Development of a Standard Baseline Calculation Protocol for Demand Response

Miriam L. Goldberg and Ken Agnew, KEMA-XENERGY Inc., Madison, WI
Michael Messenger, California Energy Commission, Sacramento, CA

ABSTRACT

The California Energy Commission has sponsored the development of a standardized measurement and verification protocol for use in calculating demand reductions by participants in demand response (DR) programs. The basis of this protocol is the calculation of the baseline electricity load profile against which peak load reductions are calculated. Completion of the protocol is aimed at increasing participation in DR programs from small- and medium-sized customers by reducing the barriers related to inconsistency and confusion about baseline methods and ensuring that only real and verifiable peak load reduction receives payments.

A standard protocol is recommended based on interviews with stakeholders in several different jurisdictions and states, as well as statistical testing of several alternative methods and features on a large number of data sets from around the country. The work also proposes terminology for describing baseline calculation methods.

Introduction

During the electricity crises of the last few years, a number of states and utilities within these states have developed programs to encourage customers to reduce their peak loads on short notice (under 2 to 24 hours) in exchange for some form of compensation. Such demand response (DR) programs depend on a credible operational procedure for determining the magnitude of load reductions. Fundamental to determining the magnitude of load reductions is the estimate of baseline load, the load that would have existed if there had been no reduction.

The use of inconsistent methods for calculating baselines and corresponding load reductions has caused both confusion and dissatisfaction among participating customers. For weather-sensitive loads, commonly used methods can provide low and even decreasing incentives during normal periods of extreme temperatures when DR programs are most likely to be in operation. For example, in California, the use of simple averages in the third day of a heat storm often estimated little or no savings because the baseline shape was calculated during relatively cool weather. The lack of a standard measurement procedure may be reducing the number of customers willing to participate in DR programs, particularly in smaller- and medium-sized commercial customers.

This work is intended to provide the foundation for a protocol that may be adopted as part of the International Performance Measurement and Verification Protocol (IPMVP; California Energy Commission 2003). The intent is not to provide a prescriptive set of steps and rules. Rather, the goal is to establish a clear vocabulary to use in describing the methods used to estimate peak load reductions delivered in response to emergency or price conditions, and to offer guidelines on good practice and the pros and cons of alternative calculation methods.

Background

The focus of this study is on calculations of DR (peak savings in kW) using whole-premise interval load data from a sample of 646 customers that vary across account type and geographical location. DR is calculated as the difference between the *baseline* and the actual metered load in each interval (Figure 1). In the event of DR program involvement, actual metered load would be lower for the whole premise as a result of active curtailment, or lowering, of load levels at the site. The baseline is the estimate of what the load would have been in each interval in the absence of the curtailment. Thus, the key question for the DR calculation is how the baseline is determined.

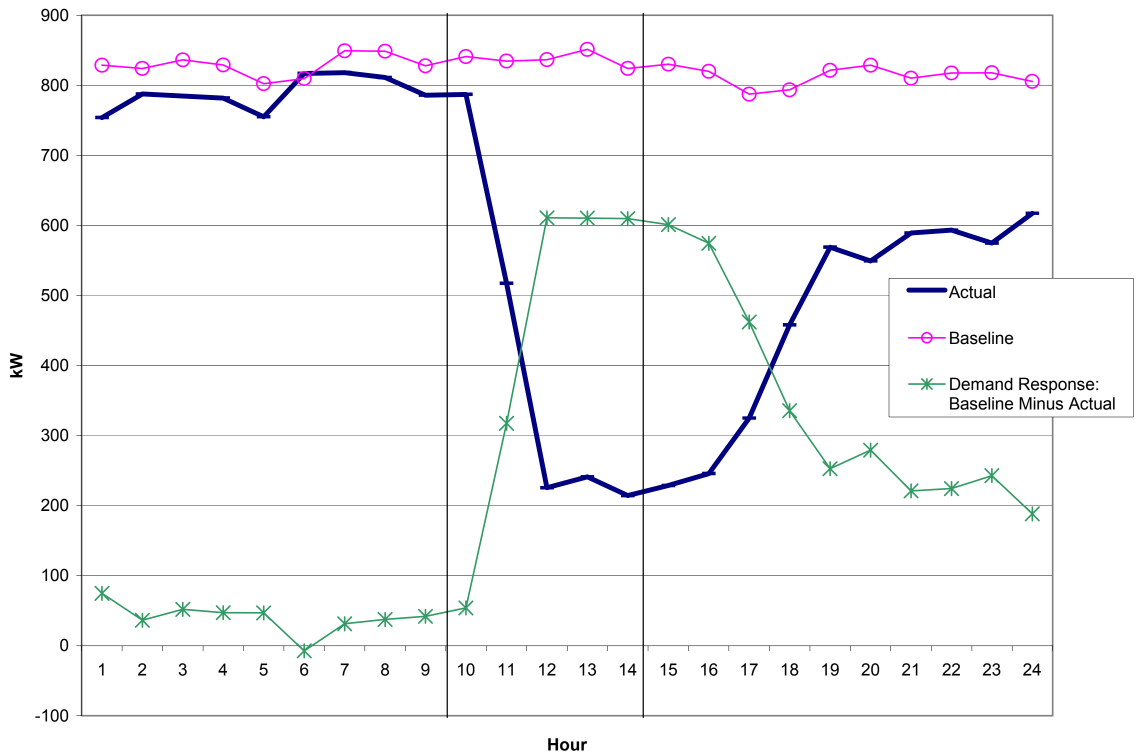

Figure 1. Example of Demand Response Calculation from Baseline and Actual Loads

In the interviews conducted with ISO and utility DR program personnel, we asked for desirable features in a DR baseline calculation method. All customer baseline developers understood that the baseline methodology they chose was a compromise. Criteria that are balanced in developing a baseline include:

- Simplicity;
- Ease of use for program participants and administrators;
- Ease of understanding;
- Verifiability;
- Accuracy;
- Lack of bias (i.e., no systematic tendency to over- or understate reductions);
- Ability to handle weather-sensitive accounts fairly;
- Minimization of gaming;
- Ability to be known prior to customer's commitment to a particular curtailment amount and event;
- Costs for participant and operator to implement; and
- Consistency with other ISOs.

The development of a standard baseline calculation protocol must take into consideration all of these criteria. The body of this paper focuses on the one criteria amenable to quantitative testing, namely accuracy, including both bias and variability. The results and recommendations address the wider list of criteria.

Classifying Baseline Calculation Methods

Baseline calculation methods based on whole-premise interval-metering data can be described in terms of three fundamental components:

- **Data selection criteria** determine what days and time periods of data will be used in the baseline calculation.
- The **estimation method** is a calculation procedure that determines the provisional baseline load at each interval for the curtailment day, using the data selected by the data selection criteria.
- The **adjustment method** shifts or scales the provisional baseline load shape to align it with known conditions of the curtailment day.

Working with these three components provides a framework within which the wide range of baseline method options can be systematically organized. Furthermore, baseline methods presently in use by ISOs and utilities can be characterized in terms of these components.[1]

Data Selection Criteria

Choosing the period from which to draw data requires the balancing of competing priorities. There is a tension between using more data and using more representative data. Customer type, data availability, data handling cost, and the estimation technique to be applied to the data all come into play. In actuality, most programs fall into one of two camps. The initial selection of days for the ISO programs all involve a variation on using the last 2 weeks of business days. Methods used by utilities for program evaluation tend to involve using data for a full season. Within the initial selection of days, it is common to choose a subset based on some sort of ranking, primarily load.

All existing methods delete from the data selection any days that had a control or curtailment event (i.e., a request to reduce load). Some replace these excluded days by going farther back in time as needed. Similarly, some methods screen out days of low or extreme output based on varying criteria and with or without replacement. For this analysis, all excluded control days were replaced with the next eligible day. No extreme output exclusions were tested. If extreme output exclusions were used, they would likely affect high-variability account results. Variations we tested within these general rules are indicated below.

Estimation Method

The two broad types of methods used to estimate baseline load shapes are averaging and regression. We describe these methods more specifically below. In the analysis conducted, all load data were available on an hourly basis, and the analysis was conducted on this basis. In some markets, load data are collected and DR is calculated on a finer time interval, such as half-hour or 15-minute intervals. Our discussion refers to analysis of hourly data. However, the same principles would apply with data at a finer time increment.

---

[1] Sources for ISO baseline protocols are listed in the references.

Averages. Averaging means that the baseline for each time interval of the curtailment day is calculated as the simple average, across all the days chosen by the data selection criteria, of the loads at that time interval. For example, the baseline for the hour ending 1:00 PM is the average over all the selected days of the loads on those days for the hour ending 1:00 PM.

Regression Models. Regression can take a wide variety of forms. The models included in this study can be understood as extensions of simple averaging that allow for the inclusion of weather variables. It is beyond the scope of this work to attempt to determine the best general regression model for this application.

In the context of calculating DR load baselines, the regression model uses the data selected for a particular account and event. The model is fit to those data, and applied to the conditions during the event, to estimate the load that whole premise would have in the absence of the control or curtailment, at each time increment in the event. In all applications reviewed and all methods tested, the model is fit separately for each account.

In most cases, the model is also fit separately for each curtailment event, because different data are selected. The exception is model fitting based on a full season of data. In these cases, the same model fit applies to all events. The estimated loads vary by event because the control-day conditions vary.

In these models, each observation corresponds to a particular day and hour (or finer time interval). The dependent variable is the account's load at that day and hour. In almost all the applications reviewed, a different set of coefficients is estimated for each hour of the day. The predictor variables are typically weather variables and possibly day type. Thus, each observation consists of the account's load for a particular day and hour together with the corresponding weather variables.

Data on factors such as production output, occupancy, or number of shifts operated could potentially be useful predictors for some accounts, particularly those that are not weather-sensitive. However, meaningful, objective, customer-specific variables that track activity by day are typically not available. As a result, we did not test any models with such variables.

Weather variables can be included in a regression in a number of ways. Outdoor temperature can be included on an hourly basis or as an average over the day. Both hourly and day-average temperatures can be included either directly or in degree-day form, representing the difference from some base temperature. A lagged temperature variable can also be included. Humidity can be included either separately or combined with temperature in a temperature-humidity index. We tested model versions with all these weather variables. Other variables such as hours of daylight, sunshine, wind, or precipitation could also be considered.

All the models we tested had hourly varying coefficients including an hourly intercept. That is, sets of 24 hourly coefficients were fit.

Description of Models Tested. The models tested are indicated in Table 1, with the variable definitions indicated in Table 2. Each model was tested with each of the decision rules indicated in Table 3.

Table 1. Model Forms Tested

A	Average. No variables besides the intercept term.	$L_{dh} = \alpha_h$
B	Daily temperature.	$L_{dh} = \alpha_h + \beta_h T_d$
C	Hourly temperature.	$L_{dh} = \alpha_h + \beta_h T_{dh}$
D	Daily heating and cooling degree-days.	$L_{dh} = \alpha_h + \beta_h HDD_d + \gamma_h CDD_d$
E	Hourly heating and cooling degree-hours.	$L_{dh} = \alpha_h + \beta_h HDH_{dh} + \gamma_h CDH_{dh}$
F	Hourly heating and cooling degree-hours with lagged degree-hours.	$L_{dh} = \alpha_h + \beta_{1h} HDH_{dh} + \gamma_{1h} CDH_d + \beta_{2h} LHDH_{dh} + \gamma_{2h} LHDH_{dh}$
G	Hourly temperature-humidity index.	$L_{dh} = \alpha_h + \beta_h THI_{dh}$

Table 2. Variable Definitions

Variable	Definition
L_{dh}	Load at hour *h* on day *d*.
T_d	Daily average temperature (average of daily minimum and maximum) on day *d.*
HDD_d	Heating degree-days base 65°F on day *d.*
CDD_d	Cooling degree-days base 65°F on day *d.*
HDH_{dh}	Heating degree-hours base 65°F at hour *h* on day *d.*
CDH_{dh}	Cooling degree-hours base 65°F at hour *h* on day *d.*
$LHDH_{dh}$	Lagged heating degree-hours base 65°F at hour *h* on day *d.*
$LCDH_{dh}$	Lagged cooling degree-hours base 65°F at hour *h* on day *d.*
THI_{dh}	Temperature-humidity index for hour *h* on day *d.*
$\alpha_h, \beta_h, \gamma_h$	Coefficients determined by the regression, *h* = 1, 2, …, 24.

Table 3. Selection Rules Tested

Code	Label	Selection Rule
1	Previous 10	Previous 10 business days beginning on d0-1 (California and New England ISOs) (California ISO 2001; ISO-NE 2002).
2	Previous 11	Previous 11 business days beginning on d0-1.
3	Previous 10 starting d0-2	Previous 10 business days beginning on d0-2.
4	Previous 20	Previous 20 business days beginning d0-1.
5	Previous 10 and Next 10	20 business days from d0-10 to d0+10.
6	High 10 of 11	Highest 10 of the last 11 business days, beginning d0-1.
7	High 5 of 10, starting d0-2	Highest 5 of the last 10 business days, beginning d0-2 (New York and PJM ISOs) (NYISO 2001, 2002; PJM 2002a, 2002b).
8	Full season	Entire season that includes the control day.
9	Full previous season	Entire season from the previous year that includes the control day.

In model F, the lag degree-day terms are based on lagged temperature calculated as

$$LT_{dh} = \sum_{k=1}^{48} T_{d,h-k} e^{-k/48} \Big/ \sum_{k=1}^{48} e^{-k/48} ,$$

where LT_{dh} is lagged hourly temperature and T_{dh} is hourly temperature and e is the natural base.

There are many different temperature-humidity indexes. For model G, we use PJM's method:

$$THI_{dh} = T_{dh} - 0.55(1\text{-}RH_{dh}/100)(T_{dh}\text{–}58),\ T_{dh} > 58 \text{ , or}$$
$$T_{dh},\ T_{dh} \leq 58.$$

where RH_{dh} is relative humidity at hour d of day h. The temperature-humidity index increases with humidity as well as with temperature, and therefore may be a better predictor of load increases related to cooling.

Conditional Weather Models. Each of the weather models B through G was fit as a "conditional weather model." That is, the weather terms were kept in the model only if certain model diagnostics indicated that these weather terms were physically meaningful (at least positive in total) and statistically well-determined (F-statistic for including the set of cooling [or heating] coefficients significant at the 0.10 significance level).

The full set of cooling coefficients were either all retained or all dropped from the model. Likewise, the full set of hourly heating coefficients were either all retained or all dropped from the model. If both the heating and cooling terms were dropped from the model based on these criteria, the model reduced to a simple average by hour of the day.

Adjustment Method

We tested three different adjustment methods, using two different time periods for the adjustment where possible. Adjustments are designed to update the provisional estimated baseline load curve with data from just prior to the curtailment period. The provisional estimated baseline is simply the unadjusted baseline as estimated by any of the estimation methods described above.

- **Additive adjustment.** A constant is added to the provisional baseline load for each hour of the curtailment period. For simple additive adjustment, the constant is calculated as the *difference* between the actual load and the provisional baseline load for some period prior to the curtailment.
- **Scalar adjustment.** The provisional baseline load for each hour of the curtailment period is multiplied by a fixed scalar. For simple scalar adjustment, the scalar multiplier is calculated as the *ratio* of the actual load to the provisional baseline load for some period prior to the curtailment.
- **Weather-based adjustment.** A model of load as a function of some weather parameter is fit to historical load data. The fitted model is used to estimate load (a) for the weather conditions of the days included in the provisional baseline, and (b) for the weather conditions of the curtailment day. The *difference* or the *ratio* of these two estimates is calculated and applied to the provisional baseline as an additive or scalar adjustment.

Both the additive and scalar adjustments are calculated using two different time periods: the two hours prior to the start of the curtailment, or the third and fourth hours prior to the start of the curtailment. Only full-hour data unaffected by the curtailment are used.

Curtailed Accounts vs. Uncurtailed Accounts

Each account for which load data were available was classified as either curtailed or uncurtailed for each year of data. Curtailed accounts were those that had at least one curtailment period during the year. Uncurtailed accounts had no curtailment events during the year. The latter were either accounts that were in a curtailment program that had no control events that year, or were in the same size class as the accounts in the curtailment program.

For curtailed accounts, the test days were the actual curtailment days. For uncurtailed accounts, the test days were the curtailment days for the similar accounts from the same utility that were in a curtailment program, or else extreme hot or cold days. For uncurtailed accounts, we compared the load estimated by each baseline method for each test day with the actual observed load on that day. For curtailed accounts, we compared each baseline method's estimate of what the load would have been in the absence of curtailment with the estimate given by the "best" method.

Account Type Classification

For this analysis, we solicited interval load data from several parts of the U.S. for both curtailed and uncurtailed accounts. Regions represented included California, the Northeast, Northwest, MidAtlantic, Midwest, Southeast, and Southwest. A total of 646 accounts were used.

Results were generated separately for curtailed and uncurtailed accounts because error is measured differently for the two types. Also, summer and non-summer curtailment period results were produced separately to allow for season-related baseline performance issues. Furthermore, because account type (commercial, industrial) was not known for all accounts, accounts were, instead, classified as weather-sensitive or non-weather-sensitive and low or high variability.

Weather-Sensitivity

Accounts were classified as weather-sensitive or not based on the diagnostics from a weather model fit. The weather model used was an hourly degree-hour model based on a full year of load data, with the degree-day bases estimated as part of the model. This is model "E" as defined above, except that the degree-day bases are parameters estimated by the model. The classification model diagnostics determined if the heating and/or cooling coefficients should be dropped from the model. If cooling coefficients were retained, the account was considered weather-sensitive for the summer analysis. If heating coefficients were retained, the account was considered weather-sensitive for the non-summer analysis.

Load Variability

Accounts were also classified as high or low variability. Variability was assessed not in terms of how "flat" the load was across the day, but how much the load at a given hour varied from day to day. For loads that are more highly variable in this sense, any projection based on previous days is likely to have greater error. That is, baselines and corresponding demand reduction estimates for these accounts will be subject to greater uncertainty.

The account variability was measured in terms of the root-mean-square deviation of load in each hour from the corresponding mean for that hour, relative to the root-mean-square load during these hours. This statistic is similar to a coefficient of variation for load during peak hours.

Within each season and curtailment type (summer or non-summer, curtailed or uncurtailed) the cut-off between high and low variability was set so that approximately one-quarter of the accounts were in the high-variability group.

The numbers of accounts and median size by weather-sensitivity, variability, and sector if known are indicated in Table 4.

Table 4. Accounts Used in the Analysis

Weather-Sensitive	Variability	Number of Accounts: Non-Industrial	Industrial	Unknown	Total	Range of Median Load (kW)[a]
Yes	Low	62	77	123	262	495 – 2,381
Yes	High	7	19	20	46	276 – 1,416
No	Low	42	127	71	240	985 – 4,915
No	High	3	67	28	98	591 – 3,870

[a] Range of median loads across four groups: non-summer/summer by curtailed/uncurtailed

Performance Measures

The tests were run for several alternative methods on several different data sets for several hours on several days. Developing meaningful measures of method performance is essential to provide a basis for conclusions. Performance measures provided are somewhat different for uncurtailed and curtailed accounts.

The goal of the performance tests is to assess the accuracy of the various baseline methods tested. Accuracy has two aspects. One is lack of bias. Bias is a systematic tendency to over- or understate the baseline and the corresponding demand reduction. The second aspect of accuracy is variability. A method may be close to unbiased; that is, to be close to correct on average, yet have a high variance, meaning it tends to have large errors in either direction. Methods that have high variance are unreliable and add risk to program participation and operations.

The performance measures are based on the hourly error. For uncurtailed accounts, this is the difference between estimated and actual load. For curtailed accounts, where no actual uncurtailed load exists, it is necessary to choose one method to be the standard by which other methods are compared. Thus, for curtailed accounts, the error is difference between the estimated load and the chosen method estimate of load.

Because accounts in this study are of widely varying sizes, when looking at the range of errors across accounts, we need to normalize them. For uncurtailed accounts, the error for each hour for each account is expressed as a fraction of the actual load for that hour and account. For curtailed accounts, the error for each hour for each account is expressed as a fraction of the chosen method's estimated load for that hour and account.

Bias

As a key measure of bias, we focus on the median relative hourly error. If the median, across all accounts and curtailment hours, of these relative hourly errors is positive, then more often than not the baseline load shape will be overstated and the magnitude of DR is overstated. If the median of the relative hourly errors is negative, then more often than not the baseline is understated and the magnitude of DR is understated.

Overall Error Magnitude

As a key measure of the total magnitude of error, we consider Theil's U statistic for each account. This statistic is a "relative root-mean-square error." It is calculated for each account as the ratio of the root-mean-square error to the root-mean-square load.

The root-mean-square error is like a standard deviation, and represents the typical error magnitude for the account. This root-mean-square error reflects both systematic error, or bias, and the level of variability around the typical error.

The root-mean-square load is a corresponding "typical" load level. Normalizing the root-mean-square error by the root-mean-square load is something like calculating a correlation coefficient. It provides a normalized measure of variability, regardless of different load levels. However, the U statistic may be greater than 1, since errors can be greater than the loads they estimate.

Theil's U statistic calculated for a given account indicates the typical relative error magnitude for that account. The distribution of this statistic across accounts indicates the range of performance. We look at this distribution in terms of both the median and an extreme, the 95th percentile. The median Theil's U statistic indicates the typical relative error magnitude for a typical account. The 95th percentile indicates typical performance for the accounts where the performance is generally worse.

Results

The following graphs provide an example of the kinds of results generated in this analysis. The codes for the different combinations of model forms and selection rules are indicated in Tables 1 through 3 above.

As discussed, each data selection rule/estimation method combination can be adjusted. The six different adjustments are indicated by different symbols in the graph below (Figure 2). As indicated above, the temperature-humidity index-based adjustment is only appropriate for models without a weather variable in the estimation method.

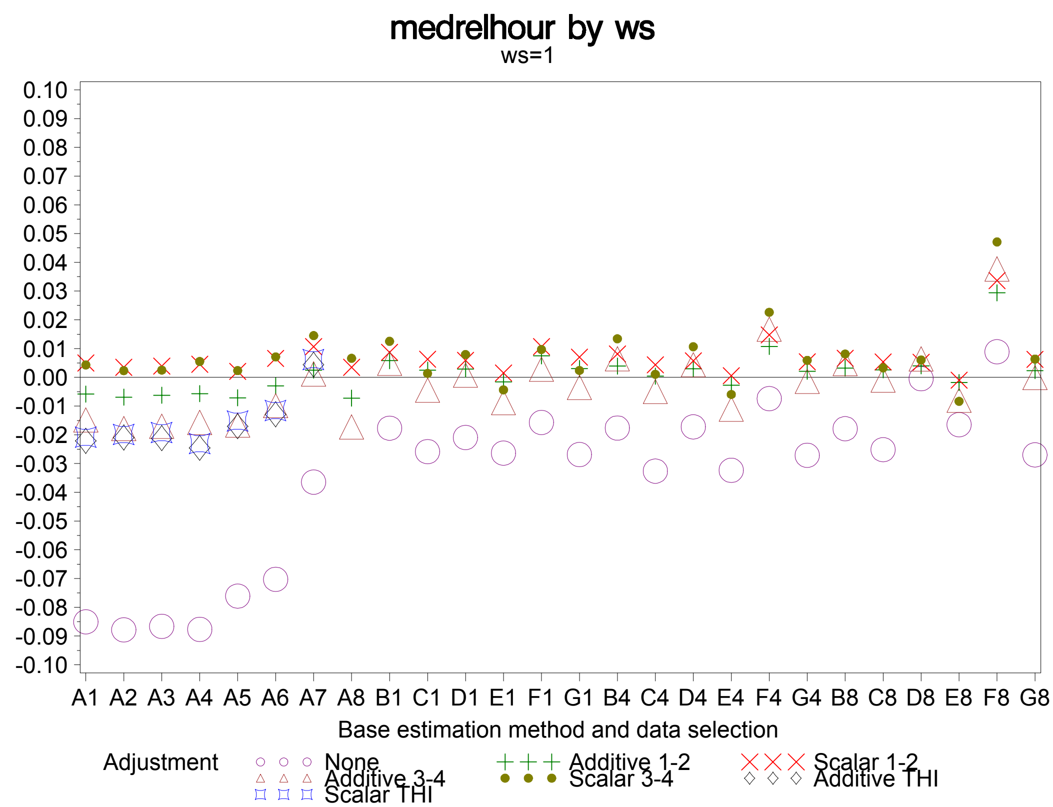

Figure 2. Median Relative Hourly Error Summer Uncurtailed Weather-Sensitive Low-Variability Accounts

Figure 2 provides the median bias results for the various protocols tested. These results apply to summer uncurtailed weather-sensitive low-variability accounts, one of the 16 different account types studied. The most striking result illustrated in this graph is downward bias of unadjusted averaging methods.

Figure 3 provides the Theil's U results for the same account type and in the same basic format. Theil's U statistic is a measure of variability, similar to a relative root-mean-squared error. These results indicate that adjustments not only generally decrease bias but also lower the variability as well. Conclusions from the test results across the various account types are summarized below.

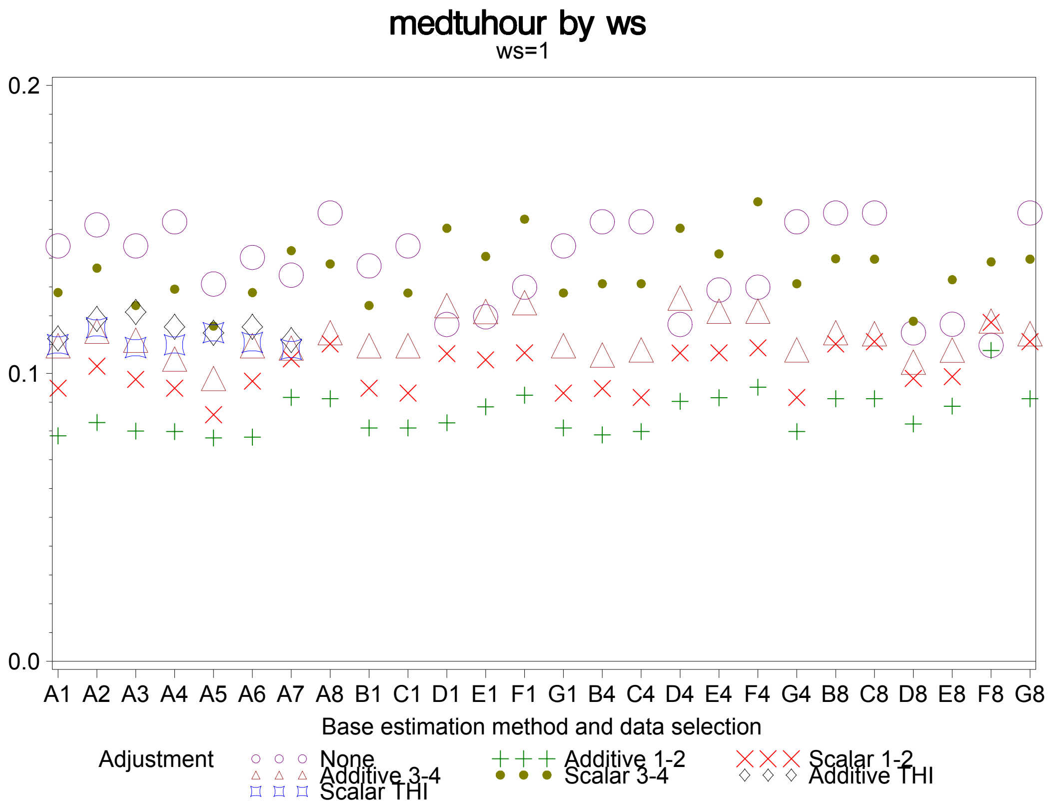

Figure 3. Median Account Theil's U Summer Uncurtailed Weather-Sensitive Low-Variability Accounts

Overall Results Summarized

The findings described below indicate the effects of various method features on bias and variability as measured in this study for the accounts and specific methods tested. These results offer general guidelines, but the performance of a particular method in a particular situation may be different.

Adjustments

An additive adjustment to 2 hours before curtailment can often reduce the bias and variability of almost all methods, including weather models, for weather-sensitive or non-weather-sensitive, high- or low-variability accounts. Other types of adjustments can improve the performance of averages, but generally with higher bias and variability.

With this additive adjustment, simple averages in most cases perform essentially as well as complex weather models, even for weather-sensitive accounts. Without adjustment, most averages tend to understate the baseline load and therefore understate the load reduction during the curtailment period.

On the other hand, additive adjustment to the last 2 hours can be problematic for several reasons:

- It opens the possibility of gaming by deliberately increasing load just before the curtailment period to boost the baseline.
- Legitimate pre-cooling in response to a curtailment notice or expectation will also erroneously increase the estimate of the baseline load.
- Conversely, an operation that achieves its curtailment target promptly upon notification and before the beginning of the required curtailment period will have a severely understated baseline.

A next step under consideration for this study is to examine the effectiveness and fairness of alternative rules for screening out both gaming and appropriate operational effects that result in distorted baselines.

Data Selection

The bias and variability of weather models tend to be reduced by longer input data series, but not dramatically. The decreased variability with longer input series is more noticeable for conditional weather models applied to non-weather-sensitive accounts, particularly high-variability accounts.

The different averaging methods performed similarly in terms of bias and variability, except for those that select a subset of days based on high load. For summer loads, the High 5 of 10 average generally reduces the otherwise negative bias. For summer loads using additive adjustment, High 5 of 10 gave the lowest bias measure of any of the averages, for both weather-sensitive and non-weather-sensitive accounts, and comparable variability. The High 10 of 11 average gave some bias reduction, but not as much.

For non-summer loads, however, the High 5 of 10 average inflates an already positive bias. The other averages perform better and roughly comparably to each other, in terms of both bias and variability, for both weather-sensitive and non-weather-sensitive accounts. The High 10 of 11 is somewhat better than the others in terms of the bias and variability measured in this study.

Weather Modeling

For summer weather-sensitive accounts, weather models tend to perform somewhat better than averages, but the difference is not dramatic. For non-summer loads, weather models do not perform better than averages.

For summer non-weather-sensitive accounts, use of a “conditional” weather model does not increase bias or variability. The conditional weather model automatically deletes weather terms if the statistical diagnostics based on the load data indicate these terms are inappropriate for a particular account. Use of such diagnostics protects against retaining terms in the model that are not well-determined and are likely not to be meaningful. Thus, if weather models are used, a single methodology can be applied to both weather-sensitive and non-weather-sensitive accounts.

Pros and Cons of Alternative Approaches

Advantages and disadvantages of key method features in terms of the criteria indicated in the Background Section are summarized in Table 5. This table is based on both qualitative considerations and the results of the performance tests.

Table 5. Advantages and Disadvantages of Key Baseline Method Features Based on Qualitative Considerations and Test Results

Baseline Method	Variant	Pros	Cons
Average	Any	Simple, easy to use and understand, low cost	Tends to understate baseline for weather-sensitive loads, especially if unadjusted
	High 5 of last 10 days	Partial adjustment for weather-sensitive loads	Still tends to understate baseline for weather-sensitive loads
			Can allow windfall load reduction credit on cool days
Regression	Any	Provides baseline corresponding to particular weather conditions of curtailment day	More complex, harder to understand, higher cost
			If observations don't include conditions as extreme as the curtailment day, model estimate may be inaccurate
			If account isn't weather-sensitive, may be less accurate than simpler methods
	Full Season	Adequate data and range of variation to yield accurate coefficients	Operating conditions from the period data are taken from may be different from curtailment day
	Recent 10 days	Operating conditions more likely to be similar to curtailment day	Model based on limited data may be inaccurate
	Lag temperature/ degree-day	Tends to reduce bias for weather-sensitive accounts	Tends to increase variability of baseline estimate.
	Conditional	Allows same general form and procedure to be used for weather-sensitive and non-weather-sensitive accounts, without pre-screening. Doesn't add much error for non-weather-sensitive accounts.	More complex. May give less consistent results across events for an account, if weather terms are sometimes retained and sometimes not.
Adjustment to precurtailment hours	Any	Simple, easy to use and understand, low cost	May be potential for gaming behavior during day-of-curtailment adjustment period
		Adjusts to weather and operating conditions of curtailment day	Appropriate pre-curtailment increase in load (e.g., pre-cooling) will result in overstated baseline
		Limits potential for collecting windfall credits for planned shut-downs	Pre-curtailment decrease in load in response to curtailment request (e.g., long ramp-down, canceling a shift) will result in understated baseline
	Additive	May adjust well for load change that is constant throughout day (e.g., industrial processes)	May not be appropriate if load changes during curtailment period (ratio adjustment may be better suited)
	Scalar	May adjust well for load change that is function of exogenous factor throughout day (e.g., higher levels of occupancy)	May not be appropriate if the day-to-day load variation is constant over the day (additive adjustment may be better suited)
	to last 2 hours before curtailment period	If load in these hours is unaffected by anticipated or initiated curtailment, provides best accuracy	If substantial curtailment is initiated in these hours, severely understates baselines
	to 3rd and 4th hour before curtailment period	Less potential for understated baseline due to pre-curtailment-period demand response	More variability than adjustment to last 2 hours
Weather-Based Adjustment	Any	Explicitly takes into account weather conditions	Adjustment may not be known to customer until after curtailment period (i.e., until after weather conditions are known for the day)
		No opportunity for gaming as with adjustment to precurtailment hours	If no observations are available for extreme conditions, estimates used for adjustment may be outside range of model
			Will badly predict load reductions if the buildings are dominated by internal loads
			Less accurate than alternative adjustments or weather model for both weather-sensitive and non-weather-sensitive accounts

Recommendations

In developing our recommendations, we did not attempt to score each method or feature with respect to each of the desirable features indicated above, nor assign explicit weights to the criteria. In general, our approach is:

- allow for options that recognize different customer or premise circumstances,
- favor simplicity if the potential accuracy gains of greater complexity appear to be slight, and
- indicate alternatives and trade-offs with respect to the criteria.

Proposed Approaches by Account Type

Offering Options. A general recommendation is that baseline calculation protocols should provide for alternatives based on customer load characteristics and operating practices. One way to simplify the provision of options is to establish a default method and allow certain deviations.

The basis for the selection of a method should be not just the customer's business type, but also the load patterns evident in the data as well as the customer's description of operating practices. Thus, for example, a customer who indicates a desire to be able to cancel a shift in advance of the control period should have access to a baseline calculation method that is not distorted by this practice.

At the same time, the program operator should have some discretion to bar customers from using an approach that they appear to have manipulated in the past. Thus, if there is evidence that a particular customer tends to inflate the baseline load after notification, beyond what would reasonably be expected for pre-cooling, that customer might not be able to use a method that includes adjustment to the 2 pre-curtailment hours.

A Practical Default Baseline Calculation Method. A method that generally works well for a range of account types is the simple average of the last 10 days, with additive adjustment to the load shape 2 hours prior to the curtailment period (in Figures 2 and 3, method A1 with additive 1-2 adjustment). This method can be recommended for both weather-sensitive and non-weather-sensitive accounts, with both low and high variability, for summer and non-summer curtailments. For most account types, the method gets high marks in terms of the criteria listed in the Background Section above. Table 6 describes in detail the practical default baseline calculation method.

Table 6. Step-by-Step Explanation of the Recommended Default Baseline[a]

Unadjusted Baseline Calculation	
Get interval load data for the previous 30 days. As little as two weeks of data may be needed, but this will depend on the number of curtailment exclusions.	L_{dh} = load at hour h of day d I = initial set of days for which data are taken
Remove weekends, holidays and previous curtailment days. Curtailment-day exclusions can be determined either by general program implementation or specific participation by the account in question. The latter is more flexible but also more complicated.	E = set of eligible days after these exclusions $E \subset I$
Keep the ten most recent days of the remaining interval data.	F = final set of 10 most recent days $F \subseteq E$
For each hour of the day, calculate the average load at that hour across the ten days. This is the unadjusted baseline.	$\bar{L}_h$ = unadjusted baseline for hour h $= (1/10)\Sigma_{d \in F} L_{dh}$

– continued –

Table 6 (cont). Step-by-Step Explanation of the Recommended Default Baseline[a]

Adjustment Calculation	
Get interval load data for the curtailment day.	L_{0h} = curtailment-day load at hour h
Based on the announced curtailment period, identify the last two, full hourly intervals immediately preceding curtailment start. These two hours will be the adjustment hours, hours 1 and 2 prior to the curtailment (e.g., hour ending 12 and 13 for a curtailment starting at 13:15).	h_0 = first hour of curtailment period h_{-1} = first hour preceding curtailment h_{-2} = second hour preceding curtailment
Average the curtailment-day load for the two adjustment hours.	$L_{-0\text{-}12} = (1/2)\,(L_{0h\text{-}1} + L_{0h\text{-}2})$
Average the unadjusted baseline load (from #4) for the two adjustment hours.	$\bar{L}_{-12} = (1/2)\left(\bar{L}_{h-1} + \bar{L}_{h-2}\right)$
Subtract the baseline average (#7) from the curtailment-day average (#6) to produce the additive adjustment increment.	$A = L_{-0\text{-}12} - \bar{L}_{-12}$
Adjusted Baseline Calculation	
Add the adjustment increment to the unadjusted baseline load at each hour. As an easy visual check, a properly adjusted baseline will cross the actual curtailment-day load between the two adjustment hours.	L_{Bh} = Baseline load at hour h $= \bar{L}_h + A$
Demand Response Calculation	
For each hour of the curtailment period, subtract the curtailment-day load from the adjusted baseline load.	DR_h = demand reduction at hour h $= L_{Bh} - L_{0h}$

[a] This explanation uses hourly intervals, but the same process can be applied at finer intervals.

This method is not recommended for accounts that tend to reduce load in advance of the required period in response to a curtailment notice. It is also not recommended for situations where the potential for gaming is a strong concern, whether across the program or for particular customers. Alternatives that can be used for different types of accounts are indicated in Table 7.

Table 7. Suggested Default and Alternative Methods for Estimating a Customer- and Event-Specific Baseline Energy-Use Profile, by Account Type and Season

Season:	**Summer**				**Nonsummer**			
Weather Sensitivity:	**Weather-Sensitive**		**Non-Weather-Sensitive**		**Weather-Sensitive**		**Non-Weather-Sensitive**	
Variability:	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**
DEFAULT: Simple average with additive adjustment to the 1st and 2nd hour prior to curtailment	X	X	X	X	X	X	X	X
ALTERNATIVES:								
Simple average with additive adjustment to the 3rd and 4th hour prior to curtailment			X	X			X	X
Weather model without adjustment, but with diagnostics determining whether heating and/or cooling terms are kept	X	X	X					
Use only the highest five of the last 10 days in the averages, with scalar adjustment based on a Temperature-Humidity Index load model.	X	X				X		
Use only the highest five of the last 10 days in the averages, without adjustment to the control day.					X			
Use the highest 10 of the last 11 days in the averages, without adjustment to the control day.			X	X			X	X

Additive adjustment: The adjusted baseline is calculated by adding a fixed amount A to the unadjusted baseline for each hour.

Scalar adjustment: The adjusted baseline is calculated by multiplying the unadjusted baseline for each hour by a fixed amount S.

Adjustment to k to k+1 hours before curtailment: The adder A or scalar S is calculated so that the adjusted baseline matches the average observed load over the period k to k+1 hours before the start of curtailment.

Weather model: Hourly load data from non-curtailed days is fit by a regression model using weather and calendar variables. The baseline is the fitted model applied to the observed conditions of the curtailment period.

Conclusion

A practical and credible baseline calculation method can improve participation and confidence in DR programs. To meet this objective, a baseline method needs to balance a number of practical considerations as well as prediction accuracy. Different methods are appropriate for different types of accounts, and according to the importance assigned to the different considerations. In developing options and recommendations, our approach has been

1. allow for options that recognize different circumstances, and
2. favor simplicity if the potential accuracy gains of greater complexity appear to be slight.

This work is intended to provide the foundation for a protocol that may be adopted as part of the IPMVP Protocol. An IPMVP Protocol, by nature, will offer options and guidance rather than be prescriptive. This first step toward a Protocol can serve as the basis for establishing specific rules and procedures within a jurisdiction. Equally important, this work can provide a common language for describing, debating, and understanding these procedures.

References

California Energy Commission. 2003. *Protocol Development for Demand Response Calculation — Findings and Recommendations*. 400-00-062: Sacramento, Calif. Prepared by KEMA-XENERGY Inc. Available at www.energy.ca.gov.

California Independent System Operator (ISO) Corporation. 2001. *Second Request for Bids to Provide Demand Relief (Load) for Summer 2001*. Available at http://www.caiso.com/docs/2001/03/2001033009195918940.pdf.

ISO-NE. 2002. *ISO-NE Load Response Program Manual (draft)*. 05-07-2002. Holyoke, Mass.

New York Independent System Operator (NYISO). 2001. *NYISO Day-Ahead Demand Response Program Manual*. Schenectady, N.Y.

New York Independent System Operator (NYISO). 2002. *NYISO Day-Ahead Demand Response Program Manual*. Schenectady, N.Y.

PJM. 2002a. *PJM 2001–2002 Load Response Pilot Program*. Valley Forge, Pa.

PJM. 2002b. *PJM Economic Load Response Program*. Fourth Revised Volume No.1. Valley Forge, Pa.