Quality Assurance Guidelines For
Statistical and Engineering Models

Final Report

by
Pacific Consulting Services
1320 Solano Avenue, Suite 203
Albany, CA 94706
510/526-3123

Contributing Authors
Richard Ridge, Pacific Consulting Services
Dan Violette, Xenergy
Don Dohrman, ADM

prepared for
The California Demand Side Management Advisory Committee:
The Subcommittee on Modeling Standards for End Use Consumption
and Load Impact Models

December 1994

TABLE OF CONTENTS

SECTION PAGE

1. INTRODUCTION

The California Public Utilities Commission (CPUC) recently adopted the *Protocols and Procedures for the Verification of Costs, Benefits, and Shareholder Earnings from Demand-Side Management Programs (Protocols)* for the measurement and evaluation (M&E) of DSM programs. These guidelines focus on the critical elements of M&E such as load impact estimation models, sampling, and metering and are specific to various combinations of customer sectors, program types, and end uses. These standards are understood to be minimal and are in many cases quite general. For example, the Protocols state that the load impact models for commercial retrofit programs may be some variant of allowable CDA model types[1], or a calibrated engineering model, both possibly supplemented by an engineering simulation model. In addition, both participants and non participants must be examined to estimate net program load impacts, and the sample sizes must be at least 450 for each group. However, the Protocols are for the most part silent regarding such detailed methodological issues as the actual specification of CDA models, testing of statistical assumptions underlying CDA models, and power analysis. CE models and engineering models also lack any methodological guidance. Thus, simply adhering to these minimal standards contained in the Protocols is no guarantee that an analyst is doing a professionally respectable job.

While one could simply ask analysts to guarantee that they adhered to the methodological guidelines contained in standard textbooks, this may not be sufficiently reassuring either to utility or regulatory staff. Thus, rather than simply trust analysts to follow the guidance contained in the basic methodological textbooks, our preference has been to develop what is called the Quality Assurance Guidelines (QAG) that requires analysts to indicate specifically how they addressed basic methodological issues. This approach is clearly consistent with the white paper prepared by the ADSMP Subcommittee on Evaluation Standards and Guidelines and the *Program Evaluation Standards* prepared in 1994 by the Joint Committee on Standards for Educational Evaluation, in that it is not very prescriptive. That is, the Subcommittee members have thus far prepared *practice and reporting standards* rather than highly prescriptive *methodological standards*. Their

[1] For a more detailed definition of the various model types currently under discussion, please see "An Evaluation of Statistical and Engineering Models for Estimating Gross Energy Impacts" by Ridge et al., 1994.

preference has been to require analysts to describe how they addressed certain key issues rather than to require analysts to address these issues in a specific way. For example, while there are many varieties of regression-based analyses, there are very basic methodological issues that often come up, such as collinearity, and that must be addressed if one is to do a professionally respectable analysis. The guidelines only require analysts to test for collinearity but does not tell them how to test for it, and, if present, does not prescribe the appropriate remedy. This is the sort of guidance that occupies a position somewhere between the minimal standards represented by the Protocols and the highly detailed guidelines contained in basic methodological texts. The QAG also asks *where* certain information such as sample dispositions can be found in the report.

It follows that the QAG must focus on those methodological issues on which there is general agreement regarding their importance within the social science and engineering communities. The QAG will also refer analysts to texts in which more detailed guidance can be found regarding all the issues addressed. Adherence to such guidelines still allows the final models to be shaped by the interaction of the situation, the data and the analyst. It is this very interaction and the resulting plethora of legitimate methodological choices that prohibited the creation of a more detailed and prescriptive QAG.

While the QAG addresses many of the key issues surrounding the estimation of both gross and net impacts, it does not address the use of discrete choice models for estimating net impacts. The Base Efficiency and Net-to-Gross Estimation Subcommittee is currently comparing a variety approaches for estimating net impacts including traditional analysis of covariance using an inverse Mills ratio and discrete choice analysis including nested logit. A report is due by the first quarter of 1995. Based on recommendations in this report, an addendum will be prepared to the QAG addressing these issues.

The QAG can be used in several ways. First, they could be included as a part of every M&E request for proposals (RFP) so that prospective bidders will know that they will be held accountable for conducting a sound analysis. Second, utility project managers and regulators reviewing an evaluation report containing a completed QAG can quickly assess whether the analyst at least addressed the most basic methodological issues. This latter point is especially important since neither utilities nor regulators have the time or personnel to carefully scrutinize every written evaluation report let alone attempt to replicate the results of all these studies. Of course, the details of how they addressed these issues should be contained either in the very detailed documentation that would be contained in the technical appendix of any evaluation report or in the work papers.

Finally, they can be used to create a common language to facilitate communication among utilities, regulators and consultants.

Analysts should not be expected to provide information on every model they estimated during the analysis. One can get much of this detailed information from the analysis logs that every competent analyst should keep or from the computer output itself. Rather, the purpose of the QAG is to characterize what was typically done for the final models within each model type.

Included in this report are drafts of the QAGs for statistical and engineering models. There are several features of these QAGs that merit discussion. First, the issues addressed are issues that a variety of basic social science and engineering methodological texts also address. That is, there appears to be a consensus that these issues are important. Second, because the QAG is supposed to save time, it should not simply be an exact replication of what is in the report itself. On the other hand, for the same reason, it should not simply refer to the appropriate part(s) of the report. The answers, while brief, should provide enough information to reassure reviewers that a given methodological issue was recognized and dealt with in a professionally responsible manner. Of course, only a pretest can determine whether this format will work. Finally, because some respondents may not be familiar some of the issues addressed or the terms used, references have been provided that should provide reasonably clear explanations.

2. QUALITY ASSURANCE GUIDELINES FOR STATISTICAL MODELS

QUALITY ASSURANCE GUIDELINES FOR CONDITIONAL DEMAND ANALYSIS (CDA) MODELS[2]

The QAG for CDA models is presented on the following pages. It is designed to cover the estimation of both net and gross impacts. With respect to net impacts, the issues addressed are well within the traditional research design framework involving the comparison of kWh consumption of participants and nonparticipants while attempting to control statistically for any compositional differences. Throughout the QAG, the observations participating in a regression model, or a sample, or in any other analysis framework are referred to simply as subjects, whether they are customers, accounts, or buildings and whether they are participant or comparison group members. Thus, how the questions are answered will depend on the type of study conducted.

This QAG should be completed for every CDA model type used in a given M&E study. However, a utility is not required to complete this form for every model attempted throughout the entire study within a given model type. One can get such detailed information from the analysis logs that every competent analyst should keep or from the computer output itself. Rather, in most cases, the purpose of the QAG is to characterize what was done for the final model(s) within a given model type. You should answer *each* of the questions briefly on separate pieces of paper.[3] Please keep your answers *brief*. You may refer the reviewer to specific sections of the evaluation report itself for more detail or perhaps for a complete and coherent answer to the question. Having said this, the reviewer should not have to piece together the answer from more than one section of the report. In other words, if the answer is in more than one section of the report, you must attempt to integrate the information from the report and provide the answer in a brief

[2] The definition of CDA is a collection of regression-based approaches that specifiy energy consumption as conditional on any number of measured variables, but not a complete inventory of equipment or other demand sources. All of the regression-based approaches described in "An Evaluation of Statistical and Engineering Models for Estimating Gross Energy Impacts" by Ridge et al., 1994. Other model types such as the statistical comparison method (SCM) and the calibrated engineering method (CEM) were considered sufficiently different from CDA models as to warrant their own guidelines.

[3] Each utility will provide a diskette containing all of the questions listed in the QAG. One can use this diskette to record all responses.

response. Remember, your summary should be much shorter than the discussion contained in the full report.

Note that some of these questions may not be relevant for a given study, thus making "not applicable (NA)" a legitimate response. For example, if you conducted a cross-sectional analysis, you should check "NA" for those questions relating to serial correlation.

Finally, you may not be familiar with certain terms or concepts contained in the QAG. To assist you in completing the QAG, numbers are placed next to some of the section headings and/or questions that refer analysts to one or more methodological references in which the particular issue raised in the section or question is addressed. When appropriate, page numbers are provided. Of course, there are other references that could be used but the ones listed were considered adequate to describe the basic issues and their relevance as well as to provide methodological guidance in handling any related problems that may arise.

Quality Assurance Guidelines:

Conditional Demand Model Types

Date ________________________

Utility Program _____________________________________

Utility Project Manager_______________________________

Lead Analyst _______________________________________

Employer ____

CPUC Study Identification Number __________

Sector(s) ____

Please indicate the sectors, programs, end uses, and measures for which estimates of gross and/or net impacts are provided. If impacts were estimated for other combinations of variables (e.g., weather zone or building type) please specify.

Are any of the impacts, adjusted for spillover? If yes, please describe.

____

Period of Time Covered by the Analysis ___________________________________

Applicable Table(s) From M&E Protocols ________________________________

Frequency of data (e.g., hourly, daily, monthly) ____________________________

A. MODEL TYPES

1. Please check the model types used [1]

a. Classic-Conditional Demand Analysis (C-CDA) using cross-sectional data and dummy variables to capture the impact of the program or installations _____

b. C-CDA using cross-sectional data and incorporating prior engineering estimates of impacts _____

c. C-CDA using cross-sectional time series (CSTS) data and dummy variables to capture the impact of the program or installations _____

d. C-CDA CSTS data and incorporating prior engineering estimates of impacts _____

e. Conditional Demand Analysis (CDA) using CSTS data and dummy variables to capture the impact of the program or installations _____

f. CDA using CSTS data and incorporating prior engineering estimates of impacts _____

g. CDA with pre/post design and dummy variables to capture the impact of the program or installations _____

h. CDA with pre/post design and incorporating prior engineering estimates of impacts _____

i. Other types of regression models used (please describe):

____

____

____

____

B. MODELS

1. Please indicate where the forms of all the final models that were used can be found. Also, the forms of all the competing models that were used in the final stages of the analysis but were not selected as the final models are of interest. Please indicate where these can be found.

C. SAMPLE

1. Did you attempt to estimate models using the population of subjects or a sample? If a sample was used, describe the sample design. For example, what were stratification variables, if any, was the sample random, was the sample proportional, and how were the weights calculated?
2. What was the size of the outbound sample? For example, how many questionnaires were initially mailed out, telephone contacts attempted, on-sites attempted?
3. What was the size of the achieved sample? For example, how many completed questionnaires were returned, telephone interviews completed, on-sites completed?
4. What were the response rates for each of the major data collection efforts? For example, the response rate to a mail survey might be 50%, while for a telephone survey it might be 65%, and for on-site surveys it might be 85%.
5. Please indicate where more detail can be found on the sample dispositions for all major data collection efforts such as telephone interviews or mail surveys, on-site surveys, and billing data extractions. A sample disposition is simply a description of what happened to each effort to collect data (e.g., no telephone number, language barrier, refused, completed, etc., missing data in program tracking, billing or weather databases).
6. Describe any efforts to estimate the extent of non-response bias. For example, in order to measure any bias, did you compare the kWh consumption or other customer characteristics for respondents versus non respondents?
7. Describe your efforts to correct for non-response bias. For example, were respondents weighted in any way to correct for any bias?
8. Were procedures used to determine the size of the samples in order to achieve to specific levels of precision at given levels of confidence? If yes, what assumptions, i.e., expected variance or error ratio if model based sampling is used, or effect size in traditional power analysis, were used?

9. Describe *key* characteristics of subjects that you used in final models. For example, were they all installers of efficient equipment or were they simply exposed to some treatment such as an audit? If residential were they single family or multi-family? What was the average income? What was the distribution across weather zones? If nonresidential, what building types and weather zones were represented?

D. DATA

1. Describe the data that were collected to support the analysis
2. Describe the source(s) and method(s) of collecting these data.
3. Indicate where the description can be found of how these data were manipulated in order to create the analysis datasets. Also, describe what screens were used to eliminate customers from the analysis and how many customers were eliminate as the result of each screen.
4. Where can all data collection instruments be found?
5. Where in the report can a flowchart be found illustrating the direct and indirect relationships of the data collected to each other and to the final estimates of impacts. If one is not available, please provide one.

E. SPECIFICATION AND ERROR

Misspecification

1. What were the initial specifications of the models and their rationale? If the specifications of these final models are different than these initial specifications, please explain what prompted the change. For example, were changes prompted by too much missing data for key variables, or the emergence of logical or theoretical inconsistencies?
2. Explain what you did to address the problem of misspecification. Describe the diagnostics carried out, the solutions attempted and their effects. If left untreated, please explain why.

Measurement Error

3. Were there substantial errors in measuring important independent variables? If so, what was done to minimize this problem. For example, was a weighted regression approach or an instrumental variables approach used?

Autocorrelation

4. If time series models were estimated, was autocorrelation a problem. If left uncorrected, biased estimates of standard errors may result. Under certain conditions, biased estimates of program impacts may also result. Please explain what you did to identify the problem in both the initial, intermediate, and final stages of the analysis and what you did to mitigate its effects. Describe the diagnostics carried out, the solutions attempted and their effects. If left untreated, please explain why.

5. What was done to ensure the stability of the solution to serial correlation during the final estimation stages.

6. Were any checks done to determine if the pattern of autocorrelation differs by customer or building type and thus require a different type of treatment. For example, schools may have a different pattern than large office buildings. If differences were found among different sub groups, was autocorrelation treated differently for each of these groups?

7. Did the solution for autocorrelation negatively affect the solution for heteroskedasticity ? If so, what was done?

Heteroskedasticity

8. If heteroskedasticity was a problem, please explain what you did to mitigate its effects. Describe the diagnostics carried out, the solutions attempted and their effects. If left untreated, please explain why.

9. Did the solution for heteroskedasticity negatively affect the solution for autocorrelation? If so, what was done?

10. If the solution to heteroskedasticity involved reweighting of data, how did this weighting process interact with the relative weights or expansion weights developed on the basis of the sampling plan and nonresponse problems?

F. COLLINEARITY

1. Explain what you did to address the problem of collinearity as it may have surfaced in the initial, intermediate, and final stages of your analysis. Describe the diagnostics carried out, the solutions attempted and their effects.

2. What level of collinearity did you find acceptable and why? For example, some collinearity among regressors may be acceptable when the regressors are theoretically required, or when the regressors are necessary to represent a polytomy.

G. TESTS FOR EXOGENEITY

1. Tests to determine the exogeneity/endogeneity of variables are not routinely done, but, depending on the situation, they can be useful. For example, if any bias were suspected due to self-selection, such tests, described by Kennedy (1992) might be called for.

H. INFLUENTIAL DATA

1. Describe any influential-data diagnostics that were performed in order to identify outliers?
2. If outliers were identified, how were they identified, how many were there, and how were they handled?

J. MISSING DATA

1. Describe how missing data were handled. For example, were cases with missing data dropped? Was mean substitution used to address the missing data problem or were other more acceptable techniques used?

K. TRIANGULATION

1. If more than one estimate of impact is provided, how have the results been combined to form a single estimate?

L. WEATHER

1. Describe how weather normalization was handled. For example, were the kWh values weather-normalized prior to initiating the analysis or were models first estimated using the original kWh data and recorded temperature and later evaluated using long-run temperature data? In either case what was the source of the long-run weather?
2. Did the normalization adjust for heating degree-days only, cooling degree-days only, or both?

3. What degree-day base was used for heating and for cooling? If the base was customer-specific, how was the base selected?
4. Are there potential seasonal biases related to the pre- and post- period dates? For example, if one or more cooling seasons exists in the pre period while none exists in the post period, the savings estimates may be overestimated.
5. On a customer-specific basis, how was the choice made between a heating-only, cooling-only, or heating-cooling normalization model?

M. ENGINEERING PRIORS

1. If prior engineering estimates of usage or savings were used in the models, what was the source(s) of the priors?

N. Precision

1. Where are the methods for calculation of key savings parameters and their standard error reported? For example, using standard statistical software, standard errors are always available on key parameters in a regression model while standard errors for other parameters like net-to-gross ratios are often calculated during some post-processing of regression results.

O. Comparison Group

1. If a comparison group[4] *was not* used to help estimate gross savings, describe what was done to control for the effects of background variables such as economic and political activity that may account for any increase or decrease in consumption in addition to the DSM program itself.
2. If you used a comparison group to estimate either gross or net impacts, describe what was done to control for any compositional differences and any suspected self-selection bias.

[4] The M&E Protocols do not require a comparison group for estimating gross savings.

QUALITY ASSURANCE GUIDELINES FOR CALIBRATED ENGINEERING METHODS (CEM)

DEFINITION

Calibrated Engineering Methods use initial engineering estimates of impacts combined with a “statistical verification” step. This verification step produces an estimated realization rate. The application of these methods involves drawing a sample of program participants; then, an in-field metering or enhanced/in-depth engineering analysis based on other measures of customer consumption is conducted at each participant site. These analyses essentially “verify” or serve as “audited” values of the initial engineering estimates. A ratio is calculated between the audited values and initial engineering estimates. For example, if the audited values are, on average, 75% of the initial engineering estimates, then the ratio is .75. If the sample of customers is drawn randomly, then the best estimate of what the evaluator (or "auditor") would have found if the analysis could have been conducted on the entire population is .75 times the sum of the initial estimates for the population. One strength of this method is that as long as the sampling is random, it is a relatively robust estimator. The types of assumptions required in the development of a regression model are not required by this method.[5]

A. PROGRAM RELATED AND MEASURE RELATED QUESTIONS

1. What energy efficiency measures are included in this program?
2. Were realization rates calculated for each measure? If no, what packages or combinations of measures were addressed?
3. What period of time is represented by the estimated impacts, i.e., what program interval is being estimated by this analysis?

B. SAMPLE AND SAMPLING (to be completed for each estimated realization rate)

1. What sample size was used to calculate the verified ratio?

[5] It is important to note that CEM is different in concept from the calibration of engineering based energy simulation models. These engineering models provide estimates of levels of energy use and therefore the models are calibrated to observed data on usage levels (e.g., billing data or load research data). Impacts are then calculated by two runs of the engineering model—a baseline model run and a model run incorporating the energy efficiency measures.

2. Explain what procedures were used to help ensure that a random sample was drawn.

3. Were any tests or comparisons made to examine whether the drawn sample was "representative" of the population of participants? If yes, please explain.

4. Were any adjustments to the sampling plan made to make the sample more "representative?" If yes, please explain.

5. Was a stratified sampling procedure used? If yes, please describe briefly, with rationale for stratification choices.

6. Were procedures used to size the samples to target specific precision levels at a given level of confidence? If yes, what assumptions, i.e., expected variance or error ratio in the case of model based sampling, were used?

C. VERIFICATION/MEASUREMENT METHOD

1. What procedures were used to verify the kWh and kW impacts for specific sites?

   a) pre/post end-use interval metering? If yes, what was the duration?

   b) spot watt metering pre and interval metering post? If yes, what was the duration?

   c) pre/post spot watt metering with post run-time metering? If yes, what was the duration?

   d) engineering analyses? If yes, please explain.

2. Were weather related/seasonal effects estimated? If yes, please explain how.

3. Were interaction effects addressed, for example the heating penalty associated with lighting efficiency improvements? If yes, please explain.

4. Were changes made to the baseline energy use from which the impacts are estimated, e.g., were changes made to account for burned out lights, expansions of space (adding on a new wing), changes in use of space, etc.? If yes, please explain.

5. While on-site, were issues of snap back, free-ridership, spillover addressed with the customer? If yes, please explain.

D. REPRESENTATION OF RESULTS

1. Within the sample, was one realization rate estimated for the entire sample, or were realization rates allowed to vary by any factor (e.g., magnitude of savings, type of building, total energy consumption)? If yes, please explain.

2. After the sample was drawn, were the strata boundaries and associated case weights adjusted to reflect the most current information on the population of participants? If yes, please explain.

3. Was an analysis of influential data points conducted? If yes, please explain.

   [Note to reviewers: For example, the sample sizes are usually small and it is easy to exclude each observation and re-estimate the realization rates to determine whether a single observation greatly influences the realization rate estimate. Similarly, potential outliers, e.g., savings estimates more that three standard deviations above the mean, can be excluded and the realization rate re-estimated. The distribution of savings tends to be a skewed distribution often with a few sites having extremely large savings estimates. A realization rate is a straight line fitting process, the y/x ratio as the slope, and the effect of several points well outside the general range of observations could influence the estimate. Many interesting estimation issues are involved, e.g., do the sets of observations essentially come from different distributions; that is, are they generated from a different underlying process and therefore belong in a different analysis.]

4. If influential data points were identified, were they analyzed to see if they were unique cases, i.e., did not fall within the definition of the program being analyzed? If yes, please explain.

5. Were any drawn sample sites dropped from the analysis for any reason? If yes, please explain.

6. Did your analysis present the mean, median, standard deviation and an example precision level and confidence interval for each realization rate estimate?

QUALITY ASSURANCE GUIDELINES FOR STATISTICAL COMPARISON METHODS (SCM)

DEFINITION

These approaches, sometimes termed "simple" comparison approaches, involve pre/ post comparisons of energy use among program participants. This involves subtracting the annual consumption in the "post" period from the annual consumption in the "pre" period. However, before this subtraction is done, the energy consumption data must first be weather-normalized: the effects of atypical weather are removed to produce what is called normalized annual energy consumption (NAC). The simple equation for calculating gross impacts for participants is presented below.

$$\text{Savings} = \text{NAC}_{\text{Pre}} - \text{NAC}_{\text{Post}}$$

One of the more commonly used methods for weather normalization is PRISM, which, like many of these comparison methods, typically does not use any data other than consumption data and weather data.

A. PROGRAM EFFICIENCY MEASURES

1. What energy efficiency measures are included in this program?
2. What proportion of program participants had each energy efficiency measure?
3. Were savings estimated separately for different participant groups? If yes, how were the different groups defined (by measures, by timing of participation, geographically, by characteristics known from the customer information system, etc.)?
4. What was the timing of program participation for the estimated savings?

B. COMPARISON GROUP

With respect to the estimates of gross savings, the M&E Protocols do not require a comparison group. The rationale for its exclusion is provided on pages 2-4 to 2-6 of *An Evaluation of Statistical and Engineering Models for Estimating Gross Energy Savings*. Thus, the SCM as defined does not include a comparison group. However, some have recommended the use of a comparison group to help control for exogenous changes related to prices, political factors, technological changes, or any systematic bias in the weather normalization procedure. Of course, a comparison group is often required for estimating net savings. The questions below are relevant if one used a comparison group for estimating either gross or net savings.

1. Was a comparison group used in the analysis?
2. How was the comparison group defined?
3. How were pre- and post- periods defined for comparison group customers?
4. Were any tests or comparisons made of the similarity between the participants and comparison group in the "pre" period? (yes/no) If yes, describe.
5. Were any adjustments made for differences between participants and the comparison group? If yes, describe.

C. SAMPLE AND SAMPLING

1. How many participating (and how many comparison) customers were used to estimate the savings?
2. Were these customers selected from the total pool of participating (nonparticipating) customers:
   - randomly
   - by census
   - by other means (explain)?
3. What screens were used to eliminate customers from the analysis?
4. How many participants (comparison cases) were eliminated as a result of each screen?
5. Were any tests or comparisons made to examine whether the drawn sample was "representative" of the population of participants (comparison population)? (explain)
6. Was a stratified sample used? (yes/no) If yes, how were strata defined and how was the allocation to strata determined?
7. Was the sample weighted in the analysis? If yes, what was the basis for the weighting?

D. WEATHER NORMALIZATION

1. What weather normalization model was used?
2. What time period defined the "normal" weather?
3. What was the source of the weather data used for the analysis?
4. What pre- and post-participation dates were included in the analysis?
5. Are there potential seasonal biases related to the pre- and post- period dates?
6. Did the normalization adjust for heating degree-days only, cooling degree-days only, or both?
7. On a customer-specific basis, how was the choice made between a heating-only, cooling-only, or heating-cooling normalization model?
8. What degree-day base was used for heating and for cooling? If the base was customer-specific, how was the base selected?
9. What accuracy measures are reported for the normalization model fits?

E. DIAGNOSTICS AND ACCURACY

1. Were the normalized savings examined for outliers?
2. How were cases identified as outliers handled?
3. Were any comparisons or tests made of the sensitivity of the results to inclusion or exclusion of outliers? If yes, describe.
4. Is the standard error of the estimated savings reported?
5. Is a confidence interval for the estimate reported? If yes, at what confidence level?
6. Is there a discussion of potential biases in the analysis?

3. QUALITY ASSURANCE GUIDELINES FOR ENGINEERING MODELS[6]

DOE-2

DOE-2 is a state-of-the-art building energy analysis computer model that is particularly well-suited for analyzing and evaluating energy use in buildings with relatively sophisticated HVAC equipment (e.g., commercial buildings). A Quality Assurance Checklist (QAC) for simulation analysis with the DOE-2 energy analysis model is provided here. This QAC provides for scrutiny of the two main steps involved in developing a DOE-2 analysis run—input data preparation and simulation calibration.

A. INPUT DATA PREPARATION

The data to be used for developing the inputs for a DOE-2 analysis usually need to be collected on-site, although the extent of on-site data collection may be determined by the complexity of the building to be simulated. In some cases, it may be appropriate to use data from building plans, from program default values, or from other sources of data for input values rather than conducting a site inspection. The purpose of the QAC for input data preparation is to ensure that summary information is readily available regarding the source of the input data and the extent to which program default values are used for a simulation. If input values are used that are not based on data collected on-site, the items where this occurs should be noted and the source of the input values identified.

On the following checklists, the source of the input values for the major items of data needed for a DOE-2 analysis should be identified. (If an item is not applicable for a particular analysis, n/a should be entered in Other column to indicate item is not applicable.)

[6] While the focus in this section is on the two energy simulation models most commonly used in California, DOE2 and Micropas, the issues addressed and the questions asked are also generally applicable to other models. Until a later study is conducted that addresses any possible issues that are unique to models other than DOE2 and Micropas, analysts should rely on the guidelines for engineering models contained in this report.

Building Characteristics:

	Site Data	Bldg. Plan	Program Default	Other (Specify Source)
Building Types				
Building Location				
Site Orientation				
Overall Building Size and Configuration				
Number of Occupants				
Floor Plan Layout and Elevations with respect to Orientations				
Floor Area of Conditioned Spaces				
Floor Area of Indirectly Conditioned Spaces				
Floor Area of Non-Conditioned Spaces				
Exterior Wall and Roof Areas and Orientations				
Construction Materials of various Opaque Surfaces				
Construction Assembly layers of various opaque surfaces				
Interior Walls, Types and Areas (if applicable)				
Window Types, Areas, and Orientations				
Glass Types				
Exterior Shading Characteristics				
Interior Shading Characteristics				
Overhang and Side Fin Characteristics				
Interior Thermal Mass				

Equipment Characteristics:

	Site Data	Bldg. Plan	Program Default	Other (Specify Source)
Type, Quantity, Size, Fuel Type, and Efficiency of Primary Heating Equipment				
Type, Quantity, Size, Fuel Type, and Efficiency of Secondary Heating Equipment				
Type, Quantity, Size, Fuel Type, and Efficiency of Primary Cooling Equipment				
Type, Quantity, Size, Fuel Type, and Efficiency of Secondary Cooling Equipment				
Type, Quantity, Size, Fuel Type, and Efficiency of Water Heating Equipment				
Type and Quantity of Lighting Fixtures, Lamps, and Ballasts				
Type, Quantity, Size, and Efficiency of Refrigeration Equipment				
Type, Quantity, Size, and Fuel Type of Kitchen Equipment				
Type, Quantity, Size, Location(s), and Efficiency of Motors related to process operation as well as HVAC applications				
Miscellaneous Equipment, including computers, office equipment, etc.				

Operating Schedules:

	Site Data	Bldg. Plan	Program Default	Other (Specify Source)
Annual Business				
Occupancy				
Lighting				
Receptacles				
Miscellaneous Equipment				
Cooling Equipment				
Heating Equipment				
Cooking Equipment				
Refrigeration Equipment				
Fans				
Thermostat Heating and Cooling Set Points				

Before simulations are performed, input values that have major effects on the results of a DOE-2 simulation analysis should be further checked and verified for accuracy. At a minimum the following items should be verified for any simulation run. (The questions are organized according to the major sections of DOE-2.) Values of any items for which the question is answered "No" should of course be corrected.

Loads Section of DOE-2
How was the weather data for the simulation chosen? How well do the chosen weather data correspond to the geographic location and climate conditions of the building?
How was the orientation of the building defined?
How were the location and control strategies of thermostat and terminal boxes used in determining HVAC zones?
How does the definition of model thermal zones correspond to grouping of the HVAC zones?
What method was used to define conditioned zones, unconditioned zones, and indirectly conditioned zones?
Were there any interior walls defined to separate zones with different thermostat set points or dead bands?
How were the construction material and assembly layers defined?
How were the opaque surface areas, tilts, and orientations defined?
How were the glass input values and window orientations defined with respect to wall definition of the XYZ coordinates?
How were the interior and exterior shadings defined?
How were the shading schedules defined?
How was the window MULTIPLIER command applied, especially for the modeling of daylighting features?
How were the custom weighting factors used for construction assembly inputs?
How were occupancy heat dissipation and density and occupancy profiles in various day types and weeks specified?

How were interior lighting type(s), density, and operating profiles specified?
How was miscellaneous equipment identified? How were their density and operating profiles specified?
How were thermal zone temperatures specified?
How were assumptions about infiltration rates made?
How were the power and operating schedules for auxiliary equipment not located in the conditioned spaces specified?

System Section of DOE-2
How were the exhaust fans kW for each thermal zone specified?
How was the type of air distribution system defined?
How was the minimum CFM ratio for the VAV boxes specified?
How do the maximum and minimum diffusive temperatures of the supply air match against observed data?
How were the economizer type and operation limits specified and modeled?
How was the ventilation air flow rate specified?
How were the fan kW and operating schedules specified and modeled?
Was the fan power included in the cooling Electric Input Ratio?
Was the return air path a plenum?
How were the system type, efficiency, and capacity specified?
If performance curves are applied to the system, how was the capacity curve fit specified?
Did the assignment of all thermal zones to the system include plenums?

Plant Section of DOE-2
How were the water heating boiler, type, quantity, capacity, efficiency, fuel type, and operating schedule specified?
How were the space heating boiler, type, quantity, capacity, efficiency, fuel type, and stand-by losses specified?
How were the chiller type, quantity, capacity, efficiency, fuel type, chilled water supply temperature, and performance curves (when data are available) specified?
How were the cooling tower, quantity, fan control type, tower water set-point, tower motor efficiency, tower design wet-bulb and tower pump head specified?
How were the chilled and hot water supply design temperatures, heads, motor efficiencies, losses, and speed control types specified?

B. SIMULATION CALIBRATION AND DIAGNOSTICS

Generally, the actual simulations of a building with DOE-2 should be calibrated to ensure that the energy use estimates from the simulations are reconciled against actual data on the building's energy use. Quality assurance checking for this calibration involves the following:

- Specifying the criteria that are used to judge whether the model has been appropriately calibrated. The actual criteria may differ from analysis to analysis, but the criteria actually used should be indicated. An example of such criteria might be to have annual simulated energy use within ±10% of annual energy use as indicated in utility billing data. Other comparisons that can be used as calibration criteria include the following:
    - Whole-building energy intensities (EUIs) against actual data
    - End-use EUI comparison against actual data
    - End-use breakdown fractions
    - Whole-building hourly energy use profile against actual data
    - End-use hourly energy use profiles against actual data
- Indicating the input values that were changed to bring the simulation into calibration and giving the reason why a value was changed. It is well known that there is a wide variety of input values that can be changed to modify the results of a DOE-2 simulation analysis. Appropriate quality assurance requires keeping a record of the values changed for purposes of calibration and explaining the diagnostic checks that were applied to determine what input values should be changed. Examples of diagnostic questions for DOE-2 calibrations include the following:

	Yes	No
Were the engineering parameters in the DOE-2 output within the acceptable range (e.g., for supply CFM/square foot or tons/square foot)?		
Were the loads met? (Can be determined from checking PS-D report.)		
Was the outside air flow modeled correctly? (Can be checked with hourly report No. 39, which gives the ratio of outside to total cfm and the effects of any economizer).		
Were proper values input for the equipment efficiencies? (Can be checked against SV-A or PV-E reports).		
Were the inside temperatures in the thermal zones maintained within the acceptable design range? (Can be checked with DOE-2's scattered temperature plot in the SS-O report).		

- In some cases, it may not be possible to achieve calibration because of idiosyncrasies of the particular facility (e.g., discontinuous occupancy patterns). Such cases should be noted and the reason why calibration failed explained.

MICROPAS

MICROPAS is a building energy analysis model that has been certified by the California Energy Commission for showing compliance with the State of California's 1992 Energy Efficiency Standards for low-rise residential buildings. It has also been adapted for more general use as a tool for analyzing energy use in residential buildings. (Complete details on MICROPAS and its application are provided in MICROPAS4, v4.0 User's Manual, ENERCOMP, Inc., Sacramento, California.)

A Quality Assurance Checklist (QAC) for simulation analysis with the MICROPAS residential energy analysis model is provided here. This QAC provides for scrutiny of the two main steps involved in developing a MICROPAS analysis run—input data preparation and simulation calibration.

A. INPUT DATA PREPARATION

The data to be used for developing the inputs for a MICROPAS analysis usually need to be collected on-site, although the extent of on-site data collection may be determined by the complexity of the building to be simulated. In some cases, it may be appropriate to use data from building plans, from program default values, or from other sources of data for input values rather than conducting a site inspection. The purpose of the QAC for input data preparation is to ensure that summary information is readily available regarding the source of the input data and the extent to which program default values are used for a simulation. If input values are used that are not based on data collected on-site, the items where this occurs should be noted and the source of the input values identified.

On the following checklists, the source of the input values for the major items of data needed for a MICROPAS analysis should be identified. (If an item is not applicable for a particular analysis, n/a should be entered in Other column to indicate item is not applicable.)

Building Characteristics:

	Site Data	Bldg. Plan	Program Default	Other (Specify Source)
Building Types				
Building Location				
Site Orientation				
Overall Building Size and Configuration				
Number of Occupants				
Floor Plan Layout and Elevations with respect to Orientations				
Floor Area of Conditioned Spaces				
Floor Area of Indirectly Conditioned Spaces				
Floor Area of Non-Conditioned Spaces				
Exterior Wall and Roof Areas and Orientations				
Construction Materials of various Opaque Surfaces				
Interior Walls, Types and Areas (if applicable)				
Window Types, Areas, and Orientations				
Glass Types				
Exterior Shading Characteristics				
Interior Shading Characteristics				
Overhang and Side Fin Characteristics				
Interior Thermal Mass				

Equipment Characteristics:

	Site Data	Bldg. Plan	Program Default	Other (Specify Source)
Type, Quantity, Size, Fuel Type, and Efficiency of Heating Equipment				
Type, Quantity, Size, Fuel Type, and Efficiency of Cooling Equipment				
Type, Quantity, Size, Fuel Type, and Efficiency of Water Heating Equipment				
Type and Quantity of Lighting Fixtures				
Type, Quantity, Size, and Fuel Type of Kitchen Equipment				
Miscellaneous Equipment, including home office equipment, shop equipment, etc.				
Miscellaneous Equipment, including home office equipment, shop equipment, etc.				

Operating Schedules:

	Site Data	Bldg. Plan	Program Default	Other (Specify Source)
Occupancy				
Lighting				
Receptacles				
Miscellaneous Equipment				
Kitchen Equipment				
Cooling Equipment				
Heating Equipment				
Thermostat Heating and Cooling Thermostat Set Points				

Before simulations are performed, input values that have major effects on the results of a MICROPAS simulation analysis should be further checked and verified for accuracy. At a minimum the following items should be verified for any simulation run.

How was the building type specified?
How was the building orientation defined?
Is the number of stories correct?
How was the weather data for the simulation chosen? How well do the chosen weather data correspond to the geographic location and climate conditions of the building?
How were the conditioned zones, unconditioned zones, and indirectly conditioned zones defined?
How was the occupancy level determined?
How was the appliance heat gain determined?
How was the infiltration rate determined?
How was the floor area determined?
How were the opaque surface areas, tilts, and orientations defined?
How were the opaque surface characteristics defined?
How was the ratio of glazing area to floor area determined?
Are there any interior walls defined to separate zones?
How were the glass characteristics defined?
How were the window areas and orientations defined?
How were the interior and exterior shadings defined?
How were the overhangs defined?
How were the thermal mass characteristics and areas defined?
How were the space heating system type, quantity, efficiency, and fuel type specified?
How were the space cooling system type, quantity, efficiency, and fuel type specified?
How were the duct system and duct insulation levels specified?
How were the water heater type, quantity, capacity, energy factor, fuel type, and distribution system specified?
How were the schedules defined?
How was the thermostat type defined?

B. Simulation Calibration and Diagnostics

Generally, the actual simulations of a building with MICROPAS should be calibrated to ensure that the energy use estimates from the simulations are reconciled against actual data on the building's energy use. Quality assurance checking for this calibration involves the following:

- Specifying the criteria that are used to judge whether the model has been appropriately calibrated. An example of such criteria might be to have annual simulated energy use within ±10% of annual energy use as indicated in utility billing data. The peak demands (when available) and month-to-month usage profiles can also be compared. The actual criteria may differ from analysis to analysis, but the criteria actually used should be indicated.

- Indicating the input values that were changed to bring the simulation into calibration and giving the reason why a value was changed. It is well known that there is a wide variety of input values that can be changed to modify the results of a simulation analysis. Appropriate quality assurance requires keeping a record of the values changed for purposes of calibration and explaining the diagnostic checks that were applied to determine what input values should be changed.

- In some cases, it may not be possible to achieve calibration because of idiosyncrasies of the particular facility (e.g., discontinuous occupancy patterns). Such cases should be noted and the reason why calibration failed explained.

REFERENCES

ASHRAE, "Energy Estimating Methods." Chapter 28 in *The 1993 ASHRAE Handbook: Fundamentals*. Published by the American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Inc., 1993.

Berk, R. A. "A Primer on Robust Estimation." In *Modern Methods of Data Analysis*, edited by Fox, J. and Long, J. S., Newbury Park, CA: Sage Publications, 1990.

Belsey, D. A., E. Kuh, and R. E. Welsch. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley. 1980.

Campbell, Donald T. and Julian C. Stanley. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally College Publishing, 1963.

Cochran, William G. *Sampling Techniques*. New York: John Wiley & Sons, 1977.

Cohen, Jacob and Patricia Cohen. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. New York: John Wiley & Sons, 1975.

Cook, Thomas D. and Donald T Campbell. *Quasi-Experimentation: Design & Analysis Issues for Field Settings*. Boston, MA.: Houghton Mifflin Company, 1979.

Dubin, J., and D. Rivers. "Selection Bias in Linear Regression, Logit and Probit Models." Chapter 10, *Modern Methods of Data Analysis*. California: Sage Publications. 1990

Electric Power Research Institute (a). *Impact Evaluation of Demand-Side Management Programs: A Guide to Current Practice*. EPRI CU-7179, Volume 1. 1991.

Electric Power Research Institute (b). *Impact Evaluation of Demand-Side Management Programs: Case Studies and Applications*. EPRI CU-7179, Volume 2. 1991.

Electric Power Research Institute (c). *Impact Evaluation of Demand-Side Management Programs*, (Volumes 1 & 2), EPRI CU-7179s, 1991.

Johnston, J. *Econometric Methods*. New York: McGraw-Hill Book Company, 1984

Kerlinger, F. N., and E. J. Pedhazur. *Multiple Regression in Behavioral Research*. New York: Holt, Rinehart, and Winston. 1973.

Kennedy, Peter. *A Guide to Econometrics*. Cambridge, MA: The MIT Press, 1992

Kish, Leslie. *Survey Sampling*. New York: John Wiley & Sons, 1965.

Kraemer, Helena Chmura and Sue Thiemann. *How Many Subjects*: *Statistical Power Analysis in Research*. Newbury Park, CA: Sage Publications, 1987.

Lipsey, Mark W. *Design Sensitivity: Statistical Power for Experimental Research*. Newbury Park, CA: Sage Publications, 1990.

Madow, William G., Harold Nisselson, Ingram Olkin. *Incomplete Data in Sample Surveys*.. New York: Academic Press, 1983

Parti, C. and M. Parti. "The Total and Appliance-Specific Conditional Demand Analysis for Electricity in the Household Sector." Bell Journal of Economics, Spring 1980.

Pollard, W.E. *Bayesian Statistics for Evaluation Research*. California: Sage Publications. 1986.

Ridge, Richard, Kirtida Parikh, Dan Violette and Don Dohrman. "An Evaluation of Statistical and Engineering Models for Estimating Gross Energy Impacts." Sponsored by the CADMAC Modeling Subcommittee, 1994.

Rossi, Peter and Howard E. Freeman. *Evaluation: A Systematic Approach.* Newbury Park, California: SAGE Publications, 1989.

Sayrs, Lois. *Pooled Time Series Analysis*. Newbury Park, CA: SAGE Publications, 1989.

SCE(c), *A Review and Critique of Statistical Techniques for Estimating Net Impacts, Volume 8*, 1993

Violette, D., and M. T. Ozog. "Correction for Self-Selection Bias in the Estimation of Audit Program Impacts." In *Proceedings of the ACEEE 1990 Summer Study in Energy Efficiency in Buildings*, American Council for and Energy Efficient Economy, Vol. 6, 1990.

First footnote on page 1

The National Association of Energy Service Companies (NAESCO) standards for metering and monitoring have been adopted as interim protocols subject to (1) adjustments to baseline energy use to reflect equipment or building minimum standards, and (2) adjustment to gross savings consistent with the M&E Protocols. The utilities should employ the adjusted NAESCO standards when doing so is cost-effective relative to the employment of other measurement approaches provided for under the Protocols. The CADMAC will consider refinements and improvements to metering and monitoring standards as part of the statewide study efforts. Any proposals for modification must be presented before the CPUC for review and final adoption in an Annual Earnings Assessment Proceeding.

Second Footnote on page 4

While the focus in this volume is on the use of metering and monitoring data to improve initial engineering-based estimates of savings, it is possible that the initial estimates of savings can be enhanced through the use of more sophisticated energy simulation models or even engineering algorithms that have benefited from much improved input data.