Reimagine tomorrow.

California ISO Baseline Accuracy Assessment

November 20, 2017

Principal authors

Josh Bode
Adriana Ciccone

Figures

Tables

1 Introduction

Historically, California has been a leader in the use of Demand Response and dynamic pricing to offset the need for additional peaking generation capacity, which is driven by system peak loads. A large share of DR resources, totaling roughly 1,700 MW, are enrolled in programs and contracts administered via the three California investor owned utilities – Pacific Gas & Electric (PG&E), San Diego Gas & Electric (SDG&E), and Southern California Edison (SCE). The market for demand response and battery storage is changing in three fundamental ways. First, the level of participation by third-parties (non-utilities) is expected to increase. Second, to receive credit for peaking capacity – also known as resource adequacy – DR resources must be bid into the CAISO markets. Third, the need for resources is increasingly becoming bidirectional; resources are needed to reduce demand (or inject power) during peak periods and to increase demand (or reduce power production) during periods when there is a surplus of power.

A key issue for incorporating DR resources into markets is accurate measurement of demand reductions for settlement. Measurements for settlement and operations need to be conducted much faster than traditional program evaluations, which are conducted on an annual basis. Settlements must also be transparent, relatively easy to understand, and simple to implement.

To estimate demand reductions, it is necessary to estimate what energy consumption would have been in the absence of DR dispatch — a baseline or counterfactual. The change in energy use is calculated as the difference between the baseline and consumption during the event. There are a variety of approaches for measuring the magnitude of curtailments with different degrees of complexity. While highly accurate results are desirable, there is often a tradeoff between simplicity and incremental accuracy.

Before 2017, settlement of DR resources at CAISO was based on using the same hour average for the 10 non-event weekdays immediately prior to dispatch of the resource – an approach known as a 10 of 10 baseline with a 20% adjustment cap – which was developed primarily based on analysis of large and mid-large non-residential customers. Prior baseline research has shown that the 10 of 10 baseline works reasonably well for the large and medium commercial customers who are not highly weather sensitive. However, research has also shown that the current baseline significantly underestimates residential demand response resources and non-residential weather sensitive customers. In addition, little research has been done on the performance of the current 10 of 10 baselines for customers participating in emergency demand response programs such as the Baseline Interruptible Program and Agricultural Pump load control.

Because the CAISO performs settlement by product type in specific geographic areas – known as sub Load Aggregation Points (subLAPs), it is critical to understand the extent to which the number of participants enrolled influence the accuracy and precision of settlements. Just as important is effect of more frequent event days, which reduce the number of control days that can be used to develop baselines.

The purpose of this study is to assess different baseline alternatives and rules that allow for accurate estimates of broad range of DR resources, including weather sensitive and less weather sensitive resources. A key outcome from this study is baseline proposal – subject to FERC approval – that allows a broader range of demand response programs/products to be bid into the CAISO market and be settled accurately. As part of the study, we obtained input regarding baseline variations to asses for accuracy from stakeholders, including from CAISO, PG&E, SCE, SDG&E and demand response and battery storage third party vendors. The three California utilities allowed the use of hourly smart meter or interval data from over 500,000 sites enrolled in eight distinct DR programs for the baseline accuracy assessment.

1.1 Key Research Questions

The study addresses several research questions, including:

- What are the most accurate and precise baselines by program type and customer class?
- How accurately and precisely do the best baselines perform?
- How much variation is there in accuracy and precision across geographic areas?
- Does the accuracy and precision vary depending on the number of customers (sample size) or event days?
- What is the effect of various baseline adjustments rules on the accuracy and precision of baseline estimates?
- Are approaches relying on control groups feasible and accurate and, if so, what are the implications of more granular sample sizes?

1.2 Aggregated versus Customer Specific Baselines

The settlement with CAISO is implemented at the resource level. Aggregator and utilities pool customers into a resource which delivers a specific product in a predefined area and bid the resource in to the CAISO market. Individual customer loads for events and non-event days are aggregated to the resource level before the baselines are calculated. While some jurisdictions estimate baselines for individual customer accounts first and then aggregate the resources to the resource level, this is not the case at CAISO. In this report, all accuracy and precision metrics are for baselines calculated for aggregated resources.

1.3 Baselines Included in Testing

At a high level, the baseline settlement methods tested for accuracy 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 curtails demand while the control group does not. The control group is used to establish the baseline of what load patterns would have been absence the curtailment event. This approach is the primary method for settlement of residential AC cycling and thermostat programs by Texas' system operator, ERCOT.

- 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 that did not experience an event. The process involves setting rules for the eligible days and rules for how the days used to estimate the baseline are selected from the eligible days. A subset of non-event days in close proximity to the event day are identified and averaged to produce baselines.

- Weather Matching — The process for weather matching baselines is similar to day-matching except that the baseline load profile is based on non-event days with similar temperature conditions. 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 23 day-matching, 12 weather-matching, and randomly assigned control groups were included in the accuracy assessment, for a total of 36 different baseline types.

Error! Reference source not found. and **Error! Reference source not found.** 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 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 1-1: 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 1-2: 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

1.4 Baseline Rules, Frequency, and Aggregation included in Testing

There several rules regarding baselines and option in the accuracy assessment which influence accuracy and precision. These include:

- Limits on baseline adjustments – Baseline adjustments are calculated by comparing actual loads and unadjusted baselines during non-event periods and using that information to calibrate the baseline. If the difference between the unadjusted baseline and the actual load is truly due to baseline estimation error, the adjustment process reduces those errors. Typically baseline adjustments are limited. As part the assessment, 10 baseline adjustments, including unlimited adjustment and no adjustments were tested.

- Adjustment buffers - To avoid contamination of the baseline with intentional changes to loads, a buffer period between adjustment periods and event dispatch hours is typically employed. Buffer periods reduce 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. The default buffer is two hours before and after and event, but as part of the assessment, the use of buffer or 1, 2, and 3 hours was tested for residential air conditioner programs.

- Use of hours before and after the event in the baseline adjustment calculation. Historically, baseline adjustments have been calculated using only pre-event hours. At the request of a stakeholder, the study assessed the inclusion of hours before and after the events to calculate the baseline adjustment. This was done only for residential air conditioner programs where results for large control groups were available.

- The number of event days called. When more events are called, it limits the days available for baseline calculations. The sole exception is control groups, which are unaffected by frequent events. To assess the impact of event days on baseline accuracy, the study assessed baseline accuracy when 3, 5, 10, or 15 events were called per summer.

- How many sites are aggregated into a resource. More aggregation of diverse resources tends to smooth out idiosyncrasies, leading to more accurate baselines. The less resources are aggregated, the lower the accuracy of baselines. To assess the impact of aggregation, the baseline accuracy was assessed using different amount of aggregation. For mass market programs such as air conditioner cycling or connected loads and agricultural pumps, the accuracy was estimated for resources of 200, 500, 1,000 and 2,000 sites. For large C&I customers, the accuracy was estimate for resources of 20, 50, 100, 200, and 300 sites.

- The timing of the event. The assessment analyzed events starting at 2 pm, 3 pm, and 5 pm and lasting four hours each.

When combined with the baselines, over 12,000 combinations of baselines, adjustment rules, aggregation, and event dispatch were tested for each of the program types included in the assessment.

1.5 Baseline Accuracy versus Demand Response 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. There are two basic approaches:

- Assess the accuracy of baselines themselves — This involves estimating the baseline and comparing it to actual unperturbed load during non-event days. While this is useful for identifying the best performing baseline, it is not a direct assessment of how accurately the signal—the demand reduction—is measured. An emphasis on baseline accuracy is analogous to assessing which method is better at reducing noise.
- Assess the accuracy of the demand reductions produced by the baseline — Baselines are simply a means to produce demand reductions estimates. They are tools to filter out noise (or explain variation) and allow the effect or impact to be more easily detected. The focus, however, is on how accurately the demand reductions are detected. If actual demand reductions are 20%, a baseline that is biased by 2% will estimate demand reductions of 22%, or estimate 110% of the actual demand reductions. Accuracy of baselines is clearly different than the accuracy of the demand reductions estimated by baselines.

Throughout this report, the focus of the analysis is on the accuracy of the baselines. For individual market participants, accuracy of settlement will depend on the amount of resources aggregated, the diversity of those resources (i.e., whether or not a single participant dominates results), and the percent demand reduction delivered.

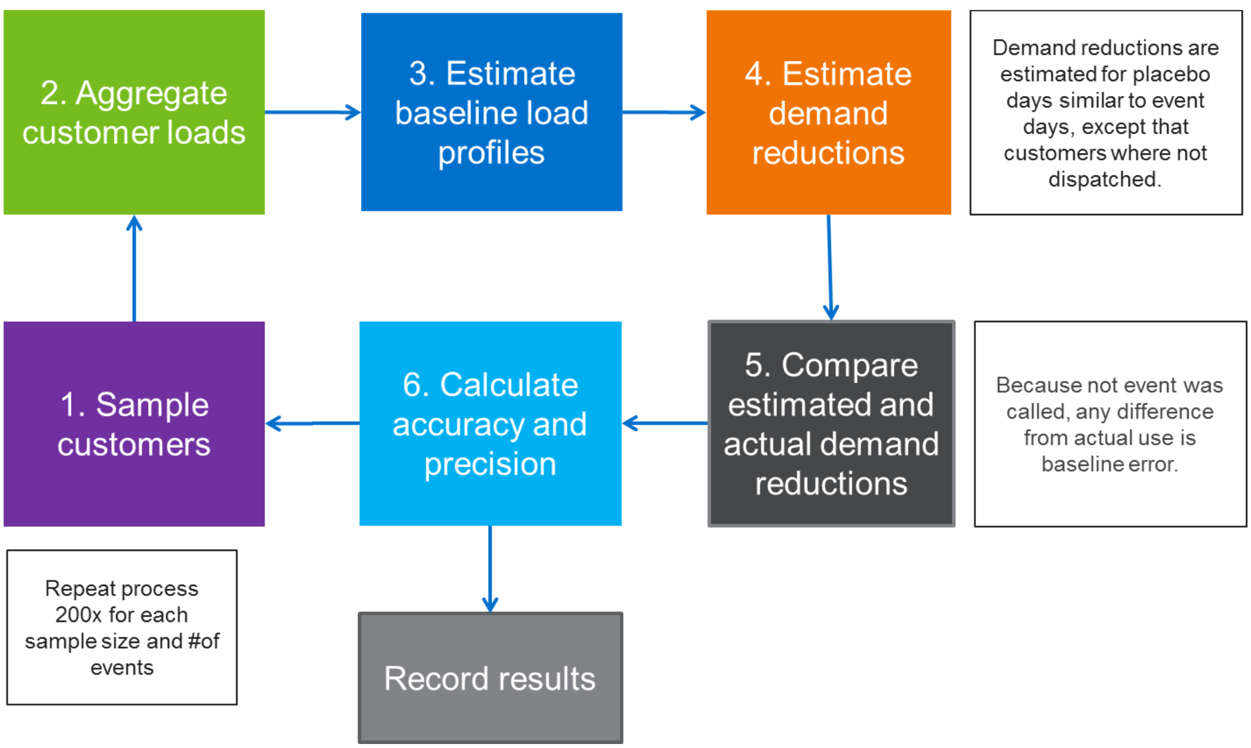
2 Methodology

2.1 Assessing Baseline Error

To assess the baseline error, 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 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.

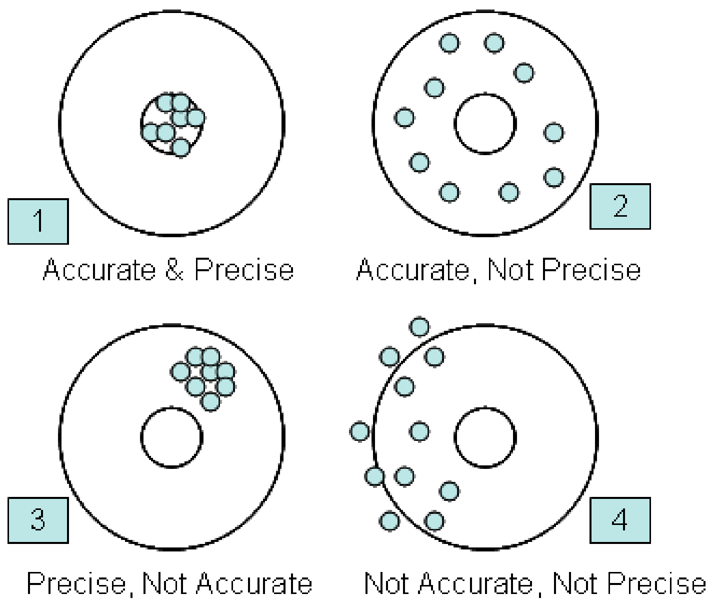
The only instance where placebo events were not used was in assessing the use of both pre and post event hours in the baseline adjustment calculation. The analysis was implemented using an air conditioner cycling program residential sample and actual event days when load was expected to be higher after events due to snapback. Because PG&E withholds a large randomly assigned control group of over 14,000 customers for each of its events, the control group estimate of the counterfactual is highly precise and nearly error free, providing a basis against which day and weather matching baselines could be compared.

2.2 Accuracy and Precision Metrics

The terms accuracy and precision have a very specific meaning to statisticians and data scientists. Accuracy refers to metrics for bias; the tendency to over or under predict. Precision refers to metrics for how close typical predictions are to actual answers.

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 (#4)

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



Throughout this report, the performance of baseline rule options was summarized using two metrics: one for accuracy (or bias) and one for precision (or goodness-of-fit). The equations and formal description are included in the methodology section, but it is important to understand how to interpret these metrics.

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.	$CV(RMSE) = \frac{RMSE}{\bar{y}}$

2.3 Data Sources

Table 2-2 summarizes the data provide by PG&E, SCE, and SDG&E for the baseline accuracy assessment. In total, hourly data over 2 years from nearly 104,000 customers was used for the baseline accuracy assessment. All sites were current or recent participants in utility programs and, in nearly all cases, the full population of participants was employed in the analysis.

Table 2-2: Data Sources for Analysis

Program Type	Utility Program	Number of accounts	Time frame
Weather Sensitive	PG&E Residential AC cycling	84,159	Jan 2015 to Oct 2015
	SDG&E Residential AC cycling (100%)	1,064	Jan 2015 to Oct 2015

	SDG&E Residential AC Cycling (50%)	1,110	Jan 2015 to Oct 2015
	SCE Commercial AC cycling	10,760	Jan 2015 to Oct 2015
	SDG&E Commercial AC Cycling	4,467	Aug 2015 to Oct 2015
Industrial and Agricultural	PG&E Baseline Interruptible Program	299	Nov 2013 to Sep 2015
	SCE Baseline Interruptible program	633	Nov 2013 to Sep 2015
	SCE Agricultural pumps	1,285	Nov 2013 to Sep 2015

2.4 Selection of Placebo Event Days

Baseline accuracy was assessed on placebo days. Because no event was called, any deviation between the baseline and actual loads is due to error. Actual event days were removed from the analysis datasets to ensure the baselines calculation did not include days where customers were delivering demand reductions.

The placebo events were based high net loads to better account for the high penetration of utility scale renewables in California, which is affecting when, how often, and for how long resources are needed. Different frequency of events was simulated, from as little as 3 events per year to as many as 15 events per year, in order to assess if the frequency of dispatch influenced the accuracy of baselines - with high frequency dispatch, fewer days are available to baseline settlement calculations.

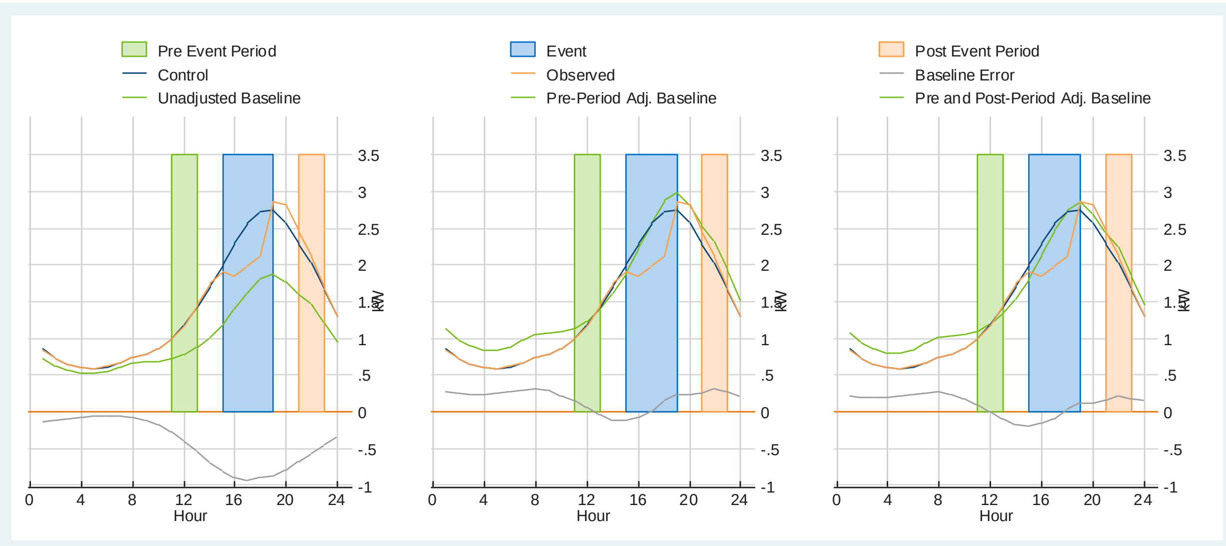
2.5 Baseline Adjustments

Another key issue is the use of baseline adjustments – are they used and, if so, what are the rules for around those adjustments? The concept relies on comparing actual loads and unadjusted baselines during non-event periods and using that information to calibrate the baseline. The underlying assumption is that differences during non-event periods are due to measurement error. That is, if the difference between the unadjusted baseline and the actual load is truly due to baseline estimation error, the adjustment process reduces those errors.

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. If, during non-event 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 non-event adjustment hours, it is adjusted downwards. To avoid contamination of the baseline with perturbed event hours, a buffer period between adjustment periods and event dispatch hours is typically employed. Buffer periods reduce 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. Same-day adjustments are often capped to reduce the variance of estimates and to limit the potential for manipulation of loads to influence baselines.

Error! Reference source not found. illustrates the concept of baseline adjustments. 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 large 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.

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



3 Results

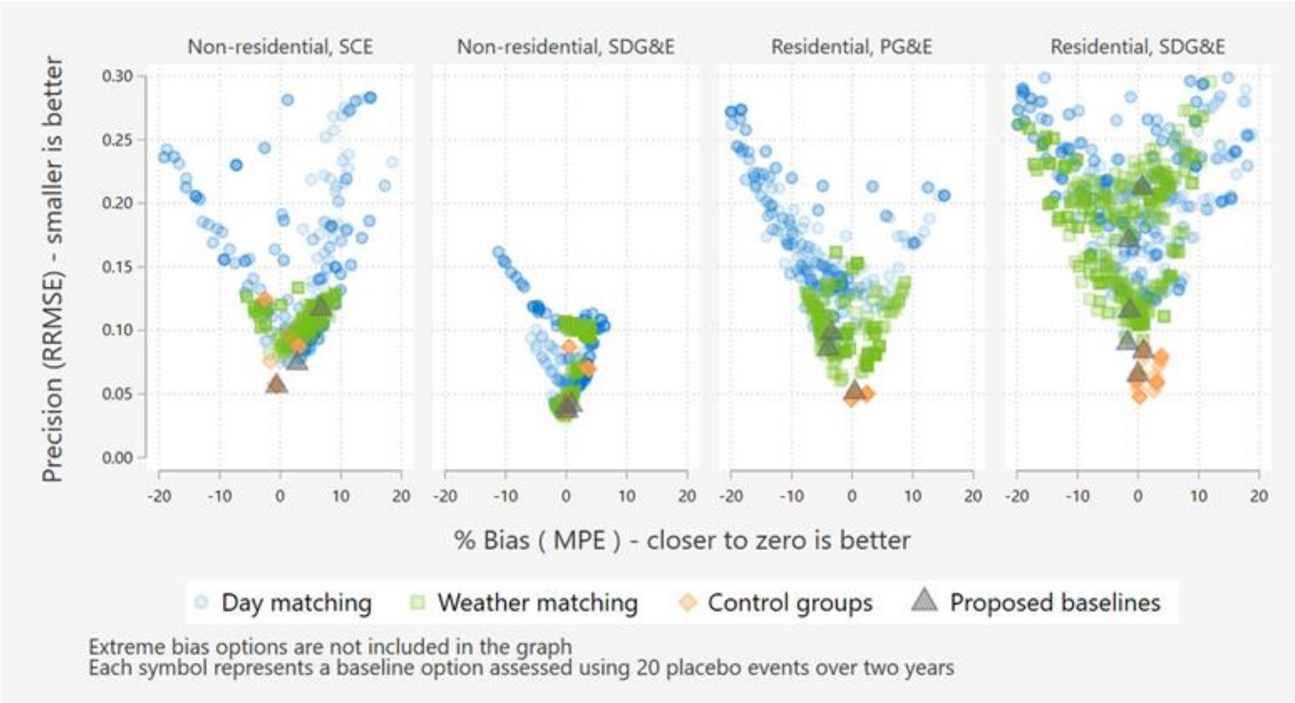
The goal of this study was to assess different baseline alternatives and rules and identify options that accurately estimate impacts of a broad range of DR resources, including weather sensitive and less weather sensitive resources. Over 120,000 combinations of baselines, adjustment rules, aggregation level, and event dispatch frequency and timing were tested on each of eight utility programs. Due to the volume, the results are presented in a summary format holding the number of events (20 over 2 years), timing of events (3 pm to 7 pm), and aggregation level (100 or 500 sites for commercial and residential types, respectively) constant, and assuming baseline adjustments are based on pre-treatment data. The effect the number of events and the amount of resource aggregation on the accuracy of baselines is presented separately. Unless otherwise indicated, accuracy was assessed using placebo event days – event like days when resources where not dispatched – allowing error to be calculated by comparing baselines against actual loads. For more detailed results, please refer to Appendix E, where the top ten best baselines for each program, utility and baseline type are listed, along with their bias and precision metrics.

A key finding of the analysis is that multiple baseline rules can deliver sufficiently unbiased and precise baselines. The proposed baselines were arrived at based on input from CAISO, third party stakeholders, and the three investor owned utilities in California.

3.1 Accuracy and Precision Metrics for Existing Programs

Figure 3-1 summarizes the baseline accuracy results for the weather sensitive air conditioner programs analyzed. Each symbol represents the bias and precision of a baseline rule option assessed over 20 placebo events over the course of two years. The best approaches have little or no bias – the tendency to over or under predict on average – and are more precise – the typical magnitude of errors for individual events periods is smaller. On the graph, the best baselines are at the bottom of the “V” shape.

Figure 3-1: Bias and Precision for Weather Sensitive Residential and Non-Residential Customers



Control groups methods consistently outperformed weather matching and day matching baselines, delivering baselines that were unbiased and more precise. Overall, weather matching methods typically outperformed day matching baselines. The chart also shows the proposed baselines. The chart does not show the degree to which inclusion of post event hours in the baseline adjustment improves the accuracy of results (it does). Because this analysis was implemented on a subset of data, it is discussed separately in section 3.2.

Figure 3-2 shows the baseline accuracy results for the Baseline Interruptible Program, which is mainly comprised of large industrial customers, and for agricultural pumps. Control groups were not assessed for these options since they are fewer in number and loads vary more widely across customers. The results are shown using the same scale as the weather sensitive loads to allow direct comparisons. While loads for these customers can be seasonal (particularly for agricultural pumps), they are less sensitive to day to day variation in weather conditions. In general, baselines for less weather sensitive customers are more precise.

Figure 3-2: Bias and Precision for Industrial (BIP) and Agricultural Customers

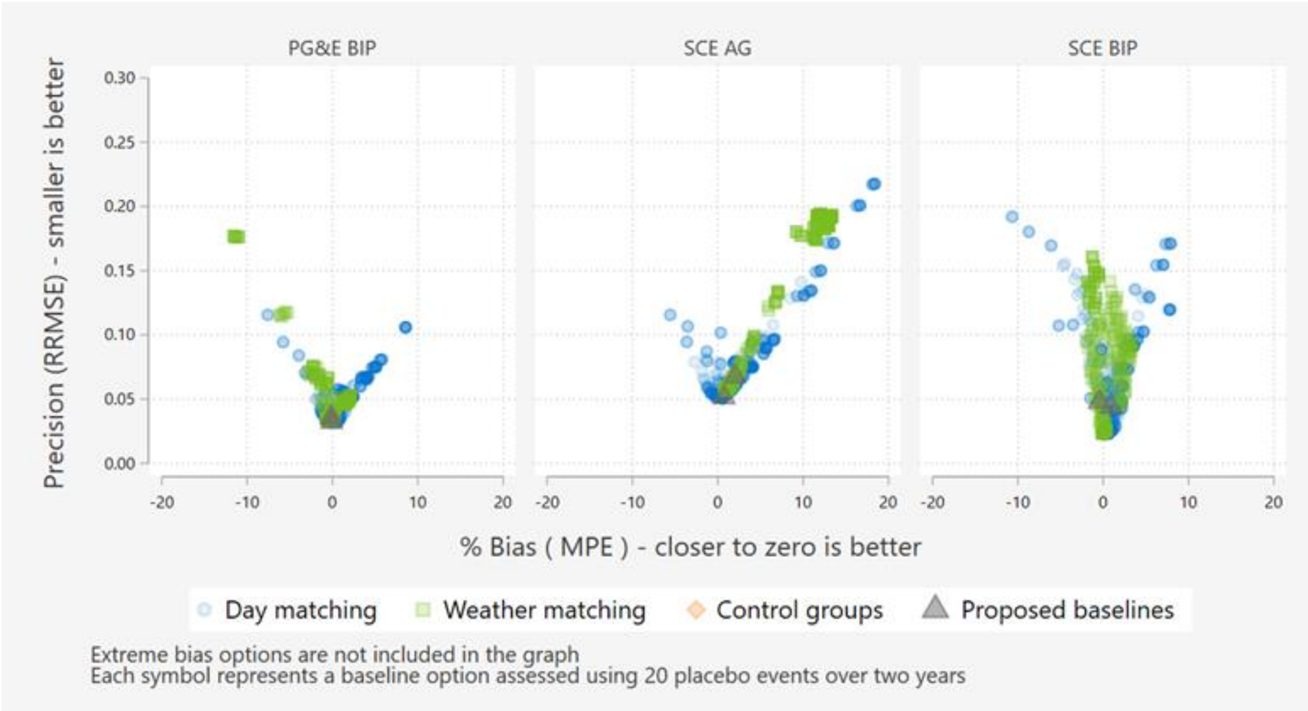


Table 3-1 shows the bias and precision metrics for the proposed and current baselines for each program assessed. For residential weather sensitive programs as whole, the current baseline is downwardly biased by 12% to 14% and event to event magnitude of errors is sufficiently large to occasionally nullify actual reductions. The proposed baselines reduce the tendency to under predict and improve precision for dispatch hours. As discussed later, the baselines of weather sensitive customers are improved further by including post event hours in the baseline adjustment calculation. For commercial customers, the existing baseline performed relatively well but can be improved on, especially by using control groups.

Table 3-1: Bias and Precision for Proposed and Current Baselines

Program Type	Utility Program	Baseline type	Proposed		Current Baseline	
			Bias (MPE)	Precision (CVRMSE)	Bias (MPE)	Precision (CVRMSE)
Weather Sensitive	PG&E Residential AC cycling	Day matching	-4.0%	0.086	-13.1%	0.179
		Weather matching	-3.4%	0.098		
		Control group	0.4%	0.051		
	SDG&E Residential AC cycling (100%)	Day matching	-1.5%	0.171	-12.7%	0.240
		Weather matching	0.7%	0.212		
		Control group	0.9%	0.084		
	SDG&E Residential AC Cycling (50%)	Day matching	-1.8%	0.090	-13.7%	0.205
		Weather matching	-1.4%	0.116		

		Control group	-0.1%	0.065		
	SCE Commercial AC cycling	Day matching	2.8%	0.074	2.8%	0.074
		Weather matching	6.7%	0.117		
		Control group	-0.6%	0.056		
	SDG&E Commercial AC Cycling	Day matching	0.9%	0.041	0.9%	0.041
		Weather matching	-0.1%	0.040		
		Control group	0.1%	0.037		
Industrial and Agricultural (not weather sensitive)	PG&E Baseline Interruptible Program	Day matching	-0.1%	0.032	-0.1%	0.032
		Weather matching	-0.2%	0.036		
	SCE Baseline Interruptible program	Day matching	0.9%	0.044	0.9%	0.044
		Weather matching	-0.4%	0.048		
	SCE Agricultural pumps	Day matching	0.7%	0.051	0.7%	0.051
		Weather matching	2.0%	0.068		

3.2 Inclusion of Post Event Hours in the Baseline Adjustments

Historically, baseline adjustments for day and weather matching baselines have been calculated using only pre-event hours. At the request of a stakeholder, the study assessed the use of hours before and after the events to calculate baseline adjustments. The drier California weather leads to limited use of air conditioning until the late afternoon and evening hours. As result, post event hours can include information useful for calibrating the baselines that is not available during pre-event hours.

The impact of including post event hours in the baseline calculation was studied using actual events and data from PG&E and SDG&E, both of which rely on control groups to estimate the baseline. Actual event days were employed to account for small increases in load that occur when control of air conditioners is released – a phenomenon known as snapback. The baselines were compared to the control group loads. While this is technically a comparison of one estimate – a baseline – to another – the counterfactual produced by the control group – the control groups used were large enough that any sampling error was minimal.

Figure 3-3: Effect of Including Post Event Hours in Baseline Adjustment Calculation

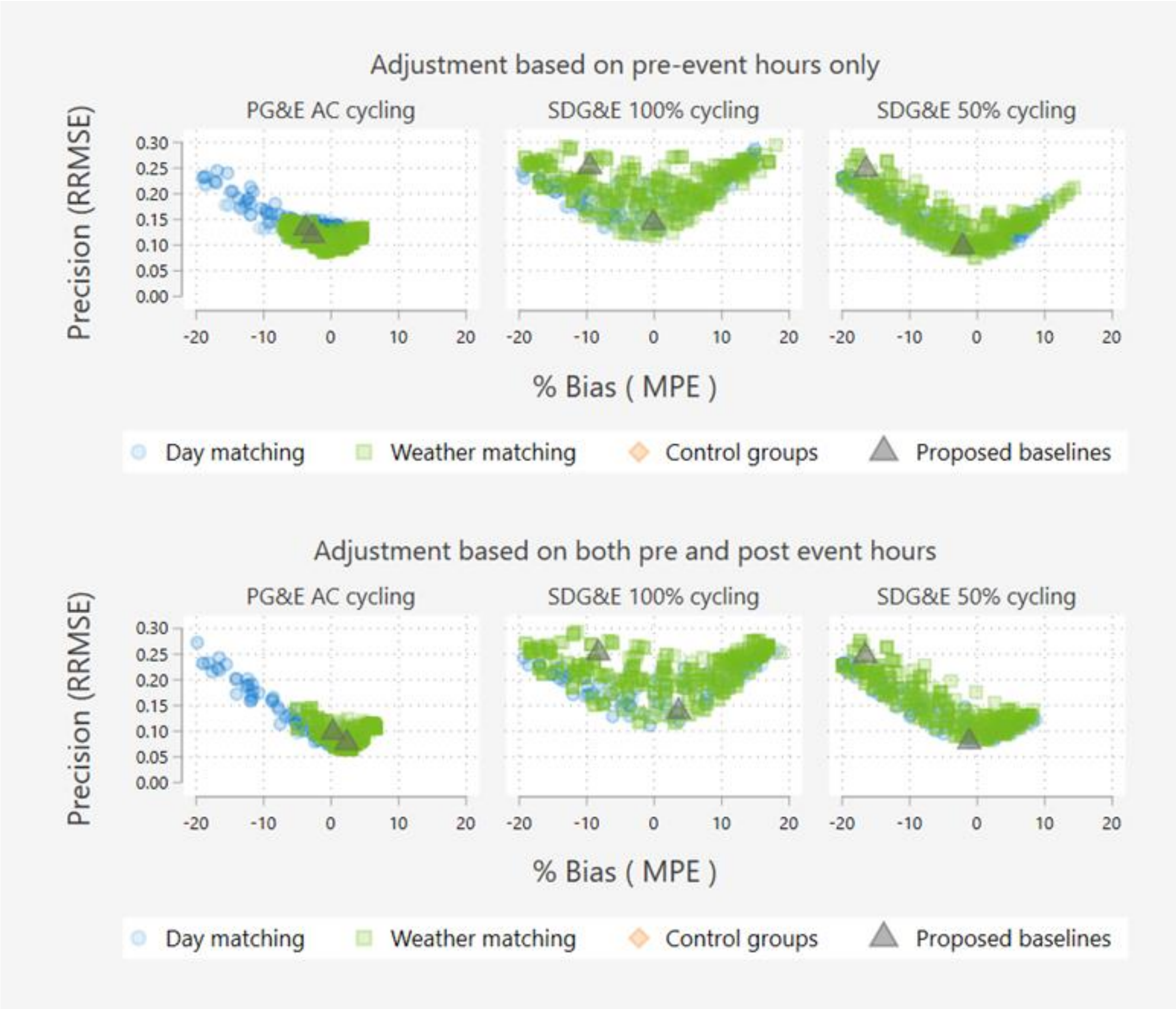
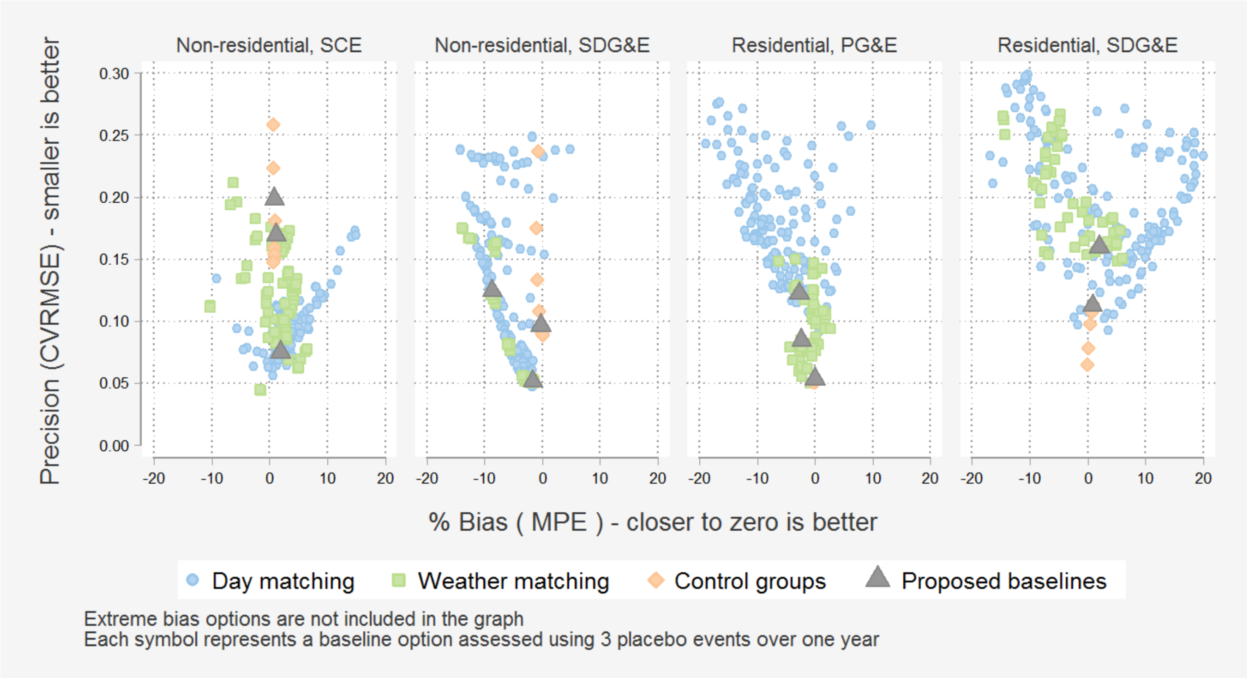


Figure 3-3compares the baselines with and without the inclusion of the post event hours in the baseline calculation. Adding the post-event hours to the baseline adjustment, reduced bias and improves the precision of the impacts for nearly all baselines tested, however the improvement is slight, compared to the improvements seen with including a pre-event adjustment at all.

3.3 Accuracy, Precision, and Post Event Hours for Weekends

Figure 3-4 summarizes the baseline accuracy results for weather sensitive air conditioner programs analyzed on weekends. The proposed weekend baselines differ from the proposed weekday baselines because the patterns of weekend use may differ substantially from weekdays. Using weekday use to predict weekend use for customer classes that vary in loadshape across days of week would substantially reduce the accuracy of the baseline. The results below are shown using the same scale as the weekday baselines to allow direct comparisons. Unlike weekday results which were simulated using ten placebo events, the weekend baselines were calculated using 3 placebo event days over the course of one year.

Figure 3-4: Bias and Precision for Weather Sensitive Residential and Non-Residential Customers on Weekends



As with weekdays, control groups consistently delivered less biased baselines. Overall, weather matching baselines typically outperformed day matching baselines, consistent with the weekday results. Since no events were called on weekends for the programs and summers of data available, there was no additional analysis on how the inclusion of post-event hours in the baseline adjustment improves the accuracy of results for weekends, nor on how the proposed baselines performed during actual events

Figure 3-5 shows the baseline accuracy results for agricultural customers and customers enrolled in the Baseline Interruptible Program on weekends. As with weekdays, control groups were not assessed for these options since they are fewer in number and loads vary more widely across customers. The results are shown using the same scale as the weather sensitive and weekday groups to allow direct comparison. As these customers are generally less weather-sensitive with more stable loads, both weather and day-matching methods performed similarly.

Figure 3-5: Bias and Precision for Industrial (BIP) and Agricultural Customers on Weekends

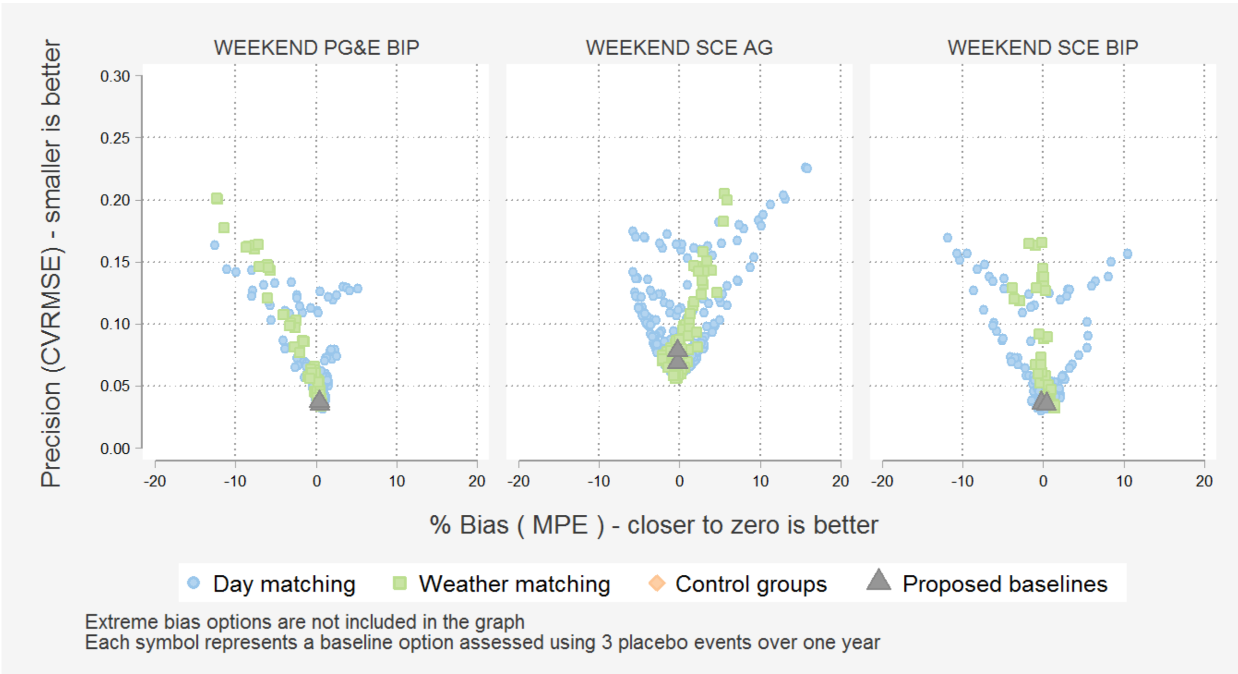


Table 3-2 shows the bias and precision metrics for the proposed and current weekend baselines for each program assessed. For residential weather sensitive programs as whole, the current baseline is downwardly biased by 6% and upwardly biased to 9% and event to event magnitude of errors is sufficiently large to occasionally nullify actual reductions. The proposed baselines reduce the tendency to over or under predict and improve precision for dispatch hours. For commercial customers, the existing baseline performed relatively well but can be improved on, especially by using control groups.

Table 3-2: Bias and Precision for Proposed and Current Baselines

Program Type	Utility Program	Baseline type	Proposed		Current Baseline	
			Bias (MPE)	Precision (CVRMSE)	Bias (MPE)	Precision (CVRMSE)
Weather Sensitive	PG&E Residential AC cycling	Day matching	-2.7%	0.122	-5.7%	0.172
		Weather matching	-2.4%	0.084		
		Control group	0.02%	0.053		
		Weather matching	2.0%	0.160		
		Control group	0.8%	0.112		
	SDG&E Residential AC Cycling	Day matching	22.5%	0.248	8.8%	0.126
		Weather matching	-2.0%	0.160		
		Control group	0.8%	0.112		
	SCE Commercial AC	Day matching	1.8%	0.075	1.8%	0.060

	cycling	Weather matching	1.1%	0.169		
		Control group	-0.8%	0.199		
	SDG&E Commercial AC Cycling	Day matching	-8.6%	0.124	-8.6%	0.124
		Weather matching	-1.6%	0.051		
		Control group	-0.3%	0.096		
Industrial and Agricultural (not weather sensitive)	PG&E Baseline Interruptible Program	Day matching	0.4%	0.036	0.4%	0.036
		Weather matching	0.4%	0.037		
	SCE Baseline Interruptible program	Day matching	-0.2%	0.036	-0.2%	0.036
		Weather matching	0.3%	0.035		
	SCE Agricultural pumps	Day matching	-0.2%	0.078	-0.2%	0.078
		Weather matching	-0.3%	0.068		

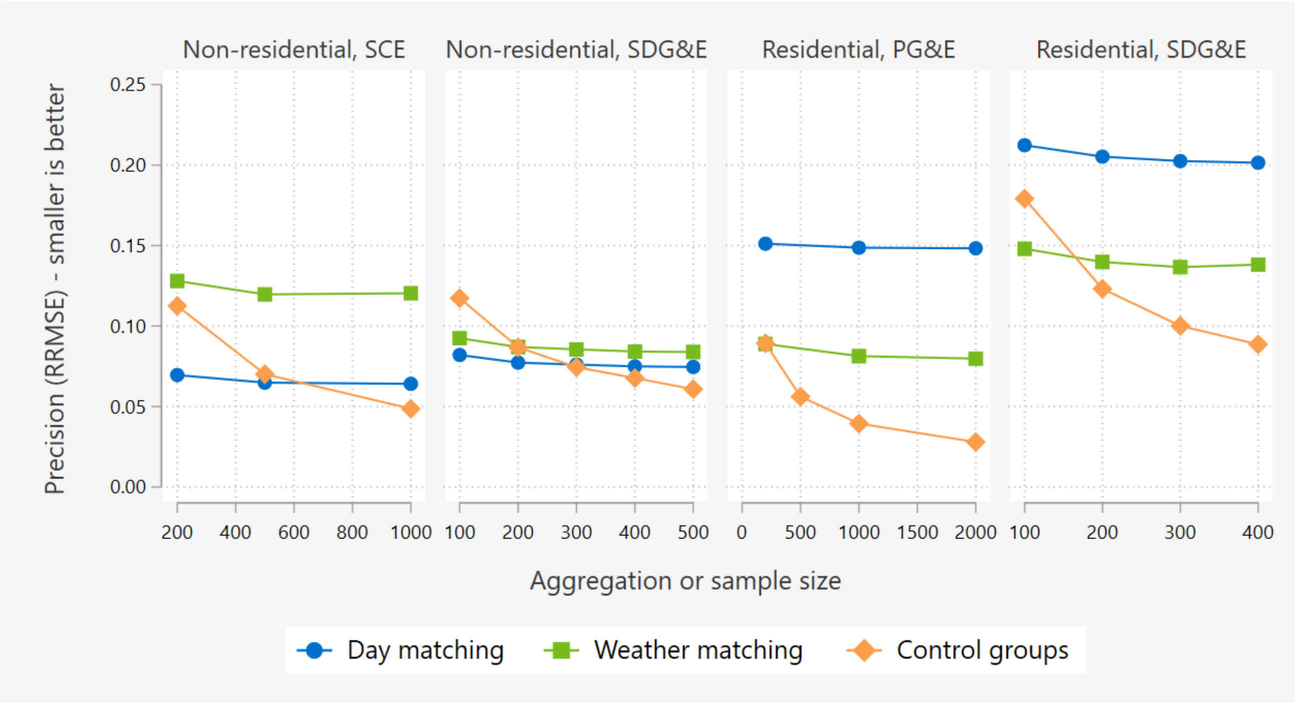
3.4 Impact of Aggregation on the Precision of Baselines

Baseline methods perform better when there are larger numbers of customers and those customers are diverse. This is true both for methods that rely exclusively on non-event data and for baselines that rely on a control group. Baselines tend to perform more poorly when there are fewer participants or when loads and demand reductions are highly concentrated on a handful of customers.

Because the focus is on settlement by product type in specific geographic areas, it is critical to understand the extent to which the number of participants enrolled influence the precision of settlements. While baselines that reply on a control groups are generally more precise, they require withholding some customers from event dispatch. The question is how many. For newer market participants, it may require a considerable share of their resources, especially because the sample sizes need to be adequate with each of the 20 geographic settlement areas

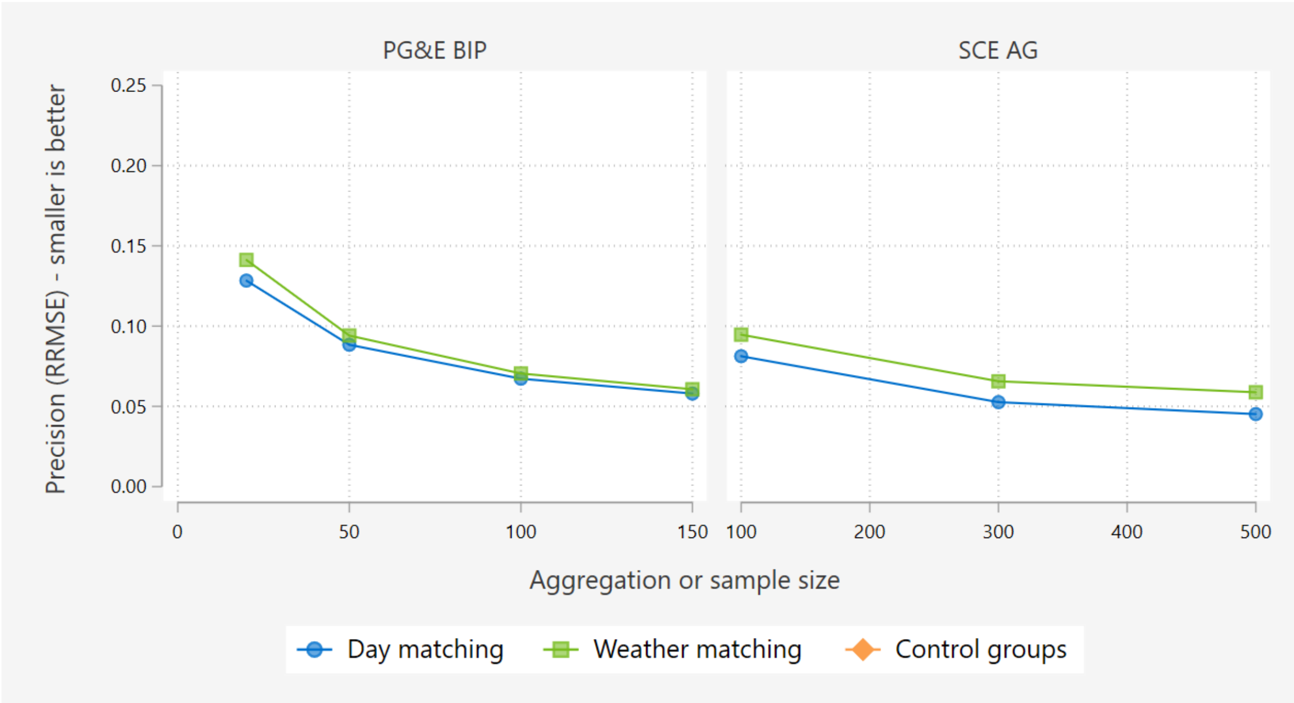
Figure 3-6 and Figure 3-7 show how the precision of baseline methods improves with aggregation or, in the case of control groups, the sample size. The aggregation levels tested for the different programs varied due to the available data, but some patterns emerge.

Figure 3-6: Effect of Aggregation or Sample Size on Precision for Weather Sensitive Customers



Day and weather matching baselines perform better for non-residential customers than for residential ones. Once control group sizes exceed approximately 200 customers, they outperform weather and day matching methods. The larger the control group, the more precise estimates produced. With 500 customers, control groups are more than twice as precise as day and weather matching baselines. However, day and weather matching methods are typically more precise than control groups when control groups are less than 200. We also observe that aggregation leads to improvement in precision for day and weather matching methods, especially with smaller groups. However, the gains of more aggregation are more pronounced with control groups.

Figure 3-7: Effect of Aggregation or Sample Size for Industrial (BIP) and Agricultural Customers



4 Recommendations

Table 4-1 shows the recommended 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 precise than 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 demand response providers 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 4-1: 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.

4.1 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 4-2 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 4-2: 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 assignment 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 the California ISO deems it necessary, DRPs will be required to securely provide the control and treatment

Component	Explanation
	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

4.2 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 4-3: 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 4-4: Non-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	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	
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	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.	
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	

4.3 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 4-5: 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 4-6: 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	

5 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 precise than day or weather matching baselines.

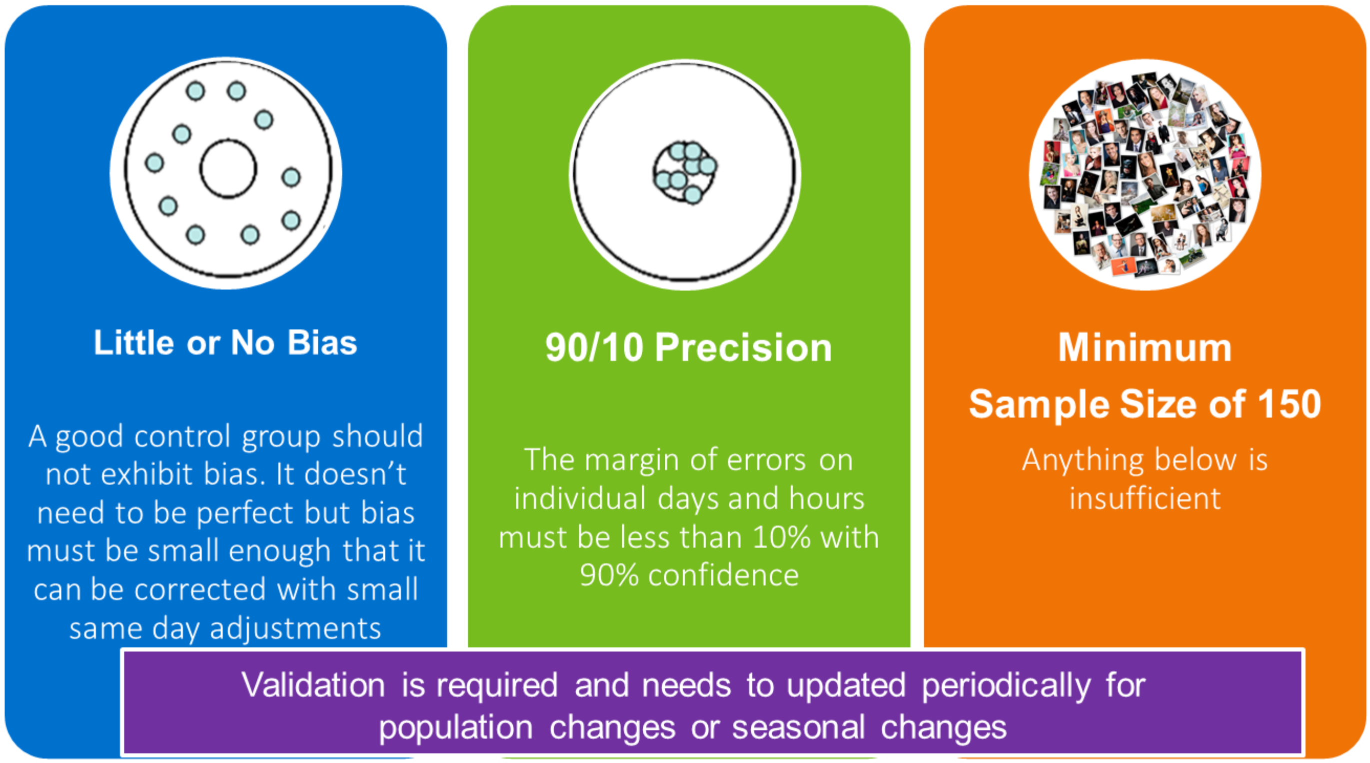
Control groups involve using a set of customers who did not experience events to establish a baseline. A control group should be made of customers who are statistically indistinguishable from the participant group on non-event days to act as a comparison on event days, instead of relying on participants' 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 demand response providers are able to establish a control group by randomly assigning and withholding a subset of participant resource sites from dispatch. However, not all demand response providers 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 5-1 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 5-1: Control Group Requirements



5.1 Statistical Checks Necessary to Demonstrate Control Group Validity

Demand response providers 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 demand response provider to develop the control group and demonstrate equivalence. The control group(s) developed are subject to audit by the CAISO.

1. The demand response provider 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%.

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 100 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.

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

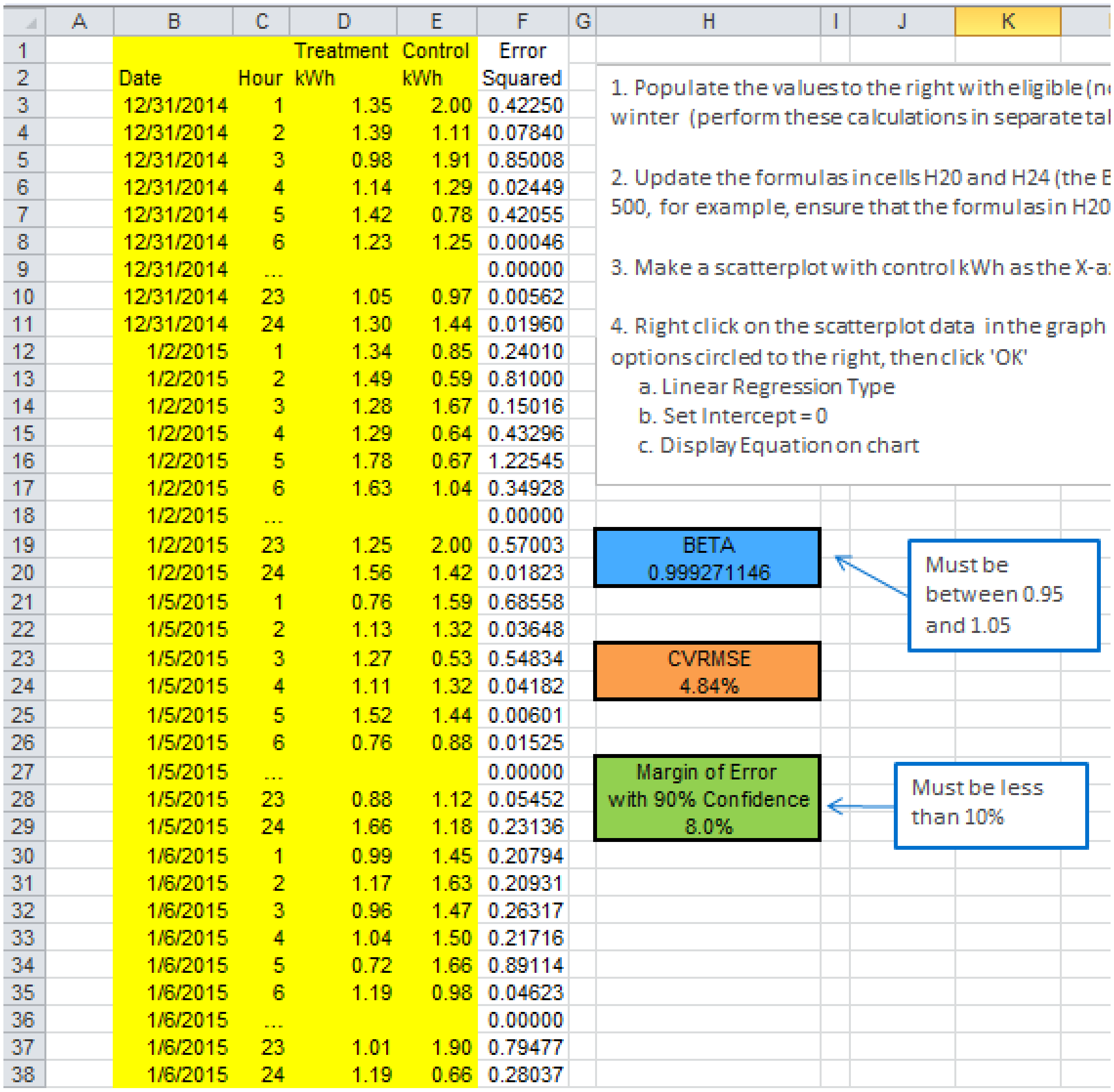
4. Arrange the data in the appropriate format.

Table A-4: Average Daily Treatment and Control Usage

Date	Hour	kWh_Treat	kWh_Control
12/31/2014	1	1.35	2.00
	2	1.39	1.11
	3	0.98	1.91
	4	1.14	1.29
	5	1.42	0.78
	6	1.23	1.25
	…		
	23	1.05	0.97
	24	1.30	1.44
1/2/2015	1	1.34	0.85
	2	1.49	0.59
	3	1.28	1.67
	4	1.29	0.64
	5	1.78	0.67
	6	1.63	1.04
	…		
	23	1.25	2.00
	24	1.56	1.42
1/5/2015	1	0.76	1.59
	2	1.13	1.32
	3	1.27	0.53
	4	1.11	1.32
	5	1.52	1.44
	6	0.76	0.88
	…		
	23	0.88	1.12
	24	1.66	1.18
1/6/2015	1	0.99	1.45
	2	1.17	1.63
	3	0.96	1.47
	4	1.04	1.50
	5	0.72	1.66
	6	1.19	0.98
	…		
	23	1.01	1.90
	24	1.19	0.66

5. Regress average treatment hourly load against average control hourly load during event hours with no constant by filling in the attached template and updating formulas in cells H20 and H24 to include the full range of the data added to columns B through E.

Figure A-1: Regression and Validation Template



	A	B	C	D	E	F
1				Treatment	Control	Error
2		Date	Hour	kWh	kWh	Squared
3		12/31/2014	1	1.35	2.00	0.42250
4		12/31/2014	2	1.39	1.11	0.07840
5		12/31/2014	3	0.98	1.91	0.85008
6		12/31/2014	4	1.14	1.29	0.02449
7		12/31/2014	5	1.42	0.78	0.42055
8		12/31/2014	6	1.23	1.25	0.00046
9		12/31/2014	...			0.00000
10		12/31/2014	23	1.05	0.97	0.00562
11		12/31/2014	24	1.30	1.44	0.01960
12		1/2/2015	1	1.34	0.85	0.24010
13		1/2/2015	2	1.49	0.59	0.81000
14		1/2/2015	3	1.28	1.67	0.15016
15		1/2/2015	4	1.29	0.64	0.43296
16		1/2/2015	5	1.78	0.67	1.22545
17		1/2/2015	6	1.63	1.04	0.34928
18		1/2/2015	...			0.00000
19		1/2/2015	23	1.25	2.00	0.57003
20		1/2/2015	24	1.56	1.42	0.01823
21		1/5/2015	1	0.76	1.59	0.68558
22		1/5/2015	2	1.13	1.32	0.03648
23		1/5/2015	3	1.27	0.53	0.54834
24		1/5/2015	4	1.11	1.32	0.04182
25		1/5/2015	5	1.52	1.44	0.00601
26		1/5/2015	6	0.76	0.88	0.01525
27		1/5/2015	...			0.00000
28		1/5/2015	23	0.88	1.12	0.05452
29		1/5/2015	24	1.66	1.18	0.23136
30		1/6/2015	1	0.99	1.45	0.20794
31		1/6/2015	2	1.17	1.63	0.20931
32		1/6/2015	3	0.96	1.47	0.26317
33		1/6/2015	4	1.04	1.50	0.21716
34		1/6/2015	5	0.72	1.66	0.89114
35		1/6/2015	6	1.19	0.98	0.04623
36		1/6/2015	...			0.00000
37		1/6/2015	23	1.01	1.90	0.79477
38		1/6/2015	24	1.19	0.66	0.28037

1. Populate the values to the right with eligible (n
winter (perform these calculations in separate ta

2. Update the formulas in cells H20 and H24 (the B
500, for example, ensure that the formulas in H20

3. Make a scatterplot with control kWh as the X-a

4. Right click on the scatterplot data in the graph
options circled to the right, then click 'OK'
a. Linear Regression Type
b. Set Intercept = 0
c. Display Equation on chart

BETA
0.999271146

Must be between 0.95 and 1.05

CVRMSE
4.84%

Margin of Error with 90% Confidence
8.0%

Must be less than 10%

6. The statistics of interest are in cells H20, H24, and H29.

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.* If using Stata, the validation statistics can be calculated easily using the two commands underlined in green. The coefficient β is the value circled in orange. The 90% limit on the CVRMSE can be calculated using the output (circled in blue) from the same two commands as shown in Figure A-2.

Figure A-2: Stata Commands to Calculate Equivalence Statistics

```
          reg kwh_treat kwh_control, noconstant

      Source |       SS       df       MS              Number of obs =    5568
-------------+------------------------------           F(  1,  5567) =       .
       Model |  3792.8973     1  3792.8973             Prob > F      =  0.0000
    Residual |  10.197965  5567  .00183186             R-squared     =  0.9973
-------------+------------------------------           Adj R-squared =  0.9973
       Total | 3803.09527  5568  .683027167            Root MSE      =   .0428

------------------------------------------------------------------------------
   kwh_treat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 kwh_control |    1.00539   .0006987  1438.93   0.000     1.004021     1.00676
------------------------------------------------------------------------------


          sum kwh_treat

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
   kwh_treat |      5568    .7518921    .3430839   .1965188   3.313407


          di in red "the RMSE is " e(rmse)
the RMSE is .04280023

          di in red "the average treatment kWh is " r(mean)
the average treatment kWh is .75189212

          di in red "the 90% confidence limit of the CVRMSE is " 1.645 * (e(rmse)/r(mean)) * 100 "%"
the 90% confidence limit of the CVRMSE is 9.3638946%

          di in red "it can also be manually entered like this: " 1.645 * (.04280023/.75189212) * 100 "%"
it can also be manually entered like this: 9.3638936%
```

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.

Appendix E Best Baseline Results by Program and Utility

Results shown here are the top 10 baselines by utility, program, and baseline category, chosen by finding baselines with absolute bias less than 10%, and then sorted by low CVRMSE. For consistency, all results shown here are for events with the following features:

- Event window from HE16-HE19 (or as close as possible to this for event-based results).
- Residential programs use samples of 500 customers, and commercial programs use 100
- Simulated over 10 events per summer for proxy weekday results and 3 events for proxy weekend results
- All post-event adjustments include a 2-hour buffer between the end of the event and the post-adjustment period.

Table A-5: Proxy Weekday Results

Program	Baseline Category	Type	Adjustment Cap	Adjustment Type	MPE (%)	CVRMSE (%)	Recommended
PG&E BIP	Day matching	10/20	+/-1.4x	No Post Adjustment	0.53	3.20	
		10/20	Unlimited	No Post Adjustment	0.53	3.20	
		10/20	+/-1.5x	No Post Adjustment	0.53	3.20	
		10/20	+/-2x	No Post Adjustment	0.53	3.20	
		10/20	+/-1.3x	No Post Adjustment	0.54	3.20	
		10/10	+/-1.5x	No Post Adjustment	-0.11	3.22	Same Type as Proposed
		10/10	+/-1.4x	No Post Adjustment	-0.11	3.22	Same Type as Proposed
		10/10	+/-2x	No Post Adjustment	-0.11	3.22	Same Type as Proposed
		10/10	Unlimited	No Post Adjustment	-0.11	3.22	Same Type as Proposed
		Bottom 10/10	+/-1.5x	No Post Adjustment	-0.11	3.22	
	Weather matching	Bins based on CDD	Unlimited	No Post Adjustment	-0.42	3.29	
		Bins based on CDD	+/-2x	No Post Adjustment	-0.42	3.29	
		Bins based on CDD	+/-1.5x	No Post Adjustment	-0.43	3.30	
		Bins based on Sum of CDH	+/-2x	No Post Adjustment	-0.65	3.41	
		Bins based on Sum of CDH	Unlimited	No Post Adjustment	-0.65	3.41	
		Bins based on Max Temp	Unlimited	No Post Adjustment	-0.67	3.41	
		Bins based on Max Temp	+/-2x	No Post Adjustment	-0.67	3.41	
		5 Day Match on Sum of CDH	+/-1.3x	No Post Adjustment	-0.13	3.41	
		5 Day Match on Sum of CDH	+/-2x	No Post Adjustment	-0.13	3.41	
		5 Day Match on Sum of CDH	+/-1.5x	No Post Adjustment	-0.13	3.41	
PG&E Res AC Cycling	Control group	Control group	+/-1.1x	Pre & Post Adjustment	0.21	3.84	Same Type as Proposed
		Control group	+/-1.4x	Pre & Post Adjustment	0.21	3.86	Proposed

		Control group	+/-1.3x	Pre & Post Adjustment	0.21	3.86	Same Type as Proposed
		Control group	+/-1.5x	Pre & Post Adjustment	0.21	3.86	Same Type as Proposed
		Control group	Unlimited	Pre & Post Adjustment	0.21	3.86	Same Type as Proposed
		Control group	+/-1.2x	Pre & Post Adjustment	0.21	3.86	Same Type as Proposed
		Control group	+/-2x	Pre & Post Adjustment	0.21	3.86	Same Type as Proposed
		Control group	+1.1x	Pre & Post Adjustment	1.65	4.31	Same Type as Proposed
		Control group	+1.3x	Pre & Post Adjustment	1.66	4.32	Same Type as Proposed
		Control group	+1.4x	Pre & Post Adjustment	1.66	4.32	Same Type as Proposed
	Day matching	5/20	+/-1.4x	Pre & Post Adjustment	-0.50	5.21	
		5/20	Unlimited	Pre & Post Adjustment	-0.51	5.21	
		5/20	+/-1.5x	Pre & Post Adjustment	-0.51	5.21	
		5/20	+/-2x	Pre & Post Adjustment	-0.51	5.21	
		3/20	+/-1.5x	Pre & Post Adjustment	-0.08	5.36	
		3/20	+/-2x	Pre & Post Adjustment	-0.08	5.36	
		3/20	Unlimited	Pre & Post Adjustment	-0.08	5.36	
		5/20	+/-1.3x	Pre & Post Adjustment	-0.20	5.37	
		3/20	+/-1.4x	Pre & Post Adjustment	0.04	5.40	
		10/20	+/-1.5x	Pre & Post Adjustment	-1.93	5.49	
	Weather matching	5 Day Match on CDD	+/-1.1x	Pre & Post Adjustment	-2.38	5.39	
		4 Day Match on Sum of CDH	+/-1.1x	Pre & Post Adjustment	-1.58	5.39	
		5 Day Match on Sum of CDH	+/-1.1x	Pre & Post Adjustment	-1.66	5.51	
		3 Day Match on Sum of CDH	+/-1.1x	Pre & Post Adjustment	-1.45	5.74	
		4 Day Match on CDD	+2x	Pre & Post Adjustment	2.23	5.77	
		4 Day Match on CDD	+1.5x	Pre & Post Adjustment	2.23	5.77	
		4 Day Match on CDD	+1.3x	Pre & Post Adjustment	2.23	5.77	
		4 Day Match on CDD	+1.4x	Pre & Post Adjustment	2.23	5.77	
		4 Day Match on CDD	+1.2x	Pre & Post Adjustment	2.23	5.77	
		5 Day Match on CDD	+2x	Pre & Post Adjustment	2.32	5.80	
SCE Agricultural	Day matching	Bottom 10/10	+/-1.4x	No Post Adjustment	0.51	5.05	
		10/10	+/-1.4x	No Post Adjustment	0.51	5.05	Same Type as Proposed
		Bottom 10/10	+/-1.5x	No Post Adjustment	0.50	5.05	
		10/10	+/-1.5x	No Post Adjustment	0.50	5.05	Same Type as Proposed
		Bottom 10/10	Unlimited	No Post Adjustment	0.50	5.05	
		Bottom 10/10	+/-2x	No Post Adjustment	0.50	5.05	
		10/10	+/-2x	No Post Adjustment	0.50	5.05	Same Type as Proposed
		10/10	Unlimited	No Post Adjustment	0.50	5.05	Same Type as Proposed
		Bottom 10/10	+/-1.3x	No Post Adjustment	0.55	5.06	
		10/10	+/-1.3x	No Post Adjustment	0.55	5.06	Same Type as Proposed
	Weather	Bins based on Sum of CDH	+/-2x	No Post Adjustment	0.81	5.64	

	matching	Bins based on Sum of CDH	Unlimited	No Post Adjustment	0.81	5.64	
		5 Day Match on CDD	+/-2x	No Post Adjustment	1.06	5.66	
		5 Day Match on CDD	Unlimited	No Post Adjustment	1.06	5.67	
		5 Day Match on Sum of CDH	+/-2x	No Post Adjustment	1.04	5.69	
		5 Day Match on Sum of CDH	Unlimited	No Post Adjustment	1.03	5.69	
		Bins based on CDD	+/-2x	No Post Adjustment	0.99	5.75	
		Bins based on CDD	Unlimited	No Post Adjustment	0.99	5.75	
		5 Day Match on Max Temp	+/-2x	No Post Adjustment	1.11	5.76	
		4 Day Match on CDD	+/-2x	No Post Adjustment	1.15	5.76	
SCE BIP	Day matching	Bottom 10/20	+/-2x	No Post Adjustment	0.16	2.22	
		Bottom 10/20	Unlimited	No Post Adjustment	0.16	2.22	
		Bottom 5/20	Unlimited	No Post Adjustment	-0.03	2.28	
		Bottom 5/20	+/-2x	No Post Adjustment	-0.03	2.28	
		Bottom 3/3	+/-1.4x	No Post Adjustment	0.61	2.32	
		Bottom 3/3	Unlimited	No Post Adjustment	0.61	2.32	
		Bottom 3/3	+/-2x	No Post Adjustment	0.61	2.32	
		3/3	+/-1.4x	No Post Adjustment	0.61	2.32	
		3/3	Unlimited	No Post Adjustment	0.61	2.32	
		3/3	+/-2x	No Post Adjustment	0.61	2.32	
	Weather matching	Bins based on CDD	+/-2x	No Post Adjustment	-0.08	2.31	
		Bins based on CDD	Unlimited	No Post Adjustment	-0.08	2.31	
		Bins based on Max Temp	Unlimited	No Post Adjustment	-0.13	2.34	
		Bins based on Max Temp	+/-2x	No Post Adjustment	-0.13	2.34	
		Bins based on CDD	+/-1.5x	No Post Adjustment	-0.09	2.34	
		Bins based on Sum of CDH	Unlimited	No Post Adjustment	0.08	2.37	
		Bins based on Sum of CDH	+/-2x	No Post Adjustment	0.08	2.37	
		5 Day Match on Max Temp	Unlimited	No Post Adjustment	0.15	2.40	
		5 Day Match on Max Temp	+/-2x	No Post Adjustment	0.15	2.40	
		Bins based on Max Temp	+/-1.5x	No Post Adjustment	-0.17	2.43	
SCE Comm AC Cycling	Control group	Control group	+/-1.5x	Pre & Post Adjustment	-0.18	4.24	Same Type as Proposed
		Control group	Unlimited	Pre & Post Adjustment	-0.18	4.24	Same Type as Proposed
		Control group	+/-2x	Pre & Post Adjustment	-0.18	4.24	Same Type as Proposed
		Control group	+/-1.4x	Pre & Post Adjustment	-0.17	4.25	Proposed
		Control group	+/-1.3x	Pre & Post Adjustment	-0.21	4.29	Same Type as Proposed
		Control group	+/-1.2x	Pre & Post Adjustment	-0.54	4.78	Same Type as Proposed
		Control group	+/-1.3x	No Post Adjustment	-0.66	5.64	Same Type as Proposed
		Control group	+/-2x	No Post Adjustment	-0.59	5.64	Same Type as Proposed
		Control group	Unlimited	No Post Adjustment	-0.59	5.64	Same Type as Proposed
		Control group	+/-1.5x	No Post Adjustment	-0.59	5.64	Same Type as Proposed

	Day matching	3/20	Unlimited	Pre & Post Adjustment	-0.41	6.79	
		5/20	Unlimited	Pre & Post Adjustment	0.39	6.94	
		5/20	+/-2x	Pre & Post Adjustment	0.42	6.95	
		10/20	Unlimited	Pre & Post Adjustment	1.57	6.96	
		10/20	+/-2x	Pre & Post Adjustment	1.57	6.96	
		10/20	+/-1.5x	Pre & Post Adjustment	1.58	6.96	
		10/10	+/-1.2x	Pre & Post Adjustment	1.09	7.11	Proposed
		Bottom 10/10	+/-1.2x	Pre & Post Adjustment	1.09	7.11	
		10/20	+/-1.4x	Pre & Post Adjustment	1.81	7.11	
		10/10	+/-1.2x	No Post Adjustment	2.79	7.43	Proposed
	Weather matching	Bins based on Max Temp	+/-1.3x	Pre & Post Adjustment	0.29	7.55	
		Bins based on Max Temp	+/-1.4x	Pre & Post Adjustment	0.98	7.82	
		Bins based on Max Temp	+/-1.5x	Pre & Post Adjustment	1.01	7.84	
		Bins based on Max Temp	+/-2x	Pre & Post Adjustment	1.01	7.84	
		Bins based on Max Temp	Unlimited	Pre & Post Adjustment	1.01	7.84	
		3 Day Match on Max Temp	+/-1.3x	Pre & Post Adjustment	2.80	7.85	
		Bins based on CDD	+/-1.3x	Pre & Post Adjustment	1.25	7.86	
		Bins based on Max Temp	+/-1.3x	No Post Adjustment	0.57	7.94	
		3 Day Match on Max Temp	+/-1.2x	Pre & Post Adjustment	1.83	7.96	
		Bins based on Max Temp	+/-1.2x	Pre & Post Adjustment	-0.54	8.02	
SDG&E Comm AC Cycling	Control group	Control group	+/-1.3x	Pre & Post Adjustment	0.08	2.98	Same Type as Proposed
		Control group	+/-2x	Pre & Post Adjustment	0.08	2.99	Same Type as Proposed
		Control group	Unlimited	Pre & Post Adjustment	0.08	2.99	Same Type as Proposed
		Control group	+/-1.4x	Pre & Post Adjustment	0.08	2.99	Proposed
		Control group	+/-1.5x	Pre & Post Adjustment	0.08	2.99	Same Type as Proposed
		Control group	+/-1.2x	Pre & Post Adjustment	0.08	3.10	Same Type as Proposed
		Control group	+/-2x	No Post Adjustment	0.14	3.69	Same Type as Proposed
		Control group	Unlimited	No Post Adjustment	0.14	3.69	Same Type as Proposed
		Control group	+/-1.4x	No Post Adjustment	0.14	3.69	Proposed
		Control group	+/-1.5x	No Post Adjustment	0.14	3.69	Same Type as Proposed
	Day matching	5/10	+/-1.5x	Pre & Post Adjustment	-0.25	3.59	
		5/10	+/-2x	Pre & Post Adjustment	-0.25	3.59	
		5/10	Unlimited	Pre & Post Adjustment	-0.25	3.59	
		5/10	+/-1.4x	Pre & Post Adjustment	-0.25	3.59	
		Bottom 4/5	+/-1.2x	Pre & Post Adjustment	-0.86	3.63	
		5/10	+/-1.3x	Pre & Post Adjustment	-0.14	3.66	
		Bottom 10/20	+/-1.2x	Pre & Post Adjustment	-1.05	3.67	
		3/5 weighted	+/-1.3x	No Post Adjustment	-0.54	3.67	
		3/5 weighted	+/-2x	Pre & Post Adjustment	-0.31	3.68	
		3/5 weighted	+/-1.5x	Pre & Post Adjustment	-0.31	3.68	

	Weather matching	5 Day Match on Sum of CDH	+/-1.4x	No Post Adjustment	-0.18	3.18	
		5 Day Match on Sum of CDH	Unlimited	Pre & Post Adjustment	-0.36	3.19	
		5 Day Match on Sum of CDH	+/-2x	Pre & Post Adjustment	-0.36	3.19	
		5 Day Match on Sum of CDH	+/-1.5x	Pre & Post Adjustment	-0.36	3.19	
		5 Day Match on Sum of CDH	+/-1.4x	Pre & Post Adjustment	-0.35	3.20	
		5 Day Match on CDD	Unlimited	Pre & Post Adjustment	-0.62	3.28	
		5 Day Match on CDD	+/-1.5x	Pre & Post Adjustment	-0.62	3.28	
		5 Day Match on CDD	+/-2x	Pre & Post Adjustment	-0.62	3.28	
		5 Day Match on CDD	+/-1.4x	Pre & Post Adjustment	-0.61	3.28	
		4 Day Match on Sum of CDH	+/-1.4x	No Post Adjustment	-0.09	3.35	
SDG&E Res 100% AC Cycling	Control group	Control group	+/-1.1x	Pre & Post Adjustment	-0.27	5.29	Same Type as Proposed
		Control group	+/-1.2x	Pre & Post Adjustment	-0.26	5.32	Same Type as Proposed
		Control group	+/-2x	Pre & Post Adjustment	-0.26	5.32	Same Type as Proposed
		Control group	+/-1.5x	Pre & Post Adjustment	-0.26	5.32	Same Type as Proposed
		Control group	Unlimited	Pre & Post Adjustment	-0.26	5.32	Same Type as Proposed
		Control group	+/-1.3x	Pre & Post Adjustment	-0.26	5.32	Same Type as Proposed
		Control group	+/-1.4x	Pre & Post Adjustment	-0.26	5.32	Proposed
		Control group	+1.1x	Pre & Post Adjustment	1.49	5.56	Same Type as Proposed
		Control group	+2x	Pre & Post Adjustment	1.53	5.58	Same Type as Proposed
		Control group	+1.2x	Pre & Post Adjustment	1.53	5.58	Same Type as Proposed
	Day matching	3/5 weighted	+1.5x	Pre & Post Adjustment	4.24	12.76	
		3/5 weighted	+2x	Pre & Post Adjustment	4.29	12.80	
		3/5 weighted	+1.4x	Pre & Post Adjustment	2.70	13.10	
		3/5 weighted	+/-1.5x	Pre & Post Adjustment	3.28	13.32	
		3/5 weighted	+/-2x	Pre & Post Adjustment	3.33	13.36	
		3/5 weighted	Unlimited	Pre & Post Adjustment	3.33	13.36	
		3/5 weighted	+/-1.4x	Pre & Post Adjustment	1.75	13.65	
		3/3 weighted	+2x	Pre & Post Adjustment	3.63	13.77	
		3/3 weighted	Unlimited	Pre & Post Adjustment	2.73	14.32	
		3/3 weighted	+/-2x	Pre & Post Adjustment	2.73	14.32	
	Weather matching	4 Day Match on CDD	+/-1.1x	Pre & Post Adjustment	-2.59	15.94	
		5 Day Match on CDD	+/-1.1x	Pre & Post Adjustment	-2.34	16.08	
		Bins based on Max Temp	+1.5x	Pre & Post Adjustment	-5.44	16.23	
		Bins based on Max Temp	+/-1.5x	Pre & Post Adjustment	-5.44	16.23	
		4 Day Match on CDD	+1.1x	Pre & Post Adjustment	-1.67	16.25	
		5 Day Match on CDD	+1.1x	Pre & Post Adjustment	-1.17	16.33	
		Bins based on CDD	+1.5x	Pre & Post Adjustment	-0.20	16.54	
		Bins based on CDD	+2x	Pre & Post Adjustment	-0.18	16.54	
		Bins based on CDD	+/-1.5x	Pre & Post Adjustment	-0.26	16.55	

		Bins based on CDD	+/-2x	Pre & Post Adjustment	-0.24	16.55	
SDG&E Res 50% AC Cycling	Control group	Control group	+/-1.1x	Pre & Post Adjustment	0.04	4.19	Same Type as Proposed
		Control group	+/-1.2x	Pre & Post Adjustment	0.02	4.23	Same Type as Proposed
		Control group	+/-1.4x	Pre & Post Adjustment	0.02	4.23	Proposed
		Control group	Unlimited	Pre & Post Adjustment	0.02	4.23	Same Type as Proposed
		Control group	+/-1.3x	Pre & Post Adjustment	0.02	4.23	Same Type as Proposed
		Control group	+/-2x	Pre & Post Adjustment	0.02	4.23	Same Type as Proposed
		Control group	+/-1.5x	Pre & Post Adjustment	0.02	4.23	Same Type as Proposed
		Control group	+1.1x	Pre & Post Adjustment	1.85	4.66	Same Type as Proposed
		Control group	+1.2x	Pre & Post Adjustment	1.89	4.68	Same Type as Proposed
		Control group	+2x	Pre & Post Adjustment	1.89	4.68	Same Type as Proposed
	Day matching	3/5 weighted	+2x	Pre & Post Adjustment	2.81	8.20	
		3/5 weighted	+1.5x	Pre & Post Adjustment	2.59	8.38	
		4/5	Unlimited	Pre & Post Adjustment	1.23	8.50	
		4/5	+/-2x	Pre & Post Adjustment	1.23	8.50	
		3/5 weighted	+/-2x	Pre & Post Adjustment	1.65	8.55	
		3/5 weighted	Unlimited	Pre & Post Adjustment	1.65	8.55	
		4/5	+2x	Pre & Post Adjustment	1.47	8.58	
		3/5	Unlimited	Pre & Post Adjustment	2.34	8.62	
		3/5	+/-2x	Pre & Post Adjustment	2.34	8.62	
		3/5 - +5%	Unlimited	Pre & Post Adjustment	2.34	8.62	
	Weather matching	Bins based on CDD	+/-2x	Pre & Post Adjustment	-1.57	8.26	
		Bins based on CDD	Unlimited	Pre & Post Adjustment	-1.57	8.26	
		Bins based on CDD	+2x	Pre & Post Adjustment	-1.56	8.26	
		Bins based on CDD	+/-1.5x	Pre & Post Adjustment	-1.59	8.28	
		Bins based on CDD	+1.5x	Pre & Post Adjustment	-1.58	8.28	
		4 Day Match on CDD	+/-1.5x	Pre & Post Adjustment	-1.38	8.78	
		4 Day Match on CDD	+/-1.4x	Pre & Post Adjustment	-1.38	8.78	
		4 Day Match on CDD	+/-1.3x	Pre & Post Adjustment	-1.38	8.78	
		4 Day Match on CDD	+/-2x	Pre & Post Adjustment	-1.38	8.78	
		4 Day Match on CDD	Unlimited	Pre & Post Adjustment	-1.38	8.78	

Table A-6: Event Day Results

Program	Baseline Category	Type	Adjustment Cap	Adjustment Type	MPE (%)	CVRM SE (%)	Recommended
PG&E Res AC Cycling	Day matching	10/20	Unlimited	Pre & Post Adjustment	0.66	6.84	
		10/20	+/-2x	Pre & Post Adjustment	0.66	6.84	
		10/20	+2x	Pre & Post Adjustment	0.66	6.84	
		10/20	+/-1.5x	Pre & Post Adjustment	0.52	6.89	

		10/20	+1.5x	Pre & Post Adjustment	0.52	6.89	
		5/10	Unlimited	Pre & Post Adjustment	1.60	7.13	Same Type as Proposed
		5/10	+/-2x	Pre & Post Adjustment	1.60	7.13	Same Type as Proposed
		5/10	+2x	Pre & Post Adjustment	1.76	7.31	Same Type as Proposed
		5/20	+/-1.4x	Pre & Post Adjustment	2.59	7.36	
		5/20	Unlimited	Pre & Post Adjustment	2.59	7.36	
	Weather matching	Bins based on Sum of CDH	+/-1.4x	Pre & Post Adjustment	1.06	7.37	
		Bins based on Sum of CDH	Unlimited	Pre & Post Adjustment	1.06	7.37	
		Bins based on Sum of CDH	+/-1.5x	Pre & Post Adjustment	1.06	7.37	
		Bins based on Sum of CDH	+/-2x	Pre & Post Adjustment	1.06	7.37	
		Bins based on Sum of CDH	+/-1.3x	Pre & Post Adjustment	1.06	7.38	
		Bins based on Max Temp	+/-1.3x	Pre & Post Adjustment	2.40	7.40	
		Bins based on Max Temp	+/-1.5x	Pre & Post Adjustment	2.27	7.42	
		Bins based on Max Temp	+/-2x	Pre & Post Adjustment	2.27	7.42	
		Bins based on Max Temp	+/-1.4x	Pre & Post Adjustment	2.27	7.42	
		Bins based on Max Temp	Unlimited	Pre & Post Adjustment	2.27	7.42	
SDG&E Res 100% AC Cycling	Day matching	10/20	+1.4x	No Post Adjustment	-0.76	10.50	
		10/20	+/-1.4x	No Post Adjustment	-0.76	10.51	
		10/20	+/-1.4x	Pre & Post Adjustment	0.06	10.89	
		10/20	+1.4x	Pre & Post Adjustment	0.06	10.89	
		10/20	+1.5x	No Post Adjustment	3.80	11.49	
		10/20	+/-1.5x	No Post Adjustment	3.80	11.49	
		5/20	+/-1.3x	No Post Adjustment	1.67	11.93	
		5/20	+1.3x	No Post Adjustment	2.02	12.04	
		10/20	+/-1.5x	Pre & Post Adjustment	4.99	12.11	
		10/20	+1.5x	Pre & Post Adjustment	4.99	12.11	
	Weather matching	Bins based on CDD	+1.5x	No Post Adjustment	2.24	15.68	
		Bins based on CDD	+1.4x	No Post Adjustment	-0.14	16.26	
		5 Day Match on CDD	+/-1.4x	No Post Adjustment	-0.83	17.04	
		5 Day Match on CDD	+1.4x	No Post Adjustment	-0.83	17.04	
		5 Day Match on CDD	+1.4x	Pre & Post Adjustment	0.22	17.35	
		5 Day Match on CDD	+/-1.4x	Pre & Post Adjustment	0.22	17.35	
		Bins based on CDD	+1.3x	No Post Adjustment	-3.14	17.52	
		5 Day Match on CDD	+1.5x	No Post Adjustment	3.12	17.56	
		5 Day Match on CDD	+/-1.5x	No Post Adjustment	3.12	17.56	
		5 Day Match on CDD	+/-1.5x	Pre & Post Adjustment	4.31	17.95	
SDG&E Res 50% AC Cycling	Day matching	5/10	+1.4x	Pre & Post Adjustment	-0.52	7.69	Same Type as Proposed
		5/10	+/-1.4x	Pre & Post Adjustment	-0.64	7.74	Proposed
		5/20	+1.3x	Pre & Post Adjustment	-1.54	8.36	
		5/20	+/-1.3x	Pre & Post Adjustment	-1.84	8.37	

		5/20	+1.4x	Pre & Post Adjustment	2.89	8.41	
		5/20	+/-1.4x	Pre & Post Adjustment	2.59	8.42	
		5/20	+1.4x	No Post Adjustment	2.63	8.60	
		10/20	+1.5x	Pre & Post Adjustment	0.12	8.65	
		10/20	+/-1.5x	Pre & Post Adjustment	0.12	8.65	
		3/20	+/-1.3x	Pre & Post Adjustment	1.21	8.66	
	Weather matching	4 Day Match on CDD	+1.4x	No Post Adjustment	-0.01	8.71	
		4 Day Match on CDD	+/-1.4x	No Post Adjustment	-0.13	8.89	
		4 Day Match on CDD	+1.4x	Pre & Post Adjustment	0.15	9.12	
		4 Day Match on CDD	+/-1.4x	Pre & Post Adjustment	0.15	9.12	
		3 Day Match on CDD	+1.4x	No Post Adjustment	-0.52	9.79	
		5 Day Match on CDD	+/-1.5x	Pre & Post Adjustment	0.36	9.88	
		5 Day Match on CDD	+1.5x	Pre & Post Adjustment	0.36	9.88	
		4 Day Match on CDD	+1.3x	No Post Adjustment	-4.59	10.07	
		4 Day Match on CDD	+/-1.3x	No Post Adjustment	-4.72	10.23	
		3 Day Match on CDD	+/-1.4x	No Post Adjustment	-1.00	10.26	

Table A-7: Proxy Weekend Results

Program	Baseline Category	Type	Adjustment Cap	Adjustment Type	MPE (%)	CVRM SE (%)	Recommended
PG&E BIP	Day matching	5/5	+/-1.8x	Pre & Post Adjustment	0.26	2.99	
		5/5	+/-1.4x	Pre & Post Adjustment	0.26	2.99	
		Bottom 5/5	+/-2x	Pre & Post Adjustment	0.26	2.99	
		5/5	+/-1.9x	Pre & Post Adjustment	0.26	2.99	
		Bottom 5/5	+/-1.5x	Pre & Post Adjustment	0.26	2.99	
		5/5	+/-2x	Pre & Post Adjustment	0.26	2.99	
		Bottom 5/5	+/-1.9x	Pre & Post Adjustment	0.26	2.99	
		Bottom 5/5	+/-1.8x	Pre & Post Adjustment	0.26	2.99	
		Bottom 5/5	+/-1.7x	Pre & Post Adjustment	0.26	2.99	
		Bottom 5/5	+/-1.6x	Pre & Post Adjustment	0.26	2.99	
	Weather matching	5 Day Match on Sum CDH	+/-1.8x	Pre & Post Adjustment	-0.06	3.00	
		5 Day Match on Sum CDH	+/-1.7x	Pre & Post Adjustment	-0.06	3.00	
		5 Day Match on Sum CDH	+/-1.9x	Pre & Post Adjustment	-0.06	3.00	
		5 Day Match on Sum CDH	+/-1.6x	Pre & Post Adjustment	-0.06	3.00	
		5 Day Match on Sum CDH	+/-2x	Pre & Post Adjustment	-0.06	3.00	
		5 Day Match on Sum CDH	+/-1.5x	Pre & Post Adjustment	-0.07	3.03	
		5 Day Match on CDD	+/-1.6x	Pre & Post Adjustment	-0.02	3.04	
		5 Day Match on CDD	+/-2x	Pre & Post Adjustment	-0.02	3.04	
		5 Day Match on CDD	+/-1.8x	Pre & Post Adjustment	-0.02	3.04	

		5 Day Match on CDD	+/-1.7x	Pre & Post Adjustment	-0.02	3.04	
PG&E Res AC Cycling	Control group	Control group	+/-1.1x	Pre & Post Adjustment	-0.13	4.14	Same Type as Proposed
		Control group	+/-2x	Pre & Post Adjustment	-0.13	4.16	Same Type as Proposed
		Control group	+/-1.4x	Pre & Post Adjustment	-0.13	4.16	Proposed
		Control group	+/-1.2x	Pre & Post Adjustment	-0.13	4.16	Same Type as Proposed
		Control group	+/-1.8x	Pre & Post Adjustment	-0.13	4.16	Same Type as Proposed
		Control group	+/-1.6x	Pre & Post Adjustment	-0.13	4.16	Same Type as Proposed
		Control group	+/-1.7x	Pre & Post Adjustment	-0.13	4.16	Same Type as Proposed
		Control group	+/-1.9x	Pre & Post Adjustment	-0.13	4.16	Same Type as Proposed
		Control group	+/-1.3x	Pre & Post Adjustment	-0.13	4.16	Same Type as Proposed
		Control group	+/-1.5x	Pre & Post Adjustment	-0.13	4.16	Same Type as Proposed
	Day matching	1/4	+/-1.4x	No Post Adjustment	-0.68	7.84	
		1/4	+/-1.7x	No Post Adjustment	-0.69	7.84	
		1/4	+/-2x	No Post Adjustment	-0.69	7.84	
		1/4	+/-1.5x	No Post Adjustment	-0.69	7.84	
		1/4	+/-1.9x	No Post Adjustment	-0.69	7.84	
		1/4	+/-1.8x	No Post Adjustment	-0.69	7.84	
		1/4	+/-1.6x	No Post Adjustment	-0.69	7.84	
		1/5	+/-1.8x	No Post Adjustment	-0.46	7.98	
		1/5	+/-1.9x	No Post Adjustment	-0.46	7.98	
		1/5	+/-2x	No Post Adjustment	-0.46	7.98	
	Weather matching	3 Day Match on Max Temp	+/-1.2x	No Post Adjustment	-0.88	5.02	
		5 Day Match on Max Temp	+/-1.2x	No Post Adjustment	-2.21	5.51	
		5 Day Match on Max Temp	+/-1.2x	Pre & Post Adjustment	1.64	5.87	
		5 Day Match on Max Temp	+/-2x	Pre & Post Adjustment	1.70	5.88	
		5 Day Match on Max Temp	+/-1.7x	Pre & Post Adjustment	1.70	5.88	
		5 Day Match on Max Temp	+/-1.9x	Pre & Post Adjustment	1.70	5.88	
		5 Day Match on Max Temp	+/-1.3x	Pre & Post Adjustment	1.70	5.88	
		5 Day Match on Max Temp	+/-1.6x	Pre & Post Adjustment	1.70	5.88	
		5 Day Match on Max Temp	+/-1.4x	Pre & Post Adjustment	1.70	5.88	
		5 Day Match on Max Temp	+/-1.5x	Pre & Post Adjustment	1.70	5.88	
SCE Agricultural	Day matching	5/5	+/-1.6x	No Post Adjustment	-0.60	6.02	
		Bottom 5/5	+/-1.6x	No Post Adjustment	-0.60	6.02	
		5/5	+/-1.7x	No Post Adjustment	-0.64	6.02	
		Bottom 5/5	+/-1.7x	No Post Adjustment	-0.64	6.02	
		Bottom 5/5	+/-1.9x	No Post Adjustment	-0.67	6.03	
		5/5	+/-1.9x	No Post Adjustment	-0.67	6.03	
		5/5	+/-1.8x	No Post Adjustment	-0.66	6.03	
		Bottom 5/5	+/-1.8x	No Post Adjustment	-0.66	6.03	
		5/5	+/-2x	No Post Adjustment	-0.68	6.03	
		Bottom 5/5	+/-2x	No Post Adjustment	-0.68	6.03	

	Weather matching	Bins based on CDD	+/-2x	Pre & Post Adjustment	-1.10	5.30	
		Bins based on CDD	+/-1.9x	Pre & Post Adjustment	-1.10	5.30	
		Bins based on CDD	+/-1.8x	Pre & Post Adjustment	-1.10	5.31	
		Bins based on CDD	+/-1.7x	Pre & Post Adjustment	-1.09	5.31	
		Bins based on CDD	+/-1.6x	Pre & Post Adjustment	-1.09	5.31	
		Bins based on CDD	+/-1.5x	Pre & Post Adjustment	-1.06	5.32	
		Bins based on CDD	+/-1.4x	Pre & Post Adjustment	-0.98	5.36	
		Bins based on CDD	+/-1.3x	Pre & Post Adjustment	-0.74	5.52	
		Bins based on Sum of CDH	+/-2x	No Post Adjustment	-0.49	5.56	
		Bins based on Sum of CDH	+/-1.9x	No Post Adjustment	-0.49	5.56	
SCE BIP	Day matching	Bottom 5/5	+/-2x	Pre & Post Adjustment	-0.05	2.07	
		Bottom 5/5	+/-1.9x	Pre & Post Adjustment	-0.05	2.07	
		Bottom 5/5	+/-1.8x	Pre & Post Adjustment	-0.05	2.07	
		5/5	+/-1.6x	Pre & Post Adjustment	-0.05	2.07	
		5/5	+/-1.5x	Pre & Post Adjustment	-0.05	2.07	
		Bottom 5/5	+/-1.7x	Pre & Post Adjustment	-0.05	2.07	
		Bottom 5/5	+/-1.6x	Pre & Post Adjustment	-0.05	2.07	
		5/5	+/-2x	Pre & Post Adjustment	-0.05	2.07	
		5/5	+/-1.8x	Pre & Post Adjustment	-0.05	2.07	
		Bottom 5/5	+/-1.5x	Pre & Post Adjustment	-0.05	2.07	
	Weather matching	Bins based on Max Temp	+/-1.6x	Pre & Post Adjustment	-0.73	2.05	
		Bins based on Max Temp	+/-2x	Pre & Post Adjustment	-0.73	2.05	
		Bins based on Max Temp	+/-1.7x	Pre & Post Adjustment	-0.73	2.05	
		Bins based on Max Temp	+/-1.9x	Pre & Post Adjustment	-0.73	2.05	
		Bins based on Max Temp	+/-1.5x	Pre & Post Adjustment	-0.73	2.05	
		Bins based on Max Temp	+/-1.8x	Pre & Post Adjustment	-0.73	2.05	
		Bins based on Max Temp	+/-1.4x	Pre & Post Adjustment	-0.72	2.05	
		Bins based on CDD	+/-1.4x	Pre & Post Adjustment	-0.71	2.06	
		Bins based on CDD	+/-1.7x	Pre & Post Adjustment	-0.71	2.06	
		Bins based on CDD	+/-1.9x	Pre & Post Adjustment	-0.71	2.06	
SCE Comm AC Cycling	Control group	Control group	+/-2x	Pre & Post Adjustment	0.91	11.72	Same Type as Proposed
		Control group	+/-1.9x	Pre & Post Adjustment	1.02	12.19	Same Type as Proposed
		Control group	+/-1.8x	Pre & Post Adjustment	1.12	12.85	Same Type as Proposed
		Control group	+/-1.7x	Pre & Post Adjustment	1.25	13.71	Same Type as Proposed
		Control group	+/-2x	No Post Adjustment	0.65	14.77	Same Type as Proposed
		Control group	+/-1.6x	Pre & Post Adjustment	1.38	14.84	Same Type as Proposed
		Control group	+/-1.9x	No Post Adjustment	0.71	15.01	Same Type as Proposed
		Control group	+/-1.8x	No Post Adjustment	0.75	15.41	Same Type as Proposed
		Control group	+/-1.7x	No Post Adjustment	0.81	15.98	Same Type as Proposed
		Control group	+/-1.5x	Pre & Post Adjustment	1.52	16.43	Same Type as Proposed

	Day matching	Bottom 2/4	+/-1.1x	Pre & Post Adjustment	0.79	4.75	
		Bottom 3/4	+/-1.8x	Pre & Post Adjustment	1.40	4.76	
		Bottom 3/4	+/-1.9x	Pre & Post Adjustment	1.40	4.76	
		Bottom 3/4	+/-1.5x	Pre & Post Adjustment	1.40	4.76	
		Bottom 3/4	+/-1.6x	Pre & Post Adjustment	1.40	4.76	
		Bottom 3/4	+/-1.7x	Pre & Post Adjustment	1.40	4.76	
		Bottom 3/4	+/-2x	Pre & Post Adjustment	1.40	4.76	
		Bottom 3/4	+/-1.4x	Pre & Post Adjustment	1.40	4.76	
		Bottom 3/4	+/-1.3x	Pre & Post Adjustment	1.40	4.76	
		Bottom 3/4	+/-1.2x	Pre & Post Adjustment	1.44	4.79	
	Weather matching	Bins based on Sum of CDH	+/-1.2x	Pre & Post Adjustment	0.91	3.35	
		Bins based on Sum of CDH	+/-1.6x	Pre & Post Adjustment	0.95	3.37	
		Bins based on Sum of CDH	+/-1.7x	Pre & Post Adjustment	0.95	3.37	
		Bins based on Sum of CDH	+/-1.4x	Pre & Post Adjustment	0.95	3.37	
		Bins based on Sum of CDH	+/-1.3x	Pre & Post Adjustment	0.95	3.37	
		Bins based on Sum of CDH	+/-1.8x	Pre & Post Adjustment	0.95	3.37	
		Bins based on Sum of CDH	+/-2x	Pre & Post Adjustment	0.95	3.37	
		Bins based on Sum of CDH	+/-1.9x	Pre & Post Adjustment	0.95	3.37	
		Bins based on Sum of CDH	+/-1.5x	Pre & Post Adjustment	0.95	3.37	
		Bins based on Max Temp	+/-1.2x	Pre & Post Adjustment	0.92	3.37	
SDG&E Comm AC Cycling	Control group	Control group	+/-2x	Pre & Post Adjustment	0.23	7.35	Same Type as Proposed
		Control group	+/-1.9x	Pre & Post Adjustment	0.22	7.35	Same Type as Proposed
		Control group	+/-1.8x	Pre & Post Adjustment	0.21	7.39	Same Type as Proposed
		Control group	+/-1.7x	Pre & Post Adjustment	0.14	7.42	Same Type as Proposed
		Control group	+/-1.6x	Pre & Post Adjustment	0.03	7.51	Same Type as Proposed
		Control group	+/-1.5x	Pre & Post Adjustment	-0.09	7.75	Same Type as Proposed
		Control group	+/-1.4x	Pre & Post Adjustment	-0.26	8.36	Proposed
		Control group	+/-2x	No Post Adjustment	0.06	8.83	Same Type as Proposed
		Control group	+/-1.9x	No Post Adjustment	0.04	8.86	Same Type as Proposed
		Control group	+/-1.8x	No Post Adjustment	0.01	8.89	Same Type as Proposed
	Day matching	3/4	+/-1.5x	No Post Adjustment	-1.78	4.72	
		3/4	+/-1.6x	No Post Adjustment	-1.60	4.78	
		3/4	+/-1.7x	No Post Adjustment	-1.60	4.78	
		3/4	+/-1.8x	No Post Adjustment	-1.60	4.78	
		3/4	+/-1.9x	No Post Adjustment	-1.60	4.78	
		3/4	+/-2x	No Post Adjustment	-1.60	4.78	
		2/4	+/-1.4x	No Post Adjustment	-2.09	4.84	
		2/4	+/-1.5x	No Post Adjustment	-1.34	4.88	
		2/4	+/-1.6x	No Post Adjustment	-1.30	4.89	
		2/4	+/-1.8x	No Post Adjustment	-1.30	4.89	

	Weather matching	5 Day Match on CDD	+/-1.8x	Pre & Post Adjustment	-1.20	3.82	
		5 Day Match on CDD	+/-2x	Pre & Post Adjustment	-1.20	3.82	
		5 Day Match on CDD	+/-1.5x	Pre & Post Adjustment	-1.20	3.82	
		5 Day Match on CDD	+/-1.4x	Pre & Post Adjustment	-1.20	3.82	
		5 Day Match on CDD	+/-1.6x	Pre & Post Adjustment	-1.20	3.82	
		5 Day Match on CDD	+/-1.9x	Pre & Post Adjustment	-1.20	3.82	
		5 Day Match on CDD	+/-1.7x	Pre & Post Adjustment	-1.20	3.82	
		5 Day Match on CDD	+/-1.3x	Pre & Post Adjustment	-1.50	4.07	
		4 Day Match on CDD	+/-1.7x	Pre & Post Adjustment	-1.27	4.09	
		4 Day Match on CDD	+/-1.4x	Pre & Post Adjustment	-1.27	4.09	
SDG&E Res AC Cycling	Control group	Control group	+/-1.1x	Pre & Post Adjustment	-0.04	6.00	Same Type as Proposed
		Control group	+/-1.6x	Pre & Post Adjustment	0.02	6.17	Same Type as Proposed
		Control group	+/-1.2x	Pre & Post Adjustment	0.02	6.17	Same Type as Proposed
		Control group	+/-1.4x	Pre & Post Adjustment	0.02	6.17	Proposed
		Control group	+/-1.9x	Pre & Post Adjustment	0.02	6.17	Same Type as Proposed
		Control group	+/-2x	Pre & Post Adjustment	0.02	6.17	Same Type as Proposed
		Control group	+/-1.5x	Pre & Post Adjustment	0.02	6.17	Same Type as Proposed
		Control group	+/-1.3x	Pre & Post Adjustment	0.02	6.17	Same Type as Proposed
		Control group	+/-1.8x	Pre & Post Adjustment	0.02	6.17	Same Type as Proposed
		Control group	+/-1.7x	Pre & Post Adjustment	0.02	6.17	Same Type as Proposed
	Day matching	Bottom 4/5	+/-1x	Pre & Post Adjustment	3.56	9.26	
		Bottom 4/5	Unadjusted	No Post Adjustment	3.56	9.26	
		Bottom 4/5	+/-1x	No Post Adjustment	3.56	9.26	
		Bottom 4/5	Unadjusted	Pre & Post Adjustment	3.56	9.26	
		4/4	+/-1.3x	Pre & Post Adjustment	0.44	9.63	
		Bottom 4/4	+/-1.3x	Pre & Post Adjustment	0.44	9.63	
		Bottom 3/4	+/-1.1x	Pre & Post Adjustment	-1.63	9.71	
		Bottom 3/4	+/-1.1x	No Post Adjustment	-1.63	9.71	
		Bottom 2/3	+/-1.2x	Pre & Post Adjustment	2.32	9.78	
		5/5	+/-1.2x	Pre & Post Adjustment	3.85	9.79	
	Weather matching	5 Day Match on Sum CDH	+/-1.2x	No Post Adjustment	5.24	14.83	
		5 Day Match on Sum CDH	+/-1.3x	No Post Adjustment	6.05	15.05	
		5 Day Match on Sum CDH	+/-1.5x	No Post Adjustment	6.05	15.05	
		5 Day Match on Sum CDH	+/-1.6x	No Post Adjustment	6.05	15.05	
		5 Day Match on Sum CDH	+/-1.8x	No Post Adjustment	6.05	15.05	
		5 Day Match on Sum CDH	+/-1.9x	No Post Adjustment	6.05	15.05	
		5 Day Match on Sum CDH	+/-2x	No Post Adjustment	6.05	15.05	
		5 Day Match on Sum CDH	+/-1.7x	No Post Adjustment	6.05	15.05	
		5 Day Match on Sum CDH	+/-1.4x	No Post Adjustment	6.05	15.05	
		5 Day Match on Max Temp	+/-1.4x	No Post Adjustment	-0.07	15.35	