FEDERAL ENERGY REGULATORY COMMISSION

Demand Response in Wholesale Markets **Docket No. AD07-11-000**
Technical Conference
April 23, 2007

Part 3
Measurement and Evaluation of Demand Response Resources

Miriam L. Goldberg
KEMA Inc.
Panelist

Panelists have been asked to address the following questions.

1. What are the current best practices to measure, verify, evaluate and forecast demand resources?
2. What is the latest research and opinion on the firmness, sustainability and reliability of demand resources; measurement and verification (M&V) protocols; customer load forecasts; customer baseline estimation; and the potential for free-ridership?
3. Are the approaches and principles that have been used historically to measure and verify savings from energy efficiency applicable to wholesale demand response resources? If they are not, what changes are needed to make them applicable?
4. If demand resources are providing capacity resources or serving as an alternative or complement to transmission expansion, what are the key factors that need to measured and verified, and at what level of precision? Can the current form and precision of measurement and verification approaches be utilized or are new approaches needed?
5. Should there be uniformity on how demand resources are measured and verified across markets?

KEMA has conducted an extensive investigation of customer baseline methods for demand response (DR). The discussion below emphasizes this issue, and also touches on some key points related to the other questions.

1. Best Practices

What are the current best practices to measure, verify, evaluate and forecast demand resources?

General Issues

Measurement and verification, evaluation, and forecasting are distinct functions. Best practices for these different functions are related, but distinct. In addition, the best practices depend on the types of demand resources.

The International Performance Measurement & Verification Protocols (IPMVP, Efficiency Valuation Organization) is referenced as a standard in many contexts. However, it is important to recognize that these protocols were not designed as a set of prescriptive standards. Rather, they were designed as a set of broad guidelines and practical approaches for creating mutually acceptable verification terms for performance contracts. Thus, by itself a requirement to follow the IPMVP does not specify methods or accuracy standards.

Baseline Estimation is critical for calculation of both energy efficiency and demand resources. Meaningful definition and calculation of savings or demand reduction has to start with agreement on “compared to what.” In principle, the desired baseline is what would have occurred in the absence of the program or program action. If this hypothetical no-program condition could be known perfectly for each customer or resource, savings would simply be the difference between the observed usage and the baseline. This calculation would capture all program effects and attribution effects (free ridership, free drivers, spillover, market effects).

In practice, baselines are usually defined by standards or conventions. Attribution effects are then separately addressed.

Attribution is the extent to which savings or demand reduction can be attributed to the program. Attribution analysis adjusts savings relative to the prescribed baseline downward for free ridership (savings that would have occurred without the program) and upward for free drivers and spillover (savings that occurred because of the program but are not tracked by it.

Some stakeholder argue that adjustments for program attribution are unnecessary. One argument is that the positive effects of free drivers/spillover more than outweighs the negative effects of free ridership. While this relationship can be true in some contexts, in others it is demonstrably not true. These effects depend on the particular conditions of a market, program offering, and delivery mechanism. The effects change over time as markets evolve. Investigation and analysis are needed to assess attribution effects for the particular conditions of a program.

A second argument for ignoring attribution effects is that the savings are present in the electric system regardless of whether the end-use customer would have paid for them without the program. However, this argument implicitly assumes that the savings that are relevant to the system are the savings relative to the prescribed baseline. If ratepayer money is used to pay for the reductions, the reductions must be determined relative to the load the system would have seen if that money had not been spent this way. If the customer’s load would have been the same with or without the program incentive, there are no savings to the system.

Baselines for demand response depend on the demand response approach.
In rate-based approaches, the end-use customers pay for what they use, at rates that may vary by time of day or critical conditions. M&V for settlement with the end-use customer requires only the observed load, not an estimate of load without the DR "event." However, evaluation of the effect of the program or rate still requires estimation of what the load would have been in the absence of the alternative rate structure or specific DR event. This analysis may be done at an aggregate level, for example via estimation of price elasticities, rather than at the level of individual customers. Examples of such programs include Time-of-Use (TOU), Critical-Peak Pricing (CPP), Real-Time Pricing (RTP), and firm load programs.

In demand reduction approaches, the end-user receives incentives for demand reduction relative to a defined baseline. M&V for settlement with end use customer requires the baseline calculation according to a pre-agreed procedure. Evaluation of overall program effectiveness may use alternative baseline calculation methods.

KEMA Study on Baseline Protocols

KEMA conducted a study on DR baseline protocols for the California Energy Commission (KEMA-XENERGY, 2003). This work included interviews with stakeholders on the development and desired features of baseline methods, as well as technical performance assessment of a large number of methods using load data from utilities across the United States. Key findings and recommendations from that study are summarized here.

DR baselines goals and criteria

Common goals described for DR baseline protocols include:

- Reflection of load that would have been used absent the program
- Ease of use for program participants
- Ease of use for program administrators
- Deterrence of gaming.

Critera that must be balanced in developing a baseline are:

- ***Simplicity***, including ease of use, ease of understanding, and low costs for participant and operator to implement
- ***Accuracy***, including lack of bias (i.e., no systematic tendency to over- or under-state reductions), appropriate handling of weather-sensitive accounts, and verifiability
- ***Minimization of gaming*** by customers
- ***Predictability***, the ability for customers to know the baseline before committing to a particular curtailment amount and event
- ***Consistency*** with other ISO methods.

Baseline protocol taxonomy

Fundamental components of baseline calculation methods based on whole-premise interval metering are the following.

- ***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 to align it with known conditions of the curtailment day.

Data Selection Criteria

Common starting points for data selection include

- Use of the last 10 to 20 uncurtailed business days
- Use of a subset of the last 10 or 11 business days that had the highest load
- Use of a full season of data.

Selection criteria include varying procedures for excluding days from the starting point and replacing excluded days, sometimes in an iterative process.

Estimation Method

Most estimation methods can be characterized as either an average by hour of the day and sometimes day type, or some form of weather-based regression model.

Adjustment Methods

Basic adjustment methods (used to adjust interval meter on the day of the curtailment) are:
Additive—shift the unadjusted baseline up or down so that the adjusted baseline load matches the observed load one or more hours prior to the start of the event
Scalar—multiply the unadjusted baseline load by a fixed scalar so that the adjusted baseline load matches the observed load one or more hours prior to the start of the event.

Findings

- Relatively simple methods can work reasonably well for many if not most kinds of accounts.
- No one baseline method works well for all types of accounts.
- For accounts with highly variable loads, no method based only on historic load and weather data is likely to work well.
- For accounts with weather-sensitive loads, methods that do not reflect the event day's weather are likely to be inaccurate. Methods based on average load for prior days without adjustment to the current day will tend to understate baselines and savings if events occur on particularly hot days.
- Simple averages with adjustments can work nearly as well as formal weather models in many cases.

Recommendations

Offering Options

A general recommendation is that baseline calculation protocols used for demand response programs should provide for alternatives based on customer load types and operating practices. One way to simplify the provision of options is to establish a default method and allow certain deviations in the method for special types of customer accounts.

The basis for the selection of a method should be not just the 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 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 pre-curtailment hours.

Practical Default and Alternative Baseline Calculation Methods

A method that generally works well for a range of load types is the simple average of the last 10 days, by hour of the day, with additive adjustment to the load observed 2 hours prior to the curtailment period. This method can be recommended for both weather-sensitive and non-weather-sensitive accounts, with both low and high variability, for summer and nonsummer events.

More Recent Findings and Recommendations

Evaluation efforts since the publication of KEMA's protocol assessment have had experience generally in line with the findings from that work. (See the list in Section 6.) One example is the evaluation of California's Large Customer DR program (Quantum Consulting and Summit Blue, 2006), which followed recommendations from the CEC study.

- Like KEMA, Quantum found that 10-day adjusted baselines performed well in most, but not all cases.

- Like KEMA, Quantum found a class of large, customers with highly variable load patterns for which no baseline method appeared to perform well.

- Quantum used a 3-day baseline, and found this baseline to be biased high. This finding was in the context of a program where customers could choose to bid DR or not, after calculating the baseline that would apply for the bid day. In this context, one would expect higher bid behavior on days when the baseline was serendipitously high. By contrast, the KEMA study was not focused on bid programs. For most of the data sets examined by KEMA, demand response actions would not have been affected by the baseline calculation. (The KEMA study did not examine a 3-day baseline.)

- The Quantum result was based on a year with mild weather. The KEMA results included data from hot summers in several locations. KEMA identified the understatement of baseline and impacts using unadjusted representative baselines as a natural consequence of having event days typically called on days with more extreme weather than the representative days used in the baseline. In a milder summer, this relationship would not necessarily be the case, and opposite results might occur.

Possible approaches to dealing with highly variable loads in DR programs where customers are paid incentives for reductions relative to a baseline include the following.

- Establish predictability or stability criteria for customers to participate. For example, the credited reduction must be greater than the typical level of variation based on historic load data. Thus, an acceptable level of variability could be defined not relative to the customer's peak load or estimated reduction, but relative to the targeted reduction credit. (This approach requires a program structure where there is an explicit target reduction level or commitment.)
- Payments might be made for reductions only to the extent they exceed some deadband of natural variation for the customer.

- o Require customers with highly variable loads (based on historic load data) to notify the program a certain number of days in advance of a major shutdown or load increase. This advance notification of change from typical patterns would be used in the baseline calculation, and could reduce the variability relative to that baseline to an acceptable level. Customers who exhibit large changes without advance notice would be ineligible for continued participation.

2. Reliability, M&V, Load Forecasts, Baselines, and Free Ridership

What is the latest research and opinion on the firmness, sustainability and reliability of demand resources; measurement and verification (M&V) protocols; customer load forecasts; customer baseline estimation; and the potential for free-ridership?

Firmness, sustainability and reliability of demand resources

A new program with no established track record is necessarily an uncertain resource. The more the program is modeled on existing, proven programs, the more it can be relied on. Factors that increase confidence in a DR resource include the following:

- o It has been operated and evaluated over a number of years under a consistent set of rules.
- o The evaluation has not only determined impacts under the particular conditions under which DR events were called (e.g., time of day, weather, event duration), but has also projected impacts under other conditions.
- o To the extent there have been opportunities to compare projected impacts with observed impacts, these have been found to be consistent.
- o Controlling for the conditions of the DR events, the evaluations have shown consistent results over time.

A benefit of many demand resources in terms of reliability is diversification. A supply-side resource is typically either fully available or not available at all. For a demand resource consisting of a large number of customers participating in a program, the exact reduction that will be delivered may be less certain, but some minimum resource can be counted on with virtual certainty.

Measurement and Verification (M&V) Protocols

See above

Customer Load Forecasts

Forecasts of customer load are intrinsically linked to the question of how baselines are defined and how program attribution is determined. Program-level savings must be defined relative to the aggregate load that would have existed without the program. The system load forecast without a particular program includes an existing level of efficiency and demand response, including the effects of previously and currently existing programs and price structures. The value of the incremental program depends on the incremental load reduction.

For an individual customer, the baseline calculation for determining demand reduction is a short-term forecast of load in the absence of the demand reduction event.

Customer Baseline Estimation

See above

Potential for Free-Ridership

See Attribution discussion above.

3. Applying Energy Efficiency M&V Approaches to Wholesale DR

Are the approaches and principles that have been used historically to measure and verify savings from energy efficiency applicable to wholesale demand response resources? If they are not, what changes are needed to make them applicable?

The basic approaches and principles that have been used for energy efficiency programs apply also to DR resources. Following are some additional considerations in the context of DR.

Portfolio level accuracy. Energy efficiency programs are typically evaluated a single program at a time, and the accuracy of the impacts determined is assessed for individual programs or even program components. When these programs are considered as wholesale demand resources, it is the portfolio-level accuracy that is important. In general, the accuracy of the sum of many portfolio elements will be better (in relative terms) than the accuracy for any one element.

Impact load shapes. Energy efficiency programs often focus on energy savings and only incidentally consider peak demand savings. For use of these resources in peak demand reduction, it is important to know how the energy reductions are distributed over time, and in particular what the kW savings are likely to be at hours the demand resource will be needed. This need increases the importance of determining impact load shapes accurately.

Lower bound savings estimates. Energy efficiency program impact evaluation is often based on statistical estimates. These estimates can often provide a basis for calculating lower bounds achieved with high confidence, even if the "best" point estimate is not tightly determined. A lower bound analysis could be helpful operationally if the level of response from a planned event is critical to system reliability.

4. M&V for DR To Avoid Transmission Expansion

If demand resources are providing capacity resources or serving as an alternative or complement to transmission expansion, what are the key factors that need to measured and verified, and at what level of precision? Can the current form and precision of measurement and verification approaches be utilized or are new approaches needed?

The primary additional consideration in the context of transmission expansion is the potential need for more location-specific estimates.

5. Uniformity for DR M&V

Should there be uniformity on how demand resources are measured and verified across markets?

Consistency in measurement and verification methods can be helpful to customers interested in participating in programs. Currently, there is not necessarily consistency across programs available from different parties within a given market. Customers who have facilities in multiple markets would welcome consistency of program rules, including baseline calculation methods. In addition, comparison of programs by evaluators and policy setters could benefit from consistent calculation methods.

At the same time, different programs serve different customer types. Markets have different histories. Different methods may be effective or may be trusted by stakeholders. Program impact evaluation needs to consider all available information and apply the methods that appear to be most meaningful and accurate in a given context. Overly constraining these methods can limit information and methodological advances.

One product of KEMA's protocol study for the CEC was the development of a taxonomy of baseline calculation methods. Several reviewers found this taxonomy to be a key step forward in understanding, describing, and comparing methods. Even if uniform calculation protocols are not defined, consistent terminology in describing methods would be welcome.

6. Selected References

Charles River Associates, 2006. *California's Statewide Pricing Pilot: Commercial & Industrial Analysis Update*, Working Group 3.

Efficiency Valuation Organization http://www.evo-world.org

KEMA-XENERGY, 2003. *Protocol Development for Demand Response Calculation — Findings and Recommendations*, California Energy Commission.

Lawrence Berkeley National Laboratory 2005. *Development and Evaluation of Fully Automated Demand Response in Large Facilities*, California Energy Commission.

Lawrence Berkeley National Laboratory and Utilipoint International, 2007. *Estimating Demand Response Market Potential among Large Commercial and Industrial Customers: A Scoping Study*, US DOE. (Authors Charles Goldman, Nicole Hopper, and Ranjit Bharvirkar, Lawrence Berkeley National Laboratory; Bernie Neenan and Peter Cappers, Utilipoint International)

Quantec, 2006. *Evaluation, Measurement, and Verification Services for the Enhanced Automation Initiative –Program 1287-04,* CALMAC, www.calmac.org.

Quantum Consulting and Summit Blue, 2006. *Evaluation of 2005 Statewide Large Nonresidential Day-Ahead and Reliability Demand Response Programs*, Southern California Edison.

TecMarketWorks and Megdal & Associates, 2004. *The California Evaluation Framework*. Study ID SCE0147.01. CALMAC, www.calmac.org.