The Building Energy Simulation Test for Existing Homes (BESTEST-EX) Methodology

Preprint

Ron Judkoff, Ben Polly, and Marcus Bianchi
National Renewable Energy Laboratory

Joel Neymark
J. Neymark & Associates

To be presented at the Building Simulation 2011 Conference
Sydney, Australia
November 14-16, 2011

Conference Paper
NREL/CP-5500-51655
November 2011

Contract No. DE-AC36-08GO28308

NOTICE

The submitted manuscript has been offered by an employee of the Alliance for Sustainable Energy, LLC (Alliance), a contractor of the US Government under Contract No. DE-AC36-08GO28308. Accordingly, the US Government and Alliance retain a nonexclusive royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for US Government purposes.

This report was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or any agency thereof.

Available electronically at http://www.osti.gov/bridge

Available for a processing fee to U.S. Department of Energy
and its contractors, in paper, from:

U.S. Department of Energy
Office of Scientific and Technical Information

P.O. Box 62
Oak Ridge, TN 37831-0062
phone: 865.576.8401
fax: 865.576.5728
email: mailto:reports@adonis.osti.gov

Available for sale to the public, in paper, from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
phone: 800.553.6847
fax: 703.605.6900
email: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/help/ordermethods.aspx

Cover Photos: (left to right) PIX 16416, PIX 17423, PIX 16560, PIX 17613, PIX 17436, PIX 17721

Printed on paper containing at least 50% wastepaper, including 10% post consumer waste.

THE BUILDING ENERGY SIMULATION TEST FOR EXISTING HOMES (BESTEST-EX) METHODOLOGY

Ron Judkoff[1], Joel Neymark[2], Ben Polly[1], and Marcus Bianchi[1]
[1]National Renewable Energy Laboratory, Golden, Colorado, U.S.
[2]J. Neymark & Associates, Golden, Colorado, U.S.

ABSTRACT

The authors developed a method for testing the reliability of models that predict retrofit energy savings, including their associated calibration methods. The test suite applies the new Building Energy Simulation Test for Existing Homes (BESTEST-EX) Methodology, developed by the National Renewable Energy Laboratory in consultation with home energy retrofit industry software developers. BESTEST-EX includes building physics test cases with fully known inputs, and calibrated energy savings test cases with specified base-case monthly utility billing data and uncertainty ranges for selected inputs. The test cases apply a variety of retrofit scenarios in a heating and a cooling climate.

INTRODUCTION

A number of computerized energy auditing systems use utility bill data and a variety of calibration methods with the objective of tuning their audit models to more accurately predict energy savings from retrofits. Performance-based tax incentives for U.S. home energy retrofits may increase; thus, procedures need to be established to test the accuracy of building energy audit software used to predict retrofit energy savings. Consequently, the National Renewable Energy Laboratory (NREL) began developing a process for testing the reliability of models that predict retrofit energy savings, including their associated calibration methods. NREL is conducting the work in phases; this paper summarizes the initial “Phase 1” test procedure and example acceptance criteria, which focus on building thermal fabric test cases (Judkoff et al. 2010a, 2010b, 2011).

Background: Building Energy Simulation Test and Diagnostic Method (BESTEST)

NREL has developed a number of building energy simulation test (BESTEST) suites for evaluating and diagnosing errors in software used for energy analysis of residential and commercial buildings. ASHRAE Standard 140, *Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs* (ANSI/ASHRAE 2007) either adopted, or are in the process of adopting, the BESTEST suites. Many entities have adopted or cited Standard 140 and/or the component BESTEST suites: the Internal Revenue Service (2008) (for certifying software used to determine tax deductions), ASHRAE building energy efficiency Standards 90.1 and 189.1 (ANSI/ASHRAE/IESNA 2007, ANSI/ASHRAE 2009), RESNET (2007), COMNET (Eley 2011), the International Energy Agency (Judkoff and Neymark 2009), and the European Community under their Energy Performance Directive (European Union 2002). These methods include software-to-software comparative testing, verification versus analytical solutions, and validation versus vetted empirical data. The theoretical basis for the BESTEST procedures is further described in the literature (ASHRAE 2009, Judkoff 1988, Judkoff et al. 2008, Judkoff and Neymark 2006).

Overview of the BESTEST-EX Phase 1 Test Suite

The test suite represents a set of cases applying the new Building Energy Simulation Test for Existing Homes (BESTEST-EX) Methodology developed by NREL. (Judkoff et al. 2010a). The NREL team developed the test cases in consultation with the home retrofit industry (BESTEST-EX Working Group 2009), and adjusted the test specifications in accordance with information supplied by a participant with access to large utility bill datasets (Blasnik 2009). BESTEST-EX includes two kinds of test cases:

- Building physics test cases with fully known inputs
- Calibrated energy savings test cases with specified base-case monthly utility bill data and uncertainty ranges for selected inputs.

The calibrated energy savings tests represent a new methodological development. The “Methods” section describes further the physics and calibration test procedures.

The cases test the ability of software to model space heating loads in a representative heating climate and space cooling loads in a representative cooling climate. The following retrofit cases are included: infiltration air sealing, attic insulation, wall insulation, programmable thermostat, low-e windows, low exterior solar-absorptance roof (cool roof), and external solar shading (physics cases only). Combined retrofit cases are also included as appropriate for both the heating and cooling climates.

To help avoid user input errors, the input for the test cases is as simple as possible, and represents "typical" constructions and thermal and physical properties. The BESTEST-EX base building is based on HERS BESTEST (Judkoff and Neymark 1995). Typical building descriptions and physical properties published by sources such as the U.S. Department of Energy (DOE), the National Association of Home Builders, the American Society of Heating Refrigerating and Air-Conditioning Engineers (ASHRAE), and the National Fenestration Rating Council are used for the test cases. The development team used empirical information from several large utility bill studies (Blasnik 2009), in consultation with industry participants (BESTEST-EX Working Group 2009), to modify some of the thermal inputs (e.g., surface heat transfer coefficients) to be more appropriate for poorly insulated older buildings.

METHODS

This section summarizes the BESTEST-EX method fundamentals. Further details about specific aspects of the methodology are included in the next section ("Application of the Method"). Italicized terms used in the following discussion are defined near the end of the paper under "Definitions."

Building Physics Test Cases With Known Inputs

The building physics test cases are a direct application of software-to-software comparative test methods (Judkoff and Neymark 2006). A given audit model is tested using specified inputs; resulting outputs are compared with reference results from three detailed simulation programs (EnergyPlus, DOE-2.1E, and SUNREL). Tested program results may also be compared with accompanying example acceptance criteria (Judkoff et al. 2010b), or with other results generated using the test procedure.

Utility Bill Calibration Test Cases

After running the building physics cases, diagnosing results disagreements, and correcting all found modeling errors, a given audit model (and associated calibration methods) can be tested by comparing utility-bill-calibrated energy savings predictions to results from the reference programs. The calibrated energy savings tests required NREL to create a new method (see Figure 1), described as follows.

1. ***Introduce input uncertainty into the test specification (this represents the uncertainty associated with developing inputs from audit survey data):***
   a. Perform sensitivity tests on inputs with potentially high uncertainties to determine their relative effects on outputs; select the inputs that have the greatest uncertainties and effects on outputs as *approximate inputs*.
   b. Specify an uncertainty range (*approximate input range*) for each *approximate input.*
2. ***Develop reference simulation results (this is done by the test developers):***
   a. Generate base-case synthetic utility bill data using the same state-of-the-art reference simulation programs used in the building physics test cases.
      i. For the reference simulations, inputs that are randomly selected from within the specified *approximate input ranges* are designated as *explicit inputs*.
      ii. All reference simulations use the same or equivalent *explicit inputs* for a given calibration scenario.
      iii. The synthetic utility bill data are taken as the average of the reference simulations results.
   b. Generate reference energy savings results by adjusting appropriate base case inputs, including *explicit inputs*, as specified for each retrofit case.
3. ***Develop tested program results (this is done by the test takers):***
   a. Develop the preliminary uncalibrated base-case model for a given calibration scenario.
   b. Predict energy savings via one of the following:
      i. Calibrate the base-case model inputs using the synthetic utility bills (described in 2a), then apply the specified retrofit cases to the calibrated model.
      ii. Apply the specified retrofit to the uncalibrated base case model and then calibrate or correct energy savings predictions using the synthetic utility bills (without adjustment to base-case model inputs), e.g., as (calibrated savings) = (predicted savings) × (base case actual bills)/(base case predicted bills).
      iii. Other calibration methods. The test cases make no recommendation about how to perform calibrations.
4. ***Compare results of tested programs (and their calibration techniques) versus reference simulation base-case energy use and retrofit energy savings projections:***
   a. Example acceptance criteria (further described under "Results") may be used to facilitate the comparison.

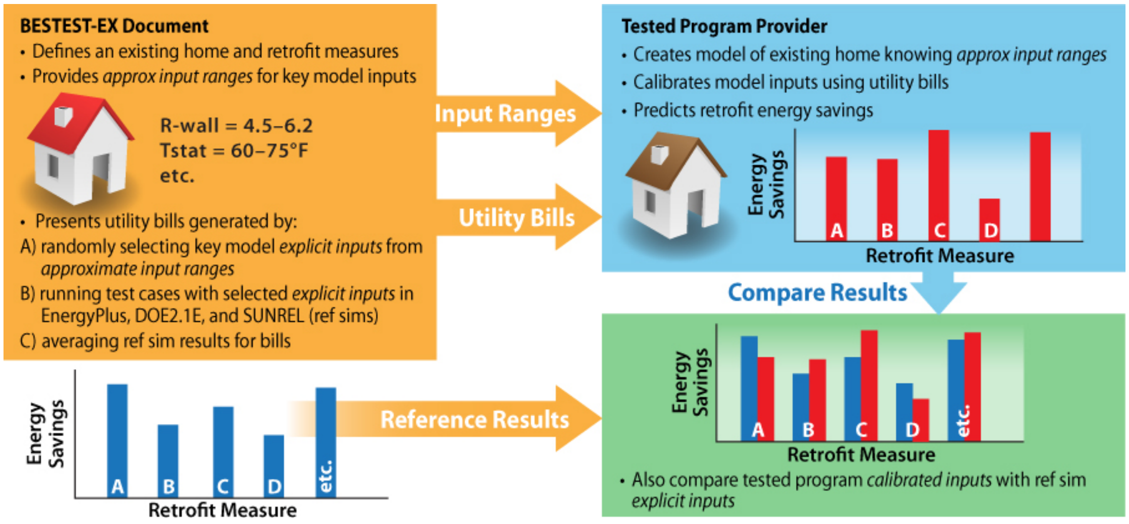


Figure 1 Calibration Cases Conceptual Flow (Bianchi et al. 2010; Judkoff 2008; Neymark and Norton 2009)

Development of the method was facilitated by convening a technical committee of software developers (the "BESTEST-EX Working Group") to help estimate approximate input ranges and develop tested program results (see Step 1b and Step 3). The test procedure and example acceptance criteria were developed in an iterative process that enabled us to improve the test specification during the simulation trials and helped trial participants improve their software.

In its purest form, the calibration test would be implemented without using the reference simulation programs. Instead, synthetic utility billing data would be generated with the tested program. Such a pure calibration test requires a) automated calibration or b) that the modeler running the calibration test does not know the explicit inputs used to develop the synthetic utility bills. Either method is acceptable, but the latter is impractical for certifying organizations. Further discussion of self testing is included elsewhere (see Judkoff et al. [2011], Appendix B).

APPLICATION OF THE METHOD

Approximate Input Ranges

Approximate input ranges in the test specification represent the uncertainty associated with developing model inputs based on audit survey data. *Approximate input ranges* are provided for selected base-case model input parameters (see Table 1). The selected parameters strongly affect energy use predictions, and are commonly known to have pre-retrofit audit uncertainty.

Nominal inputs are used for the building physics test cases (approximate input ranges are ignored for the physics tests). The *nominal input* values are also used in the calibration tests for developing uncalibrated models before they are adjusted for model calibration.

Random Selection of Explicit Inputs

Where *approximate input ranges* are provided, *explicit inputs* for the reference programs are randomly selected from within the *approximate input ranges* assuming a triangular probability distribution (see Figure 2). For this distribution the probability of selection is greatest at the nominal value and decreases linearly to zero at the minimum and maximum values. The triangular distribution may be either symmetric or asymmetric; an asymmetric distribution is shown in Figure 2.

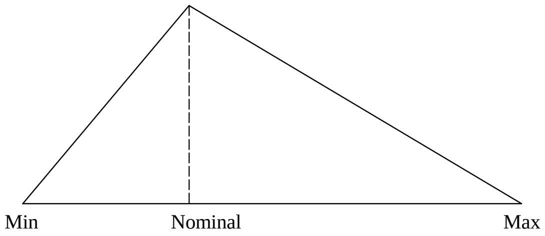


Figure 2 Triangular probability distribution assumed for random generation of explicit input (Judkoff et al. 2010a, Kotz and van Dorp 2004)

Calibration Scenarios

Feedback from the BESTEST-EX Working Group generated during the simulation trials indicated a need for multiple calibration scenarios to cover a reasonable range of calibration problem types. Three types of explicit input sets were generated for the L200EX-C base case scenarios: *targeted high*, *targeted low*, and *fully random* space-conditioning energy consumptions. Given *approximate input ranges* (min, max) for selected inputs (see Table 1), sets of *explicit inputs* were generated using the following approaches:

Table 1 Approximate Input Ranges (AIRs), Nominal Inputs, and Portions of AIRs Used for Generating Explicit Input Sets Corresponding to Targeted Low, Targeted High, and Fully-Random Selected Space-Conditioning Energy Consumptions (Judkoff et al. 2010a)

				Portion of AIR Used for Given Scenario Type		
Input	**Min**	**Max**	**Nominal**	**Targeted High (C1)**	**Targeted Low (C2)**	**Fully Random Selected (C3-C7)**
Wall R (h·ft^2·°F/Btu)	4.5	6.2	5.091	Lower	Upper	Entire
Attic R (h·ft^2·°F/Btu)	7.1	19.3	13.673	Lower	Upper	Entire
Leak area (in.^{2})	137.4	215.9	196.3	Upper	Lower	Entire
Occupant gains (Btu/d)	4347	13041	8694	Entire	Entire	Entire
Electric gains (Btu/d)	18234	80000	36468	Entire	Entire	Entire
Use to gain (%)	60.0	90.0	75.0	Entire	Entire	Entire
Gas gains (Btu/d)	7464	22392	14928	Entire	Entire	Entire
Use to gain (%)	20.0	35.0	27.5	Entire	Entire	Entire
Exterior solar absorptance	0.5	0.8	0.6	Entire	Entire	Entire
Seasonal % annual load	90	99	95	Upper	Lower	Entire
Heating set point (°F)	60.0	75.0	68.0	Upper	Lower	Entire
Furnace efficiency (%)	60.0	80.0	70.0	Lower	Upper	Entire
Cooling set point (°F)	71.0	86.0	78.0	Lower	Upper	Entire
Cooling COP	2.5	3.5	3.0	Lower	Upper	Entire

Notes:

- Abbreviations and nomenclature used here are defined under "Nomenclature for Table 1" near the end of the paper.
- All explicit inputs are selected independently for each space heating and space cooling base-case scenario, except heating set point and furnace efficiency are selected for space heating cases only, and cooling set point and cooling COP are selected for space cooling cases only.

1. **Targeted High**: *Explicit inputs* were selected randomly from the portion of the *approximate input range* (upper or lower) that led to increased space conditioning energy consumption versus *nominal input* values. For inputs that have different effects in Las Vegas and Colorado Springs on the space conditioning loads (internal gains and solar absorptance), the entire range was used.
2. **Targeted Low:** *Explicit inputs* were selected randomly from the portion of the *approximate input range* (upper or lower) that led to decreased space-conditioning energy consumption versus *nominal input* values. For inputs that have different effects in Las Vegas and Colorado Springs on the space conditioning loads (internal gains and solar absorptance), the entire range was used.
3. **Fully Random**: *Explicit inputs* were selected randomly from the entire range for each variable.

Using the methodology described above, seven calibrated base-case scenarios were developed, each for space heating and space cooling scenarios, respectively (for a total of 14 scenarios):

- L200EX-C1, *targeted high* space conditioning use
- L200EX-C2, *targeted low* space condition use
- L200EX-C3, *fully random* selection, near-nominal space conditioning use
- L200EX-C4, *fully random* selection, high space conditioning use
- L200EX-C5, *fully random* selection, low space conditioning use
- L200EX-C6, *fully random* selection, mid-high space conditioning use
- L200EX-C7, *fully random* selection, mid-low space conditioning use.

The process for developing the *fully random* scenarios was to generate 20 sets of explicit inputs each for the space heating and space cooling cases (40 total sets). Each case was then simulated in EnergyPlus. The *fully random* cases were ranked according to annual space heating/cooling consumptions (sets of heating and cooling cases were considered separately). The cases with space heating/cooling consumptions corresponding to the closest to nominal, maximum, and minimum consumptions were selected for cases C3, C4, and C5, respectively.

For selecting the mid-high cases (C6), and mid-low cases (C7), mid-high and mid-low space heating/cooling consumption target values were calculated by averaging space heating/cooling consumptions according to:

$$(C6) = \frac{((L200EXP) + (C4))}{2}$$

$$(C7) = \frac{((L200EXP) + (C5))}{2}$$

where L200EXP is the result from the building physics test base case (Case L200EX-P) applying *nominal inputs*, and C4 and C5 are the maximum and minimum results, as selected above.

Further details about developing calibration scenarios are provided in Appendix F of Judkoff et al. (2010a).

Synthetic Utility Billing Data

Synthetic utility billing data are generated using the average of the results from the reference simulations for a given calibration scenario. Gas bills aggregate space heating and domestic hot water use. Electric bills aggregate space cooling and base electricity (lighting, appliances, etc.) use. Multiple calibration scenarios are applied, and 13 months of utility data (as requested by the BESTEST-EX Working Group) are provided for each scenario.

RESULTS

Example results and acceptance criteria for the building physics heating tests are included in Figure 3. Reference results were developed using:

- DOE-2.1E version JJ Hirsch PC 2.1En136 (*DOE-2 Reference Manual* 1981, *DOE-2 Supplement* 1994)
- EnergyPlus version 3.1 (*EnergyPlus Input Output Reference* 2009)
- SUNREL version 1.14 (Deru et al. 2002).

Example Acceptance Criteria

Figure 3 shows example acceptance range maxima and minima, indicated by “range” bars (shown adjacent to each set of example results bars for each test case). For each test case shown, a tested tool result is satisfactory if it falls within the greatest maximum and least minimum defined by the blue and green range bars. Example acceptance criteria include statistically based criteria (blue range bars) and alternative economic threshold criteria (green range bars). We developed an alternative economic threshold criterion because, for some sensitivity tests, the resulting statistically based acceptance ranges represent insignificant utility cost disagreements. These disagreements should not be cause for evaluating a tested software result as unsatisfactory. Judkoff et al. (2010b) discuss additional details about development and example application of example acceptance criteria.

Figure 3 shows only the results and acceptance ranges for the building physics test cases. Reference simulation results, example acceptance criteria, and randomly selected *explicit inputs* used in the reference simulations for the calibrated energy savings test cases are included in Judkoff et al (2011).

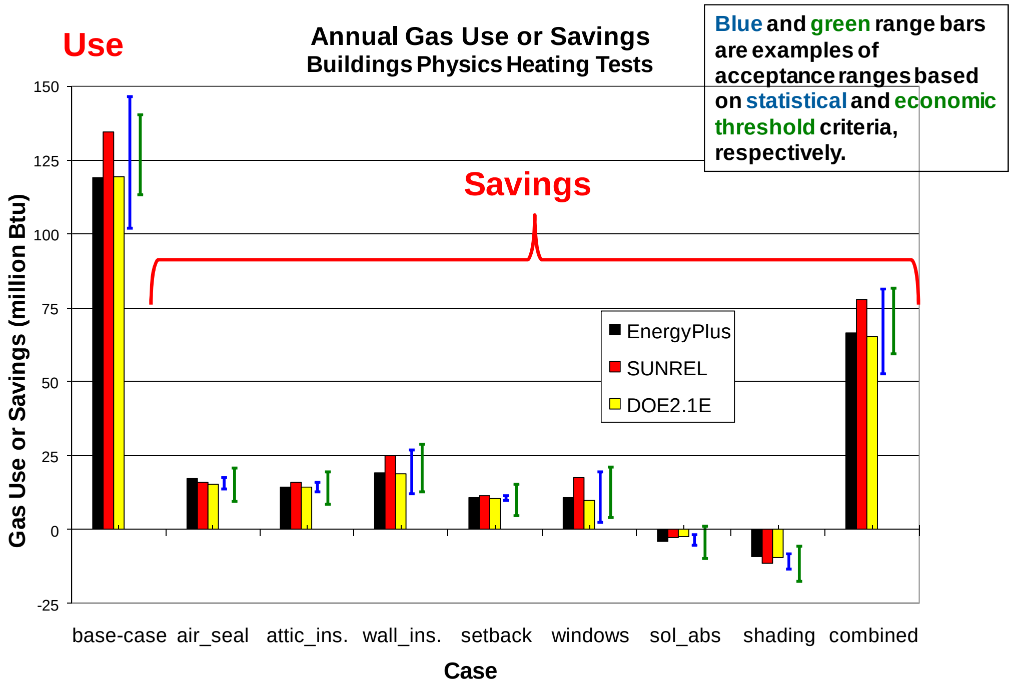
Improvements to Tested Software and Importance of Simulation Trials

As a result of the BESTEST-EX simulation trials, the working group participants documented eight software revisions and two input errors. The proprietary nature of participant programs did not allow disclosure of details. However, the participants indicated that the diagnostic logic associated with specific parameter variations of the test cases helped to isolate problems. NREL also clarified parts of the test specification related to the documented input errors. Therefore, the simulation trials were beneficial because they drove improvements to retrofit software and the test procedure.

Benefit of Calibration

BESTEST-EX working group participant results from a preliminary field trial were analyzed to assess the effect of calibrating pre-retrofit models to utility bills on the accuracy of energy savings estimates. The general approach was to compare the errors (difference between the average of participant simulation results and the average of the reference results) of energy savings predictions using models calibrated to pre-retrofit utility bills versus the errors of predictions using uncalibrated (building physics test) models. This statistical approach was one of a number of possible methods, and the results are specific to the base-case house and the approximate input ranges defined in the field test. Further details of the analysis are provided in Appendix G of the final report (Judkoff et al. 2010a).

Based on the preliminary analysis, energy savings predictions were generally improved by calibration. Improvement was not seen for every retrofit measure and calibration scenario, but calibration tended to improve predictions for:



Blue and green range bars are examples of acceptance ranges based on statistical and economic threshold criteria, respectively.

Figure 3 Reference results and example acceptance criteria for building physics heating tests (Judkoff et al. 2010b)

- Scenarios with a large difference between utility bills and the energy consumption predicted using an uncalibrated model (where a greater degree of calibration was required)
- Individual retrofit measures with robust energy savings (e.g., insulation in heating climate, windows in cooling climate)
- Combinations of retrofit measures that maximize energy savings.

The preliminary estimate of accuracy improvement attributable to calibration for combined retrofits in the 10 *fully random* selection scenarios ranged from \$39-\$234/y for heating and from \$34-\$61/y for cooling. Calibration had the least benefit for near-nominal scenarios, and most benefit for random high and random low scenarios. Further analysis is needed, as the benefit of calibration can be quantified using other methods.

CONCLUSIONS

A procedure (“BESTEST-EX”) was developed for testing the reliability of models that predict retrofit energy savings, including their associated calibration methods.

Accomplishments

The major accomplishment of BESTEST-EX is the development of a new methodology for evaluating energy retrofit software. This includes the testing of calibration methods, which use pre-retrofit utility billing data to calibrate pre-retrofit base case models or to adjust energy savings predictions.

Example procedures for establishing acceptance-range criteria were developed to evaluate tested program results, including statistically based and alternative economic threshold range setting procedures.

The development of BESTEST-EX improved some of the previous HERS BESTEST building physics test assumptions.

Preliminary results from industry participant simulation trials indicate that calibrating models to utility bills improves energy savings estimates in some circumstances.

The test procedure and example acceptance criteria were developed in an iterative process that enabled improvement of the test specification during the simulation trials and helped simulation trial participants to improve their software. Based on tested program results disagreements found during the simulation trials, simulation participants documented a total of eight model revisions and correction of two input errors. This indicates that BESTEST-EX is useful for improving models and calibration methods.

Future Work

The development of BESTEST-EX is planned as multigenerational, beginning with the Phase-1 test procedure. For further development of BESTEST-EX, NREL intends to add features that may include retrofit measures such as HVAC equipment, duct sealing, domestic hot water, lighting, appliances, foundation insulation, and others. Future test cases may include selected cross-referenced cases from

HERS BESTEST and other test procedures. NREL, in collaboration with the BESTEST-EX working group, is also investigating using empirical data from existing audited homes to estimate the accuracy of building energy simulation tools when used for modeling older poorly insulated buildings, and retrofits to those buildings. Based on this work, refinements to BESTEST-EX to better match empirical data, along with other refinements to the current cases, may be considered. BESTEST-EX has been submitted for consideration to be included (and maintained) in ASHRAE Standard 140. Appendix I of the full final report (Judkoff et al. 2010a) provides more detail about recommendations for future work.

ACKNOWLEDGEMENTS

For feedback and comments on the test procedure provided by BESTEST-EX Working Group participants during preliminary simulation trials, we wish to thank the following companies and individuals that ran simulations and responded to questionnaires: Apogee Interactive Inc.: Joel Gilbert, Eric Shewbridge, Lei Wang; Architectural Energy Corporation: Brian Christensen, Rob Salcido; Conservation Services Group: Bruce Harley, Mick Rookwood, David Weitz, Michael Blasnik (consultant); Florida Solar Energy Center: Philip Fairey; ICF International: Brian Dean, Mike L'Ecuyer; Oak Ridge National Laboratory: Mark Ternes; Performance Systems Development LLC: Chris Balbach, Ethan MacCormick, John McCartney, Greg Thomas.

Ren Anderson of NREL provided managerial support. Nick Long of NREL provided assistance with automating reference simulations. Paul Norton, formerly of NREL (now a private consultant) provided initial organizational and management support during project startup.

We appreciate the support and guidance of David Lee and Chris Early of DOE, and former DOE personnel (recently retired) including Ed Pollock, Terry Logee, and Lew Pratsch.

DEFINITIONS

Approximate input: an input for which *an approximate input range* has been defined. Also see *approximate input range.*

Approximate input range: the specified range of possible values for an *approximate input* that forms the basis uncertainty range for selecting *calibrated inputs* for the tested programs, and from which *explicit inputs* are randomly selected. Also see *calibrated input* and *explicit input.*

Calibrated input: input for tested programs that are determined based on specified *approximate input ranges* and *nominal input* values using calibration to obtain agreement with base-case reference utility billing data. Also see *approximate input range* and *nominal input.*

Explicit input*:* inputs for simulations used to develop reference utility billing data that are randomly selected from within specified *approximate input ranges*. Also see *approximate input range.*

Fully random: Calibration scenario where *explicit inputs* are selected randomly from the entire *approximate input range*. Also see *approximate input range* and *explicit input.*

Nominal input*:* an input value as specified for the building physics base case.

Targeted low: Calibration scenario where *explicit inputs* are selected randomly from the portion of the *approximate input range* (upper or lower) that leads to decreased space conditioning consumption versus *nominal input* values. Also see *approximate input range*, *explicit input*, and *nominal input.*

Targeted high: Calibration scenario where *explicit inputs* are selected randomly from the portion of the *approximate input range* (upper or lower) that leads to increased space conditioning consumption versus *nominal input* values. Also see *approximate input range*, *explicit input*, and *nominal input.*

NOMENCLATURE FOR TABLE 1

Entire: range defined by “Min” and “Max”

Lower: range defined by “Min” and “Nominal”

Min: approximate input range minimum value

Max: approximate input range maximum value

Seasonal % annual load: percentage of annual load used to define the length of a heating or cooling season

Upper: range defined by “Nominal” and “Max”

Use to gain: ratio of electricity or gas consumption to related internal gains

REFERENCES

ASHRAE Handbook Fundamentals. 2009. Atlanta, GA, USA: American Society of Heating, Refrigerating and Air-Conditioning Engineers.

ANSI/ASHRAE. 2007. ANSI/ASHRAE Standard 140-2007, *Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs*. Atlanta, GA, USA: American Society of Heating, Refrigerating and Air-Conditioning Engineers.

ANSI/ASHRAE. 2009. ANSI/ASHRAE Standard 189.1-2009, *Standard for the Design of High-Performance Green Buildings*. Atlanta, GA, USA: American Society of Heating, Refrigerating and Air-Conditioning Engineers.

ANSI/ASHRAE/IESNA. 2007. ANSI/ASHRAE/IESNA Standard 90.1-2007, *Energy Standard for Buildings Except Low-Rise Residential Buildings*. Atlanta, GA, USA: American Society of Heating, Refrigerating and Air-Conditioning Engineers.

BESTEST-EX Working Group. 2009–2010. Various BESTEST-EX Working Group discussions. January 2009–April 2010. Meeting notes available on request.

Bianchi, M., Judkoff, R., Neymark, J., Polly, B. 2010. “Putting Software Through the Test.” Presented at ACI, Austin, TX, USA, April 23, 2010.

Blasnik, M. 2009. M. Blasnik and Associates. Personal communication with J. Neymark, J. Neymark & Associates, Golden, CO, USA, March 27, 2009.

Deru, M.; Judkoff, R.; Torcellini, P. 2002. *SUNREL Technical Reference Manual*. NREL/BK-550-30193. Golden, CO, USA: National Renewable Energy Laboratory.

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

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

Eley, C. 2011. “COMNET Commercial Energy Services Network.” Presented at 2011 RESNET Building Performance Conference, Lake Buena Vista, FL, USA, February 28 – March 2, 2011.

EnergyPlus Input Output Reference. 2009. Urbana, IL: University of Illinois, and Berkeley, CA, USA: Lawrence Berkeley National Laboratory. http://apps1.eere.energy.gov/buildings/energyplus/pdfs/ InputOutputReference.pdf/. Last accessed January 2010.

European Union. 2002. *On the Energy Performance of Buildings*. Directive 2002/91/EC of the European Parliament and of the Council, Official Journal of the European Communities, December.

Internal Revenue Service. 2008. *Amplification of Notice 2006-52, Deduction for Energy Efficient Commercial Buildings*. IRS Notice 2008-40. Appears in Internal Revenue Bulletin No. 2008-14, April 7, 2008. Washington, DC, USA: U.S. Department of Treasury, Internal Revenue Service.

Judkoff, R. 1988. “Validation of Building Energy Analysis Simulation Programs at the Solar Energy Research Institute.” *Energy and Buildings*, 10(3):235. Lausanne, Switzerland: Elsevier Sequoia.

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

Judkoff, R.; Neymark, J. 2006. “Model Validation and Testing: The Methodological Foundation of ASHRAE Standard 140.” *ASHRAE Transactions* 112(2):367–376. Atlanta, GA, USA: American Society of Heating, Refrigerating and Air-Conditioning Engineers.

Judkoff, R. 2008. “Thoughts on Validation for Retrofit: Calibration/True Up.” PPT file included with personal communication with Ed Pollock, U.S. Department of Energy, April 2008.

Judkoff, R.; Neymark, J. 2009. “What Did They Do in IEA 34/43? Or How to Diagnose and Repair Bugs in 500,000 Lines of Code.” *Proceedings of Building Simulation 2009*. July 27–30, Glasgow, Scotland. International Building Performance Simulation Association.

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

Judkoff, R.; Polly, B.; Bianchi, M.; Neymark, J. 2010b. *Example Procedures for Developing Acceptance-Range Criteria for BESTEST-EX*. NREL/TP-550-47502. Golden, CO, USA: National Renewable Energy Laboratory.

Judkoff, R.; Polly, B.; Bianchi, M.; Neymark, J.; Kennedy, M. 2011. *Building Energy Simulation Test for Existing Homes (BESTEST-EX): Instructions for Implementing the Test Procedure, Calibration Test Reference Results, and Example Acceptance Range Criteria.* NREL/TP-500-52414. Golden, CO, USA: National Renewable Energy Laboratory.

Judkoff, R.; Wortman, D.; O’Doherty, B.; Burch, J. 2008. *A Methodology for Validating Building Energy Analysis Simulations*. NREL/TP-550-42059. Golden, CO, USA: National Renewable Energy Laboratory. Based on unpublished report of 1983 with same authors and title, previously referenced as SERI/TR-254-1508.

Kotz, S.; van Dorp, J. R. 2004. *Beyond Beta, Other Continuous Families of Distribution with Bounded Support and Applications.* Hackensack, NJ, USA: World Scientific Publishing Co.

Neymark, J.; Norton, P. 2009. “Accuracy Tests for Software that Model Retrofits in Existing Homes”. Presented at RESNET, New Orleans, LA, USA, February 16, 2009.

RESNET 2007. *Procedures for Verification of International Energy Conservation Code Performance Path Calculation Tools*. RESNET Publication No. 07-003. September 2007. Oceanside, CA, USA: Residential Energy Services Network, Inc. www1.resnet.us/standards/iecc/procedures.pdf.