An Evaluation of Statistical and Engineering Models For Estimating Gross Energy Impacts

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

June 1994

1. INTRODUCTION

Recently, the California Public Utilities Commission (CPUC) adopted the Protocols and Procedures for the Verification of Costs, Benefits, and Shareholder Earnings from Demand-Side Management (DSM) 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. In addition, the California Demand-Side Management Advisory Committee (CADMAC) was established with a set of objectives that includes the continuing development of the Protocols. The responsibility for dealing with specific issues such as those related to end-use consumption and load impact models or end-use metering was assigned to various subcommittees.

The evaluation studies submitted in late 1993 by California utilities as well as certain perceived deficiencies within the Protocols have motivated the CADMAC Subcommittee on Modeling Standards for End-Use Consumption and Load Impact Models (Modeling Subcommittee) to request that the following tasks be addressed:

1. Produce a list of statistical and engineering models that are now in use in the DSM evaluation field or that would reasonably be expected to occur.
2. Evaluate each of the models with respect to its requirements, outputs, strengths, and weaknesses.
3. For each of the models that are selected for inclusion in the Protocols, develop a quality assurance checklist that outlines the basic dos and don'ts.
4. Suggest, if possible, methodologically sound mechanisms for combining end-use metering information from small numbers of customers with statistical models applied over larger samples in billing analyses.
5. Investigate the cost-effectiveness of providing estimates of pre- and post-program unit energy consumption (UEC) for both participants and nonparticipants as required by Table 6 in the Protocols.
6. Investigate alternatives to providing the above UECs.

The current report presents the results for the first two tasks. Subsequent reports will present the results of the remaining tasks.

Because the motives for taking on these first two tasks are complex, they merit some discussion. Currently, the Protocols require either a conditional demand analysis (CDA) or calibrated engineering methods (CEM) for estimating gross load impacts for many of the utility-sponsored residential and nonresidential demand-side management (DSM) programs. There seems to be general agreement that, in addition to the CEM approach, engineering approaches such as DOE2 or Micropas should be explored as possible additions. However, the motives for exploring different statistical models are mixed depending on one's interpretation of CDA.

Some define CDA strictly as a very specific and complex regression-based approach that should include, among other independent variables, a complete inventory of all major energy-using equipment. Such a model can be used to estimate both pre- and post-intervention unit energy consumption (UECs), and by subtracting the post from the pre, an estimate of gross unit energy savings (UESs) can be produced for energy conservation measures (ECM).

There are at least two bases for this interpretation. The first is that for over a decade the letters "CDA" have been understood in this strict, classic, sense throughout the utility industry. The second basis is one of the two quantitative relationships for estimating both gross and net load impacts, depicted in Table 5, section B.2 of the Protocols. This model requires "base usage" and "usage in the impact year" for the "energy-using equipment" for both participant and comparison groups. In other words, it requires both pre- and post-intervention period UECs for both the participant and comparison groups.

Other impact evaluation specialists define CDA less restrictively as a collection of regression-based approaches that specify energy consumption as conditional on any number of measured variables, but not a complete inventory of equipment or other demand sources. Generally, these models can be used to estimate UESs directly, i.e., without first estimating base usage and impact year usage. Also, such variants on CDA are unlikely to decompose either consumption or savings by equipment type or end use. In such models, the intervention is identified either nominally or in interval form as an indicator that savings are to be "expected" at a particular time and foreword. The coefficients on the savings indicator variable(s) are subsequently used in calculating gross load impacts. These variants on CDA include a number of strategies for handling the variation consumption across customers, time, and end uses. One of the more well known variations is SDG&E's "Simplified CDA" or "S-CDA," which specifies that a number of energy-relevant parameters may take on different values across customers, in an effort to better isolate the unique impact of the intervention within a customer's billing trajectory. However, most DSM impact evaluations, and virtually all non-residential evaluations, fall into the general category of non-classic, less restrictive CDA.

There are three bases for this less restrictive interpretation. First, on page A-3 of the Protocols, statistically adjusted engineering models (SAE) are defined as a special type of CDA model in which gross unit energy savings can be directly estimated without first estimating both pre- and post-installation UECs. The second basis for this interpretation is that to our knowledge a classic CDA has never been used successfully in evaluating any commercial DSM program. This may have suggested to some that if they wanted to employ a classic CDA in evaluating residential DSM programs where it has been shown to work reasonably well, they could do so; but in the nonresidential sector, they could estimate non-classic CDA models. Finally, the third basis is that the pursuit of UEC (or UEC-like) estimates through classic CDA may not be a worthy objective if the demanding data collection requirements detract from the precision obtainable for savings estimates.

Given these different interpretations, the motives for the first two tasks are different. After having reviewed numerous recent evaluations of DSM programs sponsored by California utilities, those who define CDA strictly have come to suspect that perhaps the "classic" CDA is too limiting and that other equally sound techniques should be explored for possible inclusion in the Protocols. On the other hand, those who have a less strict definition of CDA assert that the less complex UES models are *already* permitted in the Protocols. What they are interested in is a systematic evaluation of the various modeling options, from the most complex to the least complex, so that they can make informed decisions when selecting a given option. Thus, whatever one's perspective on the CDA issue, accomplishing these first two tasks is very important.

Recognizing this fundamental difference in interpretation, for convenience, the authors of this report will adopt the less restrictive definition of CDA with the stricter version of CDA referred to as classic CDA (C-CDA). A useful nomenclature for the remaining model types will then be developed. It is left to the members of the CADMAC to resolve the issue of which interpretation of CDA is correct.

Because such a wide range of statistical and engineering models will be evaluated, each model type cannot receive a detailed and technical evaluation. Instead, this report offers a general evaluation guided by the attributes that are considered relevant to the Protocols. For a detailed and technical evaluation of the various models, please refer to the Bibliography.

The work on the first two tasks will not address a number of relatively detailed questions, involving, for example, weather normalization, load shape impacts, net savings, or what are in fact important variations within model types. The latter include, within the general framework of cross-sectional time series models, blocking of the variance in consumption between customers, constraining of certain coefficients to equality across customers, etc. Also included in this category are a variety of estimation issues, largely relegated to the forthcoming quality assurance document.

The remainder of this report is divided into two sections, the first dealing with statistical models and the second dealing with engineering models. Within each section, evaluation designs and attributes are first defined. Next, the types of models are described and then evaluated in terms of the attributes. Finally, because this report was not designed to be either comprehensive or highly detailed, a bibliography of standard references that address statistical and engineering modeling as well as evaluation design is provided.

2. STATISTICAL MODELS

In this section, the relevant evaluation designs used for capturing gross impacts will be briefly defined; these are followed by a discussion of the basic types of statistical models. Both must be discussed since the model types are usually defined by a combination of the evaluation designs and complementary statistical methods, e.g., a CDA using a pre/post evaluation design. Finally, the types of models are described and then evaluated in terms of their attributes.

Development of Model Types

The first task is to identify the model types in current use. In order to do this, a set of general characteristics for standard regression models was developed, which, if combined in different ways, form the basic model types. Thus, for standard regression models, there are at least three very general characteristics:

1. Number of billing periods per customer/participant, which discriminates among cross-sectional, pre/post, and time series analyses. This characteristic affects how responsive the design is to variation across customers and across time.
2. How the variables of interest are coded, which discriminates among models that use binary variables and those that incorporate engineering-based estimates of load impacts.
3. Number of variables in the model, which discriminates among fully-saturated models like C-CDA and less complex CDA models.

These various model characteristics were then combined to form eight regression-based model types and two other model types which do not directly involve regression analysis in the estimation of gross impacts. These ten model types are as follows with definition of key terms provided on the following pages:

1. C-CDA using cross-sectional design[1] with binary indicators of equipment installations (C-CDA cross-sectional with dummies)

---

1 A cross-sectional design is one that is based on a group of subjects or customers at a single point in time.

2. C-CDA using cross-sectional design with engineering priors (C-CDA cross-sectional with priors)[2]
3. C-CDA using cross-sectional time-series (CSTS) design with binary indicators (C-CDA CSTS with dummies)
4. C-CDA using cross-sectional time-series design with engineering priors (C-CDA CSTS with priors)
5. CDA using cross-sectional time-series design with binary indicators (CDA CSTS with dummies)
6. CDA using cross-sectional time-series design with engineering priors (CDA CSTS with priors)
7. CDA using pre/post design using binary indicators (CDA pre/post with dummies)
8. CDA using pre/post design with engineering priors (CDA pre/post with priors)
9. Statistical comparison methods (SCM)
10. Calibrated engineering methods (CEM)

Development of Model Attributes

The list of attributes by which one could evaluate these ten different model types is presented below:

1. UEC production
2. Input data requirements
3. Error
4. Cost
5. Robustness

Using these ten basic model types and five attributes, a fifty-cell matrix can be formed Later, as each model type is evaluated with respect to each of the five attributes, each cell of this matrix will be completed. However, in order to better understand the ten model types and the five attributes, the different evaluation designs, analytical methods, and attributes must first be defined.

2 Model types 1 and 2 are included since they are most consistent with the second quantitative relationship illustrated in Table 5, Section B.2, of the Protocols.

Definitions of Evaluation Designs

Before proceeding any further, consider the following definition of gross load impact on page A-5 of the Protocols:

> The change in energy consumption and/or demand that results directly from program-related actions taken by participants in the DSM program, regardless of why they participated.

This definition is clearly focused, appropriately enough, on the load impacts that can be directly attributed to the installation of the ECM. Of course, in order to estimate this impact, one must attempt to control for all the confounding factors that may affect energy consumption in addition to the installation of the ECM. This is the primary purpose of any evaluation design, i.e., to accurately attribute some observed outcome such as reduced energy consumption to some intervention such as the installation of an ECM.

The evaluation designs that are possible for a utility depend on both the data that are routinely collected and the frequency with which they are collected. There are two basic forms of data. Time series data are simply data that are collected repeatedly over time. Cross-sectional data are data collected from participants at only one point in time. The combinations of time series and cross-sectional data determine the portfolio of evaluation designs and data analysis approaches available to a utility.

Time Series Design. Utilities routinely collect energy consumption data on a monthly basis or in a form that can be easily and reliably be transformed to a monthly basis. In addition, weather data specific to each customer's geographic region are available on a monthly basis. The record of each customer also contains such information as their standard industrial classification code (SIC) and the price of electricity. Additional data, some of which change over time (e.g. installations of energy efficient equipment through the program and an engineering-based estimate of first-year annual savings), are often available for participants in DSM programs from the program tracking systems. At a single point in time, information on business operations, building(s), existing energy-using equipment, and recent installations of energy-using equipment outside the DSM program are often collected from both participants and nonparticipants.

These data permit the use of a very powerful quasi-experimental research design, the time series design. The essence of a time series design is the presence of a periodic measurement process on some group and the introduction of an experimental change into this time series of measurements. This design can control for all the confounding factors except for history. Such historical factors include climate as well as social, political and economic events that occurred between the pre- and post-installation periods unrelated to the installation of the equipment itself or other changes in the customer or building.

Together these factors have been referred to in the research methodology literature as "history" by Campbell and Stanley (1966)or simply "interfering events" by Rossi and Freeman (1989). However, these historical factors can at least be partially controlled. Examples of historical effects that can be controlled statistically are temperature, changes in customer square footage and operating hours, inflation, factory utilization, and interest rates (Campbell and Stanley, 1966). It should be noted that the argument that using cross-sectional time series data inflates the t-statistics is valid only if the problems of heteroscedasticity and autocorrelation exist but are ignored.

Pre/Post Design. Other utilities have taken the monthly time series data and collapsed it into annual data. For example, they create average monthly pre-installation value and average monthly post-installation value for such month-based variables as energy consumption, weather (cooling degree days), price, and inflation. The change in participant energy consumption from the pre to the post period after controlling for other confounding variables is the gross load impact.

Like the time series design, the pre/post design can control for some of the historical effects. However, to collapse a time series to pre and post averages for all the variables is throwing away valuable information, thus reducing the ability of the pre/post design to control for the effects of history. Second, when using an SAE model, if one incorporates an annual value for the engineering prior rather than allowing the value to vary by month or quarter, then one also throws away valuable information.

Within a pre/post design framework, some utilities have used a comparison group of non-installers to refine their estimates of gross load impacts. The wisdom of using a comparison group for this purpose for either the time series or pre/post evaluation designs is discussed below.

Comparison Groups. While not required by the Protocols, some utility evaluation experts have recommended that the basic time series and pre/post designs be enhanced through the use of a comparison group of customers. For example, assume that members of a participant group installed energy efficient central air conditioners (ACs) while members of a comparison group did not install central ACs, either standard or efficient. The argued purpose of the comparison group is to control for certain factors other than the installation or other customer-related variables in explaining any increases or decreases in participants' AC consumption.

With respect to the structure of the model within a time series design framework, in addition to the obvious inclusion of the installation variable, one should also include other variables that may influence consumption since one is interested in the impact of the installation after controlling for these other factors. Such variables as temperature, changes in the price of electricity faced by each customer, changes in square footage, changes in operating hours, installation of new equipment, replacement of old equipment, as well as other economic factors such as commercial interest rates, inflation rate, gross state product, population increases, and retail sales.

However, even after controlling for many of the historical factors, there is always the possibility that something else happened in the environment that affected consumption but could not be easily modeled. To attempt to control for any of these remaining historical effects, one could include a comparison group of customers who did not replace their ACs during the period being studied. Unfortunately, while it may be desirable to control more completely for the effects of history in estimating net impacts due to the installation of the efficient ACs, it may not always be practical because of methodological and budgetary barriers. Moreover, there are other less expensive and equally effective methods for controlling for these remaining historical effects.

First, consider the methodological barriers. Whenever one mixes two groups together that have different characteristics, these differences must be statistically controlled. If the observed differences between the two groups cannot be effectively controlled statistically, then the effect of history is imperfectly captured and the resulting net impact due to the installation of the equipment is biased in unknown ways. The significance of this problem in this context is no less than that encountered in attempting to estimate net energy reduction due to the program. The recent debates regarding the most effective ways to control self-selection bias have, while not producing a consensus (SCE(d), 1993; Violette, 1990) have served to convince many analysts of the complexity of the problem and the danger of underestimating its potential biasing effects.

Second, there are also the additional costs of collecting the necessary data from nonparticipant non-installers. Of primary interest are those variables that one would reasonably expect to affect energy consumption and that discriminate between participants and nonparticipants. However, given the current budgetary climate, the costs to further refine the net impact due to the installation of the equipment may be viewed as a methodological luxury. This is especially true in light of the inevitable confounding factors that one may only imperfectly control for in carrying out the analysis.

However, there are alternatives that one could use in attempting to control for the effects of history. For example, one could simply look at the trend in consumption over the same period of time for a random sample of nonparticipants, the vast majority of whom should be non-installers. Because the source of data is the utility billing records, which include among other variables SIC, price, and energy consumption, there is essentially no cost to collect these data. This analysis may reveal, for example, that the trend in nonparticipant consumption is downward. This may suggest that the estimated reduction in energy consumption due to the installation of the ECM is an overestimate. The savings estimate could then be adjusted downward in a methodologically defensible manner to reflect this.

A non-statistical approach is recommended by Stanley and Campbell (1966). It consists of simply identifying any impact-causing events occurring in the periods just before and after the installation of the equipment that could influence participants' energy consumption. If there are none, then history is less plausible as a threat. Examples of such events are a stock market crash, a sudden increase in interest rates by the Federal Reserve, an OPEC-sponsored oil embargo, a war in the Middle East, a dramatic cutback in military expenditures, or an earthquake.

In short, while we agree that including a comparison group would yield some benefits, investment of M&E dollars in other areas, such as improving the quality of data collected for participants, has a greater chance of increasing the accuracy and precision of *gross* load impact estimates.

Definitions of Model Types

A variety of model types have been used to analyze data across participants and over time.

Conditional Demand Analysis (CDA). The object of a regression model, such as CDA, is to take some measurement that varies over customers, time, or both (e.g., the level of energy consumption for some sample of commercial customers) and attempt to estimate the "true" contribution to that variation which is attributable to the variables of interest, e.g., the installation of an ECM.

Because data are available both over time (e.g., monthly energy consumption and installation of ECMs) for a number of different participants, some utilities have used a cross-sectional time series analysis, sometimes called a pooled time series analysis. The model below will serve as an example. It is a variant of the general model used in the 1990 Southern California Edison Energy Management Services and Hardware Rebate Program Evaluation.

$$E_{i\,t} = \alpha + \beta_1 INSTALL_{it} + \beta_2 PRICE_{it} + \beta_3\, SQFT_i + \beta_4 HOURS_i + + \beta_5 INFLATION_t + \varepsilon_{it} \quad (1)$$

where

E_{it} = energy consumption for the i^{th} customer at time t

α = a constant that captures the energy consumed through a set of unspecified equipment

$INSTALL_{it}$ = installation of the ECM by the i^{th} customer at time t

$PRICE_{it}$ = price of energy faced by the i^{th} customer at time t

$SQFT_i$ = square footage of the i^{th} customer

$HOURS_i$ = operating hours of the i^{th} customer

$INFLATION_t$ = the rate of inflation at time t

β_k = a vector of k coefficients that reflect the energy change associated with a one-unit change in the k^{th} explanatory variables.

ε_{it} = captures the differences in energy consumption among the various customers that are not explained by the model.

Note that INSTALL is a binary variable that takes the value of 0 before the installation of the ECM and a value of 1 afterwards. Such a model is referred to as a UES model since it directly estimates the gross load impacts.

Of course, one could use a pre/post design, in which case the model could be specified as follows:

$$E_{i,Post} = \alpha + \beta_1 INSTALL_i + \beta_2 E_{i,Pre} + \sum_{k=1}^{K} \beta_k X_{ik} + \varepsilon_i \qquad (2)$$

where

$E_{i,Post}$ = post-measure period energy consumption for the i^{th} customer

α = a constant that captures the energy consumed through a set of unspecified equipment

$INSTALL_i$ = installation of the ECM for the i^{th} customer (participant installer = 1 and nonparticipant non-installer = 0)

$E_{i,Pre}$ = energy consumption in the pre-installation period for the i^{th} customer

X = a vector of other explanatory variables, such as changes in the price of energy, square feet, operating hours, and the rate of inflation for the i^{th} customer

β_k = a vector of k coefficients that reflect the energy change associated with a one-unit change in the k^{th} explanatory variables.

ε_{it} = captures the differences in energy consumption among the various customers that are not explained by the model.

A version of this model was used most recently by PG&E in evaluating its nonresidential rebate program (PG&E(b), 1993). The specification included the use of a comparison group in an attempt to control for the effects of history that could not be statistically controlled. The X vector included a number of consumption related variables. In addition, a selectivity correction term could be included to mitigate any bias due to compositional differences emerging between participant and comparison groups. That this bias be controlled is critical since the difference between the two groups should reflect only the effects of history and not be distorted by confounding factors. The objective of including the inverse Mills ratio is to control for any compositional differences between the two groups.

However, there has been a fair amount of methodological controversy regarding whether this is an appropriate method of controlling for selection bias. The details of this debate, which centers around the correct specification and the number of nonlinear models required to model participation, has taken place among a number of evaluators including, most recently, Kenneth Train and Dan Violette. Clearly, this issue is far too complex to be addressed in this report. Others have argued that the most straightforward way to deal with this bias is simply to include in the CDA a number of relevant variables that distinguish participants from members of the comparison group. This assumes that the relationships between these variables and participation is linear rather than nonlinear as in the case of discrete choice models. This methodological controversy will be addressed in subsequent projects managed by the Base Efficiency Subcommittee.

CDA approaches can vary widely, and legitimately. "Simplified" CDA (or S-CDA) developed at SDG&E (1993a) is an important, particularly savings-focused variation on a cross-sectional time series framework, in which relatively few variables are assumed to have effects constant over customers. Other reasonable approaches introduce more constraints on parameters, or may involve various transformations of model variables, importantly including "differenced" or change models of consumption.

Classic Conditional Demand Analysis (C-CDA). A C-CDA model is a special type of regression model that is based on the simple identity that total energy consumption is equal to the sum of the energy consumption for each piece of equipment in a house or building. C-CDA, then, is used to estimate the energy consumption associated with each piece of major equipment for a given customer or building. Of course, to provide such estimates, C-CDA models incorporate a great deal of site-specific data on customers as well as energy consumption data. For example, a C-CDA model can be defined as:

$$E_i = \alpha + \sum_{k=1}^{K} \beta_k X_{ik} + \varepsilon_i \tag{3}$$

where

E_i = energy consumption for the i^{th} customer

α = a constant that captures the energy consumed through a set of unspecified equipment

X_{ik} =a vector of all the major energy using equipment for the i^{th} customer

β_k =a vector of k coefficients that reflect the energy change associated with a one-unit change in the k^{th} explanatory variable.

ε_i = captures the differences in energy consumption among the various customers that are not explained by the model.

A perhaps clearer example of how a C-CDA model could be defined is presented below:

$$E_i = \alpha + \beta_1 AC_i + \beta_2 WH_i + \beta_3 CD_i + \ldots \beta_k X_{ki} + \varepsilon_i \quad (4)$$

where

E_i = energy consumption for the i^{th} customer

α = a constant that captures the energy consumed through a set of unspecified equipment

AC_i = space heating for the i^{th} customer

WH_i = water heating for the i^{th} customer

CD_i = clothes drying for the i^{th} customer

β_k = a vector of k coefficients that reflect the energy change associated with a one-unit change in the k^{th} explanatory variables.

ε_i = captures the differences in energy consumption among the various customers that are not explained by the model.

Appliance variables are coded as dummy variables, i.e., 1 if the equipment is present and 0 otherwise.

Such a model is called a UEC model because it is used to obtain separate estimates of pre- and post-intervention UECs. Through simple subtraction of these two cross-sectional sets of estimates, gross load impacts can be obtained for the measures of interest. The cross-sectional C-CDA is discussed since it is most consistent with the second quantitative relationship illustrated in Table 5, Section B.2 of the Protocols.

The pre- and post-intervention UECs could be estimated in a single model using a cross-sectional time series analysis. Such a model might look like the following.

$$E_{it} = \alpha_{pre} \bullet PRE + \alpha_{post} \bullet POST + \beta_{1,pre} AC_{it} \bullet PRE + \beta_{1,post} AC_{it} \bullet POST + \beta_{2,pre} WH_{it} \bullet PRE + \beta_{2,post} WH_{it} \bullet POST + \beta_{3,pre} CD_{it} \bullet PRE + \beta_{3,post} CD_{it} \bullet POST + ... + \beta_{k,pre} X_{k,it} \bullet PRE + \beta_{k,post} X_{k,it} \bullet POST + \varepsilon_{it} \quad (5)$$

where

α_{pre} = a constant in the pre-installation period that captures the energy consumed through a set of unspecified equipment

α_{post} = a constant in the post-installation period that captures the energy consumed through a set of unspecified equipment

$\beta_{k,post}$ = the kWh associated with the k^{th} energy-using equipment after ECM installation, i.e., UEC of energy-using equipment in the pre-installation period

$\beta_{k,pre}$ = the kWh associated with the k^{th} energy-using equipment prior to installation of ECM, i.e., UEC of energy-using equipment in the post-installation period

PRE and POST = dummies indicating the pre- and post-installation periods.

The difference between the UEC in the pre- and post-installation period would indicate gross impact. The UEC of any particular ECM may depend on installation of some other ECM as well. For example, the UEC of AC depends upon installation of energy efficient lights. Such secondary effects can be captured by expanding the above model to include such interactions between ECMs.

In order to estimate the UEC of any energy-using equipment with the help of only one model specification, information about the complete set of equipment in the pre- and post-installation periods for the same sample of customers is required.

The advantage of such a model, although to our knowledge it has never been tried, is that one could avoid estimating two separate UEC models to estimate gross load impacts.

Statistical Comparison Methods (SCM). 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 average 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.

$$Savings = NAC_{Pre} - NAC_{Post} \quad (6)$$

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.

Calibrated Engineering Methods (CEM). These methods, as they have been applied in impact evaluation, use initial engineering estimates of impacts combined with a "statistical verification" step. The verification step produces an estimated realization rate. CEM approaches differ from SAE models in that regression equations are not used; like SAE methods they require that initial engineering estimates be developed for the population of program participants (or a very large fraction of the participants). A sample is drawn from these participants; then an in-field metering or enhanced/in-depth engineering analysis is conducted. These analyses essentially "verify" or serve as "audited" values of the engineering estimates. A ratio is calculated between the audited 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 CDA are not required by this method.

It is important to note that CEM is different in concept from the calibration of engineering based energy simulation models. These simulation models provide data on usage levels (e.g., billing data or load research data). After recorded end-use level data and/or whole-building data is used to calibrate these models, gross load impacts are then calculated by two runs of the engineering model—a baseline model run and a model run incorporating the energy efficiency measures.

Incorporation of Prior Information into Models

There are various techniques for incorporating prior information regarding gross impacts into many of the model types already discussed. The two techniques described below are statistical engineering estimates (SAE) and Bayesian methods.

SAE Estimates. An SAE model is a type of regression model, either a CDA or a C-CDA, that has been enhanced by including engineering estimates of *use* or the *gross load impact* as an independent variable rather than a dummy variable. Thus, whether engineering estimates are directly incorporated into models is a defining characteristic in model types 2, 4, 6, and 8 listed on pages 2-1 and 2-2.

There appear to be no restrictions on the source of these engineering estimates, often called engineering priors. For example, they can be based on engineering algorithms applied by a utility field representative using customer-specific information; or the evaluation analyst can estimate the prior by specifying an engineering function using location-specific data (such as square footage obtained through a mail survey) and weather data. Equation (7) is an example of the use of engineering priors in the context of a cross-sectional time series design.

$$E_{it} = \alpha + \beta_1 ENG_{it} + \beta_2 PRICE_{it} + \beta_3\ SQFT_i + \beta_4 HOURS_i + \beta_5 INFLATION_t + \varepsilon_{it} \tag{7}$$

where

$E_{i\,t}$ = energy consumption for the i^{th} customer at time t

α = a constant that captures the energy consumed through a set of unspecified equipment

ENG_{it} = engineering estimate of the gross load impact for the i^{th} customer at time t

$PRICE_{it}$ = price of energy faced by the i^{th} customer at time t

$SQFT_i$ = square footage of the i^{th} customer

$HOURS_i$ = operating hours of the i^{th} customer

$INFLATION_t$ = the rate of inflation faced by the i^{th} customer at time t

β_k = a vector of k coefficients that reflect the energy change associated with a one-unit change in the k^{th} explanatory variables.

ε_i = captures the differences in energy consumption among the various customers that are not explained by the model.

Using a pre/post evaluation design in conjunction with a comparison group of non-installers, PG&E recently used the following basic functional form to estimate gross load impacts:

$$E_{i,Post} = \alpha + \beta_1 ENG_i + \beta_2 E_{i,Pre} + \beta_2 MILLS_{\,i} + \sum_{k=1}^{K} \beta_{\,k} X_{ik} + \varepsilon_i \tag{8}$$

where

$E_{i,Post}$ = post-measure period energy consumption for the i^{th} customer

α = a constant that captures the energy consumed through a set of unspecified equipment

ENG_i = prior engineering-based estimate for first-year gross load impact associated with the installation of the ECM for the i^{th} customer (equaling zero for nonparticipants)

$E_{i,Pre}$ = pre-measure period energy consumption for the i^{th} customer

$MILLS_i$ = the selectivity correction factor for the i^{th} customer

X = a vector of other explanatory variables, such as changes in the price of energy, square feet, operating hours and the rate of inflation for the i^{th} customer

β_k = a vector of k coefficients that reflect the energy change associated with a one-unit change in the k^{th} explanatory variables.

ε_i = captures the differences in energy consumption among the various customers that are not explained by the model.

The interpretation of the coefficient of the ECM installation variable is different than in the non-SAE-models which rely on binary identification of the ECM. In the SAE model, the coefficient is interpreted as an adjustment factor on the engineering prior. For example, a coefficient of .75 means that only 75% of the engineering estimate was realized.[3] In a C-CDA version of an SAE model, the engineering prior is an estimate of the UEC. In such a model, the interpretation of the coefficient is the same as above, i.e., a coefficient of .75 means that only 75% of the engineering estimate of the UEC was realized. In a non-SAE models, the coefficients represent the estimated level of savings or use rather than an adjustment of an engineering prior.

Bayesian Inference Within Statistical Models. One concern with current evaluation efforts is that each effort essentially starts over, ignoring the results of previous evaluations. Bayesian methods of inference are one way to incorporate prior information within an estimation framework. These methods start with initial estimates of impacts (which could be last year's evaluation results adjusted for known program changes). The prior information can take the form of constraints, e.g., that impacts be greater than zero and less than 30%. The prior information can instead be expressed as a distribution, as in more traditional Bayesian analyses. The use of prior information in a Bayesian framework assumes that the collection of information is cumulative. Researchers start with existing estimates of impacts and then revise those estimates based on new information.

All of the statistical methods considered in this document are amenable to Bayesian treatment, which represents a change from classically-based methods of making inferences about observed data. In contrast to classical inference, in which "statistical significance" is assessed without regard to prior findings or informed beliefs, confidence in a

[3] If the vector of explanatory variables, X, contains variables used in the calculation of the engineering prior, ENG, the derivation of the realization rate is slightly more complex.

given estimate depends both upon observed results and quantified prior beliefs or results and upon estimates of the credible variance of these prior beliefs or results. It should be stressed that Bayesian analysis has no direct connection with the use of engineering priors in SAE models, although this would be fertile territory for its use. The approach has made its way, via decision analysis, into various evaluation and policy arenas.

Definitions of Model Attributes

The list of attributes by which one could evaluate different types of statistical models is presented below.

UECs/EUIs. Some models, like C-CDA, can produce estimates of gross impacts by first estimating the pre-and post-installation UECs or energy utilization indices (EUIs) for given pieces of energy using equipment. The next step is to subtract the post UEC from the pre UEC to produce a UES, the estimate of the gross load impact. Other models, like the CDA, can directly estimate UESs. If a utility chooses the latter approach to estimating UESs, the Protocols still require the utility to produce participant- and nonparticipant-based estimates of pre- and post-installation UECs/EUIs for energy using equipment. The primary reason for this requirement is that the change in UEC/EUI can serve as a check on the plausibility of the UES that is directly estimated. There is an additional, although perhaps secondary, role for UECs. Any UECs estimated for a random sample of the participants or UECs estimated for random samples of the participants and nonparticipants, properly weighted, can also be used in both short- and long-range forecasting models. Therefore, the first attribute that will be used to evaluate each of the various model types is whether it produce both pre- and post-installation UECs or EUIs as a by-product of estimating UESs.

Also, recall that the fifth and sixth objectives of the larger study described in the Introduction are to investigate the cost-effectiveness of providing estimates of pre- and post-installation UECs/EUIs for both participants and nonparticipants and to investigate alternatives to providing these estimates. Depending on the outcome of these two tasks, the need for the UEC/EUI attribute could vanish.

Data Collection Requirements. Some models require more data to estimate than others. For example, C-CDA models require more data than CDA models since C-CDA models require an inventory of the major energy-using equipment. On the other hand, if a utility decides to use an SAE model in a CDA framework, then it must obtain the required engineering priors. However, this effort may be minimal if these engineering priors are produced routinely as a part of the implementation of the program. So, whether the data requirement associated with using an SAE approach is great or small depends on whether the data already exist. Of course, the highest data requirements could occur for a

C-CDA model with engineering priors. One can also choose to estimate a CDA non-SAE model which can have smaller data requirements. Finally, one can choose a SCM approach, which has the smallest data requirements of all the models discussed since it depends on already existing data.

Error. In general, there are five sources of error: (1) measurement error, (2) specification error, (3) sampling error, (4) the precision with which the independent variables are measured, and (5) non-response error. The first and second sources may produce biased but efficient estimates. The third and fourth sources are concerned with the precision of the estimate regardless of whether it is biased. Finally, the fifth is concerned with the extent to which the results can be generalized to the population of participants.

Measurement error undermines the reliability and validity of the observed variables participating in the regression model, thus empirically weakening the estimates of the variables' impact on energy use or demand. Prime examples include customer reports of square footage or operating hours or the attribution of global, single weather station temperature measurements to all customers in a zone that is actually temperature-heterogeneous. The results of measurement error are basically two: (1) the estimated coefficients will be biased, and (2) the calculated F and t statistics are not correct and the tests are not valid. In other words, measurement error may lead the analyst to draw incorrect conclusions about excluding variables from the model, perhaps to the point of excluding important variables and vice-versa.

The second source of error is specification error, which results when theoretically relevant variables in a model are omitted or, conversely, irrelevant variables are included. When *relevant* variables are omitted, the result is that the coefficients will very likely be biased. Including *irrelevant* variables can sometimes result in a larger standard error of the estimate.

The third source of error, sampling error, occurs when one analyzes data from a subset of a larger population. Even a well-specified model involving well-measured indicators will be subject to sampling error, except in the extreme case of a model estimated over an entire population. In such a case—for example, a well specified model involving data drawn from all participant records in one's tracking system—sampling error is reduced to zero. Thus, the results are straight forwardly generalizable to the population, and in fact simply describe the population.

The fourth source of error is a function of the precision with which the independent variables are measured. Certain models do not use a simple dummy variable to define the installation of the efficient equipment. Rather, they directly incorporate a prior engineering-based estimate of the expected savings as the indicator of an installation. This is important since, generally, the more valid the indicator of the underlying concept the greater the precision of the estimate. There is recognition of this principle in Volume 1 of the Impact Evaluation of Demand-Side Management Programs produced by the Electric Power Research Institute. It states:

> If, in fact, this information is hierarchically accurate (that is, the house is shown to have a lower savings potential by the engineering model and does in fact achieve lower savings), then the precision of the estimated savings should increase. If the engineering model is so inaccurate as to not even be able to hierarchically rank houses by savings potential, then the model estimates may be less precise. (p. 7-18)

Finally, whenever one attempts to obtain information from a sample of participants, the response rate is almost always less than 100%. A problem emerges if the reasons for non-response such as "refusal" and "no longer in business" are systematic rather than random. If this is the case, then the estimated gross impacts may not be generalizable to the larger population of participants. It is sometimes possible to estimate the bias and appropriately adjust the sampling weights before the analysis begins.

The implications of these various types of errors for the precision of the impact estimates and the generalizability of the results to the larger population of participants can now be briefly addressed. First of all, the more data one collects, the greater the probability that some of these data will be incorrectly measured. Moreover, the more data that are required to estimate a given model, the more likely it is that the data can be collected from only a sample of participants. This introduces sampling error, introduces the potential for non-response bias, and decreases the precision of the estimated impact. However, it may also be the case that estimating parsimonious models using only tracking system data allows one to include the entire population of participants, thus eliminating sampling error, insuring generalizability, and minimizing measurement error. However, in such models, the likelihood that relevant variables are omitted may increase, thus increasing the likelihood of specification errors. Thus, because error is everywhere, analysts are forced to use their professional skills and judgment to make tradeoffs among the different types of error with the objective of achieving a net overall minimization of error.

Because the combined effect of each of these sources of error is in some cases impossible to quantify; because every analyst must make tradeoffs among the various types of error; and because the quality and quantity of the data available for any model cannot be predicted, the magnitude of the uncertainty for each of the model types is described in only very general terms.

Cost. The variation in the cost of using the various types of models derives from the input data requirements, the methods of data collection, and the number of models that one must estimate. The cost of obtaining all the data necessary to estimate a C-CDA model is indeed very high. Also, the cost of obtaining prior estimates of gross impacts based on an engineering analysis using customer-specific data can be equally high. Of course, the cost of a C-CDA model with engineering priors could be astonishingly high. However, sometimes engineering priors are available at no additional costs; some utilities routinely estimate engineering-based gross impacts as a part of normal program implementation. Models using only available energy consumption and weather data, as in the SCM use data that are essentially free.

The methods of data collection and data collectors chosen can also have an enormous impact on costs. The least expensive sources of data are those maintained routinely by the utility, either as a part of its billing and weather databases or as a part of its program tracking system. Other forms of data collection, moving from the least to the most costly, are mail surveys, telephone interviews, on-site surveys, and metering. Finally, regardless of the data collection method, the technical skills of those involved in collecting the data will also have serious cost implications. Generally, the more skilled the data collectors, the greater the costs.

Finally, the number of models that one must estimate explains an additional part of variation in cost. For example, if one chooses to run a CDA model that directly produces an estimate of the UES, then, in order to produce estimates of pre and post UECs, as currently required by the Protocols, another completely separate modeling effort would be required.

The costs associated with each of the model types are only described in very general terms, since there can be great variation with respect to how much data a given utility routinely collects, the methods of data collection for each type of data, the skills of those involved in collecting the data, and the types of modeling approaches chosen. As a result of these uncertainties, cost estimates associated with different model types are only relative. For example, while the costs may high in absolute terms, the CDA cross-sectional/time series models are generally less expensive than any C-CDA model.

Robustness. An estimator (a formula or recipe by which the data are transformed into an actual estimate) is said to be robust

> ...if its desirable properties are not sensitive to violations of the conditions under which it is optimal. In general, a robust estimator is applicable to a wide variety of situations, and is relatively unaffected by a small number of bad data values (Kennedy, 1991, p. 32).

All things being equal, robustness is a desirable quality, as the strict assumptions of most estimators are rarely met by the data available to analysts.

Model Evaluation

To fill in each cell of this matrix, each of the ten models is evaluated in terms of each of the five attributes. Tables 1 through 5 present these evaluation results These results of these evaluations should not be interpreted as *categorical pronouncements* but only as *tendencies* for certain model types to manifest certain characteristics.

Table 1. Evaluation Matrix For Statistical Model Types: UECs

<table>
<tr><td>1. C-CDA Cross-sectional w/ dummies</td><td rowspan="4">For C-CDA models, UECs are produced in the process of estimating UESs.</td></tr>
<tr><td>2. C-CDA Cross-sectional w/ engineering priors</td></tr>
<tr><td>3. C-CDA CSTS w/ dummies</td></tr>
<tr><td>4. C-CDA CSTS w/ engineering priors</td></tr>
<tr><td>5. CDA CSTS w/ dummies</td><td rowspan="4">For CDA models, UECs are not produced in the process of estimating UESs. A modeling effort separate from the M&E study is required to produce UECs.</td></tr>
<tr><td>6. CDA CSTS w/ engineering priors</td></tr>
<tr><td>7. CDA Pre/Post w/ dummies</td></tr>
<tr><td>8. CDA Pre/Post w/ engineering priors</td></tr>
<tr><td>9. Statistical Comparison Methods (SCM)</td><td>UECs are not produced in the process of estimating UESs.</td></tr>
<tr><td>10. Calibrated Engineering Method (CEM)</td><td>In general, UECs are not produced. However, if the initial estimates are for a specific end-use and end-use load research is used to audit or calibrate the initial estimate then before and after UECs can be produced.</td></tr>
</table>

Table 2. Evaluation Matrix For Statistical Model Types: Input Data Requirements [4]

1. C-CDA Cross-sectional w/ dummies	A full equipment inventory is required for n customers for both the pre- and post-installation periods. Also required are other variables that influence UECs and vary across customers.
2. C-CDA Cross-sectional w/ engineering priors	Full equipment inventory is required for n customers for both the pre- and post-installations periods. Also required other variables that influence UECs and vary across customers. Finally, this model type requires initial engineering estimates of UECs, both pre and post.
3. C-CDA CSTS w/ dummies	A full equipment inventory is required for n customers for both the pre- and post-installation periods. Also required are other variables that influence UECs and vary within and across customers.
4. C-CDA CSTS w/ engineering priors	A full equipment inventory is required for n customers for both the pre- and post-installation periods. Also required are other variables that influence UECs and vary within and across customers. In addition, an engineering estimate for each UEC, for each month, is required.
5. CDA CSTS w/ dummies	Only information on equipment installed by the program, plus standard tracking system data, is needed. However, one might want to collect data on changes in other energy-using equipment to test for potentially confounding factors as well as other factors that influence energy use.
6. CDA CSTS w/ engineering priors	Only information on equipment installed by the program, plus standard tracking system data, is needed. However, one might want to collect data on changes in other energy-using equipment to test for potentially confounding factors as well as other factors that influence energy use. In addition, engineering priors are required for equipment installed through the program.
7. CDA Pre/Post w/ dummies	Requires same data as CDA CSTS with dummies. Monthly data must be collapsed to lower frequency, i.e., average monthly energy consumption.
8. CDA Pre/Post w/ engineering priors	Requires same data as CDA CSTS with dummies. Monthly data must be collapsed to lower frequency, i.e., average monthly energy consumption.
9. Statistical Comparison Methods (SCM)	The only data required are consumption data and weather data on a set of participants.
10. Calibrated Engineering Method (CEM)	Information is needed on equipment installed by the program *and* initial engineering estimates of energy savings by measure and by site.

4 In general, whether using a pre/post evaluation design or a cross-sectional time series design, the more information that one has on variables that change over time, the better. For example, such data such as equipment inventory could be collected monthly. However, budgetary realities may dictate that while the models could certainly benefit from such data, many models must settle for less. Sometimes, it is possible to elicit from customers at a single point in time information on equipment, business practices, or building operation that have changed over time. Nevertheless, many models continue to use a great deal of data that, except for the ECM and a few other variables, does not change over time.

Table 3. Evaluation Matrix For Statistical Model Types: Error

1. C-CDA Cross-sectional w/ dummies 2. C-CDA Cross-sectional w/ engineering priors 3. C-CDA CSTS w/ dummies 4. C-CDA CSTS w/ engineering priors	In general, a C-CDA model, because of its greater data requirements, will experience greater measurement error, sample error, and non-response error than a model that has less demanding data requirements. However, these same data requirements also mean that it will be less likely to omit a relevant variable. Finally, if a C-CDA model uses engineering priors, it will manifest a smaller standard error than one that uses a binary indicator variable.
5. CDA CSTS w/ dummies 6. CDA CSTS w/ engineering priors 7. CDA Pre/Post w/ dummies 8. CDA Pre/Post w/ engineering priors	Models that have much less demanding data requirements than CDA models will experience less measurement error, sampling error, and non-response error. However, these same data requirements mean that there is a greater likelihood that a relevant variable will be omitted. If a less complex model uses engineering priors, it will generally manifest a smaller standard error than one that uses a binary indicator variable.
9. Statistical Comparison Methods (SCM)	SCM models, because they have the smallest data requirements, will experience the smallest amount of measurement error, sampling error, and non-response error. However, they will also be the most likely to omit important variables and have the lowest level of precision.
10. Calibrated Engineering Method (CEM)	CEMs can be subject to sampling error which, in general, increases as the sample size for verification of the prior (CE) Method engineering estimates decreases. They are also subject to non-response error and measurement error both of which are related to the intensity of the data requirements but which may not be significant. However, there may be some error associated with the engineering priors that should not be ignored when discussing CEM uncertainty. Also, standard errors can be provided on both program and an end-use saving basis (if end-use data are available). When two measures are installed on a single end-use, measure-specific impact estimates and standard errors cannot be developed.

Table 4. Evaluation Matrix For Statistical Model Types: Cost

1. C-CDA Cross-sectional w/ dummies	Costs are high because of the large amount of data required and because two UEC models must be estimated.
2. C-CDA Cross-sectional w/ engineering priors	Costs are high because of the large amount of data required and because two UEC models must be estimated. There may be additional costs if the utility, as part of normal program implementation, does not normally provide customers with engineering-based estimates of UECs. In such a case, these engineering priors will have to estimated for those customers in the analysis, which is very costly.
3. C-CDA CSTS w/ dummies	Costs are high because of the large amount of data required. However, CSTS models allow for the possibility of estimating pre- and post-UECs in one model.
4. C-CDA CSTS w/ engineering priors	Costs are high because of the large amount of data required. However, CSTS models allow for the possibility of estimating pre- and post installation UECs in one model. There may be additional costs if the utility, as part of normal program implementation, does not normally provide customers with engineering-based estimates of gross impacts. In such a case, these engineering priors will have to be estimated for those customers in the analysis. However, CSTS models allow for the possibility of estimating pre- and post-installation UECs in one model.
5. CDA CSTS w/ dummies	Costs are low because of the relatively small amount of data required.
6. CDA CSTS w/ engineering priors	Costs are low because of the relatively small amount of data required. However, there may be additional costs if the utility, as part of normal program implementation, does not normally provide customers with engineering-based estimates of gross impacts. In such a case, these engineering priors will have to be estimated for those customers in the analysis.
7. CDA Pre/Post w/ dummies	Costs are low because of the relatively small amount of data required.
8. CDA Pre/Post w/ engineering priors	Costs are low because of the relatively small amount of data required. However, there may be additional costs if the utility, as part of normal program implementation, does not normally provide customers with engineering-based estimates of gross impacts. In such a case, these engineering priors will have to be estimated for those customers in the analysis
9. Statistical Comparison Methods (SCM)	Costs are the lowest due to the reliance on data already collected.
10. Calibrated Engineering Method (CEM)	The costs are usually high relative to other methods, due to the expenses associates with on-site end-use metering or on-site in-depth engineering studies.

Table 5. Evaluation Matrix For Statistical Model Types: Robustness

1. C-CDA Cross-sectional w/ dummies	Regression models under certain assumptions are said to be the best in the sense that they are a minimum-variance estimator. Unfortunately, because these "best" estimators have been designed to exploit these assumptions, violations of the assumptions affect them much more than they do other, sub-optimal estimators.
2. C-CDA Cross-sectional w/ engineering priors	
3. C-CDA CSTS w/ dummies	
4. C-CDA CSTS w/ engineering priors	
5. CDA CSTS w/ dummies	
6. CDA CSTS w/ engineering priors	
7. CDA Pre/Post w/ dummies	
8. CDA Pre/Post w/ engineering priors	
9. Statistical Comparison Methods (SCM)	This method is a statistical method and does not incorporate any control variables other than weather. As a result, it is robust to the extent that certain social, economic, and political factors remain fairly stable over the period of time for which participants are analyzed . If one chooses to use a comparison group to control for such background factors, the robustness of the method is predicated on the composition of the participant and comparison groups . Regression analysis can control for variables whose sampling fractions are different from that in the population as long as those variables are incorporated into the model. The validity of simple statistical comparisons depends upon the randomness and representativeness of the samples selected.
10. Calibrated Engineering Method (CEM)	This approach requires few assumptions other than the drawing of a random, stratified sample of participants; however, the approach is dependent upon the randomness and representatives of the sample for its robustness. There are no modeling assumptions as there are with regression frameworks (e.g., assumptions regarding error terms, correlations among independent variables, omitted variable bias, etc.). This is both a strength and weakness of this method. As was explained the discussion of simple comparison approaches, regression analysis can control for variables whose sampling fractions are different from that in the population, or different between the participant and comparison groups, as long as those variables are incorporated into the model. The validity of the CEM approaches depends upon the randomness and representativeness of the samples selected. The random samples for pre/post metering required by CE Method approaches can be hard to obtain in a DSM setting, particularly when there is no pre-notification prior to customer participation.

3. ENGINEERING MODELS

Development of Model Types

Identifying a taxonomy of the engineering methods that can be used for estimating gross load impacts for space heating and cooling in buildings is much more straightforward than in the case of statistical models. Such a taxonomy is already provided in the ASHRAE Handbook of Fundamentals. The engineering models that are used most commonly to calculate energy savings are based on either of the two general methods within this taxonomy. One type of engineering model uses a relatively simple steady-state method while the other is a more complex dynamic simulation method.

Engineering algorithms constitute a fourth, least complex approach which can address end uses other than heating and cooling such as motors, refrigerators, and water heating.

Thus, four model types will be examined. The first type is the simplest, engineering algorithms. The next two types are steady-state methods while the fourth type is a dynamic simulation method.

1. Engineering algorithms
2. Degree-day methods
3. Bin methods
4. Dynamic simulation methods

Development of Model Attributes

Each type of model is evaluated in terms of the same five major attributes that were used to evaluate statistical models:

1. UEC production
2. Input data requirements
3. Error
4. Cost
5. Robustness.

Using these four basic model types and five attributes, a twenty-cell matrix can be formed. Later, as each model type is evaluated with respect to each of the five attributes, each cell of this matrix will be completed. However, in order to better understand the four model types and the five attributes, the different evaluation designs, model types, and attributes must first be defined.

Definitions of Evaluation Designs

Pre/Post. When energy use in a building is analyzed through simulations with an engineering model, it is straightforward to estimate the savings associated with any particular energy efficiency measure. First, an energy analysis simulation of a facility is constructed under conditions where the energy efficiency measure is not installed or under conditions where the measure at least meets federal or state efficiency standards. Second, an energy analysis simulation of the facility is then constructed with all conditions the same but with the energy efficiency measure now installed. Third, the results of the two simulations are compared to determine the energy savings attributable to the energy efficiency measure. This design has been used for the evaluation of both retrofit programs and new construction programs

Static Group Comparison. The lack of a "before" condition with new construction projects requires that a suitable energy use baseline be established as the basis for comparison to the energy use associated with the efficiency measures encouraged by the DSM program. One often-used baseline is defined by the level of energy efficiency established by the California Building Energy Efficiency Standards, known as Title 24. However, other baselines can be defined by considering the measures installed and the energy used in buildings that did not participate in a new construction program. If one chooses nonparticipating buildings as the appropriate point of comparison, then one is using the static group comparison evaluation design.

With a sufficiently large population, with homogeneity among buildings and in decision making, and with close matching of participant and nonparticipant buildings, the use of nonparticipant buildings to define baseline energy use is relatively straightforward. However, with commercial new construction, there are generally small populations, heterogeneous projects, and diverse building types. Accordingly, the use of nonparticipants requires that care be taken to control for extraneous effects that could unduly influence the impact evaluation. Essentially, nonparticipants need to be matched against participants on a number of important characteristics, including type, geographic location, and size.

When using nonparticipants in an evaluation of a new construction program, there are also practical difficulties in finding nonparticipant buildings whose major characteristics match those of participants and in inducing them to cooperate in the data collection activities. To the extent that a group of nonparticipant buildings cannot be found whose characteristics match against those of the participant buildings, then the engineering and econometric analyses must be designed to control as much as possible for these differences in order to avoid seriously biasing the results. The effort is worthwhile, however, because of the information that nonparticipant buildings provide about the nature of baseline construction practices. Without this information, an evaluation has to rely solely on inferences about what the participants would have done absent the program, based on what those same participants are willing or able to reveal about their own past decision making processes and construction practices. For a process as complex as commercial new construction, this may be expecting too much.

Definitions of Model Types

Engineering Algorithms. In some cases, it is possible to use relatively simple engineering algorithms to calculate the savings that result from the installation of energy efficiency measures. This is particularly true for measures that affect non-conditioning end uses, such as lighting. For such non-HVAC end uses, a basic algorithm for estimating energy use is given by the following equation:

$$kWh = kW \bullet FRAC \bullet U \tag{9}$$

where

kWh = estimated energy use

kW = the connected capacity

FRAC = the estimated operating fraction

U = a utilization multiplier to account for cycling and diversity.

This calculation can be made for different types of days. Moreover, if a load is not constant, then the values of the factors need to reflect these variations as well. The connected load can be determined in a relatively straightforward fashion by using equipment inventory information (i.e., counts of the types and sizes of equipment found at the site). For example, for lighting, a fixture listing gives the number of fixtures and the lamp and ballast types. However, information on operating profiles and on utilization factors are also needed in order to have an engineering algorithm provide a good estimate of energy use and energy savings. Operating profile information is needed that shows the percent of equipment that is operating by hour and type of day. Key issues here are the percent of equipment that is on during the day, the percent that is on at night, and the transition profile between the two states. Utilization factor information provides a bridge between equipment rated capacity and expected average load when operating. Utilization factors account for equipment cycling, wait-state energy requirements, usage patterns, and diversity.

Engineering Models Based On Degree-Day Methods. Degree-day methods of energy estimation represent one variant of steady-state methods. In general, steady-state methods are applicable under conditions where building use and the efficiency of HVAC equipment are constant. They are generally not applicable when indoor temperature or interior heat gains or losses vary. There are two types of degree-day methods: the traditional degree-day method and the variable base degree-day method.

The traditional degree-day method for estimating heating or cooling energy requirements is based on the assumption that, on a long-term average, solar and internal heat gains for a building will offset heat loss when the mean daily outdoor temperature is 65°F. Under this assumption, net heating or cooling load is therefore proportional to the difference between the mean daily temperature and 65°F. With the traditional degree-day method, a degree day is measured by the difference between the mean daily outdoor temperature and 65°F.

The variable base degree-day method calculates a degree-day measure using a base temperature that is specific to a building. That is, the variable base degree-day method is the same in principle as the traditional degree-day method, with the exception that it uses the daily average temperature at which a building experiences neither heating nor cooling (zero load) as the base temperature for computing degree days. This base temperature, which is referred to as the balance temperature or break-point temperature, is the outdoor temperature at which the internal heat sources plus solar heat gain exactly offset the envelope heat loss due to conduction and air leakage. Because of differences in climatological conditions and indoor setpoint temperatures between summer and winter, there are two balance temperatures defined for a building.

Various investigations that have used metered data on heating or cooling energy use to evaluate the accuracy of degree-day methods have shown that degree-day measures based on temperatures other than 65°F (i.e., variable base degree days) generally explain energy use better the traditional degree-day measure based on 65°F. This is particularly true, for example, for energy efficient buildings where air leakage is reduced and the indoor temperature set lower than 75°F but where the internal heat generated by appliances and lighting is increased. Because balance temperatures can be expected to vary significantly across a set of buildings, the ASHRAE Handbook of Fundamentals concludes: "... in general, degree days with the traditional base of 65°F are not to be used."

The variable base degree-day method was implemented as a computerized model in the Simplified Calculation Methodology that the California Energy Commission developed and, until the revision of the standards in 1992, certified as one of the procedures that could be used to comply with California's Title 24 Nonresidential Building Energy Efficiency Standards.

Engineering Methods Based on Bin Methods. With the bin method, energy use is calculated by evaluating energy use separately for different temperature interval "bins" and time periods. A temperature "bin" is usually defined to include temperatures within a 5°F interval (e.g., 70°F–74°F, etc.). The result of the calculation at each condition is multiplied by the number of hours of occurrence for that condition, and these results are then summed to arrive at the overall estimate of energy use. One common refinement of the bin method is to use coincident wet-bulb temperature for each bin.

When applied to houses, the bin method can be simplified through the development of one linear relationship representing heat loss (or heat gain) per degree of outdoor temperature. The sensible and latent loads due to infiltration and ventilation are not included in that relationship; rather, they are calculated by using a bin's average temperature and coincident wet-bulb temperature for each bin.[5]

The bin method of energy use estimation has been implemented in a computerized form in the ASHRAE TC 4.7 Energy Analysis Model and in some proprietary models.

Dynamic Simulation Engineering Models. Dynamic simulation engineering models represent the state of the art for estimating and analyzing energy use in buildings. Such models (e.g., DOE2, BLAST) implement the refined procedures that have been developed and published by ASHRAE for calculating heating and cooling energy use in a building. Using these procedures, the computerized dynamic simulation energy analysis models calculate the amount of heat entering or leaving a building for each hour of the year. The models consider the hourly variations in weather and internal schedules that produce variable instantaneous (e.g., light or people) or delayed (e.g., heat flow through an exterior wall) heat gains or losses in a space inside a building.

Model Attributes

As mentioned earlier, all the attributes are the same as those used for the evaluation of the statistical models.

---

5 In engineering calculations of energy use, a distinction is made between "sensible" and "latent" loads. Sensible loads are associated with heat transfer that occurs because of differences in temperature between a heat source and the surrounding space or air. This type of heat transfer, which is manifested by a rise in the tempeature of the air, includes direct radiation to the air and direct heating of circulated air. Latent loads are associated with the heat transfer that occurs when water is transferred to atmospheric air (i.e., in humidification). Latent heat transfer is manifested primarily by vaporization or condensation.

Model Evaluation

The strengths and weaknesses of engineering models that are based on the different methods are outlined in the following tables. The tables begin with the least complex models and then moves through models of increasing complexity. Tables 6 through 10 provide the evaluation of each of the four model types in terms of the five attributes.

Once again, these evaluations should not be interpreted as absolutes, but as tendencies for certain characteristics to hold true.

Table 6. Evaluation Matrix For Engineering Model Types: UEC

1. Engineering Algorithms	Engineering algorithms can produce estimates of energy use that can be used to calculate UECs.
2. Degree-Day Methods	The variable base degree-day method produces estimates of energy use that can be used to calculate UECs.
3. Bin Methods	The bin method produces estimates of energy use that can be used to calculate UECs.
4. Dynamic Simulation Methods	The dynamic simulation energy analysis models produce estimates of energy use that can be used to calculate UECs.

Table 7. Evaluation Matrix For Engineering Model Types: Data Collection Requirements

1. Engineering Algorithms	As noted, estimation of energy use for a non-HVAC end use with an engineering algorithm requires data on connected load, operating profile, and utilization factors.
2. Degree-Day Methods	Estimation of energy use with the variable base degree-day method requires data on daily average temperatures, building envelope construction details, characteristics of heating and cooling equipment and systems, and operational and occupancy conditions and schedules.
3. Bin Methods	Estimation of energy use with the bin method requires data on building envelope construction details, characteristics of heating and cooling equipment and systems, and operational and occupancy conditions and schedules. Weather data for applying the bin method have been published by ASHRAE and by the United States Air Force.
4. Dynamic Simulation Methods	Estimation of energy use with an energy analysis simulation model requires two major types of data. One type of data includes detailed data on the physical, thermal, and operational characteristics of a building. These data generally need to be more detailed than the data needed for the degree-day or bin methods. The second type of data needed includes hourly data on weather conditions; this weather data generally includes dry- and wet-bulb temperature, wind velocity, cloud type, and cloud cover index. Such weather data are available from the National Climatic Center or can be developed from data collected at utility weather stations.

Table 8. Evaluation Matrix For Engineering Model Types: Error [6]

1. Engineering Algorithms	Because engineering algorithms have fewer data requirements, they are subject to less error on these variables; samples can be larger since data collection is less onerous; and non-response error is less of a problem. However, because there are fewer data requirements, these models must make assumptions that may or may not be reasonable. Moreover, for HVAC measures, this method is the least sophisticated and therefore potentially the least accurate.
2. Degree-Day Methods	Because degree-day methods have somewhat greater data requirements than engineering algorithms, they are subject to slightly more error on these variables; samples may be smaller since data collection is more onerous; and non-response error is therefore more of a problem. However, because there are more data requirements, these models must make fewer assumptions that may or may not be reasonable.
3. Bin Methods	Because bin methods have somewhat greater data requirements than degree-day methods, they are subject to slightly more error on these variables; samples may be smaller since data collection is more onerous; and non-response error is therefore more of a problem. However, because there are more data requirements, these models must make fewer assumptions that may or may not be reasonable.
4. Dynamic Simulation Methods	Because dynamic simulation methods have the greatest data requirements of all the methods, they are subject to the most error on these variables; samples may be smaller since data collection is most onerous; and non-response error is therefore more of a problem. However, because there are more data requirements, these models must make fewer assumptions. In addition, for HVAC measures, these methods are the most sophisticated and therefore potentially the most accurate.

6 In general, there is a trade-off between multiple runs of simpler models versus fewer runs of more complex models. In effect, this concern is related to more general issues about the precision and the validity of results. These issues arise because estimates of savings are subject to both random error and systematic error. In this context, precision refers to reducing random error, while validity refers to reducing systematic error. Random error can be reduced (i.e., precision improved) most generally by increasing the number of cases studied. However, reducing systematic error requires more accurate measurement. Given a fixed budget for an evaluation study, a trade-off arises because increasing the number of cases studied may require that their energy use and savings be estimated with a simpler, less costly, and less accurate means of measurement. The dimensions of this trade-off as they pertain to the evaluation of savings from DSM programs have not been extensively examined. That is, while there is a good understanding of how sample sizes should be determined to meet precision criteria, the role of measurement error in the analysis of savings has not been addressed to any great extent.

Table 9. Evaluation Matrix For Engineering Model Types: Cost

1. Engineering Algorithms	The cost of calculating energy savings with engineering algorithms is generally low. The major costs are incurred in obtaining the data needed to make the calculations.
2. Degree-Day Methods	The variable base degree-day method is the least costly of the engineering methods of energy savings calculation. However, it may be more costly than statistical methods because data on building construction details and HVAC equipment are needed for each building to which the method is applied.
3. Bin Methods	The bin method of energy estimation is more costly than statistical methods because data on building construction details and HVAC equipment are needed for each building to which the method is applied. The bin method may be slightly more costly to apply than the variable base degree-day method because of the greater detail of the weather data that is used, but it is less costly to apply than dynamic simulation engineering models.
4. Dynamic Simulation Methods	Developing estimates of energy use through simulations with a dynamic simulation energy analysis model has generally been more costly than with the degree-day or bin methods. First, the cost of collecting data is somewhat greater because data of greater detail are needed. Second, the cost of analysis has also been higher because of the analyst's time that is required to set up the input files for the simulation. (However, pre-processor programs have been developed that reduce the amount of time and effort required to prepare DOE2 input files.)

Table 10. Evaluation Matrix For Engineering Model Types: Robustness

1. Engineering Algorithms	The robustness of engineering algorithms for estimating energy use (and hence savings) depends primarily on the accuracy of the data that are used. While relatively accurate data can generally be obtained on connected loads, obtaining accurate data on operating hours and on utilization factors presents more problems. Self-reported information on operating hours and utilization collected through survey procedures may not always be accurate. While information on these factors can be obtained through end-use monitoring, the cost of such monitoring is relatively high (offsetting a major advantage of using the algorithms in the first place).
2. Degree-Day Methods	As a steady-state method of calculation, the variable base degree-day method is not robust for estimating energy use (and hence savings) when the assumption of steady-state conditions is not warranted. As pointed out in the ASHRAE Handbook of Fundamentals: "For many applications, the degree-day method should not be used, even with the variable base method, because the heat loss coefficient, the efficiency of the HVAC system, or the balance point temperature may not be sufficiently constant." Moreover, degree-day methods have been shown to be less accurate in estimating cooling energy use than in estimating heating energy use. They also have lower reliability for estimating energy use during mild weather, where occupant behavior assumes a more prominent role in affecting energy use. Finally, degree-day methods are not well-suited for estimating energy use for multi-zone HVAC systems; they are mainly applicable to single-zone systems.
3. Bin Methods	Although the bin method, like the degree-day methods, is a steady-state method for estimating energy use, it is somewhat more robust than the degree-day method in being able to handle variations in conditions when those variations follow a pronounced pattern. The bin method has this somewhat greater robustness because it allows energy use to be calculated separately for different temperature intervals and time periods. Studies have shown that the bin method can be used to yield good estimates of annual energy use when variations in conditions do follow a pronounced pattern. Moreover, the bin method is better able to estimate energy use for multi-zone HVAC systems than are the degree-day methods. However, the bin method does rely on aggregated weather data and therefore is less accurate than dynamic simulation engineering models that use hourly weather data to calculate estimates of energy use. The bin method does not capture the dynamic effects of factors affecting energy use and therefore is less accurate than the hourly simulation models. Moreover, the time aggregation with the bin method means that estimates of monthly peak energy use may not be accurate. Some inaccuracy may also arise in bin method calculations because solar radiation is assumed to have a fixed relationship to temperature, and this relationship may not be accurately depicted for all conditions.
4. Dynamic Simulation Methods	Dynamic simulation energy analysis models are the most robust of the engineering models for calculating estimates of energy use in buildings. They address hourly energy use in a building and can therefore give more account to the dynamic effects of factors affecting energy use. This is particularly important for evaluating energy use and savings in commercial buildings, which often have high internal loads but which also may vary from hour to hour and season to season. Thus, both energy use and peak demands can be estimated more accurately with dynamic simulation models than with the more simplified engineering models.

BIBLIOGRAPHY

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.

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.

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

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

PG&E(a), *Residential New Construction 1992 Impact Evaluation*, 1993

PG&E(b), *CIA Rebate Program—Billing Analysis*, 1993

PG&E(c), *CIA Rebate Program—Double Ratio Analysis*, 1993

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, pp. 309–321.

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

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.

SDG&E(a). *AEI: 1991 Residential Compact Fluorescent Lighting Program—Impact and Cost-Effectiveness*, 1992

SDG&E(b). *Commercial & Industrial Incentives Program Evaluation Final Report*, 1991

Southern California Edison (a). "Measure Analysis and Recommendation System" (MARS), 1994.

SCE(b). *SCE EMS and Hardware Rebate Program Evaluation, Volume 2*, 1993

SCE(c). *Estimating Gross kWh Savings and kW Reductions Using DOE2, Volume 4*, 1993

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

SCE(e). *First Year Load Impacts of SCE's Residential New Construction Program*, 1993

SCGas(a). *Consumption Impact of SCGas' Residential Conservation Program*, 1993

SCGas(b). *First Year Load Impacts of SCGas' Commercial DSM Programs*, 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*, Volume 6, pp. 131–140. American Council for an Energy-Efficient Economy, 1990