

AN IMPROVED PROCEDURE FOR DEVELOPING CALIBRATED HOURLY SIMULATION MODELS

Tarek E. Bou-Saada & Jeff S. Haberl
Energy Systems Laboratory
Texas A&M University System
College Station, TX

ABSTRACT

In order to improve upon previous calibration techniques, this paper presents new calibration methods including a temperature bin analysis to improve hourly x-y scatter plots, a 24-hour weatherdaytype bin analysis to allow for the evaluation of hourly temperature and schedule dependent comparisons, and a 52-week bin analysis to facilitate the evaluation of long-term trends. In addition, architectural rendering is suggested as a means of verifying the building envelope dimensions and external shading placement. Several statistical methods are also reviewed to evaluate the goodness-of-include percent difference calculations, mean bias error (MBE), and the coefficient of variation of the root mean squared error (CV(RMSE)).

The procedures are applied to a case study building located in Washington, D.C. Nine months of hourly whole-building electricity data and site-specific weather data were measured and used with the DOE2AD building simulation program to test the new techniques. Use of the new calibration procedures were able to produce an hourly MBE of -0.7% and a CV(RMSE) of 23.1% which compare favorably with the most accurate hourly neural network models.

INTRODUCTION

Computers and programmable calculators have been used extensively during the past three decades as effective heating, ventilating, and air-conditioning (HVAC) design tools to supplement tedious manual energy calculations. Initially, super mini-computers or mainframe computers were required. Unfortunately, simulation was restricted to research organizations supported by public funding, large consulting firms and manufacturers, and a relatively few academic research institutions because of large expenses incurred for hardware and the limited availability of both software and computer personnel (Ayers and Stamper 1995). As computing technology has become affordable, engineers and architects have taken advantage of hourly simulation programs on desktop

personal computers (PCs) that can inexpensively and quickly perform load calculations (ASHRAE 1991). Eventually, simulation packages began to be used for retrofit evaluation in existing buildings and required calibration to measured data.

With increased use of building energy simulation programs for evaluating energy conservation retrofits, calibration of the simulation program to measured data has been recognized as an important factor in substantiating how well the model fits data from a real building. The calibration of a simulation to measured monthly utility data has been the preferred method for many years. Recently, studies have reported calibrated models using hourly measured data. Most of the previous methods have relied on very simple comparisons including bar charts, monthly percent difference time-series graphs, and monthly x-y scatter plots. Unfortunately, at hourly levels of calibration, many of the traditional graphical calibration techniques become overwhelmed with too many data points which makes it difficult to determine the central tendency of the black cloud of data points. A few advanced methods have been proposed including carpet plots and comparative 3-D time-series plots.

Calibrating computer models to actual metered data is not a new practice. As early as 1970, recommendations were made to calibrate models based on measured data (Ayers and Stamper 1995). Some researchers and engineers have attempted to compile "how to" manuals and methods in order to simplify this task; however, in almost all cases the end result typically falls short of a complete procedure (Hsieh 1988; Kaplan et al. 1992; Bronson et al. 1992). To date, no consensus standards have been published on calibration procedures and methods that can generically be used on a wide variety of buildings. Historically, actual calibration has been an art form that inevitably relies heavily on user knowledge, past experience, statistical expertise, engineering judgment, and an abundance of trial and error.

One 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 accuracy for electricity is near 5% per month), the author does not reveal the techniques used other than stating the final result. Hourly or daily error values are seldomly reported. Even in cases when error estimates are presented, the methods and equations used to obtain the comparisons are not. A complete review of the methods is provided in Bou-Saada (1994a).

METHODOLOGY

To simulate and calibrate a computer model to measured energy data, several stages were completed. First, site specific hourly weather data were collected to improve the calibration. The weather data included dry bulb temperature, relative humidity, and peak wind speed gathered from the nearby National Weather Service (NWS) station at the Washington National Airport and global solar radiation data measured on-site. A routine was developed to convert global solar radiation measured on an 18° south facing tilt into global horizontal beam and diffuse radiation (Bou-Saada 1994a; 1994b). The weather data were then joined into a single datafile and packed onto a TRY weather tape (Bronson 1992) for use with the DOE-2 simulation program. Next, energy use and solar data from two dataloggers made by different manufacturers were merged with the NWS data for use with the graphical techniques described shortly.

The calibration procedure also entailed creating a DOE-2 input file based on information obtained from site visits and architectural as-built plans. The hourly output report from the simulation was processed with post-processing routines (Bronson 1992) specially modified for this work. This avoids the long-term data averaging encountered with monthly simulation comparisons (Hinchev 1991).

Information Required for Calibrating DOE-2

Figure 1, adapted from Bronson (1992), is a general overview of the calibration process used for this research. A detailed discussion of the data processing routines is contained in Bou-Saada (1994a). The grouping at the top of Figure 1 includes all the required input information to produce a DOE-2 simulation including: DOE-2 reference manuals, as-built drawings, information from site visits, utility billing data, on-site measured energy consumption data and weather data. A typical DOE-2 input file may be produced using any number of computerized ASCII text editors. Any DOE-2 simulation usually requires a visit to the standard set of DOE-2 reference manuals to observe correct BDL syntax (the DOE-2 input format)

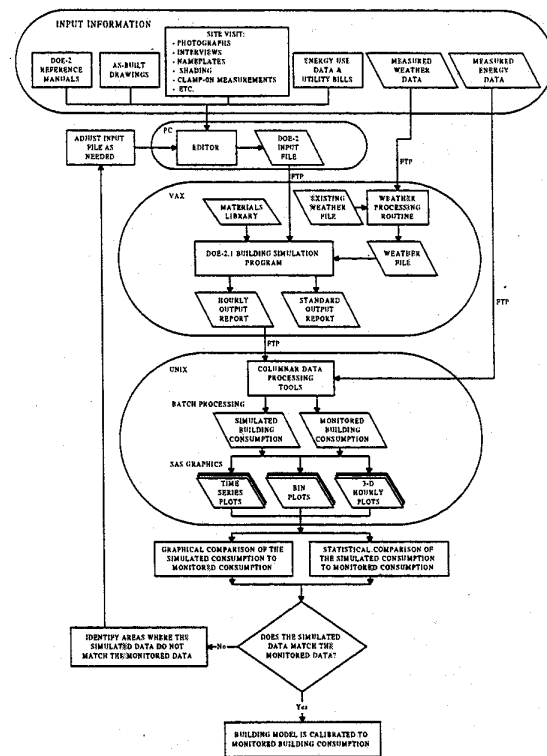


Figure 1 DOE-2 Calibration Procedure.

and mandatory BDL requirements (LBL 1980; 1981; 1982; 1989). As-built drawings help to correctly dimension the building and calculate lighting and equipment levels. A site visit is generally essential to verify lighting counts and equipment nameplate data as well as to verify dimensions and any other miscellaneous discrepancies. The site visit should also include photographs of the building's surroundings for establishing shading calculations and detailed interviews with occupants, engineers, architects, and building operations personnel. Also included in the site visit should be shading measurements, and one-time RMS clamp-on Watt measurements of key pieces of equipment to verify actual power requirements.

A major part of the site visit includes the gathering of energy use data and/or monthly utility bills. Either is acceptable, but neither is a strict requirement to compile a first pass at the input file. An HVAC system air balance report is also helpful when describing the zone air flow rates. On-site weather data measured for the simulation period have also been shown to significantly improve the simulation by Haberl et al. (1995). In cases where no weather data are available, standard weather tapes such as TRY and TMY may be used. Finally, prior experience with the DOE-2 program plays a crucial factor that can benefit the user

in avoiding commonly made mistakes. Many problems with the input file may be avoided simply by having prior knowledge of program expectations as well as a thorough engineering understanding of HVAC systems and buildings in general.

After a simulation was performed in the current procedure, the Statistical Analysis Software (SAS 1989) program was used to analyze the goodness-of-fit and produce graphical feedback of the simulation progress. This includes time-series plots, bin plots, and three-dimensional hourly plots for further analysis. This allowed for a graphical comparison of the simulated consumption to the monitored consumption. Also included in this paper is a technique for DOE-2 calibration that supplements the graphical comparison with a statistical comparison of the simulated and measured consumption which is described later with more detail. With this information now processed, the user can then decide if the model is calibrated to an acceptable level, and of equal importance, where the remaining mismatch may be located. This second feature is accomplished with the assistance of the calibration tools described here. If it is determined that a simulation is not fully calibrated, the areas where the simulated data do not match the measured data must be identified and adjusted in the input file. The DOE-2 program is run again and the data processed for comparison until an acceptable calibration is reached.

Architectural Rendering

Several software programs have recently become available for purposes of architectural rendering or viewing of building simulation input files. One such program, DrawBDL (Huang 1993), was used to verify the building envelope descriptions used in the DOE-2 input file. The building shown in Figure 2 shows a view of the case study building using the DrawBDL program. The software also includes such capabilities as rotating the building to allow for viewing from any direction with a three-dimensional perspective, a plan view, an elevation view, and a wire frame view. With a BDL visualization tool, each case study building envelope surface and shading surface can be inspected for proper placement, size, and orientation. This type of checking could not easily be done prior to the creation of such architectural rendering tools.

New Calibration Graphing Methods

Formerly, DOE-2 users were confined to using simple time-series plots (Hsieh 1988; Hunn et al. 1992; Reddy 1993) where simulated and actual data are

superimposed upon the same graph for a short period of time. Although two-dimensional time-series plots are useful for determining certain features, a special problem exists when plotting long-term hourly time-series data. In such cases, direct comparison becomes ineffective for all practical purposes because it is very difficult to identify individual hourly data points.

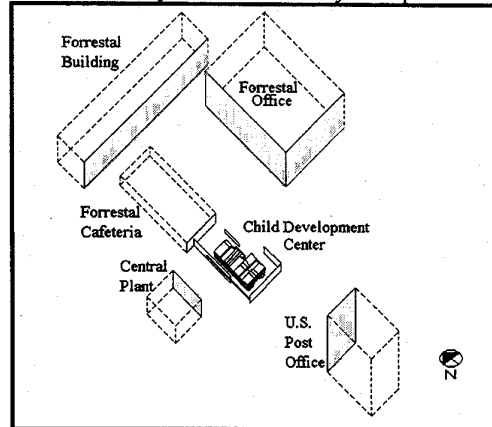


Figure 2 The U.S.D.O.E. Forrestal complex and Surrounding Area. The Solid Planes Represent Shading from Buildings, Walls, and Trees.

One improvement over past graphical techniques is shown in Figure 3 which shows an example of a binned analysis that was modified for this paper from indices developed by Abbas (1993). The superimposed and juxtaposed binned box-whisker-mean plots 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 this plot is that the data are statistically binned by temperature. This feature allows for the bin-by-bin goodness-of-fit to be evaluated. By using the box-whisker-mean plot combined with a scatter plot, one can visualize the data as a whole while simultaneously seeing the effects of the outliers in specific situations (Tukey 1977; Cleveland 1985).

In Figure 3 only the weekday data are plotted using a technique developed by Abbas (1993) and modified for this paper that includes a combination of vertical and horizontal juxtapositioning, temperature-based box-whisker-mean bins and super-positioning of the mean bin line in the lower right graph. Similar analysis can be performed with weekend and holiday data (Bou-Saada 1994a). In the upper left graph the hourly measured whole-building electricity use is plotted against hourly ambient temperature. In the

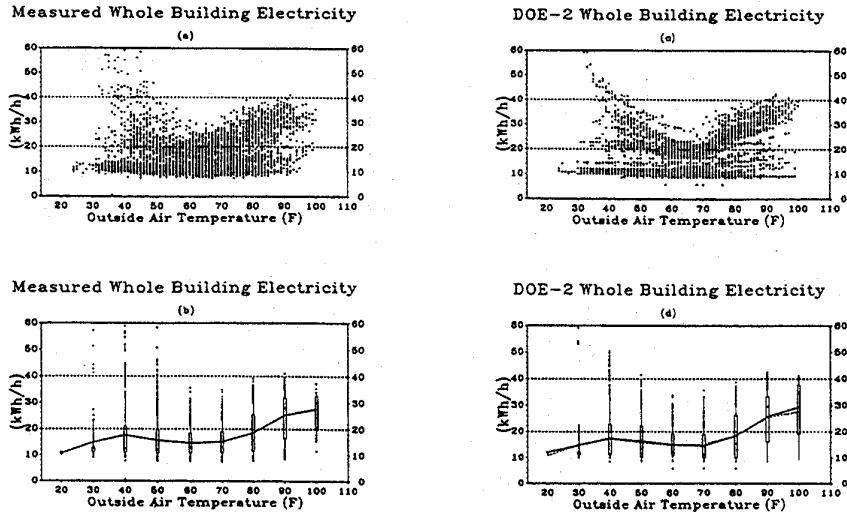


Figure 3 Weekday Temperature Bin Calibration Plots.

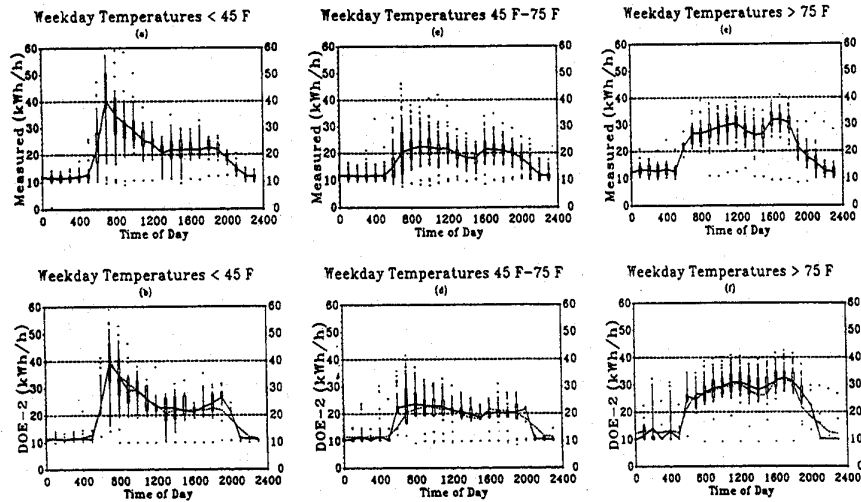


Figure 4 Weekday 24-Hour Weather Daytype Box-whisker-mean Plot.

upper right graph, the corresponding DOE-2 simulated electricity data for the same period are shown. Below each scatter plot are binned box-whisker-mean plots. These plots show the whole-building electricity consumption as a function of outdoor temperature bins divided into 10° F segments. One final feature of these plots is that the measured data mean is superimposed as a dashed line onto the calibrated DOE-2 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 interquartile range (i.e., the distance between the 25th and 75th percentiles) represents the hourly variation in a given bin.

Figure 4 is an example of a newly developed weekday 24-hour weather daytype box-whisker-mean

plot that shows the whole-building electricity use versus the hour-of-the-day for both the measured data and the DOE-2 simulated data in three weather daytypes. A similar analysis can also be performed with weekend and holiday data. The weather daytypes arbitrarily divide the measured data into temperatures below 45°F, between 45°F and 75°F, and above 75°F. The original concept for this plot can be traced to the weather daytype analysis developed by Hadley (1993).

This additional calibration procedure allows a DOE-2 user to view and analyze the weather and schedule dependent hourly energy use. The solid line in parts (b), (d), and (f) is the simulated mean. The dashed line is the measured mean line from parts (a), (c), and (e) that is superimposed onto the simulated data. These plots confirmed that the building's 24-

hour electricity profiles are strongly influenced by the ambient temperature. The plots also provide a more efficient method of viewing the data based on heating only, no heating or cooling, and cooling only modes.

Comparative 3-D Surface Calibration Plots

In Figure 5 comparative 3-D surface plots show the monitored data in part (a) and the DOE-2 simulated data in part (b) for the April to December 1993 period. Figure 5(c) shows positive-only values of the measured data subtracted from the DOE-2 predicted data while Figure 5(d) shows positive only values of the DOE-2 predicted data subtracted from the measured data.

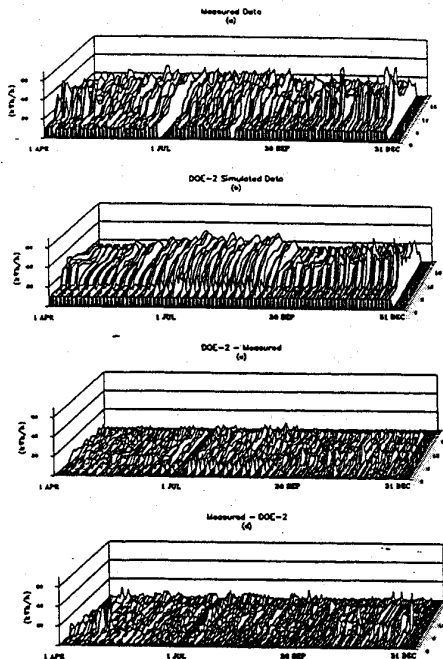


Figure 5 Comparative Three-dimensional Plots. (a) Measured Data. (b) Simulated Data. (c) Simulated-Measured Data. (d) Measured-Simulated Data.

Individual hourly differences may be visually detected over the entire simulation period allowing the user to recognize patterns in the comparisons such as DOE-2's over-predictions in the spring and fall mornings and afternoons and both over- and under-predictions in the late evening throughout the year. An obvious benefit of such plots are their ability to aid in the identification of oversights such as a daylight savings shift or misalignment of 24-hour holiday profiles (Bronson et al. 1992). One negative drawback associated with these graphs is the difficulty in viewing exact details such as a specific hour on a specific day.

By inspecting Figure 5(d), it is clear that DOE-2 is under-predicting early morning unoccupied electricity

use in the Spring between April 1 and July 1 1993. One possible cause for this is the continuous operation of the building's cross-wired exhaust fans that were supposed to be off, but operated continuously. This draws in cool outdoor air and activates the heating system when the temperature falls below the setback temperature of 55° F. DOE-2 will not simulate this condition properly because the program cannot account for negative building pressures. Therefore, no extra heating is evident in the simulated plot. It is felt that this also causes the under-prediction of energy use in the early morning hours of April and December as shown in Figure 5(a). The ragged on/off "pickets" during July through September are a characteristic of DOE-2's hourly calculation algorithm.

Figure 6 shows the same data as the 3-D surface plots in Figures 5(a) and (b). However, Figure 6 displays the energy usage using 52-week time-series box-whisker-mean bins instead of temperature bins. The measured data are shown in part (a) and the DOE-2 simulation is shown in part (b). The x-axis in Figure 6 is the simulation week number; for this simulation, week "0" begins on April 1. The y-axis shows the whole-building electricity use in both (a) and (b).

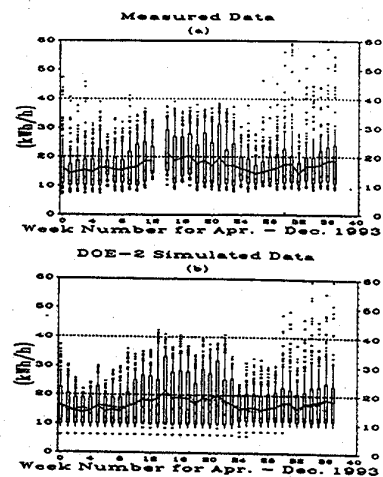


Figure 6 52-Week Binned Box-whisker-mean Plots.

Figure 6 utilizes graphical superposition of the mean line from Figure 6(a) (dashed line) upon Figure 6(b) to further improve the viewing efficiency of the graph. It follows a markedly similar path traced by the simulated data mean line represented by the solid line. Fine differences can also be seen in those points above the 90th percentile in weeks 0-5. Those points represent the hours of electrical resistance heating in the measured data that occurred when the staff manually switched on the baseboards in the toddler rooms to preheat the rooms prior to the arrival of the children.

When evaluating each set of paired weekly bins, it may be concluded that the average weekly simulated data seems to consistently track the average weekly measured data since the simulated box sizes (i.e., inter-quartile range) do not deviate significantly from the measured box sizes. However, the minimum simulated data limits are consistently lower in the DOE-2 plot (Figure 6(b)) further emphasizing the difficulties encountered in predicting the nighttime building shut down schedule. The consistency of the medium and minimum points seen in the winter is a characteristic of the DOE-2's rigid scheduling. It would appear that beginning in approximately week 12 and ending in week 17, the measured nighttime HVAC setback mode was overridden due to the zone temperature exceeding the control system upper setpoint temperature limit.

Statistical Calibration Methods

In the previous research, Haberl et al. (1995) and Bronson (1992) summed monthly simulation results and verified the calibration via a percent difference. Torres-Nunci (1989) and Hinchey (1991) only declared the model "calibrated", submitted hourly graphs to demonstrate the goodness-of-fit, and provided numerical differences only in the form of \pm monthly differences. The problem with this approach is that the \pm monthly difference does not provide a fine enough goodness-of-fit indicator and is in fact misleading because it can indicate a near perfect fit when there is still considerable hourly error in the calibration. Therefore, in the interest of furthering the calibration procedures, several statistical calculations were compared including a monthly mean difference, an hourly mean bias error (MBE) for each month, an hourly root mean squared error (RMSE) reported monthly, and an hourly coefficient of variation of the root mean squared error (CV(RMSE)) (Kreider and Haberl 1994a; 1994b). These indices were previously shown to be useful in comparing hourly neural network models against measured hourly use.

The percent difference is a simple calculation whereby a difference for each monthly measured and simulated energy use total is taken and divided by the measured monthly total consumption. This index is the typical value reported for most DOE-2 predictions (Diamond and Hunn 1981; Kaplan et al. 1990; Bronson 1992; McLain et al. 1993; Haberl et al. 1995).

The mean bias error, MBE (%) (Kreider and Haberl 1994a; 1994b), is a method with which to determine a non-dimensional bias measure (the sum of errors), between the simulated data and the measured data for each individual hour (Katipamula 1994).

The root mean squared error is typically referred to as a measure of variability, or how much spread exists in the data. For every hour, the error, or difference in paired data points is calculated and squared. The sum of squares errors (SSE) are then added for each month and for the total periods and divided by their respective number of points yielding the mean squared error (MSE); whether for each month or the total period. A square root of the result is then reported as the root mean squared error (RMSE).

The coefficient of variation of the root mean squared error, CV(RMSE) (%) (Draper and Smith 1981) is essentially the root mean squared error divided by the measured mean of the data. It is often convenient to report a non-dimensional result. CV(RMSE) allows one to determine how well a model fits the data; the lower the CV(RMSE), the better the calibration (the model in this case is the DOE-2-predicted data). Therefore, a CV(RMSE) is calculated for hourly data and presented on both a monthly summary and total data period.

The purpose of calculating the CV(RMSE) and comparing the results with the standard percent difference is to demonstrate that a percent difference report may be misleading. Since these calculations are usually shown for monthly simulations or even total simulation periods, the reader is never certain if the model is a true representation of the actual building or if the \pm errors have canceled out. If one examines the hour-by-hour data results, it would be evident that each pair of points would in all likelihood be dissimilar and in some cases be significantly different, despite using the same measured weather data to drive the simulation model. Reporting monthly data therefore does not take into account the canceling out of individual differences observed when the simulation over-predicts during one hour and under-predicts during the next hour by approximately the same amount. More information on the statistical indices used in this paper can be found in Bou-Saada (1994a).

The statistical results for the DOE-2 model can be seen in Table 1. Additional tables of weekday occupied, weekday unoccupied, and weekend statistics can be found in Bou-Saada (1994a). This table shows monthly and total measured kWh, measured mean kWh, total simulated kWh, and simulated mean kWh. More importantly, the table compares the monthly and total percent difference to the hourly MBE, hourly RMSE, and hourly CV(RMSE) for each month and the total period. The MBE, RMSE, and CV(RMSE) statistics are calculated for each hour and summed for each month and for the total period, the weekday occupied period, the weekday unoccupied period, and

the weekend period. This section explains why the percent difference statistics alone typically reported in most calibrations may not be true representations of the measured data, and why the hourly CV(RMSE) calculations may be more appropriate.

Table 1 Total Period Statistics Summary.

Month	Total Measured (kWh)	Mean (kWh)	Total Simulated (kWh)	Mean (kWh)	Total Difference %	Hourly MBE %	Hourly RMSE (kWh)	Hourly CV(RMSE) %
April	10,606	15.3	10,243	15.0	-3.4	-3.4	4.2	26.9
May	11,507	15.5	10,948	14.7	-4.9	-4.9	3.4	21.8
June	11,568	17.4	11,313	17.0	-2.2	-2.2	3.5	20.2
July	11,247	19.9	11,052	19.6	-1.7	-1.7	4.9	24.7
August	12,621	18.6	12,531	18.5	-0.7	-0.7	4.0	21.7
September	11,885	16.5	12,012	16.7	1.1	1.1	3.7	22.5
October	11,216	15.1	11,413	15.3	1.8	1.8	3.5	23.2
November	11,944	16.7	12,202	17.0	2.2	2.2	4.0	23.7
December	9,220	18.3	9,337	18.5	1.3	1.3	4.2	22.8
Total	101,814	16.9	101,051	16.8	-0.7	-0.7	3.9	23.1

Table 1 is a summary of the statistics for the total calibration period from April 1, 1993 through December 21, 1993. The contrast between percent difference and CV(RMSE) clearly can be seen when comparing May and July. In May, a simulation under-prediction exists to the tune of 4.9% while the CV(RMSE) is a relatively low 21.8% (i.e., low when compared to the other months in the table). July shows an under-prediction difference of 1.7% whereas the CV(RMSE) is a high 24.7%. The hourly peak and trough cancellation effect described earlier is evident here showing an overall low percent difference in July and an hourly high CV(RMSE). Despite all the months generally showing a low percent difference, the hourly CV(RMSE) consistently remained in the 20 to 27% range. This is still well within the range reported in a recent ASHRAE sponsored simulation competition that compared various modeling techniques including the winning model, a principle component, Bayesian non-linear complex neural network (Kreider and Haberl 1994a; 1994b; Mackay 1994). Clearly, calibrations should be performed using both hourly CV(RMSE) and hourly MBE calculations rather than solely with the total percent difference calculation. These statistical indices serve to shed light on improving calibration techniques well into the future.

Model Fine Tuning Progress

To represent the adequacy of the DOE-2 models from an initial workable model to a final calibrated model, each major change to the input file was documented through the research phase. A condensed set of iterations is graphically displayed in Figure 7 which shows the impact of most of the major modifications made to the model. The corresponding history of the changes is listed in Table 2 which details the hourly total MBE and CV(RMSE) for each model and modifications. In reality, about 100 different iterations were run, however due to space constraints,

summary groups of the runs are shown in the figure and table.

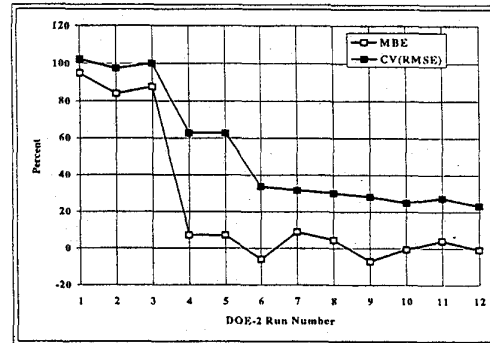


Figure 7 Tuning Progress with Input Modification.

The first DOE-2 model to run without errors was considered Run #1 and is labeled as the base model. This run consisted of the best information that could be obtained from as-built drawings, equipment lists, audit findings, and DOE-2 default values. As Table 2 shows, the error is large with a 94.8% MBE and a 102.4% CV(RMSE). Run #2 showed slight improvements with the change of infiltration from 0.0 to 0.1 air changes per hour (ACH), extraction of hourly data from the PLANT sub-program instead of from the SYSTEMS sub-program, and modifications made to the indoor use of indoor lighting and equipment. Correcting the cooling and heating capacities to the manufacturer's specifications in Run #3 worsened the MBE and CV(RMSE) slightly calling for adjustments to the model elsewhere. The most significant modification to the model occurred with Run #4 when indoor lighting and equipment schedules were lowered to match the measured daily profiles. Clearly, one may conclude for this building that the lighting and equipment schedules had a significant impact on the simulation. This is consistent with advice given by Hsieh (1988), Kaplan et al. (1992), and Griffiths and Anderson (1994).

Run #5 was the first simulation to make use of the DrawBDL software to verify the envelope. As was the case with this simulation, corrections to wall locations were minor because they had been placed fairly close to where they belong in the first place. Run #5 also did not include the effects of shading at the site, thus not affecting the output significantly. Run #6 was another notable improvement in the process in which a large change in the MBE and CV(RMSE) was observed. This run incorporated a combination of adjustments to the ventilation air changes per hour, exhaust air volume, and power level corrections as specified by the balance report, adjustment of lighting

power in one zone, addition of building shading, addition of specified ground temperatures, and the adjustment of cooling and heating schedules from 24-hour operation to "off at night" schedules. The cooling and heating load adjustment would in all probability be the most influential factor of this list by reducing the PLANT load significantly in the night setback mode.

Table 2 DOE-2 Input File Changes with Each Run.

Run	Change Made To Model	MBE (%)	CV(RMSE) (%)
1	Base model	94.8	102.4
2	Adjusted infiltration from 0 to 0.1 ACH Added PLANT sub-program to model for hrly reports Modified indoor light & indoor equipment schedules from 10% night/100% day to 35% night/100% day	84.0	97.4
3	Set correct cooling & heating capacities (as per spec.)	87.6	100.0
4	Modified indoor light & indoor equipment schedules (adjusted schedules downward)	7.2	62.8
5	Corrected walls as per DrawBDL Added sizing option -> "ADJUST-LOADS"	7.1	62.8
6	Adjusted air changes from 0.1 ACH to 0.6 ACH Set exhaust CFM as per balance report Set exhaust kW as per balance report Adjusted lighting kW in 1 zone Added surrounding building shading to input file Added ground temperatures to input file Adjusted cooling & heating schedules from 24-hour operation to off at night	-5.9	33.7
7	Adjusted indoor equipment schedules	8.9	31.7
8	Set thermostat settings as per EMCS specs. (summer:74-day, 80-night; winter:72-day, 55-night) Added baseboard heat (7504 BTUH) in north zone	4.6	30.1
9	Adjusted indoor light & indoor equipment schedules Modified HVAC schedule to winter/summer setting Added DHW schedule Adjusted fan schedule from 24-hour operation to off at night	-6.9	28.2
10	Removed two buildings from building shading (they were not close enough to affect shading) Adjusted indoor light & indoor equipment schedules Adjusted DHW schedule	-0.4	25.0
11	Adjusted indoor light & indoor equipment schedules Increased return air CFM as per balance report (this corrected an error of not simulating enough fans)	4.1	27.4
12	Added return air fan as per balance report Adjusted indoor light & indoor equipment schedules	-0.7	23.1

As Figure 7 and Table 2 show, the remainder of the runs were fine-tuning runs with no large scale change in either the MBE or CV(RMSE). Most of the changes made to the model included minor schedule adjustments until Run #12. By reviewing the input file, an error was detected in the original assumptions concerning the SYSTEMS input file. It was discovered that the incorrect air volume was originally specified by the inadvertent omission of the return air fan and volume. Correcting this error required a recalibration effort by adjusting the lighting and equipment schedules to compensate for the return air fan correction. Run #12, the final run, was then considered the "calibrated" version of which all the graphs in this thesis are based on. The MBE of -0.7% and an hourly CV(RMSE) of 23.1% were considered acceptable for the research. Previous work reported by Kreider and Haberl (1994a; 1994b) showed that even the very best empirical models (i.e. artificial neural

networks used with a large commercial building) were only capable of producing CV(RMSE) in the 10 to 20% range. When comparing the neural network results to the daycare center, minute in size comparison, the CV(RMSE) of 23.1% is favorable.

For this paper, both CV(RMSE) and MBE did not appear to further improve without major modifications to the input file such as additional schedules to account for differences in the occupants' behavior.

CONCLUSIONS

This paper has investigated techniques for improving calibrated computer building energy simulation methods and has presented several new techniques for improving calibrations. The new methods include new graphical procedures and statistical goodness-of-fit parameters for quantitatively comparing simulated data to measured data. A four zone, single story electrically heated and cooled case study building was simulated with DOE-2.1D and calibrated using hourly measured whole-building electricity data and ambient weather conditions to demonstrate the new techniques.

Findings from Applying the New Techniques

The important new calibration features that were developed (Bou-Saada 1994a) include:

- The use of architectural rendering software allows for visual verification of size and placement of the building's exterior surfaces and shading surfaces to those of the actual building.
- The combined use of a statistical analysis and data display improves the assessment of non-weather dependent and weather-dependent calibrations.
- The use of a DOE-2 zone temperature comparison to measured indoor temperature confirms simulated HVAC schedules.
- The use of solar beam and diffuse data synthesized from on-site measured global solar radiation data improves the simulation.

These new techniques significantly improved the previous DOE-2 calibration methods. The long-term goal of this type of research is to eventually lead to a standardized calibration procedure that could be used on a wide variety of buildings and simulation codes. The next section provides additional details concerning recommendations for future work.

Recommendations and Calibration Procedure

During this research effort, many valuable lessons were learned by refining the simulation. As a guide to

future DOE-2 users, recommendations are provided so that calibration efforts can be improved.

The first step to simulating a building should include a site visit to acquire the following building information:

- A complete set of as-built drawings (architectural, mechanical, and electrical). Information gathered from as-built drawings includes envelope description and placement, HVAC zoning, lighting loads, and control specifications.
- An HVAC air balance report including supply and return air temperature measurements.
- Information concerning actual thermostat settings including day/night setback. If an EMCS is installed, a print of the program settings is helpful.
- Measured indoor temperature in each zone during normal operations and nighttime for heating, cooling, and intermediate seasons.
- Hourly HVAC schedules, hourly interior/exterior lighting schedules, hourly occupancy schedules, and hourly equipment schedules. Note the general operation schedule for miscellaneous activity such as dishwashing frequency, etc.
- Perform blower door test to check infiltration rates.
- Evaluate exterior shading surrounding the building including relative distance from building off-south orientation, height, and angle of shading objects.
- Measure at least seven to nine months of hourly whole-building electricity, cooling, heating, and major equipment end-uses.
- For smaller electrical equipment such as appliances and small motors, clamp-on RMS Watt measurements for 24-hours should be made.
- Survey equipment and note all specifications.
- An accurate light fixture count should include Wattage and ballast power level. Note how many lamps are typically off for extended periods and verify power with clamp-on RMS Watt meter.
- Measure local hourly weather data including relative humidity, dry bulb temperature, wind speed, and global horizontal solar radiation. An alternative can include the NWS, preferably with a weather station as close to the site as is possible. However, NWS solar data are not very useful for similar purposes, and NWS wind data are peak hourly gusts which have almost no correlation to average hourly wind speed.
- Photograph the building's exterior, interior, equipment, and surrounding area to compare with computerized rendering.
- Contact HVAC and internal equipment representative for off-peak performance specifications.

Then create the DOE-2 simulation input file by matching the building HVAC system to the nearest fixed schematic HVAC system in the DOE-2 reference manual. It is imperative to use as many measured details as possible. All DOE-2 default values should be investigated for appropriateness.

Use data from the DOE-2 hourly reports corresponding to the measured data from the building. Process the data into a single ASCII columnar datafile compatible with the comparative graphical routines presented in this paper. Finally, use the tools similar to those described in this paper (Bou-Saada 1994a) and those previously developed (Bronson 1992; Hinchey 1991) to compare the simulated energy use to the measured energy use. Iterate until the measured data matches the simulated data to a suitable level as evaluated with hourly MBE, RMSE, and CV(RMSE).

ACKNOWLEDGEMENTS

Work discussed in this paper was sponsored by the United States Department of Energy through Sandia National Laboratory under contract number AK-8010. We deeply appreciate all those involved for their contributions.

REFERENCES

- Abbas, M. 1993. Development of graphical indices for building energy data, *M.S. Thesis*, Energy Systems Rpt No. ESL-TH-93/12-02, Texas A&M University, College Station, TX.
- ASHRAE. 1991. ASHRAE handbook: 1991 HVAC applications volume. *American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Inc.*, Atlanta, GA.
- Ayers, J.M. and E. Stamper. 1995. Historical development of building energy calculations, *ASHRAE Transactions Preprint*. 101(1).
- BLAST. 1993. BLAST users manual. BLAST Support Office, University of Illinois Urbana-Champaign.
- Bou-Saada, T.E. 1994a. An improved procedure for developing a calibrated hourly simulation model of an electrically heated and cooled commercial building, *M.S. Thesis*, Energy Systems Rpt No. ESL-TH-94/12-01, Texas A&M University, College Station, TX.
- Bou-Saada, T.E. 1994b. Advanced DOE-2 calibration procedures: A technical reference manual. Energy Systems Laboratory Rpt No. ESL-TR-94/12-01, Texas A&M University, College Station, TX.
- 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): 636-

652. Energy Systems Lab Rpt. No. ESL-PA-91/05-01, Texas A&M University, College Station, TX.
- Bronson, J.D. 1992. Calibrating DOE-2 to weather and non-weather dependent loads for a commercial building. *M.S. Thesis*, Energy Systems Rpt No. ESL-TH-92/04-01, Texas A&M University, College Station, TX.
- Cleveland, W.S. 1985. The elements of graphing data. Wadsworth & Brooks/Cole, Pacific Grove, CA.
- 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.
- Draper, N. and H. Smith. 1981. Applied regression Analysis. 2nd. Ed. John Wiley & Sons, NY, NY.
- EPRI. 1992. Building energy analysis program: User's guide and technical reference manual. Electric Power Research Institute. Palo Alto, CA 94304.
- Griffiths, D.M. and K.J. Anderson. 1994. Commercial building/system characteristics sensitivity analysis. *Proceedings from the ACEEE 1994 Summer Study on Energy Efficiency in Buildings*, Vol. 5, pp. 5.105-5.112. Washington, DC.
- Haberl, J.S., J.D. Bronson, and D.L. O'Neal. 1995. An evaluation of the impact of using measured weather data versus TMY weather data in a DOE-2 simulation of an existing building in central Texas. *ASHRAE Transactions*. 101(1). ESL Rpt No. ESL-TR-93/09-02. College Station, TX.
- Hadley, D.L. 1993. Daily variations in HVAC system electrical energy consumption in response to different weather conditions. *Energy and Buildings*. 19: pp. 235-247.
- Hinchey, S.B. 1991. Influence of thermal zone assumptions on DOE-2 energy use estimations of a commercial building. *M.S. Thesis*, Energy Systems Rpt No. ESL-TH-91/09-06, Texas A&M University, College Station, TX.
- Hsieh, E.S. 1988. Calibrated computer models of commercial buildings and their role in building design and operation. *M.S. Thesis*, PU/CEES Rpt No. 230, Princeton University, Princeton, NJ.
- Huang & Associates. 1993. DrawBDL user's guide. 6720 Potrero Ave., El Cerrito, California, 94530.
- Hunn, B.D., J.A. Banks, and S.N. Reddy. 1992. Energy analysis of the Texas Capitol restoration. *The DOE-2 User News*. 13(4): 2-10.
- Kaplan, M.B., B. Jones, and J. Jansen. 1990. 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.
- Katipamula, S. 1994. Personal communication, Richland, WA: Pacific Northwest Laboratory.
- Kreider J.F. and J.S. Haberl. 1994a. Predicting hourly building energy use: The great energy predictor shootout-overview and discussion of results. *ASHRAE Transactions*. 100(2):1104-1118.
- Kreider J.F. and J.S. Haberl. 1994b. Predicting hourly building energy usage. *ASHRAE Journal*. pp. 72-81.
- LBL. 1980. DOE-2 User Guide, Ver. 2.1. Lawrence Berkeley Laboratory and Los Alamos National Laboratory, Rpt No. LBL-8689 Rev. 2; DOE-2 User Coordination Office, LBL, Berkeley, CA.
- LBL. 1981. DOE-2 Engineers Manual, Ver. 2.1A, Lawrence Berkeley Laboratory and Los Alamos National Laboratory, Rpt No. LBL-11353; DOE-2 User Coordination Office, LBL, Berkeley, CA.
- LBL. 1982. DOE-2.1 Reference Manual Rev. 2.1A. Lawrence Berkeley Laboratory and Los Alamos National Laboratory, Rpt No. LBL-8706 Rev. 2; DOE-2 User Coordination Office, LBL, Berkeley, CA.
- LBL. 1989. DOE-2 Supplement, Ver 2.1D. Lawrence Berkeley Laboratory, Rpt No. LBL-8706 Rev. 5 Supplement. DOE-2 User Coordination Office, LBL, Berkeley, CA.
- Mackay, D.J. 1994. Bayesian nonlinear modeling for the energy prediction competition. *ASHRAE Transactions*. 100(2):1053-1062.
- McLain, H.A., S.B. Leigh, and J.M. MacDonald. 1993. Analysis of savings due to multiple energy retrofits in a large office building. Oak Ridge National Laboratory, Rpt No. ORNL/CON-363, Oak Ridge, TN.
- Reddy, S.N. 1993. Determination of retrofit energy savings using a calibrated hour-by-hour building energy simulation model. *M.S. Thesis*. The University of Texas at Austin, Austin, TX.
- SAS Institute, Inc. 1989. SAS reference manual. SAS Institute, Inc., SAS Circle, P.O. Box 8000, Cary, NC, 27512-8000.
- Torres-Nunci, N. 1989. Simulation modeling energy consumption in the Zachry Engineering Center. *M.S. Project Report*, Texas A&M University, College Station, TX.
- Tukey, J.W. 1977. Exploratory data analysis. *Addison-Wesley Publishing Co.* Reading, MA.