The BESTEST Method for Evaluating and Diagnosing Building Energy Software

Ron Judkoff
National Renewable Energy Laboratory, Golden, CO
and
Joel Neymark
J. Neymark & Associates, Golden, CO

ABSTRACT

The increasing power of personal computers has encouraged a proliferation of building energy simulation software packages. The prospective user has no way to quantitatively judge the appropriateness of a given software package for a given design problem. To address this issue, the Building Energy Simulation Test (BESTEST) procedure was developed for systematically testing whole building energy simulation programs and diagnosing sources of predictive disagreement. The BESTEST procedure takes a "comparative testing" approach where a program is compared to itself or to other programs. It focuses on testing a software package's ability to model thermal processes associated with the building envelope. Field trials of its approach were conducted with a number of detailed state-of-the-art programs by researchers from nations participating in International Energy Agency (IEA) Solar Heating and Cooling (SHC) Programme Task 12 and Annex 21. The approach consists of a number of carefully specified test case buildings that progress systematically from extremely simple to relatively realistic. The outputs from tested programs are evaluated according to diagnostic logic to determine the algorithms responsible for predictive differences. The procedure has proven very effective at revealing bugs, faulty algorithms, and input errors in a variety of detailed and simplified software for commercial and residential buildings. Since it was first published in 1995, BESTEST has been used and/or adopted by a number of nations, states, certifying bodies, and universities. Recently BESTEST was augmented to better accommodate simplified software and software for buildings in hot humid climates. Current work conducted in collaboration with IEA Solar Heating and Cooling Programme Task 22, and the American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc. (ASHRAE) Standards Project Committee 140 includes development of tests for heating, ventilation, and air-conditioning (HVAC) algorithms. This paper describes the BESTEST procedure, recent additions to the procedure, and results obtained by those who have used it to evaluate software.

Introduction

The increasing power and attractive pricing of personal computers has created a proliferation of building energy simulation software packages. An on-line directory sponsored by the U.S. Department of Energy (DOE) lists more than 120 software tools developed worldwide that have thousands of users (DOE 1998). There is little if any objective quality control of this software. An evaluation of a number of design tools conducted by nations participating in IEA's Solar Heating and Cooling (SHC) Programme Task 8 showed large unexplained predictive differences between these tools, even when run by experts

(Rittelmann, Ahmed 1985b). However, it is important that the design industry not become disillusioned with simulation tools because of the great potential for energy savings and comfort improvements through their use.

The effort begun under IEA SHC Programme Task 8 was continued in SHC Task 12 and Annex 21 to develop a quantitative procedure for evaluating and diagnosing building energy simulation software packages (Judkoff et al. 1983, 1988 and 1989; Bloomfield 1989). The effort produced the International Energy Agency Building Energy Simulation Test and Diagnostic Method (IEA BESTEST). This paper summarizes BESTEST and its "comparative testing" approach. Recent additions to BESTEST and results obtained from its users are also discussed.

Background

An overall validation methodology consists of three parts:

- Analytical Verification—compares the output from a program, subroutine, or algorithm to the result from a known analytical or generally accepted numerical solution for isolated heat transfer mechanisms under very simple boundary conditions.
- Empirical Validation—compares calculated results from a program, subroutine, or algorithm to monitored data from a real structure, test cell, or laboratory experiment.
- Comparative Testing—compares a program to itself or to other programs that may be considered better validated or more detailed and presumably, more physically correct. The comparative approach includes "sensitivity testing" and "intermodel comparisons."

Each of these approaches has different strengths and weaknesses (Judkoff 1988). The procedures presented here take the comparative testing approach. A range of results from a number of detailed public domain models, considered state-of-the-art in the United States and Europe, is provided as the basis for comparison. These reference model results do not necessarily represent "truth;" however, they are representative of what is commonly accepted as state-of-the-art in whole-building energy simulation.

The National Renewable Energy Laboratory (NREL) has been developing a class of procedures for systematically testing whole-building energy simulation programs and diagnosing the sources of predictive disagreement. These procedures and their capabilities are summarized in Table 1.

IEA BESTEST, the first of these procedures, was developed in conjunction with IEA SHC Programme Task 12b/21c. It is designed to test a program's ability to model the building envelope and provides detailed diagnosis of sources of disagreement among programs. ASHRAE Standard 140P is a proposed ASHRAE Standard Method of Test for building energy simulation software that is based on IEA BESTEST. This will be the first standard method of test for software issued by ASHRAE. Home Energy Rating Systems (HERS) BESTEST was designed to test simplified tools such as those currently used for home energy rating systems. Currently it tests just the ability to model the building envelope. Although, HERS BESTEST has a more realistic base building than IEA BESTEST, it's ability to diagnose sources of differences among results is not as detailed. Additional discussion comparing IEA BESTEST and HERS BESTEST is included in a previous paper (Neymark, Judkoff 1997b). Florida-HERS BESTEST is based on HERS BESTEST but is revised for the hot and humid climate of Orlando, Florida. Finally, HVAC BESTEST is a new procedure being developed to test a program's ability to model mechanical equipment as part of IEA SHC Programme Task 22.

Table 1. Capabilities of the BESTEST Procedures

Comparative Test Procedure	Algorithms Tested Building Envelope	 Mechanical Equipment	Appropriate SoftwareType Detailed	 Simplified
IEA BESTEST (Judkoff, Neymark 1995a)	yes	limited*	yes	limited**
ASHRAE Standard 140 P (ASHRAE, 1998)	yes	limited*	yes	limited**
HERS BESTEST (Judkoff, Neymark 1995b)	yes	no	sometimes***	yes
Florida - HERS BESTEST (Judkoff, Neymark 1997)	yes	no	sometimes***	yes
HVAC BESTEST (Neymark, Judkoff 1997a)	no	yes	yes	maybe

*Thermostat settings and mechanical ventilation only
**Some of the tests may not be possible to perform with some simplified software.
***See Neymark and Judkoff 1997b.

NREL developed the theoretical basis for building energy software comparative testing (Judkoff et al. 1983) implemented by IEA BESTEST (Judkoff and Neymark 1995a). For that work, NREL led a group of experts from the IEA SHC Task Programme 12b and the IEA Buildings and Community Systems Programme Task 21c. The 5-year international research effort included the participants listed below. The participants' country and software they used are listed in parenthesis.

- Building Research Establishment (United Kingdom, SERIRES 1.2)
- De Montfort University (United Kingdom, ESP-RV8)
- Electricite de France (France, CLIM2000)
- Lund Institute of Technology (Sweden, DEROB-LTH)
- National Renewable Energy Laboratory (United States, BLAST 3.0, DOE-2.1D, SERIRES/ SUNCODE 5.7
- Politecnico Torino (Italy, BLAST 3.0)
- Tampere University (Finland, TASE)
- University of Sevilla (Spain, S3PAS)
- Vrije Univeriteit (Belgium, TRNSYS 13.1)

Important conclusions of the IEA BESTEST effort include:

- The BESTEST method trapped bugs and faulty algorithms in every program tested.
- The IEA Task 12b/21c experts unanimously recommend that no energy simulation program be used until it is "BESTESTed."
- BESTEST is an economic means of testing, in several days, software that has taken many years to develop.
- Even the most advanced whole-building energy models show a significant range of disagreement in the calculation of basic building physics.
- Improved modeling of building physics is as important as improved user interfaces.

The list of BESTEST users continues to grow and NREL has received requests for and mailed out several hundred copies of test procedures. Some of the organizations that use BESTEST are:

- ASHRAE (Standards Project Committee - SPC140)
- Australian Nationwide House Energy Rating Scheme
- Building Research Establishment (United Kingdom)
- Canada Department of Natural Resources
- California Energy Commission
- Carrier Corporation
- Electric Power Research Institute
- Electricite de France
- Florida Solar Energy Center and Florida Energy Office
- Iowa State Department of Natural Resources, Energy Bureau
- Lawrence Berkeley National Laboratory
- National HERS Program (United States, Cited by NOPR 10 CFR 437 and HERS Council Guidelines)
- National Weatherization Program (United States)
- The Netherlands (as part of a national software validation program)
- New Zealand (for certifying software used for energy code compliance)
- Owens Corning Corporation
- Pacific Northwest National Laboratory
- TNO Building and Construction Research (The Netherlands)
- University of Strathclyde, Energy Systems Division (United Kingdom)

IEA BESTEST

Specification of Test Cases

IEA BESTEST has 40 comparative test cases that are organized into a "qualification" series and two "diagnostic" series. The series of buildings proceed from the thermally simple to the realistic approximately one parameter at a time. The cases are defined so that thermal properties, geometric proportions, and thermal responses are meaningful in terms of actual envelope load dominated buildings. The specification is a compromise between U.S. and European construction. **Figure 1** shows the basic building geometry which remains similar for all cases with minimal changes to allow investigation of sensitivity to various changes.

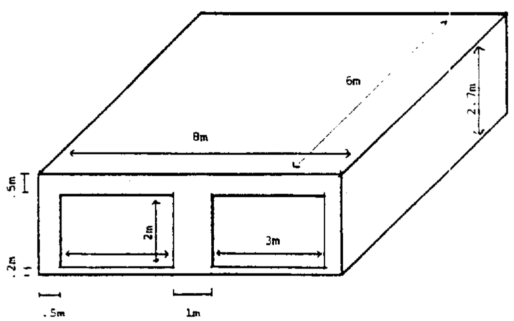

Figure 1. IEA BESTEST Base Building with Unshaded South Facing Windows

The qualification series is relatively realistic. It tests such features as thermal mass, direct gain windows, window shading, window orientation, sunspaces, ground coupling, night ventilation, and dead-band and set-back thermostat control. The overall heat transmission coefficient was equivalent in both the heavy and lightweight cases so that the sensitivity to thermal mass could be assessed under a number of parametric variations. Cases that required simulation of mechanical systems were not defined. The equipment was assumed 100% efficient and adequately sized to meet peak loads.

The first of the two diagnostic series (A-series) is designed to provide excitation of a particular heat transfer mechanism while suppressing signals from other mechanisms. These diagnostics minimize interacting effects. The second diagnostic series (B-series) is required because not all programs can model the A-series diagnostic cases. The B-series diagnostics have more interacting effects but can be used by more programs.

Input Equivalency. The test cases were specified so that equivalent input files could be defined for a variety of detailed and simplified whole-building energy programs. To minimize interpretive problems, input information was provided at several levels of physical detail. All participants were instructed to model the test cases at the most detailed level their programs would allow. To minimize input errors, input files for each program were independently developed by two experienced modelers or developed by one and checked by another.

Diagnostic Results

Simulations were performed for each test case with the participating computer programs using annual hourly Typical Meteorological Year (TMY) weather data from Denver, Colorado. The Denver climate is characterized by cold clear winters (3600 °Kelvin-degree-days), hot dry summers, and large diurnal temperature variations throughout the year. At each stage of the exercise, output data from the simulations were compared to each other according to the diagnostic logic of the test cases. A number of bugs, faulty algorithms, and input errors were uncovered, isolated, and corrected via this process. Several examples follow. The accompanying figures, illustrating IEA BESTEST comparative results (**Figures 2-5**), are included to show how problems are detected. For a more complete explanation of the contents of these figures, see IEA BESTEST (Judkoff and Neymark 1995a).

TRNSYS. Early in the study, Belgium implemented the TRNSYS 12.2v1 program (Klein 1990). TRNSYS is considered the most advanced program for simulation of active solar systems sponsored by DOE. Belgium's results, as shown in **Figure 2**, disagreed markedly from those of the other programs for many of the qualification cases involving high thermal capacitance. By following the diagnostic flow logic provided with IEA BESTEST, researchers confirmed that the algorithm causing the problem was related to the calculation of the thermal mass effect. They were then able to trace the problem to the TRNSYS "BID" module that contains the transfer function coefficients (Mitalas and Arsenault 1971). Inspection of the module revealed that two sets of coefficients were transposed. Rearranging the coefficients eliminated the discrepancies (**Figure 3**). This problem would neither have been detected in any of the lightweight cases, nor in case 930 (high mass with east and west facing windows, **Figure 2**) where compensating errors between the handling of thermal mass and the shading of east and west windows conceals the problem. The power of IEA BESTEST is that a program is tested over a broad range of parametric interactions. This problem was corrected in all subsequent versions of TRNSYS.

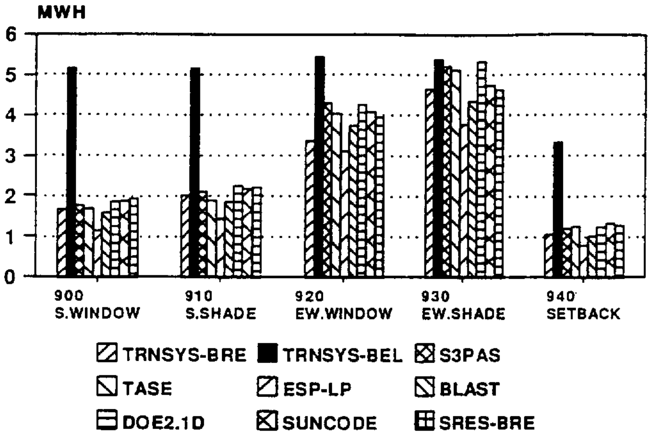

Figure 2. IEA BESTEST 900-Series High Mass Annual Heating Loads Indicating Discovered TRNSYS Disagreement

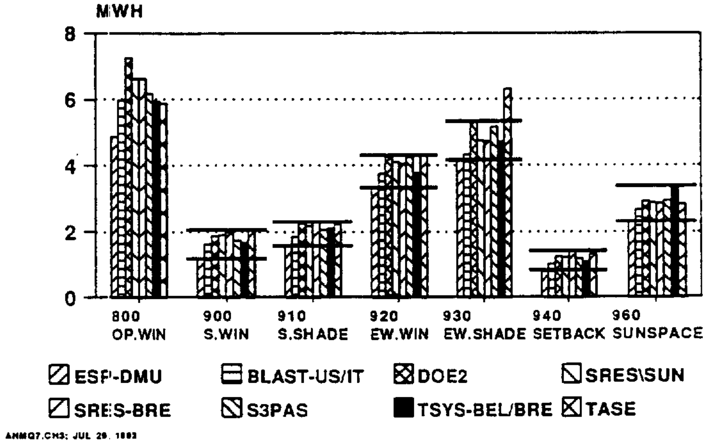

Figure 3. IEA BESTEST 900-Series High Mass Annual Heating Loads

DOE-2. The DOE-2 program is considered to be the most advanced program sponsored by the U.S. Department of Energy. Diagnostic tests revealed a problem in the DOE-2.1D014 computer program concerning treatment of solar absorptivity on exterior surfaces (DOE-2 Manuals 1981, 1989). This is illustrated in **Figure 4** using Cases 250 and 220 (A-series diagnostics with exterior solar absorptance of 0.9 and 0.1 respectively). As shown for Case 250 and for the difference between Case 250 and Case 220 ("250-220"), the annual cooling load output from DOE-2 appears less sensitive to a change in exterior solar absorptivity than the other programs. This was traced to a bug in the solar absorptance algorithm associated with surfaces defined as doors. The bug has been repaired in all subsequent versions of DOE-2 and now yields results comparable to the other programs (Winkelmann 1991-1993; Hirsch 1992-1997).

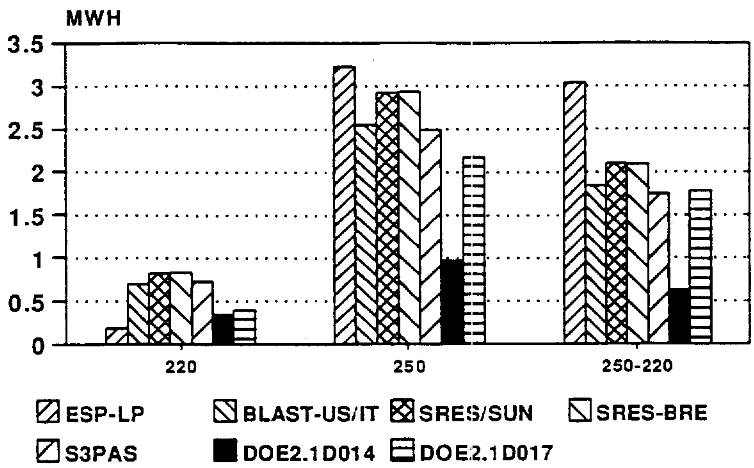

Figure 4. IEA BESTEST Exterior Shortwave Absorptance Effect (Cases 220,250) Annual Cooling Loads Indicating Discovered DOE-2.1D Disagreement

ESP. ESP (ESRU, 1993) is the program used by the research arm of the Council of European Communities (CEC) as a reference program for building energy research. ESP predicted relatively low annual heating loads for the qualification test cases. The source of the difference was isolated using Cases 270 and 280 (A-series diagnostics with interior solar absorptances of 0.1 and 0.9 respectively). Sensitivity results for the cases in **Figure 5** revealed a problem with interior solar absorptance or cavity albedo. The ESP development group at Strathclyde, United Kingdom was then able to locate and correct the responsible algorithm.

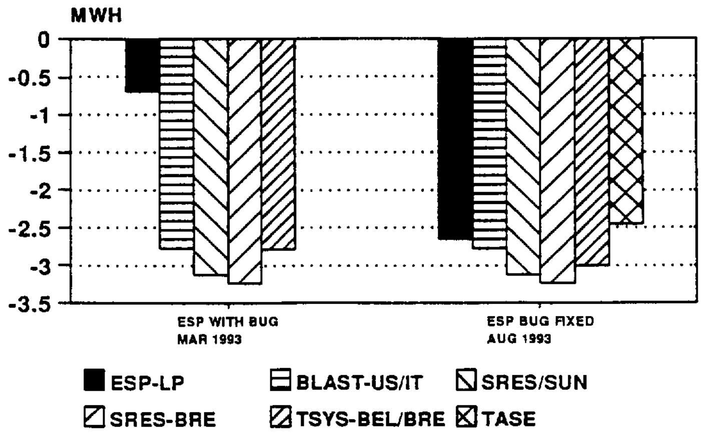

Figure 5. IEA BESTEST Interior Shortwave Absorptance Effect (Sensitivity Results 280-270) Annual Cooling Loads Indicating Discovered ESP-r Disagreement

Overall Results. It is only possible to present a brief example of the results from the full length IEA BESTEST report (Judkoff and Neymark 1995a). **Figure 3** shows an example of the annual heating load results from the final set of test case simulations performed with a group of programs that may be considered state-of-the-art in whole-building energy modelling as of 1993. Horizontal bands delimiting reference ranges are superimposed on the results to compare models with these results. Discussion of how bandwidths were set is included in the full report. The results show varying amounts of disagreement among the programs for the different cases and output types. The reference ranges reflect this disagreement. A more detailed discussion regarding the ranges of disagreement is included in a previous paper (Judkoff and Neymark 1995c).

There is no truth standard in this type of exercise. For any given case, a program that yields values in the middle of the range should not be perceived as better or worse than a program that yields values at the borders of the range. The ranges represent algorithmic differences in state-of-the-art programs. Programs that fall outside the range are producing results different from the state-of-the-art as defined by our group of international experts. Investigating the source(s) of the difference(s) is worthwhile, but the existence of a difference does not necessarily mean a program is faulty. Collective experience in this task has indicated that when programs show major disagreement with a range, there is often find a bug, a questionable algorithm, or an input error caused by faulty or ambiguous documentation.

HERS BESTEST

HERS BESTEST (Judkoff and Neymark 1995b) is a comparative test method for evaluating the credibility of more simplified building energy software such as the programs used by home energy rating systems. The method provides the technical foundation for "certification of the technical accuracy of building energy analysis tools used to determine energy efficiency ratings" as called for in the U.S. Energy Policy Act of 1992 (Title I, Subtitle A, Section 102, Title II, Part 6, Section 271). The following set of test cases represents the Tier 1 and Tier 2 Tests for Certification of Rating Tools as described in DOE 10 CFR Part 437 and the HERS Council *Guidelines for Uniformity* (HERS Council). Although HERS BESTEST was not developed in conjunction with an international collaboration, it is beginning to be used outside the United States.

Specification of the Test Cases

The HERS BESTEST base building is a 1539-ft² (143.0-m²) single-story house with one conditioned zone (the main floor), an unconditioned attic, and a vented crawl space. **Figure 6** shows the basic building geometry. The geometry remains similar for all cases with minimal changes to allow the investigation of sensitivity to the features noted below. Key thermal and physical properties of the base building are based on U.S. sources that include:

- American Society of Heating Refrigerating and Air Conditioning Engineers (*ASHRAE Handbook Fundamentals*)
- Lawrence Berkeley National Laboratory (LBNL)
- National Association of Home Builders (NAHB)
- National Fenestration Rating Council (NFRC)
- U.S. Department of Energy (*Housing Characteristics, 1990*).

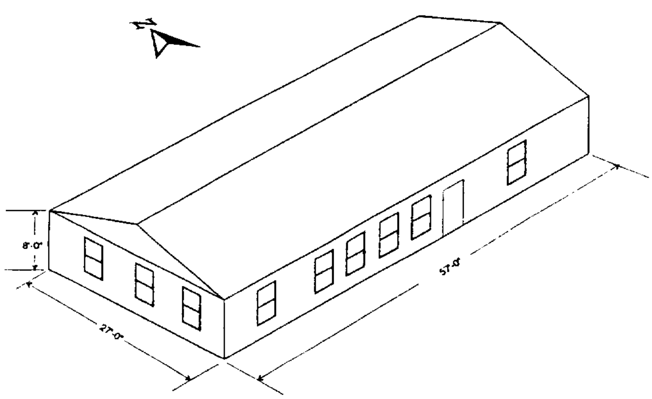

Figure 6. HERS BESTEST Base Building Axonometric

Separate U.S. weather sites, one in Colorado Springs, Colorado, and one in Las Vegas, Nevada, are used for heating and cooling loads respectively. Thermostat settings are for either heating only or cooling only in the respective climates. Avoiding a deadband thermostat makes the test better for non-hourly simulation tools.

The Tier 1 tests consist of a basic house with typical glazing and insulation. Specific cases are designed to test a building energy computer program with respect to the following components of heat and mass transfer: infiltration; wall and ceiling R-Values; glazing physical properties, area, and orientation; south overhang; internal gains; exterior surface color; energy-inefficient building; crawl space; uninsulated and insulated slab; and uninsulated and insulated basement. The Tier 2 tests consist of the following additional elements related to passive solar design: direct gain, passive solar home, variation in mass, glazing orientation, east and west shading, glazing area, and south overhang. A third tier of tests, including mechanical equipment, active and passive solar, and utility rate structure tests, are planned for future development.

Example Results

Comparative testing as applied in the HERS BESTEST method includes a set of results from public domain reference programs that have already been subjected to extensive analytical, empirical, and intermodel testing. The BLAST 3.0 Level 215, DOE-2.1E-W54, and SERIRES/SUNCODE 5.7 programs were used to generate reference results.

Figure 7 shows an example graph of maximum and minimum results for annual heating loads. A full description of example results is included in HERS BESTEST (Judkoff and Neymark 1995b).

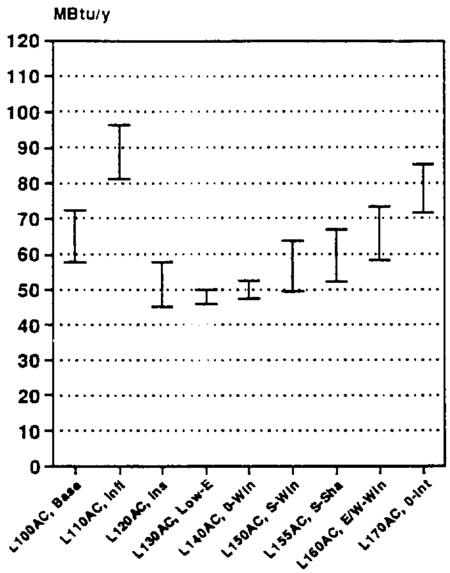

Figure 7. HERS BESTEST Tier 1 Reference Results Annual Heating Loads

HERS BESTESTing of HOT2000™ by Natural Resources Canada

Natural Resources Canada (NRCan) has documented their running of HOT2000™ through HERS BESTEST (Haltrecht and Fraser 1997). As a result of their work with HERS BESTEST, NRCan revised their modeling of internal gains that caused HOT2000™ to underpredict cooling loads. The effect of this change, noted in **Table 2** for Case L170 (like L100 base case but with zero internal gains), brings annual cooling load results up closer to the low end of the range of the HERS BESTEST example results. Also, NRCan had some large disagreements for initial results that were caused by a user input error. This resulted in adding an automated input check and a clarification to the HOT2000™ user manual (Haltrecht, 1998). The effect of eliminating this user input error is shown in **Table 2** for Case L100 (base case).

Using results generated with revised software, NRCan reached the following conclusions regarding the performance of HOT2000™ (Haltrecht and Fraser 1997):

- Results for all but the "highly passive solar" cases were acceptable

- Examine their model for solar utilization and improve it if necessary
- Rerun HOT2000™ through IEA BESTEST to help indicate sources of current discrepancies
- Compare results from the latest version of HOT2000™ with data from monitored houses

Table 2. Effect of Improvements to HOT 2000™ by NRCan

HOT2000™ Revision:	HERS BESTEST Results Case	Annual Cooling Loads (Mbtu/y) BLAST	DOE-2	SERIRES	HOT2000™ Before fix	After fix
Internal Gains Model Correction	**L170** (0 Internal Gains)	45.83	49.06	49.31	38.36	43.32
Thermostat Input Check and Documentation Fix	**L100** (Base Building)	54.66	60.80	59.32	46.29	53.01

Additionally, HERS BESTEST is now part of the standard procedure for testing revisions to HOT2000™ (Haltrecht 1998).

NRCan also had the following recommendations for improvements to HERS BESTEST.

- Basement ground coupling reference results should be expanded to include more accurate methods provided by recent research, and better foundation models should be incorporated into the software used to develop the reference results.
- Cases should be incorporated that test actual infiltration models rather than simply varying a given rate of exchange.
- It is not necessary to use the most detailed model (as is recommended in HERS BESTEST) to achieve a good test.

NRCan also noted that an important test of accuracy is versus empirical data, and perhaps it would be worthwhile to have an empirical form of BESTEST (Haltrecht and Fraser 1997).

Florida-HERS BESTEST

In collaboration with the State of Florida, NREL developed a climate-specific version of HERS BESTEST for the state entitled *Home Energy Rating System Building Energy Simulation Test for Florida* (Judkoff and Neymark 1997). The cases for Florida-HERS BESTEST were adopted from HERS BESTEST and applied for Orlando, Florida TMY weather data. Minor changes were made to portions of the original HERS BESTEST User Manual to create a climate-specific version including:

- Local average windspeed
- Altitude and related air density and infiltration rates for softwares that do not automatically correct for altitude (previous data were listed for higher altitude sites)
- ASHRAE slab-on-grade loss coefficients
- Solar distribution fractions for Cases P100, P105, and P110 (the high mass passive solar cases with respectively: unshaded south glass, south overhang, and reduced mass).

There were no changes to building geometry or construction materials from those used in the original HERS BESTEST. However, variation of exterior surface coefficients resulted in minor changes to equivalent soil thicknesses corresponding to below grade wall and floor components used for more detailed ground coupling simulations.

Florida-HERS BESTESTing of EnGauge 2.0 by Florida Solar Energy Center. The Florida Solar Energy Center (FSEC) used Florida HERS BESTEST to test their correlation-method-based EnGauge 2.0 software (Fairey et al. 1998). FSEC's EnGauge 2.0 generally agrees with the BESTEST example results. The only disagreement noted was for Case 200 (energy-inefficient construction) where EnGauge heating load results (not shown) are slightly outside the example-pass/fail range that FSEC generated using example range expansion algorithms developed by NREL (see Appendix H of HERS BESTEST or Florida-HERS BESTEST, Judkoff and Neymark, 1995 and 1997).

In the process of upgrading EnGauge, FSEC modified the modeling of glazing systems, roof overhangs, and raised wood floors. FSEC provides an example (see **Figure 8**) of a specific change to EnGauge related to the raised-wood-floor model (Fairey 1998). The data in **Figure 8** apply floor "point multiplier" data that were generated over 15 years ago for Florida's energy code. Based on their BESTEST work, FSEC revised the raised-wood-floor "point multipliers" using detailed hourly simulation models and incorporated these multipliers into EnGauge. **Figure 9** indicates results using the new multipliers, which now give much better agreement with the Florida-BESTEST example results. Note in **Figures 8 and 9** that "Max, pass" and "Min, pass" incorporate the above mentioned example-pass/fail range setting criteria.

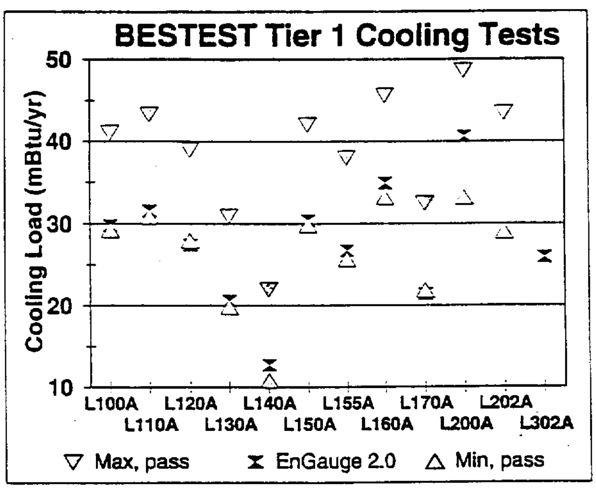

Figure 8. Comparison of Cooling Loads for EnGauge 2.0 versus HERS BESTEST Results Using Example Range Setting Criteria for Obsolete Raised-Wood-Floor Coefficients by FSEC.

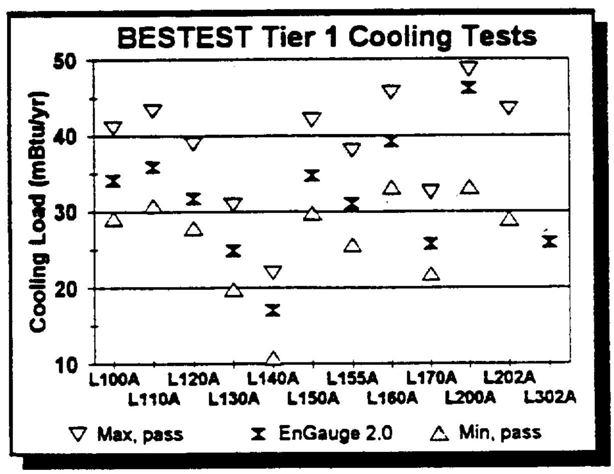

Figure 9. Comparison of Cooling Loads for EnGauge 2.0 versus HERS BESTEST Results Using Example Range Setting Criteria for Revised Raised-Wood-Floor Coefficients by FSEC.

As a result of this work, FSEC proposed the following improvements for Florida's EnGauge software:

- Reconsider development of load coefficients for exterior surface color impacts
- Re-evaluate wall and raised-wood-floor point multipliers

- Conduct detailed analysis of Florida's slab-on-grade floor load coefficients.

Additionally, FSEC proposes that changes to Florida's 1993 Energy Code and Rating System be expedited because the older procedures it specifies cannot pass Florida-HERS BESTEST.

Regarding suggestions for improvements to BESTEST, FSEC noted that because 95% of Florida home construction is slab on grade, NREL should develop a separate BESTEST test case for summer cooling-load impacts for slab-on-grade floor construction. Additionally, NREL should add separate test cases for uninsulated walls and uninsulated floors (these are currently lumped as a single energy-inefficient case) and raise the winter thermostat set point assumption for Florida to 72°F (22°C).

HVAC BESTEST

NREL has completed a preliminary draft of fundamental test cases for comparative testing of mechanical equipment. This test, the International Energy Agency Building Energy Simulation Test and Diagnostic Method for Mechanical Equipment (HVAC BESTEST), is being field tested as part of IEA SHC Programme Task 22 (Neymark and Judkoff 1997a). The first round of tests is for residential and small commercial cooling equipment commonly used in the United States and Europe. For the fundamental test cases, this equipment is applied to highly controlled load and ambient conditions using a near-adiabatic (very well insulated) test cell with specified sensible and latent internal gains along with artificial weather data.

There are 14 cases that represent a set of fundamental mechanical equipment tests. These cases test a program's ability to model mechanical equipment steady-state performance under controlled load and weather conditions by employing artificial whether data. The configuration of the base case building is a near-adiabatic rectangular single zone with only user specified internal gains to drive cooling loads. The geometric and materials specifications are purposely kept as simple as possible to minimize the opportunity for input errors on the part of the user. Mechanical equipment specifications represent a simple unitary vapor-compression cooling system, or more precisely, a split system, air-cooled condensing unit with indoor evaporator coil. Currently, only four parameters are varied to develop the remaining cases: internal sensible gains, internal latent gains, indoor dry-bulb temperature, and outdoor dry-bulb temperature. Systematic variation of these parameters also allows for the testing of the effect of varying part load ratio and sensible heat ratio. Additionally, there is a performance test at Air-Conditioning and Refrigeration Institute (ARI) equipment rating conditions.

Preliminary Results

Preliminary results of the first HVAC BESTEST field trial have been submitted by the following participants (country and software in parenthesis):

- CIEMAT (Spain, DOE-2.1E)
- Electricite de France (France, CLIM2000)
- Klimasystemtechnik (Germany, PROMETHEUS)
- National Renewable Energy Laboratory (United States, DOE-2.1E)
- Technische Universitat, Dresden (Germany, TRNSYS 14.2).

Additional results are expected as the task proceeds.

The preliminary test matrix with coefficient of performance (COP) results for the above software are included in **Table 3**. These are included primarily to illustrate: the diversity of the cases; the large change in COP that is caused by varying the equipment operating parameters; and areas where problems may be found. It is important to emphasize that these results are preliminary and may include user input errors. Additionally, the test specification requires some changes that could affect the results.

Table 3. HVAC BESTEST Case Descriptions and Preliminary COP Results

	Zone			Weather	Preliminary COP				
	Internal Gains*		Setpoint						
Case #	Sensible (W)	Latent (W)	IDB (°C)	ODB (°C)	IEA1	IEA2	IEA3	IEA4	IEA5
dry zone series									
E100	5400	0	22.2	46.1	2.55	2.39	2.24	2.72	2.40
E110	5400	0	22.2	29.4	3.47	3.56	3.05	3.95	3.44
E120	5400	0	26.7	29.4	3.65	3.71	3.53	3.97	3.67
E130	270	0	22.2	46.1	2.22	1.45	1.19	2.74	2.08
E140	270	0	22.2	29.4	3.09	1.98	1.53	3.97	3.04
humid zone series									
E150	5400	1100	22.2	29.4	3.70	3.69	3.34	3.95	3.64
E160	5400	1100	26.7	29.4	3.92	3.91	3.57	3.97	3.87
E165	5400	1100	23.3	40.6	2.95	2.86	2.83	3.12	2.90
E170	2100	1100	22.2	29.4	3.56	3.73	3.42	3.97	3.54
E180	2100	4400	22.2	29.4	4.03	4.28	3.86	3.96	4.10
E185	2100	4400	22.2	46.1	2.80	2.88	2.60	3.96	2.84
E190	270	550	22.2	29.4	3.62	3.24	2.78	3.99	3.69
E195	270	550	22.2	46.1	2.44	2.26	1.96	2.75	2.49
full load test at ARI conditions									
E200	6036	1817	26.7	35.0	3.61	3.57	3.55	3.60	3.56

Abbreviations: IDB = indoor dry-bulb temp; ODB = outdoor dry-bulb temp
COP = coefficient of performance
ARI = Air Conditioning and Refrigeration Institute
*Internal Gains are internally generated sources of heat and humidity that are not related to operation of the mechanical cooling system or its air distribution fan.

Even in its early stages, HVAC BESTEST has uncovered problems in software or its documentation. For example, in 1994, some initial work with DOE-2.1D RESYS2 system yielded some unexpectedly high COPs that were eventually traced to a bug that was corrected. More recent work with DOE-2 uncovered a documentation ambiguity regarding the meaning of one of the DOE-2 keywords for modeling COP degradation caused by equipment cycling (Hirsch 1992-1997).

Conclusion

The BESTEST procedures, developed for systematically comparing whole-building energy simulation software and determining the algorithms responsible for prediction differences, have a variety of uses, including:

- Comparing a set of building energy simulation programs with each other to determine the amount of disagreement associated with using one program versus another
- Diagnosing the algorithmic sources of differences in predictions observed between a set of building energy simulation programs
- Comparing the predictions from a given simplified building energy program to the reference results from the detailed building energy simulation programs presented in this report
- Checking a program against a previous version of itself after internal code modifications to ensure that only the intended changes actually resulted
- Checking a program against itself after a single algorithmic change to understand the sensitivity between algorithms.

The BESTEST procedures give building energy software developers and users the ability to test a program for reasonableness of results and determine if a program is appropriate for a particular application. The BESTEST procedures have proved very effective at isolating the sources of predictive differences. The IEA BESTEST diagnostic procedures revealed bugs, faulty algorithms, or documentation problems that can lead to input errors in every one of the building energy computer programs (among the best in the world) used in the IEA BESTEST study. Furthermore, users who have documented their work with HERS BESTEST, such as NRCan and FSEC, have made corrections to their software as a result of using that procedure.

Finally, the BESTEST procedures would have been very difficult to develop without international collaboration. International collaboration allows the work to draw from the largest possible pool of modeling expertise and allows each participating country to leverage its funding investment. It would be much more difficult to obtain the large number of high quality example results essential for developing the "foundation" comparative test procedures of IEA BESTEST and HVAC BESTEST if that work were not part of an IEA task.

Recommendations

For future work, a third Tier of tests not currently part of HERS BESTEST is also planned as described in the *HERS Council Guidelines*. HVAC BESTEST, which is currently being developed under IEA SHC Task 22, is the foundation for the HERS BESTEST HVAC tests.

Many of the groups using BESTEST have sent suggestions for improvement. Some of these suggestions can be handled by a simple update sheet, and some would require publication of an updated document. Some of the more interesting improvements might include:

- Re-running reference programs using the new TMY2 weather data after it becomes commonly used by simulation developers
- Periodically re-running reference programs using the newest versions of the programs

- Add cases that specify a combined convective/radiative thermostat, especially where there are large peak load effects such as with thermostat setback/setup
- Revise modeling requirements to specify consistency rather than most detailed models
- Cases should be incorporated that test actual infiltration models rather than simply varying a given rate of exchange
- Basement ground coupling reference results in HERS BESTEST should be expanded to include more accurate methods provided by recent research, and better foundation models should be incorporated into the software used to develop the reference results
- Develop a separate HERS BESTEST test case for summer cooling-load impacts for slab-on-grade floor construction
- Raise the Florida-HERS BESTEST winter thermostat set point assumption for Florida to 72°F (22°C)
- Add a non-deadband set point strategy to IEA BESTEST so that non-hourly tools would be easier to test
- Tabulate monthly results in IEA BESTEST for additional diagnostics so that non-hourly tools with seasonal cut-offs would be easier to test.

Additionally, heat transfer mechanisms that we have not yet tested, but which we believe may contribute to major predictive uncertainties include:

- Interzone and intrazone natural convection, and stratification
- Latent loads and moisture migration in buildings.

There is clearly a need for further development of whole-building energy simulation models combined with a substantial program of testing and validation. Such an effort should contain all the elements of an overall validation methodology, including analytical verification, empirical validation, and comparative testing and diagnostics. An empirical form of BESTEST would be especially desirable.

Acknowledgements

We greatly appreciate the efforts of the following DOE personnel responsible for funding the BESTEST procedures, including: Mary Margaret Jenior, Richard Karney, and John Reese.

The following members of the IEA SHC Programme Task 12b and Conservation Programme Annex 21c contributed to the IEA BESTEST work by providing many of the simulation results: S. Hammond and F. Parand at Building Research Establishment, Watford, United Kingdom; H. Eppel and K. Lomas at De Montfort University, Leicester, United Kingdom; V. Bocchio and A. Mazza at Politecnico di Torino, Torino, Italy; T. Haapala, T. Kalema, and S. Kataja at Tampere University, Tampere, Finland; S. Alvarez and E. Rodriguez at Universidad de Sevilla, Sevilla, Spain; and R. van de Perre and P. Verstraete at Vrije Universiteit, Brussels, Belgium. Also, we appreciate the efforts of Michael Holtz and David Bloomfield, operating agents for Task 12 and Annex 21, respectively; Mary Margaret Jenior, DOE Program Manager for Task 12 and DOE Representative to the IEA SHC Programme Executive Committee.

The following members of the IEA SHC Programme Task 22 contributed to the HVAC BESTEST work by providing many of the simulation results: J. Travesi, CIEMAT, Madrid, Spain; G. Guyon,

Electricite de France, Moret-sur-Loing, France; M. Behne and T. Kretschmer, Klimasystemtechnik, Berlin, Germany; H. Le, Technische Universitat at Dresden, Dresden, Germany.

The following researchers submitted additional results that were used in this paper: Philip Fairey, FSEC, Cocoa, Florida, US; Debra Haltrecht, Natural Resources Canada, Ottawa, Ontario, Canada.

References

Affordable Housing Through Energy Conservation-A Guide to Designing and Constructing Energy Efficient Homes (Technical Support Document). 1989. Berkeley, Calif.: Lawrence Berkeley National Laboratory.

ASHRAE Proposed Standard 140P, Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs. 1998. Working Draft 98/2. Atlanta, GA. American Society of Heating, Refrigerating, and Air-Conditioning Engineers, June 3.

ASHRAE Handbook Fundamentals. 1993. Atlanta, GA. American Society of Heating, Refrigerating, and Air-Conditioning Engineers: 25.10-25.12.

BLAST User Reference, Volume 1 and Volume 2. 1991. BLAST Support Office. Urbana, Ill.: University of Illinois.

Bloomfield, D. editor. 1989. *Design Tool Evaluation: Benchmark Test Cases*, Report # T8B4, BRE. Garston, U.K., International Energy Agency Solar Heating and Cooling Program.

Building Energy Tools Directory. 1998. Washington, DC: United States Department of Energy. Worldwide Web address "http://www.eren.doe.gov" (follow "EERE Programs & Offices" to "Building Technology, State and Community Programs").

DOE-2 Reference Manual (Version 2.1A) Part 1. 1981. D. York and C. Cappiello, eds. Berkeley, Calif.: Lawrence Berkeley Laboratory, May.

DOE-2 Supplement (Version 2.1D). 1989. The Simulation Research Group, Energy and Environment Division, Lawrence Berkeley Laboratory, One Cyclotron Rd., Berkeley, Calif. 94720, June.

DOE-2 Supplement (Version 2.1E). 1994. Berkeley, CA: Lawrence Berkeley Laboratory, January.

ESRU. 1993. *ESP-r User Guide Version 8.* ESRU Manual U93/2. Glasgow, UK: University of Strathclyde.

Fairey, P., M. Anello, L. Gu, D. Parker, M. Swami, R. Vieira. 1998. *Comparison of EnGauge 2.0 Heating and Cooling Load Predictions with the HERS BESTEST Criteria.* FSEC-CR-983-98. Cocoa, Fla.: Florida Solar Energy Center.

Fairey, P., personal communications with author. February 1998.

Haltrecht, D. and K. Fraser. 1997. "Validation of HOT2000™ Using HERS BESTEST." *Proceedings Building Simulation 97, Fifth International IBPSA Conference*. Prague, Czech Republic. International Building Performance Simulation Association, September 8-10.

Haltrecht, D. personal communications with author. March 1998.

Hirsch, J. (James J. Hirsch & Associates) Camarillo, Calif. Personal communications with J. Neymark. 1992 through 1997.

Home Energy Ratings Systems (HERS) Council. 1995. *Guidelines for Uniformity: Voluntary Procedures for Home Energy Ratings*. Washington, D.C.: The Home Energy Rating Systems Council.

Housing Characteristics 1990. Washington, D.C.: U.S. Department of Energy, Energy Information Administration.

Judkoff, R., D. Wortman, B. O'Doherty, and J. Burch. 1983. *A Methodology for Validating Building Energy Analysis Simulations*. SERI/TR-254-1508. Golden, Colo.: Solar Energy Research Institute (now NREL).

Judkoff, R. 1988. "Validation of Building Energy Analysis Simulation Programs at the Solar Energy Research Institute." *Energy and Buildings*, 10(3):235.

Judkoff, R., S. Barakat, D. Bloomfield, B. Poel, R. Stricker, P. van Haaster, and D. Wortman. 1988. "International Energy Agency Design Tool Evaluation Procedure," *PROCEEDINGS ASES, MIT*, SERI/TP-254-3371, DE88001173, Cambridge Mass..

Judkoff, R. 1989. *IEA Task 12b/21c Multi-Year Research Plan*. NREL.

Judkoff, R., J. Neymark. 1995a. *International Energy Agency Building Energy Simulation Test (BESTEST) and Diagnostic Method*. NREL/TP-472-6231. Golden, Colo.: National Renewable Energy Laboratory.

Judkoff, R., J. Neymark. 1995b. *Home Energy Rating System Building Energy Simulation Test (HERS BESTEST)*. NREL/TP-472-7332. Golden, Colo.: National Renewable Energy Laboratory.

Judkoff, R., J. Neymark. 1995c. "A Procedure for Testing the Ability of Whole Building Energy Simulation Programs to Thermally Model the Building Fabric." *Journal of Solar Energy Engineering* February: 117:7-15.

Judkoff, R., J. Neymark. 1997. *Home Energy Rating System Building Energy Simulation Test for Florida (Florida-HERS BESTEST)*. NREL/TP-550-23124. Golden, Colo.: National Renewable Energy Laboratory.

Kennedy, M., L. Palmiter, T. Wheeling. 1992. *SUNCODE-PC Building Load Simulation Program*. Ecotope, Inc., 2812 E. Madison, Seattle, Wash., 98112, (206) 322-3753. This software is based on

SERIRES-1.0 developed at NREL. (See also Palmiter et al.)

Klein, S.A. 1990. *TRNSYS A Transient System Simulation Program*," Madison, Wisc., Solar Energy Laboratory, University, September. Cited as TRNSYS.

Mitalas, G., J. Arsenault. 1971. "Fortran IV Program to Calculate Z-Transfer Functions for the Calculation of Transient Heat Transfer through Walls and Roofs." *Use of Computer for Environmental Engineering Related to Buildings*. Washington, D.C.: National Bureau of Standards, Building Science Series 39: 633-668.

National Association of Home Builders (NAHB) Research Center. 1992. *1980 Base Case and Feasibility Analysis*. Prepared for National Renewable Energy Laboratory, Golden, Colo.

National Fenestration Rating Council (NFRC). 1993 . *Certified Products Directory*. First Edition. Silver Spring, MD: National Fenestration Rating Council.

Neymark, J., R. Judkoff. 1997a . *International Energy Agency Building Energy Simulation Test and Diagnostic Method for Mechanical Equipment (HVAC BESTEST)*. Draft, 1997. Golden, Colo.: National Renewable Energy Laboratory, September.

Neymark J. R. Judkoff. 1997b. "A Comparative Validation Based Certification Test for Home Energy Rating System Software." *Proceedings Building Simulation 97, Fifth International IBPSA Conference*. Prague, Czech Republic. International Building Performance Simulation Association, September 8-10.

Palmiter, M.L., T. Wheeling, R. Judkoff, B. O'Doherty, D. Simms, D. Wortman. 1983. *Solar Energy Research Institute Residential Energy Simulator (Version 1.0)*. Golden, Colo.: Solar Energy Research Institute (now NREL).

Rittelmann, P.R. S.F. Ahmed, 1985a, Design Tool Survey, IEA Task VIII, Burt Hill Kosar Rittelmann Assoc, Butler, PA.

Rittleman, P.R., S. F. Ahmed. 1985b. *Design Tool Comparison*, IEA Task VIII Working Document, Burt Hill Kosar Rittelmann Assoc, Butler, PA.

F. Winkelmann. 1991-1993. Lawrence Berkeley Laboratory, Berkeley, CA. Personal communications. December 1991 through February 1993.