FREEMAN, SULLIVAN & CO.

A MEMBER OF THE FSC GROUP

2011 Statewide Evaluation of California Aggregator Demand Response Programs

Volume II: Baseline Calculation Rules and Accuracy

June 1, 2012

Prepared for:
Pacific Gas and Electric Company
Southern California Edison
San Diego Gas & Electric

Prepared by:
Stephen George, PhD.
Josh Bode, MPP
Dries Berghman, MPP

Freeman, Sullivan & Co.
101 Montgomery St., 15th Floor
San Francisco, CA 94104
fscgroup.com

Table of Contents

Executive Summary

This report provides an assessment of the extent to which modifications to the current settlement baseline calculation rules improves settlement accuracy for aggregator demand response (DR) programs operated by the three California investor-owned utilities (IOUs), Pacific Gas and Electric (PG&E), Southern California Edison (SCE) and San Diego Gas and Electric (SDG&E). Specifically, the assessment includes the statewide Capacity Bidding Program (CBP), which is operated by all three IOUs, PG&E's Aggregator Managed Portfolio (AMP) and SCE's Demand Response Resource Contracts (DRRC). Collectively, these programs are referred to as aggregator programs.

This report is the second of two volumes documenting the ex post impact evaluation of the aggregator programs for 2011. Volume 1 contains the ex post impact estimates for each 2011 event for each program and resource option (e.g., day-of or day ahead). In this report, various baseline methods are tested on both actual 2011 event days, in which the baseline impact estimates are compared to the regression based estimates presented in Volume I, and on event-like days, in which baseline estimates are compared with simulated impacts that are known with certainty. In total, 32 baseline methods are evaluated. In each case, the baseline is calculated using the same-hour average of the 10 weekdays immediately preceding the curtailment event in question, excluding other event days and holidays – also known as a 10-in-10 baseline. However, the baselines vary across the following three parameters:

- *Aggregate versus individual baseline calculation*: With the individual approach, the baseline is calculated for each individual customer, and the same-day adjustments (and caps) are applied at the individual customer level. The individual customer results are then aggregated for each settlement portfolio. With the aggregate calculation approach, customer loads are first aggregated for each settlement portfolio. The baseline is then calculated based on the aggregated loads and the same-day adjustments (and caps) are applied using the aggregated loads.
- *Universal versus optional application of same-day adjustments:* Baselines are either adjusted for all settlement portfolios, or the aggregator can elect prior to events whether or not to apply same-day adjustments.
- *The magnitude of same-day adjustment caps:* A total of eight adjustment caps are tested, including no adjustment, ±20%, ±30%, ±35%, ±40%, ±50%, -50% to 200%, and unlimited adjustments.

For most aggregator program options, the current baseline underestimates the demand reductions both in comparison to the 2011 evaluation results and under idealized simulation settings where baseline bias can be calculated. For most aggregator program options, the difference between the current baseline estimates and the 2011 evaluation results parallels the estimated baseline bias. However, no single change reduces the degree of baseline bias across all aggregator program options.

Key findings include:

- *Applying same-day adjustments universally increases accuracy.* Almost every single table included in this report shows that applying any kind of same-day adjustment improves results. A substantial share of the underestimation by the current baseline method is, in fact, due to aggregator decisions not to apply same-day adjustment for all customers in their portfolios. Making same-day adjustments universal generally improves the accuracy of the demand reduction estimate. It almost always *increases* the amount of the overall reduction measured by the baseline.

- *Calculating adjustments at the settlement portfolio level generally does not noticeably decrease or increase bias in the impact estimate, but it does reduce the magnitude of same-day adjustments.* Aggregating loads first reduces much of the individual customer volatility, makes loads more predictable and leads to smaller same-day adjustments. The higher the number of accounts in a portfolio, the less volatile the load. That is, less extreme same-day adjustments are required if an aggregator portfolio is large. This is because the load for a large, aggregated group of customers is much easier to predict than the load for one individual customer.
- *Increasing the adjustment cap generally improves the accuracy of the results*. This result varies by program option. For some program options, they make no or little difference. When a higher adjustment cap affects accuracy, it typically leads to more accurate results. However, removing the adjustment cap altogether is not advisable because it can cause extreme adjustments for some customers on some days. These outliers can be large enough to affect the overall program results.
- *For SCE and SDG&E's programs, most of the discrepancy between regression-based impacts and baseline-measured impacts can be explained by bias in the baseline method*. In PG&E's case, some of this discrepancy is also due to the fact that some customers change their consumption behavior in anticipation of an event: they either begin reducing load early, or increase their load in pre-event hours, affecting the same-day adjustment.
- *Baseline bias is minimized when customers and aggregators commit to larger demand reductions, as a percentage of their load, during event windows*.

1 Introduction and Purpose of Study

This report is Volume 2 of the 2011 ex post impact evaluation for the aggregator programs offered by PG&E, SCE and SDG&E. The analysis covers the day ahead and same-day notification options for the statewide Capacity Bidding Program (CBP), PG&E's Aggregator Managed Portfolio (AMP) and SCE's Demand Response Resource Contracts (DRRC). This volume provides an assessment of the extent to which modifications to the current baseline rules would improve the accuracy of the demand reductions calculated for settlement purposes. The assessment is conducted with both actual 2011 event days, which reflect customer behavior, and event-like days where the degree of bias can be estimated under laboratory-like settings. The analysis is intended to comply with the CPUC decision D.12-04-045, which directed California's investor owned utilities (IOUs) to compare the load impact evaluation results to the demand reductions calculated for settlement purposes.

In this volume, we answer the following questions about baseline methodology:

- Does allowing aggregators to elect whether or not in-day adjustments are applied affect the accuracy of the demand reductions estimated for settlement? Does accuracy improve or deteriorate if same-day adjustments are applied universally?
- Does calculating the baselines and applying the same-day adjustment cap in aggregate rather than individually affect the accuracy of the demand reductions estimated for settlement? Does it affect the magnitude of same-day adjustments?
- Does increasing the cap on the magnitude of in-day adjustments affect the accuracy of the demand reductions calculated for settlement?

The remainder of this section provides an introduction to how baselines are calculated and adjusted; a brief summary of the current baseline method used by the IOUs; a description of the baselines methods tested; and the structure for the remainder of this report.

1.1 Introduction to Baselines

Historically, demand reductions for settlement have been calculated by using information about electricity use patterns during the days preceding an event. Electricity use on days preceding an event is averaged to create an estimate – a *baseline* – of what electricity use *would have been* if the event had not occurred.[1] Using this approach, demand reductions are calculated by subtracting the load on the event day from the baseline. In other words, baselines are a tool to estimate demand reductions, which cannot be observed directly.

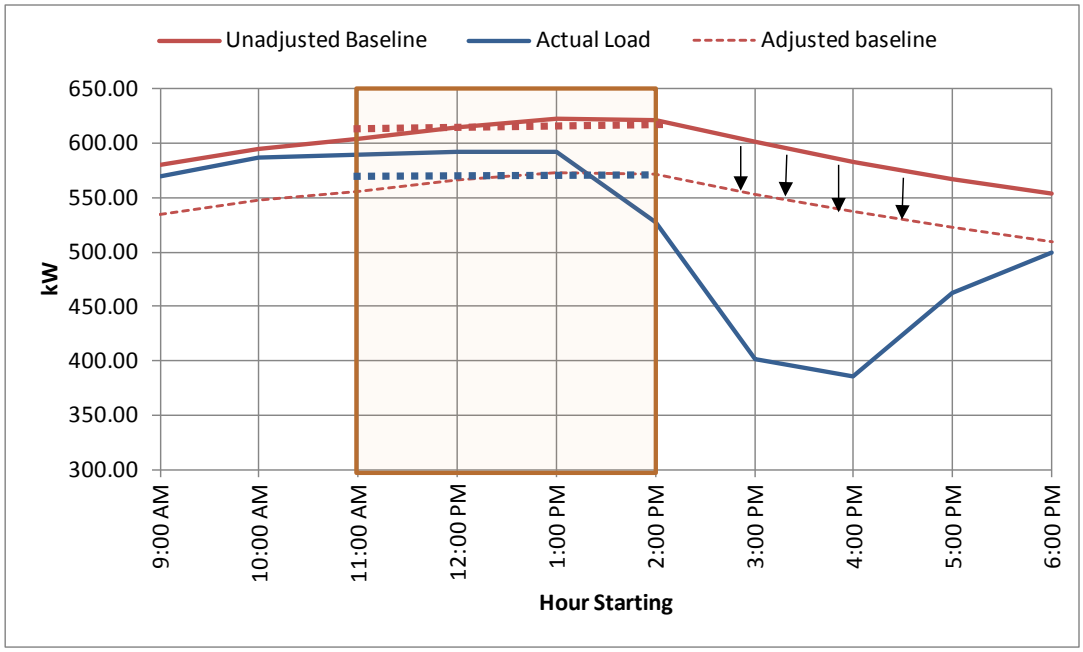
Baseline estimates of electricity use during an event period can be adjusted up or down based on electricity use patterns during the hours leading up to an event. This procedure is known as *same-day adjustment*. If, during pre-event hours, the baseline is less than the actual load, it is adjusted upwards. Similarly, if the pre-event baseline is above the actual load before the event, it is adjusted downwards. To adjust the load, the initial baseline value is multiplied by the ratio between the unadjusted baseline and the actual load during pre-event hours. In other words, the baseline is calibrated to match actual usage patterns in the hours leading up to the event. Note that the same-

[1] This approach is also sometimes referred to as the "representative day approach" or as a "day-matching baseline." For simplicity we refer to them as "baselines" throughout this report.

day adjustment procedure implicitly assumes that differences between the baseline and actual loads during hours leading up to an event are due to predictive error and, *not* due to customer behavior such as shifting of production to pre-event hours or implementing demand reductions early.

Figure 1-1 illustrates the baseline adjustment process. In the example, the event starts at 3 PM. The first three of the four hours leading up to the event, from 11 AM to 2 PM, are used to calculate the adjustment. The blue line represents the actual load for the day. The red line reflects the calculated baseline prior to the application of same-day adjustments. In this example, in the hours leading up to the event, the unadjusted baseline is higher than the actual load. The baseline adjustment process assumes this difference is due to error. To correct for this difference, the baseline is calibrated downward by roughly 8%, as reflected by the red dotted line.

Figure 1-1: Example of Baseline Same-day Adjustment



$$Adjustment = \frac{Avg.kW\ during\ adjusment\ period}{Avg.unadjusted\ baseline\ over\ adjustment\ period} = \frac{571}{619} = 92.2\%$$

If the difference between the unadjusted baseline and the actual load is truly due to baseline estimation error, the adjustment process reduces those errors. However, if the difference is actually caused by a behavioral change, the procedure can lead to incorrect adjustments. For example, customers can shift production to pre-event hours in anticipation of a curtailment event or, conversely, can start reducing demand early in order to have enough lead time to fully deliver their commitment. In both cases, the same-day adjustment confounds legitimate changes in electricity use during pre-event hours with error, leading to incorrect adjustments. When customers shift load to pre-event hours, increasing their load relative to their normal usage pattern, same-day adjustment can upwardly bias the baseline and overstate the demand reduction. When customers start reducing demand levels during the adjustment window, same-day adjustment can downwardly bias the

baseline and understate demand reduction. This latter example is illustrated in Figure 1-1. It is conceivable that the drop in load observed between 1 and 2 PM is in fact a demand reduction in anticipation of the event. Perhaps the customer needed to ramp down a production process gradually, thus initiating their load reduction early. However, the adjustment process treats such a behavioral change as estimation error and adjusts downward too much.

Same-day adjustments are often capped because adjustment can introduce the potential for manipulation of pre-event loads to bias baselines.[2] The concern is that participants may be able to "game" the system by increasing their electricity use during the adjustment period, leading to baselines that are too high and that overestimate actual demand reductions. Capping the magnitude of the adjustment limits the potential for this kind of abuse.

1.2 Current Baseline Rules

Currently, settlement baselines for most aggregators are based on the same-hour average of the 10 weekdays immediately preceding the curtailment event in question, excluding other event days and holidays.[3] The baseline is calculated separately for each customer and can be adjusted up or down by 20%, based on the first three of the four hours immediately before an event. The same-day adjustment and adjustment cap are applied individually for each customer. Then, each customer's baseline is summed to obtain the result for the aggregator settlement portfolio. Aggregators can elect whether or not same-day adjustments are applied, but must make their decisions in advance for each individual customer in their portfolio. For aggregators with utility contracts, the decision about whether or not to apply same-day adjustments is made at the start of each summer season. For aggregators participating in the Capacity Bidding Program, the decision about whether or not to apply same-day adjustments for specific customers is made a month in advance.

The current baseline rules were adopted by the California Public Utilities Commission in 2009 (D.09-08-027) and were implemented starting in 2010. In reaching its decision, the CPUC cited two studies that concluded that the 10-in-10 individual baseline with a same-day adjustment was highly accurate. The first study was the KEMA 2003 baseline study[4] and the second was the Quantum 2006 baseline study.'[5] However, both of those studies focused mainly on the set of days used to construct the baseline (e.g., top 10-in-10, top 15-in-20, etc.). Neither study analyzed the effect of the 20% cap on same-day adjustments. The KEMA 2003 study did not limit the magnitude of the baseline adjustment and the Quantum study allowed baselines to be adjusted upward by 100% and downward by 50%. In 2011, a study[6] focusing on baseline accuracy for customers with highly volatile loads found that a

[2] This report does not attempt to determine whether manipulation of baselines has or will occur and takes no position on the issue. i.

[3] The exception is one PG&E aggregator contract that calculates the baseline based on the highest 3 of the past 10 non-event weekdays, without same-day adjustments.

[4] KEMA, Inc. 2003. *Protocol Development for Demand Response Calculation—Findings and Recommendations.* Prepared for the California Energy Commission. http://www.energy.ca.gov/reports/2003-03-10_400-02-017F.PDF

[5] Quantum Consulting, Inc. 2004. *Working Group 2 Demand Response Program Evaluation – Program Year 2004*. Prepared for the Working Group 2 Measurement and Evaluation Committee. http://calmac.org/publications/2004-12-21_WG2_2004_REPORT.pdf

[6] Christensen Associates Energy Consulting. 2010. *Highly Volatile-Load Customer Study*. Prepared for Southern California Edison, Pacific Gas and Electric Company, and San Diego Gas & Electric.

20% cap on the adjustment for the 10-in-10 individual baseline was not large enough to accommodate some customers inaccurate and underestimated program performance and aggregator payments.

1.3 Baseline Variations Studied

In this study, we assessed the accuracy of load impacts for each event in 2011 using 32 different settlement baseline methods. These 32 variations are summarized in Table 1-1. The baseline rules vary in the magnitude of same-day adjustment caps, aggregate versus individual calculation, and whether or not same-day adjustments are applied universally.

Baselines allow the demand reduction calculation to be conducted quickly and easily. However, baseline rules have a direct impact on the settlement and payment process, because an aggregator's payments are directly linked to their ability to meet their demand reduction commitment when called upon. The magnitude of that demand reduction is measured by the baseline. If an aggregator meets their commitment in full, they receive payment in full. But if they fail to provide the full reduction, they receive lower payments. Thus, it is very important to fully understand the implications of choosing different baseline rules.

http://www3.sce.com/sscc/law/dis/dbattach3e.nsf/0/270D74C09E563CC2882577CE007D951A/$FILE/A.08-06-001+DR+App_Joint+Utilities+HVLC+Report+20101027+final.pdf.

Table 1-1: Baseline Calculation Methods Tested

No.	Days Used to Develop Baseline	Aggregation	Adjustment Option	Adjustment Cap		
				Name	Lower Bound	Upper Bound
1	**10-in-10** Baseline is based on the same-hour average of the 10 weekdays immediately preceding the curtailment event in question, excluding other event days and holidays.	**Individual** The baseline is calculated for each individual customer, and the in-day adjustment and cap are applied at the individual customer level. The individual customer results are then aggregated for each settlement portfolio.	**Aggregator Choice** The aggregator decides prior to events whether or not to apply same-day adjustments to the baseline.	Unadjusted	100%	100%
2				± 20%	80%	120%
3				± 30%	70%	130%
4				± 35%	65%	135%
5				± 40%	60%	140%
6				± 50%	50%	150%
7				2x[7]	50%	200%
8				Unlimited	-	-
9			**Universal** The same-day adjustment is applied to all customers and settlement portfolios.	Unadjusted	100%	100%
10				20%	80%	120%
11				30%	70%	130%
12				35%	65%	135%
13				40%	60%	140%
14				50%	50%	150%
15				2x	50%	200%
16				Unlimited	-	-
17		**Aggregate** Customer loads are first aggregated for each settlement portfolio. The baseline is then calculated using the aggregated loads and the same-day adjustment and caps are applied using the aggregated loads.	**Aggregator Choice** The aggregator decides prior to events whether or not to apply same-day adjustments to the baseline.	Unadjusted	100%	100%
18				20%	80%	120%
19				30%	70%	130%
20				35%	65%	135%
21				40%	60%	140%
22				50%	50%	150%
23				2x	50%	200%
24				Unlimited	-	-
25			**Universal** The same-day adjustment is applied to all customers and settlement portfolios.	Unadjusted	100%	100%
26				20%	80%	120%
27				30%	70%	130%
28				35%	65%	135%
29				40%	60%	140%
30				50%	50%	150%
31				2x	50%	200%
32				Unlimited	-	-

1.4 Report Focus and Contents

This report focuses on the accuracy of the demand reduction estimates calculated using various baseline methods rather than the accuracy of the baseline itself. The reason a baseline is calculated in the first place is to estimate demand reductions. Aggregators, customers and system operators are

[7] The inclusion of the "2x" cap recognizes that, in theory, "absolute" adjustments, such as "± 20%," skew adjustments downward. This is because baseline adjustments are calculated as ratios; thus, a cap of 50% limits ratio adjustments to 0.5 and 1.5, producing a downward adjustment of *half* the original value ($\frac{1}{1-.5} = \frac{1}{2}$) while producing an upward adjustment of only 50% ($1 * (1.5) = 1.5$). This intuition is confirmed by considering a cap of 100%; this allows downward adjustments to turn an unadjusted baseline to 0 while limiting upward adjustments to twice the original baseline. If applied symmetrically, a cap of 100% should yield unlimited upward adjustments: $\frac{1}{1-1} = \frac{1}{0} \approx \infty$.

concerned with how well a baseline estimates load reduction because this impacts system operations and how much customers are compensated. In other words, baselines are a means to an end, not the end in itself. Thus, we assess each baseline method on how accurately it estimates demand reductions, not on how well it estimates the load itself.[8]

In Section 2 of this volume, the results for each of the 32 baseline methods summarized in Table 1-1 are compared to the 2011 evaluation results presented in Volume I, which were estimated using regression methods. Note that this is a comparison between two estimates: one that uses more sophisticated regression methods and another that uses less sophisticated baseline methods. Comparing results from actual event days is useful because it provides insight into how baselines perform in the real world; for example, customers may shift load to pre-event hours, or may reduce demand early in anticipation of a load curtailment period.

Section 3 includes a systematic assessment of how accurately each baseline method estimates demand reductions under a set of idealized conditions. This is achieved by simulating a load reduction on an event-like day and then calculating baselines and impacts for this event-like day. In this way, one can assess how well the baseline accords with the unperturbed load, and how well the baseline impact accords with the true impact, because both the unperturbed load and the true impact are known. Note that there are two key assumptions embedded within this analysis: first, that customers provide a complete load reduction, and second, that there are no load increases or decreases in anticipation of the event. This allows an investigation of baseline accuracy without having to worry about confounding factors that can be difficult for a baseline to discern. Section 3 outlines the framework used to carry out this analysis and presents the analysis results for PG&E, SCE and SDG&E, respectively.

[8] In addition to different baseline rules, two other factors that improve the accuracy are greater aggregation and greater demand reductions. The effect of these two factors is discussed in Appendix A.

2 Comparison of EM&V and Baseline Results

This section compares the 2011 load impact evaluation estimates (documented in Volume 1) for each event day with the demand reductions estimated using the current settlement baseline rules and the alternative baseline rules outlined in Table 1-1. In each case, the baseline load impacts are presented as a percentage of the ex post (regression-based) evaluation results. This is done for each of the demand response programs evaluated in this report. This section is designed to satisfy requirements set forth by the California Public Utilities Commission in D.12-04-045.

Although comparing EM&V and baseline results is a comparison of two estimates – that is, neither one provides a perfect, true answer – this comparison is useful because it reflects customer behavior and weather patterns during actual event days. In particular, the regressions can detect whether a customer is changing their load patterns in anticipation of an event. In contrast, the baseline adjustment processes can confound early load reductions or load shifting with estimation error and improperly adjust for those behaviors.

Table 2-1 summarizes the comparison between the evaluation results and the demand reductions estimated using the current baseline method for all three utilities. It presents results by program and advance dispatch notification. In each instance, the settlement baselines were calculated based on the same-hour average of the 10 weekdays immediately preceding the curtailment event in question, excluding other event days and holidays. The baselines were calculated individually for each customer, the same-day adjustments (if selected) and cap were applied individually for each customer and then the results for each customer were summed to obtain the aggregate estimates.

Table 2-1 also includes three potential explanations for the difference between evaluation load impact estimates and the baseline estimates:

- The current baseline tends to over or under predict impacts when the true demand reductions are known (absent load shifting to adjustment period or early demand reductions) – that is, it exhibits bias.[9]
- Customers increase load in the pre-event hours used to calculate the adjustment, leading to upward adjustments that are too high and that over estimate demand reduction;
- Customers provide early demand reductions and decrease load in the pre-event hours used to calculate the adjustment, leading to downward adjustments that are too low and that under estimate the demand reduction; and

The results for program options that have few customers or were dispatched infrequently in 2011 should be interpreted with caution, since they typically have more uncertainty around the demand reduction estimates. For programs and resource options where more events are called and more customers are aggregated, patterns are more stable and provide more reliable information.

[9] To assess accuracy, it is necessary to know the correct demand reductions. We used a simulation based on actual load data and each customers demand reduction history precisely because the correct demand reductions amounts are known with a simulation. The process used to estimate the degree of bias (or lack thereof) in the baseline demand reduction calculations is detailed in Section 3.1 of this volume.

Table 2-1: Comparison of Evaluation Results and Demand Reductions Estimated Using the Current Baseline[10]

Utility	Program	Advance Notification	Average Event Comparison				Baseline Bias (simulation results)	Change in Customer Behavior During Pre-event Adjustment Period (Compared to Behavior on Non-event Days)		
			Event Hours	Evaluation Load Reduction (MW)	Baseline Load Reduction (MW)	% of Evaluation Results		Load Increase	Load Decrease	No Statistically Significant Change
PG&E	CBP	Day Ahead	15	13.6	15.2	112.0%	-22.3%	X		
		Day-of	3	14.2	12.5	87.7%	-9.7%			X
	AMP	Day Ahead	4	52.8	43.3	82.1%	-22.8%		X	
		Day-of	4	111.2	104.5	94.0%	-5.1%		X	
SCE	CBP	Day Ahead	48	4.0	3.2	80.0%	-15.0%			X
		Day-of	10	15.2	13.6	89.5%	-11.3%			X
	DRRC	Day Ahead	4	17.4	17.0	97.7%	-3.3%			X
		Day-of	8	81.4	81.8	100.5%	-2.8%			X
SDG&E	CBP	Day Ahead	19	11.0	8.6	78.2%	-12.5%			X
		Day-of	21	11.4	9.9	86.8%	-18.0%			X

[10] The "current baseline" is estimated based on the same-hour average of the 10 weekdays immediately preceding the curtailment event in question, excluding other event days and holidays. The baseline is calculated individually for each customer, the same-day adjustments and cap (±20%) is applied individually for each customer, and then the results for each customer are summed to obtain the result for the aggregator settlement portfolio. The results reflect aggregator decisions about whether or not in-day adjustments are applied, which are made in advance of events. This baseline is used for all programs and aggregators, with the exception of one PG&E aggregator under contract. However, the table applies the same baseline to allow for a direct comparison.

In eight out of ten cases, the current baseline *underestimates* the reduction measured by the regressions; in one case, PG&E's CBP day-ahead option, the baseline *overestimates* the reduction. Unless the customer changes their behavior during the adjustment window, under- or over-estimation of the demand reduction measured by the baseline is due to upward or downward bias in the baseline. In seven out of ten cases, underestimates in the baseline load reductions are explained solely by downward bias in the baseline, *not* by changes in behavior during the adjustment window.

This is not the case for customers in PG&E's CBP day-ahead program; they receive notice a day in advance and consistently shift load to the pre-event hours, including those used to adjust baselines, compounding the effects of baseline bias. As noted earlier, the baseline adjustment mechanism treats changes in electricity use during pre-event hours leading to the event as error and adjusts for it, leading to incorrect *upward* adjustments when customers shift loads to those hours. Moreover, in the case of PG&E's AMP program, customers reduce their load in pre-event hours, including those used to adjust baselines, leading to incorrect *downward* adjustments.

2.1 PG&E Event Day Comparison

This section compares the evaluation results to the estimates produced by applying the 32 baseline methods shown in Table 1-1. It also discusses the impact of adjusting baselines at the individual level and at the settlement portfolio level on the magnitude of in-day adjustments.

2.1.1 Detailed Comparison of Evaluation and Baseline Results

Table 2-2 shows the estimated demand reductions for each baseline method for PG&E's AMP and CBP programs. The day-ahead and day-of resource options are presented separately. The first row in the table shows the impacts estimated using regression methods. In addition, estimates using the current baseline method are shown in bold. The main finding is simple: for AMP-DO, CBP-DO and CBP-DA, changing the baseline rules does not help align the settlement demand reductions with the evaluation results. For AMP-DA, the baseline results better align with the evaluation results when same-day adjustments are applied to all customers (the "universal" adjustment option baselines).

For the AMP-DO resource, the current baseline produces load impact estimates that are 94.0% of the evaluation results – that is, the baseline results are lower by 6.0%. Some of this difference can be explained by bias caused by baseline methods; the remaining difference between the evaluation and baseline results is explained primarily by customer behavior during the hours used to calculate the adjustment. Specifically, some customers start reducing load early, leading to downward adjustments that are too large and demand reductions that are too low. All the baseline rules tested produce similar results for PG&E's AMP program contracts with day-of notification, as long as some type of day adjustment is applied.

For both the AMP-DO and AMP-DA resources, relaxing the baseline adjustment cap does not lead to improvements relative the evaluation results, as one would expect. While we do not see an improvement during actual event days, we do see an improvement when we measure the accuracy using a simulation. This suggests that we don't see improvements because customers are shifting load to pre-event hours or reducing load in anticipation of the event during actual events. Such behavior can affect baseline adjustments and counter the improvements expected from using same-day adjustments.

Table 2-2: Comparison of Evaluation Results and Demand Reductions Estimated Using the Alternate Baseline Rules

Application of Baselines	Adjustment Option	Adjustment Cap	Aggregator Managed Portfolio				Capacity Bidding Program			
			Day-of		Day-ahead		Day-of		Day-ahead	
			Estimated Impacts	% of M&E Results	Estimated Impacts	% of M&E Results	Estimated Impacts	% of M&E Results	Estimated Impacts	% of M&E Results
M&E Result			111.2	-	52.8	-	14.2	-	13.6	-
Individual	Aggregator Choice	Unadjusted	105.0	94.4%	43.7	82.8%	10.2	71.7%	12.8	94.4%
		20% [1]	**104.5**	**94.0%**	**43.3**	**82.1%**	**12.5**	**87.7%**	**15.2**	**112.0%**
		30%	103.7	93.3%	43.3	82.1%	12.6	88.5%	15.4	113.2%
		35%	103.3	92.9%	43.3	82.1%	12.6	88.7%	15.4	113.3%
		40%	102.9	92.6%	43.3	82.1%	12.6	88.8%	15.4	113.3%
		50%	102.7	92.3%	43.3	82.1%	12.6	88.8%	15.4	113.2%
		2x	103.5	93.1%	43.3	82.1%	12.7	89.2%	15.8	116.5%
		Unlimited	103.5	93.1%	43.3	82.1%	12.9	90.3%	15.8	116.0%
	Universal	Unadjusted	105.0	94.4%	43.7	82.8%	10.2	71.7%	12.8	94.4%
		20%	104.1	93.6%	45.3	85.8%	12.5	87.8%	15.2	111.8%
		30%	103.3	92.9%	45.3	85.9%	12.6	88.6%	15.4	112.9%
		35%	102.9	92.6%	45.3	85.7%	12.6	88.8%	15.4	112.9%
		40%	102.5	92.2%	45.2	85.7%	12.7	88.9%	15.4	112.9%
		50%	102.3	92.0%	45.5	86.2%	12.7	88.9%	15.3	112.8%
		2x	103.2	92.8%	46.6	88.2%	12.7	89.3%	15.8	116.1%
		Unlimited	103.3	92.9%	46.7	88.6%	12.9	90.4%	15.7	115.4%
Aggregate	Aggregator Choice	Unadjusted	104.0	93.5%	43.7	82.8%	10.7	75.2%	13.7	100.7%
		20%	104.1	93.6%	43.4	82.2%	12.9	90.7%	16.0	117.4%
		30%	104.1	93.6%	43.4	82.2%	12.9	90.7%	15.9	116.6%
		35%	104.1	93.6%	43.4	82.2%	12.9	90.7%	15.9	116.6%
		40%	104.1	93.6%	43.4	82.2%	12.9	90.7%	15.9	116.7%
		50%	104.1	93.6%	43.4	82.2%	12.9	90.7%	15.9	116.7%
		2x	104.1	93.6%	43.4	82.2%	12.9	90.7%	15.9	116.7%
		Unlimited	104.1	93.6%	43.4	82.2%	12.9	90.7%	15.9	116.7%
	Universal	Unadjusted	104.0	93.5%	43.7	82.8%	10.7	75.2%	13.7	100.7%
		20%	103.9	93.5%	46.6	88.3%	12.9	90.8%	15.9	117.2%
		30%	103.9	93.5%	46.6	88.3%	12.9	90.8%	15.8	116.3%
		35%	103.9	93.5%	46.6	88.3%	12.9	90.8%	15.8	116.3%
		40%	103.9	93.5%	46.6	88.3%	12.9	90.8%	15.8	116.3%
		50%	103.9	93.5%	46.6	88.3%	12.9	90.8%	15.8	116.2%
		2x	103.9	93.5%	46.6	88.3%	12.9	90.8%	15.8	116.2%
		Unlimited	103.9	93.5%	46.6	88.3%	12.9	90.8%	15.8	116.1%

For the AMP-DA resource, the current baseline produces load impact estimates that are 82.1% of the evaluation results – that is, the baseline results are lower by 17.9%. The baseline results better align with the evaluation results when same-day adjustments are universally applied to all customers. Put differently, the difference between the baseline and evaluation results is in part due to aggregator decisions about whether or not to apply same-day adjustments. These findings are consistent with the simulation analysis of baseline accuracy discussed in Section 3, where the biggest improvement is from applying same-day adjustments universally for AMP-DA.

For the CBP day-of notification option, the current baseline produces load impact estimates that are 87.7% of the evaluation results – that is, the baseline results are lower by 12.3%. When assuming that customers do not change their behavior during the adjustment window, the current baseline produces estimates that are, on average, 9.7% too low.[11] Changing the baseline method does not help align the settlement demand reductions with the evaluation results. Neither increasing limits on the adjustment cap, universal application of same-day adjustments, nor calculating baselines after aggregating loads helps remove the downward bias completely.

For CBP with day-ahead notification, the current baseline method demand reduction estimates that are 112% of the evaluation results – that is, the baseline estimates are 12% higher than the regression-based estimates. At first glance the results appear counterintuitive because the current baseline produces estimates that are, on average, 22.3% too low under idealized conditions – that is, assuming no shifting to pre-event hours or early demand reductions.[12] However, a closer inspection of the evaluation results indicates that these customers shift load to the pre-event hours in response to the event, increasing use during the period used to calculate the baseline adjustments. This leads to baseline adjustments that are too high and that over estimate demand reductions.

2.1.1 Magnitude of Same-day Adjustments

The idea behind the same-day adjustment is that observed differences between the baseline and actual loads in the hours leading up to an event can help correct prediction errors in the unadjusted baseline. As noted earlier, the adjustment process implicitly assumes that electricity use during the adjustment period is unperturbed and any differences between the unadjusted baseline and actual loads are solely due to prediction error. The adjustments are typically capped, to help limit the potential for gaming. Large adjustments are required when electricity use in prior days are a poor indicator of electricity use during actual event days. In this section, we assess the impact of same-day adjustments based on individual versus aggregate baseline calculation rules and calculate how often different baseline caps are exceeded.

Under the individual approach, baselines are calculated for each individual customer and the same-day adjustment (and cap) is applied at the individual customer level. The individual customer results are then aggregated for each settlement portfolio. This is how baselines are currently calculated.

Under the aggregated approach, customer loads are first aggregated for each settlement portfolio. The baseline is then calculated using the aggregated loads and the same-day adjustment (and caps) is

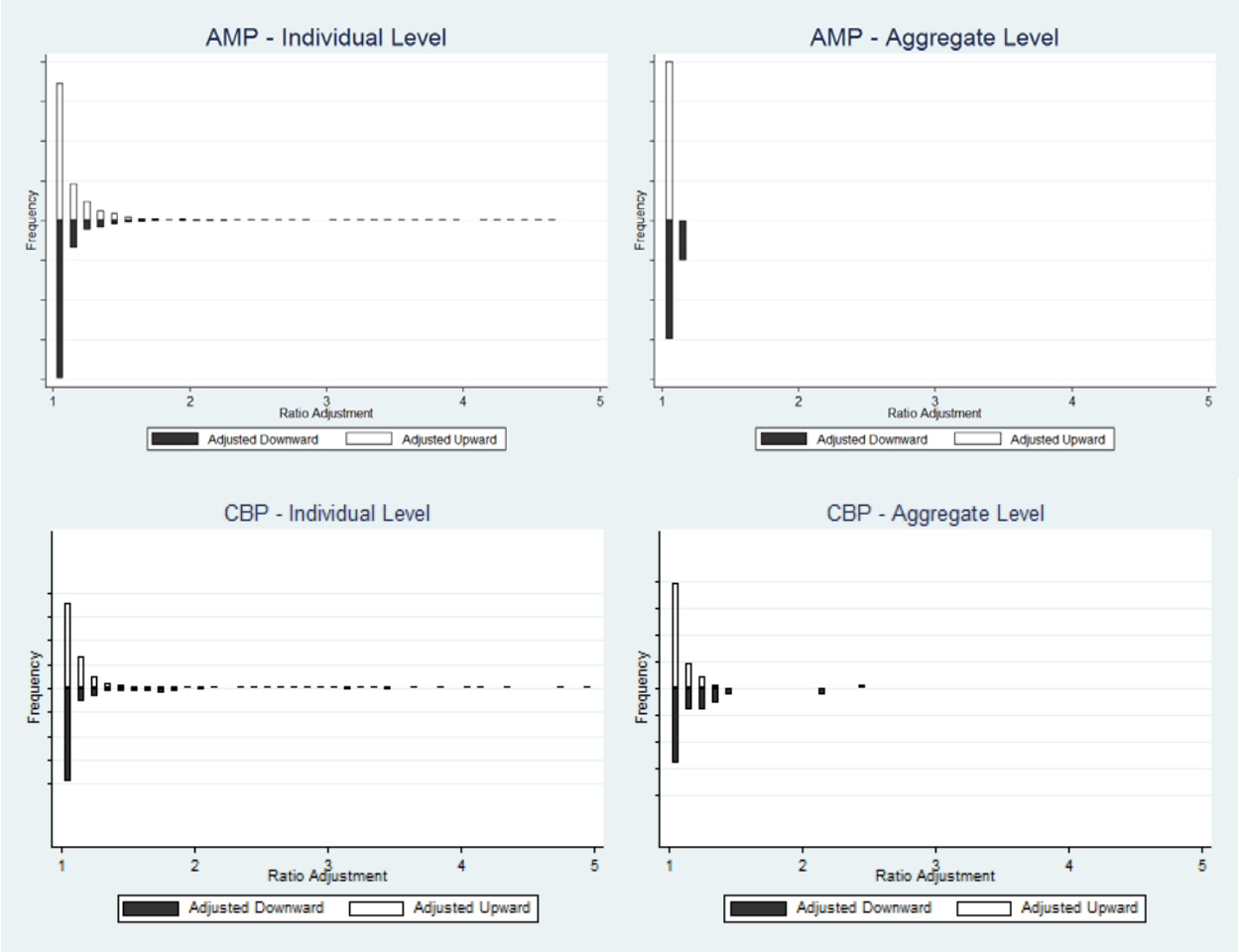
[11] See Section 3 for details on this result.

[12] Ibid.

applied using the aggregated values. The logic for using the aggregated approach is that individual customer idiosyncrasies and volatility are smoothed out in the aggregation process, leading to smaller same-day adjustments. Under the aggregated approach, the magnitude of the same-day adjustment depends on the number of customers in the aggregator's portfolio. A settlement portfolio that has one site is more likely to require large adjustments than a settlement portfolio with 100 sites.

Figure 2-1 combines two sets histograms of ratio adjustments: the first set shows adjustments calculated for individual customers for AMP and CBP, while the second shows adjustments calculated when the baselines are estimated using aggregated data for each settlement portfolio. Recall that the ratio adjustment is calculated by taking the measured load during the adjustment window and dividing it by the baseline during the adjustment window. If the resulting ratio is greater than 1, the baseline is adjusted upward; if it is less than 1, it is adjusted downward. In Figure 2-1, the x-axis shows the magnitude of such ratio adjustments; as one moves to the right side of the axis, adjustments become more extreme. (Note that downward adjustments have been inverted so they may be more easily compared; that is, a downward adjustment of 0.5 is shown as "2.") The y-axis denotes the relative frequency of each ratio adjustment.

Figure 2-1: PG&E Distribution and Magnitude of Same-day Ratio Adjustments (Uncapped)



The figure highlights several key findings. First, adjustments occur in both directions. It is not the case that adjustments are upward because they are needed mostly for weather sensitive loads. The same-day adjustments calibrate the unadjusted baselines to avoid over and under predictions for both weather insensitive and weather sensitive customers. Second, the adjustments tend to be smaller when loads are aggregated prior to the baseline calculation. Third, relatively few settlement portfolios and customers require more than a twofold adjustment.

Table 2-3 shows the share of observations by program and dispatch option exceeding different same-day adjustment caps. The table reinforces the point that large adjustments are rarely required. It also shows that adjustment caps are exceeded less often when baselines are calculated after aggregating loads for each settlement portfolio, particularly when settlement portfolios have a large number of customers such as those in the aggregator contracts.

Table 2-3: Share of Observations Exceeding Same-day Adjustment Thresholds

Program	Adjustment Level	Adjustment Direction	Adjustment Cap			
			± 20%	± 30%	± 40%	± 50%
Aggregator Contracts	Aggregate	Upward	0.0%	0.0%	0.0%	0.0%
		Downward	0.0%	0.0%	0.0%	0.0%
		TOTAL	0.0%	0.0%	0.0%	0.0%
	Individual	Upward	12.6%	7.9%	5.7%	4.0%
		Downward	7.6%	5.0%	3.8%	3.2%
		TOTAL	20.2%	12.9%	9.4%	7.2%
CBP	Aggregate	Upward	7.9%	3.1%	1.6%	1.6%
		Downward	15.7%	3.9%	3.9%	3.9%
		TOTAL	23.6%	7.1%	5.5%	5.5%
	Individual	Upward	16.5%	11.4%	9.2%	7.3%
		Downward	19.3%	18.3%	17.3%	16.3%
		TOTAL	35.8%	29.7%	26.6%	23.7%

2.2 SCE Event Day Comparison

This section compares the evaluation results to the estimates produced by applying the 32 methods summarized in Table 1-1. It also discusses the impact of adjusting baselines at the individual level and at the settlement portfolio level on the magnitude of in-day adjustments.

2.2.1 Detailed Comparison of Evaluation and Baseline Results

Table 2-4 shows the demand reductions estimated for each baseline method for SCE's DRCC and CBP programs. The first row in the table shows the evaluation results and the row in bold shows the current baseline results.

For the DRRC-DO and DRRC-DA resources, the baseline and regression results are almost equivalent. Any gap between the evaluation and baseline results is minimal. Both day-of and day-ahead customers receive nearly 100% of the EM&V-measured reduction using the current baseline methods. Thus, using an alternate baseline calculation method would not help these customers.

For the CBP-DO resource, the current baseline produces demand reduction estimates that are 89.3% of the evaluation results – that is, the baseline results are lower by 10.7%. The gap closely matches the downward baseline bias of 11.3% calculated for the current baseline under idealized conditions. The baseline estimates better align with the evaluation results when the same-day adjustments are universally applied, although this leads to overestimation of demand reductions by roughly 5%. Accuracy results for universally adjusted baselines presented in Section 3 show that they also overestimate demand reductions for event-like days under laboratory-like conditions.

The gap between the evaluation and current baseline results is largest for SCE's CBP-DA resource. The current baseline produces estimates that are, on average, 20% lower than the evaluation results.

These customers would benefit from being required to apply same-day adjustment. If aggregators had elected to apply the same-day adjustment, the baseline would have better mirrored the EM&V results: applying the same-day adjustment universally leads to demand reduction estimates that are 12% lower than the evaluation results, as opposed to 20% lower with the optional adjustment. These customers would also benefit from loosening the adjustment cap.

Table 2-4: Comparison of Evaluation Results and Demand Reductions Estimated Using the Alternate Baseline Rules

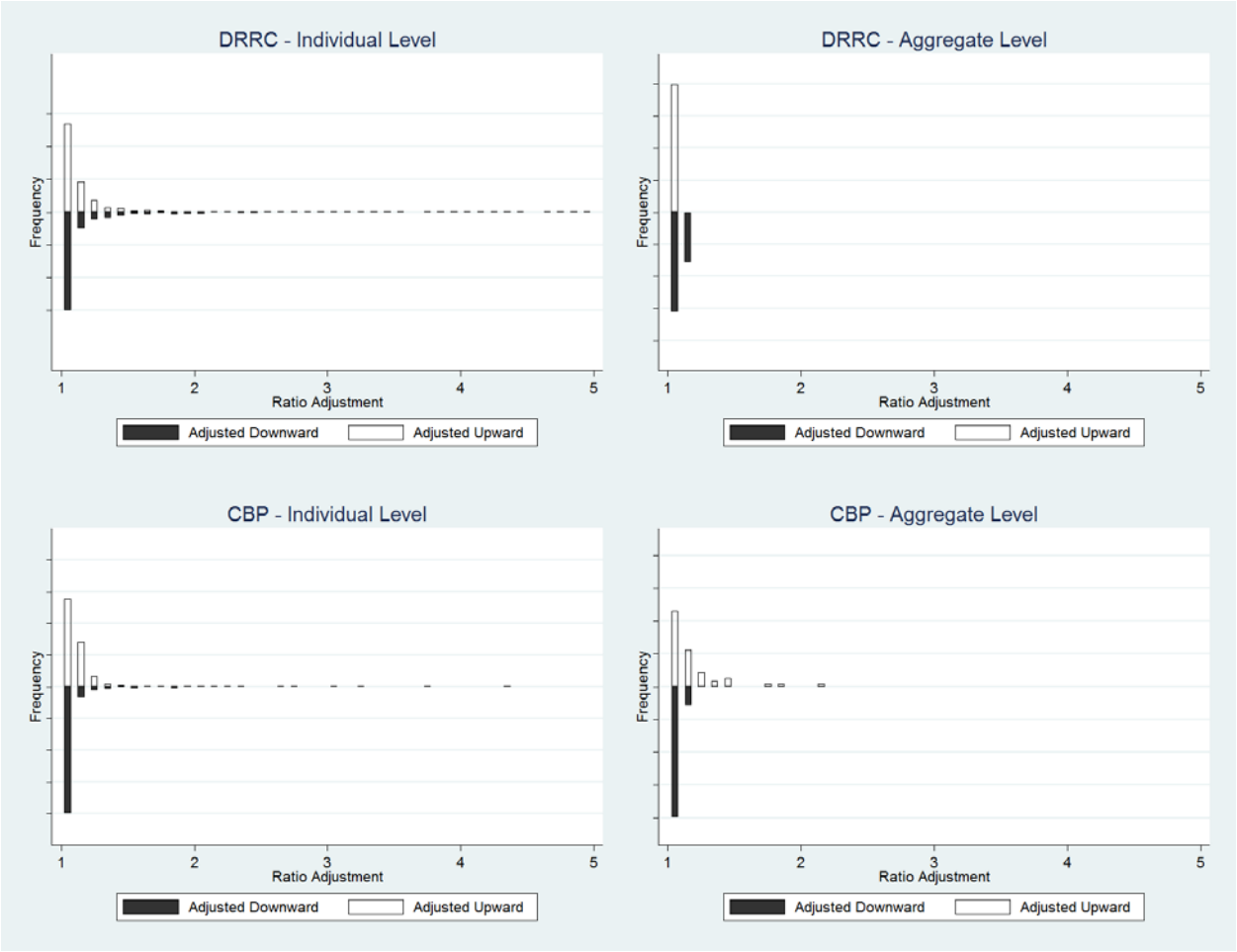
Application of Baselines	Adjustment Option	Adjustment Cap	Demand Response Resource Contracts				Capacity Bidding Program			
			Day-of		Day-ahead		Day-of		Day-ahead	
			Estimated Impacts	% of M&E Results	Estimated Impacts	% of M&E Results	Estimated Impacts	% of M&E Results	Estimated Impacts	% of M&E Results
M&E Result			81.4	-	17.4	-	15.2	-	4.0	-
Individual	Aggregator Choice	Unadjusted	72.7	89.3%	16.4	94.2%	11.1	72.8%	2.1	52.2%
		20% [1]	**81.8**	**100.5%**	**17.0**	**97.9%**	**13.6**	**89.3%**	**3.2**	**79.9%**
		30%	82.3	101.0%	16.8	96.7%	13.6	89.6%	3.2	81.7%
		35%	82.3	101.0%	16.7	96.4%	13.6	89.6%	3.3	82.0%
		40%	82.2	100.9%	16.7	96.2%	13.6	89.7%	3.3	82.4%
		50%	82.1	100.8%	16.6	95.7%	13.7	89.8%	3.3	82.9%
		2x	83.2	102.2%	16.9	97.2%	13.7	89.8%	3.4	84.4%
		Unlimited	84.3	103.5%	16.4	94.6%	13.7	89.8%	3.4	84.6%
	Universal	Unadjusted	72.7	89.3%	16.4	94.2%	11.1	72.8%	2.1	52.2%
		20%	81.8	100.5%	17.0	97.9%	15.8	104.0%	3.4	86.8%
		30%	82.3	101.0%	16.8	96.7%	16.0	105.0%	3.6	91.8%
		35%	82.3	101.0%	16.7	96.4%	16.0	105.1%	3.7	93.3%
		40%	82.2	100.9%	16.7	96.2%	16.0	105.2%	3.8	94.5%
		50%	82.1	100.8%	16.6	95.7%	16.0	105.3%	3.8	96.0%
		2x	83.2	102.2%	16.9	97.2%	16.0	105.3%	4.0	100.6%
		Unlimited	84.3	103.5%	16.4	94.6%	16.0	105.3%	4.0	101.5%
Aggregate	Aggregator Choice	Unadjusted	72.7	89.3%	16.4	94.2%	11.1	72.8%	2.1	52.3%
		20%	81.9	100.6%	16.3	93.8%	13.6	89.3%	3.2	81.6%
		30%	81.9	100.6%	16.3	93.8%	13.6	89.3%	3.3	82.0%
		35%	81.9	100.6%	16.3	93.8%	13.6	89.3%	3.3	82.0%
		40%	81.9	100.6%	16.3	93.8%	13.6	89.3%	3.3	82.1%
		50%	81.9	100.6%	16.3	93.8%	13.6	89.3%	3.3	82.2%
		2x	81.9	100.6%	16.3	93.8%	13.6	89.3%	3.3	82.2%
		Unlimited	81.9	100.6%	16.3	93.8%	13.6	89.3%	3.3	82.2%
	Universal	Unadjusted	72.7	89.3%	16.4	94.2%	11.1	72.8%	2.1	52.3%
		20%	81.9	100.6%	16.3	93.8%	15.9	104.7%	3.5	88.6%
		30%	81.9	100.6%	16.3	93.8%	15.9	104.7%	3.7	92.1%
		35%	81.9	100.6%	16.3	93.8%	15.9	104.7%	3.7	93.3%
		40%	81.9	100.6%	16.3	93.8%	15.9	104.7%	3.7	94.3%
		50%	81.9	100.6%	16.3	93.8%	15.9	104.7%	3.8	95.3%
		2x	81.9	100.6%	16.3	93.8%	15.9	104.7%	3.9	98.4%
		Unlimited	81.9	100.6%	16.3	93.8%	15.9	104.7%	3.9	99.0%

[1] Current baseline

2.2.2 Magnitude of Same-day Adjustments

Figure 2-2 combines two sets histograms of ratio adjustments: the first set shows adjustments calculated for individual customers for DRRC and CBP, while the second shows adjustments calculated for aggregated portfolios for DRRC and CBP. Recall that the ratio adjustment is calculated by taking the measured load during the adjustment window and dividing it by the baseline during the adjustment window. If the resulting ratio is greater than 1, the baseline is adjusted upward; if it is less than 1, it is adjusted downward. In Figure 2-2, the x-axis shows the magnitude of such ratio adjustments; as one moves to the right side of the axis, adjustments become more extreme. (Note that downward adjustments have been inverted so they may be more easily compared; that is, a downward adjustment of 0.5 is shown as "2.") The y-axis denotes the relative frequency of each ratio adjustment.

Figure 2-2: SCE Distribution and Magnitude of Same-day Ratio Adjustments (Uncapped)



Key findings are similar to those for PG&E. First, very few customers need extreme adjustments. This helps to explain the finding that the unlimited baseline rarely does much to improve accuracy; most customers simply don't need very large adjustments, and thus fall below successively higher adjustment caps. Second, note that aggregated baselines need much lower adjustments than individually calculated baselines. Finally, adjustments occur in both directions; baselines are just as

likely to be adjusted upward, perhaps because they are weather sensitive (use more electricity on warmer days), as they are to be adjusted downward.

Table 2-5 shows the share of observations by program and dispatch option exceeding different same-day adjustment caps. The table reinforces the point that large adjustments are rarely required. It also shows that adjustments caps are exceeded less often when baselines are calculated after aggregating loads for each settlement portfolio, particularly when settlement portfolios have a large number of customers.

Table 2-5: Share of Observations Exceeding Same-day Adjustment Thresholds

Program	Adjustment Level	Adjustment Direction	Adjustment Cap			
			± 20%	± 30%	± 40%	± 50%
Aggregator Contracts	Aggregate	Upward	19.8%	14.4%	9.0%	5.4%
		Downward	0.0%	0.0%	0.0%	0.0%
		TOTAL	19.8%	14.4%	9.0%	5.4%
	Individual	Upward	10.0%	3.7%	2.2%	1.6%
		Downward	1.1%	0.7%	0.6%	0.4%
		TOTAL	11.1%	4.4%	2.8%	2.0%
CBP	Aggregate	Upward	0.0%	0.0%	0.0%	0.0%
		Downward	0.0%	0.0%	0.0%	0.0%
		TOTAL	0.0%	0.0%	0.0%	0.0%
	Individual	Upward	14.2%	8.9%	6.9%	5.2%
		Downward	10.7%	9.0%	7.8%	6.7%
		TOTAL	24.9%	17.9%	14.8%	11.9%

2.3 SDG&E Event Day Comparison

This section compares the evaluation results to the estimates produced by applying the 32 baseline types shown in Table 1-1. It also discusses the impact of adjusting baselines at the individual level and at the settlement portfolio level on the magnitude of in-day adjustments.

2.3.1 Detailed Comparison of Evaluation and Baseline Results

Table 2-6 shows the demand reductions calculated by each set of baseline rules for SDG&E's CBP program. As with the tables presented for PG&E and SCE, the first row in the table shows impacts calculated by the regression, while the bolded row shows the impacts calculated by the current baseline.

Table 2-6: Comparison of Evaluation Results and Demand Reductions Estimated Using the Alternate Baseline Rules

Application of Baselines	Adjustment Option	Adjustment Cap	Capacity Bidding Program			
			Day-of		Day-ahead	
			Estimated Impacts	% of M&E Results	Estimated Impacts	% of M&E Results
M&E Result			11.4	-	11.0	-
Individual	Aggregator Choice	Unadjusted	7.2	63.1%	7.4	67.0%
		20% [1]	**9.9**	**86.3%**	**8.6**	**77.9%**
		30%	10.1	88.1%	8.9	80.3%
		35%	10.1	88.5%	8.9	81.0%
		40%	10.1	88.8%	9.0	81.3%
		50%	10.2	89.1%	9.0	81.7%
		2x	10.2	89.4%	9.1	82.4%
		Unlimited	10.2	89.1%	9.1	82.5%
	Universal	Unadjusted	7.2	63.1%	7.4	67.0%
		20%	10.8	94.8%	8.4	75.9%
		30%	11.1	97.4%	8.7	79.0%
		35%	11.2	97.9%	8.9	80.9%
		40%	11.2	98.2%	9.1	82.2%
		50%	11.2	98.4%	9.1	82.7%
		2x	11.3	98.7%	9.4	85.3%
		Unlimited	11.2	98.0%	10.5	95.6%
Aggregate	Aggregator Choice	Unadjusted	6.6	57.6%	7.4	67.0%
		20%	9.4	82.2%	8.9	81.0%
		30%	9.5	82.8%	9.1	82.2%
		35%	9.5	82.8%	9.1	82.2%
		40%	9.5	82.8%	9.1	82.2%
		50%	9.5	82.8%	9.1	82.2%
		2x	9.5	82.8%	9.1	82.2%
		Unlimited	9.5	82.8%	9.1	82.2%
	Universal	Unadjusted	6.6	57.6%	7.4	67.0%
		20%	10.3	90.4%	9.8	88.5%
		30%	10.4	91.4%	9.9	90.1%
		35%	10.4	91.4%	9.9	90.0%
		40%	10.4	91.4%	9.9	90.0%
		50%	10.4	91.4%	9.9	89.8%
		2x	10.4	91.4%	9.9	90.1%
		Unlimited	10.4	91.4%	9.9	89.9%

[1] Current baseline

For the CBP-DO resource, the current baseline underestimates demand reductions by 14% when compared to the EM&V results. This result can be improved by loosening the cap and applying same-

day adjustments universally (at the individual level). Adjusting the baseline at the settlement portfolio level is not as beneficial for these customers, although the universally adjusted baselines still present a significant improvement over the current alternative.

For the CBP-DA resource, the current baseline *underestimates* load impacts by 22% when compared to the EM&V results. While any adjusted baseline presents an improvement over the current alternative, applying the adjustment at the settlement portfolio level gives the greatest improvement. However, note that baselines for SDG&E's CBP-DA resource produce demand reduction estimates with a wide margin of error, in part because the resource includes a fairly small number of large customers with relatively volatile loads.

2.3.2 Magnitude of Same-day Adjustments

Figure 2-3 combines two histograms of ratio adjustments: the first set shows adjustments calculated for individual CBP customers, while the second shows adjustments calculated when the baselines are estimated using aggregated data for each settlement portfolio. Recall that the ratio adjustment is calculated by taking the measured load during the adjustment window and dividing it by the baseline during the adjustment window. If the resulting ratio is greater than 1, the baseline is adjusted upward; if it is less than 1, it is adjusted downward. In Figure 2-3, the x-axis shows the magnitude of such ratio adjustments; as one moves to the right side of the axis, adjustments become more extreme. (Note that downward adjustments have been inverted so they may be more easily compared; that is, a downward adjustment of 0.5 is shown as "2.") The y-axis denotes the relative frequency of each ratio adjustment.

Very few customers in any program or adjustment level need extreme adjustments. This helps to explain the finding that the unlimited baseline rarely does much to improve accuracy; most customers simply don't need very large adjustments, and thus fall below successively higher adjustment caps. Second, note that aggregated baselines need much lower adjustments than individually calculated baselines. Finally, adjustments occur in both directions; baselines are just as likely to be adjusted upward as they are to be adjusted downward.

Figure 2-3: SDG&E Distribution and Magnitude of Same-day Ratio Adjustments (Uncapped)

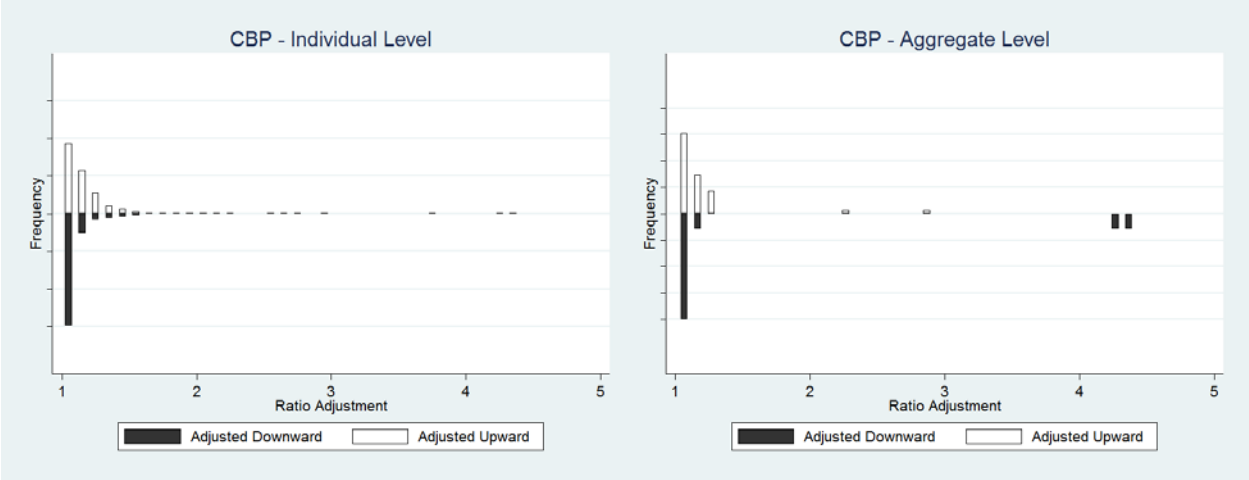


Table 2-7 shows the share of observations by program and dispatch option exceeding different same-day adjustment caps. The table reinforces the point that large adjustments are rarely required. It also shows that adjustments caps are exceeded less often when the baselines are calculated after aggregating loads for each settlement portfolio, particularly when settlement portfolios have a large number of customers such as those in the aggregator contracts.

Table 2-7: Share of Observations Exceeding Same-day Adjustment Thresholds

Program	Adjustment Level	Adjustment Direction	Adjustment Cap			
			± 20%	± 30%	± 40%	± 50%
CBP	Aggregate	Upward	22.8%	3.5%	3.5%	3.5%
		Downward	3.5%	3.5%	3.5%	3.5%
		TOTAL	26.3%	7.0%	7.0%	7.0%
	Individual	Upward	20.3%	9.1%	4.8%	2.1%
		Downward	2.2%	1.2%	0.7%	0.6%
		TOTAL	22.5%	10.3%	5.5%	2.7%

3 Impact Simulation Analysis

Section 2 summarized the assessment of baseline accuracy by comparing impact estimates for each baseline option with regression based estimates (documented in Volume 1) for each event day in 2011. In this section, baseline accuracy is assessed by simulating a load reduction on an event-like day and observing how well the baseline calculation performs for this hypothetical event. This method makes it possible to assess how well the baseline accords with the unperturbed load, and how well the baseline impact accords with the true (simulated) impact, because both the unperturbed load and the true impact are known. Furthermore, it allows for an investigation of the bias and goodness of fit for the load impact estimates for each baseline option without having to worry about confounding factors that can be difficult for a baseline to discern, like load shifting to pre-event hours, reducing load in anticipation of the event, or other ways in which aggregators try to manage a particular baseline. In other words, the analysis presented in this section allows the reader to make a clear judgment about each baseline method's performance under idealized conditions.

The first subsection below describes impact simulation analysis applied here. Detailed assessments of each baseline option for each utility, program and resource option are presented in the remainder of the section.

3.1 Framework for Assessing Accuracy

To implement the assessment framework described above, we:

1. *Selected proxy event days:* The proxy days were selected to match, as closely as possible, actual event conditions observed in 2011 for each of the three utilities.
2. *Identified realistic demand reductions for each individual customer*: For both the CBP program and the aggregator contracts, commitments are made at the aggregate level, not at the individual level. Because aggregators do not produce estimates of the expected load reductions for each customer, we used customer specific values produced during the M&E evaluation. The reductions are a realistic approximation of individual customers' load reduction. By design, the reductions varied across customers, but were the same across event days for each customer.
3. *Applied the demand reductions to unperturbed loads during proxy event days:* For each of the proxy curtailment event periods, the demand reductions were subtracted from the unperturbed loads. The same individual customer-specific demand reduction is applied to every proxy event day. Because the demand reductions were simulated, the true counterfactual and demand reductions are known, allowing us to test how close the baseline methods are to correctly estimating the demand reductions.
4. *Calculated the baselines and load impacts using each of 32 baseline methods:* As described in Table 1-1, the baselines were calculated individually and in aggregate, assuming universal and elective application of same-day adjustments, and with different caps on same-day adjustments. Impacts were calculated as the difference between the baseline and the loads with the simulated demand reduction.
5. *Assessed the accuracy of each of the settlement alternatives:* To standardize the comparison, we used metrics designed to assess if each baseline method systematically over or under-estimates demand reductions (bias) and metrics that summarize how close the estimates are to the true (simulated) demand reductions (goodness-of-fit).

How and Why Do Baseline and Demand Reduction Accuracy Differ?

How well a baseline method estimates demand reductions is closely related to how well it estimates customer loads. However, these two concepts differ in subtle ways. A good baseline explains much of the variation in electricity use, reducing the noise inherent in daily patterns of electricity use, thereby allowing for better detection of the signal of interest – the demand reduction. Nevertheless, aggregator and system operators are concerned with how accurately demand reductions are estimated; they are less concerned with how accurately the baseline itself is estimated.

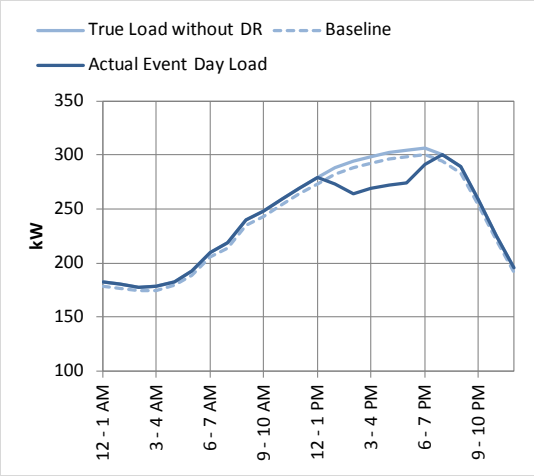
Aggregator and customer behavior plays a role in how well baselines estimate demand reductions. An aggregator can help improve baseline estimates by aggregating across more customers. By including more customers and different types of loads into settlement accounts, aggregated loads and demand reductions become more predictable and less volatile. The magnitude of the demand reduction also plays a key role in determining impact accuracy. *The larger the load impact, the better the impact accuracy.* The more demand reduction that aggregators deliver, the better that baselines estimate them. A baseline can perform equally well at predicting loads for two customers, but perform differently in estimating demand reductions if they reduce a different share of their load.

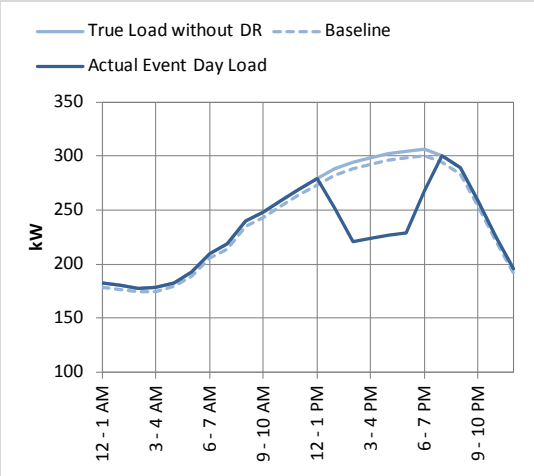
To illustrate, consider an example where the baseline method estimates actual loads equally well for two very similar customers – Store A and Store B. Assume that we know what the customers would have used had they not curtailed loads, and that we also know the true demand reductions they delivered.

For both customers, the baseline is too low. It underestimates what the customers *would have used* if an event hadn't occurred by 2%. During the event, Store A reduces their demand by 10%, or 30 kW, from 300 kW to 270 kW. Because the baseline is downwardly biased by 2%, however, they are credited with 24 kW (294 kW -270 kW), or 80% of their true demand reduction of 30kW – underestimating the true reduction by 20%.

Store B's baseline has the same bias, -2%. However, they reduce their demand by a greater amount than Store A: they reduce by 25%, or 75 kW, reducing load from 300 kW to 225 kW. Because of the downward bias in the baseline, they are only credited for a drop of 69 kW (294 kW – 225 kW) instead of 75 kW. While the magnitude of the baseline error is the same in absolute and percentage terms for both customers – 6 kW or 2% – Customer B is credited for 92% of their true demand reduction, while Customer A is only credited for 80%. Thus, by aggregating and delivering larger percent reductions (or avoiding customers that don't deliver much), aggregators can help minimize the chance of inaccurate payments.

Appendix A summarizes the empirical data on how the amount of aggregation and magnitude of the demand reductions affect the accuracy of baseline estimates.





The remainder of this subsection provides more details regarding each step. In the the remainder of the report, we focus on the accuracy of the demand reductions calculated using baselines (as opposed to baseline accuracy). The reason a baseline is calculated in the first place is to estimate load reductions. Aggregators and customers are not concerned with how well a baseline estimates their load in the absence of a load reduction, *per se*; they are concerned about how well a baseline estimates their load reduction, because this impacts how much they are compensated. In other words, baselines are means to an end, not the end in itself. Thus, we assess each set of baseline rules based on how accurately it estimates demand reductions, not just by how well it estimates the load itself.

3.1.1 Selection of Proxy Event Days and Hours

The proxy days were selected to match actual event conditions observed in 2011 as closely as possible based on system load and weather conditions. This is a critical step in the method because the proxy events are used to assess how accurately the baselines calculate demand reductions under event-like conditions. Tables 3-1, 3-2 and 3-3 compare the actual and proxy event day conditions based on annual system load ranking and average daily temperature, which reflects overall heat intensity better than the daily maximum temperature. To facilitate the comparison, the actual and proxy event days are sorted by the annual ranking of system peak load for that day. Overall, the proxy event days we selected were highly comparable to the actual event days for each utility and program option.

Table 3-1: Comparison of PG&E Actual and Proxy Events

Event	Actual				Proxy			
	Annual System Peak Rank	System Load (MW)	Annual Temp Rank	Avg. Daily Temp	Annual System Peak Rank	System Load (MW)	Annual Temp Rank	Avg. Daily Temp
1	2	17,749	4	75.1	4	17,324	5	73.5
2	3	17,700	2	75.6	7	17,013	15	72.4
3	5	17,269	17	72.0	11	16,831	9	72.8
4	10	16,831	13	72.6	15	16,699	11	72.8
5	16	16,687	19	71.4	17	16,601	6	73.5
6	23	16,263	25	70.7	19	16,352	31	70.1
7	25	16,082	28	70.5	22	16,344	22	71.0
8	27	15,921	43	69.1	26	15,996	34	69.5
9	29	15,761	29	70.5	35	15,457	16	72.0
Average	**15.6**	**16,696**	**20.0**	**71.9**	**17.3**	**16,513**	**16.6**	**72.0**

Table 3-2: Comparison of SCE Actual and Proxy Events[13]

Program	Advance Notification	Day Type	Annual System Peak Rank	System Load (MW)	Annual Temp Rank	Average Daily Temp
CBP	**Day-ahead**	Proxy	15.3	18,679	14.6	77.7
		Actual	16.9	18,552	16.9	77.2
	Day-of	Proxy	15.0	18,621	17.6	77.1
		Actual	12.0	19,782	11.0	78.9
DRRC	**Day-ahead**	Proxy	15.6	18,644	17.9	77.1
		Actual	17.5	18,619	15.5	77.8
	Day-of	Proxy	15.0	18,603	14.6	77.4
		Actual	29.7	17,668	21.7	76.6

Table 3-3: Comparison of SDG&E Actual and Proxy Events

Number	Actual				Proxy			
	Annual System Peak Rank	System Load (MW)	Annual Temp Rank	Avg. Daily Temp	Annual System Peak Rank	System Load (MW)	Annual Temp Rank	Average Daily Temp
1	1	4,372	1	79.6	2	4,320	2	78.9
2	4	3,865	3	75.9	3	3,906	4	75.8
3	6	3,849	5	75.2	5	3,851	14	72.4
4	9	3,709	8	74.7	7	3,772	7	74.7
5	10	3,683	6	74.8	11	3,671	15	71.7
6	12	3,663	10	73.9	13	3,616	13	72.5
7	45	3,036	53	66.5	14	3,614	11	73.7
Average	**12.4**	**3739.6**	**12.3**	**74.4**	**7.9**	**3821.4**	**9.4**	**74.2**
Average (no outlier)	**7.0**	**3856.8**	**5.5**	**75.7**	**7.9**	**3821.4**	**9.4**	**74.2**

3.1.2 Identifying Customer Demand Reductions

Since demand reductions were simulated on unperturbed customer loads, it was necessary to identify realistic estimates of individual customer demand reductions. Commitments for both the CBP program and the aggregator contracts are made at the settlement portfolio level; aggregators do not produce estimates of expected load reductions for each customer. As a result, we relied on average demand reductions for each individual customer produced from the M&E evaluation. On average, these values

[13] Because SCE called as few as 2 or as many as 19 events for each program and DA/DO type, results are presented in a different format than for SDG&E and PG&E. Each combination of program and DA/DO had nine proxy events, some of which were shared by each combination.

are a realistic approximation of individual customers' reductions. At the individual level, the estimates include some noise, especially when a customer experienced few events.

However, the most important aspect of this exercise is that the demand reductions used in the simulation are known. Identifying customer demand reductions provides an "answer key" for the test, which then allows an assessment of which baseline methods are most accurate.

3.1.3 Application of Demand Reductions to Unperturbed Loads

Next, for the proxy event hours, the demand reductions were subtracted from the unperturbed electric load data of the corresponding customer. Through this process, the true loads with and without the simulated curtailment, as well as the demand reduction, are known.

3.1.4 Metrics for Assessing Accuracy

Accuracy refers to how close a measurement is to the actual value. It can be analyzed at the program, settlement portfolio and individual customer level. Typically, accuracy results at one level are similar to those at another level; a baseline that performs well for individual customers usually also performs well for settlement portfolios. We will focus on the accuracy of the program wide results, but also include information on how the baseline performs across settlement portfolios.

Table 3-4 summarizes the metrics used to test bias and goodness-of-fit in the assessment for each baseline method. It includes a brief description of each summary statistic and the corresponding mathematical equations used to calculate the test values.

Table 3-4: Metrics Used to Assess Baseline Accuracy

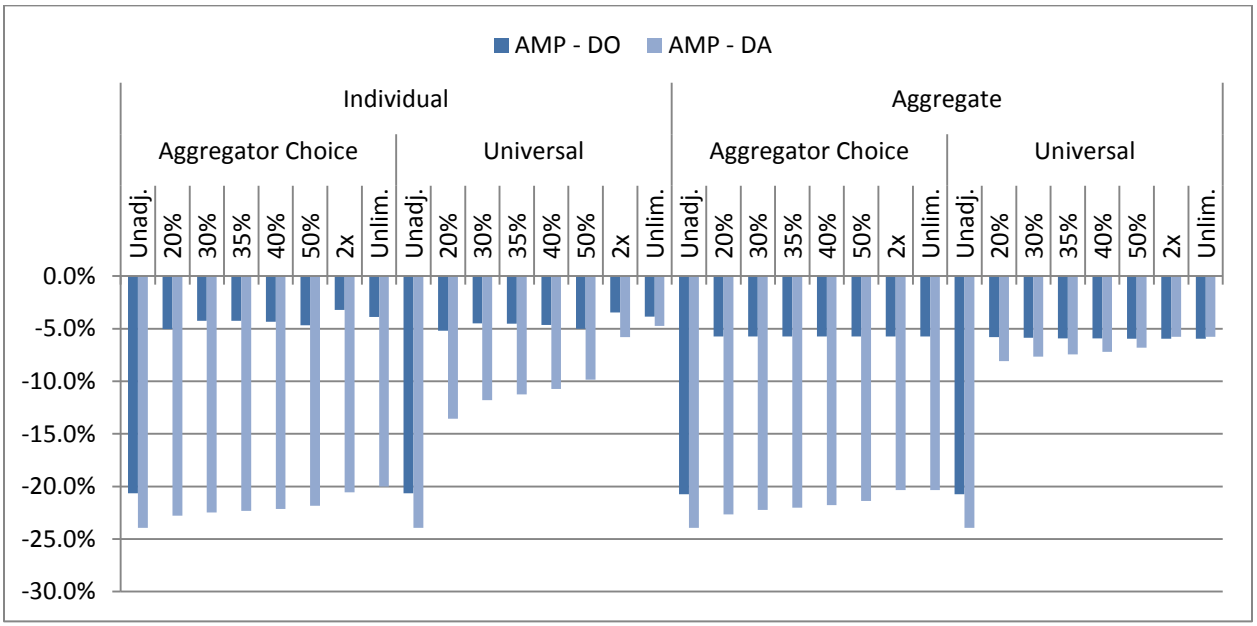
Type of Metric	Metric	Description	Mathematical Expression
Bias	Mean Percentage Error (MPE)	Mean percentage error (MPE) indicates the percentage by which measurement, on average, tends to over or underestimate the true demand reduction.	$MPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{y}_i - y_i}{\bar{y}}$
Goodness-of-fit	Mean Absolute Percentage Error (MAPE)	Mean absolute percentage error (MAPE) is a measure of the relative magnitude of errors across event days, regardless of positive or negative direction. It is normalized, allowing for comparison of results across different data sources.	$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{\hat{y}_i - y_i}{y_i}\right\rvert$
	Root Mean Squared Error (RMSE)	Root mean squared error (RMSE) is sensitive to larger errors. The squaring process gives disproportionate weight to very large errors, which are then recalibrated by taking the square root. This metric is not normalized and can only be compared between models whose errors are measured in the same units.	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$
	CV(RMSE)	This metric normalizes the RMSE by dividing it by the average of the actual demand reduction.	$CV(RMSE) = \frac{RMSE}{\bar{y}}$

Throughout the baseline accuracy sections, we will use two types of accuracy measurements. The first is bias; the second is goodness of fit. Bias denotes the error in baselines and impacts, on average. Thus, if three impacts show an error of -4%, +1%, and +6% when compared to the expected value, the bias is defined as the average of these three values, or +1%. We use three types of goodness-of-fit measures in this report, the simplest of which (mean absolute percent error) takes the average of the absolute value, or 3.6%. These two types of accuracy metrics recognize that it is just as important to be accurate for the average customer or the average event as it is to minimize the general amount of error present for individual customers or individual events.

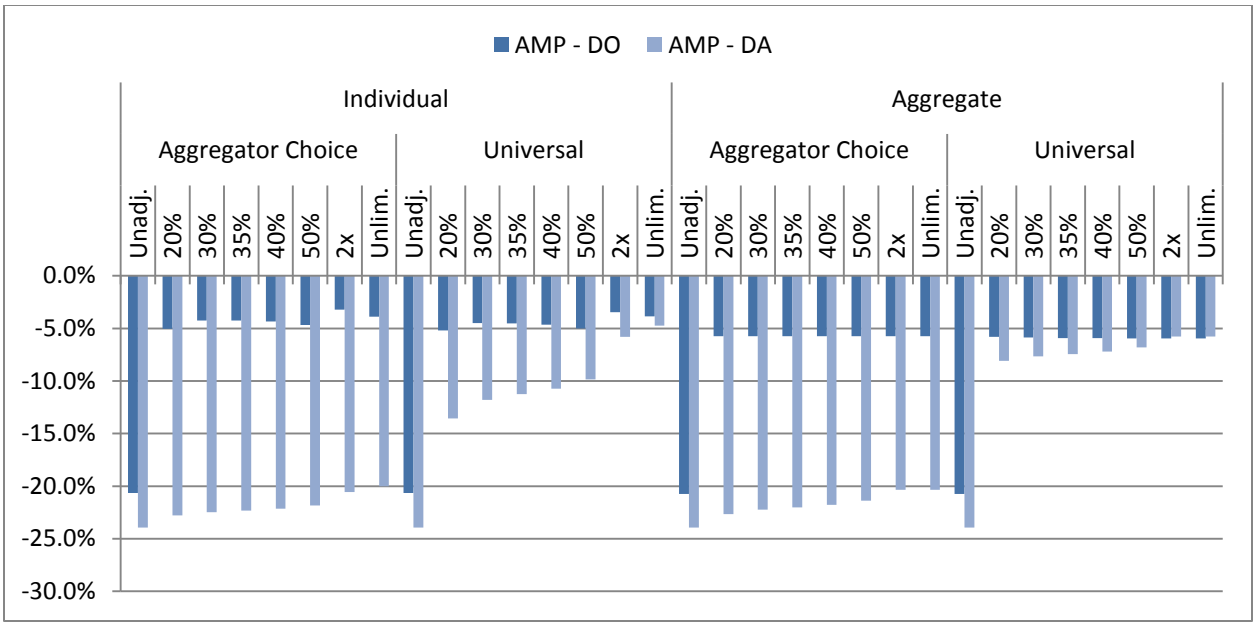
3.2 PG&E Settlement Baseline Accuracy Analysis

Figure 3-1 shows the extent to which each baseline method underestimates demand reductions for PG&E's AMP program for both the by day-of and day-ahead resources. Table 3-5 shows the same information in tabular form. A concrete example is helpful for understanding how to interpret the figure. For example, the individual, universally adjusted unlimited baseline for DA has an error of -4.7%, meaning that if an aggregator provides 1 MW of load reduction, they only get credit for 0.953 MW of that load reduction.

Figure 3-1: Bias by Baseline Type for PG&E's AMP Program[14]



[14] Table 1-1 summarized the baseline options displayed in this and other figures and tables in this report section.

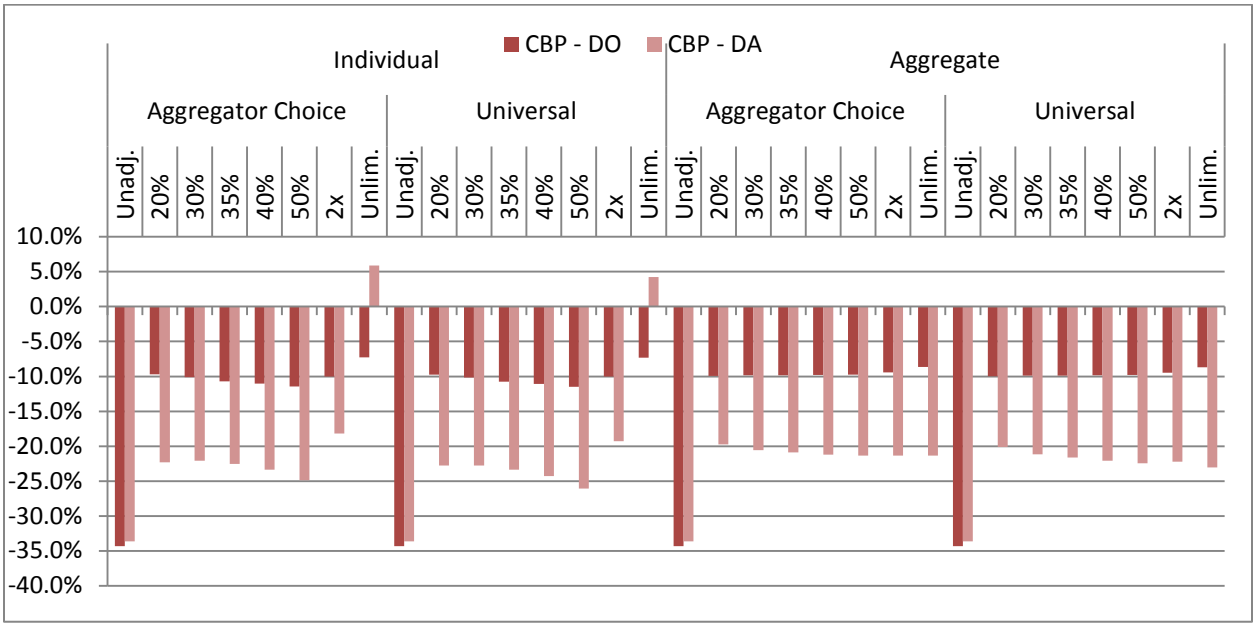


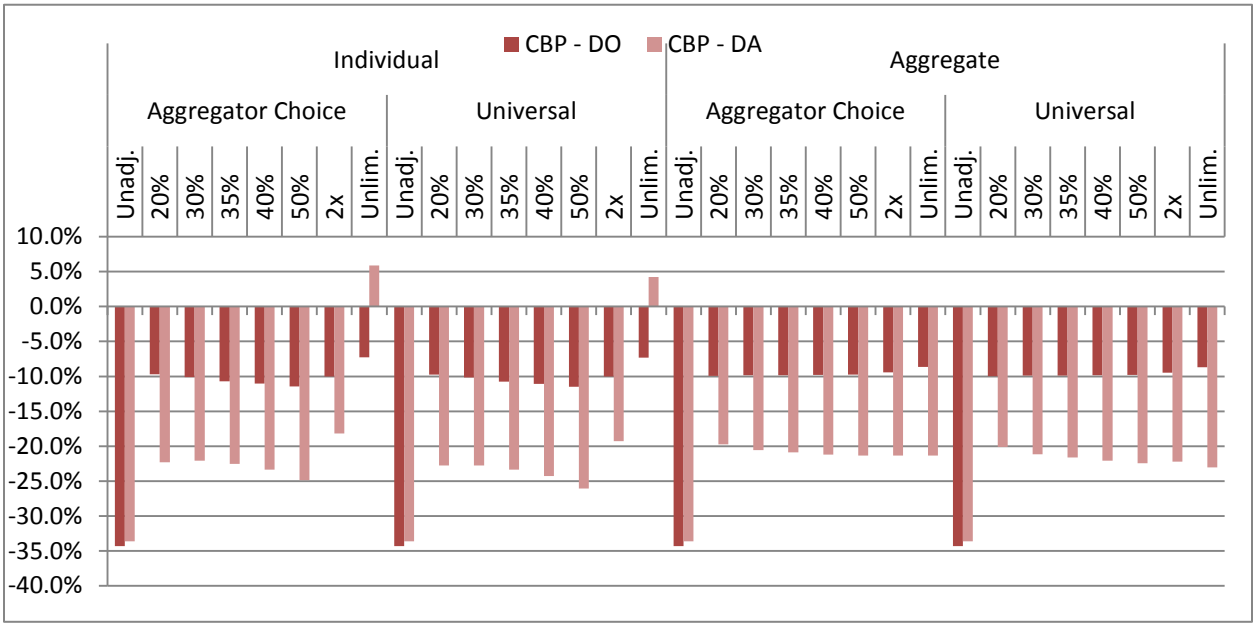
Each set of baseline rules underestimates the true impact, but the degree of bias varies. As the adjustment cap is loosened, the tendency to underestimate reductions decreases; moreover, as adjustments are applied at the settlement portfolio level and become universal, error decreases further. This is in keeping with the expected pattern. Loosening the adjustment cap allows for those baselines needing a more extreme in-day adjustment to be corrected, thus decreasing the error. Furthermore, making the adjustment universal decreases errors because it forces all accounts to apply in-day adjustments, which usually have the effect of decreasing error.[15] Finally, applying adjustments at the settlement portfolio level can also decreases errors because aggregation smoothes out the errors observed at the individual customer level. Calculating baselines for aggregated loads usually helps when underlying loads are volatile. However, when customer loads are stable and predictable, individual baselines can produce more accurate results.

Figure 3-2 shows the extent to which each baseline method under or overestimates demand reductions for PG&E's CBP Program for both the day-of and day-ahead resources. Table 3-5 shows the same information in tabular form. Impacts are underestimated for almost all baseline options except for the aggregator choice, unlimited adjustment baselines for the day-ahead resource and the universal, unlimited adjustment option for day-of resources.

[15] This is true as long as customers do not shift load to pre-event hours or reduce load early – that is, if they do not perturb the loads during the period used to calculate the same-day adjustment.

Figure 3-2: CBP Bias by Baseline Type





Note that for DA customers, the bias *does not* generally decrease as the adjustment becomes mandatory, is applied at the settlement portfolio level, and the adjustment cap is loosened. Bias does not decrease when the adjustment is mandatory because the vast majority of DA customers already choose the same-day adjustment; thus, making the adjustment mandatory does not reduce the errors further. In addition, applying the adjustment at the settlement portfolio level has very little impact on the errors because a few customers provide the bulk of the impacts, thus limiting the efficacy of adjusting baselines at the settlement portfolio level. The largest CBP DA customers are about 120 times the size of the median CBP DA customer, meaning that the impacts are highly concentrated among a handful of customers.

Table 3-5: Bias by Baseline Type for PG&E's AMP and CBP Programs

Application of Baselines	Adjustment Option	Adjustment Cap	Aggregator Contracts				Capacity Bidding Program			
			Day-Of		Day-Ahead		Day-Of		Day-Ahead	
			Est. Impacts	Error (MPE)	Est. Impacts	Error (MPE)	Est. Impacts	Error (MPE)	Est. Impacts	Error (MPE)
Simulated Impact			102.8	-	44.3	-	8.1	-	9.8	-
Individual	Aggregator Choice	Unadj.	81.6	-20.7%	33.7	-23.9%	5.3	-34.3%	6.5	-33.6%
		20%	97.6	-5.1%	34.2	-22.8%	7.3	-9.7%	7.6	-22.3%
		30%	98.5	-4.2%	34.3	-22.5%	7.3	-10.1%	7.6	-22.1%
		35%	98.5	-4.2%	34.4	-22.3%	7.2	-10.7%	7.6	-22.5%
		40%	98.4	-4.3%	34.5	-22.2%	7.2	-11.0%	7.5	-23.4%
		50%	98.0	-4.7%	34.6	-21.8%	7.1	-11.5%	7.4	-24.9%
		2x	99.5	-3.2%	35.2	-20.6%	7.3	-10.0%	8.0	-18.2%
		Unlim.	98.8	-3.9%	35.4	-20.0%	7.5	-7.3%	10.4	5.9%
	Universal	Unadj.	81.6	-20.7%	33.7	-23.9%	5.3	-34.3%	6.5	-33.6%
		20%	97.5	-5.2%	38.3	-13.6%	7.3	-9.7%	7.6	-22.8%
		30%	98.2	-4.5%	39.1	-11.8%	7.3	-10.2%	7.6	-22.8%
		35%	98.2	-4.5%	39.3	-11.2%	7.2	-10.8%	7.5	-23.3%
		40%	98.1	-4.7%	39.5	-10.7%	7.2	-11.1%	7.4	-24.3%
		50%	97.7	-5.0%	39.9	-9.8%	7.1	-11.5%	7.2	-26.1%
		2x	99.3	-3.5%	41.7	-5.8%	7.3	-10.0%	7.9	-19.3%
		Unlim.	98.9	-3.9%	42.2	-4.7%	7.5	-7.3%	10.2	4.2%
Aggregate	Aggregator Choice	Unadj.	81.5	-20.7%	33.7	-23.9%	5.3	-34.3%	6.5	-33.6%
		20%	96.9	-5.7%	34.2	-22.7%	7.3	-9.9%	7.9	-19.8%
		30%	96.9	-5.7%	34.4	-22.2%	7.3	-9.9%	7.8	-20.5%
		35%	96.9	-5.7%	34.5	-22.0%	7.3	-9.8%	7.7	-20.9%
		40%	96.9	-5.7%	34.6	-21.8%	7.3	-9.8%	7.7	-21.2%
		50%	96.9	-5.7%	34.8	-21.4%	7.3	-9.7%	7.7	-21.4%
		2x	96.9	-5.7%	35.3	-20.4%	7.3	-9.4%	7.7	-21.4%
		Unlim.	96.9	-5.7%	35.3	-20.4%	7.4	-8.6%	7.7	-21.4%
	Universal	Unadj.	81.5	-20.7%	33.7	-23.9%	5.3	-34.3%	6.5	-33.6%
		20%	96.9	-5.8%	40.7	-8.1%	7.3	-10.0%	7.8	-20.1%
		30%	96.8	-5.9%	40.9	-7.7%	7.3	-9.9%	7.7	-21.2%
		35%	96.8	-5.9%	41.0	-7.4%	7.3	-9.9%	7.7	-21.6%
		40%	96.7	-5.9%	41.1	-7.2%	7.3	-9.8%	7.6	-22.1%
		50%	96.7	-5.9%	41.3	-6.8%	7.3	-9.8%	7.6	-22.4%
		2x	96.7	-5.9%	41.7	-5.8%	7.3	-9.5%	7.6	-22.2%
		Unlim.	96.7	-5.9%	41.7	-5.8%	7.4	-8.7%	7.5	-23.0%

3.2.1 Detailed Results for PG&E's AMP Program

Table 3-6 compares the estimated impacts to the simulated reductions for each baseline method for PG&E's AMP day-ahead resource. The table echoes the findings shown above: bias (MPE) is minimized by making the adjustment mandatory, by applying the adjustment at the settlement portfolio level, and by loosening the adjustment cap.

The table also shows two goodness-of-fit statistics: mean average percent error (MAPE) and the normalized root mean squared error (CV RMSE). These statistics summarize the amount of error present, regardless of the direction of the error. Both statistics indicate that loosening the adjustment cap reduces the overall amount of error and improves the overall goodness-of-fit in addition to reducing the negative bias present in the baseline impact estimates. They also show that how well baseline methods estimate demand reductions varies from event to event. Even the most accurate set of baseline rules tends to produce errors for individual event days that on average err by at least 7% (as described by MAPE). The current baseline produces errors for individual days that are off on average by 28.1%.

Table 3-7 shows the accuracy and goodness of fit statistics for each proxy event day for PG&E's AMP day-of resource. Again, loosening the adjustment caps decreases bias and improves goodness-of-fit, as does making the adjustment mandatory. However, applying the adjustment at the settlement portfolio level has a limited effect, likely because several settlement portfolios are dominated by very large customers. In fact, one settlement portfolio has several customers that are at least 20 times as large as the median customer. Thus, applying adjustments at the settlement level only has a limited effect.

The baseline results for the AMP day-of resource are more accurate than they are for the day-ahead option. They have less tendency to underestimate demand reductions relative to EM&V results overall and are also better able to predict impacts for individual days. The current baseline produces errors for individual event hours that average 6.1%.

Table 3-6: Daily Impacts and Errors by Baseline Type, PG&E AMP, Day-ahead

Application of Baselines	Adjustment Option	Adjustment Cap	Impact (MW) by Proxy Event Day									Avg	Bias	Goodness-of-Fit	
			1	2	3	4	5	6	7	8	9		MPE	MAPE	CV RMSE
Simulated Impact			44.3									-	-	-	-
Individual	Aggregator Choice	Unadjusted	50.6	48.9	34.8	28.6	22.9	17.9	27.3	32.8	39.5	33.7	-23.9%	29.4%	0.336
		20%	50.4	48.8	34.7	28.9	23.4	18.8	28.5	34.8	39.5	34.2	-22.8%	28.1%	0.322
		30%	50.3	48.7	34.7	29.0	23.6	19.3	29.1	34.8	39.5	34.3	-22.5%	27.7%	0.317
		35%	50.2	48.6	34.7	29.1	23.7	19.5	29.4	34.8	39.5	34.4	-22.3%	27.5%	0.315
		40%	50.2	48.6	34.6	29.1	23.8	19.7	29.8	34.8	39.5	34.5	-22.2%	27.3%	0.312
		50%	50.1	48.5	34.6	29.3	24.1	20.2	30.4	34.9	39.5	34.6	-21.8%	26.9%	0.307
		2x	50.1	48.5	34.6	30.3	25.3	22.5	31.0	34.9	39.5	35.2	-20.6%	25.6%	0.288
		Unlimited	49.6	48.2	34.5	32.1	26.0	23.1	30.9	34.9	39.5	35.4	-20.0%	24.6%	0.278
	Universal	Unadjusted	50.6	48.9	34.8	28.6	22.9	17.9	27.3	32.8	39.5	33.7	-23.9%	29.4%	0.336
		20%	49.2	49.8	44.0	36.5	28.4	22.0	34.8	40.6	39.2	38.3	-13.6%	18.8%	0.237
		30%	48.3	49.5	44.7	38.3	29.7	23.5	36.8	41.5	39.3	39.1	-11.8%	16.7%	0.215
		35%	48.0	49.1	44.9	38.7	30.1	24.3	37.7	41.6	39.3	39.3	-11.2%	15.9%	0.206
		40%	47.8	48.7	45.0	39.0	30.5	25.0	38.7	41.8	39.3	39.5	-10.7%	15.1%	0.197
		50%	47.4	48.0	44.9	39.7	31.1	26.4	40.4	42.0	39.3	39.9	-9.8%	13.6%	0.181
		2x	47.7	48.0	45.0	42.4	34.1	32.4	43.8	42.5	39.6	41.7	-5.8%	9.7%	0.130
		Unlimited	46.1	46.4	44.7	46.1	35.7	34.0	43.9	42.8	40.0	42.2	-4.7%	7.8%	0.110
Aggregate	Aggregator Choice	Unadjusted	50.6	48.9	34.8	28.6	22.9	17.9	27.3	32.8	39.5	33.7	-23.9%	29.4%	0.336
		20%	50.0	48.4	34.4	29.3	23.7	19.2	28.9	34.8	39.6	34.2	-22.7%	27.6%	0.317
		30%	49.8	48.1	34.4	29.7	24.1	19.8	29.7	34.8	39.6	34.4	-22.2%	26.9%	0.309
		35%	49.6	48.1	34.4	29.8	24.3	20.1	30.1	34.8	39.6	34.5	-22.0%	26.6%	0.305
		40%	49.5	48.1	34.4	30.0	24.5	20.4	30.5	34.8	39.6	34.6	-21.8%	26.3%	0.302
		50%	49.4	48.1	34.4	30.4	24.9	21.0	30.7	34.8	39.6	34.8	-21.4%	25.9%	0.296
		2x	49.4	48.1	34.4	31.7	25.8	22.9	30.7	34.8	39.6	35.3	-20.4%	24.8%	0.281
		Unlimited	49.4	48.1	34.4	31.7	25.8	22.9	30.7	34.8	39.6	35.3	-20.4%	24.8%	0.281
	Universal	Unadjusted	50.6	48.9	34.8	28.6	22.9	17.9	27.3	32.8	39.5	33.7	-23.9%	29.4%	0.336
		20%	45.4	45.8	44.2	43.2	33.4	30.2	40.9	42.9	40.2	40.7	-8.1%	9.4%	0.141
		30%	45.2	45.5	44.2	43.6	33.8	30.9	41.8	42.9	40.2	40.9	-7.7%	8.7%	0.134
		35%	45.0	45.4	44.2	43.8	34.0	31.2	42.2	42.9	40.2	41.0	-7.4%	8.4%	0.131
		40%	44.9	45.4	44.2	43.9	34.2	31.5	42.6	42.9	40.2	41.1	-7.2%	8.1%	0.128
		50%	44.9	45.4	44.2	44.3	34.6	32.1	42.7	42.9	40.2	41.3	-6.8%	7.7%	0.122
		2x	44.9	45.4	44.2	45.6	35.5	34.0	42.7	42.9	40.2	41.7	-5.8%	7.3%	0.108
		Unlimited	44.9	45.4	44.2	45.6	35.5	34.0	42.7	42.9	40.2	41.7	-5.8%	7.3%	0.108

Table 3-7: Daily Impacts and Errors by Baseline Type, PG&E AMP, Day-of

Application of Baselines	Adjustment Option	Adjustment Cap	Impact (MW) by Proxy Event Day									Avg	Bias	Goodness-of-Fit	
			1	2	3	4	5	6	7	8	9		MPE	MAPE	CV RMSE
Simulated Impact			102.8									-	-	-	-
Individual	Aggregator Choice	Unadjusted	98.8	86.1	61.8	81.7	92.8	90.9	92.0	58.6	71.7	81.6	-20.7%	20.7%	0.245
		20%	104.8	102.3	95.9	105.9	96.7	92.0	90.7	95.5	95.0	97.6	-5.1%	6.1%	0.071
		30%	103.4	102.1	99.6	108.1	96.9	90.4	90.9	98.3	96.7	98.5	-4.2%	5.5%	0.067
		35%	102.4	101.8	100.1	108.4	97.2	89.6	90.8	99.1	96.9	98.5	-4.2%	5.4%	0.068
		40%	101.7	101.4	100.4	108.5	97.2	89.0	90.7	99.6	97.0	98.4	-4.3%	5.6%	0.069
		50%	100.2	100.4	100.7	108.5	97.1	88.0	90.6	100.0	96.9	98.0	-4.7%	5.9%	0.072
		2x	101.4	101.8	102.4	111.4	98.2	89.3	91.6	101.8	97.9	99.5	-3.2%	5.1%	0.067
		Unlimited	98.9	101.9	102.1	115.6	98.1	85.8	90.3	100.6	96.5	98.8	-3.9%	6.6%	0.086
	Universal	Unadjusted	98.8	86.1	61.8	81.7	92.8	90.9	92.0	58.6	71.7	81.6	-20.7%	20.7%	0.245
		20%	104.9	103.5	95.9	105.1	96.2	91.7	91.3	94.5	94.4	97.5	-5.2%	6.3%	0.073
		30%	103.4	103.3	99.6	107.3	96.3	90.0	91.5	96.8	95.7	98.2	-4.5%	5.7%	0.069
		35%	102.5	103.1	100.1	107.6	96.5	89.3	91.4	97.4	95.8	98.2	-4.5%	5.6%	0.070
		40%	101.7	102.8	100.4	107.7	96.6	88.7	91.3	97.6	95.8	98.1	-4.7%	5.7%	0.071
		50%	100.2	101.8	100.7	107.7	96.4	87.6	91.3	98.0	95.7	97.7	-5.0%	6.0%	0.074
		2x	101.5	103.4	102.5	110.7	97.6	89.0	92.3	99.9	96.8	99.3	-3.5%	5.3%	0.068
		Unlimited	99.0	103.7	101.8	115.1	97.6	85.3	93.4	98.7	95.3	98.9	-3.9%	6.7%	0.083
Aggregate	Aggregator Choice	Unadjusted	98.6	86.1	61.2	81.7	92.8	90.9	92.0	58.6	71.7	81.5	-20.7%	20.7%	0.246
		20%	95.4	99.5	100.0	111.1	98.5	86.3	87.3	99.0	95.2	96.9	-5.7%	7.5%	0.089
		30%	95.4	99.5	100.0	111.1	98.5	86.3	87.3	99.0	95.2	96.9	-5.7%	7.5%	0.089
		35%	95.4	99.5	100.0	111.1	98.5	86.3	87.3	99.0	95.2	96.9	-5.7%	7.5%	0.089
		40%	95.4	99.5	100.0	111.1	98.5	86.3	87.3	99.0	95.2	96.9	-5.7%	7.5%	0.089
		50%	95.4	99.5	100.0	111.1	98.5	86.3	87.3	99.0	95.2	96.9	-5.7%	7.5%	0.089
		2x	95.4	99.5	100.0	111.1	98.5	86.3	87.3	99.0	95.2	96.9	-5.7%	7.5%	0.089
		Unlimited	95.4	99.5	100.0	111.1	98.5	86.3	87.3	99.0	95.2	96.9	-5.7%	7.5%	0.089
	Universal	Unadjusted	98.6	86.1	61.2	81.7	92.8	90.9	92.0	58.6	71.7	81.5	-20.7%	20.7%	0.246
		20%	95.6	101.2	99.6	110.6	98.1	85.9	88.4	98.1	94.5	96.9	-5.8%	7.5%	0.088
		30%	95.6	101.2	99.6	110.6	98.1	85.9	88.5	97.7	94.1	96.8	-5.9%	7.5%	0.088
		35%	95.6	101.2	99.6	110.6	98.1	85.9	88.5	97.5	94.0	96.8	-5.9%	7.6%	0.089
		40%	95.6	101.2	99.6	110.6	98.1	85.9	88.5	97.2	94.0	96.7	-5.9%	7.6%	0.089
		50%	95.6	101.2	99.6	110.6	98.1	85.9	88.5	97.2	94.0	96.7	-5.9%	7.6%	0.089
		2x	95.6	101.2	99.6	110.6	98.1	85.9	88.5	97.2	94.0	96.7	-5.9%	7.6%	0.089
		Unlimited	95.6	101.2	99.6	110.6	98.1	85.9	88.5	97.2	94.0	96.7	-5.9%	7.6%	0.089

3.2.2 Detailed Results for PG&E's CBP Program

Table 3-8 compares estimated impacts for each proxy event with the simulated demand reductions by baseline option for PG&E's CBP day-ahead resource. As described in the beginning of this chapter, bias *does not* generally decrease as the adjustment becomes mandatory, when the adjustment is applied at the settlement portfolio level, and when the adjustment cap is loosened. Bias does not decrease when the adjustment is universal because the vast majority of day-ahead customers already choose the same-day adjustment; as such, making the adjustment universal does not have any effect. In addition, applying the adjustment at the settlement portfolio level has very little impact on the tendency of baselines to underestimate. A few customers provide the bulk of the impacts, thus limiting the effect of calculating and adjusting baselines at the settlement portfolio level.

While the unlimited baseline is less likely to consistently under predict reductions (i.e., is the least biased) when applied individually, it produces estimates that err by a wide margin for individual event days. The unlimited baseline does well for most customers that have stable, predictable loads, but does very poorly for a small subset of customers. These few customers experience very large errors and affect the program level results. Thus, the unlimited adjustment may not be a desirable choice. The other baseline options tested do not substantially reduce bias or improve goodness-of-fit. The current baseline produces errors for individual event hours that average 22.3%.

Table 3-9 shows the accuracy and goodness of fit statistics for PG&E's CBP day-of resource. The results show the usual pattern of a decrease in the tendency to underestimate impacts when the adjustment is mandatory, the cap is loosened and the settlement calculate is done at the portfolio level. Loosening the adjustment cap has a limited impact when the adjustment is applied at the settlement portfolio level and is mandatory because less extreme adjustments are required when the load has been aggregated; individual volatility has been smoothed out.

The baseline results for the CBP day-of resource are more accurate than they are for the day-ahead option. They have less tendency to underestimate demand reductions overall (i.e., less bias) and are better able to predict impacts for individual days (i.e., better goodness-of-fit). The current baseline tends to produce errors for individual event hours that average 11.0% (MAPE).

Table 3-8: Daily Impacts and Errors by Baseline Type, PG&E CBP, Day-ahead

Application of Baselines	Adjustment Option	Adjustment Cap	Impact (MW) by Proxy Event Day									Avg	Bias	Goodness-of-Fit	
			1	2	3	4	5	6	7	8	9		MPE	MAPE	CV RMSE
Simulated Impact			9.8									-	-	-	-
Individual	Aggregator Choice	Unadjusted	7.9	7.1	1.8	8.9	5.3	7.0	9.4	5.0	6.1	6.5	-33.6%	33.6%	0.403
		20%	8.8	8.6	5.1	9.5	6.6	6.0	8.5	7.8	7.4	7.6	-22.3%	22.3%	0.264
		30%	8.5	8.5	5.3	9.8	7.0	5.5	8.3	8.1	7.6	7.6	-22.1%	22.2%	0.262
		35%	8.3	8.5	5.3	9.8	7.1	5.3	8.2	8.2	7.5	7.6	-22.5%	22.7%	0.267
		40%	8.1	8.4	5.3	9.7	7.1	5.1	8.1	8.3	7.5	7.5	-23.4%	23.4%	0.275
		50%	8.0	8.3	5.3	9.3	6.9	4.6	8.0	8.4	7.4	7.4	-24.9%	24.9%	0.289
		2x	9.0	8.7	6.2	9.6	7.4	5.4	8.4	9.5	7.8	8.0	-18.2%	18.2%	0.229
		Unlimited	9.9	9.1	8.0	8.4	7.0	25.6	8.2	9.2	7.9	10.4	5.9%	30.2%	0.559
	Universal	Unadjusted	7.9	7.1	1.8	8.9	5.3	7.0	9.4	5.0	6.1	6.5	-33.6%	33.6%	0.403
		20%	8.7	8.5	5.2	9.5	6.5	5.9	8.4	7.9	7.5	7.6	-22.8%	22.8%	0.266
		30%	8.3	8.4	5.4	9.8	6.9	5.4	8.2	8.1	7.6	7.6	-22.8%	22.8%	0.267
		35%	8.1	8.3	5.4	9.8	7.0	5.1	8.1	8.2	7.5	7.5	-23.3%	23.3%	0.273
		40%	7.8	8.2	5.4	9.6	6.9	4.8	8.0	8.3	7.5	7.4	-24.3%	24.3%	0.282
		50%	7.7	8.0	5.5	9.3	6.8	4.3	7.8	8.4	7.4	7.2	-26.1%	26.1%	0.299
		2x	8.7	8.5	6.4	9.5	7.2	5.1	8.3	9.5	7.8	7.9	-19.3%	19.3%	0.238
		Unlimited	9.5	8.8	8.2	8.4	6.7	25.1	8.3	9.1	7.8	10.2	4.2%	30.5%	0.545
Aggregate	Aggregator Choice	Unadjusted	7.9	7.1	1.8	8.9	5.3	7.0	9.4	5.0	6.1	6.5	-33.6%	33.6%	0.403
		20%	9.4	9.4	7.4	8.3	7.4	5.3	7.7	8.9	6.8	7.9	-19.8%	19.8%	0.236
		30%	9.4	9.4	7.5	8.3	7.4	4.7	7.6	8.9	6.8	7.8	-20.5%	20.5%	0.250
		35%	9.4	9.4	7.5	8.3	7.4	4.4	7.6	8.9	6.8	7.7	-20.9%	20.9%	0.257
		40%	9.4	9.4	7.5	8.3	7.4	4.2	7.6	8.9	6.8	7.7	-21.2%	21.2%	0.264
		50%	9.4	9.4	7.5	8.3	7.4	4.0	7.6	8.9	6.8	7.7	-21.4%	21.4%	0.267
		2x	9.4	9.4	7.5	8.3	7.4	4.0	7.6	8.9	6.8	7.7	-21.4%	21.4%	0.267
		Unlimited	9.4	9.4	7.5	8.3	7.4	4.0	7.6	8.9	6.8	7.7	-21.4%	21.4%	0.267
	Universal	Unadjusted	7.9	7.1	1.8	8.9	5.3	7.0	9.4	5.0	6.1	6.5	-33.6%	33.6%	0.403
		20%	9.2	9.3	7.5	8.3	7.3	5.2	7.7	9.0	6.8	7.8	-20.1%	20.1%	0.238
		30%	9.2	9.3	7.6	8.3	7.3	4.6	7.6	9.0	6.8	7.7	-21.2%	21.2%	0.255
		35%	9.1	9.3	7.6	8.3	7.3	4.2	7.6	8.9	6.8	7.7	-21.6%	21.6%	0.263
		40%	9.1	9.2	7.6	8.3	7.3	3.9	7.6	8.9	6.8	7.6	-22.1%	22.1%	0.271
		50%	9.0	9.2	7.6	8.3	7.2	3.7	7.6	8.9	6.8	7.6	-22.4%	22.4%	0.276
		2x	9.0	9.2	7.6	8.3	7.2	3.7	7.7	8.9	6.8	7.6	-22.2%	22.2%	0.275
		Unlimited	9.0	9.0	7.6	8.3	7.1	3.5	7.7	8.8	6.8	7.5	-23.0%	23.0%	0.284

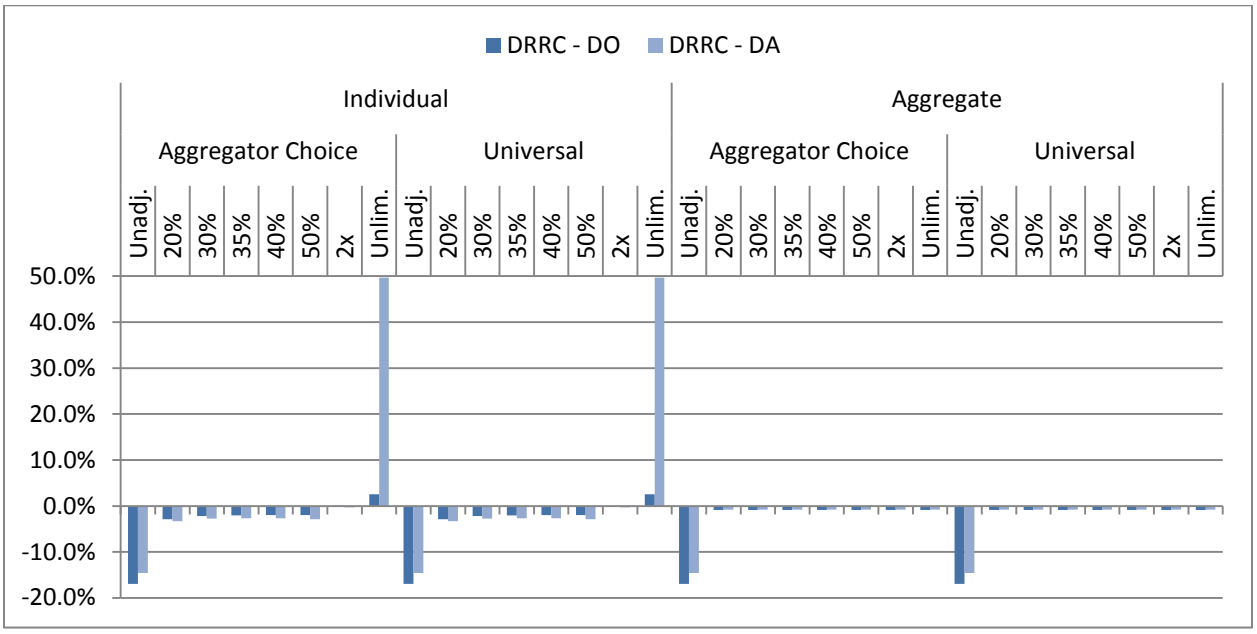
Table 3-9: Daily Impacts and Errors by Baseline Type, PG&E CBP, Day-of

Application of Baselines	Adjustment Option	Adjustment Cap	Impact (MW) by Proxy Event Day									Avg	Bias	Goodness-of-Fit	
			1	2	3	4	5	6	7	8	9		MPE	MAPE	CV RMSE
Simulated Impact			8.1									-	-	-	-
Individual	Aggregator Choice	Unadjusted	4.8	3.1	2.0	3.8	8.7	9.2	7.7	3.8	4.8	5.3	-34.3%	38.9%	0.456
		20%	6.9	6.0	6.3	7.5	8.1	6.8	8.5	8.0	7.5	7.3	-9.7%	11.0%	0.139
		30%	7.0	6.3	6.7	7.6	7.9	6.2	8.4	8.1	7.1	7.3	-10.1%	11.3%	0.138
		35%	7.0	6.3	6.7	7.6	7.8	6.0	8.4	8.2	6.9	7.2	-10.7%	11.8%	0.143
		40%	7.0	6.3	6.8	7.5	7.8	5.9	8.3	8.2	6.8	7.2	-11.0%	12.1%	0.147
		50%	7.1	6.4	6.8	7.5	7.7	5.7	8.3	8.2	6.7	7.1	-11.5%	12.4%	0.151
		2x	7.3	6.5	7.1	7.6	8.0	5.7	8.3	8.2	6.7	7.3	-10.0%	11.0%	0.141
		Unlimited	7.5	7.2	7.0	8.3	9.6	5.2	8.1	8.1	6.4	7.5	-7.3%	12.2%	0.163
	Universal	Unadjusted	4.8	3.1	2.0	3.8	8.7	9.2	7.7	3.8	4.8	5.3	-34.3%	38.9%	0.456
		20%	6.9	6.0	6.3	7.5	8.1	6.8	8.5	8.0	7.5	7.3	-9.7%	11.1%	0.140
		30%	7.0	6.3	6.7	7.6	7.9	6.2	8.4	8.1	7.1	7.3	-10.2%	11.4%	0.139
		35%	6.9	6.3	6.7	7.6	7.8	6.1	8.4	8.2	6.9	7.2	-10.8%	11.9%	0.144
		40%	7.0	6.4	6.7	7.5	7.8	5.9	8.3	8.2	6.8	7.2	-11.1%	12.1%	0.147
		50%	7.1	6.4	6.8	7.5	7.7	5.7	8.3	8.2	6.7	7.1	-11.5%	12.4%	0.152
		2x	7.2	6.5	7.0	7.6	8.0	5.7	8.3	8.2	6.7	7.3	-10.0%	11.1%	0.142
		Unlimited	7.5	7.2	7.0	8.3	9.6	5.2	8.1	8.1	6.4	7.5	-7.3%	12.3%	0.164
Aggregate	Aggregator Choice	Unadjusted	4.8	3.1	2.0	3.8	8.7	9.2	7.7	3.8	4.8	5.3	-34.3%	38.9%	0.456
		20%	7.0	7.6	6.7	7.6	8.3	5.3	7.9	8.1	6.9	7.3	-9.9%	10.8%	0.148
		30%	7.0	7.6	6.7	7.6	8.4	5.3	7.9	8.1	6.9	7.3	-9.9%	10.8%	0.148
		35%	7.0	7.6	6.7	7.6	8.4	5.3	7.9	8.1	6.9	7.3	-9.8%	10.8%	0.148
		40%	7.0	7.6	6.7	7.6	8.4	5.3	7.9	8.1	6.9	7.3	-9.8%	10.9%	0.148
		50%	7.0	7.6	6.7	7.6	8.4	5.3	7.9	8.1	6.9	7.3	-9.7%	10.9%	0.148
		2x	7.0	7.6	6.7	7.7	8.6	5.3	7.9	8.1	6.9	7.3	-9.4%	11.0%	0.149
		Unlimited	7.0	7.6	6.7	8.2	8.7	5.3	7.9	8.1	6.9	7.4	-8.6%	10.7%	0.148
	Universal	Unadjusted	4.8	3.1	2.0	3.8	8.7	9.2	7.7	3.8	4.8	5.3	-34.3%	38.9%	0.456
		20%	7.0	7.6	6.7	7.6	8.4	5.3	7.9	8.1	6.9	7.3	-10.0%	10.9%	0.148
		30%	7.0	7.6	6.7	7.6	8.4	5.3	7.9	8.1	6.9	7.3	-9.9%	10.9%	0.149
		35%	7.0	7.6	6.7	7.6	8.4	5.3	7.9	8.1	6.9	7.3	-9.9%	10.9%	0.149
		40%	7.0	7.6	6.7	7.6	8.4	5.3	7.9	8.1	6.9	7.3	-9.8%	10.9%	0.149
		50%	7.0	7.6	6.7	7.6	8.4	5.3	7.9	8.1	6.9	7.3	-9.8%	10.9%	0.149
		2x	7.0	7.6	6.7	7.7	8.6	5.3	7.9	8.1	6.9	7.3	-9.5%	11.1%	0.149
		Unlimited	7.0	7.6	6.7	8.2	8.7	5.3	7.9	8.1	6.9	7.4	-8.7%	10.8%	0.149

3.3 SCE Settlement Baseline Accuracy Analysis

Figure 3-3 shows the extent to which each baseline method underestimates or overestimates demand reductions for SCE's DRRC program for both the day-of and day-ahead resources. Table 3-10 shows the same information in tabular form.

Figure 3-3: Bias by Baseline Type for SCE's DRRC Program



Except for the individual, unlimited aggregator choice baselines, each baseline method underestimates the true impact, although the magnitude of the underestimation varies. As the adjustment cap is loosened, the tendency to underestimate reductions decreases; moreover, as adjustments are applied at the settlement portfolio level and are mandatory, error decreases further. This is in keeping with the patterns seen for PG&E's AMP program.

The one exception to the pattern is the unlimited aggregator choice baseline, which *overestimates* demand reduction by nearly 50%. This is because one particular day-ahead settlement portfolio has a disproportionate number of accounts with enormous unlimited ratio adjustments; one account has an adjustment ratio of 450%, while four others have ratio adjustments greater than 200%. These massive adjustments are not needed and produce an enormous amount of bias in the baselines for these customers when baselines are calculated at the individual level, which, when averaged together with the bias for all 158 customers, creates an average upward bias of about 50%. When aggregating loads and *then* applying the adjustment, the error is drastically reduced, because it is much easier to estimate an accurate baseline for a large group of customers than for one individual customer.

Figure 3-4 shows the extent to which each baseline method underestimates demand reductions for SCE's CBP program for both the day-of and day-ahead resource. Table 3-10 shows the same information in tabular form. Making the adjustment universal reduces bias for day-of customers, but has little effect for day-ahead customers, since most ay ahead settlement portfolios already apply the adjustment. In addition, loosening the adjustment cap has little effect for day-of or day-ahead

customers; once any kind of adjustment is applied, loosening the cap does not markedly reduce bias. This is likely because the adjustments fall below progressively higher adjustment caps.

Figure 3-4: Bias by Baseline Type SCE's CBP Program

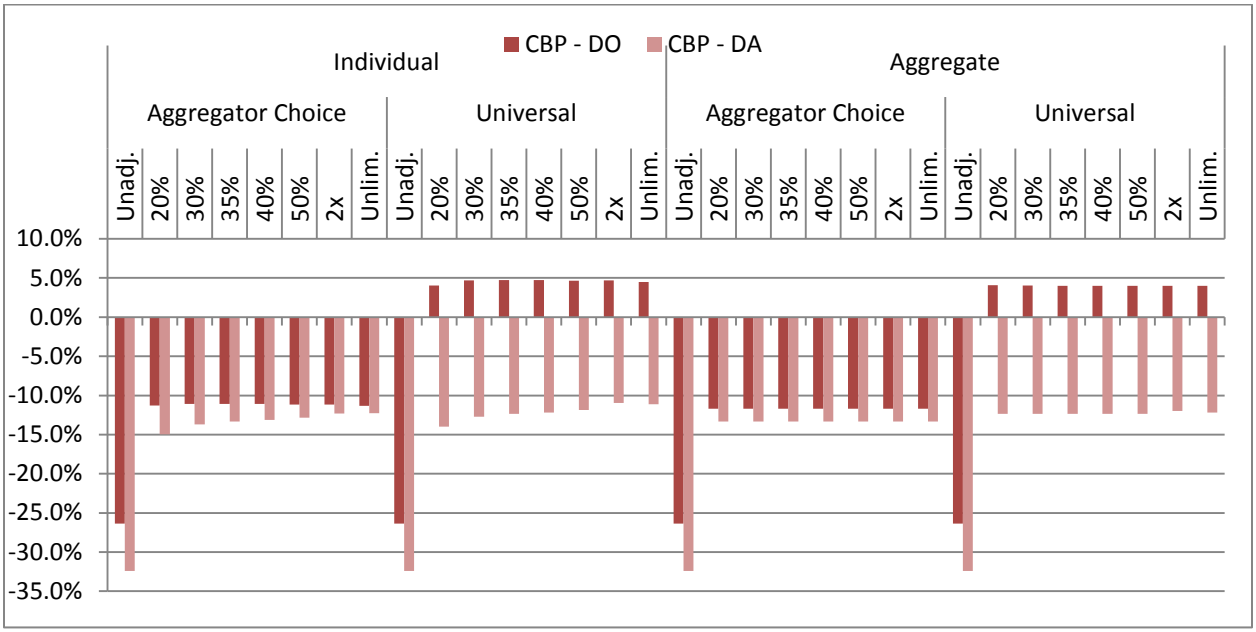


Table 3-10: Bias by Baseline Type for SCE's DRRC and CBP Programs

Application of Baselines	Adjustment Option	Adjustment Cap	Demand Response Resource Contracts				Capacity Bidding Program			
			Day-of		Day-ahead		Day-of		Day-ahead	
			Estimated Impacts	Error (MPE)	Estimated Impacts	Error (MPE)	Estimated Impacts	Error (MPE)	Estimated Impacts	Error (MPE)
Simulated Impact			132.7	-	18.3	-	15.0	-	4.7	-
Individual	Aggregator Choice	Unadjusted	110.2	-16.9%	15.7	-14.6%	11.0	-26.4%	3.2	-32.4%
		20%	128.9	-2.8%	17.7	-3.3%	13.3	-11.3%	4.0	-15.0%
		30%	129.8	-2.1%	17.9	-2.7%	13.3	-11.1%	4.1	-13.7%
		35%	130.0	-2.0%	17.9	-2.7%	13.3	-11.1%	4.1	-13.3%
		40%	130.0	-2.0%	17.9	-2.7%	13.3	-11.1%	4.1	-13.1%
		50%	130.1	-1.9%	17.8	-2.9%	13.3	-11.1%	4.1	-12.8%
		2x	132.8	0.1%	18.3	-0.3%	13.3	-11.1%	4.1	-12.3%
		Unlimited	136.0	2.5%	27.5	49.7%	13.3	-11.3%	4.1	-12.2%
	Universal	Unadjusted	110.2	-16.9%	15.7	-14.6%	11.0	-26.4%	3.2	-32.4%
		20%	128.9	-2.8%	17.7	-3.3%	15.6	4.0%	4.1	-14.0%
		30%	129.8	-2.1%	17.9	-2.7%	15.7	4.7%	4.1	-12.7%
		35%	130.0	-2.0%	17.9	-2.7%	15.7	4.7%	4.1	-12.4%
		40%	130.0	-2.0%	17.9	-2.7%	15.7	4.7%	4.1	-12.2%
		50%	130.1	-1.9%	17.8	-2.9%	15.7	4.7%	4.2	-11.9%
		2x	132.8	0.1%	18.3	-0.3%	15.7	4.7%	4.2	-11.0%
		Unlimited	136.0	2.5%	27.5	49.7%	15.7	4.5%	4.2	-11.1%
Aggregate	Aggregator Choice	Unadjusted	110.2	-16.9%	15.7	-14.6%	11.0	-26.4%	3.2	-32.4%
		20%	131.5	-0.9%	18.2	-0.8%	13.2	-11.7%	4.1	-13.3%
		30%	131.5	-0.9%	18.2	-0.8%	13.2	-11.7%	4.1	-13.3%
		35%	131.5	-0.9%	18.2	-0.8%	13.2	-11.7%	4.1	-13.3%
		40%	131.5	-0.9%	18.2	-0.8%	13.2	-11.7%	4.1	-13.3%
		50%	131.5	-0.9%	18.2	-0.8%	13.2	-11.7%	4.1	-13.3%
		2x	131.5	-0.9%	18.2	-0.8%	13.2	-11.7%	4.1	-13.3%
		Unlimited	131.5	-0.9%	18.2	-0.8%	13.2	-11.7%	4.1	-13.3%
	Universal	Unadjusted	110.2	-16.9%	15.7	-14.6%	11.0	-26.4%	3.2	-32.4%
		20%	131.5	-0.9%	18.2	-0.8%	15.6	4.1%	4.1	-12.3%
		30%	131.5	-0.9%	18.2	-0.8%	15.6	4.0%	4.1	-12.3%
		35%	131.5	-0.9%	18.2	-0.8%	15.6	4.0%	4.1	-12.3%
		40%	131.5	-0.9%	18.2	-0.8%	15.6	4.0%	4.1	-12.3%
		50%	131.5	-0.9%	18.2	-0.8%	15.6	4.0%	4.1	-12.3%
		2x	131.5	-0.9%	18.2	-0.8%	15.6	4.0%	4.2	-12.0%
		Unlimited	131.5	-0.9%	18.2	-0.8%	15.6	4.0%	4.1	-12.2%

3.3.1 Detailed Results for SCE's DRRC Program

Table 3-11 compares the estimated impacts to the simulated reductions for each baseline type for SCE's DRRC day-ahead resource. The table echoes the findings discussed above: bias is minimized by applying the adjustment at the aggregate level. Individual unlimited adjustments produce an enormous amount of bias because a few customers have extreme ratio adjustments. Note that the vast majority of the error in the unlimited individual baseline comes from the second proxy event day, showing that unlimited baselines do very well for most days, but can produce very volatile results for individual event days and individual customers.

The table also contains values for two goodness-of-fit statistics: MAPE and CV RMSE. Both of these statistics show that, generally, baselines with the lowest amount of bias also have the best goodness-of-fit.

Table 3-12 shows the estimated reductions for each proxy event day and compares these with the simulated demand reductions for each baseline method for SCE's DRRC day-of resource. Again, the results echo what was shown in the previous section. Bias decreases as the caps are loosened, when adjustments are mandatory, and when adjustments are made at the settlement portfolio level. Note that for individually calculated baselines, bias improves when loosening the cap, but goodness-of-fit does not. This indicates that some extreme errors for individual customers on individual days are retained.

Table 3-11: Daily Impacts and Errors by Baseline Type, SCE DRRC, Day-ahead

Application of Baselines	Adjustment Option	Adjustment Cap	Impact (MW) by Proxy Event Day									Avg.	Bias	Goodness-of-fit	
			1	2	3	4	5	6	7	8	9		MPE	MAPE	CV RMSE
Simulated Impact			18.3												
Individual	**Aggregator Choice**	**Unadjusted**	15.1	16.3	16.0	15.0	14.9	15.6	15.7	16.3	16.0	15.7	-14.6%	14.6%	0.149
		20%	18.3	17.5	18.1	16.1	16.5	17.1	16.5	19.9	19.6	17.7	-3.3%	6.7%	0.077
		30%	18.6	17.4	18.3	16.3	16.6	17.2	16.6	20.1	19.5	17.9	-2.7%	6.6%	0.075
		35%	18.8	17.4	18.3	16.3	16.7	17.2	16.6	20.1	19.4	17.9	-2.7%	6.6%	0.074
		40%	18.9	17.4	18.3	16.4	16.7	17.1	16.5	20.1	19.3	17.9	-2.7%	6.6%	0.075
		50%	19.0	17.3	18.3	16.4	16.7	17.1	16.5	20.1	19.1	17.8	-2.9%	6.7%	0.075
		2x	20.1	17.7	18.8	16.7	17.1	17.7	16.7	20.4	19.4	18.3	-0.3%	6.7%	0.073
		Unlimited	20.1	101.9	18.8	16.7	17.1	17.9	16.6	19.7	18.4	27.5	49.7%	55.8%	1.519
	Universal	**Unadjusted**	15.1	16.3	16.0	15.0	14.9	15.6	15.7	16.3	16.0	15.7	-14.6%	14.6%	0.149
		20%	18.3	17.5	18.1	16.1	16.5	17.1	16.5	19.9	19.6	17.7	-3.3%	6.7%	0.077
		30%	18.6	17.4	18.3	16.3	16.6	17.2	16.6	20.1	19.5	17.9	-2.7%	6.6%	0.075
		35%	18.8	17.4	18.3	16.3	16.7	17.2	16.6	20.1	19.4	17.9	-2.7%	6.6%	0.074
		40%	18.9	17.4	18.3	16.4	16.7	17.1	16.5	20.1	19.3	17.9	-2.7%	6.6%	0.075
		50%	19.0	17.3	18.3	16.4	16.7	17.1	16.5	20.1	19.1	17.8	-2.9%	6.7%	0.075
		2x	20.1	17.7	18.8	16.7	17.1	17.7	16.7	20.4	19.4	18.3	-0.3%	6.7%	0.073
		Unlimited	20.1	101.9	18.8	16.7	17.1	17.9	16.6	19.7	18.4	27.5	49.7%	55.8%	1.519
Aggregate	**Aggregator Choice**	**Unadjusted**	15.1	16.3	16.0	15.0	14.9	15.6	15.7	16.3	16.0	15.7	-14.6%	14.6%	0.149
		20%	19.8	17.9	19.1	16.8	17.8	18.9	16.3	19.5	17.8	18.2	-0.8%	5.5%	0.062
		30%	19.8	17.9	19.1	16.8	17.8	18.9	16.3	19.5	17.8	18.2	-0.8%	5.5%	0.062
		35%	19.8	17.9	19.1	16.8	17.8	18.9	16.3	19.5	17.8	18.2	-0.8%	5.5%	0.062
		40%	19.8	17.9	19.1	16.8	17.8	18.9	16.3	19.5	17.8	18.2	-0.8%	5.5%	0.062
		50%	19.8	17.9	19.1	16.8	17.8	18.9	16.3	19.5	17.8	18.2	-0.8%	5.5%	0.062
		2x	19.8	17.9	19.1	16.8	17.8	18.9	16.3	19.5	17.8	18.2	-0.8%	5.5%	0.062
		Unlimited	19.8	17.9	19.1	16.8	17.8	18.9	16.3	19.5	17.8	18.2	-0.8%	5.5%	0.062
	Universal	**Unadjusted**	15.1	16.3	16.0	15.0	14.9	15.6	15.7	16.3	16.0	15.7	-14.6%	14.6%	0.149
		20%	19.8	17.9	19.1	16.8	17.8	18.9	16.3	19.5	17.8	18.2	-0.8%	5.5%	0.062
		30%	19.8	17.9	19.1	16.8	17.8	18.9	16.3	19.5	17.8	18.2	-0.8%	5.5%	0.062
		35%	19.8	17.9	19.1	16.8	17.8	18.9	16.3	19.5	17.8	18.2	-0.8%	5.5%	0.062
		40%	19.8	17.9	19.1	16.8	17.8	18.9	16.3	19.5	17.8	18.2	-0.8%	5.5%	0.062
		50%	19.8	17.9	19.1	16.8	17.8	18.9	16.3	19.5	17.8	18.2	-0.8%	5.5%	0.062
		2x	19.8	17.9	19.1	16.8	17.8	18.9	16.3	19.5	17.8	18.2	-0.8%	5.5%	0.062
		Unlimited	19.8	17.9	19.1	16.8	17.8	18.9	16.3	19.5	17.8	18.2	-0.8%	5.5%	0.062

Table 3-12: Daily Impacts and Errors by Baseline Type, SCE DRRC, Day-of

Application of Baselines	Adjustment Option	Adjustment Cap	Impact (MW) by Proxy Event Day									Avg	Bias	Goodness-of-fit	
			1	2	3	4	5	6	7	8	9		MPE	MAPE	CV RMSE
Simulated Impact			136.7												
Individual	**Aggregator Choice**	**Unadjusted**	89.8	110.6	107.6	120.2	104.3	118.4	106.4	128.4	106.1	110.2	-16.9%	16.9%	0.187
		20%	128.2	124.3	132.0	135.1	128.1	129.3	131.0	127.0	125.1	128.9	-2.8%	3.3%	0.037
		30%	131.2	124.7	132.7	134.9	128.9	129.0	134.2	127.4	125.2	129.8	-2.1%	2.8%	0.034
		35%	131.8	124.8	133.1	134.8	129.0	129.1	134.7	127.5	124.9	130.0	-2.0%	2.8%	0.034
		40%	132.1	124.9	133.3	134.8	128.9	129.1	134.8	127.6	124.7	130.0	-2.0%	2.8%	0.034
		50%	133.1	125.0	133.7	134.8	128.7	129.1	134.8	127.5	124.2	130.1	-1.9%	2.9%	0.035
		2x	136.2	128.6	137.5	137.4	132.0	130.5	137.4	129.4	125.9	132.8	0.1%	2.9%	0.032
		Unlimited	141.2	128.4	139.3	141.6	133.1	129.1	150.4	130.4	130.7	136.0	2.5%	4.6%	0.059
	Universal	**Unadjusted**	89.8	110.6	107.6	120.2	104.3	118.4	106.4	128.4	106.1	110.2	-16.9%	16.9%	0.187
		20%	128.2	124.3	132.0	135.1	128.1	129.3	131.0	127.0	125.1	128.9	-2.8%	3.3%	0.037
		30%	131.2	124.7	132.7	134.9	128.9	129.0	134.2	127.4	125.2	129.8	-2.1%	2.8%	0.034
		35%	131.8	124.8	133.1	134.8	129.0	129.1	134.7	127.5	124.9	130.0	-2.0%	2.8%	0.034
		40%	132.1	124.9	133.3	134.8	128.9	129.1	134.8	127.6	124.7	130.0	-2.0%	2.8%	0.034
		50%	133.1	125.0	133.7	134.8	128.7	129.1	134.8	127.5	124.2	130.1	-1.9%	2.9%	0.035
		2x	136.2	128.6	137.5	137.4	132.0	130.5	137.4	129.4	125.9	132.8	0.1%	2.9%	0.032
		Unlimited	141.2	128.4	139.3	141.6	133.1	129.1	150.4	130.4	130.7	136.0	2.5%	4.6%	0.059
Aggregate	**Aggregator Choice**	**Unadjusted**	89.8	110.6	107.6	120.2	104.3	118.4	106.4	128.4	106.1	110.2	-16.9%	16.9%	0.187
		20%	134.1	128.6	136.5	137.7	127.8	128.6	134.5	129.1	126.9	131.5	-0.9%	2.9%	0.031
		30%	134.1	128.6	136.5	137.7	127.8	128.6	134.5	129.1	126.9	131.5	-0.9%	2.9%	0.031
		35%	134.1	128.6	136.5	137.7	127.8	128.6	134.5	129.1	126.9	131.5	-0.9%	2.9%	0.031
		40%	134.1	128.6	136.5	137.7	127.8	128.6	134.5	129.1	126.9	131.5	-0.9%	2.9%	0.031
		50%	134.1	128.6	136.5	137.7	127.8	128.6	134.5	129.1	126.9	131.5	-0.9%	2.9%	0.031
		2x	134.1	128.6	136.5	137.7	127.8	128.6	134.5	129.1	126.9	131.5	-0.9%	2.9%	0.031
		Unlimited	134.1	128.6	136.5	137.7	127.8	128.6	134.5	129.1	126.9	131.5	-0.9%	2.9%	0.031
	Universal	**Unadjusted**	89.8	110.6	107.6	120.2	104.3	118.4	106.4	128.4	106.1	110.2	-16.9%	16.9%	0.187
		20%	134.1	128.6	136.5	137.7	127.8	128.6	134.5	129.1	126.9	131.5	-0.9%	2.9%	0.031
		30%	134.1	128.6	136.5	137.7	127.8	128.6	134.5	129.1	126.9	131.5	-0.9%	2.9%	0.031
		35%	134.1	128.6	136.5	137.7	127.8	128.6	134.5	129.1	126.9	131.5	-0.9%	2.9%	0.031
		40%	134.1	128.6	136.5	137.7	127.8	128.6	134.5	129.1	126.9	131.5	-0.9%	2.9%	0.031
		50%	134.1	128.6	136.5	137.7	127.8	128.6	134.5	129.1	126.9	131.5	-0.9%	2.9%	0.031
		2x	134.1	128.6	136.5	137.7	127.8	128.6	134.5	129.1	126.9	131.5	-0.9%	2.9%	0.031
		Unlimited	134.1	128.6	136.5	137.7	127.8	128.6	134.5	129.1	126.9	131.5	-0.9%	2.9%	0.031

3.3.2 Detailed Results for SCE's CBP Program

Table 3-13 compares estimated impacts for each proxy event day with the simulated reductions by baseline method for SCE's CBP day-ahead resource. Bias is reduced by loosening the adjustment cap, applying the adjustment on the aggregate level and making the adjustment universal. While bias is reduced by loosening the adjustment cap on the individual level, note that goodness of fit does not improve very much when the adjustment cap is loosened. Note also that there is considerable volatility across proxy event days; the baseline performs much better on certain days than on others.

Table 3-14 summarizes the baseline assessment for SCE's CBP day-of resource. Bias and goodness-of-fit are slightly better for the day-of program than for the day-ahead program. Otherwise, the results for each program are very similar.

Table 3-13: Daily Impacts and Errors by Baseline Type, SCE CBP, Day-ahead

Application of Baselines	Adjustment Option	Adjustment Cap	Impact (MW) by Proxy Event Day									Avg	Bias	Goodness-of-fit	
			1	2	3	4	5	6	7	8	9		MPE	MAPE	CV RMSE
Simulated Impact			4.7												
Individual	**Aggregator Choice**	**Unadjusted**	4.0	3.2	3.7	4.2	4.0	4.3	1.6	1.6	2.0	3.2	-32.4%	32.4%	0.394
		20%	5.0	4.5	3.8	3.9	4.9	5.0	2.9	3.1	3.0	4.0	-15.0%	18.4%	0.231
		30%	5.1	4.5	3.8	3.9	4.9	5.0	3.0	3.2	3.2	4.1	-13.7%	17.9%	0.217
		35%	5.1	4.5	3.8	3.9	4.9	5.0	3.0	3.3	3.2	4.1	-13.3%	17.6%	0.212
		40%	5.1	4.6	3.8	3.9	4.9	5.0	3.0	3.4	3.2	4.1	-13.1%	17.4%	0.210
		50%	5.1	4.6	3.7	3.9	5.0	5.0	3.0	3.4	3.3	4.1	-12.8%	17.2%	0.206
		2x	5.1	4.6	3.7	3.9	5.0	5.0	3.1	3.5	3.3	4.1	-12.3%	16.7%	0.200
		Unlimited	5.1	4.6	3.7	3.8	5.0	5.0	3.1	3.6	3.4	4.1	-12.2%	16.6%	0.199
	Universal	**Unadjusted**	4.0	3.2	3.7	4.2	4.0	4.3	1.6	1.6	2.0	3.2	-32.4%	32.4%	0.394
		20%	4.9	4.7	3.9	3.9	5.1	5.1	2.7	3.1	3.1	4.1	-14.0%	18.5%	0.232
		30%	4.9	4.7	3.8	3.9	5.1	5.1	2.9	3.3	3.2	4.1	-12.7%	17.7%	0.218
		35%	4.9	4.8	3.8	3.9	5.1	5.1	2.9	3.3	3.3	4.1	-12.4%	17.4%	0.213
		40%	4.9	4.8	3.8	3.9	5.2	5.1	2.9	3.4	3.3	4.1	-12.2%	17.3%	0.210
		50%	4.8	4.9	3.8	3.9	5.2	5.1	2.9	3.5	3.4	4.2	-11.9%	17.1%	0.207
		2x	4.9	5.0	3.8	3.9	5.2	5.1	2.9	3.6	3.4	4.2	-11.0%	17.1%	0.202
		Unlimited	4.8	5.0	3.8	3.9	5.2	5.1	2.9	3.6	3.4	4.2	-11.1%	16.8%	0.201
Aggregate	**Aggregator Choice**	**Unadjusted**	4.0	3.2	3.7	4.2	4.0	4.3	1.6	1.6	2.0	3.2	-32.4%	32.4%	0.394
		20%	5.1	4.6	3.8	3.9	4.9	5.0	3.0	3.4	3.2	4.1	-13.3%	17.2%	0.210
		30%	5.1	4.6	3.8	3.9	4.9	5.0	3.0	3.4	3.2	4.1	-13.3%	17.2%	0.210
		35%	5.1	4.6	3.8	3.9	4.9	5.0	3.0	3.4	3.2	4.1	-13.3%	17.2%	0.210
		40%	5.1	4.6	3.8	3.9	4.9	5.0	3.0	3.4	3.2	4.1	-13.3%	17.2%	0.210
		50%	5.1	4.6	3.8	3.9	4.9	5.0	3.0	3.4	3.2	4.1	-13.3%	17.2%	0.210
		2x	5.1	4.6	3.8	3.9	4.9	5.0	3.0	3.4	3.2	4.1	-13.3%	17.2%	0.210
		Unlimited	5.1	4.6	3.8	3.9	4.9	5.0	3.0	3.4	3.2	4.1	-13.3%	17.2%	0.210
	Universal	**Unadjusted**	4.0	3.2	3.7	4.2	4.0	4.3	1.6	1.6	2.0	3.2	-32.4%	32.4%	0.394
		20%	4.9	4.7	3.8	3.9	5.1	5.1	2.9	3.5	3.3	4.1	-12.3%	17.3%	0.211
		30%	4.9	4.7	3.8	3.9	5.1	5.1	2.9	3.5	3.3	4.1	-12.3%	17.1%	0.211
		35%	4.9	4.8	3.8	3.9	5.1	5.1	2.9	3.5	3.3	4.1	-12.3%	17.1%	0.211
		40%	4.8	4.8	3.8	3.9	5.1	5.1	2.9	3.5	3.3	4.1	-12.3%	17.1%	0.211
		50%	4.8	4.9	3.8	3.9	5.1	5.1	2.9	3.5	3.3	4.1	-12.3%	17.1%	0.211
		2x	4.8	5.0	3.8	3.9	5.1	5.1	2.9	3.5	3.3	4.2	-12.0%	17.4%	0.211
		Unlimited	4.7	5.0	3.8	3.9	5.1	5.1	2.9	3.5	3.3	4.1	-12.2%	17.4%	0.211

Table 3-14: Daily Impacts and Errors by Baseline Type, SCE CBP Day-of

Application of Baselines	Adjustment Option	Adjustment Cap	Impact (MW) by Proxy Event Day									Avg	Bias	Goodness-of-fit	
			1	2	3	4	5	6	7	8	9		MPE	MAPE	CV RMSE
Simulated Impact			15.0												
Individual	**Aggregator Choice**	**Unadjusted**	8.7	9.4	11.8	13.7	8.7	13.1	12.5	13.1	8.5	11.0	-26.4%	26.4%	0.297
		20%	13.1	13.3	13.4	14.3	11.7	14.7	13.3	14.3	11.7	13.3	-11.3%	11.3%	0.131
		30%	13.2	13.4	13.3	14.3	11.7	14.7	13.3	14.3	11.8	13.3	-11.1%	11.1%	0.129
		35%	13.2	13.4	13.3	14.3	11.7	14.7	13.2	14.3	11.8	13.3	-11.1%	11.1%	0.129
		40%	13.2	13.4	13.3	14.3	11.7	14.7	13.2	14.3	11.8	13.3	-11.1%	11.1%	0.129
		50%	13.2	13.4	13.3	14.3	11.7	14.7	13.2	14.3	11.8	13.3	-11.1%	11.1%	0.129
		2x	13.2	13.4	13.3	14.3	11.7	14.7	13.2	14.3	11.8	13.3	-11.1%	11.1%	0.129
		Unlimited	13.2	13.4	13.3	14.3	11.7	14.7	13.1	14.3	11.7	13.3	-11.3%	11.3%	0.131
	Universal	**Unadjusted**	8.7	9.4	11.8	13.7	8.7	13.1	12.5	13.1	8.5	11.0	-26.4%	26.4%	0.297
		20%	16.8	16.8	15.3	15.4	15.1	15.9	14.4	15.3	15.5	15.6	4.0%	5.0%	0.064
		30%	17.2	17.0	15.2	15.3	15.3	15.9	14.3	15.3	15.7	15.7	4.7%	5.7%	0.074
		35%	17.3	17.0	15.2	15.3	15.3	15.9	14.3	15.3	15.7	15.7	4.7%	5.7%	0.075
		40%	17.3	17.0	15.2	15.3	15.3	15.9	14.3	15.3	15.7	15.7	4.7%	5.8%	0.075
		50%	17.3	17.0	15.2	15.3	15.3	15.8	14.3	15.3	15.7	15.7	4.7%	5.8%	0.076
		2x	17.3	17.0	15.2	15.3	15.3	15.8	14.3	15.3	15.7	15.7	4.7%	5.8%	0.076
		Unlimited	17.3	17.1	15.2	15.3	15.3	15.8	14.1	15.3	15.6	15.7	4.5%	5.7%	0.076
Aggregate	**Aggregator Choice**	**Unadjusted**	8.7	9.4	11.8	13.7	8.7	13.1	12.5	13.1	8.5	11.0	-26.4%	26.4%	0.297
		20%	13.1	13.3	13.3	14.3	11.7	14.7	13.1	14.3	11.5	13.2	-11.7%	11.7%	0.136
		30%	13.1	13.3	13.3	14.3	11.7	14.7	13.1	14.3	11.5	13.2	-11.7%	11.7%	0.136
		35%	13.1	13.3	13.3	14.3	11.7	14.7	13.1	14.3	11.5	13.2	-11.7%	11.7%	0.136
		40%	13.1	13.3	13.3	14.3	11.7	14.7	13.1	14.3	11.5	13.2	-11.7%	11.7%	0.136
		50%	13.1	13.3	13.3	14.3	11.7	14.7	13.1	14.3	11.5	13.2	-11.7%	11.7%	0.136
		2x	13.1	13.3	13.3	14.3	11.7	14.7	13.1	14.3	11.5	13.2	-11.7%	11.7%	0.136
		Unlimited	13.1	13.3	13.3	14.3	11.7	14.7	13.1	14.3	11.5	13.2	-11.7%	11.7%	0.136
	Universal	**Unadjusted**	8.7	9.4	11.8	13.7	8.7	13.1	12.5	13.1	8.5	11.0	-26.4%	26.4%	0.297
		20%	17.2	16.9	15.2	15.3	15.2	15.8	14.1	15.3	15.4	15.6	4.1%	5.4%	0.072
		30%	17.2	16.9	15.1	15.2	15.2	15.8	14.1	15.3	15.4	15.6	4.0%	5.3%	0.072
		35%	17.2	16.9	15.1	15.2	15.2	15.8	14.1	15.3	15.4	15.6	4.0%	5.3%	0.072
		40%	17.2	16.9	15.1	15.2	15.2	15.8	14.1	15.3	15.4	15.6	4.0%	5.3%	0.072
		50%	17.2	16.9	15.1	15.2	15.2	15.8	14.1	15.3	15.4	15.6	4.0%	5.3%	0.072
		2x	17.2	16.9	15.1	15.2	15.2	15.8	14.1	15.3	15.4	15.6	4.0%	5.3%	0.072
		Unlimited	17.2	16.9	15.1	15.2	15.2	15.8	14.1	15.3	15.4	15.6	4.0%	5.3%	0.072

3.4 SDG&E Settlement Baseline Analysis

Figure 3-5 shows the extent to which each baseline method underestimates or overestimates demand reductions for SDG&E's CBP program for both the day-of and day-ahead resources. Table 3-15 shows the same information in tabular form. Loosening the adjustment cap reduces bias for both resource options. For the day-ahead customers the unadjusted baseline has a negative bias and underestimates reductions by nearly 40%. Applying adjustments on a universal basis decreases the bias substantially. However, loosening the cap too much leads to overestimation of the reductions. Making the adjustment mandatory decreases bias for both resource options. While aggregators and direct participant can elect to apply same day adjustments (in advance), several choose not to apply the adjustment. It is clearly in the interest of aggregators to elect to apply the adjustment since it reduces the amount of load reduction underestimation.

Figure 3-5: Bias by Baseline Type SDG&E's CBP Program

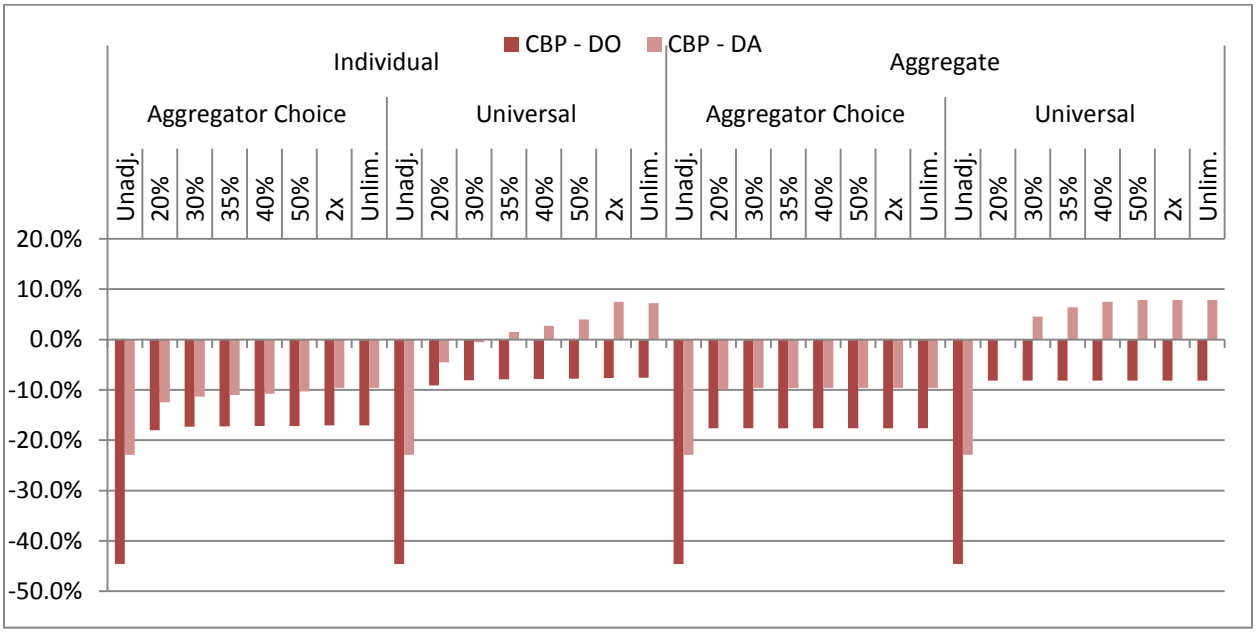


Table 3-15: Bias by Baseline Type for SDG&E's CBP Program

Application of Baselines	Adjustment Option	Adjustment Cap	Capacity Bidding Program			
			Day-of		Day-ahead	
			Estimated Impacts	Error (MPE)	Estimated Impacts	Error (MPE)
Simulated Impact			8.7		6.2	
Individual	Aggregator Choice	Unadjusted	4.9	-44.5%	6.2	-22.9%
		20%	7.2	-18.0%	7.1	-12.5%
		30%	7.2	-17.3%	7.2	-11.3%
		35%	7.2	-17.2%	7.2	-11.0%
		40%	7.2	-17.2%	7.2	-10.7%
		50%	7.2	-17.2%	7.3	-10.3%
		2x	7.3	-17.1%	7.3	-9.6%
		Unlimited	7.3	-17.0%	7.3	-9.6%
	Universal	Unadjusted	4.9	-44.5%	6.2	-22.9%
		20%	8.0	-9.1%	7.7	-4.6%
		30%	8.0	-8.1%	8.1	-0.5%
		35%	8.1	-7.9%	8.2	1.5%
		40%	8.1	-7.8%	8.3	2.7%
		50%	8.1	-7.8%	8.4	4.0%
		2x	8.1	-7.6%	8.7	7.5%
		Unlimited	8.1	-7.6%	8.7	7.2%
Aggregate	Aggregator Choice	Unadjusted	4.9	-44.5%	6.2	-22.9%
		20%	7.2	-17.6%	7.3	-9.8%
		30%	7.2	-17.6%	7.3	-9.6%
		35%	7.2	-17.6%	7.3	-9.6%
		40%	7.2	-17.6%	7.3	-9.6%
		50%	7.2	-17.6%	7.3	-9.6%
		2x	7.2	-17.6%	7.3	-9.6%
		Unlimited	7.2	-17.6%	7.3	-9.6%
	Universal	Unadjusted	4.9	-44.5%	6.2	-22.9%
		20%	8.0	-8.1%	8.1	0.0%
		30%	8.0	-8.1%	8.5	4.6%
		35%	8.0	-8.1%	8.6	6.4%
		40%	8.0	-8.1%	8.7	7.5%
		50%	8.0	-8.1%	8.7	7.9%
		2x	8.0	-8.1%	8.7	7.9%
		Unlimited	8.0	-8.1%	8.7	7.9%

3.4.1 Detailed Results for SDG&E's CBP Program

Table 3-16 compares the estimated reductions for each proxy event day with the simulated impacts by baseline method for SDG&E's CBP day-ahead resource. Bias is reduced by loosening the adjustment cap, by applying the adjustment on the aggregate level, and by making the adjustment universal. While bias is reduced by loosening the adjustment cap on the individual level, goodness of fit does not improve very much when the adjustment cap is loosened. There is considerable volatility between proxy event days; the baseline performs much better on certain days than it does on others.

Table 3-17 shows the assessment for SDG&E's CBP day-of resource. CBP day-of baselines are most improved by making the adjustment mandatory because most settlement portfolios do not already apply the same-day adjustment. Otherwise, loosening the adjustment cap has limited effects because the adjustment ratio already falls below successively higher adjustment caps.

Table 3-16: Daily Impacts and Errors by Baseline Type, SDG&E CBP Day-ahead

Application of Baselines	Adjustment Option	Adjustment Cap	Impact (MW) by Proxy Event Day							Avg.	Bias	Goodness-of-fit	
			1	2	3	4	5	6	7		MPE	MAPE	CV RMSE
Simulated Impact			8.1										
Individual	**Aggregator Choice**	**Unadjusted**	5.6	11.0	5.5	5.1	6.3	7.5	2.7	6.2	-22.9%	33.0%	0.370
		20%	6.5	11.3	6.4	6.0	6.9	8.8	3.7	7.1	-12.5%	26.3%	0.300
		30%	6.6	11.3	6.5	6.0	7.0	9.0	3.8	7.2	-11.3%	25.9%	0.294
		35%	6.7	11.3	6.5	6.1	7.0	9.1	3.9	7.2	-11.0%	25.7%	0.291
		40%	6.7	11.2	6.5	6.1	7.0	9.1	3.9	7.2	-10.7%	25.5%	0.288
		50%	6.7	11.2	6.6	6.1	7.0	9.2	4.0	7.3	-10.3%	25.3%	0.284
		2x	6.7	11.2	6.7	6.1	7.0	9.5	4.1	7.3	-9.6%	25.5%	0.284
		Unlimited	6.7	11.2	6.7	6.1	7.0	9.5	4.1	7.3	-9.6%	25.5%	0.284
	Universal	**Unadjusted**	5.6	11.0	5.5	5.1	6.3	7.5	2.7	6.2	-22.9%	33.0%	0.370
		20%	8.1	10.5	7.8	6.0	8.3	8.2	5.3	7.7	-4.6%	14.0%	0.198
		30%	8.5	9.8	8.1	6.3	9.0	8.4	6.2	8.1	-0.5%	12.5%	0.154
		35%	8.6	9.5	8.3	6.5	9.4	8.6	6.7	8.2	1.5%	12.3%	0.139
		40%	8.8	9.2	8.5	6.5	9.6	8.8	6.9	8.3	2.7%	12.6%	0.136
		50%	9.1	8.5	8.8	6.5	9.6	9.2	7.3	8.4	4.0%	12.6%	0.134
		2x	9.2	8.5	9.2	6.5	9.6	9.7	8.2	8.7	7.5%	13.1%	0.148
		Unlimited	9.2	8.4	9.2	6.5	9.6	9.7	8.2	8.7	7.2%	12.8%	0.147
Aggregate	**Aggregator Choice**	**Unadjusted**	5.6	11.0	5.5	5.1	6.3	7.5	2.7	6.2	-22.9%	33.0%	0.370
		20%	6.7	11.2	6.7	6.1	7.0	9.4	4.1	7.3	-9.8%	25.3%	0.283
		30%	6.7	11.2	6.7	6.1	7.0	9.5	4.1	7.3	-9.6%	25.6%	0.284
		35%	6.7	11.2	6.7	6.1	7.0	9.5	4.1	7.3	-9.6%	25.6%	0.284
		40%	6.7	11.2	6.7	6.1	7.0	9.5	4.1	7.3	-9.6%	25.6%	0.284
		50%	6.7	11.2	6.7	6.1	7.0	9.5	4.1	7.3	-9.6%	25.6%	0.284
		2x	6.7	11.2	6.7	6.1	7.0	9.5	4.1	7.3	-9.6%	25.6%	0.284
		Unlimited	6.7	11.2	6.7	6.1	7.0	9.5	4.1	7.3	-9.6%	25.6%	0.284
	Universal	**Unadjusted**	5.6	11.0	5.5	5.1	6.3	7.5	2.7	6.2	-22.9%	33.0%	0.370
		20%	8.6	9.4	8.2	7.0	8.5	9.2	5.8	8.1	0.0%	11.9%	0.144
		30%	9.3	8.7	9.0	7.0	9.3	9.3	6.7	8.5	4.6%	13.4%	0.137
		35%	9.3	8.7	9.4	7.0	9.5	9.3	7.2	8.6	6.4%	13.6%	0.139
		40%	9.3	8.7	9.6	7.0	9.5	9.3	7.6	8.7	7.5%	13.1%	0.138
		50%	9.3	8.7	9.6	7.0	9.5	9.3	7.8	8.7	7.9%	12.7%	0.137
		2x	9.3	8.7	9.6	7.0	9.5	9.3	7.8	8.7	7.9%	12.7%	0.137
		Unlimited	9.3	8.7	9.6	7.0	9.5	9.3	7.8	8.7	7.9%	12.7%	0.137

Table 3-17: Daily Impacts and Errors by Baseline Type, SDG&E CBP, Day-of

Application of Baselines	Adjustment Option	Adjustment Cap	Impact (MW) by Proxy Event Day							Avg	Bias	Goodness-of-fit	
			1	2	3	4	5	6	7		MPE	MAPE	CV RMSE
Simulated Impact			8.7										
Individual	**Aggregator Choice**	**Unadjusted**	2.9	4.2	5.8	5.7	6.5	6.1	2.8	4.9	-44.5%	44.5%	0.475
		20%	6.1	6.9	7.9	8.4	7.7	8.0	5.2	7.2	-18.0%	18.0%	0.219
		30%	6.2	7.0	7.9	8.4	7.7	8.0	5.3	7.2	-17.3%	17.3%	0.211
		35%	6.2	7.1	7.9	8.4	7.7	8.0	5.3	7.2	-17.2%	17.2%	0.210
		40%	6.3	7.1	7.9	8.4	7.7	8.0	5.3	7.2	-17.2%	17.2%	0.210
		50%	6.3	7.1	7.9	8.4	7.7	8.0	5.2	7.2	-17.2%	17.2%	0.210
		2x	6.3	7.1	7.9	8.4	7.7	8.0	5.3	7.3	-17.1%	17.1%	0.208
		Unlimited	6.3	7.1	7.9	8.4	7.7	8.0	5.3	7.3	-17.0%	17.0%	0.207
	Universal	**Unadjusted**	2.9	4.2	5.8	5.7	6.5	6.1	2.8	4.9	-44.5%	44.5%	0.475
		20%	7.2	7.8	8.5	9.3	8.3	8.7	5.9	8.0	-9.1%	10.8%	0.150
		30%	7.5	7.9	8.6	9.3	8.3	8.7	6.0	8.0	-8.1%	10.0%	0.140
		35%	7.5	7.9	8.6	9.3	8.3	8.8	6.0	8.1	-7.9%	9.8%	0.139
		40%	7.6	7.9	8.6	9.3	8.3	8.8	6.0	8.1	-7.8%	9.7%	0.138
		50%	7.6	7.9	8.6	9.3	8.3	8.8	6.0	8.1	-7.8%	9.7%	0.138
		2x	7.6	7.9	8.6	9.3	8.3	8.8	6.1	8.1	-7.6%	9.6%	0.135
		Unlimited	7.6	7.9	8.6	9.3	8.3	8.8	6.1	8.1	-7.6%	9.5%	0.133
Aggregate	**Aggregator Choice**	**Unadjusted**	2.9	4.2	5.8	5.7	6.5	6.1	2.8	4.9	-44.5%	44.5%	0.475
		20%	6.3	7.0	7.9	8.4	7.7	8.0	5.2	7.2	-17.6%	17.6%	0.213
		30%	6.3	7.0	7.9	8.4	7.7	8.0	5.2	7.2	-17.6%	17.6%	0.213
		35%	6.3	7.0	7.9	8.4	7.7	8.0	5.2	7.2	-17.6%	17.6%	0.213
		40%	6.3	7.0	7.9	8.4	7.7	8.0	5.2	7.2	-17.6%	17.6%	0.213
		50%	6.3	7.0	7.9	8.4	7.7	8.0	5.2	7.2	-17.6%	17.6%	0.213
		2x	6.3	7.0	7.9	8.4	7.7	8.0	5.2	7.2	-17.6%	17.6%	0.213
		Unlimited	6.3	7.0	7.9	8.4	7.7	8.0	5.2	7.2	-17.6%	17.6%	0.213
	Universal	**Unadjusted**	2.9	4.2	5.8	5.7	6.5	6.1	2.8	4.9	-44.5%	44.5%	0.475
		20%	7.5	7.9	8.5	9.3	8.3	8.7	6.0	8.0	-8.1%	9.9%	0.139
		30%	7.5	7.9	8.5	9.3	8.3	8.7	6.0	8.0	-8.1%	9.9%	0.139
		35%	7.5	7.9	8.5	9.3	8.3	8.7	6.0	8.0	-8.1%	9.9%	0.139
		40%	7.5	7.9	8.5	9.3	8.3	8.7	6.0	8.0	-8.1%	9.9%	0.139
		50%	7.5	7.9	8.5	9.3	8.3	8.7	6.0	8.0	-8.1%	9.9%	0.139
		2x	7.5	7.9	8.5	9.3	8.3	8.7	6.0	8.0	-8.1%	9.9%	0.139
		Unlimited	7.5	7.9	8.5	9.3	8.3	8.7	6.0	8.0	-8.1%	9.9%	0.139

4 Recommendations

The analysis in this report tested the effect of various potential modifications to the settlement baseline rules, including calculating baseline in aggregate rather individually, universal application of same-day adjustments, and various same-day adjustment caps.

Based on the evidence, we recommend the following:

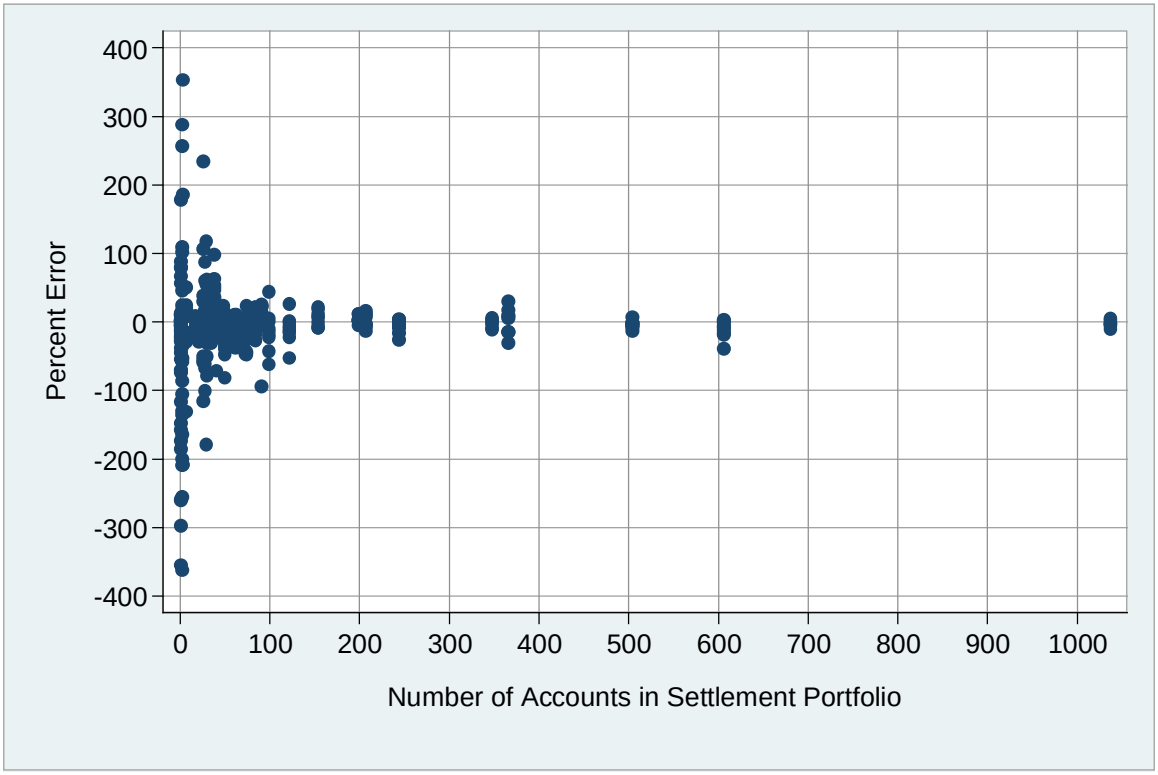
- *Make same day adjustments the default option for settlement baselines.* Currently, aggregators can elect to apply same-day adjustments but must take proactive action to do so. Making the same-day adjustment the default ensures aggregators retain an element of choice, but at the same time guides them towards the option that is known to produce the most accurate results.
- *Increase the adjustment cap to ±30%.* Increasing the same day adjustment cap generally improves the accuracy of the results, though for some program options, it makes no or little difference. When it does affects accuracy, a higher adjustment cap typically leads to more accurate results. However, we advise against removing the adjustment cap altogether it can cause extreme adjustments for some customers that are large enough to affect the overall program results. We do not recommend higher adjustment caps of ±40% or ±50% for two reasons: relatively few accounts require large adjustments; and customers that require large adjustment are less predictable and instead should be encouraged to enroll in DR programs that do not rely on a baseline for settlement.

Improving settlement accuracy is not just a function of the settlement baseline rules but of the degree to which aggregators aggregate and manage their portfolios. The higher the number of accounts in a portfolio, the less volatile and more predictable the loads become. It is much easier to estimate accurate baselines for large, aggregated groups of customers than it to accurately estimate baselines for individual customers or small settlement portfolios. Likewise, baselines are more accurate for portfolios that deliver larger percent demand reductions. No single change in baseline rules improves accuracy for settlement portfolios with a small number of customers and/or small percent reductions. Aggregators can improve the accuracy of the impact estimates used for settlement by applying same day adjustments universally, aggregating across more accounts and procuring customers who reduce a large share of their loads.

Appendix A. Other Factors Influencing Impact Error

In addition to choosing a different baseline methodology, two other factors can help to reduce impact error. All else being equal, settlement portfolios with many customers tend to have smaller errors than settlement portfolios with fewer customers. This intuition is confirmed in Figure A-1. The figure shows the percent error for each proxy event day and settlement portfolio at each of the three utilities. As discussed earlier, to estimate accuracy, the correct answer needs to be known. This accomplished though simulating impacts – based on the demand reductions observed for each customer during historical events – during hot non-event days. As can be seen, the errors are more pronounced for settlement portfolios with fewer customers and more accurate for those with more customers.

Figure A-1: Individual Event Day Error by Size of Settlement Portfolios

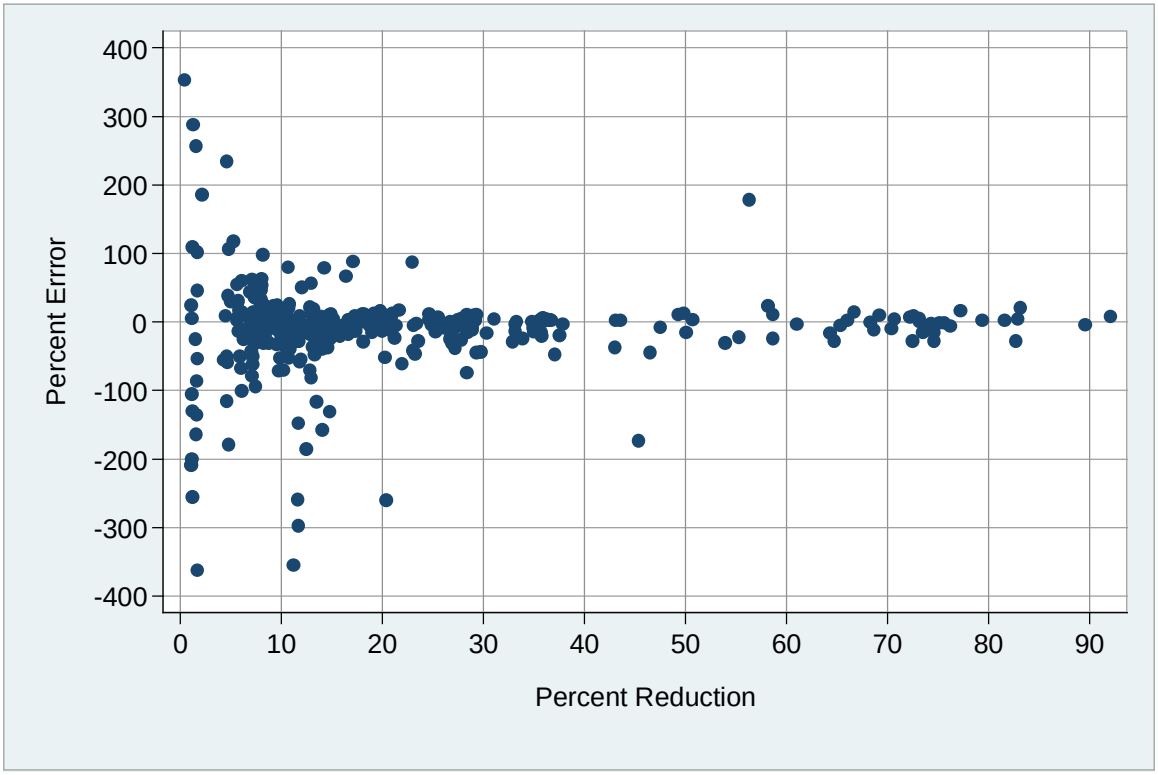


A second factor that helps to reduce impact error is the magnitude of the impacts. All else equal, a settlement portfolio that delivers a larger percent demand reduction has lower errors than a settlement portfolio that delivers a smaller percent reduction. This relationship is clearly expressed by the following equation:

$$\%\ Impact\ Error = \frac{\%\ Baseline\ Error}{\%\ True\ Load\ Reduction}$$

Increasing the load reduction relative to the counterfactual reduces impact error. Figure A-2 shows empirical evidence of this. Again, each dot represents the percent error for each settlement portfolio on each proxy event day. The figure summarizes all settlement portfolios across all three utilities.

Figure A-2: Individual Event Day Error by Percent Demand Reduction



While choosing a different baseline methodology may help to reduce event impact errors, encouraging aggregators to increase the number of customers per settlement portfolio and percent load reductions can also reduce impact errors considerably.