Reimagine tomorrow.

California ISO

Baseline Accuracy Work Group Proposal

June 6, 2017

Figures

Tables

1 Introduction

Currently, the proxy demand resource (PDR) and reliability demand response resource (RDRR) use a 10 of 10 baseline with a 20% same day adjustment to estimate the load impact achieved by the resource. While research has shown this baseline to be accurate for many medium and large commercial and industrial customers, research has also shown that this baseline is not accurate for all customer types. The purpose of the Baseline Analysis Working group (BAWG) is to identify additional settlement methods which, when offered in addition to the 10 of 10 baseline, will enable the load impacts from a wider variety of demand response resources to be accurately estimated.

The BAWG identified three major areas of research.

- The use of alternative traditional baseline methods to estimate the load impact of current demand response resources.
- The option of using control groups rather than traditional baselines to estimate the load impacts of demand response resources.
- Ways to accurately measure load impacts of resources that are frequently dispatched.

1.1 Traditional baselines methodologies for current demand response resources

The research objective has been to identify additional traditional baselines which accurately estimate the load impacts of existing demand response resources that are not accurately estimated by the current CAISO-approved 10 of 10 baseline. Research has shown that the 10 of 10 baseline underestimates the load impact from residential customers, so identifying baselines for residential customers was an important task. In order to address this issue, analysis was done using data from the air-conditioning cycling programs of all three utilities. The analysis estimated the effectiveness of the current 10 of 10 baseline and tested the effectiveness of alternative baseline methodologies. In addition, the effectiveness of the 10 of 10 baseline on estimating the load impacts of reliability programs such as the Base Interruptible Program (BIP), Agricultural Pump Interruptible Program and small commercial AC load control has not been rigorously tested and these customers currently do not rely on a 10 of 10 baseline for their retail compensation.

The working group also addressed the issue of how to determine which baseline should be applied to which resources. Offering more than one baseline option raises the issue of whether or not all baseline options should be available to all customer types. For example, if a particular baseline is more accurate for residential customers than it is for commercial customers, the baseline might only be made available to resources consisting of residential customers. The working group also identified other operational barriers that may arise due to offering more than one baseline option. Ultimately, the working group recommended one day matching, one weather matching, and one control group option for both residential and non-residential customers for both weekdays and weekends. This provides flexibility for DRPs to rely on the baseline that is the most accurate for their population while ensuring that the number of baselines available does not proliferate.

1.2 Control Groups

Control groups provide an alternative to traditional baseline methodologies for the estimate of load impacts. Control group methodologies use the energy use of a group of customers who do not participate in the demand response event to compare to that of those who do. There are two main types of control groups: 1) a randomized controlled trial (RCT) and, 2) a matched control group. In the RCT a subset of participants is randomly selected in advance and withheld from curtailment during the event period. A matched control group consists of non-participants with similar characteristics to participants. The working group studied control group settlement methodologies already in use by other independent system operators and determined if they can be implemented by the CAISO. Questions that were addressed in this area include:

1. What requirements would need to be put in place to ensure the energy use of the control group accurately reflects the energy use of the treatment group?
2. What requirements regarding samples sizes or precision should be established?
3. How will the control groups be identified operationally?
4. Is it feasible to allow control groups to vary by events/rotate?
5. How can control group methodologies be established that work for both utilities and third party demand response providers (DRPs)?

1.3 Frequent Dispatch

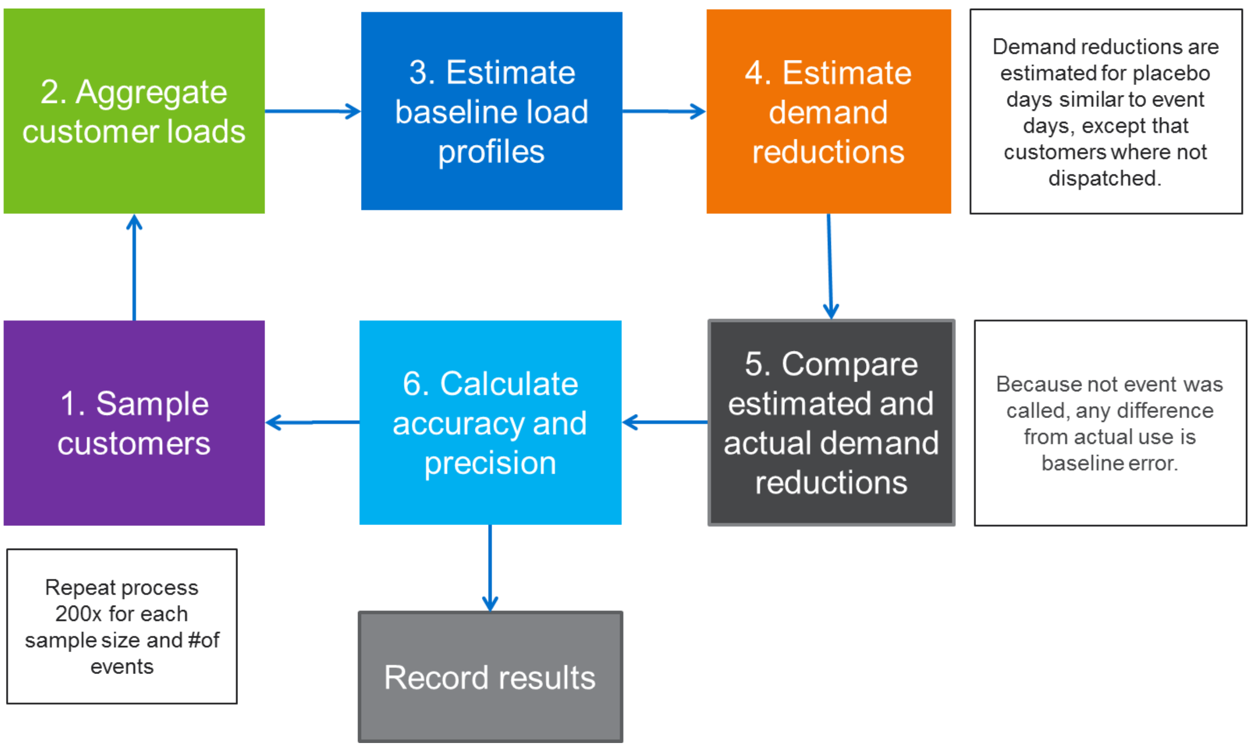
The current 10 of 10 PDR baseline methodology relies upon historical non-event day data in order to estimate a baseline. It may be challenging to find 10 previous non-event days for resources which are frequently dispatched during a period within a reasonable proximity of the event day. In particular, behind the meter storage which is not separately metered and participating in a PDR or RDRR product may participate frequently in the market. The working group explored how the load impact of frequently dispatched resources can be accurately estimated using only data from the premise. Cases in which meter generator output is available and used for settlement will be considered out of the scope of this working group because it has been addressed in the ESDER Phase 1 initiative. Research was conducted to examine how many days are necessary to establish an accurate baseline.

2 Assessing Baseline Accuracy

To assess the accuracy of the estimated values, one needs to know the correct values. When the correct answers are known, it is possible to assess if each alternative settlement option correctly measures the demand reduction and, if not, by how much it deviates from the known values. Figure 2-1 summarizes the approach for assessing accuracy and precision. The basic approach is used to address all three primary areas of research.

The objective is to test different baselines with different samples of participants using actual data from participants in order to identify the most accurate analysis method. Baseline accuracy is assessed on placebo days, which are treated as event days. Because no event was called, any deviation between the baseline and actual loads is due to error.

Figure 2-1: Method for Testing Baseline Accuracy



The process is repeated hundreds of times, using slightly different samples – a procedure known as bootstrapping – to construct the distribution of baseline errors. In addition, the accuracy of the baselines is tested at granular geographic levels, such as subLAPs, to mimic market settlement. A key question is the degree to which more or less aggregation influences the accuracy and precision of the estimates. This is assessed by repeating the below process using different subsets of customers so the relationship between the amount of aggregation and baseline accuracy is quantified. Another important question is how high frequency dispatch, which limits baseline days, affects baseline accuracy. This is assessed by repeating the same process described below for different number of event days per year, thus producing a plot of accuracy and precision as a function of the number of events.

2.1 Metrics of Identifying Suitable Baselines

For both the accuracy of the baseline and the demand reduction, the BAWG identified the best baselines as those that are both accurate and precise. The figure below illustrates the difference between accuracy and precision. An ideal model is both accurate and precise (example #1). Baselines can be accurate but imprecise when errors are large but cancel each other out (#2). They can also exhibit false precision when the results are very similar for individual events but are biased (#3). The worst baselines are both imprecise and inaccurate, i.e. the individual event results vary substantially and they are also biased.

Figure 2-2: Precision versus Accuracy (Lack of Bias)

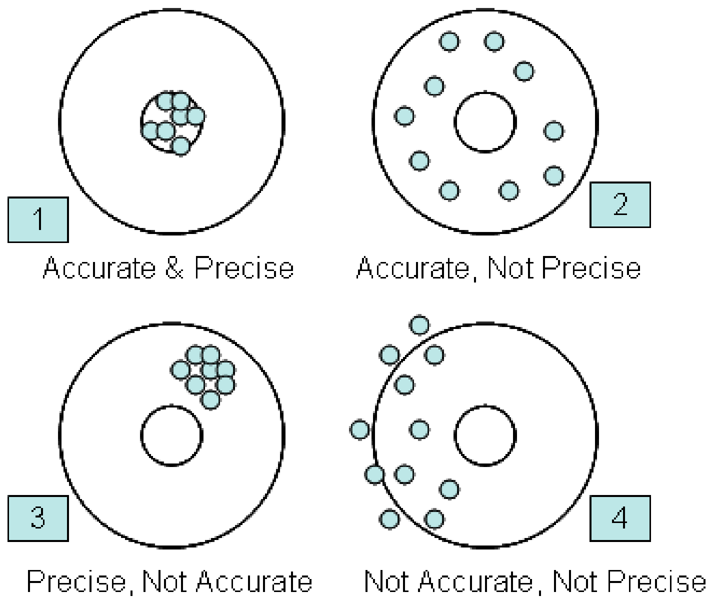


Table 2-1 summarizes metrics for accuracy (bias) and precision (goodness-of-fit) that were produced to assess the different baseline alternatives. Bias metrics measure the tendency of different approaches to over or under predict (accuracy or lack of bias) and are measured over multiple days. The BAWG used the mean percent error since it describes the relative magnitude and direction of the bias. A negative value indicates a tendency to under-predict and a positive value indicates a tendency to over-predict. This tendency is best measured using multiple days. Baselines that exhibit substantial bias were eliminated from consideration.

Precision metrics describe the magnitude of errors for individual events days and are always positive. The closer they are to zero, the more precise the results. The primary metric for precision was CVRMSE, or normalized root mean squared error. Among baselines which exhibit little or no bias, more precise metrics will be favored. Last, but not least, multiple baselines can prove to be both relatively accurate and precise. In which case, the BAWG has submitted its recommendation based on practical considerations such ease of implementation or potential for gaming.

Table 2-1: Accuracy and Precision Metrics Used to Identify Best Performing Baselines

Type of Metric	Metric	Description	Mathematical Expression
Accuracy (Bias)	Mean Percent Error (MPE)	Indicates the percentage by which the measurement, on average, over or underestimates the true demand reduction.	$MPE = \frac{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)}{\bar{y}}$
Precision (Goodness-of-Fit)	Mean Absolute Percentage Error (MAPE)	Measures the relative magnitude of errors across event days, regardless of positive or negative direction.	$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{\hat{y}_i - y_i}{y_i}\right\rvert$
	CV(RMSE)	This metric normalizes the RMSE by dividing it by the average of the actual demand reduction.	$CV(RMSE) = \frac{RMSE}{\bar{y}}$

2.2 Baselines Included for Testing

There are a variety of approaches for measuring the magnitude of demand reduction with different degrees of complexity, data sources, and metering requirements. In addition, each method can be varied based on differences in the number of eligible days used to develop baselines, the type of days used to develop baselines, caps on the magnitude of adjustments, use of different sample sizes, and the granularity of estimates. At a high level, however, the settlement methods under consideration by the BAWG can be classified under three broad categories:

- Control Groups — An ideal control group has nearly identical load patterns in aggregate and experiences the same weather patterns and conditions. The only difference is that on some days, one group has loads curtailed while the control group does not. The control group is used to establish the baseline of what load patterns would have been absent the curtailment event. This approach is the primary method for settlement of residential AC cycling and thermostat programs by Texas' system operator, ERCOT. There are three basis ways to establish control valid control groups: random assignment of customers; random assignment of clusters (for one-way devices that are not directly addressable) and matching.

- Day Matching — Day-matching baselines estimate what electricity use would have been in the absence of curtailment by relying on electricity use in the days leading up to the event. It does not include information from a control group. A subset of non-event days in close proximity to the event day are identified and averaged to produce baselines. A total of 13 day matching baselines are being tested.

- Weather Matching — The process for weather matching baselines is similar to day-matching except that the baseline load profile is selected from non-event days with similar temperature conditions and then calibrated with an in-day adjustment. In general, weather matching tends to include a wider range of eligible baseline days, which are narrowed to the ones with weather conditions closest to those observed during events. A total of 7 weather matching baselines are being tested.

2.2.1 Baselines methods tested

Table 2-2 and Table 2-3 provide additional details about the baselines tested. These baselines were identified by reviewing the best performing baselines for past studies, inside and outside of California, for residential, industrial, and commercial loads. For each baseline, a number of baseline rules were tested for using existing customers in the BIP, Agricultural pumping, residential air conditioner, and commercial air conditioner customers. These include rules include various combinations of baseline adjustment hours, adjustments caps and, when possible, assessment of accuracy and precision for actual event days (if large control groups were available) and for non-event days when net CAISO loads were high – proxy event days where the actual loads in the absence of demand response were known.

Table 2-2: Baselines Tested and Compared: Weekday

Control group	Day Matching	Weather Matching
1. Comparison of means	2. Average 3 of last 3 eligible days 3. Use 3 of last 3 eligible days; more recent days receive higher weight 4. Average the top 3 of the last 5 eligible days 5. Use top 3 of the last 5 eligible days; more recent days receive higher weight 6. Average 3 of last 5 eligible days and adjust upward by 5% for all customers 7. Average top 4 of the last 5 eligible days 8. Average top 5 of the last 5 eligible days 9. Average top 3 of the last 10 eligible days 10. Average top 5 of the last 10 eligible days 11. Average 10 of the last 10 eligible days 12. Average top 3 of the last 20 eligible days 13. Average top 5 of the last 20 eligible days 14. Average top 10 of the last 20 eligible days	15. Average 3 days with most similar weather during the last three months 16. Average 4 days with most similar weather during the last three months 17. Average 5 days with similar weather during the last three months 18. Assign days with high temperatures exceeding 80°F to 3 bins based on maximum temperature; baseline equals the average peak-period load on non-event days in a similar bin 19. Assign days with high temperatures exceeding 80°F to 3 bins based on CDD for the day; baseline equals the average peak-period load on non-event days in a similar bin 20. Assign days with high temperatures exceeding 80F to 3 bins based on the total CDH for the day; baseline equals the average peak-period load on non-event days in a similar bin

Table 2-3: Baselines Tested and Compared: Weekend

Control Group	Day Matching	Weather Matching
▪ Comparison of means	▪ 1/1 ▪ 1/2, 2/2 ▪ 1/3, 2/3, 3/3, 3/3 weighted ▪ 1/4, 2/4, 3/4, 4/4, ▪ 1/5, 2/5, 3/5, 3/5 weighted, 4/5, 5/5	▪ Matching baselines based on: – average temperature – sumCDH – maximum temperature ▪ Match on 1-5 days out of 8 prior weekend lookback

2.2.2 Same-Day Adjustments

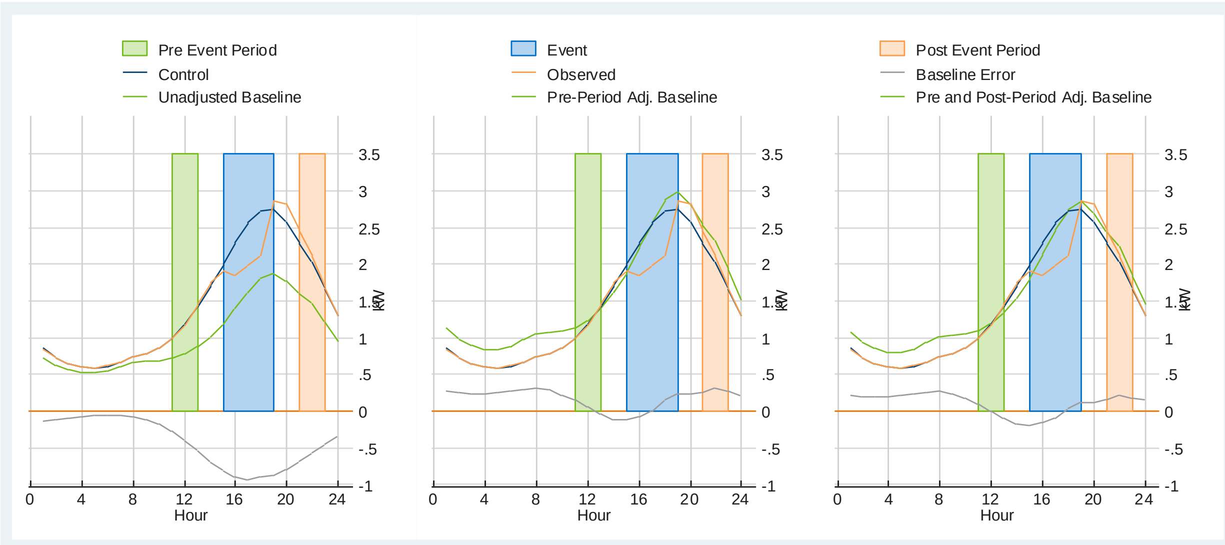
For all baseline methods, the analysis tested unadjusted baselines and the use of same-day adjustments with caps of 20%, 30%, 40%, 50%, 200%, and unlimited caps in addition to no adjustment. Same-day adjustments were tested both using pre-event data only as well as both pre- and post-event adjustments combined. Same-day adjustments calibrate the baseline to the observed non-event hours on the event day to improve precision and accuracy. Including a post-event adjustment in addition to the pre-event adjustment can scale the baseline up or down to capture additional information about the event day conditions. In both cases, the adjustments calibrate the baseline based on hours leading up to the event and after the event, with a buffer between the calibration period and the actual event.

Baseline estimates of electricity use during an event period can be adjusted up or down based on electricity use patterns during the hours leading up to an event or during both pre- and post-event hours. This procedure is known as *same-day adjustment*. If, during adjustment hours, the baseline is less than the actual load, it is adjusted upwards. Similarly, if the baseline is above the actual load in the adjustment hours, it is adjusted downwards. To adjust the load, the initial baseline value is multiplied by the ratio between the unadjusted baseline and the actual load during adjustment hours. In other words, the baseline is calibrated to match actual usage patterns in the hours leading up to the event as well as the post-event hours. In the case where both a pre- and post-event adjustment used, the calibration window includes hours both before and after the event, though the method for making the adjustment is the same. To avoid contamination of the baseline with perturbed event hours, the BAWG recommends a two-hour buffer be used for both pre- and post-event adjustments. This buffer period reduces the risk of this contamination by allowing pre-cooling and snapback to occur in the hours directly before and after the event without using those hours to adjust the baseline.

Figure 2-3 illustrates the baseline adjustment process. In the example, the event occurs from 3 PM to 6PM. With two hour buffers both before and after the event, the adjustment windows are 11AM-1PM and 8PM-10PM. The green line in each graph is the baseline, unadjusted, adjusted with the pre-event period only or adjusted with both the pre- and post-event period. The orange line is the observed load on the event day, while the black line indicates the counterfactual (modeled here by a control group). The ratio of the observed (orange) loads during the pre-event adjustment window is applied to the baseline in the center graph, while the ratio of the average observed compared to baseline loads for both the pre- and post-event periods is shown in the rightmost graph. The graph on the left shows the unadjusted result.

All the recommended baselines will have an adjustment period that includes two pre-event and two post-event hours (4 hours total), each with a two hour buffer from the event. If an event is called from 2pm to 4pm, the pre-event buffer window will be from 12am to 2pm and the post-event buffer window will be 4pm to 6pm. The pre-event buffer ensures that the adjustment window is free of any load increases that could be associated with pre-cooling, while the post-event buffer allows the increased loads associated with event snapback to diminish without contaminating the adjustment windows.

Figure 2-3: Example of Baseline Same-day Adjustment



If the difference between the unadjusted baseline and the actual load is truly due to baseline estimation error, the adjustment process reduces those errors. Same-day adjustments are often capped to reduce the variance of estimates and to limit the potential for manipulation of loads to influence baselines. To calculate a same-day adjustment once the unadjusted baseline has been calculated, the following steps are performed. A simple example that shows the mechanics of the adjustment, as well as the effect of different adjustment windows with an unlimited cap is shown in Table 2-4.

1. Calculate the average participant load in the adjustment window, factoring in the two-hour buffer. For example, if an event started at 3pm and finished at 6pm, the adjustment window would include the hours of 11am to 1pm and 8pm-10pm. Calculate the average baseline load (or control group load if using a control group) during the same window using the event baseline.

2. The ratio of participant kW during the adjustment window to that of the unadjusted baseline during that same window is the percentage adjustment.

3. Cap the ratio if using a cap. For example, if the adjustment ratio is 112% but the cap on adjustments is 10% (+/-1.1x), then the adjustment ratio will now be 110%. If no cap is being used, the adjustment ratio remains 112%. If the ratio is less than 1/1.10 = 0.91, then the adjustment cap is similarly limited to being 91%.

4. Apply the adjustment ratio to the unadjusted baseline for all hours on the event day.

5. Calculate load impacts as the difference between the adjusted baseline and the observed participant load.

Table 2-4: Adjustment Ratio Calculation

Value	Hours	No Adjustment	Pre-Event Adjustment	Pre- and Post-Event Adjustment
Pre Event Observed kW	11am-1pm	1.32		
Pre Event Unadj. Baseline kW		0.83		
Pre & Post Event Observed kW	8pm-10pm	2.28		
Pre & Post Event Baseline kW		1.54		
Ratio Calculation		None	=1.32/0.83	=(1.32 + 2.28)/(0.83+1.54)
Ratio		1.00	1.58	1.52
Event Period Observed kW	3pm-6pm	1.99		
Unadj. Baseline kW		1.51		
Event Period Baseline = (Unadj. Baseline x Ratio)		1.51	2.39	2.30

3 Baseline Recommendations

Table 3-1 shows the best performing baselines for residential and non-residential loads. Randomized control groups consistently outperformed day and weather matching baselines. With large enough sample sizes, between 200 and 400 participants, they were more than twice as precise as day or weather matching baselines. For this reason, control groups are recommended as a settlement options for both residential and non-residential customers. However, a day matching and a weather matching baseline are also options available to DRPs who may lack a sufficiently large customer base to develop a control group. The baseline option for any portfolio of resources needs to be specified for the month, in advance, and cannot be modified after the fact.

Table 3-1: Recommended Baselines for CAISO Settlement[1]

Customer Segment[2]	Weekday	Baselines Recommended	Adjustment Caps
Residential	Weekday	Control group	+/- 40%
		4 day weather matching using maximum temperature	+/- 40%
		Highest 5/10 day matching	+/- 40%
	Weekend	Control group	+/- 40%
		4 day weather matching using maximum temperature	+/- 40%
		Highest 3/5 weighted day matching	+/- 40%
Non-residential	Weekday	Control Group	+/- 40%
		4 day weather matching using maximum temperature	+/- 40%
		10/10 day matching	+/- 20%
	Weekend	Control group	+/- 40%
		4 day weather matching using maximum temperature	+/- 40%
		4 eligible days immediately prior (4/4)	+/-20%

Baseline calculations require multiple steps and definition of rules. For clarity, this section presents the baseline calculation processes and rules for control groups, weather matching baselines, and day matching baselines. Appendix A provides an applied example of control group validation and an example of how the baseline is calculated with a control group. 0 includes an applied example of a day matching baseline (the weekend residential baseline). Appendix D provides an applied example of a weather matching baseline.

3.1 Control Group Baselines

Control groups involve using a set of customers who did not experience events to establish a baseline. A control group should be made of customers who have nearly identical load patterns and experience the

[1] In the case of PDR resources that combine residential and non-residential customers, the aggregate baselines for the two customer groups should be calculated separately using the appropriate baseline for residential and non-residential customers, then added together to represent the full resource. This subdivision is not necessary if the baseline method for both residential and non-residential customers is the same, as is the case for the current recommended weather matching baselines.

[2] Residential and non-residential designations are based on customer rate class from that customer's local distribution company. That is, if a customer is served under a non-residential rate from it's LDC, that customer is classified as a non-residential customer.

same weather patterns and conditions as the resource's customers who are dispatched. During event days, the difference is that one group, known as the treatment group, experienced event dispatch while the control group did not.

Table 3-2 summarizes the control group process and rules. The process and baseline rules are identical for residential and non-residential customers and for weekdays and weekends. Section 6 includes additional discussion regarding the implementation of control group baselines. Instructions for demonstrating control group equivalence, with applied examples, are also included in the appendix to this document.

Table 3-2: Control Group Baseline Process and Rules

Component	Explanation
Baseline process	1. Determine the method for developing the control group 2. Identify the control group customers 3. Narrow data to hours and days required for validation checks (see validation options) 4. Calculate average customer loads for each hour of each day 5. Drop CAISO event days and utility program event days for programs the resource or control customers participate in. 6. Validate on the schedule described in 'Validation Options' below. Conduct validation checks and ensure all of the following requirements are met for: a. Sufficient sample size – 150 customer or more b. Lack of bias - see Section 6 c. Precision – see Section 6 7. Submit information about which sites designated as a control group and which sites will be dispatched to CAISO in advance. 8. Submit the validation checks to CAISO. 9. For event days: a. Calculate the control group average customer load for each hour of event day b. Calculate the dispatch group average customer load for each hour of the event day c. Subtract the control group load (a) from the treatment group load (b) for each hour of the event day. The difference is the change in energy use for the average customer attributable to the event response, known as the load impact. d. Multiply the load impact for each hour by the number of customers controlled or dispatched. 10. Submit summary results to CAISO and store code, analysis datasets, and results datasets. 11. Update control group validation for changes in the resource customer mix of more than +/-10% or to remain compliant with seasonal or rolling window validation requirements.
Event period	Per CAISO, the event period includes any phase-in or phase-out ramp defined by the schedule coordinator, in addition to hours where the resource is dispatched.
Method for control group development	List the method used to develop the control group – random assignment of site, random assigned of clusters, matched control group, or other. For random assignment, please retain the randomization code and set a random number generator seed value.
Replication and Audit	Control group equivalence and event days calculation are subject to audit. The results must be reproducible. The underlying customer level data, randomization files, and validation code, and event day analysis code must be retained for 3 years and be made available the CAISO within 10 business days of a request. In the case where

Component	Explanation
	the California ISO deems it necessary, DRPs will be required to securely provide the control and treatment group's interval data to recreate the bias regression coefficient and CVRMSE to ensure they meet the criteria
Validation options	Validation is performed by the DRP and subject to audit by CAISO. The validation method uses 75-day lookback period with a 30-day buffer. Validation is required as described in note e, below. The 75 days selected for validation should be chosen such that the validation is complete prior to finalizing the control group to act as the designated baseline method for that resource. a. 30 days used to collect and validate the groups b. Prior 45 days used for the validation (t-31 to t-75) c. Candidate validation days used to establish control group similarity are either non-event weekdays (if the resource is dispatched only on weekdays) or all non-event days (if the resource can be dispatched on any day) d. A minimum of 20 candidate days are required to be in the validation period. If there are not 20 non-event validation days, extend the validation period backwards (t-76 and further) until there are 20 candidate days in the validation period. e. Requires validation check updates every other month if the number of accounts in the resource does not change more than ± 10%. If the number of accounts changes by more than ± 10%, the control group must be validated monthly. f. If the validation fails, the control group method is unavailable for that resource unless the control group is updated and revalidated. Control groups may be updated monthly. g. 90% of the population must be in both the validation period and the active period
Aggregation of Control Groups across Sub Load Aggregation Points (subLAPs)	Aggregation of control groups is permissible across different subLAPs; however the same performance on intra-subLAP equivalence checks must be demonstrated. While sourcing a control group from a region with similar weather and customer mix conditions is not explicitly mandated, considerations for these attributes that affect load may help in developing an appropriate control group.
Rotation of control groups	The assignment to treatment and control groups can be updated on a monthly basis; however this assignment must be completed prior to any events. Validation of new control groups must also be completed prior to any events in concurrence with any new control group development. The assignment cannot be changed once set for the month and cannot be changed after the fact

3.2 Weather Matching Baselines

Weather-matching baselines estimate what electricity use would have been in the absence of dispatch (the baseline) by relying exclusively on electricity use data for customers who were dispatched. The load patterns during a subset of non-event days with the most similar weather conditions are used to estimate the baseline for the event day. Weather matching baselines do not include information from an external control group.

Table 3-3: Residential Weather Matching Baseline Process and Rules

	Weekday Baseline 4 Day Matching Using Daily Maximum Temperature	Weekend Baseline 4 Day Matching Using Daily Maximum Temperature
Baseline calculation process	1. Identifying eligible baseline days that occurred prior to an event 2. Calculate the aggregate hourly participant load on the event day and on each eligible baseline day during the event period hour. 3. Calculate the resource's participant weighted temperatures for each hour of each event day and eligible baseline day 4. Select the baseline days out of the pool of eligible days 5. Average hourly customer loads across the baseline days to generate the unadjusted baseline. 6. Calculate the same-day adjustment ratio based on the adjustment period hours. 7. If the same day adjustment ratio exceeds adjustment limit, limit the adjustment ratio to the cap. 8. Apply the same day adjustment ratio to the overall unadjusted baseline to produce the adjusted baseline. Application of the baseline adjustment is not optional. It must be employed to calibrate the unadjusted baseline. 9. Calculate the demand reduction as the difference between the adjusted baseline and actual electricity use for each event hour	
Eligible baseline days	Weekdays, excluding event days and federal holidays, in the 90 days immediately prior to the event.	Weekends and federal holidays, excluding event days, in the 90 days immediately prior to the event
Baseline day selection criteria	Rank eligible days based on how similar daily maximum temperature is to the event day	Rank eligible days based on how similar daily maximum temperature is to the event day
Number of days selected to develop baseline	4 days with the closest daily maximum temperature	4 days with the closest daily maximum temperature
Calculation of temperatures	1. Map the resource sites to pre-approved National Oceanic Atmospheric Association weather station based on zip code and the mapping included as Appendix B 2. Calculate the participant-weighted weather for each hour of each event and eligible baseline day. That is the weather for each relevant weather station is weighted based on the share of participant associated with the specific weather station. 3. Calculate the average temperature or daily maximum temperatures across all 24 hours in both the event day and eligible baseline days.	
Event	Per CAISO, the event period includes any phase-in or phase-out ramp defined by the schedule coordinator, in addition to hours where the resource is dispatched.	
Unadjusted baseline	The hourly average of the resource's electric load during baseline days. The unadjusted baseline includes all 24 hours in day.	
Adjustment hours	Two hours immediately prior to the event period with a two hour buffer before the event and two hours after the event with a two hour buffer. For example, if an event went from 1pm to 4pm, the adjustment hours would be 9am-11am and 6-8pm.	
Same day adjustment ratio	Calculate the ratio between the resources load and the unadjusted baseline during the adjustment hours. $\text{Adjustment ratio} = \frac{\text{Total kWh during adjusment hours}}{\text{Unadjusted baseline kWh over adjustment hours}}$	
Adjustment Limit	Cap the ratio between +/- 1.4x. If the ratio is larger than 1.4, limit it to 1.4. If the ratio is less than 1/1.4 = 0.71, limit it to 0.71	
Adjusted baseline	Apply the capped same day adjustment ratio to the unadjusted baseline to calculate the final adjusted baseline. The ratio is applied to all 24 hours of the unadjusted baseline	

Table 3-4: Non-Residential Weather Matching Baseline Process and Rules

<table>
<tr><th></th><th>Weekday Baseline
4 Day Matching Using Daily Maximum Temperature</th><th>Weekend Baseline
4 Day Matching Using Daily Maximum Temperature</th></tr>
<tr><td>Baseline calculation process</td><td colspan="2">10. Identifying eligible baseline days that occurred prior to an event
11. Calculate the aggregate hourly participant load on the event day and on each eligible baseline day during the event period hour.
12. Calculate the resource’s participant weighted temperatures for each hour of each event day and eligible baseline day
13. Select the baseline days out of the pool of eligible days
14. Average hourly customer loads across the baseline days to generate the unadjusted baseline.
15. Calculate the same-day adjustment ratio based on the adjustment period hours.
16. If the same day adjustment ratio exceeds adjustment limit, limit the adjustment ratio to the cap.
17. Apply the same day adjustment ratio to the overall unadjusted baseline to produce the adjusted baseline. Application of the baseline adjustment is not optional. It must be employed to calibrate the unadjusted baseline.
18. Calculate the demand reduction as the difference between the adjusted baseline and actual electricity use for each event hour</td></tr>
<tr><td>Eligible baseline days</td><td>Weekdays, excluding event days and federal holidays, in the 90 days immediately prior to the event.</td><td>Weekends and federal holidays, excluding event days, in the 90 days immediately prior to the event</td></tr>
<tr><td>Baseline day selection criteria</td><td>Rank eligible days based on how similar daily maximum temperature is to the event day</td><td>Rank eligible days based on how similar daily maximum temperature is to the event day</td></tr>
<tr><td>Number of days selected to develop baseline</td><td>4 days with the closest daily maximum temperature</td><td>4 days with the closest daily maximum temperature</td></tr>
<tr><td>Calculation of temperatures</td><td colspan="2">4. Map the resource sites to pre-approved National Oceanic Atmospheric Association weather station based on zip code and the mapping included as Appendix B
5. Calculate the participant-weighted weather for each hour of each event and eligible baseline day. That is the weather for each relevant weather station is weighted based on the share of participant associated with the specific weather station.
6. Calculate the average temperature or daily maximum temperatures across all 24 hours in both the event day and eligible baseline days.</td></tr>
<tr><td>Event</td><td colspan="2">Per CAISO, the event period includes any phase-in or phase-out ramp defined by the schedule coordinator, in addition to hours where the resource is dispatched.</td></tr>
<tr><td>Unadjusted baseline</td><td colspan="2">The hourly average of the resource’s electric load during baseline days. The unadjusted baseline includes all 24 hours in day.</td></tr>
<tr><td>Adjustment hours</td><td colspan="2">Two hours immediately prior to the event period with a two hour buffer before the event and two hours after the event with a two hour buffer. For example, if an event went from 1pm to 4pm, the adjustment hours would be 9am-11am and 6-8pm.</td></tr>
<tr><td>Same day adjustment ratio</td><td colspan="2">Calculate the ratio between the resources load and the unadjusted baseline during the adjustment hours.
$\text{Adjustment ratio} = \frac{\text{Total kWh during adjusment hours}}{\text{Unadjusted baseline kWh over adjustment hours}}$</td></tr>
<tr><td>Adjustment Limit</td><td colspan="2">Cap the ratio between +/- 1.4x. If the ratio is larger than 1.4, limit it to 1.4. If the ratio is less than 1/1.4 = 0.71, limit it to 0.71</td></tr>
<tr><td>Adjusted baseline</td><td colspan="2">Apply the capped same day adjustment ratio to the unadjusted baseline to calculate the final adjusted baseline. The ratio is applied to all 24 hours of the unadjusted baseline</td></tr>
</table>

3.3 Day Matching Baselines

Day-matching baselines also estimate what electricity use would have been in the absence of dispatch (the baseline) by relying exclusively on electricity use data for customers who were dispatched. The load patterns during a subset of non-event days are used to estimate the baseline for the event day.

Table 3-5: Residential Day Matching Baseline Process and Rules

	Weekday Baseline Highest 5 of 10	Weekend Baseline Highest 3 of 5 weighted
Baseline calculation process	1. Identifying eligible baseline days that occurred prior to an event 2. Calculate the aggregate hourly participant load for the event day and for each eligible baseline day 3. Calculate total MWh during the event period for each eligible baseline day 4. Rank the baseline days from largest to smallest based on MWh consumed over the event period 5. Select the baseline days out of the pool of eligible days 6. Average hourly customer loads across the baseline days to generate the unadjusted baseline. Apply weighted average, if appropriate. 7. Calculate the same-day adjustment ratio based on the adjustment period hours. 8. If the same day adjustment ratio exceeds adjustment limit, limit the adjustment ratio to the cap. 9. Apply the same day adjustment ratio to the overall unadjusted baseline to produce the adjusted baseline. Application of the baseline adjustment is not optional. It must be employed to calibrate the unadjusted baseline. 10. Calculate the demand reduction as the difference between the adjusted baseline and actual electricity use for each event hour.	
Eligible baseline days	10 weekdays immediately prior to event, excluding event days and federal holidays	5 weekend days, including federal holidays, immediately prior to the event
Baseline day selection criteria	Rank days for largest to smallest based on MWh over the event period, pick the top 5 days	Rank days for largest to smallest based on MWh over the event period, pick the top 3 days
Application of weights (if needed)	Not applicable	1. 50% - Highest load day 2. 30% - 2nd Highest load day 3. 20% - 3rd Highest load day
Event	Per CAISO, the event period includes any phase-in or phase-out ramp defined by the schedule coordinator, in addition to hours where the resource is dispatched.	
Unadjusted baseline	The weighted hourly average of the resource's electric load during baseline days. The unadjusted baseline includes all 24 hours in day.	
Adjustment hours	Two hours immediately prior to the event period with a two hour buffer before the event and two hours after the event with a two hour buffer. For example, if an event went from 1pm to 4pm, the adjustment hours would be 9am-11am and 6-8pm.	
Same day adjustment ratio	Calculate the ratio between the resources load and the unadjusted baseline during the adjustment hours. $\text{Adjustment ratio} = \frac{\text{Total kWh during adjusment hours}}{\text{Unadjusted baseline kWh over adjustment hours}}$	
Adjustment Limit	Cap the ratio between +/- 1.4x. If the ratio is larger than 1.4, limit it to 1.4. If the ratio is less than 1/1.4 = 0.71, limit it to 0.71	Cap the ratio between +/- 2x. If the ratio is larger than 2.0, limit it to 2.0. If the ratio is less than 1/2 = 0.50, limit it to 0.50
Adjusted baseline	Apply the capped same day adjustment ratio to the unadjusted baseline to calculate the final adjusted baseline. The ratio is applied to all 24 hours of the unadjusted baseline	

Table 3-6: Non-Residential Day Matching Baseline Process and Rules

	Weekday Baseline Highest 10 of 10	Weekend Baseline Highest 4 of 4
Baseline calculation process	11. Identifying eligible baseline days that occurred prior to an event 12. Calculate the aggregate hourly participant load for the event day and for each eligible baseline day 13. Calculate total MWh during the event period for each eligible baseline day 14. Rank the baseline days from largest to smallest based on MWh consumed over the event period 15. Select the baseline days out of the pool of eligible days 16. Average hourly customer loads across the baseline days to generate the unadjusted baseline. Apply weighted average, if appropriate. 17. Calculate the same-day adjustment ratio based on the adjustment period hours. 18. If the same day adjustment ratio exceeds adjustment limit, limit the adjustment ratio to the cap. 19. Apply the same day adjustment ratio to the overall unadjusted baseline to produce the adjusted baseline. Application of the baseline adjustment is not optional. It must be employed to calibrate the unadjusted baseline. 20. Calculate the demand reduction as the difference between the adjusted baseline and actual electricity use for each event hour.	
Eligible baseline days	10 weekdays immediately prior to event, excluding event days and federal holidays	4 weekend days, including federal holidays, immediately prior to the event
Baseline day selection criteria	Keep all 10 eligible days	Keep all 4 eligible days
Application of weights (if needed)	Not applicable	Not applicable
Event	Per CAISO, the event period includes any phase-in or phase-out ramp defined by the schedule coordinator, in addition to hours where the resource is dispatched.	
Unadjusted baseline	The weighted hourly average of the resource's electric load during baseline days. The unadjusted baseline includes all 24 hours in day.	
Adjustment hours	Two hours immediately prior to the event period with a two hour buffer before the event and two hours after the event with a two hour buffer. For example, if an event went from 1pm to 4pm, the adjustment hours would be 9am-11am and 6-8pm.	
Same day adjustment ratio	Calculate the ratio between the resources load and the unadjusted baseline during the adjustment hours. $\text{Adjustment ratio} = \frac{\text{Total kWh during adjusment hours}}{\text{Unadjusted baseline kWh over adjustment hours}}$	
Adjustment Limit	Cap the ratio between +/- 1.2x. If the ratio is larger than 1.2, limit it to 1.2. If the ratio is less than 1/1.2 = 0.83, limit it to 0.83	Cap the ratio between +/- 1.2x. If the ratio is larger than 1.2, limit it to 1.2. If the ratio is less than 1/1.2 = 0.83, limit it to 0.83
Adjusted baseline	Apply the capped same day adjustment ratio to the unadjusted baseline to calculate the final adjusted baseline. The ratio is applied to all 24 hours of the unadjusted baseline	

4 Implementation of Control Group Settlement Methodology

Randomized control groups consistently outperformed day and weather matching baselines for residential and commercial AC cycling programs during testing. With large enough sample sizes, between 200 and 400 participants, they were more than twice as precise as day or weather matching baselines. For this reason, the BAWG recommends that control groups be one of the settlement options for both residential and non-residential customers.

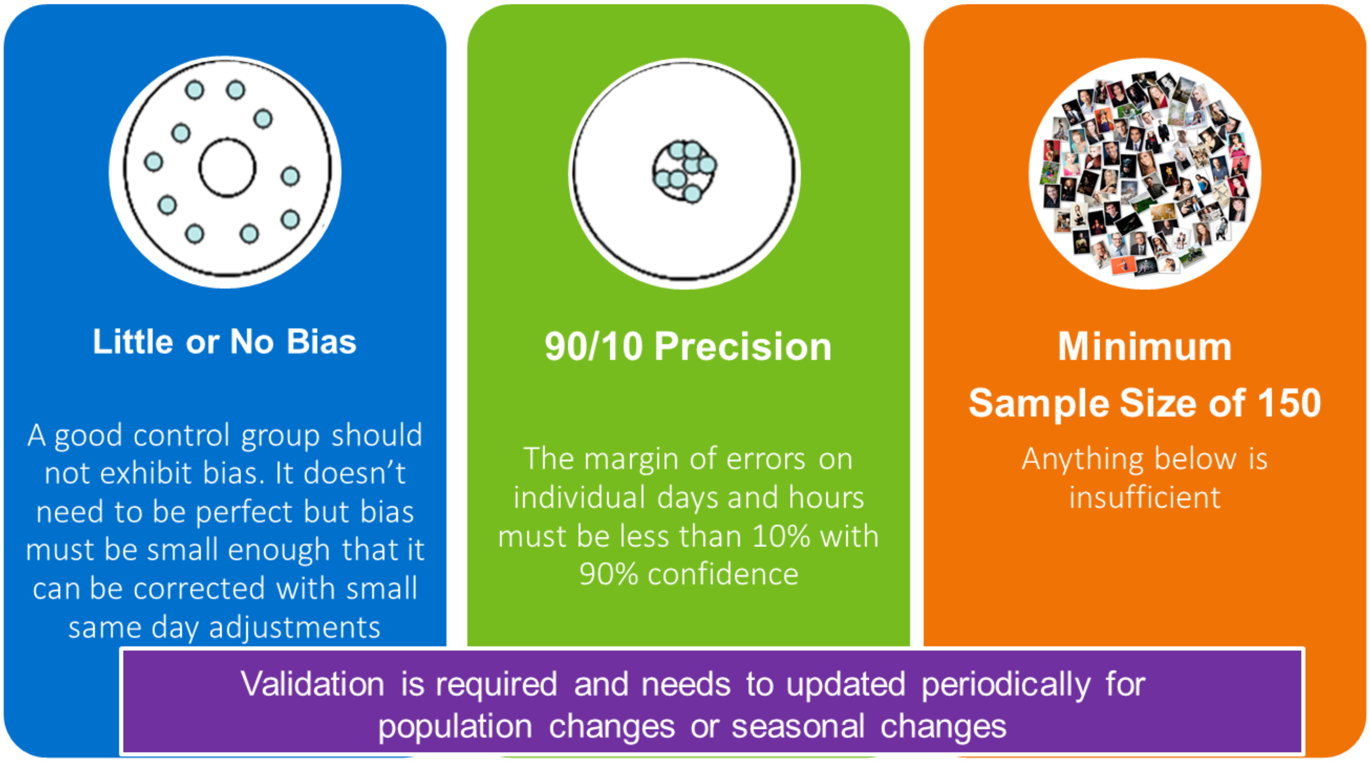
Control groups involve using a set of customers who did not experience events to establish a baseline. A control group should be made of customers who are statistically indistinguishable from the participant group on non-event days to act as a comparison on event days, instead of relying on participants' past performance. There are many ways to develop a control group, including random assignment and statistical or propensity score matching. The rules were intentionally developed so as not preclude use of alternate methods for selecting a control group. There are, however, multiple issues surrounding the development of matched control groups (e.g. data security, equal access to non-participant data, legality, and cost) that were outside of the BAWG scope. Currently, all DRP are able to establish a control group by randomly assigning and withholding a subset of participant resource sites from dispatch. However, not all DRP's have equal access to utility smart meter data for non-participants, which is necessary for development of matched control groups.

The best approach for developing a valid control group is to randomly assign a subset of customers in a resource portfolio to serve as the control group. This requires withholding a subset of participants from event dispatch, thus establishing the baseline. Because of random assignment, there are no systematic differences between the group that is dispatched and the control group, except the event dispatch. With sufficient sample sizes, differences due to random chance are minimized and the control group becomes statistically indistinguishable from the treatment group. This then means that any difference in load profiles on event days can be attributed to the effect of treatment, and that any difference between the two groups on non-event days should be negligible.

However, before a control settlement methodology can be employed it is necessary to demonstrate that the energy use of the control group is an accurate predictor of the energy use of the participants. Three high level requirements for demonstrating the validity of a control group are shown below. Instructions for demonstrating control group equivalence follow, with applied examples in the appendix to this document. Once a suitably accurate and precise baseline has been developed, it can be adjusted using same-day adjustments as described at the end of this section. However, it is the unadjusted baseline that must meet the accuracy, precision and sample size criteria.

Figure 4-1 demonstrates the three key principles for the development and validation of control groups. They must exhibit little or no bias, must be sufficiently precise, and be large enough to represent the treatment population.

Figure 4-1: Control Group Requirements



4.1 Statistical Checks Necessary to Demonstrate Control Group Validity

DRPs will need to demonstrate that the control group reflects the electricity use patterns of customers curtailed (validation). The process for demonstrating equivalence is outlined below. It is the responsibility of the DRP to develop the control group and demonstrate equivalence. The control group(s) developed are subject to audit by the CAISO.

1. The DRP Identifies a control pool of at least 150 customers to be selected via statistical matching or randomly withheld from the participant population. A single control group may be used for multiple subLAP settlement groups; however, equivalence, using the procedure outlined below, must be demonstrated for each of the treatment groups against the control group. For example, if there are five subLAPs, five equivalence checks must be completed to show that the control customers are equivalent to treatment customers in subLAPs A, B, C, D and E. Use of a different control group for each subLAP is also permitted and will be necessary if there are significant differences in weather sensitivity or other characteristics among treatment groups in different subLAPs. In those cases, equivalence must be demonstrated only between the treatment group and the control group for which it is acting as control.

2. For each resource ID, look back 75 days from when the validation occurs, and pull hourly data from the 45 earliest days (t-31 to t-75). The days included in the validation must be in this t-31 to t-75 range, excluding any days that an event has been called for this resource. If the resource is only dispatched on weekdays, the candidate weekend days may be ignored. If the resource can

be dispatched on weekdays and weekends/holidays, all non-event days must be included in the validation period. In addition, exclude event days that the customers in the resource could have participated in. If customers are dually participating in utility load modifying programs, event days of the load modifying resource may also be excluded. If there are not at least 20 available candidate days, continue looking further back (t-76 to t-85 for example) to find additional candidate days until 20 days are available for validation.

3. Average the hourly load profile for all treatment group customers and all control group customers by day and hour.

4. Filter to the appropriate hours and days. Validation is only done on the hours 12-9pm but does include weekdays, weekends, and holidays if the resource can be dispatched on those days.

5. Arrange the data in the appropriate format. For most statistical packages and Excel, regressions are easiest to perform when data is in a long format by date and hour and wide by treatment status. Note that the datasets should be separate for each treatment/control group pairing to be tested.

6. Regress average treatment hourly load against average control hourly load during event hours with no constant. This can be done in a statistical package like R or Stata, or within an Excel file or other spreadsheet application. The functional form of this model should be

$$y_{i,h}^{T} = \beta y_{i,h}^{C} + \varepsilon_{i,h}$$

   Where $y_{i,h}^{T}$ is the average kW across all treatment customers for the non-event day i and hour h, and $y_{i,h}^{C}$ is the average kW across all control customers for that same hour and day. The coefficient,β, represents the bias that exists in the control group; that is, the percent difference between the average treatment kW and the average control kW across all days and event hours. A coefficient of 1.05 means that the treatment group demand is on average 5% higher than that of the control group. Similarly, a coefficient of 0.86 means that the control group load is 86% that of the treatment group. Note that this model explicitly excludes a constant term from the regression.

7. To demonstrate lack of bias, the coefficient β should be between 0.95 and 1.05, minimizing the unadjusted absolute bias from the treatment group.

8. To demonstrate that the control group has sufficient precision, the value of the normalized root mean squared error at the 90% confidence level should be less than 10%. The normalized root mean squared error, or CVRMSE, is calculated according to

$$CV(RMSE) = \frac{\sqrt{\frac{\sum_{i,h}(y_{i,h}^{C} - y_{i,h}^{T})^2}{n}}}{(1/n)\sum_{i,h} y_{i,h}^{T}}$$

In this equation, the squared difference between treatment and control for each event hour and day is summed over all event hours and days, and then divided by the total number of event hours and days (n). The square root of that value is divided by the average treatment load across all event hours and days to normalize the error. Under the assumption that the CVRMSE is normally distributed, the 90% confidence level for this statistic is 1.645 times the CVRMSE. For example, if the CVRMSE is 0.86%, the 90% confidence level for the statistic is 1.414%.

4.2 Using Matched Control Groups to Generate a Baseline

Use of a matched control group would allow DRPs to dispatch their entire participant group during an event, while a separate group of non-participants would act as a control. Alternatively, participants that include customers both inside and outside a subLAP could act as a control group.

The BAWG is open to the possibility of a matched control group baseline option. It is the preferred option for SCE. However, PG&E, SCE, and SDG&E were concerned about customer data security, the allocation of cost to fund this option, and potential legal issues associated with having utilities involved in identifying a matched control group on behalf of other DRPs. While matched control groups are subject to the same validation criteria as randomized control groups, the use of non-participants to develop a control group is of considerable interest to DRPs that wish to dispatch their entire enrolled population during an event. However, no recommendation has been developed that would allow DRPs access to non-participant data to develop the matched control group.

However, a few agreements were reached.

- DRPs with access to non-participant interval data may have the option to utilize matched control groups. The BAWG may choose to withhold the ability to create a matched control group if the access to non-participant data is not available to all parties. These matched control groups are subject to the same validation requirements as the randomly assigned control groups, as outlined above.
- The issue of access to non-participant data is broader than its use for settlement baselines and needs be worked out at the CPUC.
- The matched control group can be updated on a monthly basis but needs to be designated in advance. It cannot be changed once it is set for the month and cannot be changed after the fact.
- The matched control group assignment is subject to audit. The purpose of audits is to assure that baselines were properly calculated and control groups met precision and validation criteria. Audits may include delivery of customer interval data with the goal of recreating bias and precision metrics assessed in the validation process.

5 Baseline Process Considerations for Settlement

5.1 Process to Calculate Sub-hourly Impacts from Hourly Baselines

The BAWG recognized that the proposed performance calculation results provided to the CAISO as Settlement Quality Meter Data ("SQMD") must be in intervals of five minutes when a PDR or RDRR offers real-time or ancillary services (non-spin and spinning reserve). The BAWG proposes to employ CAISO's current methodology for deriving these results, borrowing from the CAISO's current 10 in 10 customer load baseline methodology3. In summary, to achieve a 5-minute DR Energy Measurement 4, an hourly baseline is pro-rated to create a 5-minute baseline from which the 5-minute interval load, measured during the event, is subtracted. The CAISO would maintain its current requirement that baselines, and measured load during the event, be derived using, at maximum, a 15-minute interval load measurement when the PDR or RDRR is participating in real-time or ancillary service markets.

5.2 Negative Values Settlement

In cases where a baseline is less than the observed interval data in a given 5-minute interval, the DR energy measurements are set to zero. That is, no negative impacts are used in settlement.

6 Baseline Process Discussion

The following additional process discussion points were addressed in meetings of the full working group.

- Allowing custom or alternate baselines: CAISO does not support any recommendation for new or custom baselines.
- Who will estimate the baselines: The BAWG recommends that DRPs estimate the baselines and provide them to CAISO. CAISO will have an annual process where the DRPs attest to the accuracy of the baselines and may also audit the accuracy of the baselines on an as-needed basis.
- Managing baselines for customer transitions: Further work in this area is needed. The registration process for new PDRs needs to be fully understood by the BAWG participants to ensure that the proper recommendation is developed. A suspension period for customers transitioning to a new settlement group may be necessary to ensure there are sufficient past candidate days to develop a baseline. A method of tracking past event days for customers who transition is also required.

[3] See DRS User Guide for DR Energy Measurement Adjustment for Real Time beginning on page 160 http://www.caiso.com/Documents/DemandResponseUserGuide.pdf

[4] The resulting Energy quantity calculated by comparing the applicable performance evaluation methodology of a PDR or RDRR against its actual underlying performance for a Demand Response Event.

Appendix A Applied Examples of Control Group Validation

A.1 Using Excel

Shown below are examples of how to demonstrate equivalence between treatment and control groups in Excel. A template for performing this calculation can be found in the file called 'Randomization Validation Template.xlsx'. As described above, the steps to performing this calculation are:

1. Identify a control pool of at least 150 customers to be selected via statistical matching or randomly withheld from the participant population. Create a dataset that has the form shown in Figure A-1 with control and participant's hourly usage by date from hours ending 1 through 24.

Table A-1: Base Dataset

Participant ID	Treat	RA Season	Date	kWh1	kWh2	kWh3	kWh4	kWh5	kWh6	…	kWh23	kWh24
1	C	Winter	12/31/2014	2.00	1.11	1.91	1.29	0.78	1.25		0.97	1.44
1	C	Winter	1/1/2015	0.72	1.81	0.88	1.97	1.39	1.79		1.49	1.40
1	C	Winter	1/2/2015	0.85	0.59	1.67	0.64	0.67	1.04		2.00	1.42
1	C	Winter	1/3/2015	1.76	0.61	1.99	0.77	1.27	1.27		1.85	1.85
1	C	Winter	1/4/2015	1.60	0.66	1.55	1.08	1.86	1.57		0.68	0.83
1	C	Winter	1/5/2015	1.59	1.32	0.53	1.32	1.44	0.88		1.12	1.18
1	C	Winter	1/6/2015	1.45	1.63	1.47	1.50	1.66	0.98		1.90	0.66
2	T	Winter	12/31/2014	1.11	0.97	1.39	0.58	1.36	1.30		1.54	0.79
2	T	Winter	1/1/2015	0.65	1.04	1.38	1.31	0.81	1.68		0.80	1.47
2	T	Winter	1/2/2015	0.97	1.44	1.31	1.19	1.89	1.74		0.59	1.44
2	T	Winter	1/3/2015	1.16	1.59	1.70	1.25	1.11	1.63		0.79	0.97
2	T	Winter	1/4/2015	0.72	1.98	1.24	1.52	1.91	1.99		0.57	1.85
2	T	Winter	1/5/2015	0.56	1.20	1.19	1.34	1.33	0.50		1.23	1.38
2	T	Winter	1/6/2015	0.99	0.99	0.60	1.32	0.61	1.23		0.93	1.27
3	T	Winter	12/31/2014	1.59	1.81	0.58	1.69	1.49	1.15		0.55	1.81
3	T	Winter	1/1/2015	1.11	1.67	0.71	1.00	0.95	1.39		1.86	1.50
3	T	Winter	1/2/2015	1.71	1.54	1.26	1.40	1.67	1.52		1.90	1.67
3	T	Winter	1/3/2015	1.54	1.11	1.03	1.45	1.10	0.85		1.81	2.00
3	T	Winter	1/4/2015	1.13	0.67	1.25	0.83	1.96	1.58		0.78	0.64
3	T	Winter	1/5/2015	0.96	1.06	1.35	0.89	1.72	1.01		0.54	1.95
3	T	Winter	1/6/2015	0.99	1.35	1.32	0.75	0.82	1.16		1.08	1.11

2. Average the hourly load profile for all treatment group customers and all control group customers by day and hour.

Table A-2 Average Daily Treatment and Control Usage

Ineligible Day	Treat	RA Season	Date	kWh1	kWh2	kWh3	kWh4	kWh5	kWh6	…	kWh23	kWh24
	C	Winter	12/31/2014	2.00	1.11	1.91	1.29	0.78	1.25		0.97	1.44
Holiday	C	Winter	1/1/2015	0.72	1.81	0.88	1.97	1.39	1.79		1.49	1.40
	C	Winter	1/2/2015	0.85	0.59	1.67	0.64	0.67	1.04		2.00	1.42
Weekend	C	Winter	1/3/2015	1.76	0.61	1.99	0.77	1.27	1.27		1.85	1.85
Weekend	C	Winter	1/4/2015	1.60	0.66	1.55	1.08	1.86	1.57		0.68	0.83
	C	Winter	1/5/2015	1.59	1.32	0.53	1.32	1.44	0.88		1.12	1.18
	C	Winter	1/6/2015	1.45	1.63	1.47	1.50	1.66	0.98		1.90	0.66
	T	Winter	12/31/2014	1.35	1.39	0.98	1.14	1.42	1.23		1.05	1.30
Holiday	T	Winter	1/1/2015	0.88	1.36	1.04	1.15	0.88	1.53		1.33	1.49
	T	Winter	1/2/2015	1.34	1.49	1.28	1.29	1.78	1.63		1.25	1.56
Weekend	T	Winter	1/3/2015	1.35	1.35	1.36	1.35	1.10	1.24		1.30	1.49
Weekend	T	Winter	1/4/2015	0.92	1.33	1.25	1.18	1.93	1.79		0.68	1.24
	T	Winter	1/5/2015	0.76	1.13	1.27	1.11	1.52	0.76		0.88	1.66
	T	Winter	1/6/2015	0.99	1.17	0.96	1.04	0.72	1.19		1.01	1.19

3. Flag and remove days in which the resource is not available and event days that the customers in the resource could have participated in. Keep only hours 12pm-9pm (hour ending 13-21).

Table A-3 Average Daily Treatment and Control Usage

Treat	RA Season	Date	kWh1	kWh2	kWh3	kWh4	kWh5	kWh6	…	kWh23	kWh24
C	Winter	12/31/2014	2.00	1.11	1.91	1.29	0.78	1.25		0.97	1.44
C	Winter	1/2/2015	0.85	0.59	1.67	0.64	0.67	1.04		2.00	1.42
C	Winter	1/5/2015	1.59	1.32	0.53	1.32	1.44	0.88		1.12	1.18
C	Winter	1/6/2015	1.45	1.63	1.47	1.50	1.66	0.98		1.90	0.66
T	Winter	12/31/2014	1.35	1.39	0.98	1.14	1.42	1.23		1.05	1.30
T	Winter	1/2/2015	1.34	1.49	1.28	1.29	1.78	1.63		1.25	1.56
T	Winter	1/5/2015	0.76	1.13	1.27	1.11	1.52	0.76		0.88	1.66
T	Winter	1/6/2015	0.99	1.17	0.96	1.04	0.72	1.19		1.01	1.19

4. Arrange the data in the appropriate format.

Table A-4 Average Daily Treatment and Control Usage

Date	Hour	kWh Treatment	kWh Control
5/18/2015	13	0.93	0.58
	14	1.03	0.59
	15	1.14	0.63
	16	1.24	0.69
	17	1.33	0.79
	18	1.4	0.91
	19	1.41	0.99
	20	1.34	1
	21	1.3	1.06
5/19/2015	13	0.77	0.63
	14	0.82	0.65
	15	0.9	0.68
	16	1	0.75
	17	1.13	0.86
	18	1.26	0.98
	19	1.31	1.05
	20	1.25	1.07
	21	1.22	1.11
5/20/2015	13	0.83	0.68
	14	0.92	0.74
	15	1.03	0.8
	16	1.19	0.91
	17	1.37	1.02
	18	1.53	1.14
	19	1.58	1.2
	20	1.51	1.18
	21	1.4	1.17

5. Regress average treatment hourly load against average control hourly load during event hours with no constant, or fill out the attached template to calculate bias and precision statistics

Figure A-1: Regression and Validation Template – Bias

Date	Hour Ending	Treatment kWh	Control kWh	Numerator	Denominator
5/18/2015	13	0.93	0.58	0.539	0.336
5/18/2015	14	1.03	0.59	0.608	0.348
5/18/2015	15	1.14	0.63	0.718	0.397
5/18/2015	16	1.24	0.69	0.856	0.476
5/18/2015	17	1.33	0.79	1.051	0.624
5/18/2015	18	1.4	0.91	1.274	0.828
5/18/2015	19	1.41	0.99	1.396	0.980
5/18/2015	20	1.34	1	1.340	1.000
5/18/2015	21	1.3	1.06	1.378	1.124
5/19/2015	13	0.77	0.63	0.485	0.397
5/19/2015	14	0.82	0.65	0.533	0.423
5/19/2015	15	0.9	0.68	0.612	0.462
5/19/2015	16	1	0.75	0.750	0.563
5/19/2015	17	1.13	0.86	0.972	0.740
5/19/2015	18	1.26	0.98	1.235	0.960
5/19/2015	19	1.31	1.05	1.376	1.103
5/19/2015	20	1.25	1.07	1.338	1.145
5/19/2015	21	1.22	1.11	1.354	1.232
5/20/2015	13	0.83	0.68	0.564	0.462
5/20/2015	14	0.92	0.74	0.681	0.548
5/20/2015	15	1.03	0.8	0.824	0.640
5/20/2015	16	1.19	0.91	1.083	0.828
5/20/2015	17	1.37	1.02	1.397	1.040
5/20/2015	18	1.53	1.14	1.744	1.300
5/20/2015	19	1.58	1.2	1.896	1.440
5/20/2015	20	1.51	1.18	1.782	1.392
5/20/2015	21	1.4	1.17	1.638	1.369
5/21/2015	13	1.39	0.65	0.904	0.423
5/21/2015	14	1.6	0.65	1.040	0.423
5/21/2015	15	1.81	0.68	1.231	0.462
5/21/2015	16	2.01	0.75	1.508	0.563
5/21/2015	17	2.21	0.85	1.879	0.723
5/21/2015	18	2.36	0.98	2.313	0.960

Summed Numerator	655.5552	
Summed Denominator	638.8391	
Beta	1.026166	< Pass if between 0.95 and 1.05

$$\beta = \frac{\sum\left(Y_{i,h}^{C}\right)\left(Y_{i,h}^{T}\right)}{\sum\left(Y_{i,h}^{C}\right)^{2}}$$

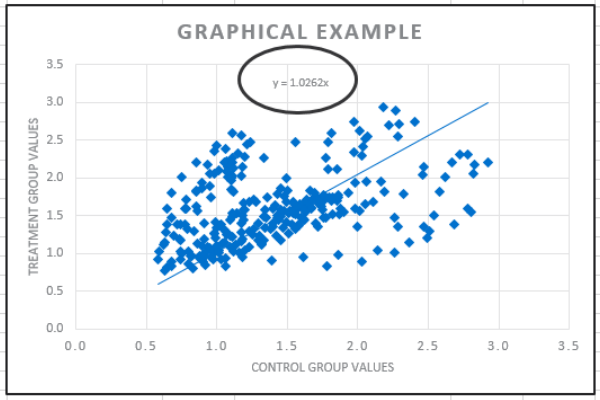


Figure A-2: Regression and Validation Template – Precision

Date	Hour Ending	Treatment kWh	Control kWh	Error Squared
5/18/2015	13	0.93	0.58	0.1225
5/18/2015	14	1.03	0.59	0.1936
5/18/2015	15	1.14	0.63	0.2601
5/18/2015	16	1.24	0.69	0.3025
5/18/2015	17	1.33	0.79	0.2916
5/18/2015	18	1.4	0.91	0.2401
5/18/2015	19	1.41	0.99	0.1764
5/18/2015	20	1.34	1	0.1156
5/18/2015	21	1.3	1.06	0.0576
5/19/2015	13	0.77	0.63	0.0196
5/19/2015	14	0.82	0.65	0.0289
5/19/2015	15	0.9	0.68	0.0484
5/19/2015	16	1	0.75	0.0625
5/19/2015	17	1.13	0.86	0.0729
5/19/2015	18	1.26	0.98	0.0784
5/19/2015	19	1.31	1.05	0.0676
5/19/2015	20	1.25	1.07	0.0324
5/19/2015	21	1.22	1.11	0.0121
5/20/2015	13	0.83	0.68	0.0225
5/20/2015	14	0.92	0.74	0.0324
5/20/2015	15	1.03	0.8	0.0529
5/20/2015	16	1.19	0.91	0.0784
5/20/2015	17	1.37	1.02	0.1225
5/20/2015	18	1.53	1.14	0.1521
5/20/2015	19	1.58	1.2	0.1444

RMSE	0.5778	
CVRMSE	0.3636	
90% Confidence Band +/-	0.5980	< Pass if < 0.1

$$CVRMSE = \sqrt{\frac{\sum\left(Y_{i,h}^{C} - Y_{i,h}^{T}\right)^{2}}{n}} \Bigg/ \left(\frac{\sum Y_{i,h}^{T}}{n}\right)$$

A.2 Applied Example of Validation Required – Using Stata

Example code that performs the control group validation can be found in the Stata do file named 'Stata Code to Validate Equivalence.do'.

The command to perform this regression is: *reg kWh_treat kWh_control, noconstant.* Both the bias value and precision value can be calculated from the cleaned average hourly kWh values as described in the code example.

Appendix B Process to Calculate Participant-Weighted Weather

B.1 Mapping of NOAA Weather Stations to ZIP codes

Weather matching baselines require weather data in order to find similar non-event days. The BAWG found that participant-weighted weather, meaning an average hourly weather profile that is the weighted average of the geographic mix of resource participants, vastly outperforms using a single weather profile for each subLAP and resource. To facilitate this process, the BAWG has put together a mapping of NOAA stations to California zip codes.

The mapping was done using distance matching by finding the closest NOAA weather station by physical distance to the centroid of each zip code. For zip codes that did not have latitude and longitude values available (the metrics used to calculate distance from the stations), a matching process was used to find the weather stations of proximate surrounding zip codes, which was then used to fill in missing values. The full list of zip codes and their associated weather stations can be found in the Excel workbook 'NOAA Station to Zip Mapping.xlsx'. This list above shall be updated by the IOUs for each of their respective territories and updated at the request of DRPs.

B.2 Calculating Participant-Weighted Weather

Once participants have been identified for a particular resource, their weather data can be compiled to calculate the participant-weighted average weather by day and hour. The process is as follows:

1. Determine the weather stations associated with the resource in question. For all the resource participants, collect their associated premise-level zip codes (ie the zip code associated with their physical location, not their billing location), and use the mapping listed above to generate a list of associated weather stations for each resource

2. Collect the last 90 days of weather data from NOAA from the weather stations in question.

    a. Data should be at the hourly level for all days and weather stations

3. Assemble the dataset of participants for the full baseline search period. The look-back period for weekday baselines is 90 days and 56 days (8 weeks) for weekend baselines. Each participant must have an associated premise zip code that indicates their physical (ie not billing) location.

4. Merge the customer-level dataset with the weather station mapping by zip code. In effect, ensure that each customer has a single weather station that is mapped to their zip code using the mapping attached above (or a subsequent update).

5. Now merge the weather data in to the customer-level dataset by weather station. This should yield a dataset that is unique by participant id, date and hour (if the dataset is long by hour).

6. Create the resource-average dataset by collapsing the participant-level dataset to an average by date and hour. No weighting is required if the dataset described in step 5 includes all the participants in the particular resource. Frequency weights should be applied to calculate the weighted average of all the weather stations in the resource (weighted by the total number of participants that are mapped to each weather station) if the dataset does not include all participants.

7. The dataset is participant-weighted and can be merged to the average hourly load data by date and hour to calculate weather-matching baselines.

Appendix C Detailed Day-Matching Calculation Process

A detailed example of how to calculate a weather matching baseline is described in the Excel workbook named 'Example_Day_Match_Workbook.xlsx'. The steps are as follows:

0. Start with hourly interval data for all participants in the program, with at least 90 days of prior data. Note this is not shown in the attached example.

1. Collapse the data to the average hourly load by day for the full set of participants. The dataset should now look something like the example shown in Tab 1 of the attached document.

2. Clean the data by removing ineligible days (weekends and holidays, already excluded from this example) and other event days that the participants were dispatched for (highlighted in grey). The event day in this example, was September 10th, 2015, when the program was called between 4-7pm (hour ending 17 to hour ending 19). Note that this dataset is slightly smaller than the 90 days of eligible data, but it does not affect the calculations required for day matching.

    a. Generate the average event load. For each of the non-event days remaining in the dataset, average the hourly load for the event hours (in this case HE17-HE19) for each day.

3. Keep the last Y eligible days. The number Y refers to the denominator of the day matching baseline. If the baseline is a top 5/10, Y = 10. If the baseline is a top 3/5, as shown in the example workbook, Y = 5. These are your eligible days

4. Sort by the average event load in decreasing order, and pick the top X largest days. These are your baseline days. The X in this case refers to the numerator of the day matching baseline. For the two baseline examples listed in Step 3, X = 5 or X = 3, respectively. In the attached example, X = 3.

5. Generate the unadjusted baseline. Two options are presented in the attached example:

    a. Top 3/5 Unweighted: The three baseline days are simply averaged to generate the baseline.

    b. Top 3/5 Weighted: The closest day to the baseline receives a weight of 50%, the next closest receives a weight of 30% and the furthest receives a weight of 20%. Note that closest in this case refers to days closest to the event day, not by the average event load sorting that was done in Step 4. The weighting is applied by multiplying the % for each day to the hourly load profiles, then summing. This is a weighted average.

6. Perform the same-day adjustment as necessary.

a. Define the adjustment window periods. In the example, the event occurs between HE17and HE19 (highlighted in blue in the example). For two-hour pre- and post-event adjustment windows with a two-hour buffer, the adjustment window hours (highlighted in orange in the example) are HE13, HE14, HE22, and HE23.

b. Average the usage across those four hours for both the baseline and the event day observed load.

c. Calculate the adjustment ratio by dividing the baseline average window value by the observed average window value. In the example, the baseline has an adjustment window value of 1.49kW and the event adjustment window value is 1.76. The ratio is then 1.18.

d. Cap the ratio at the required level. If the cap is 1.4x, as in the example, the following logic applies:

    i. If the ratio is less than 1/1.4 = 0.71, the capped ratio is now set to 0.71.

    ii. If the ratio is between 0.71 and 1.4, the ratio remains as is.

    iii. If the ratio is greater than 1.4, the capped ratio is now set to 1.4.

e. Apply the capped ratio to each hour of the baseline by multiplying the capped ratio by the hourly baseline values for each hour

f. The profile obtained in step 6e is the baseline.

7. DR Energy Measurements are calculated as the difference between the baseline and the observed load, which have already been decomposed to the 5-minute increment level, such that load reductions relative to the baseline are positive. Load increases, when the baseline is less than the observed load, should be set to 0 for settlement purposes.

Appendix D Detailed Weather-Matching Calculation Process

A detailed example of how to calculate a weather matching baseline is described in the Excel workbook named 'Example_Weather_Match_Workbook.xlsx'. The steps are as follows:

0. Start with hourly interval data for all participants in the program, with at least 90 days of prior data. Note this is not shown in the attached example.

1. Collapse the data to the average hourly load by day for the full set of participants. The dataset should now look something like the example shown in Tab 1 of the attached document.

2. Clean the data by removing ineligible days (weekends and holidays, already excluded from this example) and other event days that the participants were dispatched for (highlighted in grey). The event day in this example, was September 10th, 2015, when the program was called between 4-7pm (hour ending 17 to hour ending 19). Note that this dataset is slightly smaller than the 90 days of eligible data, but it does not affect the calculations required for day matching.

    a. Also generate the weather variable of interest for the baseline – either the maximum hourly temperature or the average daily temperature

    b. Drop any days that occur AFTER the event day for which the baseline is being calculated.

3. Sort the dataset by how similar the eligible days are to the event day, by calculating the absolute value of the difference between the event day average (or maximum) temperature and the eligible day's average (or maximum) temperature.

4. Sort by the weather variable absolute difference in decreasing order, and pick the top X largest days. These are your baseline days. The X in this case refers to number of days used to estimate the weather baseline. A 3 day weather matching baseline will have X = 3. A 5-day weather matching baseline will have X = 5.

5. Generate the unadjusted baseline by averaging the hourly kW values across the X baseline days.

6. Perform the same-day adjustment as necessary.

    a. Define the adjustment window periods. In the example, the event occurs between HE17and HE19 (highlighted in blue in the example). For two-hour pre- and post-event adjustment windows with a two-hour buffer, the adjustment window hours (highlighted in orange in the example) are HE13, HE14, HE22, and HE23.

    b. Average the usage across those four hours for both the baseline and the event day observed load.

c. Calculate the adjustment ratio by dividing the baseline average window value by the observed average window value. In the example, the baseline has an adjustment window value of 1.64kW and the event adjustment window value is 1.76. The ratio is then 1.07.

d. Cap the ratio at the required level. If the cap is 1.4x, as in the example, the following logic applies:

i. If the ratio is less than 1/1.4 = 0.71, the capped ratio is now set to 0.71.

ii. If the ratio is between 0.71 and 1.4, the ratio remains as is.

iii. If the ratio is greater than 1.4, the capped ratio is now set to 1.4.

e. Apply the capped ratio to each hour of the baseline by multiplying the capped ratio by the hourly baseline values for each hour

f. The profile obtained in step 6e is the baseline.

7. DR Energy Measurements are calculated as the difference between the baseline and the observed load such that load reductions relative to the baseline are positive. Load increases, when the baseline is less than the observed load, should be set to 0 for settlement purposes.