Literature Review on Calibration of Building Energy Simulation Programs: Uses, Problems, Procedures, Uncertainty, and Tools

T. Agami Reddy, PhD, PE
Member ASHRAE

ABSTRACT

Calibrated simulation is the process of using an existing building simulation computer program and "tuning" or calibrating the various inputs to the program so that observed energy use matches closely with that predicted by the simulation program. The two primary reasons for adopting this approach is that it allows (1) more reliable identification of energy savings and demand-reduction measures (involving equipment, operation, and/or control changes) in an existing building and (2) increased confidence in the monitoring and verification process once these measures are implemented. Historically, the calibration process has been an art form that inevitably relies on user knowledge, past experience, statistical expertise, engineering judgment, and an abundance of trial and error. Despite widespread interest in the professional community, unfortunately no consensus guidelines have been published on how to perform a simulated calibration using detailed simulation programs. ASHRAE initiated a research project (RP-1051) intended to cull the best tools, techniques, approaches, and procedures from the existing body of research and develop a coherent and systematic calibration methodology that includes both "parameter estimation" and determination of the uncertainty in the calibrated simulation. This paper provides a pertinent and detailed literature review of calibrated simulation techniques, describing their strengths, weaknesses, and applicability, thus serving as a precursor to reporting the results of the research project in subsequent papers.

BACKGROUND

Calibration as Part of ECM Identification and Monitoring and Verification

The oil shock of 1973 triggered a flurry of energy conservation activities especially by federal and state agencies. This led to the widespread initiation of demand-side management (DSM) projects especially targeted to residential and small commercial building stock. Subsequently, in the 1980s, building professionals started becoming aware of the potential and magnitude of energy conservation savings in large buildings (office, commercial, hospitals, retail, etc.). DSM measures implemented included any retrofit or operations practice, usually some sort of passive load curtailment measure during the peak hours such as installing thermal storage systems, retrofits to save energy (such as delamping, energy efficient lamping, changing constant air volume [CV] systems into variable air volume [VAV], demand meters in certain equipment [such as chillers], and energy management and control systems [EMCS] for lighting load management). The drastic spurt in activity by energy service companies (ESCOs) led to numerous papers being published in this area (such as that by Schuldt and Romberger [1998] and various other publications reviewed later in this paper) and standard development efforts by organizations such as ASHRAE (2002) and USDOE (IPMVP 2001). During the last few years, electric market transformation and utility deregulation have led to a new thinking toward more proactive load management of single and multiple buildings (Reddy and Norford 2004; Norford and Reddy 2004). Though several facilities do implement such proactive load management practices, these are yet to achieve a sufficient level of maturity at this time.

T. Agami Reddy is a professor in the Civil, Architectural, and Environmental Engineering Department, Drexel University, Philadelphia, Pa.

The proper implementation of DSM measures involved first the identification of the appropriate energy conservation measures (ECMs), and then assessing their impact or performance once implemented. This need resulted in monitoring and verification (M&V) activities acquiring a key importance. Typically, retrofits involved rather simple energy conservation measures in numerous more or less identical buildings. The economics of these retrofits dictated that associated M&V also be low-cost. This led to utility bill analysis (involving no extra metering cost) and even analyzing only a representative subset of the entire number of retrofitted residences or small commercial buildings. In contrast, large commercial buildings have much higher utility costs, and the HVAC&R devices are not only more complex but more numerous as well. Hence, the retrofits were not only more extensive but the large cost associated with them justified a relatively large budget for M&V as well. The analysis tools used for DSM projects were found to be too imprecise and inadequate, which led to subsequent interest in the development of more specialized inverse (or data-driven) modeling and analysis methods that use monitored energy data from the building along with other variables, such as climatic variables and operating schedules.

A widely used technique is the calibrated simulation approach, whereby one uses an existing building simulation computer program and "tunes" or calibrates the various physical inputs to the program so that observed energy use matches closely with that predicted by the simulation program (Stein 1997). Once that is achieved, more reliable and insightful predictions could be made than with statistical approaches. Initial attempts, dating back to the early 1980s, involved using utility bills with which to perform the calibration. Since there are many more input parameters that need to be specified (and, hence, tunable) in a detailed simulation, this severely restricted the accuracy and reliability of the calibration process. Subsequently, researchers started using instantaneous or short-term measured data of certain key end-uses in order to improve the accuracy. Finally, the use of hourly data for a month or season, or even a year, was adopted in an effort to more accurately represent building dynamics (in terms of diurnal scheduling and equipment control).

Uses of Calibrated Simulation

In practice, a large number of energy professionals are involved in performing calibrated simulations, and many more profess an active interest in this area. Calibrated simulation can be used for the following purposes:

a. To prove/improve specific models used in a larger simulation program (Clarke 1993).

b. To provide insight to an owner into a building's thermal and/or electrical diurnal loadshapes using utility bill data (Sonderegger et al. 2001).

c. To provide an electric utility with a breakdown of baseline, cooling, and heating energy use for one or several buildings based on their utility bills in order to predict impact of different load control measures on the aggregated electrical load (Mayer et al. 2003).

d. To support investment-grade recommendations made by an energy auditor who has to identify cost-effective ECMs (equipment change, schedule change, control settings, etc.) specific to the individual building and determine their payback.

e. For M&V under one or several of the following circumstances (ASHRAE 2002): (1) to identify a proper contractual baseline energy use against which to measure energy savings due to ECM implementation, (2) to allow making corrections to the contractual baseline under unanticipated changes (creep in plug load, changes in operating hours, changes in occupancy or conditioned area, addition of new equipment, etc.), (3) when the M&V requires that the effect of a end-use retrofit be verified using only whole-building-monitored data, (4) when retrofits are complex and interactive (e.g., lighting and chiller retrofits) and the effect of individual retrofits need to be isolated without having to monitor each subsystem individually, (5) either pre-retrofit or post-retrofit data may be inadequate or not available at all (e.g., for a new building or if the monitoring equipment is installed after the ECM has been implemented), and (6) when length of post-retrofit monitoring for verification of savings needs to be reduced.

f. To provide facility/building management services to owners and ESCOs with the capability of implementing (1) continuous commissioning or fault detection (FD) measures to identify equipment malfunction and take appropriate action (such as tuning/optimizing HVAC and primary equipment controls) and (2) optimal supervisory control, equipment scheduling, and operation of a building and its systems, either under normal operation or under active load control in response to real-time price signals.

Note that fulfilling purposes (d) and (e) above automatically fulfills (b) and (c).

OBJECTIVES

Standard building simulation programs typically produce electrical demand and consumption data as a program output. When modeling existing plants, if the results do not match actual monitored data, the programmer will typically "adjust" inputs and operating parameters on a trial-and-error basis until the program output matches the known data. This "fudging" process often results in the manipulation of a large number of variables that may significantly decrease the credibility of the entire simulation (Troncoso 1997). The problem is further compounded by the fact that it is a dynamic matching over one year, not a static one at one condition or time frame (such as in classical optimization problems). A final problem is a lack of proper specification of model inputs. In order to run a simulation model, one needs to know, over and above the climatic data, the exact specification of the numerous parameter set, such as envelope details, building scheduling, type, exact layout, size and operating schedules of secondary and primary systems, and specifics of the control algorithms. Further, one needs to know the exact type of model parameters used (e.g., the numerical values of the coefficients of the polynomial model for the chiller). Determination of all of these parameters and variables from calibration is subject to the "identifiability problem," i.e., the inability to identify a unique solution set since the problem is underdetermined. Hence, the analyst relies on personal judgment and the circumstances specific to these problems (for example, some of the model inputs may be better specified than others) in order to first reduce the order of the model by selecting "best-guess" values of certain parameters and then to calibrate the remaining parameters.

Thus, the main reservations with the widespread use of calibrated simulation is that it is labor intensive, time consuming, requires a high level of user skill and knowledge in both simulation and practical building operation, and is *highly dependent on the personal judgment of the analyst doing the calibration*. The last reservation implies that this is more an art than a science, and, worse, that the results are analyst-specific.

ASHRAE developed Guideline 14 (ASHRAE 2002) for determining the appropriate methods for analyzing energy and demand savings from energy conservation retrofits. Guideline 14 has defined how energy savings are to be measured and characterized following any one of four analysis methods, one of which is calibrated simulation. Because of its broader scope, Guideline 14 defines *uncertainty* in estimating savings as the standard of comparison between different energy savings calculations. Simple formulas are proposed, anchored in basic statistics, to quantify such uncertainty. Discrepancy between measured data and simulation results, therefore, has a direct bearing on uncertainty. The greater the discrepancy, the more uncertain are the savings predictions of the simulation, however accurate and detailed. Unfortunately, although Guideline 14 provides procedures for using calibrated simulation, it does not provide a methodology to calibrate a simulation to measured conditions.

Another major problem with reporting simulation accuracy rests with the calculation procedures, which have been reported in the previous work. Typically, when a model is established as being calibrated (i.e., the user states that the "accuracy" for electricity use is approximately "5% per month"), the author does not reveal the techniques used other than stating that the final result is "calibrated" or "validated." Hourly or daily error values are seldom reported. Even in cases when error estimates are presented, the methods and equations used to obtain the comparisons are not. Therefore, because of the manifest lack of uniformity and abundance of confusion in calibrating simulations to actual data, Guideline 14 identified the development of consensus procedures for comparing the results of computer simulations to measured data as an important task.

In an effort to address the above-described issues, ASHRAE initiated a research project (RP-1051) whose broad intent was to cull the best tools, techniques, approaches, and procedures from the existing body of research and develop a coherent and systematic calibration methodology that includes both "parameter estimation" and determination of the uncertainty in the calibrated simulation. Such a methodology would be of great benefit to the entire profession since it would allow more transparency, reproducibility, and uniformity in how calibrated simulations are performed. Using this research (Reddy et al. 2006), the objective of this paper is to provide a pertinent and detailed literature review of calibrated simulation techniques, describing their strengths, weaknesses, and applicability. This would set the stage for reporting the research results of RP-1051 in subsequent papers.

CALIBRATED SIMULATION PROCESS

Several of the concepts, issues, and proposed procedures pertinent to the building design process are also applicable to the calibration process. A number of overview and status-review papers on building simulation programs have been written over the years, which are further reviewed by Reddy et al. (2006). A literature search revealed a number of publications on calibrated simulation of building energy simulation programs (about 30 references have been identified). We first provide a general background of the steps involved in a calibration process based on three general publications. Next, adopting a classification proposed by Clarke et al. (1993), we have divided these references into four classes. Though the papers that deal with how to calibrate detailed simulation programs such as DOE-2 (Winklemann et al. 1993) are the primary focus of ASHRAE RP-1051, we have also reviewed relevant papers that deal with calibrating simpler computer programs such as SEAP (Knebel 1983) and subsequent offshoots of this approach, as proposed by Liu and Claridge (1998).

General References

In this category we include publications of a general nature that discuss why calibration is done to detailed building energy simulation programs, the advantages and difficulties of doing so, and some general guidelines as how to perform calibration. The last aspect, though instructive to general understanding of the entire process, is not covered in detail, so specific technical details are lacking. Other than the several sections from ASHRAE Guideline 14 (ASHRAE 2002) that deal with the calibrated simulation approach for baselining energy use in M&V projects, three publications have been identified that fall in this category.

Kaplan et al. (1992) provide an overview of guidelines that distill the experience gained from the large-scale research and demonstration of the Energy Edge project funded by Bonneville Power Administration. The intent of the guidelines was to advise conservation program managers on the use of modeling and to improve the accuracy of design-phase computer models as applicable to commercial buildings. Relevant to RP-1051 research is the fact that these guidelines spell out the main sources of discrepancies between savings predicted using calibrated simulation, which simulation inputs have the greatest impact on the results, and what modelers can do to minimize error in their work. The study is geared toward the small commercial building sector.

The well-illustrated report of about 20 pages by Stein (1997) starts the section on techniques for calibrating simulation models by stating that it is more an art than a science. He describes the technique developed by the Energy System Laboratory of Texas A&M University, which is largely graphical and heuristic and involves the use of submetering as well as hourly whole-building data. The calibration technique essentially consists of four steps: (1) collect data, (2) enter data and run simulation, (3) compare simulation model output to whole-building data, and (4) decide on whether desired accuracy has been achieved.

The book by Waltz (2000) provides an excellent introduction, sound heuristic advice, and operating guidelines of when calibrated simulation is useful and how to proceed in order to perform a calibrated simulation. The book also contains numerous anecdotal examples and case studies in the author's professional experience that are illuminating and good reading. The book is targeted toward energy professionals adopting this approach as a tool for investment-grade audits in the framework of performance contracting. It also clearly explains and dispels myths about calibrated simulation and also provides clear and simple guidelines as to how to model energy conservation measures. Though the examples and discussion are primarily geared toward the spreadsheet type of simulation approach, the thought process involved during calibration is equally appropriate for detailed simulation programs as well. The book also gives numerous actual case study examples of how proprietary detailed fixed-step simulation programs have been calibrated by the author. Extensive appendices contain standards for performing energy audits, pre-survey checklists, field survey data-gathering checklists, an ECM work-up sheet, and various other types of forms. The book is more detailed in its treatment of calibrated simulation than the report by Stein (1997). The various steps needed to be undertaken for calibrating a simulation model are broken up into more discrete steps that are explained in some detail. The book has an excellent chapter on critiquing output and model calibration, by which specific processes of checking whether the output is reasonable or not are implied.

Calibration Based on Manual, Iterative, and Pragmatic Intervention

Most studies to date are based on this approach with, however, different nuances. These nuances involve complementing utility bill data with walkthrough audits and spot and/or short-term monitoring of certain key end-uses. Though some graphical plots are used in this approach, these are essentially simple. Also, the procedure adopted to iteratively calibrate the parameters is user-specific and largely heuristic.

The earliest study on calibration of detailed building energy simulation programs we could identify was that by Diamond, Hunn, and other workers from Los Alamos National Laboratory in the late 1970s. Several reports were written that are summarized in a paper by Diamond and Hunn (1981). This paper describes the results of a project where (1) seven commercial building types were selected (restaurant, single-floor office building, retail store, hospital, multi-floor office building, school, and solar heated and cooled building), (2) monthly utility bills for an entire year were gathered, (3) several competent contractors were identified and given detailed information about the buildings and their HVAC&R systems and operating schedules and (4) asked to calibrate the DOE-2 simulation program to these buildings, and (5) the results were finally analyzed. It was found for all seven buildings taken together (a) that on an annual basis, the standard deviation for total energy use was less than 8.0% with a maximum difference of 12.0%; (b) that on an annual basis, these indices were 11.0% and 19.0%, respectively, for gas/fuel and 9.2% and 15.0%, respectively, for electricity use; and (c) that on a monthly basis, these indices were 16.7% and 24.0% for total energy use, 26.3% and 35.0% for gas/fuel use, and 18.7% and 30.0% for electricity use. These values are higher than those recommended in ASHRAE Guideline 14, but it must be realized that this study was conducted about 25 years ago at the dawn of the building simulation era and did not require any submetering whatsoever.

The results of another USDOE-funded study a few years later are documented in a report by TRC (1984). Only one large office building was selected, detailed monitoring was done to various end-uses, and the accuracy of the DOE-2 program algorithms was evaluated at hourly time scales for each of the four seasonal periods for several components, such as chillers, water pumps, cooling towers, boilers, and secondary system components. Measurement errors of the sensors were also accounted for. It was found that monthly predicted total energy use was lower by about 5% on average over the year as compared to measured energy use (though individual months showed much greater variability: underprediction in July by 6% and overprediction in March by 12%). Values of such differences for each of the equipment are also specified. It must be noted that these results are building- and location-specific. The authors conclude that the ability of the DOE-2 simulation program to predict building energy use is within the accuracy of empirical measurements of building energy use.

The approach adopted by Kaplan et al. (1990a, 1990b) in the framework of small office buildings within the Energy Edge program is to monitor several end-uses during short periods and to perform the calibration for these periods only as against a whole year. The *short tuning periods* recommended are one month during a hot period, one month during a cold period, and one month in between. The authors studied shorter periods of a week but concluded that one-month intervals tended to better smooth variability. In general, the tuning process first corrected obvious simulation errors highlighted by the discrepancies, then adjusted for internal loads, and, finally, adjusted other inputs such as HVAC end-uses. Tuning tolerances of about 10% for whole-building energy use on an annual basis were achieved.

Hunn et al. (1992) describe a study using as-built drawings, site interviews, and whole-building electric use data for the Texas Capitol Building to calibrate DOE-2 to the pre-renovation energy use status. Energy savings due to the extensive renovations could then be identified using post-renovation data. A similar methodology was adopted by Reddy et al. (1994) for a 250,000 ft^2 university building in Austin using two months of pre-retrofit monitoring and site visits to identify the most significant end-use components. They were able to achieve calibration tolerances of 4.5% (monitored data were lower) in whole-building energy use for the entire seven months of pre-retrofit period and 2.8% for heating energy use. However, large differences in cooling energy during certain months (up to 10%) were detected. Norford et al. (1994) present results of a case study in a large office building in New Jersey with a rich variety of architectural and engineering features designed to minimize energy use. The individual effects of a large number of variables, such as power density and schedule of lights and office equipment, HVAC schedules and thermostat settings, and HVAC and building shell performance, were studied. The study reiterates the need for properly accounting for internal loads and recommends that one-time measurements be adopted for reliable calibration. Finally, Lunneberg (1999) also points out the critical importance of monitoring key short-term end-use internal loads so that more realistic schedules of the building can be obtained. The results of simulated calibration to a 49,000 ft^2 commercial office building in San Diego are presented.

Pedrini et al. (2002) describe a methodology involving three steps: using as-built drawings, walk-through visits, and electric and thermal measurements. The methodology has been used to calibrate the DOE-2 simulation program with monitored data from about a dozen office buildings. They found that the first step of calibration usually results in large unacceptable differences (up to 20% at monthly time scales). Next, schedule adjustments are made by monitoring certain key end-uses for a few days. They find that this usually reduces the errors to about 5% to 6%. Next a walk-through audit is conducted where spot measurements of quantities such as lighting levels, airflow, air temperature, and on/off status of power circuits are made with handheld instruments. The mean differences can thus be brought down to less than 1%. The final calibration step involves using end-use data measurements for more accurate calibration. The paper presents monthly and diurnal plots to illustrate how the calibration process is gradually improved. Yoon et al. (2003) present a seven-step methodology for calibrated simulation: (a) base case modeling that involves gathering building data, utility bills, weather data, and as-built drawings as well as carefully considering building zoning; (b) using monitored data of several end-uses during a week, analyzing differences between simulations and measurements in the base load (i.e., weather-independent gas and electric use) using scatter plots with outdoor temperature; (c) fine tuning a simulation model during the swing season period when heating and cooling energy use is low; (d) performing additional site visits and interviews to refine lighting power densities, equipment quantities, schedules, number of occupants; (e) calibrating the heating/cooling season; (f) evaluating accuracy of the calibration by calculating statistical indices such as normalized mean bias error and coefficient of variation values as well as using graphical plots such as scatter plot and box and whisker plots; and (g) evaluating the effect of promising ECMs. The results of applying this methodology to a large 26-story commercial building of 83,000 m^2 in downtown Seoul, South Korea, are also presented.

Calibration Based on a Suite of Informative Graphical Comparative Displays

This approach is described in several papers by Haberl and colleagues (Bronson et al. 1992; Haberl et al. 1993a, 1993b, 1993c; Bou-Saada and Haberl 1995a, 1995b; Haberl et al. 1996; Haberl and Abbas 1998a, 1998b; Haberl and Bou-Saada 1998). The issue of determining where errors exist between measured and simulated performance is simple when viewed at monthly or diurnal plots. When calibration is being performed with hourly data, the analyst is overwhelmed with too many data points and is at a loss to determine exactly where the differences exist. Hence, along with the manual, iterative, and pragmatic approach to calibration described in the previous section, certain types of visual graphics can highlight differences and be useful aids to the analyst in deciding which parameters to calibrate or "tweak" for the next iteration. Such graphical plots cannot be generated with spreadsheet programs and, hence, there is a need for special toolkits. Common plots include carpet plots, three-dimensional time-series plots of energy use and residuals, superposed and juxtaposed binned box, and whisker and mean (BWM) plots in addition to the standard two-dimensional plots, such as scatter plots and time-series plots. Other practitioners have also developed similar graphics capability (e.g., McCray et al. [1995a, 1995b]).

Calibration Based on Special Tests and Analytical Procedures

In this category, we include a compendium of specialized approaches as described below.

Intrusive Blink-Tests. In case resources are limited, "blink" tests or on/off tests can be used to determine snapshot end-use measurements of lighting and motor control center electricity use. Blink tests can be performed over a weekend using a data recorder to record whole-building electricity use. A series of tests are then performed whereby groups of end-use loads are turned on and off in a controlled sequence, and their incremental power readings provide the necessary end-use information. Data are usually recorded at one- to five-minute intervals.

Soebarto (1997) presents the results of two case studies (one a university campus building consisting of laboratories, offices, and classrooms and the other a municipal building) using monthly utility bills in conjunction with site visits, building drawings, and hourly monitored whole-building electricity use and on/off "blink tests." The paper showed that only two-to-four weeks of monitoring at any period of the year was enough to provide adequate accuracy: coefficient of variation (CV) of 6.7% for hourly whole-building electricity and within 1% for chilled water use for the university building.

Shonder et al. (1998) suggest the use of detailed calibration to estimate the effect of various ECMs in types of campus buildings such as large military housing, federally subsidized low-income housing, and planned communities such as condominiums, townhouses, and senior centers. The approach has been illustrated for a large army base in Louisiana consisting of 46 buildings and 200 individual apartments. Pre-retrofit data were used to calibrate TRNSYS (SEL 2000), which was then used to predict savings, subsequently found to be within 5% of the measured post-retrofit energy use. The type of ECMs modeled involved replacement of air-source heat pumps with geothermal heat pumps, installation of low-flow shower heads, and lighting retrofits (delamping and compact light fixtures). The authors also found that only six months of data were adequate to calibrate the TRNSYS model satisfactorily.

STEM Tests. These include a protocol of short-term energy monitoring (STEM) tests along with a systematic approach of reconciling differences between measured and simulated data. The origin of this method was that developed by Subbarao (1988) for residential buildings that was extended to small-medium commercial buildings by Manke and Hittle (1996). The basic approach in both methods is similar in that it involves intrusive and controlled heating and cooling tests during a three- to five-day period. For smaller buildings portable heaters could be used, while for larger buildings the building heating and cooling system could be used. The test protocol consists of a period during which temperature is controlled to be constant over time and uniform over space and another period during which temperatures were allowed to float. The coheating period is intended to provide a good estimate of the building loss coefficient during nighttime when heat flows other than through the shell are negligible. The cooldown period is intended to provide an estimate of the effective thermal time constant of the building.

Manke and Hittle (1996) then describe how the BLAST computer program (BSL 1999) was calibrated by first identifying four or five primary building parameters and another set of secondary parameters. Next, each of the parameters is varied one at a time from 10% to 200% of their nominal values in increments of 10% while holding the other parameters constant and computing the RMSE for each sensitivity run. The calibration process starts with the parameter with the lowest root mean square error (RMSE) value and varies the next best parameter until a minimum RMSE is reached. This process is continued with each parameter, in turn, until a global minimum is reached. The results of applying this approach to two buildings are also presented.

Macro Parameter Estimation Methods. In this category, we include studies that have attempted to deduce overall or aggregate parameters (such as building U-factors, etc.) purely from non-intrusive monitored data as against performing intrusive tests, as done by Manke and Hittle (1996). A study by Reddy et al. (1999) falls in this category. The paper proposes an inverse method to estimate building and ventilation parameters from non-intrusive monitoring of heating and cooling energy use of large commercial buildings. The procedure involves first deducing the loads of an ideal one-zone building from the monitored data and then, in the framework of a mechanistic macro-model, using a multistep linear regression approach to determine the regression coefficients (along with their standard errors), which can finally be translated into estimates of the physical parameters (along with the associated errors). Several different identification schemes have been evaluated using heating and cooling data generated from a detailed building simulation program for two different building geometries and building mass at two different climatic locations. A multistep identification scheme has been found to yield very accurate results, and an explanation as to why it should be so is also given. This approach has been shown to remove much of the bias introduced in the multiple linear regression approach with correlated regressor variables. It was found that the parameter identification process is very accurate when daily data over an entire year are used. Parameter identification accuracy using twelve monthly data points and daily data over three months of the year was also investigated. Identification with twelve monthly data points seems to be fairly accurate, while that using daily data over a season does not yield very good results.

Signature Analysis Methods. This approach has been developed by Liu, Claridge, and colleagues from Texas A&M university over several years in an effort to improve operation and control schedules, to diagnose malfunction of HVAC components, to predict the effect of different control changes primarily to the secondary equipment of large buildings, and to optimize their control settings. The basis of this approach is to use the simplified engineering models based on the SEAP procedure (Knebel 1983) and to calibrate them against measured data that not only include whole-building electric and motor control center data but also whole-building *heating and cooling thermal loads* on an hourly basis. The approach has been applied to about two dozen university campuses with great success in that this approach was able to identify millions of dollars of energy savings. The simplification is done in several ways: the building is assumed to be a two-zone building (one interior or core zone and one perimeter zone), average daily data and steady-state models are used for the calibrated simulation and analysis, and one large air-handling unit (AHU) is substituted for the numerous smaller ones for each zone (this consolidation is done only with similar types of AHUs). Buildings with different types of secondary systems (CV, VAV, dual duct) have been studied. The approach started with a paper by Katipamula and Claridge (1993), later extended by Liu and Claridge (1998), which described the basis of this approach, the simplifying assumptions made, the modeling equations, the procedure of calibration, how this approach could be used to identify control changes, along with a case study example. It is important to keep in mind that since heating and cooling thermal load data are available, the calibration approach relies heavily on these data for more robust calibration. Other than this nuance in data availability, the original approach is similar to the often adopted calibration based on manual, iterative, and pragmatic intervention as discussed earlier.

A few years later, Liu and colleagues added a new dimension in the calibration process, as described by Wei et al. (1998), Haves et al. (2001), and Liu et al. (2003, 2004). Again, the primary purpose of the calibration is for performing building diagnostics during the framework of continuous commissioning. The approach has been modified into a two-step calibration process integrated with *energy signatures* of the building AHU that express the variation in energy consumption caused by varying a single parameter at a time. The energy signatures are generated as follows. First, the base case building is simulated with representative values for the major input parameters, such as floor area, interior zone fraction, envelope U-factor, air infiltration, internal heat gain, outside air intake, cold deck temperature schedule, room temperature, supply air flow rate, and minimum airflow ratio for different outdoor temperature values. Next, each of these parameters is varied one by one in small discrete steps over their range of realistic variation. Plots of the percentage differences in chilled water and hot water with outdoor temperatures are then generated under each combinatorial change, which are called "energy signatures."

Once such signatures are generated, the calibration process is undertaken as a two-step process. The first calibration focuses on the weather dependency of the model, while the second focuses on the time-dependent schedule. During the first process, the residuals between simulated heating and cooling energy consumption are plotted with outdoor air temperature on a daily basis. The patterns of these residuals are compared with previously generated graphic energy signatures to identify the primary input parameters that may be responsible for this difference. Appropriate changes are then made, and another simulation is performed. This process is continued until satisfactory calibration is achieved. The second-level calibration compares the measured and simulated energy consumption on a time-of-day or hourly basis, and calibration is done heuristically until satisfactory. Finally, it is recommended that the first-level calibration be rechecked and repeated if necessary. A case study of a large eleven-story university building with nine single-duct VAV units located in Texas is presented. The report by Liu et al. (2003) is an extension of the previous work and contains characteristic signatures suitable for use in calibrated energy simulations of large buildings with the following four types of secondary systems: single-duct CAV and VAV and dual-duct CAV and VAV as applicable to three California climates typified by Pasadena, Sacramento, and Oakland.

Analytical/Mathematical Methods of Calibration

The calibration process adopted in all of the studies described earlier involves tuning or refining the initial simulation input parameters in a heuristic manner, which depends on the experience and expertise of the user. The analytical/mathematical formulation of the calibration process is akin to an optimization problem with the objective function being the minimization of the month-by-month (or even hour-by-hour) mean square errors between measured and simulated energy use data. Such an approach, which would automatically determine which parameters to tune and by how much, has been formulated and developed to the extent that it was implemented in a commercially available detailed simulation program called RESEM (Carroll et al. 1989). This was done by Carroll and colleagues at LBNL, as reported in a series of papers summarized by Carroll and Hitchcock (1993).

Since numerical optimization is more efficient with fewer parameters to be optimized, the authors suggest that the calibration process should first start with the process of reducing the number of parameters to be calibrated. This involves using sensitivity analyses to identify them. A recommendation is that this number be brought down to about 25 or less. The entire process was evaluated both by using synthetic data from several prototypical buildings as well as from a limited sample of actual buildings. It must be mentioned that this procedure was developed for use with simplified simulation programs such as ASEAM (ASEAM 3.0, 1991) and has not been used along with detailed fixed-step programs such as DOE-2. The authors also spell out various future enhancements to this work, such as using more matching criteria to the penalty function (such as including monthly demand data) and using knowledge-based system technology.

On a final note, a recent paper by Sun and Reddy (2006) proposes an analytic framework for calibrating building energy system simulation software/programs that has a firm mathematical and statistical basis. The approach is based on the recognition that though calibration can be cast as an optimization problem, the basic issue is that the calibration problem is underdetermined, i.e., there are many more parameters to tune than can be supported by the monitored data. The proposed methodology involves several distinct concepts, namely, sensitivity analysis (to identify a subset of strong influential variables), identifiability analysis (to determine how many parameters of this subset can be tuned mathematically, and which specific ones are best candidates), numerical optimization (to determine the numerical values of this best subset of parameters), and uncertainty analysis (to deduce the range of variation of these parameters). A synthetic example involving an office building is used to illustrate the methodology with the DOE-2 simulation program.

RELEVANT ERROR AND UNCERTAINTY ISSUES

Issues of uncertainty relevant to building simulation programs as applied to the building design process have much in common with the process of calibration with measured data. Therefore, a brief literature review of uncertainty as it relates to building design is undertaken, followed by one relevant to calibrated simulation.

Relevant to Building Simulation Programs

Errors or uncertainties in building energy simulation programs (or any simulation program for that matter) arise from different sources: (1) improper input parameters that could be due to user-related lack of experience (or even negligence) and improper specification of material properties or system parameters; (2) improper model assumptions and simplifications due either to the underlying physics of the phenomenon or to the use of semi-empirical model coefficients; (3) lack of robust and accurate numerical algorithms; and (4) error in writing the simulation code.

While verification deals with determining whether the equations are solved correctly (error 3 above), validation involves solving the right equations (error 2 above). Bloomfield (1999) provides a good overview of work done on validation of computer programs for predicting the thermal performance of buildings. A more recent and concise description can be found in chapter 31 of the *2005 ASHRAE Handbook—Fundamentals* (ASHRAE 2005). Major conceptual issues are described along with outstanding problems, both pragmatic and philosophical. Finally, Bloomfield, based on several previous papers (one of the first being that by Judkoff et al. [1983]) categorizes validation techniques as follows: (1) *code checking*, which involves a series of activities designed to test the operation of the code against specified functionalities and expected behavior; (2) *analytical validation tests*, in which outputs from the program, subroutine, or algorithm are compared to results from a generally accepted numerical method for isolated heat transfer mechanisms under very simple and highly constrained boundary conditions; (3) *intermodel (or comparative) comparisons*, where the results of one program are checked against those of another that may be considered better validated or more detailed or, presumably, more physically correct; and (4) *empirical validation*, which entails comparing simulation predictions with measurements or monitored data from a real building, test cell, or laboratory experiment.

Though several papers can be found in the published literature on verification and validation of building energy analysis computer programs (e.g., Yuill and Wray [1990]), the first systematic and complete study was undertaken by researchers from NREL and called the BESTEST Intermodal comparison method, which provides both systematic model testing and diagnosing the source of predictive disagreement (Judkoff and Neymark 1999). The method comprises a series of tests that progress from extremely simple to relatively realistic. The tests are so designed that differences between the results (such as annual loads, annual maximum and minimum temperatures, peak loads, etc.) would point to a certain deficiency in the computer program, thus allowing diagnosis. The validation process has undergone refinement and currently is a three-step process. The first step involves analytical verification, followed by empirical validation and, finally, intermodel comparisons. The BESTEST procedure has been adopted as ANSI/ASHRAE Standard 140 (ASHRAE 2001) and also by the International Energy Agency (IEA). The methodology that was earlier meant for validating the models for a building envelope has also been extended to include models that simulate the behavior of unitary space-cooling equipment (Neymark et al. 2002).

The NREL methodology as it pertains to empirical validation distinguishes between different levels, depending on the degree of control exercised over the possible sources of error during the simulation. The error sources were divided into:

a. *external error types* due to differences/discrepancies between actual and simulation inputs: (1) in weather data, (2) in building operational data (such as schedules, control strategies, effects of occupant behavior, etc.), (3) in physical properties (thermal, optical, etc.) of the various building envelope and equipment components, and (4) due to user error in deriving building input files;

b. *internal error types* having to do with the accuracy of the models or algorithms: (1) due to model simplifications in how the heat, mass, and fluid flow processes are modeled, (2) from improper numerical resolution of the models, and (3) due to coding errors.

The empirical validation method has been, historically, the most adopted validation procedure. The simplest (or simplistic) level is to compare long-term energy use of a building with that calculated by the computer program. Though this approach is favored by many practitioners, the results have to be interpreted with care since individual errors may compensate or offset each other to some extent. Higher levels of validation would require finer time scale and controlling or eliminating the various combination of error types. The empirical validation approach has increased in sophistication from simple scatter and time series plot comparison to sophisticated and elaborate experimental design and statistical tests. However, given that real buildings are very complex and modeling their behavior is extremely difficult, and good experimental data are hard to gather, validation efforts have moved to simpler buildings and to test cells (Bloomfield 1999), the European PASSYS test cells being a good example. The approach adopted is to conduct very detailed and carefully designed controlled experiments on a number of passive solar components in outdoor heavily instrumented test cells. Thus, the cause of discrepancies between simulations and measurements could be better identified. Findings of the model validation and development subgroup within PASSYS are described by Clarke et al. (1993), Strachen (1993), and Jensen (1995).

Lomas et al. (1997) report on the largest-ever exercise (25 users from Australia, Europe, and the US) to validate dynamic thermal simulation programs using three single-zone test cells in the UK. The tests consisted of a ten-day period where they were intermittently heated and another ten days when they were unheated. The work produced five empirical validation benchmarks that resulted in significant practical benefits for program users, vendors, and potential users. It was recommended that empirical validation exercises should consist of an initial blind phase and then an open phase in which measurements are made available.

Statistical screening techniques are important when dynamic effects dominate. Palamo et al. (1991) propose estimating the autocorrelation and power spectrum of the residuals and determining the cross-correlation function between them and the inputs, both in time and frequency domains. This may provide a means as to whether the physical process is well represented by the simulation program. Ramdani et al. (1997) investigated the diagnostic ability of spectral analysis tools, while Del Barrio and Guyon (2003) proposed a new diagnostic method using sensitivity analysis and optimization techniques for analyzing the model parameter space. This involves identifying the changes in parameter values that are required for a significant improvement in model behavior. A paper by Tabary and Ramdani (1995) suggests different frequency ranges used for error disaggregation (from 2- to 34-hour periods) using test cell measurements.

A good review of the various sources of prediction uncertainty of thermal building energy simulation programs is provided by Macdonald and Strachan (2001). The paper also describes and illustrates how uncertainty analysis has been incorporated in the ESP-r program. This includes multiple simulations with perturbations made to salient simulation input parameters according to their uncertainty distributions.

Finally, we have been able to identify one paper (Aude et al. 2000) that deals with uncertainties associated with code input data. TRNSYS was used with measured data from several test cells. The adjoint method, which allows the determination of first- and higher-order derivatives in an exact manner (as compared to an approximate manner using finite differences), was shown to simplify the calculation of uncertainties when a large number of parameters need to be considered. The drawback is that it requires a delicate implementation for it to work properly.

Relevant to Calibrated Building Simulation and ASHRAE-14

As pointed out by Bloomfield (1999), all computer programs make approximations. Hence, any comparison between measured and predicted performance should, therefore, be seen, not in terms of whether the program agrees with the measurements, but *whether the program is good enough* for its intended purpose. The issue is then, how does one determine "good enough"? The statistical basis of the manner in which ASHRAE Guideline 14 addresses uncertainty issues has been adopted from Reddy and Claridge (2000). This *methodology applies to regression models* identified from baseline monitored data and, hence, relates to black-box and grey-box approaches. It cannot be applied as such to the calibrated simulation model approach, which was one of the reasons why the ASHRAE RP-1051 research was undertaken in the first place.

The ASHRAE Guideline 14 document sets uncertainty or tolerance limits for calibrated simulation as described below:

a. *Accuracy of the hourly computer program used*: Section 5.3.2.4 of the guideline document stipulates that the computer model being used for calibration should be accurate to within 5% for the normalized mean bias error (NMBE) and 15% for the root mean square error (CV(RMSE)) relative to monthly data. If hourly data are used for validation, these thresholds are 10% and 30%, respectively.

b. *Calibration accuracy*: Section 6.3.3.4.2.2 of the document states: "models are declared to be calibrated if they produce NMBEs within ±10% and CV-RMSEs within ±30% when using hourly data, or 5% to 15% with monthly data."

c. *Savings uncertainty*: Section 5.3.2.4 stipulates that the level of uncertainty be less than 50% of the annual reported savings at a confidence level of 68%.

No specific reason is provided as to why the above tolerance values were proposed, but they are presumably based on the practical experience of the energy modelers who performed calibrated simulations. Recall that the results of one of the first calibration studies (Diamond and Hunn 1981) were described earlier.

CALIBRATION TOOLS AND CAPABILITIES

Daytyping and Disaggregation Studies

There seems to be two lines of thought as to what the term "daytyping" implies. Some consider it to be a procedure whereby the diurnal (i.e., 24 hourly values during a day) building energy use (thermal loads and/or electrical use) during an entire year is grouped into a small number of discrete bins within which the diurnal energy use patterns are similar. Others (including us) argue that the primary objective of the day-typing is not to characterize the energy use, but to *characterize the behavior of the system's response to the driving functions* into a small number of bins. In either case, daytyping characterizes energy use on a daily profile, rather than an hourly basis. The words "patterns or profiles" include both the shape and magnitude, and, hence, the combination of energy use conditions over a 24-hour period is considered one multi-dimensional "point." Characterizing energy use and system performance on a daily profile basis helps reduce the complexity of the analysis by reducing the number of points that must be considered. The daytyping approach is advantageous in that the extensive information contained in a set of hourly data is condensed into relatively few typical days that essentially characterize the response of the building to external weather changes.

Building energy use can be broken up into weather-independent components (such as lights, plug loads, and occupancy) and weather-dependent components (such as cooling and heating secondary and primary energy use and whole-building electricity use), the relative contribution, of which are weather, building, and HVAC&R equipment specific. For *weather-independent energy use*, a widely used grouping scheme is by weekday, weekend, and holidays. A greater degree of accuracy may require finer grouping, in which case Mondays may be distinguished from Fridays, and both of them from the rest of the weekdays. The holiday group may also be divided into short (such as three- to four-day holidays) and long (such as Christmas breaks or term breaks in universities and schools). This breakup can be done purely from a statistical analysis that entails sorting the data into weekday and weekend 24-hour profiles, comparing and grouping each point by a predetermined standard deviation limit or a t-test (as done by Katipamula and Haberl 1991; Dhar et al. 1998). The procedure suggested by Katipamula and Haberl was later used by Bronson et al. (1992) to calibrate DOE-2 simulation to represent occupancy and schedules. A more refined version of this approach was proposed by Thamilseran and Haberl (1995), who used statistical measures, such as mean, median, and inter-quartile range comparisons. They showed that the use of inter-quartile ranges is a more robust goodness-of-fit indicator and is not affected by the magnitude of the hourly mean as was the Katipamula and Haberl approach.

The day-typing or grouping of *weather-dependent* loads or energy use is more complex since the annual variation of weather impacts the operation of the HVAC&R systems (due to economizer cycles, hot and cold deck reset, terminal reheat, etc.). Note that the use of load frequency distribution methods, which include load duration curves and degree-day or degree-hour methods, are only appropriate for simplified modeling of equipment whose operation is coupled directly to the load. However, more detailed information is needed if equipment operation is dependent on other factors such as outdoor temperature.

Kaplan et al. (1990 a, 1990b) describe a procedure capable of distinguishing the daily non-HVAC loads during the intermediate or swing season (i.e., no heating or cooling). Another daytyping procedure by Hadley (1993) used a principal component analysis and cluster analysis to identify distinctive weather daytypes into what he calls "representative days." HVAC system energy use data for each day were grouped by these day types, and hourly load profiles were developed for each day type. The result is data sorted according to three temperature day types (cold, moderate, and hot temperatures), which can then be segregated into 24-hour system operation schedules (heating-only mode, no heating/no cooling mode, and cooling-only mode).

Baughman et al. (1993) developed a "characteristic days method" for evaluating cool storage system performance. Daily cooling coil loads (in kBtu/day) were found to be correlated with the peak daily temperature. A set of 15 days was selected to represent the range of variation of cooling coil loads over the year. Performance of a cool storage system was estimated for each of these days using a computer model. Each month of the year was characterized according to the number of each of the characteristic day types occurring in the month. (Presumably weather data for a typical year were used to perform this characterization.) The storage system demand and energy use for each month were determined from the maximum demand and total energy use of the day types in the month.

Algorithms to disaggregate whole-building hourly electricity use into its various end-uses are also of interest to calibrated simulation research. The analyst may have interval data that can be used as an additional monitored data channel against which to calibrate the detailed hourly simulation program. If such interval data allow preliminary estimates of end-use data to be gleaned, this would be advantageous during the calibration process. Akbari et al. (1988), Akbari (1995), and Akbari and Konopacki (1998) propose a methodology called the EDA method (end-use disaggregation algorithm), which has two steps. First, regression models for each hour of the day are identified between hourly measured whole-building energy use and outdoor dry-bulb temperature for each major day type with the year divided into two seasons only—summer and winter. The regression constant of the model for each hour is then assumed to provide some indication of the weather-independent energy use of the whole building, while the slope of the model relates to the weather-dependent behavior. The next step involves disaggregating the non-weather-dependent load, and this is done in two steps. First, on-site survey data along with DOE-2 simulations are used to provide the relative contribution of each end-use toward the total. These proportions, along with a number of statistically based corrections and assumptions about cooling system efficiencies, are used to compute hourly end-use profiles. The EDA method in its current form is capable of only identifying a few end-uses (lighting, plug loads, and HVAC) and is applicable only for the cooling season. It has been applied to data from several individual buildings as well as to data from two major California utilities for use in their forecasting models (Akbari 1995).

Finally, BouSaada and Haberl (1995a) used the Hadley (1993) concept of weather day-typing described above to develop a procedure for disaggregating whole-building electricity signal into hourly end-uses. Instead of using principal component analysis to identify the day types or sophisticated statistical corrections as done by Akbari, a more intuitive and practical method is proposed. The monitored hourly whole-building data were grouped into weekdays and weekends first and then into three day types depending on the daily dry-bulb temperature: <7°C, 7°C-24°C, and >24°C. This results in six end-use categories: weekday/weekend, heating, cooling, and non-heating/non-cooling. Box-whisker plots of these 12 diurnal plots for the entire year provide an indication as to the variability of hourly data around such a classification. Subsequently, weather-independent end-use disaggregation is done based on site surveys and interviews and clamp-on meters to provide some indication of exterior lighting, interior lighting, and plug loads. Weather-dependent end-uses that include heating and cooling energy use were separated based on days when the outdoor temperature was less than 7°C (heating) or more than 24°C (cooling). Between these two temperature ranges, the energy use was attributed to end-use HVAC fan use. This approach was illustrated with monitored and DOE-2 simulated end-use data for an 8,100 ft^2 daycare case study building located in Washington, DC, where acceptable agreement was reached.

Data Visualization Tools

Being able to visualize data, whether it be monitored or simulated, is an important aspect of the calibration process since it provides a convenient graphical comparison of both types of data and suggests to the user how to proceed with the calibration process. This phase is often performed along with the calculation of the statistical indices described in the previous section. There are numerous papers on the different ways by which data can be viewed in the fields of science and engineering. We shall only present those that seem to be most pertinent and do so in order to keep this paper within stipulated page limits. A more detailed description can be found in the report by Reddy et al. (2006).

Basic Monthly and Hourly Time Series Plots. These are the simplest types of plots, widely used by practitioners (such as Waltz [2000]). They are very simple to generate using spreadsheet programs. The monthly plots would depict how the 12 utility bill data (either energy use or peak)—both measured and simulated—vary with month of the year on the x-axis. The hourly time series plot would provide similar information, but on an hourly basis for a typical day of the year. Such plots provide a quick means to determine the magnitude of the errors, their patterns, and the season during which they occur and interactively indicate to the analyst possible alternatives to the calibration or tuning process.

ASHRAE Guideline 14 (ASHRAE 2002) discusses three different graphical techniques; the determination regarding which techniques to use for any given calibration is left to the judgment of the modeler. All these are based on hourly data availability. The following three techniques are largely adopted from several papers by Haberl from Texas A&M university (Haberl et al. 1993a, 1996; Haberl and Abbas 1998a, 1998b).

Scatter Plots of Energy Use and Ambient Temperature. Scatter plots of energy use versus ambient temperature (separated into weekdays and weekends) are well known to data analysts. Various additions to such basic x-y plots have been suggested, such as box-whisker-mean (BWM) plots, which display the maximum, minimum, mean, median, 10th, 25th, 75th, and 90th percentile points for each data bin for a given period of data. These plots eliminate data overlap and allow for a statistical characterization of the dense cloud of hourly points (scatter plots are still useful in showing individual point locations). The important feature to note about such plots is that the data are statistically binned by temperature. This feature allows for the bin-by-bin goodness-of-fit to be evaluated quantitatively and graphically. Using the box-whisker-mean plot combined with a scatter plot also allows one to visualize the data as a whole while simultaneously seeing the effects of the outliers in specific situations. One final feature of these plots is that the measured data mean is superimposed as a dashed line onto the calibrated simulation data. The difference between mean lines in each bin provides a measure of how well the model is calibrated at a specific temperature bin. Likewise, the inter-quartile range (i.e., the distance between the 25th and 75th percentiles) represents the hourly variation in a given bin.

Diurnal Time Series Plots of Energy Use with Hour of Day. Weekday and weekend 24-hour weather day-type box-whisker-mean plots show the whole-building electricity use versus the hour-of-the-day for both the measured data and the simulated data in three weather day types complements BWM plots. The weather day types arbitrarily divide the measured data into groupings. For example, the summer peak weekday can be defined by selecting the five warmest non-holiday weekdays during June, July, and August using the actual weather data for the calibration period. The hourly load data for each of those identified days is then extracted from the utility data sets and the simulation output and compared. In a similar fashion, the summer average weekday data are prepared from the remaining weekday data (excluding the days used in determining the peak day data set), as are the other day types of interest.

Advanced Visualization Methods. Comparative three-dimensional surface plots allow clearer visualization of the monitored data, the simulated data, positive-only values of the measured data subtracted from the simulated data, and positive-only values of the simulated data subtracted from the measured data. Individual hourly differences may be visually detected over the entire simulation period using these plots, which allows the user to recognize patterns in the comparisons such as the simulation program's overpredictions in the spring and fall mornings and afternoons and both over- and underpredictions in the late evening throughout the year. An obvious benefit of such plots is their ability to aid in the identification of oversights, such as a daylight savings shift or misalignment of 24-hour holiday profiles. One negative drawback associated with these graphs is the difficulty in viewing exact details, such as the specific hour or specific day on which a misalignment occurs. Some model calibrators complain that three-dimensional surface plots obscure data that are behind "hills" or in "valleys." By substituting color for depth, some calibrators find they can more easily interpret the graphs (Christensen 1984).

Glaser and Ubbelohde (2001) describe novel high performance visualization techniques for reviewing time-dependent data common to building simulation. Techniques such as brushing and linking (where the user selects to investigate the behavior during a few days of the year), tessellating (dividing a two-dimensional chart into multiple smaller two-dimensional charts giving a four-dimensional view of the data such that a single value of a representative sensor can be evenly divided into smaller spatial plots arranged by time of day), magic lenses (which can zoom into a certain portion of the room), and magic brushes are described. These techniques enable rapid inspection of trends and singularities that cannot be gleaned from conventional viewing methods. The paper is geared toward visualization of daylighting and illumination simulations and not for energy use data.

Finally, animated scatter plots of energy use versus temperature (Haberl et al. 1993a) can provide compelling and clearer indication of time sequence discrepancies between simulated and measured data. Animated contour plots wherein binning of hourly data can provide better visual diagnostics have also been suggested.

SUMMARY

This paper provides a pertinent and detailed literature review of published papers on calibration of building energy simulation programs. A brief historical description of why and how calibration of building energy simulation programs as a formal discipline came into being, its advantages, and its unique problems are first reviewed. This is followed by pertinent descriptions of various calibration studies grouped in a logical structure. Subsequently, the issue of uncertainty and its implications was discussed in the framework of both building design as well as calibrated simulation (more specifically within ASHRAE 14P). Finally, various papers dealing with calibration tools and capabilities were described. Subsequent papers will describe and illustrate a coherent and systematic calibration methodology that will include both parameter estimation of the various simulation inputs as well as uncertainty in predictions using the calibrated simulation program.

ACKNOWLEDGMENTS

This research was performed as part of an ASHRAE-funded research project (RP-1051) under Technical Committee 4.7. We would like to thank the PMSC members for their close supervision of this research. Valuable advice and assistance from Drs. I. Maor, J. Sun, and C. Panjapompom are greatly appreciated. Finally, the feedback and involvement of J. Haberl, B. Ramirez, R. Sonderegger, and J. Waltz is acknowledged.

REFERENCES

Akbari, H., K.E. Heinemeier, D. Flora, and P. LeConaic. 1988. Analysis of commercial whole-building 15 minute-interval electric load data. *ASHRAE Transactions* 94(2).

Akbari, H. 1995. Validation of an algorithm to disaggregate whole-building hourly electrical load into end uses. *Energy* 20(12):1291–1301.

Akbari, H., and S.J. Konopacki. 1998. Application of an end-use disaggregation algorithm for obtaining building energy-use data. *ASME Journal of Solar Energy Eng*. 12:205–210.

ASHRAE. 2001. *ANSI/ASHRAE 140-2001, Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs*. Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.

ASEAM 3.0. 1991. A simplified energy analysis method. U.S. Department of Energy, Washington D.C., April.

ASHRAE. 2002. *ASHRAE Guideline 14-2002, Measurement of Energy and Demand Savings*. Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.

ASHRAE. 2005. *2005 ASHRAE Handbook—Fundamentals*. Chapter 31, Energy estimation and modeling methods. Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.

Aude, P., L. Tabary, and P. Depecker. 2000. Sensitivity analysis and validation of buildings's thermal models using adjoint-code method. *Energy and Buildings* 31:267–283.

Bloomfield, D.P. 1999. An overview of validation methods for energy an environmental software. *ASHRAE Transactions* 105(2).

Baughman, M.L., J.W. Jones, and A. Jacob. 1993. Model for evaluating the economics of cool storage systems. *IEEE Transactions on Power Systems* 8(2):716–722.

Bou-Saada, T., and J. Haberl. 1995a. A Weather-daytyping procedure for disaggregating hourly end-use loads in an electrically heated and cooled building from whole-building hourly data. *Proceedings of the 30th Intersociety Energy Conversion Engineering Conference*, July 31-August 4, 1995, Orlando, Florida.

Bou Saada, T., and J. Haberl, 1995b. An improved procedure for developing calibrated hourly simulation models. *Proceedings of the International Building Performance Simulation Association*, August 14-16, 1995, Madison, Wisconsin.

Bronson, D.J., S.B. Hinchey, J.S. Haberl, and D.L. O'Neal. 1992. A procedure for calibrating the DOE-2 simulation program to non-weather-dependent measured loads. *ASHRAE Transactions* 98(1).

BSL. 1999. BLAST 3.0 Users Manual, Building Systems Laboratory, Department of Mechanical and Industrial Engineering, University of Illinois, Urbana-Champaign, IL.

Carroll, W.L., B.E Birdsall, R.J. Hitchcock, and R.C. Kammerud. 1989. RESEM: An evaluation tool for energy retrofits in institutional buildings. *International Building Performance Simulation Association, Building Simulation 1989, Proceedings of the First International IBPSA Conference*, Vancouver, Canada, June 23-24.

Carroll, W.L., and R.J. Hitchcock. 1993. Tuning simulated building description to match actual utility data: Methods and implementation. *ASHRAE Transactions* 99(2):928–934.

Christensen, C. 1984. Digital and color energy maps for graphic analysis of class B hourly data. *Proceedings of the 11th National Passive Solar Conference,* Boulder, Co., June.

Clarke, J.A. 1993. Assessing building performance by simulation. *Building and Environment* 28(4):419–427.

Clarke, J.A, P.A. Strachan, and C. Pernot. 1993. An approach to the calibration of building energy simulation models. *ASHRAE Transactions* 99(2):917–927.

Del Barrio, E.P., and G. Guyon. 2003. Theoretical basis for empirical model validation using parameters space analysis tools. *Energy and Buildings*, Vol. 35.

Dhar, A., T.A. Reddy, and D.E. Claridge. 1998. Modeling hourly energy use in commercial buildings with fourier series functional forms. *ASME Journal of Solar Energy Engineering* 120:217–223.

Diamond, S.C., and B.D. Hunn. 1981. Comparison of DOE-2 computer program simulations to metered data for seven commercial buildings. *ASHRAE Transactions* 87(1):1222–1231.

Glaser, D., and S. Ubbelohde. 2001. Visualization for time dependent building simulation. *7th IBPSA Conference*, Rio de Janeiro, Brazil, Aug. 13-15, pp. 423–429.

Haberl, J., K. Kissock, R. Belur, and R. Sparks. 1993a. Improving the paradigm for displaying complex building energy consumption data. *Proceedings of the ASME-SED International Solar Energy Conference*, pp.475–485.

Haberl, J., D. Bronson, S. Hinchey, and D. O'Neal. 1993b. Graphical tools to help calibrate the DOE-2 simulation program to non-weather dependent measured loads. *ASHRAE Journal* 35(1):27–32.

Haberl, J., D. Bronson, D. O'Neal. 1993c. Impact of using measured weather data versus TMY weather data in a DOE-2 simulation. *ASHRAE Transactions* 101(2).

Haberl, J., R. Sparks, and C. Culp. 1996. Exploring new techniques for displaying complex building energy consumption data. *Energy and Buildings* 24:27–38.

Haberl, J.S., and M. Abbas. 1998a. Development of graphical Indices for viewing building energy data: Part I. *ASME Journal of Solar Energy Eng*. 120:156–161.

Haberl, J.S., and M. Abbas. 1998b. Development of graphical indices for viewing building energy data: Part II. *ASME Journal of Solar Energy Eng*. 120:162–167.

Haberl, J.S., and T.E. Bou-Saada. 1998. Procedures for calibrating hourly simulation models to measured building energy and environmental data. *ASME Journal of Solar Energy* Eng. 120:193–204.

Hadley, D.L. 1993. Daily variations in HVAC system electrical energy consumption in response to different weather conditions. *Energy and Buildings* 19:235–247.

Haves, P., T. Salsbury, D. Claridge, and M. Liu. 2001. Use of whole building simulation in on-line performance assessment: Modeling and implementation issues. *7th IBPSA Conference*, Rio de Janeiro, Brazil, Aug. 13-15, pp. 335-342.

Hunn, B.D., J.A. Banks, and S.N. Reddy. 1992. Energy analysis of the Texas capitol restoration. *Proceedings of the 8th Symposium on Improving Building Systems in Hot and Humid Climates*, May 13-14, Dallas, TX, pp. 165–173.

IPMVP. 2001. International performance measurement and verification protocol. Department of Energy, DOE/GO-102001-1187, Washington D.C., January.

Jensen, S.O. 1995. Validation of building energy simulation programs: a methodology. *Energy and Buildings* 22:133–144.

Judkoff, R., D. Wortman, B. O'Doherty, and J. Burch. 1983. A methodology for validating building energy analysis simulations, SERI/TR-254-1508. Solar Energy Research Institute (now National Renewable Energy Laboratory), Golden, CO.

Judkoff, R.D., and J.S. Neymark. 1999. Adaptation of the BESTEST intermodel comparison method for proposed ASHRAE Standard 140P: Method of test for building energy simulation programs. *ASHRAE Transactions* 105(2).

Katipamula, S., and J.S. Haberl. 1991. A methodology to identify diurnal load shapes for non-weather dependent electric end-uses. *Proceedings of the ASME-JSES International Solar Energy Conference*, March 17–21, Reno, Nevada.

Katipamula, S., and D.E. Claridge. 1993. Use of simplified systems model to measure retrofit energy savings. *ASME Journal of Solar Energy Eng.* 115(2):57–68.

Kaplan, M.B., J. McFerran, J. Jansen, and R. Pratt. 1990a. Reconciliation of a DOE2.1C model with monitored end-use data for a small office building. *ASHRAE Transactions* 96(1):981–993.

Kaplan, M.B., B. Jones, and J. Jansen. 1990b. DOE-2.1C model calibration with monitored end-use data. *Proceedings from the ACEEE 1990 Summer Study on Energy Efficiency in Buildings*, Vol. 10, pp. 10.115–10.125.

Kaplan, M.B., P. Caner, and G.W. Vincent. 1992. Guidelines for energy simulation of commercial buildings*. Proceedings from the ACEEE 1992 Summer Study on Energy Efficiency in Buildings*, Vol. 1, pp. 1.137–1.147.

Knebel, D. 1983. *Simplified Energy Analysis Method*. Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.

Liu, M., and D.E. Claridge. 1998. Use of calibrated HVAC system models to optimize system operation. *ASME Journal of Solar Energy Eng*. 120:131–138. See also Liu, M., and D.E. Claridge. 1995. Application of calibrated HVAC system models to identify component malfunction and to optimize the operation and control schedules. *Proceedings of the ASME/JSME/JSES International Solar Energy Conference*, Maui, 1:209–217.

Liu, M., D.E. Claridge, N. Bensouda, K. Heinemeier, S.U. Lee, and G. Wei. 2003. High performance commercial building systems: manual of procedures for calibrating simulations of building systems, Report HPCBS#E5P23T2b, prepared for the California Energy Commission, PIER Program, October.

Liu, M., L. Song, G. Wei, and D.E. Claridge. 2004. Simplified building an air handling unit model calibration and applications. *ASME Journal of Solar Energy Eng.* 126:601–609.

Lomas, K.J., H. Eppel, C.J. Martin, and D.P. Bloomfield. 1997. Empirical validation of building energy simulation programs. *Energy and Buildings* 26: 253–275.

Lunneberg, T.A. 1999. Improving simulation accuracy through the use of short-term electrical end-use monitoring. *IBPSA Conference*, Kyoto, Japan, Sept. 13–15.

Macdonald, I., and P. Strachan. 2001. Practical application of uncertainty analysis. *Energy and Buildings* 33: 219–227.

Manke, J., and D. Hittle. 1996. Calibrating building energy analysis models using short term test data. *Proceedings of the 1996 ASME International Solar Engineering Conference*, ASME Solar Energy Division, pp. 369–378.

Mayer, T., F. Sebold, A. Fields, R. Ramirez, B. Souza, and M. Ciminelli. 2003. DrCEUS: Energy and demand usage from commercial on-site survey data. *Proc. of the International Energy program Evaluation Conference*, Aug. 19–22, Seattle, WA.

McCray, J.A., P.L. Bailey, J.L. Parker, and R. Gillman. 1995a. Using data visualization tools for the calibration of hourly DOE-2.1 simulations. *IBPSA Conference*, Madison, WI, Aug. 14–16, pp. 461–466.

McCray, J.A., P.L. Bailey, and J.L. Parker. 1995b. Visualize-It model calibration tool kit. *IBPSA Conference*, Madison, WI, Aug. 14-16, pp. 461–466.

Neymark, J., R. Judkoff, G. Knabe, H.T. Le, M. Durig, A. Glass, and G. Zweifel. 2002. An analytical verification procedure for testing the ability of whole-building energy simulation programs to model space conditioning equipment at typical design conditions. *ASHRAE Transactions* 108(2).

Norford, L.K., R.H. Socolow, E.S. Hsieh, and G.V. Spadaro. 1994. Two-to-one discrepancy between measured and predicted performance of a "low-energy" office building: insights from a reconciliation based on the DOE-2 model. *Energy and Buildings* 21:121–131.

Norford, L.K., and T.A. Reddy. 2004. Tools and capabilities for pro-active multi-building load management: Part II—Proactive load control and management of load, power and financial data. *ASHRAE Transactions* 110(2).

Palomo, E., J. Marco, and H. Madsen. 1991. Method to compare measurements and simulations. *IBPSA Conference*, Nice, France, Aug. 20–22, pp. 570–577.

Pedrini, A., F.S. Westphal, and R. Lamberts. 2002. A methodology for building modeling and calibration in warm climates. *Building and Environment* 37:903–912; See also, Pedrini, A., and R. Lamberts. 2001. Methodology for building modeling and calibration for warm climates. *7th IBPSA Conference*, Aug 13–15, Rio de Janeiro, Brazil, pp. 1367–1372.

Ramdani, N., Y. Candau, S. Dautin, S. Delille, N. Rahni and P. Delicieux. 1997. How to improve building thermal simulation programs by use of spectral analysis. *Energy and Buildings* 25:223–242.

Reddy, S.N., B.D. Hunn, and D.B. Hood. 1994. Determination of retrofit savings using a calibrated building energy simulation model. *Proceedings of the 9th Symposium on Improving Building Systems in Hot and Humid Climates*, May 19-20, Arlington, TX, pp. 153–165.

Reddy, T.A., S. Deng, and D.E. Claridge. 1999. Development of an inverse method to estimate overall building and ventilation parameters of large commercial buildings. *ASME Journal of Solar Energy Eng.* 121:40–46.

Reddy, T.A., and D.E. Claridge. 2000. Uncertainty of "measured" energy savings from statistical baseline models. *HVAC&R Research Journal* 6(1).

Reddy, T.A., and L.K. Norford. 2004. Tools and capabilities for pro-active multi-building load management: Part 1: Customer aggregation. *ASHRAE Transactions* 110:2.

Reddy, T.A., I. Maor, C. Ponjapompom, and J. Sun. 2006. Procedures for reconciling computer-calculated results with measured energy data. ASHRAE RP-1004, final report, January.

Schuldt, M.A., and J.S. Romberger. 1998. Alternative approaches to baseline estimation using calibrated simulations. *ASHRAE Transactions* 104(2).

SEL. 2000. TRNSYS Version 15 User Manual and Documentation. Solar Energy Laboratory, Mechanical Engineering Department, University of Wisconsin, Madison.

Shonder, J.A., P.J. Hughes, and J.W. Thornton. 1998. Using calibrated engineering models to predict energy savings in large-scale geothermal heat pump projects. *ASHRAE Transactions* 104(1).

Soebarto, V.I. 1997. Calibration of hourly energy simulations using hourly monitored data and monthly utility records for two case study buildings. *IBPSA Conference Proceedings*, Prague, Czech Republic, Sept. 13–15.

Sonderegger, R., J. Avina, J. Kennedy, and P. Bailey. 2001. Deriving loadshapes from utility bills through scaled simulation. ASHRAE seminar slides, Kansas City, MI.

Strachan, P.A. 1993. Procedures for scaling and replication by simulation. *IBPSA Conference Proceedings*, Adelaide, Australia, Aug. 16–18, pp. 343–349.

Stein, J. 1997. Calibrated simulation: An improved method for analyzing building energy use. E-Source Tech Update TU-97-10, Boulder, CO, September.

Subbarao, K. 1988. PSTAR- Primary and secondary terms analysis and renormalization: A unified approach to building energy simulations and short-term testing, TR-254-3175, National Renewable Energy Laboratory, Golden, CO.

Sun, J., and T.A. Reddy. 2006. Calibration of building energy system simulation programs using the analytic optimization approach (RP-1051). Paper accepted for publication by the *International Journal of Heating, Ventilating, Air-Conditioning, and Refrigerating Research*. To appear in 12(1).

Tabary, L., and N. Ramdani. 1995. An error analysis method applied to a building simulation software: An example of applications and results. *IBPSA Conference Proceedings*, Madison, WI, Aug. 14–16, pp. 316–323.

Thamilseran, S., and J.S. Haberl. 1995. A bin method for calculating energy conservation retrofit savings in commercial buildings. *Proceedings of the Fourth ASME-JSES-JSME International Solar Energy Conference*, Maui, HI, pp. 111–123.

TRC. 1984. DOE-2: Comparison with measured data: design and operational energy studies in a new high-rise office building- Vol. 5, report prepared by Tishman Research Corporation New York to U.S. Department of Energy under contract no. DE-AC02-79CS20271, March.

Troncoso, R. 1997. A hybrid monitoring-modeling procedure for analyzing the performance of large central chilling plants. *IBPSA Conference Proceedings*, Prague, Czech Republic, Sept. 13–15.

Waltz, J.P. 2000. *Computerized Building Energy Simulation Handbook*. Lilburn, Ga.: Fairmont Press.

Wei, G., M. Liu, and D.E. Claridge. 1998. Signatures of Heating and Cooling Energy Consumption for Typical AHUs. *Eleventh Symposium on Improving Building Systems in Hot and Humid Climates*, Forth Worth, TX, June, pp. 387–402.

Winkelmann, F.C., B.E. Birdsall, W.F. Buhl, K.L. Ellington, A.E. Erdem, J.J. Hirsch, and S. Gates. 1993. DOE-2 Supplement, Version 2.1E, Lawrence Berkeley National Laboratory, November.

Yoon, J., E.J. Lee, and D.E. Claridge. 2003. Calibration procedures for energy performance simulation of a commercial building. *Journal of Solar Energy Eng.* 125:251–257.

Yuill, G.K., and C.P. Wray. 1990. Overview of ASHRAE TC 4.7 annotated guide to models and algorithms for energy calculations relating to HVAC equipment. *ASHRAE Transactions* 96(1):91–97.