TESTING AND VALIDATION OF A NEW BUILDING ENERGY SIMULATION PROGRAM

Michael J. Witte, Robert H. Henninger, and Jason Glazer
GARD Analytics, Inc.
Park Ridge, IL 60068 USA

Drury B. Crawley
U.S. Department of Energy
Washington, DC 20585 USA

ABSTRACT

Formal independent testing has been an integral component in the development of EnergyPlus, a new building energy simulation program. Testing to date has included analytical, comparative, sensitivity, range, and empirical tests. Published test suites which include reference results have been applied as much as possible in order to take advantage of the efforts of others to develop well-defined, reproducible tests. The results to date show good agreement with well-established simulation tools such as DOE-2.1E, BLAST, and ESP. Several testing utilities have been developed to help automate the task of assuring that each new version of the software is still performing properly. Selected test results are presented along with lessons learned.

INTRODUCTION

Formal independent testing has been an integral component in the development of EnergyPlus, a new building energy simulation program recently released by the U.S. Department of Energy. Comprehensive testing of building energy analysis software is a difficult task given the infinite combinations of inputs that may be entered and the difficulties in establishing truth standards for all but the simplest cases. Testing has been guided by a comprehensive test plan which includes the following types of tests:

- Analytical tests which compare against mathematical solutions,
- Comparative tests which compare against other software,
- Sensitivity tests which compare small input changes versus a baseline run,
- Range tests which exercise the program over wide ranges of input values, and
- Empirical tests which compare against experimental data.

Published test suites which include reference results have been applied as much as possible in order to take advantage of the efforts of others to develop well-defined, reproducible tests. Published suites used to date include:

- Analytical - BEPAC/Bland Conduction Tests (BEPAC 1993, Bland 1992),
- Analytical - ASHRAE 1052-RP Building Fabric (ASHRAE 2000),
- Comparative - BESTEST/ASHRAE Standard 140P (ASHRAE 2000),
- Comparative/Analytical - HVAC-BESTEST (IEA 2000),
- Empirical - IEA Validation Suite (IEA 1994).

While the scope of these tests is rather limited compared to the broad range of the software's capabilities, the user community is likely to use the results of these tests to judge the credibility of this new software. Selected test results are presented below along with discussion of lessons learned.

ANALYTICAL – BEPAC/BLAND CONDUCTION

BEPAC (the Building Energy Performance Analysis Club in the UK) has published a collection of conduction tests for validating building simulation software (BEPAC 1993, Bland 1992). The validation package includes FORTRAN routines for calculating results. This test suite can be exercised for a broad range of material properties which are subjected to step, ramp, and sine changes in outdoor dry bulb temperatures. The tests are run with a cube made of the same material on all six sides with free-floating or controlled inside temperature. One of the cases is a cube of 5 cm thick aluminum driven by a step change weather file which goes from 0 to 20 to 70 to 50 to 20 C for months at a time with a heating setpoint of 20C and a cooling setpoint of 50C. This material was included in the BEPAC/Bland test suite to test conduction calculations with thermal mass and near-zero thermal resistance. In an early Beta 1 version of EnergyPlus, this test case uncovered an accuracy problem with the conduction transfer function (CTF) calculations, as illustrated in Figure 1. EnergyPlus allows the user to select from 1 to 6 time steps per hour. With time step=1/2 hour, the inside surface and air temperatures responded as expected to the changes in outdoor temperature. But with time step=1 hour, impossible conditions occurred, for example, outdoor=70C, indoor=20C, and surfaces=18C. This error has been fixed as a result of these tests.

ANALYTICAL – ASHRAE 1052RP BUILDING FABRIC

The American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) sponsored research project 1052RP to develop analytical tests for the building fabric (ASHRAE 2000). At the time of this writing, the project was nearly complete, including documentation describing 16 tests and a software toolkit to generate analytical results. The tests cover a variety of building envelope mechanisms including conduction, convection, solar gains, shading, infiltration, internal gains, radiant transfer, and ground coupling. While a variety of analytical conduction tests have been published before, this is the first test suite to provide analytical solutions for the other areas. These tests were applied to EnergyPlus as part of the review process for the research project. Several bugs were found in EnergyPlus while applying the tests, and the results of some of the tests have raised questions requiring further investigation. Figure 2 shows the results of one of the transient conduction tests. In this test, the initial conditions are 20C outdoor dry bulb and indoor temperature is controlled to 20C. After several months at the initial condition, the outdoor temperature is stepped up to 70C. Inside and outside surface temperatures and zone cooling load are compared to the analytical solution over the next several days. For this case, the general response of the simulation is good, and the two solutions come together after 30 to 40 hours. But there are some small differences during the transition period with inside temperature differing by as much as 1.3C in the sixth hour after step-up. Yet the integrated cooling load agrees within 0.5% for the first 24 hours and within 0.05% for the first 48 hours. For some special purposes, the 1.3C surface temperature error may be significant, but for the vast majority of applications any errors seen here are insignificant. The 1052RP toolkit allows the user of the test suite to generate results for a wide range of material properties. It is important to exercise the simulation tool for a wide range of combinations of mass, conductivity, and thickness in order to uncover hidden instabilities.

Figure 3 shows results for one of the window solar gain tests with south facing glass on August 21. The solar incident on the exterior and the solar transmitted by the window are compared throughout the day. Note the time shift between the simulated and analytical results for this case. This was observed for many of the solar-related tests. Initially this was thought to be a daylight savings time error, but that has been ruled out. A possible cause may be the interpolation between hourly weather data for sub-hourly time steps. The development team is investigating this issue.

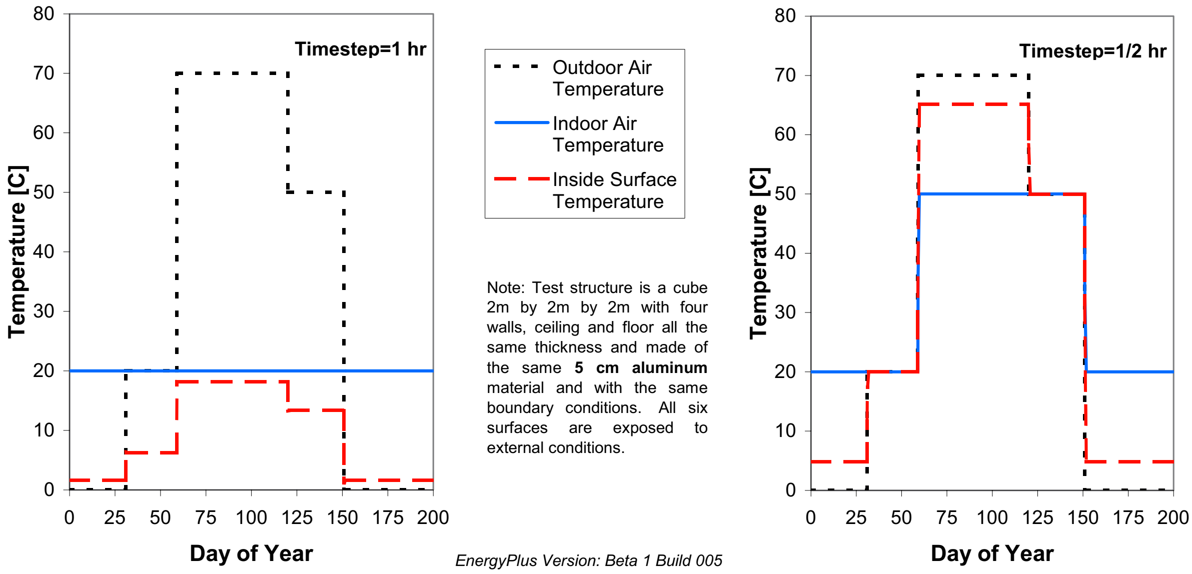

Figure 1 Results from BEPAC/Bland Conduction Test with 5cm of aluminum showing early CTF accuracy problem. *In an early Beta 1 version of EnergyPlus, the BEPAC/Bland conduction test suite uncovered an accuracy problem with the CTF calculations with a construction of 5cm of aluminum. With time step=1/2 hour, the inside surface and air temperatures responded as expected. But with time step=1 hour, impossible conditions occurred, e.g. outdoor=70C, indoor=20C, surfaces=18C. This error has been fixed.*

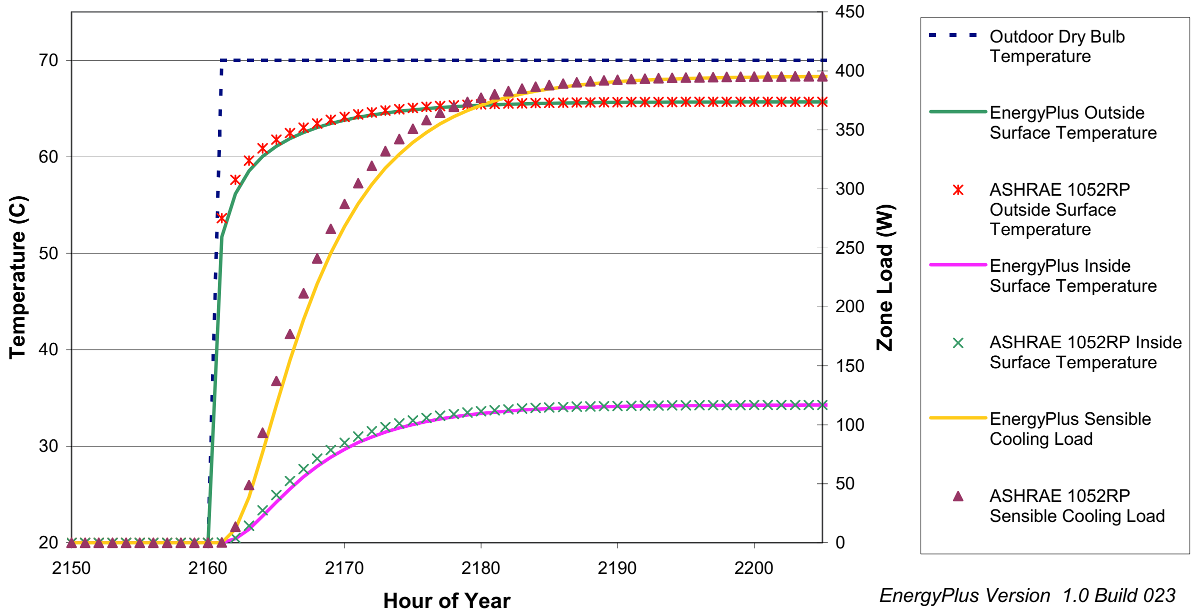

Figure 2 Results from ASHRAE 1052RP Analytical Test Suite - Transient Conduction TC2, 10 cm Wood, 50C Step Up Change in External Temperature, Indoor Temperature Constant at 20C. *For this case, excellent agreement is shown. The 1052RP toolkit allows the user of the test suite to generate results for a wide range of material properties.*

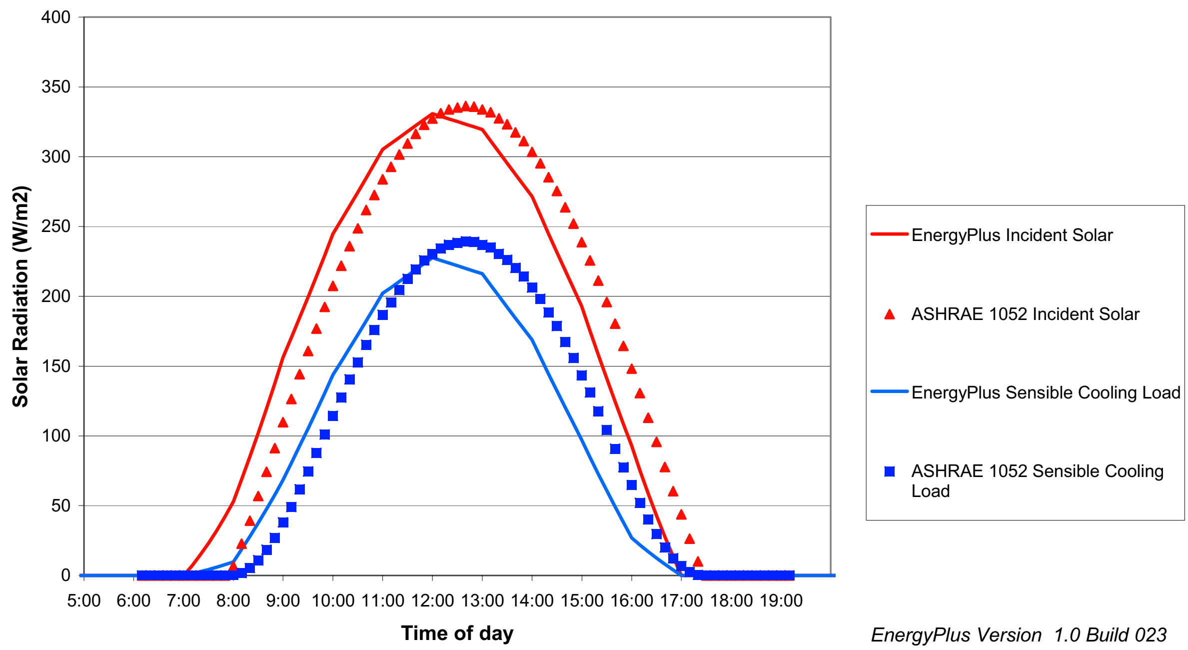

Figure 3 Results from ASHRAE 1052RP Analytical Test Suite - Window Solar Gain, Atlanta, August 21, South Facing 3.175 mm (1/8 inch) Clear Single-Pane Glass. *Note the time shift between the simulated and analytical results. Possible cause may be interpolation between hourly weather data for sub-hourly time steps.*

COMPARATIVE – BESTEST/ASHRAE STD 140

The BESTEST suite developed through an International Energy Agency (IEA) project (IEA 1995) is a comparative set of tests run on single-zone and double-zone shoebox configurations with variations in mass, windows, overhangs, and fins. BESTEST has been restated as a Standard Method of Test in ASHRAE Standard 140-2001 (ASHRAE 2001). Reference results for eight different simulation programs are included with the standard to provide a comparison point for testing other software. Although not part of the original BESTEST set of results, results for later versions of BLAST (3.0 level 334) and DOE2 (DOE-2.1E) have been added. For each test case, results for annual cooling load, annual heating load, peak cooling load, and peak heating load are compared. In addition, there are some free-floating cases where maximum and minimum temperatures are compared instead of load. Annual cooling results for low mass building construction are presented in Figure 4. Each cluster of bars shows the results from all programs for a particular case. In general, EnergyPlus results are within the range of the other tested programs. Applying this test suite helped identify several bugs and documentation deficiencies, e.g. shade fin surface coordinates were inverted. Figure 5 shows how the EnergyPlus results changed with various versions of EnergyPlus. Notice in Case 630 how the cooling load dropped when the fin shading coordinate problem was corrected in version 1-14.

COMPARATIVE/ANALYTICAL – HVAC BESTEST

The HVAC BESTEST suite, in development through an IEA project, tests the cooling loads and electric power consumption for a single-zone DX cooling system with a dry or wet coil under varying conditions of entering dry bulb, outdoor dry bulb, part-load ratio (PLR), and sensible heat ratio (SHR).

Figure 6 shows the space cooling electricity consumption (including indoor fan, compressor, and outdoor fan) comparisons for EnergyPlus versus other tested programs and three analytical solutions. While EnergyPlus results are within the range of results from the other tested programs, there are some issues related to fan energy and humidity ratio which warrant further investigation. This test suite revealed several bugs, including:

- Reporting problems for coil loads with cycling fan operation.
- Standard air density not applied as documented
- Supply fan operation needed to include cycling adjustments
- Use of both dry and wet c_p needed to be converted to all wet c_p
- Heat of vaporization assumptions needed adjustment to make latent zone loads and coil loads consistent with each other.

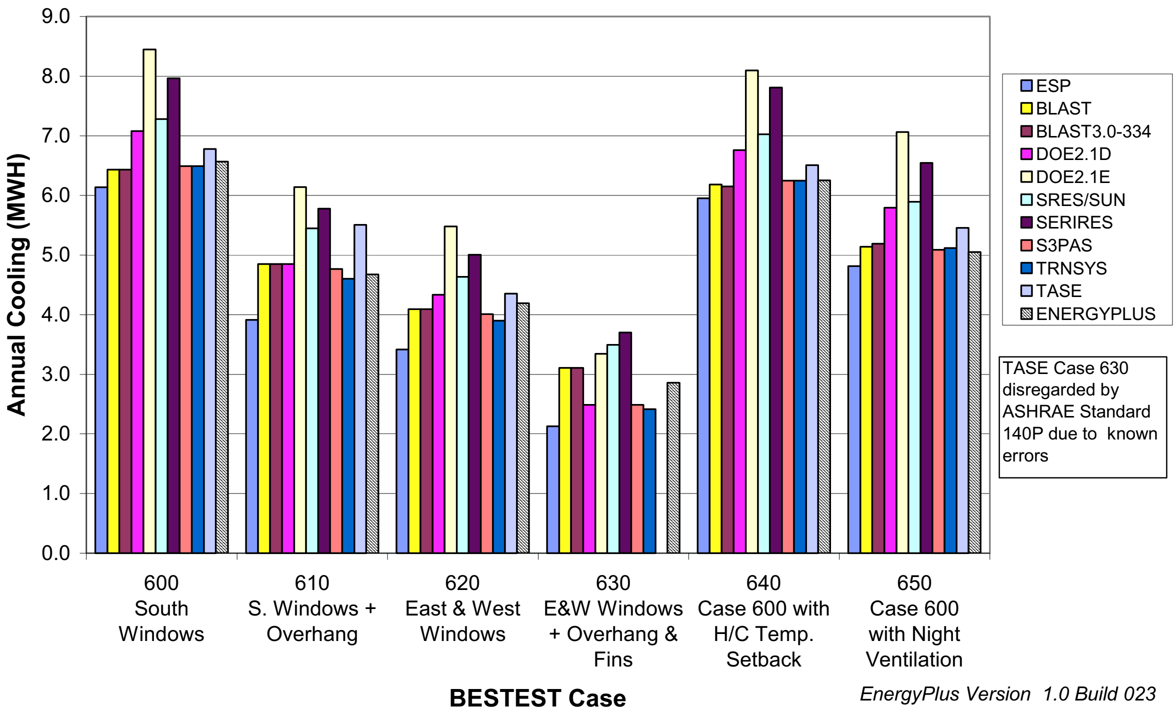

Figure 4 Results from BESTEST/ASHRAE Std 140 – Low Mass Building Annual Cooling. *EnergyPlus results are within the range of results from other tested programs.*

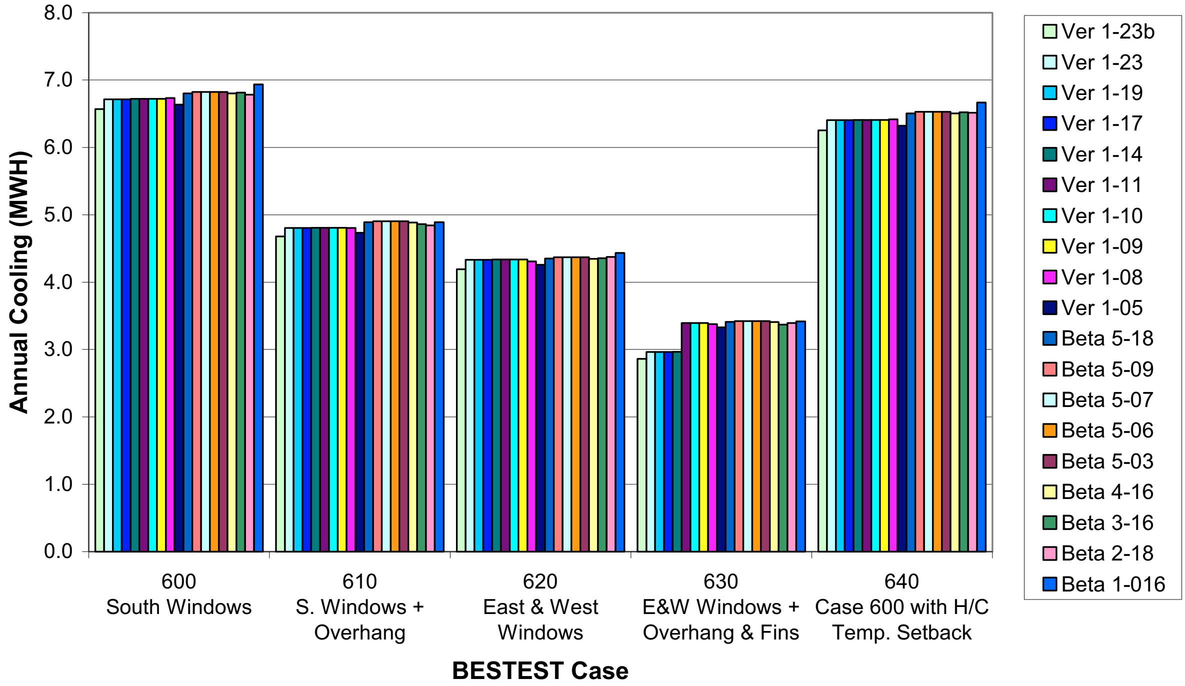

Figure 5 Comparison of Results for Various Versions of EnergyPlus for BESTEST/ASHRAE Std 140 – Low Mass Building Annual Cooling. *Notice for Case 630 that annual cooling was reduced significantly when the fin shading coordinate problem was corrected in version 1-14.*

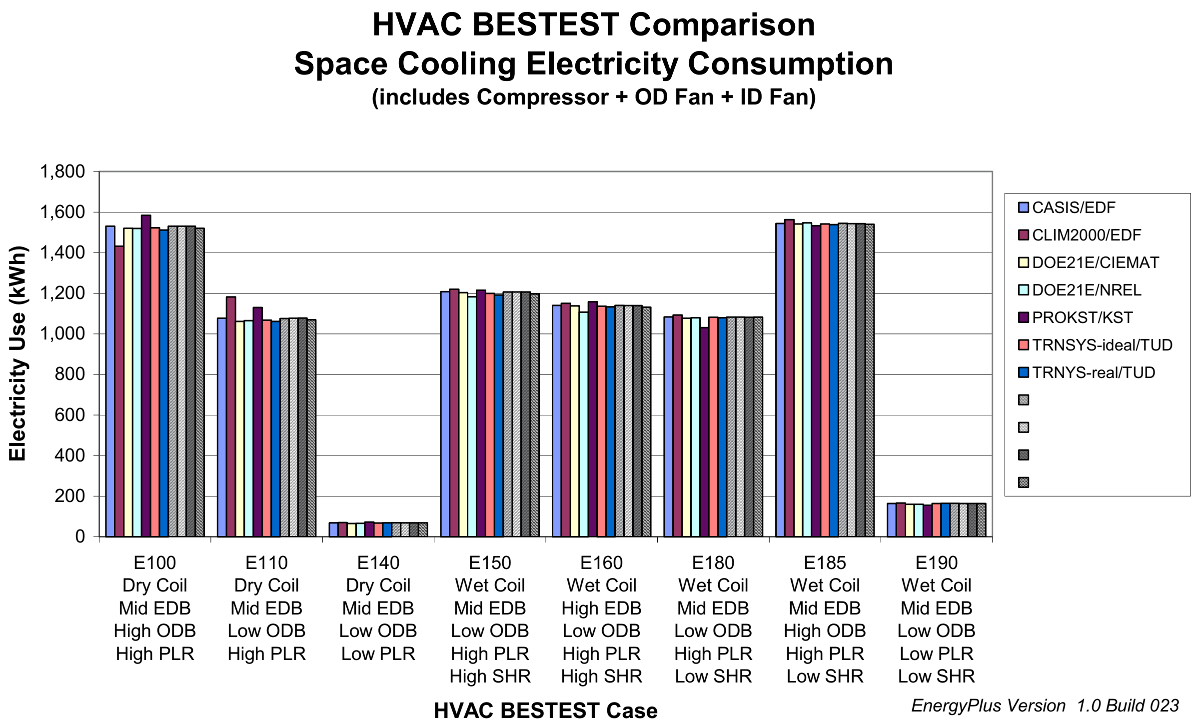

Figure 6 Electricity Consumption Results for Selected HVAC BESTEST Cases. *EnergyPlus results are within the range of results from the other tested programs, but there are some issues related to fan energy and humidity ratio which warrant further investigation.*

SENSITIVITY AND RANGE TESTS

Sensitivity and range test suites have been developed to systematically exercise many of the program inputs in order to confirm that basic program elements are functioning properly and to identify as many bugs as possible prior to public release of the software. These test suites take advantage of the macro preprocessor and allow batch processing of large numbers of runs with parameters passed to alter one or more values in the input file. An automated tool has been developed which will take any input file and exercise all fields for nine cases: slightly less than minimum, minimum allowed, nominal (the value originally in the input file), nominal minus 50%, nominal plus 50%, maximum allowed, and slightly more than maximum. The minimum and maximum allowed values are determined from specifications in the EnergyPlus input data dictionary (IDD) file. A log file is generated which summarizes whether each run completed successfully, terminated with error messages, or crashed. The log file also compares the changes in output variables for the nominal and nominal minus 50% cases and uses this to predict the results for the nominal plus 50% cases. The first time this system was used, many of the runs crashed, prompting several change requests, most of which were related to the IDD minimum and maximum specifications and the need to terminate the simulation gracefully if temperatures go out of range. With these changes implemented, crash conditions have been eliminated for the cases tested.

EMPIRICAL - IEA VALIDATION SUITE

The IEA Empirical Validation Package (IEA 1994) contains detailed information for two ten-day experiments using three highly monitored test rooms. The validation package was developed specifically for use in validating building energy simulation software. It contains detailed descriptions of the test rooms, heating equipment, and instrumentation and includes a data diskette. The three rooms were equipped with single glazing, double glazing, and an opaque insulated panel. Data for two tests are provided: heated (October 17 - 26, and unheated (May 21 - 30). These test cases exercise the following components of the simulation:

- Opaque conduction and exterior solar gains
- Simple glazing, conduction and solar gains
- Zone heat balance without internal loads
- Simple heating system

Input files have been developed for this test suite, but the meteorological data which comes with the test suite is not in a standard format. The data contains global and diffuse horizontal total radiation which must be converted to direct normal solar values for use with EnergyPlus.

OTHER TESTING TOOLS

Several testing utilities have been developed to help automate the task of assuring that each new version of the software is still performing properly. One tool, which is used extensively by the software developers, is a tool for mathematically comparing output files. Typical text file comparison tools will flag differences even when a result has only changed by a very small amount. The mathematical difference utility compares two CSV (commas separated values) format files containing simulation results. The results files will typically be for the same input file using two different versions of the software. Any differences which are greater than both a specified percentage tolerance and absolute tolerance (e.g. 0.5% and 0.001 difference) are flagged in a log file. This batch-oriented system allows large suites to be run before and after code changes are implemented to quickly see what results have changed significantly. This allows developers to gain confidence that changes made in one section of the program have not unintentionally changed results from another section of the program.

CONCLUSIONS

Formal independent testing during the development of EnergyPlus has helped produce a more robust and credible simulation tool. Application of published test suites for analytical, comparative, and empirical tests has been very useful in detecting bugs and confirming that basic modeling algorithms are working properly. Significant bugs found include:

- Conduction transfer function accuracy problem with certain material types
- Solar time shift
- Inverted coordinates for shading fins
- HVAC mass flow and c_p inconsistencies
- DX coil reporting error when cycling
- Various code crashes with extreme inputs

Even with well-documented test suites, there were often input details which were not specified or which did not translate directly into EnergyPlus inputs. EnergyPlus results to date generally show good agreement with well-established simulation tools such as DOE-2, BLAST, TRNSYS and ESP. The results of these tests may be used by the user community to judge the credibility of this new simulation software.

Additional test suites and test utilities have also been developed to aid the development and debugging process. Range and sensitivity tests have helped to eliminate code crashes, and a utility for mathematical comparisons of results has simplified the task of evaluating the impact of source code changes.

ACKNOWLEDGEMENTS

This work was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Building Technology, State and Community Programs, Office of Building Research and Standards of the U.S. Department of Energy, Drury Crawley, program manager.

REFERENCES

ASHRAE 2000. *ASHRAE Analytical Test Suite – Building Fabric*, 1052RP Draft Final Report, December 2000.

ASHRAE 2001. ASHRAE Standard 140, *Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs*, (expected publication date August 2001).

BEPAC 1993. B.H. Bland, *Conduction tests for the validation of dynamic thermal models of buildings*, Technical Note 93/1, August 1993.

Bland 1992. Bland, B.H. "Conduction in dynamic thermal models: Analytical tests for validation", *Building Services Engineering Research and Technology*, 13(4), pp 197-208.

IEA 1994. K.J. Lomas, C. Matrin, H. Eppel, M. Watson, and D. Bloomfield, *Empirical Validation of Thermal Building Simulation Programs Using Test Room Data, Volume 2: Empirical Validation Package*, September 1994.

IEA 1995. *Building Energy Simulation Test (BESTEST) and Diagnostic Method*, National Renewable Energy Laboratory, Golden, Colorado, February 1995.

IEA 2000. J. Neymark, and R. Judkoff, *International Energy Agency Building Energy Simulation Test and Diagnostic Method for HVAC Equipment Models (HVAC BESTEST) Vol. 1: Cases E100-E200* (Draft), November 2000.