

The Demand Response Baseline

Introduction

EnerNOC, Inc. is a leading developer and provider of clean and intelligent energy solutions to commercial, institutional, and industrial customers, as well as electric power grid operators and utilities.

EnerNOC supplies demand response (DR) resources in open market programs and under bilateral arrangements with vertically-integrated utilities throughout North America. As DR investments increase in many regions, baseline methodologies for DR resources become increasingly

important. This white paper provides a discussion of EnerNOC's experience working with different baseline methodologies across North America and puts forth EnerNOC's perspective on best-practices for baseline calculations.

Why Baselines Matter

A properly designed baseline methodology is perhaps the most important determinant of the success of any DR program - it enables grid operators and utilities to measure performance of DR resources. A well-designed baseline benefits all stakeholders by aligning the incentives, actions and interests of end-user participants, aggregators, utilities, grid operators and ratepayers.

While it is straightforward for a utility or third-party-provider to track facility load at regular intervals, it is more challenging to measure curtailments in response to a DR event. During a DR event, actual facility load must be compared to "business as usual" load or what the facility load would have been but for the implemented curtailment measures. "Business as usual" load is estimated using a baseline methodology. The difference between the baseline and actual load constitutes that facility's DR performance.

Because performance measurement is entirely dependent on the baseline calculation, the baseline methodology must calculate as accurately as possible the "business as usual" load. For example, if a facility's baseline is calculated as 350 kW and during a DR event its load is measured at 275 kW, then the load reduction contribution from this customer would be 350 – 275 kW, or 75 kW. A baseline methodology that systematically over-estimates the "business as usual" loads will over-value the contribution of a DR resource to the grid. Conversely, a baseline methodology that under-estimates "business as usual" load will under-value the DR resource.

The baseline represents a theoretical figure. Therefore, it should come as no surprise that several calculation methodologies have arisen representing a broad range of qualities and characteristics discussed in greater detail throughout this whitepaper.

Important Baseline Characteristics

EnerNOC has developed a four-pronged framework for considering different baseline methodologies:

Accuracy

Customers should receive credit for no more and no less than the curtailment they actually provide.

Integrity

A program should not encourage irregular consumption and irregular consumption should not influence baseline calculations; in other words, a baseline with a high level of integrity will protect against attempts to "game the system."

Simplicity

The baseline and resulting curtailment calculations should be simple enough for all stakeholders to calculate, including end-user customers.

Alignment

DR program designers should consider the goals of the DR program when choosing a baseline methodology; for example, a baseline methodology should minimize unintended consequences such as inadvertently penalizing real curtailment efforts.

Balancing the appropriate emphasis on these traits is not easy. In some cases, a baseline resistant to manipulation can be so complex as to be unworkable by program stakeholders. On the other hand, the simplest approaches could allow market participants to exploit the baseline in their favor. Baselines designed exclusively for accuracy can be both extremely complex and vulnerable to manipulation. EnerNOC's experience in DR markets across North America has resulted in our conclusion that it is critical to design a baseline that balances the qualities of accuracy, simplicity, integrity and alignment.

Program Categories

Before considering specific baseline methodologies, it is important to think about the categories of programs that require a baseline calculation. Generally, any demand side program that provides reductions off a “business as usual” load requires measurement against a baseline. Examples of such programs include the following [key differentiating traits among program categories include event triggers, frequency and deployment period¹]:

Energy Efficiency

Programs implement or promote energy saving measures which reduce a user’s day-to-day electricity consumption.

Demand Response

Resource Adequacy or Capacity

Programs are triggered to provide grid stability during periods of peak demand. Upon dispatch, on-call participants provide load reductions for several hours at a time. Event frequency is typically limited.

Economic or Energy

Programs are triggered by energy price signals,

allowing participants to bid demand reductions into the system at their discretion. Economic programs are not usually governed by the frequency and duration limitations observed in resource adequacy programs.

Balancing or Ancillary Services

Programs are triggered by system imbalances in cases of generation or transmission interruption. Ancillary services programs are characterized by high event frequency coupled with short event durations. Programs are often called upon 10 minutes notice or less.

¹NAESB, “Recommendation to NAESB Executive Committee: Review and Develop Business Practice Standards to Support DR and DSM-EE Programs”, September 29, 2008

Not surprisingly, factors such as trigger, event frequency and deployment period lead to discrepancies between the desirable baseline characteristics for different program types. For example, the appropriate baseline methodology for a permanent energy efficiency measure requires far different analysis than for a 10 minute ancillary services dispatch, called on a moment’s notice. The bulk of this paper will focus on baselines related to DR rather than energy efficiency programs.

Distributed Generation Baselines

It is also important to distinguish between distributed generation and curtailment DR resources. Measuring the performance of a generation-backed resource is relatively straightforward. Prior to start-up, the baseline output of a generator is zero. When operating, the net demand reduction value to the grid is measured directly as the total load on that generator. This is made possible by instantaneous metering and communication technology.

DR performance for resources with generation is often measured by the kW and kWh of generator output.

Although at face value this appears more accurate than a theoretical baseline calculation, difficulties of agency arise. Since both supply and demand are determined internally (at the customer site), the participant has an incentive to maximize performance by increasing load on the generator (turning on air conditioning, lighting, etc.). This additional output provides no benefit to the grid. In this case, the business as usual consumption during an event period may be better represented by a standard baseline-based calculation.

Critical Baseline Elements

DR programs create incentives for customers to change their behavior when demand reductions are needed on the electrical grid. The baseline should neither reward nor penalize a facility for the natural load variance caused by normal operations. Similarly, a baseline should appropriately account for business changes unrelated to the DR program, such as growth or contraction, as well as other factors inherent to typical business activities such as batch processing in a manufacturing facility.

A primary difficulty of developing an accurate baseline is the inherent volatility of a customer's energy consumption. Consider a small manufacturer that operates during

normal business hours, Monday through Friday. This facility's load may vary within a normal week, month, or year, depending on a business cycle; the manufacturer may produce its product on an intermittent schedule according to seasonal demand. Local weather conditions also drive hour-to-hour, day-to-day, and week-to-week variations in load.

An effective baseline measures only the real, material value that the participant provides to the electric grid. Several key elements that make up a baseline analysis are defined in the table below and outlined in greater detail in the following paragraphs.

FIGURE 1. BASELINE ELEMENTS DEFINED

Profile Baseline	Incorporates frequent granular measurement across similar days, resulting in a demand estimate that mimics the dynamic nature of a customer's demand curve over a 24 hour period.
Static Baseline	Generates a flat demand estimate representing the average demand during an extended time interval (such as a season), providing one demand estimate regardless of time of day or day of the week.
Measurement Granularity	Refers to size of time intervals used for discrete demand measurements (e.g., 5-minute).
Baseline Window	The window of time (typically days) over which demand data is collected in order to establish a baseline. ²
Exclusion Rules	Rules governing data within a baseline window that is included or excluded from the calculation (e.g. days of an event).
Baseline Adjustments	Changes to a calculated baseline based on actual demand or weather conditions on the day of a DR event.
Additive Adjustment	A fixed kW adjustment across all event time intervals. If the observed demand during an adjustment period is 20 kW above the estimated baseline, 20 kW is added to the estimated baseline for each time interval during the event.
Scalar Adjustment	A percentage multiplier across all event time intervals. If the observed demand during an adjustment period is 20% above the estimated baseline, the estimated baseline for each time interval during the event is multiplied by 120%.
Adjustment Cap	A limit on the magnitude of a baseline adjustment.
Individual Baseline	The concept of calculating performance or applying exclusion rules at the individual site level, then summing those performance calculations to calculate the performance of an entire DR portfolio.
Portfolio Baseline	The concept of calculating performance or applying exclusion rules at the portfolio level.
Average Calculation	Baseline for a given time interval is calculated as the average demand observed across a number of similar time intervals.
Regression Calculation	Baseline calculation takes an extensive data set and determines the relationship between a number of different variables, such as weather, time of day and demand, among others.

² NAESB, "Recommendation to NAESB Executive Committee: Review and Develop Business Practice Standards to Support DR and DSM-EE Programs", September 29, 2008

Profile or Static Baseline

It is important to understand the distinction between a profile baseline, often referred to as a customer baseline (CBL) and a static baseline, sometimes referred to as average peak monthly demand (APMD) or firm service level (FSL). Consider the following example (Figure 2 below) in which the same customer responds in the same manner in a hypothetical event, yet is credited with delivering dramatically different capacity levels due to differences in how the baseline is calculated.

The profile baseline, or CBL, (red line) in the event on the left was determined by using granular time interval data intended to mimic the dynamic shape of a customer’s demand. Note how the baseline follows actual demand (black line) very closely leading up to and following the event. Peak performance for the event is measured at 65 kW over a nominated capacity of 50 kW.

The static baseline in the example on the right was determined using an approach that simply averaged the peak monthly demand for this facility over the previous corresponding delivery season (previous summer or previous winter). The resulting baseline is far greater than the actual load at the facility when the event was called. Peak performance for the event is measured at 83 kW, over the same 50 kW of nominated capacity, but 32 kW (or more than half of nominated capacity) was met through what is known as “natural curtailment” or “incidental performance.” This additional capacity “delivered” by this

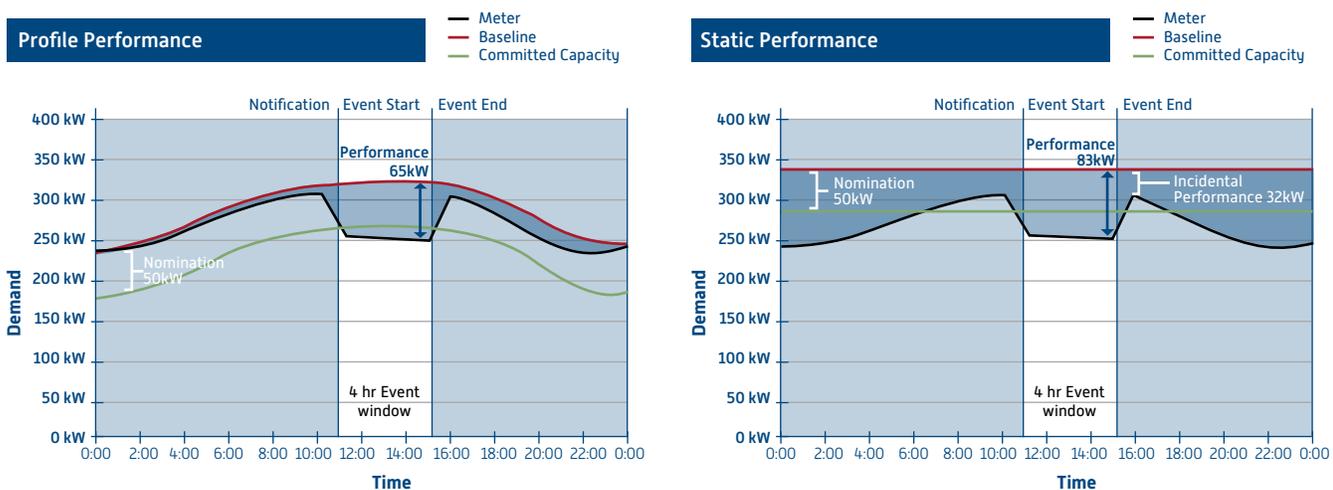
site in this event represents additional program cost to the utility or grid operator in the form of increased energy and/or capacity payments above the true value the site provided to the grid through its load reductions.

Measurement Granularity and Communication Requirements

Depending on the need being met by the DR resource, and the level of granularity preferred by the utility or system operator, an effective baseline methodology must incorporate an appropriate timing interval for data collection and calculation. In the case of capacity and energy, metering and settlement are conducted on either a five-minute, fifteen-minute, or hourly basis in current DR programs. In contrast, some ancillary services such as regulation and frequency response are measured in shorter increments, typically from a few seconds up to one minute.

Many DR programs are intended for peak load or emergency conditions. In this context, the quantity and value of DR performance is most appropriately measured in short time intervals. If cost effective, increased data granularity can only increase the value of DR through more accurate performance measurements, increased resource visibility for grid operators, and easier settlement processes. A 5-minute interval is sufficiently granular to accommodate most notification periods and event durations while modern technology makes 5-minute interval metering cost-effective a majority of the time.

FIGURE 2. PROFILE VS STATIC BASELINE



Baseline Window

In selecting the length of the baseline window, it is worth considering the implications of design extremes. There is a strong a priori argument for the use of recent data since such data better approximates what the facility load would have been during an event. However, with longer DR events and advanced notification, a short baseline window can exhibit problems of accuracy and be susceptible to manipulation.

Consider an example that uses a very short baseline window: a customer's average demand during the time interval immediately preceding a DR event. For a brief ancillary services event with a dispatch notification of 10 minutes or less, this has the advantage of automatically compensating for exogenous factors such as weather and business climate of the event day while addressing the actual load on the grid at that given time. Furthermore, due to the brief nature of an ancillary services need, the likelihood of customer demand shifting significantly over a brief time window in the absence of an event dispatch is minimal. Thus the baseline is sound.

However, for a resource capacity event that might last up to eight hours, upon two hours notice, a very short baseline window raises a number of concerns. First, a baseline that only considers very recent data may place an undue emphasis on short-term variations in load and might not accurately capture "true" demand reductions. Second, given sufficient or excessive warning and incentive to do so, a facility could actively and intentionally increase consumption prior to a DR event in order to maximize its baseline and thus overstate actual curtailment levels. A longer baseline window acts to prevent gaming such that the cost of active manipulation to elevate baseline levels outweighs the benefit as the customer's utility bills would quickly increase due to increased consumption and potentially higher demand charges.

It is generally accepted that a period of approximately 10 (non-event) business days reasonably represents consumption for normal operations and therefore makes up a preferred baseline window for resource adequacy and capacity programs. **Using a 10 day time window provides an appropriate balance – short enough to account for near-term trends and long enough to limit opportunities for manipulation.**³

Higher frequency programs, such as an economic program responding to price signals, require a window that places greater emphasis on the contemporary data that best

mimics the economic conditions driving an event. As a result, a baseline window such as five non-event days is recommended. In extreme cases, some economic programs are called on a daily basis, allowing customers to participate on consecutive days, indefinitely. With a baseline window that considers only non-event days, a customer's baseline will cease to reflect current conditions if the customer participates day in and day out. In these cases, we recommend a provision in the baseline window that forces the inclusion of event days when a baseline window would otherwise look back 30 days or 45 days in order to find a non-event day, for example.

Exclusion Rules

After determining an appropriate time window, the specific data to be evaluated within that window must be considered. For example, data should be evaluated across "like" time periods, with a sensible baseline calculation considering that weekdays are typically like other weekdays, and weekday afternoons are typically like other weekday afternoons. As discussed with regard to a profile approach, a participant's baseline should reflect the dynamic nature of a participant's load curve. This aspect of baseline design is widely, if not universally, accepted.

Simply looking across "like" time periods is not necessarily sufficient. Within a given baseline window, a calculation could consider "like" time periods on each "like" day, providing an average across those days. However, imagine a customer account in a resource adequacy program where weather is a central determinant of electrical consumption. For this participant, peak load situations often occur on extreme weather days, when events are called. In this case, the baseline window will include a number of non-event days likely displaying less extreme weather conditions, and therefore less demand. Using the average of all non-event days across the baseline window will consistently understate the participant's baseline, reducing their incentive to participate while challenging both the accuracy and alignment of the program.

To combat this understatement, many programs choose to look only at a set number of highest load days among the non-event data-set. Examples of this technique occur in California, where a "High 3 of 10" approach is employed in many IOU-sponsored programs such as the Capacity Bidding Program, and in New York, where a "High 5 of 10" approach is employed for energy payment calculations in the Emergency Demand Response Program. Among the 10 prior non-event days, the three or five days with the highest peak load are averaged for inclusion in

³ KEMA – XENERGY, "Protocol Development for Demand Response Calculation- Findings and Recommendations", February 2003.

that event’s baseline calculation reducing the typical understatement of participant baselines. A “High 5 of 10” approach addresses the issue of understated performance while incorporating more load data than a “High 3 of 10” approach, nicely balancing accuracy and alignment.

Figure 3 provides an example of the calculation of facility baseline usage for Intervals 1, 2 and N assuming that the five (5) highest demand days are days 2, 4, 6, 7 and 9. Based on this data, the facility’s baseline for time intervals 1, 2 and N would be 2,280 kW, 2,380 kW and 2,280 kW respectively. These interval calculations determine the shape of the baseline over time.

FIGURE 3. HIGH 5 OF 10 DATA

Day	Interval 1 (kw)	Interval 2 (kw)	Interval N (kw)	Average Usage (kw)
1	2,000	2,100	2,000	2,033
2	2,100	2,200	2,100	2,133
3	2,000	2,100	2,000	2,033
4	2,200	2,500	2,200	2,300
5	2,000	2,100	2,000	2,033
6	2,100	2,200	2,100	2,133
7	2,400	2,300	2,400	2,367
8	2,000	2,100	2,000	2,033
9	2,600	2,700	2,600	2,633
10	2,000	2,100	2,000	2,033
Baseline	2,280	2,380	2,280	

In programs where peak demand is not the primary consideration, such as an economic program, less exclusion is required. In fact, because such programs often use a shorter baseline window to emphasize recent information, the exclusion of fewer days actually emphasizes integrity by minimizing the ability for a participant to game the system.

Ultimately, any exclusion rule chosen for a baseline calculation must account for both accuracy in performance measurement and alignment of program goals with customer behavior and incentives.

Baseline Adjustments

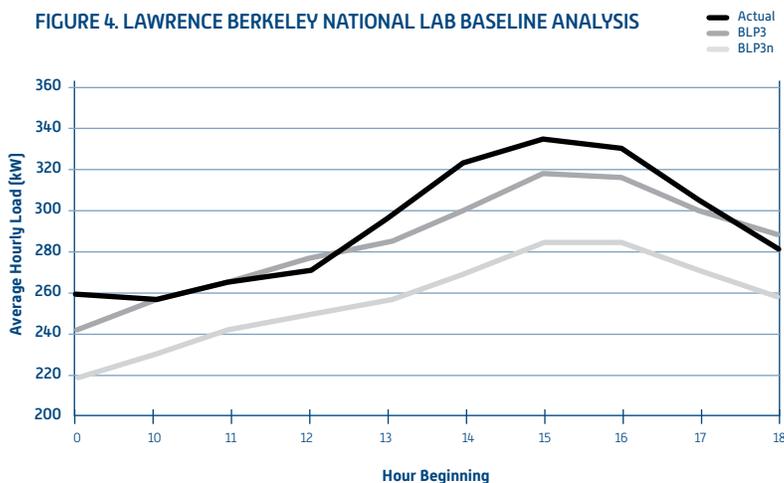
Because customer demand is often heaviest on event days, capturing day-of realities in a customer load profile is essential to delivering accurate performance calculations. A common way to address this need is

through an adjustment based on day-of event conditions. Several factors affect a customer’s load prior to a DR event. For example, the first day of the year that requires air conditioning is likely to exhibit a vastly different load profile from the preceding cooler days. An appropriate adjustment mechanism for day-of load conditions is necessary to more accurately reflect actual circumstances and avoid penalizing customers who are consuming more energy than a “like” day alone.

Such a short-term adjustment to the “like” day calculations should be based on the conditions either during, or immediately preceding, a DR event. Current programs, including those in ISO New England and the PJM Interconnection (PJM), use readily verifiable data, such as temperature or load in the period prior to an event as the basis for a day-of adjustment.

Because adjustments based on weather tend to be complex and difficult to verify, an adjustment based on the facility load immediately preceding the event is preferable. However, whenever possible, it is best to limit the influence that a customer may have over this calculation. An effective way to do this is to base an adjustment on relative energy consumption during a calibration period that occurs prior to DR event notification.

FIGURE 4. LAWRENCE BERKELEY NATIONAL LAB BASELINE ANALYSIS



In a recent study of baseline calculations, the Lawrence Berkeley National Laboratory concluded that “applying a morning adjustment factor significantly reduces the bias and improves the accuracy of all baseline load profiles examined in our sample of buildings.” The study’s results are presented in Figure 4 above, comparing actual demand with “High 3 of 10” baselines with (“BLP3”) and without (“BLP3n”) adjustments.⁴

⁴ Lawrence Berkeley National Laboratory, “Estimating Demand Response Load Impacts: Evaluation of Baseline Load Models for Non-Residential Buildings in California”, January 2008, page 25.

Adjustments are calculated either with a scalar or with an additive technique.⁵ The scalar technique is based on a percentage comparison. If load on an event day prior to notification is measured to be 130% of the calculated baseline, each time interval of the event baseline would be the product of the calculated baseline and 130%. The additive approach instead calculates the actual demand difference in kW. If load during the calculation period is 50 kW above the calculated baseline, then 50 kW is added to each interval in the actual event baseline.

It is also important to consider whether adjustments reflect demand conditions symmetrically (baseline adjusted up and down) or asymmetrically (baseline only adjusted up). The symmetric approach considers that day-of conditions can have a real impact on customer demand in both directions and therefore it can be argued that symmetric adjustments maximize the accuracy of a baseline calculation.

However, from an alignment point of view, downward adjustments represent a serious cause for concern. The reduction of a customer baseline based on day-of conditions can have damaging unintended consequences for a number of reasons, including customers penalized for taking anticipatory curtailment actions; customers with unusually low demand prior to an event lacking incentives to keep that load offline; and customers focusing their attention on administrative details and keeping load online through a notification period rather than actual curtailment procedures. Each of these challenges is outlined in greater detail below.

First, Participants should be rewarded for anticipating a DR event and taking action, rather than harmed by their responsible behavior. For example, a participant might cancel an afternoon shift or manufacturing process based on an informal ‘heads-up’ suggesting that an event is likely, or a customer may anticipate an event on their own after a string of hot days. In either case, a symmetric adjustment picks up on these decisions and reduces that customer’s baseline, limiting the customer’s ability to receive credit and disrupting the alignment between utility or grid operator goals and end-user incentives.

Participants should also be motivated to keep demand offline. If a participant happens to be in-between processes on the morning of an event or to have equipment offline for maintenance, a symmetric adjustment could take that customer’s baseline toward zero, in which case they would literally have no incentive to maintain their curtailment

during the event window, when their participation is most critical. Once again, the downward adjustment in this case upsets the alignment of incentives among DR constituents.

Finally, when an adjustment calculation includes time intervals subsequent to event notification, such as with a day-ahead notification, symmetric adjustments can cause participants to focus on counter-productive behavior, such as maintaining load through the end of an adjustment period, rather than on critical curtailment measures. Upon notification, it would be in such a customer’s best interest to first determine the exact time associated with the adjustment period in order to make sure they are not reducing their load too early. In the face of peak demand, the idea that customers would maintain load for administrative purposes does not reflect the best interests of the grid. Furthermore, it represents added complexity for all parties where simplicity is ideal. In each of these cases, a symmetric adjustment runs the risk of significant incentive and customer satisfaction consequences capable of undermining current and future DR performance.

Upward adjustments are not without challenges. It is plausible that anticipatory actions could result in an overstatement of compensation, through gaming. Furthermore, participants with unusually high load on the morning of an event-day will see an increased baseline during the event window, regardless of whether or not their morning load was correlated to likely afternoon demand. However, it is important to note that in each of these cases, the issue is about fairness in compensation, not the alignment of incentives to encourage curtailment during an event window, when performance is critical. While it can be argued that in these scenarios a participant might receive more than their fair share, these participants remain incentivized to curtail as much load as possible during a DR event, aligning their interests with those of the grid.

Furthermore and most importantly, to not provide an upward adjustment in a vast majority of cases would limit a customer’s ability to perform on the highest demand days, resulting in harmful consequences with respect to alignment of incentives among all DR interests.

In order to limit the magnitude of any adjustment, some programs use a cap. For example, a customer with a 100 kW baseline exhibits demand of 130 kW prior to event notification. Using an additive adjustment, the customer baseline throughout that day’s event would be increased

by 30 kW. However, in the presence of a cap, that additive adjustment would be limited: if the cap were 20%, then the additive adjustment would be 20 kW.

In the presence of symmetric day-of adjustments, a cap (particularly on the downside) helps to reduce the alignment issues and customer satisfaction issues described above. If capped, the impact of the downward adjustments that come with symmetry is muted. However, a cap on upward adjustments again leads to the unintended consequences of limited or understated end-user performance despite real curtailment. Consider the demand of a firm with a weather dependent peak experiencing a hot day following a stretch of cooler weather. In this case, it is reasonable to assume that actual demand is significantly higher than demand observed during the baseline window. However, in the face of a cap, such a customer may receive little or no credit despite taking curtailment action and delivering real value to the grid.

Ultimately, any demand response program must consider the impact of unintended consequences on customer incentives and behavior when weighing the symmetry of adjustments or the use of caps.

In calculating day-of adjustments, a period between two and four hours provides sufficient data to calculate a reasonable adjustment factor. By using a day-of adjustment approach, the final baseline preserves the integrity of a customer's load profile, linking their performance incentives with actual behavior while accounting for unexpected day-of deviations from historical data.

Individual Baseline or Portfolio Baseline

Baseline windows and exclusion rules can either be applied at the individual customer level or at the portfolio level. With a "High 5 of 10" method, for example, the five highest demand days can be determined for each individual customer, their individual baselines calculated accordingly, and the individual baselines summed together to arrive at the portfolio baseline. In contrast, with a portfolio baseline method, the five highest demand days are determined for the portfolio as a whole, and those days are then used in each individual customer calculation.

It is important to consider alignment as it relates to individual and portfolio baselines. The portfolio approach represents essentially a random choice of exclusion days, whereas the individual approach chooses the days in line with that customer's high demand. With a portfolio method, customers will view their performance incentives

not as the firm result of admirable curtailment efforts, but as the random result within a range of performance calculation possibilities, tied loosely to their curtailment efforts. Such a disconnect hurts all parties associated with the demand response resource.

Another key benefit of the individual approach is that the individual customer is capable of measuring their own performance in near real-time, reflecting simplicity. On the other hand, with a portfolio approach, the customer must await information from the program as a whole to know which historic days will be used to determine their individual baseline. Furthermore, if the program does not have complete data from all assets at the time an event is called, the baseline might not be able to be calculated until after an event's completion, making it impossible for a participant to calculate their ongoing performance during an event, when it matters most.

In looking at actual EnerNOC data from a March 2008 demand response event employing the "High 3 of 10" method in California (Figure 5), it is interesting to note that with a portfolio method, less than 10% of customers had their highest three demand days aligned with those of the portfolio. In other words, over 90% of participants were not only unable to calculate their baseline based on internal demand data, but also reliant on random (from the participants' perspective) information to understand

FIGURE 5. CUSTOMER VS. PORTFOLIO DATA

Customers with Individual Top Demand Days Matching Portfolio Top Demand Days		
	Number	% of Total
3 Days Matching	20	9%
2 Days Matching	103	45%
1 Day Matching	71	31%
0 Days Matching	37	16%
Total	231	

their official performance. And for 16% of the participants, the "High 3" days used to calculate their baseline included none of their top demand days for the period, highlighting the inaccuracy of this approach from an individual customer perspective.

Calculation Type

Baseline data can be analyzed using distinct calculation methodologies. In general, these can be divided into two groups, average and regression.

Average

One of the primary benefits of the average technique is simplicity, affording all stakeholders the opportunity to vet data and analysis themselves and in turn promoting transparency in the process, linking actual curtailment efforts to real, recognizable incentives.

When coupled with other appropriate baseline characteristics, averaging also promotes accuracy and integrity. To illustrate, it is helpful to consider the baseline techniques that employ the average method in many programs in California and in New York. Many California programs use a ten day baseline window, excluding all but the three highest demand non-event days. From an accuracy standpoint, California's baseline window considers only recent behavior in an attempt to best approximate actual demand conditions. Furthermore, the exclusion rules recognize that DR events are typically called on the highest demand days. By excluding all but the highest non-event demand periods in the average, the natural tendency to understate baseline demand is mitigated. Regarding integrity, the fact that the baseline considers a 10 day window prevents irregular consumption from influencing the data. In order to successfully game the system, a participant would have to control multiple weeks of demand in order to prepare for a potential event, quickly driving the costs of gaming above and beyond the potential benefits.

In contrast, New York's Average Peak Monthly Demand (APMD) approach raises a number of questions. For an APMD summer DR program, a participant baseline is

calculated based on the average peak monthly demands for June, July, August and September of the previous year. From a baseline characteristics standpoint, this is a broad baseline window with no exclusion rules, presenting accuracy and integrity challenges. Consider a manufacturer that closes a plant but is paid for demand reductions in a resource capacity program. Not only does this approach promote inaccurate data (a customer contributing zero curtailment is credited with curtailment performance), it also presents dubious integrity as a closed or downsized operations are easily able to game the system.

Regression

From an accuracy point of view, a regression model allows a DR program to use sophisticated statistical tools to calculate a baseline, promoting the highest degree of accuracy possible. Furthermore, because a regression analysis is so complex, opportunities to game the system are minimized, promoting integrity. Unfortunately, the complexity argument also makes the regression less hospitable to stakeholders, making it increasingly challenging for participants to understand the link between their actual curtailment efforts and the performance for which they are credited.⁶ Furthermore, due to the data requirements of a regression approach, it is possible that a baseline cannot be calculated until after an event's completion, limiting the ability of all constituents to understand event performance in near real-time. This focus on accuracy to the detriment of simplicity can create significant performance issues as incentives become increasingly blurred.

⁶KEMA – XENERGY, "Protocol Development for Demand Response Calculation- Findings and Recommendations", February 2003, p. 2-11.

EnerNOC Recommendation

EnerNOC asserts that the goal of a baseline should be to capture what facility load would have been the day and time of the event had the customer not taken actions as a direct result of participation in a DR program.

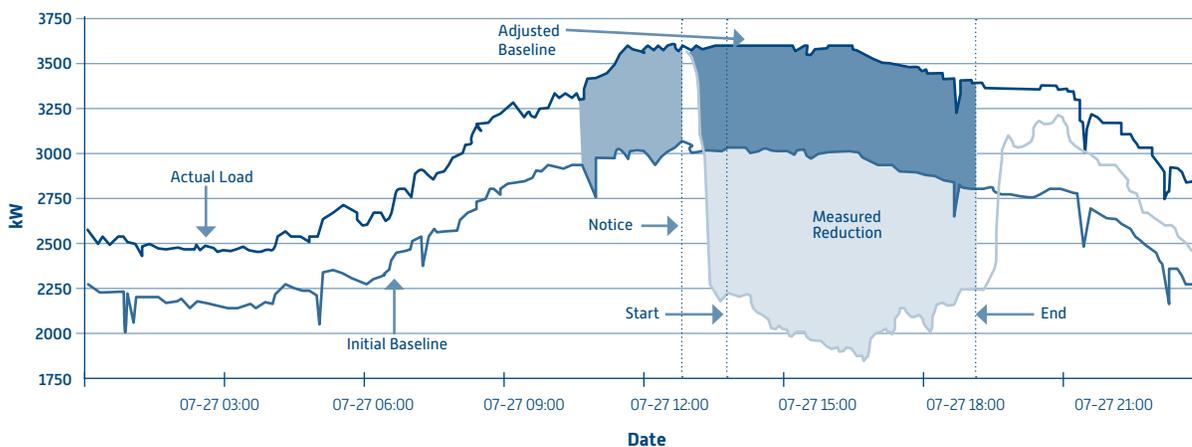
The calculation of a baseline establishes customer performance, represented by the measured difference between adjusted baseline and actual load in Figure 6 below.

Resource Adequacy / Capacity

EnerNOC recommends a baseline approach exhibiting the following attributes, ensuring that the four important qualities of a baseline are accounted for:

- 10 day baseline window (accuracy and integrity)
- High 5 of 10 exclusion rules among “like” days, excluding event days and holidays (accuracy and alignment)

FIGURE 6. EXAMPLE BASELINE AND PERFORMANCE MEASUREMENT FOR DEMAND RESPONSE ASSET



Details on calculating the initial and adjustment baselines and the measured reduction are found below.

- Additive upward day-of adjustment (accuracy and alignment)
- Individual baseline rather than a portfolio baseline (simplicity and alignment)
- Average calculation method (simplicity)

Baseline Variables

b = baseline average
 d = non-event day
 dn = nth highest energy usage day among previous 10 non-event days
 t = time interval
 c = highest kW energy consumption for a given time interval [t]

Adjustment Factor Variables

a = day-of adjustment
 t-n = time interval starting n hours prior to event notification

Performance Calculation Variables

p = total performance
 e = total time intervals during event

For a given time interval [t], initial baseline [b] is calculated as the average interval demand among the 5 highest energy usage days out of the prior 10 non-event days (this calculation is performed for each interval during the DR event, for example for each five minute window):

$$b_t = (C_{td1} + C_{td2} + C_{td3} + C_{td4} + C_{td5}) * 1/5$$

Adjustment factor [a] is calculated as the difference in observed demand and estimated baseline for a calibration

period starting two hours before event notification, with a minimum adjustment of 0:

$$a_t = \max \{ [(c_{t-1} - b_{t-1}) + (c_{t-2} - b_{t-2})] * 1/2, 0 \}$$

Total performance [p] is measured as the integrated difference between the sum of the baseline [b] and adjustment factor [a] less consumption [c] for each interval [t] over an event period beginning at time [0] and ending at time [e]:

$$p = \sum_{i=0}^e (b_i + a) - c_i$$

Capacity-setting performance [pavg] is simply the average performance during all intervals of the DR event where program rules stipulate that performance is mandatory:

$$p_{avg} = \frac{\sum_{i=0}^e (b_i + a) - c_i}{e}$$

While this methodology suggests an upward adjustment with no cap, it is important to consider the balance discussed earlier in the Adjustments section of this paper. While EnerNOC advocates here for upward adjustments as a means of preventing the understatement of customer performance, we recognize the use of symmetry in some cases. In these situations, we recommend the use of a cap in order to minimize the potential for unintended penalties on participating customers.

Note that the day-of load adjustment period is the two hours prior to event notification instead of two hours prior to the event start. Doing so makes the day-of adjustment less gameable. Also note that capacity-setting performance is measured as average performance instead of max performance. Clearly, the value of DR is best represented by the delivery of consistent, reliable performance instead of volatile, single-interval peaks.

Economic / Energy

EnerNOC recommends a baseline approach exhibiting the following attributes, ensuring that the four important qualities of a baseline are accounted for:

- 5 day baseline window (accuracy and integrity)
- High 4 of 5 exclusion rules among "like" days, excluding event days and holidays (accuracy and alignment)

- Additive upward day-of adjustment (accuracy and alignment)
- Individual baseline rather than a portfolio baseline (simplicity and alignment)
- Average calculation method (simplicity)

For a given time interval [t], initial baseline [b] is calculated as the average interval demand among the 4 highest energy usage days out of the prior 5 non-event days (this calculation is performed for each interval during the DR event, for example for each five minute window):

$$b_t = (C_{td1} + C_{td2} + C_{td3} + C_{td4}) * 1/4$$

Adjustment factor [a] is calculated as the difference in observed demand and estimated baseline for a calibration period starting two hours before event notification, with a minimum adjustment of 0:

$$a_t = \max \{ [(c_{t-1} - b_{t-1}) + (c_{t-2} - b_{t-2})] * 1/2, 0 \}$$

Total performance [p] is measured as the integrated difference between the sum of the baseline [b] and adjustment factor [a] less consumption [c] for each interval [t] over an event period beginning at time [0] and ending at time [e]:

$$p = \sum_{i=0}^e (b_i + a) - c_i$$

Capacity-setting performance [pavg] is simply the average performance during all intervals of the DR event where program rules stipulate that performance is mandatory:

$$p_{avg} = \frac{\sum_{i=0}^e (b_i + a) - c_i}{e}$$

In cases where a program may be dispatched with high frequency, it is important to consider the potential benefit of a provision in the baseline window that forces the inclusion of event days when a baseline window would otherwise look back 30 days or 45 days in order to find a non-event day, for example. Additionally, as discussed in the Resource Adequacy case previously, the important considerations of symmetry and capped adjustments apply here as well.

Note that, like in the Resource Adequacy / Capacity calculation, the day of load adjustment period is the two hours prior to event notification and the capacity-setting performance is measured as average performance instead of max performance. The same arguments regarding gaming and performance apply in this case.

Balancing / Ancillary Services

In an ancillary services event, the minimal notice and reduced event durations create a set of circumstances that require a unique baseline calculation. Generally, an ancillary services event is intended to reduce load on the grid at that moment, for a short period of time, rather than to reduce a dynamic load profile likely to fluctuate over time.

In the PJM Synchronized Reserves Program, the magnitude of a demand resource's performance is initially calculated as the difference between the demand resource's MW consumption at the start of the event and its MW consumption ten minutes after the start of the event.

In order to allow for small fluctuations and possible telemetry delays, a demand resource's consumption at the start of the event is defined as the greatest telemetered consumption between one (1) minute prior to and one (1) minute following the start of the event. Similarly, a demand resource's consumption ten minutes after the event is defined as the lowest consumption measured between nine (9) and eleven (11) minutes after the start of the event. In order to incentivize consistent response throughout the duration of an event, final performance is determined by averaging performance across each minute of the event. This is calculated as the sum of (i) 10, multiplied by the initially-measured 10-minute performance (as above), and (ii) the performance in each minute of the event beginning with the 11th minute after event dispatch (where performance during any minute is defined as the difference between the (a) demand resource's highest measured consumption between the minute prior to and the minute following the start of the event, and (b) measured consumption during that minute); divided by the total number of minutes during the event.⁷

⁷PJM Interconnection, LLC, "Amended and Restated Operating Agreement of PJM Interconnection LLC", October 14, 2008.

Summary

Since our formation in 2001, EnerNOC has been an industry leader in providing demand response services to the commercial, institutional, and industrial sectors. With deep experience across program categories, from resource adequacy to economic to ancillary services programs, EnerNOC personnel have worked with curtailment providers, utilities, grid operators and regulators to help shape the development of a number of the baseline techniques observed across North America today. Based on this experience, we believe the industry is poised for a convergence of best practices in baseline methodology in line with the preferred approaches asserted in this paper.

Regardless of specific program type, EnerNOC's experience has shown that the most successful baseline methodologies balance the essential qualities of accuracy, simplicity, integrity and alignment.

Although quantitative support is needed to fully establish the accuracy of a baseline, experience has shown that EnerNOC's preferred methodologies avoid many of the shortcomings associated with other baseline approaches, from preserving the natural shape of a customer's average load to accounting for day-of deviations from a baseline forecast.

A baseline calculation must also promote integrity through resistance to manipulation. Depending on a specific program's notification period, event durations or frequency, the choice of baseline window and exclusion rules are critical to the achievement of balance between accuracy and integrity.

With a simple approach, an end-use customer's baseline can be calculated using a straightforward spreadsheet-based tool and data of any granularity. Other methods require far more complex tools and may not be calculable without significant technological expertise. It is also important to note that overly simplified techniques, such as APMD, promote simplicity at the expense of accuracy, integrity and alignment.

Ultimately, the baseline must focus on alignment, encouraging maximum participation by end-use customers and providing a reliable mechanism for utilities and grid operators to measure the contributions of DR resources.

Boston, MA

75 Federal Street
Suite 300
Boston, MA 02110
Office: 617.224.9900
Fax: 617.224.9910

Baltimore, MD

1414 Key Highway
Suite 200-M
Baltimore, MD 21230
Office: 443.524.2880
Fax: 443.524.2887

Boulder, CO

1155 Canyon Boulevard
Suite 110
Boulder, CO 80302
Office: 303.385.0325
Fax: 303.998.1400

Houston, TX

1010 Lamar
Suite 1040
Houston, TX 77002
Office: 713.231.1021

New York, NY

14 Wall Street
Suite 6A
New York, NY 10005
Office: 212.624.0000
Fax: 212.624.0001

info@enernoc.com
www.enernoc.com

Oakville, Ontario

11-1155 North Service
Road West
Oakville, Ontario L6M 3E3
Office: 416.461.4678
Fax: 289.291.4001

San Francisco, CA

500 Howard Street
Suite 400
San Francisco, CA 94105
Office: 415.343.9500
Fax: 415.227.1645

Stamford, CT

68 Southfield Avenue
Suite 215
Stamford, CT 06902
Office: 203.961.1112
Fax: 203.961.1171

Southern CA

906 Palomares Avenue
La Verne, CA 91750
Office: 415.343.9500
Fax: 415.227.1645

©2009 EnerNOC, Inc.
All rights reserved.
No text can be reprinted
without permission.