Developing Protocols to Estimate Load Impacts from Demand Response Programs and Cost-Effectiveness Methods -- Rulemaking Work in California

Dr. Daniel M. Violette, Summit Blue Consulting, Boulder, CO
Dr. David Hungerford, California Energy Commission, Sacramento, CA

Abstract

California is engaged in a process designed to produce a set of protocols for estimating the load impacts of different types of Demand Response (DR) programs. These DR impact estimation protocols will serve as a companion document to the California Energy Efficiency Evaluation Protocols adopted in April 2006. This work was initiated by CPUC Decision D.05-011-099 resulting in Rulemaking Docket R.07-01-041. When completed, the protocols will be used to guide the processes and efforts associated with estimating the load impacts of California's demand response programs.

This paper will focus on the issues and estimation approaches that were identified in the first draft Protocols document.[1] The issues were identified as needing resolution with regard to the application of different methods for estimating load impacts from DR programs in varied settings. An initial set of "Impact Issues" were defined reflecting choices that need to be made when selecting approaches for estimating the DR program load impacts. As part of this CPUC rulemaking, the investor owned utilities (IOUs) in California are to submit "straw" proposals for addressing the DR Impact Estimation Protocols by August 2007. Estimating DR impacts across different sectors and program types has raised new program evaluation questions. This paper presents the ongoing work and results being produced by this CPUC rulemaking effort.

I. Introduction

The Public Utilities Commission of the State of California (CPUC) issued Decision (D.) 05-011-009 which called for an assessment and protocols for the application of methods to estimate DR load impacts and DR program and customer costs. Specifically, ordering paragraph Number 8 calls for:

> "More precise demand reduction estimates derived from an accepted measurement methodology are a necessary prelude to performing accurate cost-effectiveness analysis… It is our belief that until the industry develops further trust that demand response will deliver demand reductions when needed, demand response will continue to be dismissed in the resource planning and acquisition process."

This decision called for an initial report addressing the scope and issues involved in developing protocols for estimating load impacts from DR programs. This paper focuses on summarizing issues identified in the initial report and suggested evaluation approaches that might be considered within the DR Estimation Protocols document. This paper is organized into five sections:

1. Dimensioning the DR Impact Estimation Landscape -- Types of DR Programs and Implications for Load Impact Estimation Methods;
2. Characterizing DR Impact Estimation Problem – Defining Appropriate Baselines;
3. Issues and Methods for Impact Estimation of Event-Based DR Programs;
4. Select "Other Issues" viewed as Key in the Estimation of DR Impacts; and,
5. Conclusions and Status of the Ongoing DR Estimation Protocol CPUC Rulemaking.

[1] "Draft Version 1 – Protocols for Estimating the Load Impacts From DR Programs," Prepared for the California Energy Commission Working Group 2, Measurement and Evaluation Committee, Summit Blue Consulting, LLC and Quantum Consulting, Inc. April 3, 2006.

II. Dimensioning the DR Estimation Landscape -- Types of DR Programs and Implications for Load Impact Estimation Methods

One of the first issues to be addressed was whether the protocols should address only retrospective impacts, i.e., indicate what load impacts were achieved for a given event or historical time period; or should they also be designed to forecast future load impacts for planning purposes? This choice could require different approaches and model specifications. Given the increasing importance of DR in resource planning, it is important that impact estimation approaches not only address ex post impacts, but also can be used to forecast load impacts within a resource planning context.[2]

Demand response programs, most broadly defined, apply rate design, incentives, and technology to enhance the ability of customers to change demand in response to prices and/or system conditions. Two types of demand response programs were described in CPUC D.01-05-056:

1. "Price-responsive" programs (in which customers choose how much load reduction they can provide based on either the electricity price or a per-kilowatt (kW) or kilowatt-hour (kWh) load reduction incentive); and
2. "Reliability-triggered" programs (in which customers agree to reduce their load to some contractually-determined level in exchange for an incentive, often a commodity price discount).

Another distinction made in describing programs is whether the program is "dispatchable" or "non-dispatchable." A dispatchable program is one where a system or control operator calls an event-day which triggers a program.[3] For example, in a simple interruptible program, a program operator can call for an agreed-upon number of megawatts (MWs) to be dropped by a program participant.

Price response programs can also be dispatchable. For example, a program operator can call for an "event-day" at, for example, 5:00 PM the previous day. This would trigger a critical peak pricing (CPP) event where very high prices would be seen by participants at peak period during that event-day.

A non-dispatchable program is one where there are no event-days and no event-day triggers are included in the design of the program. The DR program provides price signals during all hours and customers are able to make their own decisions regarding how much to use during a specific time period based on the price in that period independent of a program operator. Load reductions can also be non-dispatchable through the use of daily timers or regular schedule shifts in energy using equipment.

One challenge facing the development of impact estimation protocols is the diversity of program types to be addressed. Two program types discussed above are:

Type 1) Event-based programs. These are dispatchable programs such as callable load reductions or event-day price response programs (e.g., CPP), as described above; and,

Type 2) Nonevent-based programs. These are non-dispatchable DR programs that may include time-differentiated pricing such as day-ahead hourly pricing to regularly scheduled load reductions (e.g., irrigation customers with each skipping a different weekday to create diversity and reduce the co-incident peak irrigation demand for the group as a whole).

Candidate DR Load Impacts to be Estimated

This section is divided into two parts: Type 1) event-based, dispatchable programs; and Type 2) impacts for nonevent-based, non-dispatchable programs as defined above.

[2] A specific historical period may not include enough extreme peak demands (e.g., for hot-weather days) to forecast how a DR program might perform under those circumstances. As a result, it is important that evaluation methods include explanatory variables that can be used to build a base forecasting model of DR impacts across different system conditions.

[3] Event-days are usually triggered by a predetermined set of system or weather conditions. These triggers may be a temperature at a given location, a reduction in available operating reserves, or demand reaching a given level.

Load Impacts to be Estimated for Event-Based Programs. Candidate load estimates (i.e., MW) to be addressed by the estimation protocols are listed below:

1. Load impacts for each event for each hour in the event. From this average event-period estimates can be calculated as needed.
2. Average hourly impacts for each hour across all events. This would provide estimates of impacts for a program across a season. In a regression analysis, information could be included such as weather and possibly type of customer, or if a bid was put in by a customer such that there is a prior load reduction target, then a realization rate variant on the statistically adjusted engineering (SAE) method could be used.
3. Load impacts for events that are viewed as particularly extreme. This could be important for forecasting impacts for event-days that are beyond the range of factors recently seen (e.g., weather or system contingencies). Impacts will vary across events, even for the same program. For example, an AC load control program will typically provide greater impacts on hotter days. Also, a DR program may provide a different impact on the highest system cost or resource-constrained day than during another event -- which can affect the DR program's capacity rating.
4. Load impacts outside of the event period. Impacts have been seen both in anticipation of the start of an event period and after the event period. The post-event period impacts can show an increase (e.g., snapback in the case of AC loads) or a decrease, as the effects of the curtailment seem to linger for some participants. In looking at pre- and post-event period impacts, consideration should be given to the magnitude of the impacts in kW and over which hours they occur (pre- or post-event).

In addition to estimating the MWs associated with a specific event, there may be other factors which need to be addressed in the estimation framework:

1. Factors influencing load impacts within events (heat storms, number of interruptions, contiguous events across multiple days, and variability in baseline loads) that may need to be controlled for in the estimation framework.
2. Different time dimensions in estimating load impacts can be important. If the program design calls for impacts to be delivered within a specified time period (e.g., a 5-minute notice), it may be relevant to estimate how many MW were delivered within a specified time limit, 2 minutes or 5 minutes or 10 minutes.
3. DR program impacts by location and customer type. This can provide input into program design, as well as assess the contribution of the DR program to alleviating system congestion.
4. Synergies across programs can be important.[4] Portfolio benefit-cost analyses requiring information on synergies (both positive and negative) in impacts across programs may need to be assessed.

Load Impacts to be Estimated for Nonevent-Day Programs. Candidate MW impacts to be addressed in the protocols for nonevent-day, non-dispatchable programs include:

1. Load impacts for scheduled load-response programs and every-day pricing programs are needed. Impact estimation for scheduled load response should be more straightforward since these programs are usually designed for regular, every-day loads (e.g., pumping), and an examination of load data on scheduled curtailment days compared to non-curtailment days should provide adequate estimates. Impacts for nonevent-based pricing programs are the most problematic of the candidate impacts to measure. For these programs, the impacts can occur in every hour of every day, and the impacts may also vary by day-type (Monday through Friday, weekends/holidays), by weather condition, and by lagged weather conditions (if load is weather dependent). Methods that control for these variables are needed for unbiased estimates.

[4] See Violette, D., R. Freeman, and C. Neil, 2006.

2. Load impacts for "stress" days[5] which could be defined by high market prices or could use the same event-day criteria as the event-based programs. This would give a point of comparison between the nonevent-day programs and the event-day programs in terms of contribution.
3. Estimation of load impacts on "extreme" days. Among the stress days, some days are much more extreme than others. These can be used to measure the true value of DR to meet contingencies that represent 1-in-10 or 1-in-20 planning criteria.

Many of the same "other" factors discussed in the section on event-based programs will also be important to include in the estimation framework for nonevent-based programs.

III. Characterizing the DR Impact Estimation Problem

Demand response impacts are "difference estimates," i.e., the difference between the observed load and the higher baseline that would have occurred if the DR program had not been offered. Since the actual load is known, the key challenge is in estimating the baseline (see Figure 1 for an example).

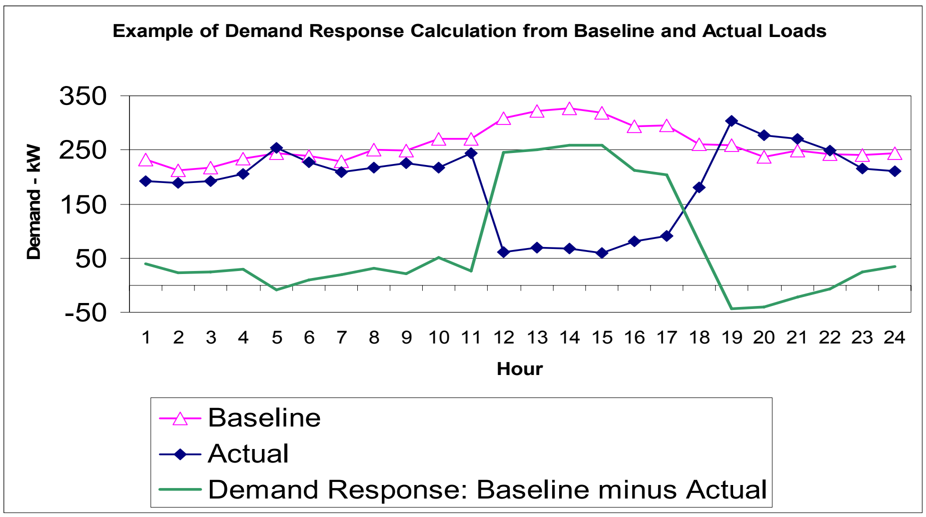

Figure 1. Demand Response Impact Estimation -- Baseline minus Actual

Actual consumption during an event-day or program period can be measured; as a result, the focus of the impact estimation problem is the selection of the most accurate baseline. A typology of approaches for estimating customer-specific DR impacts for program settlements is presented in the CEC Baselines Report (2003). Across candidate baselines, two types of baseline methods were viewed as most viable in the CEC Baselines Report and subsumed many elements of other approaches. These two methods are:

1. Estimation by Averaging – This approach is termed the "representative-day" approach in the California Energy Commission Working Group 2 (WG2) evaluation reports.[6] The representative-day approach calculates baselines for each event based on a series of recent "similar" days that are proximate to the event-day in question. This could be the average of the 10 preceding weed days, or some different combination of the hourly load shapes in a number of preceding days. A variety of approaches are analyzed in the CEC WG2 evaluations of the 2005 and 2004 DR programs and in the CEC Baselines Report (2003). In addition to different combinations of prior days, different "same-day" adjustments can be used to better match the baseline to the actual customer loads, for example:

[5] Since there are no event-days, the term "stress" day refers to a day "similar" to what would be an event-day for a callable program in terms of weather and/or system conditions.
[6] Publications by Working Group 2 are available on the California Energy Commission website under Demand Response.

- An additive same-day adjustment where the baseline load shape is moved up or down to match that customer's event-day loads in the hours preceding the event period; or,
- A weather same-day adjustment where a regression equation is used to estimate the hourly load relationship for a temperature/humidity index (THI) to adjust the customer's load for differences between the baseline and that day's weather.[7]

2. Estimation by Regression – This approach was used in the recent CEC Working Group 2 impact assessment[8] and in the Working Group 3 (WG3) assessment of DR programs in mass markets.9 Multivariate statistical regression models are used to estimate program impacts with as many days as are allowed to enter the model (e.g., a regression model can handle a full season of load data), plus a series of weather, customer, and other characteristics. Same-day adjustments can be made to regression baseline estimates in the same manner as used for the representative-day approaches. These adjustment approaches are explained in the CEC Baselines Study (2003).

Both approaches are used in the CEC Working Group 2 (2004and 2005) DR evaluation reports with a wide variety of representative-day methods tested as well as a broader set of multivariate regression approaches. One difference between the CEC Baselines Study (2003) and the latest load impact estimation protocols (2006) is that the protocols are focused on estimating program-wide impacts that can be used in benefit-cost analyses and resource planning, rather than on individual customer level impacts. Estimates of individual customer contributions are likely to be important for program design and monetary settlements, i.e., is the program appropriately paying a customer for the MWs delivered.

Settlements may be based on a different calculation method than estimates of "actual" load reductions to make payments transparent and timely. The estimation of actual load reductions for benefit-cost analyses do not have to meet the same time and transparency constraints as do approaches for calculating settlements. As a result, the need to meet customer service criteria may result in the program administrator paying for more or less MWs than those estimated to be actually delivered using a more detailed estimation method as part of the longer-term program planning cycle.

Candidate estimation methods shown in this section take advantage of the fact that many DR programs often have information on the entire population of participants so there may be initial estimates of the magnitude of a customer's response to a called event. For example, if the program is a bidding program, there is information on the amount of MWs each customer has bid. For some interruptible programs, the customer may have estimated what their load reduction would be if called, or has committed to a known firm service level. These initial impact estimates can be used in the analysis to leverage data using ratio and difference estimation approaches that produce realization rates, i.e., how much of the expected load reduction is actually realized given the observed load data.

IV. Issues and Impact Estimation Methods for Event-Based Programs

Event-based, dispatchable programs comprise the largest fraction of DR programs being implemented in California and are the focus of this section. Selected types of event-based programs are:

- Event-Based Pricing (large and mass-market customers) -- Called event-day pricing or CPP tariff programs

[7] CEC Baselines study (2003) on page 2-12 states: "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 ratio of these two estimates is calculated, and applied to the provisional baseline as an additive or scalar adjustment."

[8] "Working Group 2 Demand Response Program Evaluation – Program Year 2005", prepared for the Working Group 2, Measurement and Evaluation Committee, by Quantum Consulting, Inc. and Summit Blue Consulting, LLC, in progress, expected May 2006.

[9] "Impact Evaluation of the California Statewide Pricing Pilot," prepared for Working Group 3, by Charles River Associates. On-line: March 24, 2005.

- Large Customer Event-Based Programs -- Interruptible, bidding and other
- Mass Market Event-Based Programs -- Direct equipment control, indirect equipment control (e.g., temperature setback); and mass-market bidding (via aggregators)

For the large customer programs, the number of participating customers is often a small enough set that it is possible to work with information on the population. For mass-market programs, the number of customers can easily be in the tens or hundreds of thousands requiring that samples be used.[10]

The issue in both cases is developing an appropriate baseline for the event period against which the actual event period consumption can be compared. In EE evaluation, it is not uncommon for non-participants to be used as a control group for setting baselines. However, impact evaluations of DR event-based programs have generally used the participants themselves for the construction of the baseline. The basis for the assumption is that DR program participants themselves are the best control group, i.e., they are perfectly matched. Long-term trends in the market which would be captured by a non-participant control group are not as relevant for DR programs as they are for EE programs. It is difficult to develop a control group, particularly for large customer DR programs, that actually provides a reasonable match to a group of program participants. In addition, the general availability of hourly data provides numerous observations on participants' energy use across different hours, different days, and under different weather conditions.

This defines another issue for the Protocols and that is whether a control group comprised only of program participants is sufficient in many evaluations, i.e., a non-participant control group is not needed for estimating event-day load impacts and may even confound the results. In the case of most large customer programs, customer-specific baselines are constructed using either representative-day methods or regression methods based on that customer's consumption in nonevent-day hours. The availability of hourly data generally provides for an adequate set of data from which to estimate baselines for the individual participants in a DR program.

Estimation issues and methods are discussed using the two different approaches for estimating baselines discussed above – 1) the representative-day approach; and 2) the use of regression methods.

Issues and Methods using Representative-day Baselines for Estimating DR Impacts

Representative-day baselines for use in estimating load reductions have been assessed in a number of recent studies. These include the CEC Baselines studies (2003) and the Working Group 2 (WG2) evaluations of DR program for program years 2004 and 2005. Both sets of documents consider a variety of methods for developing the representative-day baseline. Two factors typically define the baseline:

1. The number of days immediately prior to the event-day whose load data are averaged to create a baseline for the event-day. CEC Baselines Report (2003) considered baselines that included as many as 20 prior days. The WG2 study efforts limited the prior day analysis to 10 days.
2. Event-day adjustments that can be made to better fit the baseline to the specific event-day. These adjustments are designed to match the baseline load shape to the same-day loads for one or two hours just prior to the start of an event period.

There are many methods to estimate representative-day variants that can be considered. Of the many considered in prior studies, two examples are given below. The first method (the 3 highest out of 10-day method) has been used for settlements at each of the three IOUs for the Demand Bidding Program (DBP). The second method is a 10-day average with same-day adjustment. This second method was the recommended default method in the CEC Baselines Report (with some caveats), and it was found to be the most accurate representative-day method in the WG2 DR Evaluations.

[10] Sampling is discussed in the initial draft of protocol issues (CEC, 2006) in Section 6. It is not discussed here due to space limitations.

1. 3 highest out of 10 day method – This method is calculated by first selecting a series of the most recent 10 days (excluding weekends, holidays, or other event-days). From these 10 days, the three days with the highest overall load during the event hours is selected, and the loads for each of these three days is averaged (by hour) to calculate an hourly 3-day baseline estimate.
2. 10-day adjusted method – This representative-day method simply uses the 10 most recent days (excluding weekends, holidays, or other event-days) to construct an average baseline; then this baseline is adjusted by a scalar factor determined by a series of calibration hours, e.g., the 3 or 4 hours immediately preceding the event. (See WG2 Evaluation reports for numeric example.)

Both the WG2 DR Evaluation and the CEC Baselines Report found that in general the 10-day adjusted baseline was the most robust choice of "a practical default baseline" from among the representative-day approaches considered. It is important to note that the CEC Baselines Report focused on methods for settlements. As a result, issues important for settlements with customers that factored into the recommendation included: 1) the ability to know the baseline prior to the customer making a curtailment decision; 2) the ability of the customer to know the baseline immediately after an event; 3) simplicity of calculation; and 4) minimizing burden on participants and operators (among others). Importantly, when looking at program-wide evaluation for use in resource assessments, different criteria may be needed or given more weight in developing recommendations. In particular, program-wide evaluations for use in planning likely will give greater weight to increased precision in the program's estimated resource contribution (i.e., load reduction).

Several issues from among those raised in the Draft Protocols Report related to the use of representative-day baseline approach are discussed below.

Issue 1 – Same Day Adjustments in Representative-Day Baseline Approaches. Making adjustments for same-day effects in the event-day baseline have been common in many baseline approaches. Several issues need to be considered when assessing the effects of same-day adjustments. Among these are:

1. Recent work by the WG2 measurement and evaluation committee has shown that some customers begin to curtail load prior to the actual start of the called event. Thus, if these 2 hours are used to calibrate the baseline load shape, the baseline would be inappropriately adjusted downward and the impacts underestimated.
2. The CEC Baselines Report also raised the issue of gaming where a customer might try to increase their loads in the preceding two hours to get an upward adjustment in their baseline for that day; or, since customers know their baselines in advance of the control day, they may submit a curtailment bid when they know their normally planned consumption would be less than the calculated baseline used for settlements. Gaming by customers can result in settlement values that overestimate actual DR program impacts.

It has been shown that some customers, in order to meet their load reduction obligations, may need to lower their consumption in the several hours preceding an event. Gaming with respect to the same-day adjustment has not been shown to be a substantive problem. However, the WG2 2005 analysis showed that some customers in the demand bidding program appear to submit bids when they know that their normal consumption on an event-day will be lower than the settlements baseline. As a result, there may be customer-specific circumstances (e.g., batch processing or scheduled loads that vary by day) which can make the representative-day approach inappropriate for that customer. Any average of the preceding days would not fit the specific circumstance for that event-day. In these cases, there should be some flexibility in determining appropriate baselines for customers with these load types, i.e., different baseline methods may be appropriate for different customers and different event-days.

Issue 2 -- Representative-Day Baseline Approaches for High-Variance Customers. Representative-day approaches which take an average of a set of prior days' loads will work better for customers whose load shapes are more consistent from day to day. The CEC Baselines Report separated customers into two groups – a low-variance group and a high-variance group. The report clearly shows that the representative-day approach is less accurate for the high-variance group, and even less accurate for extreme cases. (CEC Baselines Report, 2003). The CEC WG2 Evaluation Report (May 2006) showed even greater discrepancy between the low- and high-variance day-to-day load shape groups.

The CEC WG2 Evaluation Reports illustrate that there are some customers that had extremely variable day-to-day loads such that any baseline using an average of the loads from the preceding days likely will not be a "representative" baseline for a single, specific day; and second, the calculation of a 10-day baseline of "similar" days (i.e., excluding weekends, holidays and other event-days) sometimes required going back more than a month to find 10 similar days which resulted in the baseline for a specific event day being less representative.

This issue is illustrated in Figure 2, taken from the WG2 Evaluation Report (May 2006). It provides an example of a large customer with high day-to-day load variability, i.e., whose load shape varies widely during the previous 10 "similar" days. On some days this customer's usage remains fairly consistent across the whole day at around 1.5 MWs. On other days the customer's load jumps up to levels closer to 3.5 MW, and on still other days it fluctuates over the course of the day between 1.5 and 3.5 MW. This customer bid 2 MW/hour over the event period and it appears that the customer's true program impact should be close to this hourly bid amount, although the 10-Day Adjusted Representative-day approach calculates an impact that is only half of the customer's bid amount.

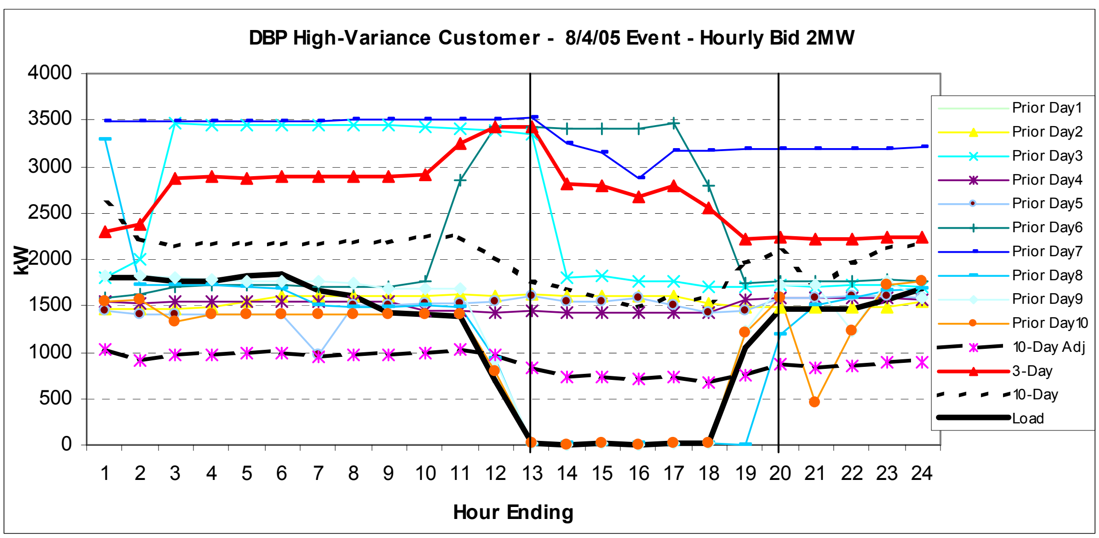

Figure 2. Daily load shapes associated with a single "high variance" customer for alternate baseline methods.

Figure 2 illustrates that some customers have such extreme day-to-day load variability that they may require methods other than the whole-facility, representative-day baseline approach: For some customers, the representative-day approach simply does not make sense. If that customer's day-to-day load shape (as shown for days 1 through 10 in Figure 2, above) varies dramatically, no averaging across those days will produce a reasonable baseline for a specified event-day. For these customers, the regression approach (discussed below) seems to work better, but still may not provide reliable estimates. Alternative, customer-specific methods for estimating load impacts may be needed for customers with high variability in day-to-day loads. These methods may involve site data collection monitoring for event-day calculations. Alternatively, it may be preferable for such customers to participate in other programs such as pricing programs (e.g., CPP) that do not need a calculated baseline for settlement.

Issue 3: Representative-day Baseline Approaches for the Largest Program Contributors. For some programs, a few large customers may account for a large fraction of the total program impacts. Figure 3, below, shows that 5% of the participants contributed about 50% of all program impact.

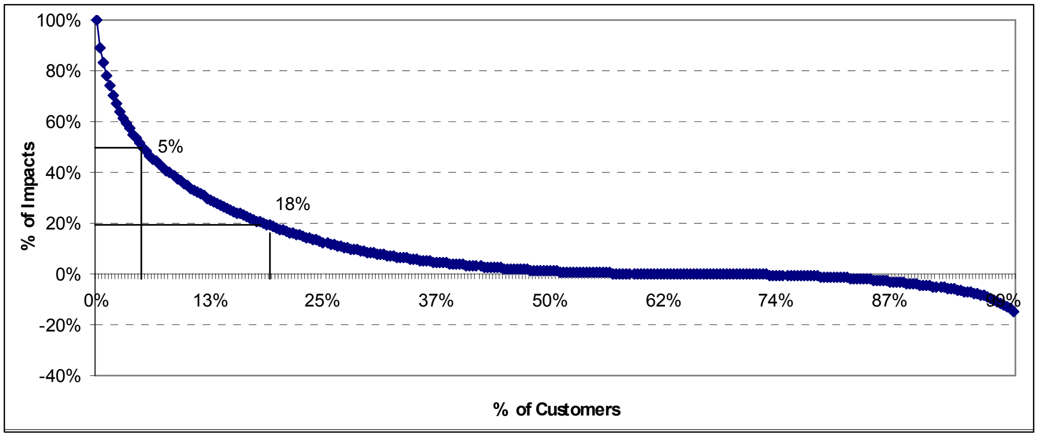

Figure 3. From CEC WG2 2005 DR Evaluation Report (May, 2006)

Accurate estimation of DR program impacts may warrant more detailed follow-up data collection on these high-impact customers to validate their large contribution to program impacts. A focus on the highest-impact customers will produce higher levels of resolution for both customer and program impacts. Also, some lessons learned from the high-impact customers may be applicable to other customers, resulting in an increase in overall program load reductions.

Regression Approaches for Estimating DR Impacts

Two basic approaches are addressed in the CEC Baselines Report (2003): representative-day approaches (discussed in the section above) and regression approaches. The representative-day discussion helps frame the discussion of regression approaches. In addition, the representative-day approach is not addressed in the California Energy Efficiency Evaluation Protocols (2006) where regression approaches are given extensive treatment. As a result, the discussion of issues related to the use of regression methods is shorter due to the treatment of typical regression issues in both the EE evaluation protocols and in the California evaluation framework report.

A regression-estimation method for DR baseline estimation involves the modeling of factors that influence a customer's hourly electricity demand. This can include factors that are not explicitly accounted for in the representative-day approach. In simple terms, a regression model determines the relationship between some observable factors and load. Using this relationship, load can be estimated for any scenario based on the levels of those factors for that scenario.

Regression approaches also have disadvantages. Regression estimation is a more complicated technique than simply averaging to get representative-day baselines. Ease of calculation and transparency are both important considerations for baselines used to calculate settlements, but they may be less appropriate for season-ending assessments of DR program impacts for use in benefit-cost and cost-effectiveness assessments. WG2 Evaluation of the 2005 DR programs (May 2006) contains a more thorough discussion of the use of regression approaches for estimating the impacts of DR programs.

Regression models can be used as an alternative to the representative-day approach, in which case one of the keys to the specification of the regression model is the number of days to be used in the regression analysis. If the representative-day approach uses the 10 prior "similar" days (excluding holidays, weekends, and other event-days), then the regression approach can use the same 10 days. If the 10 days are in close proximity to the event-day, then operating and weather conditions may be similar to the event-day. However, the WG2 2004 evaluation of DR programs showed that the 10-day approach can require going back nearly a month to get the 10 "similar" days. So, the use of 10 prior "similar" days may have some limitations. In addition, the availability of data just after the event-day may be just as important as the prior similar days.

Regressions based on a full season of data have been shown to produce more accurate estimates of seasonal impacts, but these estimates have to wait until a full season of data become available. For settlements it may be unacceptable to wait for a full season of data; but for estimates of accurate program-wide impact of DR, it may be appropriate to use a full season of data.

In addition to serving as an alternative to the representative-day approach for estimating the impacts of DR programs, baselines constructed using regression methods can also have "same-day" adjustments, just like the representative-day approaches. The CEC Baselines Report illustrates both additive and scalar adjustments to regression-based baseline estimates (see Figure 2-6, p. 2-13). This same report determined full season simple regression models (using weather data only) and generally performed better than the 10-day average representative-day approach.

The WG2 2005 DR Evaluation Report (May 2006) discusses the impact estimates developed for the utilities' critical peak pricing (CPP) program and the demand bidding program (DBP) using a full-season regression model to estimate baselines. The regression models for the pricing programs used time-series equations that were specific to each customer in a program.[11]

V. Select "Other Issues" in DR Impact Estimation

The issues addressed in this paper are only a subset of the issues raised in the initial set of load impact estimation protocols, and this paper does not address any of the cost-side estimation issues. Other issues that arose concerned:

- Treatment of free riders.
- Treatment of spillover. Is there the potential for spillover effects where there may be program-induced load reduction impacts that are not fully captured by the program accounting and load impact estimation method?
- Use of post-event-day data in the construction of baselines and how this can be used in both settlements calculations as well as in program-wide resource estimates.
- Method for calculating the statistical accuracy of representative-day approaches. This could be important as the standard approach is to select or develop one baseline (whose accuracy is unknown) and subtract the customers' actual load for that day. This method, on its own, produces no statistical information, i.e., there are no estimates of variance for any customer's impact calculation because only one estimate (i.e., one baseline) of unknown accuracy is obtained. This problem is not solved by going to large groups of participants. If data on 100 customers are available, you could calculate a variance for the impacts across the 100 customers given the specific baseline method used (one data point for each customer), but each customer still does not have any statistical information that can be used to inform estimates of program-wide impacts in terms of a confidence level. There are ways around this problem but depending upon the level of data work for each customer, some stretch assumptions may be required. The type of data analyses for each customer that would lead to estimates of accuracy at both the customer and program level might involve Monte Carlo analyses

[11] While the WG2 impact evaluation of DR programs used separate equations for each participating customer, it is likely that other schemes that employ some degree of pooling of customers into a single regression model will likely be appropriate. The "fixed effects" approach is becoming a commonly used form of pooling when the time periods are seasonal. This approach is discussed in the California Evaluation Framework (2004) and CEC WG2 2005 Evaluation Report (2006).

around potential baselines for each customer. This would produce a distribution of outcomes for that customer or the development of a set of estimates for each customer using event-like days that were not called as events to estimate the accuracy for that customer using a specific baseline method compared to the known baseline. This latter approach is useful in that all the uncertainty is contained in developing the estimates of the baseline. Actual consumption on an event-day can be measured. This problem is not as severe for regression approaches. As long as the regression model assumptions hold, the regression approach produces estimates with a variance allowing confidence intervals to be constructed.

VI. Conclusions and Ongoing Work on the California Protocols Rulemaking

The issues identified in the first document for the work towards DR Impact Estimation Protocols (CEC WG2, 2006) and the suggested DR impact methods being developed as part of the DR impact estimation protocols in California are addressing some of the difficult problems faced by evaluators of DR programs. The research already conducted and expected throughout 2007 should be transferable to most DR program types used in other regions.

It is important to recognize that there are two different uses of impact estimation. The first is the use of impact estimates for financial settlements with customers who participate in the DR program. Any settlements method needs to be transparent and easily calculated within a reasonable timeframe. Settlements typically cannot wait until data for an entire season is collected. The settlement provisions are part of the agreement made between the program administrator and the customer for payments based on a calculated amount of peak load reduction or demand response to a triggered event-day.

The second use of DR impact estimates is for application in benefit-cost analyses and resource planning. For these applications, it may be possible to use more complex methods and longer time frames for estimation than is reasonable for a customer settlements approach. This could include regression equations based on data for the entire season.[12] Estimating actual impacts for benefit-cost tests and resource planning can incorporate more factors than might be reasonable for a settlements calculation.[13]

With respect to CPUC rulemaking R.07-01-041, workshops on issues were held May 3 through 4, 2007, and a CPUC/CEC Staff Guidance Document was prepared on May 24, 2007 providing direction to parties planning to submit straw proposals for consideration. The CPUC requested that the joint utilities (PG&E, SCE, and SDG&E) should submit a single straw proposal. This joint utility straw proposal, along with efforts submitted by other parties, will be the basis of workshops held this fall. A proposed decision is currently targeted for February 2008. This proceeding can be followed through the CPUC website. The link for this rulemaking is found at http://www.cpuc.ca.gov/static/hottopics/ 1energy/_drce. htm. At the time of this draft, the next upcoming workshop on the Phase 1 load impact and estimation protocols and cost-effectiveness methods are scheduled for May 3 through 4, 2007.

References

Goldberg, M. and G. Agnew 2003. "*Protocol Development For Demand Response Calculation — Findings and Recommendations.*" California Energy Commission, Report CEC 400-02-017F, February.

[12] The CEC Baselines Report showed this to be the most accurate baseline estimation method, but it was not recommended for settlements due to the lag time that would have to occur between the time of payment to the customer and the calculation of impacts based on trend analysis across an entire season of data.
[13] Over time, it is hoped that enough could be learned such that the two approaches will tend to converge.

California Public Utilities Commission 2004. "*The California Evaluation Framework.*" TecMarket Works and Team Members, Project #: K2033910, June.

California Public Utilities Commission Energy Division. 2006. *"California Energy Efficiency Evaluation Protocols: Technical, Methodological and Reporting Requirements for Evaluation Professionals"* by The TecMarket Works Team, April.

Quantum Consulting Inc. and Summit Blue Consulting, LLC 2006. "*Evaluation of 2005 Statewide Large Nonresidential Day-Ahead and Reliability Demand Response Programs*," California Energy Commission Working Group 2 Measurement and Evaluation Committee, May 1.

Quantum Consulting, Inc. and Summit Blue Consulting, LLC, (2004) "*Working Group 2 Demand Response Program Evaluation – Program Year 2004.*" California Energy Commission Working Group 2 Measurement and Evaluation Committee, 2004.

Summit Blue Consulting, LLC and Quantum Consulting, Inc. 2006. "*Draft Version 1 -- Protocols for Estimating the Load Impacts from DR Programs.*", California Energy Commission Working Group 2 Measurement and Evaluation Committee, April 3.

Summit Blue Consulting, LLC. 2005. "*Demand Response Education Program Evaluation for 2005*," California Energy Commission Working Group 2 Measurement and Evaluation Committee, May 26.

Violette, D., R. Freeman, and C. Neil 2006. "*DR Valuation and Market Analysis -- Volume II: Assessing the DR Benefits and Costs*," International Energy Agency Demand-Side Programme, Task XIII. Summit Blue Consulting, January 6.