful when possible. Analytical or quasi-analytical solutions represent a "mathematical truth standard." That is, given the underlying physical assumptions in the case definitions, there is a mathematically correct solution for each case. In this context, the underlying physical assumptions regarding the mechanical equipment as defined in Sections 5.3 and 5.4 are representative of typical manufacturer data normally used by building design practitioners. Many whole-building simulation programs are designed to work with this type of data. It is important to understand the difference between a "mathematical truth standard" and an "absolute truth standard." In the former, we only test the solution process for a model, not the appropriateness of the model itself; that is, we accept the given underlying physical assumptions while recognizing that these assumptions represent a simplification of physical reality. An "approximate truth standard" from an experiment tests both the solution process and the appropriateness of the model within experimental uncertainty. The ultimate or "absolute" validation truth standard would be comparison of simulation results with a perfectly performed empirical experiment, with all simulation inputs perfectly defined.*

The quasi-analytical and analytical solution results presented in selected parts of Annex B16 represent a mathematical truth standard. This allows identification of bugs in the software that would not otherwise be apparent from comparing software only to other software and therefore improves the diagnostic capabilities of the test procedure. The primary purpose of also including simulation results for the cases where analytical or quasi-analytical solutions exist is to allow simulationists to compare their relative agreement (or disagreement) versus the analytical or quasi-analytical solution results to that for other simulation results. Perfect agreement among simulations and analytical or quasi-analytical solutions is not necessarily expected. The results give an indication of the degree of agreement that is possible between simulation results and the analytical or quasi-analytical solution results. Because the physical assumptions of a simulation may be different from those for analytical or quasi-analytical solutions, a tested program may disagree with such solutions without necessarily being incorrect. However, it is worthwhile to investigate the sources of differences as noted previously.