USING DATA VISUALIZATION TOOLS FOR THE CALIBRATION OF HOURLY DOE-2.1 SIMULATIONS

James A. McCray, Patrick L. Bailey, & Jedd L. Parker - RLW Analytics Inc. Sonoma, CA USA
Richard Gillman - EPRI Retail Market Tools and Services, Palo Alto, CA USA

ABSTRACT
The volume of annual, monthly, and hourly simulation output developed by building simulation packages such as DOE-2.1 presents the building modeler with significant challenges. Developing hourly total load and end-use estimates of building performance calibrated to 15 minute or hourly metered total load or end use data requires new analytic tools that allow the modeler to quickly review the results and make iterative changes to the models. This paper suggests that engineering model calibration can quickly go beyond monthly customer billing data to minimize self-canceling errors where under-predictions in one end-use cancel out overpredictions in another end-use. New visual data analysis techniques (VDA) can provide the modeler with tools that improve the accuracy of hourly total load and end end-use data, while minimizing modeling costs. This paper presents four calibration approaches and supporting statistical measures that can be used as iterative diagnostic tools for the building modeler.

INTRODUCTION

Until the releases of DOE-2.1D and later DOE-2.1E in 1994, the development of large volumes detailed 8760 hour end-use data was difficult, undocumented, and subject to the limitations of post-processor analysis tools. New, more effective post-processors can now easily access and view the hourly data, allowing the modeler to quickly assess the results and make appropriate changes in the next simulation run. As recently as 1990, the calibration of a DOE-2 model to metered end-use data might require four man-months to complete and was prohibitively expensive.[1]

EPRI's "Engineering Methods for Estimating the Impacts of Demand-Side Management Programs, Volume 1" suggests that model calibrations must go beyond matching monthly billing data to minimize unapparent but self-canceling errors in underlying end-use estimates. The manual states that under-predictions for one end use may cancel out over predictions for another end use, resulting in apparently accurate total load data that closely matches monthly data but incorrectly disaggregates hourly demands at the end-use level[2]. A recent ASHRAE Journal article suggested that model predictions can be compared with known total load and end-use metered data using standard statistical measures and graphical "Visual Data Analysis" (VDA) techniques.[3] A 1994 ACEEE paper discusses the calibration of DOE-2.1 simulations with end-use data at the monthly kWh energy level, and suggests that calibrations within +/- 10 to 20 percent have been the norm.[4]

This paper presents four VDA techniques developed to support Load Research, End-Use Data Development, and large scale DSM Impact Evaluations of Commercial New Construction Programs. Each is illustrated using a commercial DOE-2.1 modeling example where 15 minute interval total load data was available for use in model calibration. The table and figures referred to in this paper are located at the end of Section 4.

THE DOE-2.1E MODEL CALIBRATION APPROACH

Developing reliable calibrated DOE-2 models in a short time at a reasonable cost is one of the major challenges of any modeling project. To make sense, the overall cost of the data developed through this process must be significantly less than that of staging a supporting short-term end-use metering study. The simulation based approach must also deliver data in a more timely manner than the monitoring alternative. The use of the diagnostic calibration tools presented here, coupled with the knowledge gained during an on-site engineering assessment (audit), has helped

---

[1] Richards, D.. Wright, R. and Puckett, C., July 11, 1990, "End-Use Load Information in the Commercial Sector", End-Use Load Information and Its Role in DSM, p. 21, The Fleming Group, Syracuse, 1990.

[2] EPRI, Engineering Methods for Estimating the Impacts of Demand-Side Management Programs, Volume 1, Electric Power Research Institute, Palo Alto, California, 1992.

[3] Kreider, J. and Haberl, J., June 1994, "Predicting Hourly Building Energy Usage", ASHRAE Journal, pp. 72-81.

[4] Peterson, Hassl, et.al. 1994, "A Commissioning Cost Effectiveness Case Study," Commissioning, Operation, and Maintenance Proceedings, Volume 5, ACEEE Summer Study on Energy Efficiency in Buildings, pp 5.201-208

meet that challenge. When these VDA tools are coupled with statistical metrics that compare the hourly model results with metered data, models can be quickly and effectively improved in a short period of time, bringing the perception of "man-months" down to hours or days.

RLW's Visualize-IT™ DOE-2.1 Calibration Tools utilize four calibration approaches. Each approach is a diagnostic tool for the building modeler. They include: (1) monthly and annual kWh energy and peak kW demand comparisons; (2) 8760 hour load profile comparisons using residual analysis; (3) average day comparisons; and (4) total load kW demand vs. ambient temperature comparisons. The critical value of the approach is the immediacy and visual nature of the feedback to modeler receives, within minutes of simulation run completion. The tools are designed to work sequentially, or in an exploratory manner, moving from familiar monthly billing data to hourly total load profiles over a full year, to load profiles for typical days, and to the performance of temperature dependent systems. Each computer screen becomes a visual record of the model's predictions of building performance. They can also be used to document the iterative modeling decisions made during the calibration effort. Supporting statistics on goodness of fit between the actual billing or metered data and the model predictions are developed and used to chart the calibration effort. The following discussion illustrates how the calibration toolset is used to improve the accuracy of the model over two iterations.

The Commercial Building. The building modeled was a small office building located at approximately 3000 feet elevation in an arid mountainous area with large diurnal temperature swings. The building has approximately 10,000 square feet of conditioned space and was constructed using standard "stick frame" construction. The building is occupied Monday through Friday from approximately 8:00 AM to 5:00 PM by 60 employees with an additional 10 clients being in the building during most of those hours. The exact age of the building is unknown but is approximately 20 years. The space conditioning system for the building consists of a gas/electric packaged rooftop unit and a small heat pump which serves an 800 square foot 24 hour dispatch area.

The DOE-2.1E Simulation and Post-Processing. The DOE-2.1E simulations were run on a Sun Microsystems Sparcstation IPC. Both traditional loads, systems, and plant reports were developed by the model and output to the *office.out* file by the program. In addition, hourly kW data for 12 end-uses and total building load were output using hourly report options and the NOSUBDIR command. This approach produced ASCII 8760 hour data on each end-use. After each model run was completed, the 8760 data was quickly post-processed by SAS into a comma separated (CSV) format designed to facilitate quick loading into a PV-Wave based program called Visualize-IT™, developed by RLW Analytics, Inc. After the model run is completed, the process described above, plus loading into the calibration toolset takes about four minutes.

Model Iteration 1. The most familiar and basic look at the simulation results is provided by the monthly and annual kWh energy and peak kW demand comparisons. Here, the monthly billing data, as well as maximum billing or created kW demand and kWh energy are plotted with (+/- 5%) error bounds. and compared with the first iteration DOE-2.1E estimates of overall building performance. Statistics comparing the DOE-2E and metered total load results is provided in Figure 3-1. On a monthly basis peak kW demand is seen to be within 5 to 24 percent and monthly kWh is within 30 to 45%. Annual kWh consumption estimates are within 40 %. The results suggest that the off-peak schedules and equipment loads may be quite low, i.e. monthly and annual energy are low, but the modeler has little other information to use for improving the model.

The volume of data produced for 8760 hour total load and multiple end-use estimates can make calibration a difficult task. The second diagnostic tool utilizes imaging technology to develop color coded plots of 24 hour by 365 day kW demands, allowing the modeler to look at up to eight channels of 8760 data on a single compute screen. These 8760 hour plots are called "Energyprints™ " and represent metered total load data, DOE 2.1 simulated total load, and up to six DOE 2.1 simulated end uses. Figure 3-2 provides the first iteration EnergyPrints™ for the office building as modeled using DOE 2.1. Metered total load, DOE-2 total load, DOE-2 lighting, DOE-2 compressor AC, DOE-2 electric heat, DOE-2 miscellaneous and DOE-2 residuals as well as the outside air drybulb temperature for one year can be seen. The end-use profiles are used by the modeler to validate the reasonableness of the DOE output and guide any modifications to simulation variables and schedules that are apparent. These eight graphical representations contain over 70,000 data points and would be difficult to present using conventional graphic techniques.

In Figure 3-2, the second image from the right presented the residual load, representing the difference between the metered total load and the DOE-2.1 model predictions. The "goodness of fit" between the two loads, the residual analysis, and be developed using standard statistical methods as suggested by Krieder and Haberl in 1994.[5] Here, associated statistics on the overall fit of the DOE-2 hourly total load estimate Vs the metered total load data are computed and displayed. The primary metrics used in the calibration are the Coefficient of Variation (CV) and the Mean Bias Error (MBE). The objective of the model calibration then becomes to minimize the CV and use the MBE as an indicator of the directed of adjustment needed. In this first iteration, the CV for the total load was 0.3196 and the MBE was 0.3977 %. The initial mean total load was 31.93 kW while the mean of the residuals was 12.699 kW. The residual load EnergyPrint is used to understand the end-use characteristics of the calibration adjustment required. Patterns in the residual load print can be seen in the associated end-use estimates, allowing the modeler to quickly make the right changes. During the next iteration, these statistics will be used to measure the improvement in the model.

The calibration tools also provide an interactive capability for the analyst specific days can be selected for inspection, from any of the end-use and total load Energyprints™. The hourly profiles for a selected day are displayed in a kW demand vs. hour plot. An example of this is shown on the left in Figure 3-3.

Average day comparisons are used to smooth out the 8760 data into weekday and weekend profiles that provide insight into the scheduling and operating characteristics of the building, suggesting changes to the model for the next iteration. Figure 3-4 provides the average day profiles for the first iteration.

The calibration tool also produces plots of kW demand Vs ambient outdoor temperature which can be used to determine building balance points, modes of operation, and to compare metered system performance with model predictions. The plot on the right in Figure 3-3 shows the performance of the building in the heating and cooling seasons, plotted against ambient dry-bulb temperature. These plots provide insight into the actual and predicted performance curves of the buildings heating and cooling systems.

Based on the review of model results using the four diagnostic tools the modeler reduced the connected lighting and equipment loads from 3.0 W/ft2 to approximately 2.6 W/ft^2. In addition, a limited attempt was made to increase HVAC hours of operation by increasing the operating hours from approximately 14 to approximately 16 hours per day, while retaining night setbacks. The magnitude of the peak load is close to the metered total load, although the DOE-2 annual energy consumption is still considerably below the metered energy total. It is obvious from the single day end-use profile that the scheduling is still not correct. Off peak energy/demand is still higher for the metered case than for the modeled case. The approach taken in iteration 2 is to let the HVAC system run wild (i.e.: no thermostatic setbacks)

[5] Krieder, J. and Haberl, J, page 79.

Model Iteration 2. Figure 3-5 provides a comparison of metered and DOE-2.1 simulation data on a monthly and annual basis for iteration 2 in the modeling process. Here the calibration of kWh energy has improved from 40 percent to 18.8 percent with monthly fits ranging from 10% to 25%. The calibration of kW demand has improved to between -7 % to 12%. The kW results are quite close, while the kWh results can be improved more, they indicate the model is now picking up more of the off-peak loads.

Figure 3-6 provides the second iteration EnergyPrints™ for the office building as modeled using DOE 2.1. The hourly profiles are shown in Figure 3-7. In Figure 3-6, the third image from the left presented the residual load, representing the difference between the metered total load and the DOE-2.1 model predictions. The "goodness of fit" between the two loads, the residual analysis, for the second iteration indicates that the CV for the total load was 0.267% and the MBE was 0.188 %. The mean of the total load remains the same at 31.93 kW while the mean of the residuals drops to 6.022 kW a 50 % improvement, indicating that the model is still underpredicting. This represents a significant improvement in the model in the second iteration alone. Inspection of the residual load image suggests that the differences in the model are less random and may be accounted for by lighting and equipment scheduling in the next iteration.

The average day comparisons indicate that the calibration of the loads has improved dramatically and suggests that the underprediction is still related to modeled loads during off-peak unoccupied periods and weekends. Figure 3-8 provides the average day profiles for the second iteration.

The plots of kW demand Vs ambient temperature shown on the right in Figure 3-7 illustrate the performance of the building systems Vs temperature.

The results of iteration 2 show the effects of letting the cooling system run uncontrolled and the removal of daylight savings time for the DOE-2 data. Weekend and evening use are still higher for the metered data than for the DOE simulation results. The weekend use is the result of lighting, plug loads, HVAC equipment or some combination of the above. Finally, the DOE-2 total load (DOE-TL) and DOE-2 air conditioning (DOE-AC) EnergyPrints^{TM} shown in Figure 3-6 indicate:

1. During the peak summer months the metered total load data shows a gradual reduction of demand through the evening hours, not the well defined break point typical of a tightly controlled HVAC system. The DOE-2 results emulate this phenomenon. The weather data for these runs represents an arid desert climate with high daytime temperatures and cool night temperatures, explaining why the cooling load is very small even though the cooling system is running wild.

2. The DOE-2 results underestimate the energy use for several reasons. First, the variation in actual occupancy during weekends, and unpredicted building operations that cause the lights or equipment to be left on over unoccupied periods. It is difficult to model these occurrences because they are random and unintentional. The effects of such scheduling items as spikes caused by janitors turning on lights for cleaning, intermittent weekend or evening occupancy and HVAC systems that are not tightly controlled must be somehow accounted for in a calibrated model. These items are not typically discovered during traditional on-site audits and interviews. Additional probing must be attempted to uncover such occupancy factors.

Assessing DOE-2.1 Model Accuracy. The DOE-2.1 engineering model calibration approach is iterative in nature. The cost effective development of hourly end-use data requires engineering and statistical judgment regarding the calibration process. For each iteration, the modeler must make decisions in order to keep the time and budget of each simulation model within reason. The combined use of engineering judgment and statistical measures can help guide that effort.

Traditionally, the measure of model accuracy was either building energy consumption or peak kW demand. In either case, the objective was to get within "X" percent of billing data, or sometimes hourly load data for the peak day. In practice, models within +/-10 to 20 percent of annual kWh using billing data have been perceived as "good enough." Additional metrics can be used to evaluate the accuracy of each building model, assessing the over/under prediction problems discussed earlier. Two additional target statistics can be used to assess the improvement in calibration and model accuracy for each successive model. As suggested by Krieder and Haberl, when comparing hourly data, the Coefficient of Variation (CV) and Mean Bias Error (MBE) provide reasonable measures of accuracy for comparing modeled and metered loads[6].

The CV is used to assess the goodness of fit of the current model, as well as the improvement in fit over the previous iteration. The CV is defined as:

$$\mathbf{CV = SD_{residual} / MM}$$

where:

$SD_{residual}$ = the standard deviation of the hourly residuals of the metered and model estimates of load.

MM = mean of hourly metered load

The MBE statistic is used to assess the direction of the bias in the estimate, and can be used by the modeler to focus model adjustment and calibration efforts. The MBE is defined as:

$$\mathbf{MBE = MR / MM}$$

where:

MR = mean of hourly residual load

The following measure statistics can then become the standards for "goodness of fit" between the modeled results and the total load data. With each iteration of the calibration, the modeling team assesses the marginal improvement in the measures. The measure statistics include:

- <u>Annual kWh</u>. Percentage differences.
- <u>Monthly kWh</u>. Percentage differences.
- <u>Monthly Max kW</u>. Percentage differences.
- <u>Total Load Residuals</u>. 8760 hourly total load model data is compared to the actual metered total load, for hours where total load data is available for each site. This assessment of this measure will include a qualitative visual inspection of the residual EnergyPrint, looking for a minimum in variation, and a quantitative measure of the Coefficient of Variation (CV) for the predicted and metered loads.
- <u>Total Load Vs Temperature Fit</u>. The relationship between 8760 hourly total load data and outdoor drybulb temperatures in the model data is compared to the metered load and temperatures for hours where total load data is available for each site.
- <u>Average Daytype Fits</u>. The relationship between summer and winter average or peak day hourly

[6] Kreider, J. and Haberl, J, page 79.

load data will be compared to the actual metered data, where metered data is available for each site.

- Visual Inspection of All Channels. The total load and end-use load shapes developed by the DOE-2 model can be visually compared to the metered total load and end-use data. Using building specific knowledge of end-uses, schedules, and occupancy, the EnergyPrints will be reviewed as assessed for reasonableness. The overall review will represent the final check in each iteration of the calibration process.

Table 3-1 suggests possible criteria for each of the discussed above. These measures can be used as guidelines for each building, recognizing that some models will exceed the guidelines, and may not meet the criteria within a reasonable amount of modeling time. These measures are suggested levels. Project funding and actual experience using the calibration may result in the adoption of other levels as each new project progresses.

CONCLUSIONS

The paper presents a set of DOE-2.1 model calibration tools that employ new visual data analysis techniques and statistical measures of fit. These diagnostic tools can be used to rapidly improve the quality of model predictions of building performance. The techniques can be used control or decrease the costs of building modeling, and to establish when "good is good enough" in the iterative modeling process using two standard statistical techniques, the CV and MBE measures. The paper provides an commercial office modeling example, where the model was rapidly improved by using the calibration tools to assess the accuracy of the model after the first iteration, make improvements, and assess the accuracy after the second iteration. These diagnostic tools provide a large step forward in the calibration of the models, moving from using monthly and annual data, or even peak day kW demand data to a method where the goodness of fit for the full 8760 hour period is assessed after each model iteration.

REFERENCES

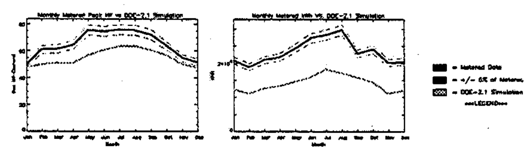
EPRI, Engineering Methods for Estimating the Impacts of Demand-Side Management Programs, Volume 1, Electric Power Research Institute, Palo Alto, California, 1992.

Kreider, J. and Haberl, J., June 1994, "Predicting Hourly Building Energy Usage", ASHRAE Journal, pp. 72-81.

Peterson, Hassl, et.al. 1994, "A Commissioning Cost Effectiveness Case Study," Commissioning, Operation, and Maintenance Proceedings, Volume 5, ACEEE Summer Study on Energy Efficiency in Buildings, pp 5.201-208.

Richards, D.. Wright, R. and Puckett, C., July 11, 1990, "End-Use Load Information in the Commercial Sector", End-Use Load Information and Its Role in DSM, p. 21, The Fleming Group, Syracuse, 1990.

TABLES AND FIGURES



Metered Total Load Data VS. DOE-2.1 Simulated Data – Monthly Comparisons

Monthly Max kW Comparison for

Month	Metered kW	DOE-2.1 kW	% Difference
Jan	50.3424	47.4210	5.80306
Feb	60.9408	49.7700	18.3306
Mar	61.2884	51.0260	16.7516
Apr	64.2818	50.9420	20.7518
May	75.6712	56.7250	24.9383
Jun	74.5344	60.4890	18.8442
Jul	78.6320	63.1950	18.9837
Aug	76.3408	63.8350	16.2717
Sep	72.1162	60.6030	15.9836
Oct	63.6904	56.8120	10.9740
Nov	54.7200	50.1670	8.28672
Dec	51.8098	47.3370	8.27968

Monthly kWh Comparison for

Month	Metered kWh	DOE-2.1 kWh	% Difference
Jan	20751.1	12280.0	40.8196
Feb	18885.4	11071.9	41.3732
Mar	20931.9	12528.1	40.1580
Apr	21737.3	13240.9	39.0887
May	24394.8	14521.8	40.4722
Jun	27342.9	15949.9	41.6672
Jul	28377.0	18277.8	35.5894
Aug	29832.4	17259.2	42.1460
Sep	22810.9	15928.7	30.1950
Oct	24082.5	14387.4	40.2585
Nov	20138.7	11080.6	46.0782
Dec	20483.2	11992.4	41.3963
Annual	279723.	168477.	39.7701

Figure 3-1: Monthly Peak kW Demand and kWh Energy Comparisons - Iteration 1.

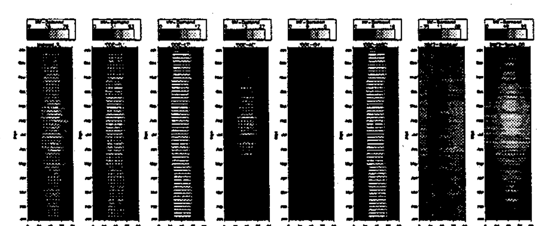


Figure 3-2: Total Load and End-Use Profiles - Iteration 1.

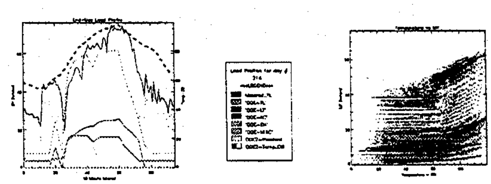


Figure 3-3: Hourly Load Profiles - Iteration 1.

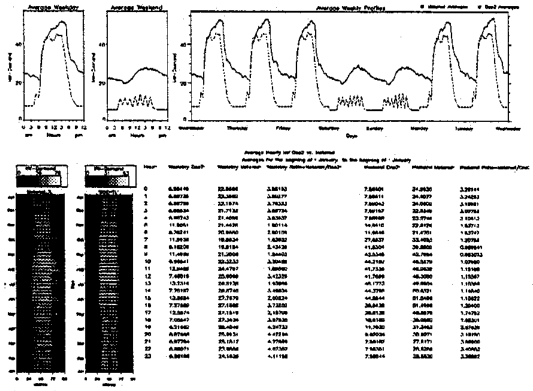


Figure 3-4: Average Day Hourly Profiles - Iteration 1.

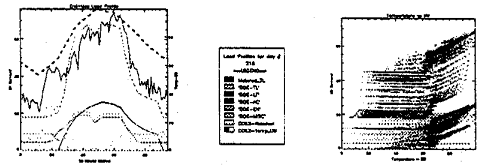


Figure 3-7: Hourly Load Profiles - Iteration 2.

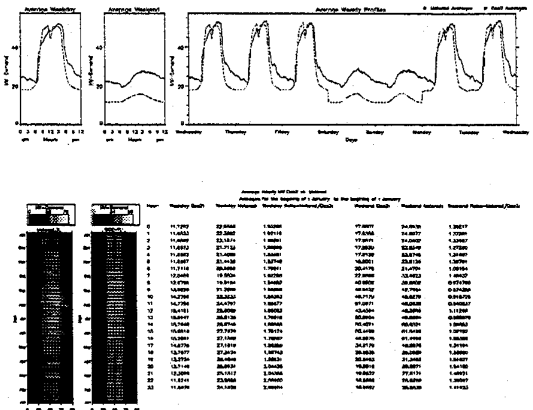


Figure 3-8: Average Day Hourly Profiles - Iteration 2.

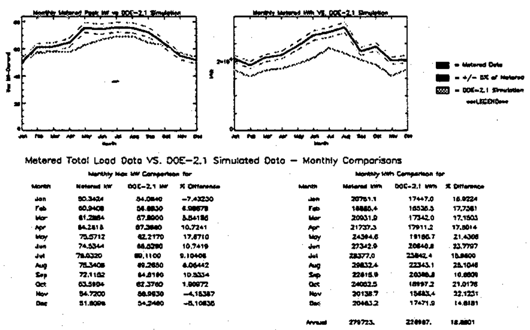


Figure 3-5: Monthly Peak kW Demand and kWh Energy Comparisons - Iteration 2.

Measure	Sample Models
Annual kWh	+/- 10%
Monthly kWh	+/- 10%
Monthly Max kW	+/- 10%
Total Load Residuals	CV=20%
Total Load Vs Temp Fit	OK
Average Daytype Profiles	OK
Visual Inspection	OK

Table 3-1: Suggested Measures for Calibrated Simulation Fits

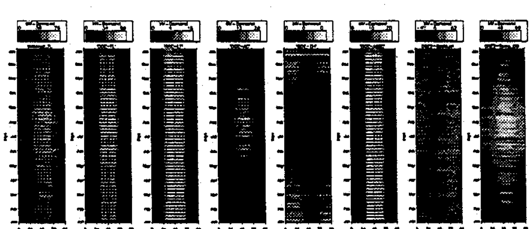


Figure 3-6: Total Load and End-Use Profiles - Iteration 2.