THE “ETNA BESTEST” EMPIRICAL VALIDATION DATA SET

J. Neymark1, P. Girault2, G. Guyon2, R. Judkoff3, R. LeBerre2, J. Ojalvo2, P. Reimer2
1J. Neymark & Associates, Lakewood, Colorado, United States. neymarkj@msn.com
2Electricité de France Research & Development, Moret sur Loing, France
3National Renewable Energy Laboratory, Golden, Colorado, United States

ABSTRACT
This paper describes a new data set appropriate for empirical validation of whole-building energy simulation software. The data has been created using artificial and natural climate configurations in the empirically characterized ETNA (Essais Thermiques en climat Naturel et Artificiel) test cells of an electric utility in France. This data set includes parametric variations based on the Building Energy Simulation Test and Diagnostic Method (BESTEST) comparative test methodology developed at a national laboratory in the United States.

INTRODUCTION
This paper summarizes recent collaborative work on empirical validation of whole-building energy analysis software (Neymark et al 2004). The test specification and empirical data has recently been released into the public domain (but no technical support for its use provided by the sponsor).

Background
There are only a few ways to evaluate the accuracy of a whole-building energy simulation program (Judkoff et al. 1983; Judkoff 1988):

- **Empirical Validation**—in which calculated results from a program, subroutine, algorithm, or software object are compared to monitored data from a real building, test cell, or laboratory experiment.

- **Analytical Verification**—in which outputs from a program, subroutine, algorithm, or software object are compared to results from a known analytical solution or a generally accepted numerical method for isolated heat transfer mechanisms under very simple and highly constrained boundary conditions.

- **Comparative Testing**—in which a program is compared to itself or to other programs that may be considered better validated or more detailed and, presumably, more physically correct.

Advantages and disadvantages of these techniques are described elsewhere (ASHRAE 2005; ASHRAE 2004; Judkoff 1988; Judkoff and Neymark 1995; Judkoff et al 1983; Neymark and Judkoff 2002). The above-cited literature indicates that work on comparative testing (including analytical verification testing where possible) is becoming relatively well developed in the area of testing models used for analysis of the building thermal fabric. However, work on developing usable and useful empirical validation test procedures is in an earlier phase of development. This is because empirical validation tests are more expensive and difficult to develop than comparative or analytical verification tests. It is also worth noting here that comparative (software-to-software) tests are good for diagnosing errors, but have no truth standard. Because there is no truth standard in comparative tests it is possible for a wide range of results disagreements based on physical model disagreements to occur without the presence of programming errors. Therefore, it is important to develop empirical validation tests where possible.

Previous major efforts regarding empirical validation of building envelope and thermal fabric models applied within whole-building energy analysis software include works by:

- National Renewable Energy Laboratory (NREL), US (Judkoff et al 1983)
- The PASSYS Project (Jensen 1989)
- IEA SHC Task 22 (Moinard and Guyon 1999).

The NREL work includes preliminary test specifications for projects that ran during the early 1980s. Some of the work was completed (e.g. Judkoff, Wortman and Burch 1983), and some of the work could not be completed because of later-year funding cuts. The NREL work includes important contributions regarding empirical validation test philosophy (including methodology for measurement practices), a detailed preliminary test specification (including description of test apparatus, measurement procedures, sensor installations, and data acquisition...
techniques), and methodology for simulation-to-data comparison.

The PASSYS empirical-validation project involved analysis of the ability to model the PASSYS exterior reference wall (which includes a window) over a wide range of climate conditions – i.e., using similarly designed test cells located throughout Europe. Empirical validation test results focused on the ability to predict zone temperature. The validation methodology was only applied to one simulation program. Differential (parametric) sensitivity analysis was performed with simulations to identify possible causes of differences between simulations and measured data. Jensen (1993) gives conclusions about some of the work for a test cell located in Stuttgart, Germany noting that because of errant or lacking measurements of local climate conditions, reasons for observed differences between measured and simulated results may remain hidden.

IEA SHC Task 12/ECBCS Annex 21 included a study using test cells in the United Kingdom. Side-by-side configurations were run among three test cells, where only the south facing façades of the rooms differed – two had different glazings and the third was opaque. Data included measurements of overall heat loss coefficient of the test cells. The study included 25 detailed whole-building energy simulation programs from 8 countries, but only a few programs were able to produce estimates of both energy consumptions and air temperatures within the estimated uncertainty bands of the experimental results. Simulation modelers identified the following differences between features of the experiment and the assumptions made by their programs: the dynamics and output of the heater; internal heat transfer coefficients; and internal air movement and stratifications. In particular the heater’s large radiant output component and near-window location differed from common simulation assumptions of purely convective heater output and well-mixed zone air.

The IEA SHC Task 22 work included a study of two experiments using test cells in France. (That work was the immediate predecessor to the project that is the primary topic of this paper.) The experiments tested the ability of models to predict temperature and energy consumption. The tests compared experimental results using a heat source meant to be an ideally convective heater located in the center of one test cell, while the other test cell used a more typical radiative heat source located under the south window. The tests were carried out in natural climate conditions, with dynamically varying internal gains in the experiment with floating zone temperature. Ten simulation programs were involved in the project. Comparison of simulation results to experimental data revealed the following issues:

- All programs predicted 10-30% lower energy consumption than the experiment, indicating a problem with characterizing the overall UA-value of the test cells
- Uncertainty regarding what film coefficients to use
- Measurements are more sensitive to solar radiation than the simulation predictions.

Recommendations for further work included:

- The global UA-value and other fundamental heat transfer properties of the test cells should be characterized experimentally
- The test specifications require a more precise definition of measured temperatures and thermostat controls
- For more detailed models a need exists to better understand the test-cell electric heaters heat distribution characteristics
- Further empirical test cell experiments are needed to expand the range of variables (parameters) that can be evaluated versus measured data, and to isolate the validity of specific algorithms applied in the simulation models.

The Task 22 project also included development of mathematical data analysis techniques that were applied to attempt to identify the importance of various thermal models in causing results disagreements. (Palomo and Guyon 2002)

The common feature in all of this work is that in general the empirical validation data were for an aggregated set of thermal analysis models, with the exceptions that heater-type effects were examined in the ETNA/Task 22 work, and glazing model effects were isolated in the Task 12 work. For these studies, if simulations results disagreed with experimental results there is only limited means for directly diagnosing the sources of the disagreement. Additionally, simulations could give results that agree with experimental results within the uncertainty of the experiment, but could still have compensating disagreements from individual algorithms used within the overall model.

Therefore, we decided to try to develop an empirical validation procedure that also incorporates the enhanced diagnostic capabilities of previously developed comparative test procedures. Such empirical validation tests would trace the validity of specific models used for thermal analysis of the building thermal fabric. A good comparative-test framework for such cases is the IEA Building Energy Simulation Test (BESTEST) and Diagnostic Method (Judkoff and Neymark 1995).
Objective

The objective of this work is to develop an empirical validation test suite that can perform BESTEST-type diagnostics with a truth standard defined by the uncertainty of real empirical data. Such a test specification tests the validity of many individual physical models used in whole-building energy simulation software.

TEST SUITE SUMMARY

Three important aspects of an empirical validation test specification include: the empirical test facility, the data collected from the empirical test facility, and the final test specification (which describes the test facility and test data to the user). A weakness in any of these aspects must, by definition, propagate through the entire empirical validation test suite. Each of these three aspects is summarized in turn, with further details available in the full test specification (Neymark et al 2004).

Empirical Test Facility

The test cell laboratory is called “ETNA” for “Essais Thermique en climat Naturel et Artificiel,” which translates as Thermal Studies in Natural and Artificial Climate. The facility is in a rural environment 75 km southeast of Paris, France, without any obstacle (detached shading) on the South façade over a length of 100 meters. From previous empirical validation work during IEA SHC Task 22 (described above), this test facility was known to be highly controllable (Guyon and Moinard 1999) and served well for these experiments.

Figures 1 and 2 show plan and elevation-section views of the test cells in their natural climate configuration. The configuration of the test facility consists of a building that contains two identically designed and oriented 16 m² test cells separated by and surrounded by individually controllable fixed guard zones (i.e., each boundary of each test cell is controlled). The “south” wall and windows actually face 30° West of South. Test cell windows have a set back depth from the exterior wall surface of 0.32 m. The south wall can also be bounded by a mobile thermal guard zone. The mobile thermal guard makes it possible to carry out tests under natural or artificial climatic conditions.

Ventilation flow rates within each test cell are individually controllable and measurable. The test cells include advanced measurement and data acquisition capability, and employed full-time engineering staff during the project.

Similarly designed base-case heating systems were installed in both of the test cells. The heating systems are designed to be purely convective and to provide well-mixed air within each zone (to emulate common simulation assumptions). The Heater with Diffusers (HWD, shown in Figure 3) is a custom-built system installed in the center of the test cell that includes intake ducts located near the ceiling, and four supply-air diffusers near the floor. The diffuser orientation is perpendicular to the floor with diffusers oriented at angles of 0°, 90°, 180° and 270° relative to the test cell length axis. Contained within the ductwork are the air distribution fan and the heater. Within the HWD, sensors are provided to measure: heating electric power, fan electric power, and fan airflow rate. The air temperature of each test cell is
calculated from the average of 20 air temperature measurements at various locations within each test cell. The HWD is equipped with a Proportional-Integral-Derivative (PID) thermostat controller that allows control of the average test cell air temperature within a range of 0.1°C.

Artificial climate tests were included primarily for the purpose of characterizing the thermal behaviour of the test cells and the heating system. Such tests included steady-state heat-loss calorimetry to evaluate overall building loss coefficient as well as the overall thermal conductance of each individual wall. Additional sensors were installed for evaluating interior surface heat transfer coefficients – measured distribution airflow rates of the heating system may also be used for estimating convective surface coefficients. Decay and heat-up tests were performed for evaluating combined internal and interior surface thermal capacitance effects.

Other artificial climate test configurations include tests for modeling:

- Internal gains (interior lights on/off)
- Typical French convective heater versus typical French radiative heater
- Controlled Mechanical Ventilation.

Data Preparation

Four separate sets of hourly data were compiled based on the four time periods during which the data were acquired (during the years 2000 and 2001). A complete list of data sensors, their locations, and further details of data acquisition are provided in the test specification. All data files are in spreadsheet (.XLS) format and each line contains data for one hour. Hourly test cell data were produced based on original data acquired at 5-minute intervals. Hourly weather data, including all solar radiation data, were produced based on original data acquired at 1-minute intervals. All data were integrated using a preceding-hour time convention to produce a 1-hour time step. This time convention is such that on any given day, hour 2 is actually the data averaged during the preceding hour period from 01:00 to 02:00. For natural climate cases the south guard zone was removed, and actual weather data is provided. Only the hourly-integrated data files are available in the public domain.

A significant part of the effort, after gathering the data itself, was comprehensive data checking with correction of occasional errant or missing data as needed. Data checking and correction covered the test cell data, non-solar weather data, and solar radiation data including incident radiation on the building itself and solar radiation transmitted through windows. Details of weather data checking, documentation of corrections, and correction techniques are provided with the test specification.
Surface UA Calculation Summary and Adjustments to Listed Material Properties

During previous work of IEA SHC Task 22 (Guyon and Moinard 1999), the project participants determined that the overall conductance (UA) of each test cell was substantially greater than what would be predicted from given material properties using typical 1-dimensional conduction analysis. For this work we planned initially to address that disagreement by making calorimetric conductance measurements of only the overall test cell and the south wall. However, because the initial overall conductance measurement indicated about 25% disagreement between the measured overall test cell UA and the overall test cell UA calculated from given material properties, we decided to attempt to evaluate the UA of each surface to isolate the source of the overall UA disagreement, and thereby be able to adjust the UA for each surface accordingly. UA values for each surface were further useful because it was necessary to maintain the guard zones at 10°C to ensure heating is always needed in the test cell in the natural climate cases (no cooling system was provided inside the test cell), and because there is some variation among the steady-state guard zone temperatures relative to each other. The technique for determining conductances of the overall test cell and each bounding surface is included in the test specification. These empirically determined UA values implicitly incorporate all as-built heat transfer effects including thermal bridging, air leakage, multi-dimensional conduction, etc. For the purpose of translating the empirically derived UA values to simulationists, the thermal conductivities of specific material layers were adjusted, as documented in the test specification.

Individual surface UA analysis showed that the surface with greatest deviation from UA estimated by one-dimensional conduction analysis was the floor, where measured UA versus initial calculated UA was 66% greater in Cell A and 88% greater in Cell B. Consistent with this variation is that the floor has the most complicated construction of all the surfaces bounding the test cell.

Recommended Additional Analysis

A number of analysis details were not possible to complete within the time frame of the project, these include:

- Refinement of conductivity adjustments to the north wall of Cell A for one of the data sets, which may be attributable to a temporary leak in the door seal.
- Comparison of conductivity adjustments of walls with windows by adjusting only the wall conductance versus the current method of adjusting both the wall and window conductances
- Adjustment of values for appropriate interior surfaces’ density and specific heat based on analysis of results of the three test cases that were performed specifically for that purpose
- Calculation of interior surface coefficients based on detailed measurements
- Estimation of exterior surface coefficients based on air velocity measurements
- Calculation of in-situ window optical properties based on existing measurements
- Correlation of existing flow meter data in ventilation test cases with existing tracer gas results
- Experimental uncertainty analysis – individual sensor uncertainties are known, but their effect on uncertainty of calculated values was not analyzed.

More details recommending how to conduct these analyses are provided in the full test specification.

PRELIMINARY SIMULATION RESULTS

Preliminary simulation results applying SERIRES/SUNCODE (Wheeling, Palmiter, and DeLaHunt 1997) indicate the importance of using empirically adjusted thermal properties based on the analysis provided in the test specification.

The current set of results shown in Figures 4 through 8 indicates the effect of adjustment of selected surface conductivity values (and therefore UA values) related to steady-state heat transfer. This adjustment is based on the results of empirical test cell characterization experiments described above. In Figures 4 through 6, it is evident that the conductivity adjustments based on the artificial climate test cell characterization cases result in improved agreement between simulation results and empirical data for the natural climate cases. Agreement is best for the test without solar gains (see Figure 4), and agreement is still much improved when solar gains are added (see Figures 5 and 6).

The effect of thermostat set back (i.e. more robust thermal dynamics) is included in the results of Figures 7 and 8 for the HWD and typical French convector respectively. In these figures, the improvement to agreement from adjusting thermal conductivity is still evident, but the more robust dynamics appear to cause greater disagreement between simulation results and empirical data than results with less robust dynamics of the preceding figures. Modeling robust dynamic behaviour requires accurate characterization of thermal mass and accurate characterization of the interaction between zone air and thermal mass. Before concluding there is a problem with the simulation model, test specification thermal mass and convective surface
Coefficient characterization analysis should be performed.

Differences in results of Figures 7 and 8 for the HWD versus the typical French convector just after each thermostat setback period indicate that to bring the zone temperature up to the controller set point the French convector (which is located below the south window) must maintain peak output for longer duration than the HWD. Duration of peak heating level in simulation results – which assume idealized control, pure convective air distribution, and well-mixed air – are more consistent with HWD behaviour.

Field trials of the test data beyond the preliminary work discussed above have not been conducted. To obtain minimally biased field trial results, it is best to conduct the initial round of field trials of a software test specification with key output results as blind to field trial participants. Specific recommendations for how to conduct such field trials are included with the test specification.
Conclusions

A new data set and test specification appropriate for empirical validation of whole-building energy simulation software were developed. The purpose of the test method is to evaluate the technical capabilities and range of application of simulation programs, by comparing simulated results with results from empirical tests. The experimental test sequences are designed so that the effects of various heat transfer phenomena can be isolated. This test method is intended to allow for a true test of the validity of many individual physical models used in whole-building energy simulation software – as well as the validity of the aggregate combination of such models – within the uncertainty of the empirical data.

The experimental data has been created using artificial and natural climate configurations in the empirically characterized ETNA test cells. This data set includes parametric variations based on the BESTEST comparative test methodology. Within the data set, parametric variations in a natural climate configuration include: dynamic thermal diffusion, solar gains, thermostat setback, variation of interior surface convective coefficient, variation of heater type, variation of thermal mass, and interactions of these. Parametric variations in an artificial climate configuration include tests for the ability to model: outside air ventilation/infiltration, internal gains, and typical wall mounted “convective” and “radiative” heaters versus a heater designed for ideal pure convective output with uniform mixing of zone air.

Considerable effort was expended to identify thermal characteristics of the test cells using empirical data. Data was gathered in the artificial climate configuration to empirically characterize: steady-state overall building heat loss coefficient; steady-state thermal conductance of individual walls, floor, ceiling, and windows; and internal thermal capacitance. Measurements were also made with the objectives of estimating interior convective surface coefficients and empirically characterizing incidence-angle-dependent window optical transmittance. Further effort has been expended to make the user’s manual for the test suite as detailed and clear as possible, consistent with other BESTEST user’s manuals. This effort includes thorough review of the data, and documentation regarding data checking and correction.

Preliminary results indicate that the test procedure is usable and that the empirical validation data can be reasonably simulated for natural climate cases with solar gains. Comparison of experimental data with simulation results for test cases with thermostat setback indicate that further analysis should be applied regarding characterization of the test cells’ thermal capacitance properties and interior convective surface coefficients; this analysis can be accomplished with the existing data.

Recommendations

In general we feel that creation of this data set represents an important step forward in the field of empirical validation of whole-building energy simulation software. However, in a more ideal world of greater funding, we may have considered first writing the test specification, and then designing and constructing the test facility. If that had been the case we may have constructed the test facility such that windows were mounted flush with the wall exterior surfaces; this would have allowed more robust solar and shading model diagnostics by including external shading devices in a separate test case. We may also have built the test cell floor with a more uniform construction.

Closing Remarks

There is an obvious need for full simulation field trials of ETNA BESTEST. Unfortunately the initial sponsor cannot provide technical support related to the test specification.

The test specification and the test data, both of which are in the public domain may be obtained from the corresponding author upon request. If you are a software developer or other software user who has not yet run the IEA BESTEST comparative cases (Judkoff and Neymark 1995; ASHRAE 2004), we recommend running those test cases before implementing the ETNA BESTEST.

Considering the relatively large amounts of funding and labor resources required to develop a usable empirical validation data set and test specification, we feel that as a minimum ETNA BESTEST should be reviewed by researchers undertaking (or considering to undertake) a project related to empirical validation of whole-building energy simulation software. This has already begun to happen; for example, researchers on two new and separate IEA projects involving empirical validation have requested and received copies of the test procedure. (Manz 2004; Beausoleil-Morrison 2004)

Acknowledgments

We would like to thank Jay Burch of NREL for his initial involvement in the early stages of test specification development.
REFERENCES


T. Wheeling, L. Palmiter, and J. DeLaHunt. (1997). SERIRES/SUNCODE-PC 6.1 User’s Manual. Seattle, Washington : Ecotope. The version used for simulations documented here is a customized version by M. Kennedy (of Port Townsend, WA) that allows input of external files for detailed hourly inputs (e.g. set points, etc.), and more flexible output analysis.