CHRISTENSEN
A S S O C I A T E S
ENERGY CONSULTING

2012 Statewide Load Impact Evaluation of California Aggregator Demand Response Programs Volume 2: Baseline Analysis

CALMAC Study ID PGE0318.02

Steven D. Braithwait
Daniel G. Hansen
David A. Armstrong

April 1, 2013

Table of Contents

Tables

EXECUTIVE SUMMARY

This second volume of the aggregator evaluation report for program-year 2012 contains an analysis of the performance of the current aggregator program baseline methods and of a number of potential alternative baseline methods. The baseline analysis was ordered in the CPUC decision regarding DR activities and budgets for 2012 through 2014.

Settlement baselines are critical components of demand response programs such as the aggregator programs, as they serve as the reference point for measuring program load reductions for which customers and aggregators receive credit payments, and utilities receive resource credits. The baseline analysis in this study focuses on differences in results under the following conditions:

1. Using aggregations of *individual customer* baselines, compared to baselines that are constructed from *aggregated* loads across customers in a relevant portfolio (*e.g.,* all of the customer accounts nominated in one aggregator's CBP DO notice portfolio for a given month);
2. Using a range of caps on day-of percentage *baseline adjustments* (*e.g.,* cap percentages of 20, 30, 40, and 50, plus unlimited adjustments); and
3. Comparing results across only those customers who actually selected the day-of baseline adjustment in 2012 to a case where baselines are adjusted for *all* enrolled and nominated customers (*i.e.,* assuming that all customers in a portfolio selected the adjustment).

The baseline analysis was conducted for both *actual* event days in 2012, as well as a set of event-like non-event days, or *simulated* events, in July, August, and September. In the case of *actual events*, the alternative baselines were compared to the baseline loads implied by the customer-level regression analyses conducted in the 2012 ex post load impact evaluation (*i.e.,* estimated load impacts are added back in to the observed event-day loads to create a "but for the event" reference load). In the case of *simulated* events, the observed loads on the event-like days serve as "true" baselines, which are then compared to all of the relevant alternative baseline methods.

ES.1 Baseline Performance Metrics

The performance of a baseline method is generally measured by how *accurate* the method is, and whether the baseline method tends to be *biased* upward or downward. The accuracy metric used in this study is the average of the *absolute values* of the percent errors over the relevant observations, such as customers and events, or Mean Absolute Percentage Error (MAPE). To measure bias, this study reports mean, median, and 10, 25, 50, 75, and 90 percentile values of percent errors, where the 50th percentile value is the median.

ES.2 Summary of Study Findings

The primary conclusions of this baseline study are the following:

- The accuracy and bias measures for the aggregated baseline methods can vary substantially across utilities, programs, and notice types, suggesting that baseline errors depend on a number of factors other than baseline type (*e.g.*, adjusted 10-in-10). These likely include the characteristics of the particular customers who participate in the programs and the nature of the events.
- The portfolio baselines, which are calculated on the basis of aggregations of the loads of individual customer accounts into their portfolio category (*e.g.*, aggregator and notice type) are often somewhat more accurate and have smaller biases than do aggregations of individual customer baselines. However, in cases where the aggregated baseline errors are relatively large and numerous, such as PG&E and SDG&E CBP DA, the portfolio baselines do not improve the results much, at least without baseline adjustments. Of some note, even in cases of large average percent errors, the typical bias as measured by the median percent error was quite small; however, the range of errors could be quite large.
- Allowing some day-of adjustment to the 10-in-10 baseline often improves accuracy and reduces bias somewhat, but unlimited adjustments often cause baseline errors to increase relative to even a 50 percent cap. However, in this study, there are few clear patterns of the degree of improvement in performance under different adjustment cap restrictions. For many of the programs, caps above 20 or 30 percent were not binding.

CA Energy Consulting

1.   INTRODUCTION AND OBJECTIVES OF THE STUDY

This second volume of the aggregator evaluation report for program-year 2012 contains an analysis of the performance of the current aggregator program baseline methods and of a number of potential alternative baseline methods. The baseline analysis was ordered in the CPUC decision regarding DR activities and budgets for 2012 through 2014, which included the following order on the topic of baselines:

"[The utilities] shall provide, as part of the Load Impact Annual Filing on June 1, 2012 and again on April 1, 2013 and 2014, an analysis that compares their baseline settlement result using both individual and aggregated baseline with cap percentage adjustments of 20, 30, 40, 50 and no cap for the months of July, August, and September of the prior year. The Utilities shall compare the annual baseline settlement results with the Measurement and Evaluation results for the same year. The comparison analysis must include service accounts for which the adjusted energy baseline option was selected in that nomination month as well as a second set of service accounts, assuming all service accounts select day-of adjustment."

CA Energy Consulting has conducted a comprehensive baseline analysis for each aggregator program and product type (*e.g.,* day-ahead and day-of), including a range of alternative baseline methods. Our understanding is that the current baseline methods for the aggregator programs are the following:
- CBP – The program baseline is the sum of individual customer baselines within a product portfolio (*e.g.,* DA 1-4), which are 10-in-10 baselines. Day-of adjustment (first 3 of 4 hours prior to event) is at the individual customer level, is optional, and is limited to 40%. Adjustments are for the window HE 12-19.
- AMP and DRC – Program baselines are agreed upon within the aggregator contracts, and are set at the beginning of the summer.

The baseline analysis in this study focuses on differences in results under the following conditions:
4. Using aggregations of *individual customer* baselines, compared to baselines that are constructed from *aggregated* loads across customers in a relevant portfolio (*e.g.,* all of the customer accounts nominated in one aggregator's CBP DO 1-4 portfolio for a given month);
5. Using a range of caps on day-of percentage *baseline adjustments* (*e.g.,* cap percentages of 20, 30, 40, and 50, plus unlimited adjustments); and
6. Comparing results across only those customers who actually selected the day-of baseline adjustment in 2012 to a case where baselines are adjusted for *all* enrolled and nominated customers (*i.e.,* assuming that all customers in a portfolio selected the adjustment).

The baseline analysis was conducted for both *actual* event days in 2012, as well as a set of event-like non-event days, or *simulated* events, in July, August, and September. In the

case of *actual events*, the alternative baselines were compared to the baseline loads implied by the customer-level regression analyses conducted in the 2012 ex post load impact evaluation (*i.e.,* estimated load impacts are added back in to the observed event-day loads to create a "but for the event" reference load). In the case of *simulated* events, the observed loads on the event-like days serve as "true" baselines, which are then compared to all of the relevant alternative baseline methods.

2. METHODOLOGY

Two categories of performance metrics were calculated for each alternative baseline method: measures of *accuracy* (*e.g.,* the average percent error, regardless of sign) and measures of *bias* (*i.e.,* the tendency of a baseline to under-state or over-state the true baseline). In both types of metrics, the calculations begin with the basic notion of a *baseline error*, which is the difference between the baseline that is "predicted" by one of the alternative baseline methods for an event window, and the "actual," or "true" baseline.[1] A common practice in baseline studies is to divide the baseline error by the level of the true baseline load to produce *percentage* errors.

2.1 Accuracy

A common accuracy metric is the average of the *absolute values* of the percent errors over the relevant observations, such as customers and events, or Mean Absolute Percentage Error (MAPE). The key feature of this measure is that the absolute values count positive and negative errors equally, rather than allowing them to cancel each other. The relevant formula is the following:

$MAPE = (1/n) \sum [|(L^P_i - L^A_i)| / L^A_i]$,

where the summation is over all observations *i* = 1 to *n* of customers and events,
L^A_h is the "actual" observed or regression-based baseline load, and
L^P_h is one of the alternative *predicted* baseline loads.

---

[1] In several previous baseline studies, we have used the convention of defining the baseline error as the difference between the actual baseline and the predicted baseline. However, this convention has the unfortunate feature that *positive* values (*i.e.*, the actual baseline is greater than the predicted baseline) imply *under-stated* baselines, which is somewhat counterintuitive. With the definition used in this study, positive errors imply *over-stated* baselines, and negative errors imply *under-stated* baselines, which seems more intuitive.

In previous baseline studies[2], we have reported an alternative accuracy metric known as the Relative Root Mean Square Error, or RRMSE, whose formula is:

$$RRMSE = [(1/n) \sum (L^P_i - L^A_i)^2]^{1/2} / [(1/n) \sum (L^A_i)^2]^{1/2} .$$

The two measures are closely related, where the primary difference between MAPE and RRMSE is that the latter metric weights larger (percent) errors more heavily than smaller errors, since it squares the errors and loads before averaging. Both metrics produce values in units of percentages. MAPE has the advantage of being somewhat more transparent and easily understood. Values of both metrics are provided in this study.

2.2 Bias

A bias metric is designed to measure the extent to which errors tend to be positive or negative, or in the present application, for baseline methods to have a tendency to over-state or under-state true baseline values. Three basic metrics have been used in previous baseline studies, including *mean percent error*, *median percent error*, and *percentiles* of the distribution of percent errors. The mean percent error is simply the average of errors across events and/or customers. The median percent error is the midpoint of the distribution of percent errors.

A principle advantage of the median, rather than the mean in the context of baseline errors, is that percentage errors in some cases can be quite large, thus dominating the mean value and producing values that are not representative of the full distribution. Presenting percentile statistics provides a more comprehensive picture of the full distribution of baseline errors than either the mean or median values alone. This study reports mean, median, and 10, 25, 50, 75, and 90 percentile values, where the 50th percentile value is the median.

---

[2] "2008 Evaluation of California Statewide Aggregator Demand Response Programs, Vol. II: Baseline Analysis of PG&E's AMP Program," CA Energy Consulting, CALMAC Study ID PGE0274.02, May 1, 2009; "2009 Impact Evaluation of PG&E's Peak Choice Demand Response Program for Commercial and Industrial Customers: Ex Post and Ex Ante Report," CA Energy Consulting, April 1, 2010 (Section 5); and "Highly Volatile-Load Customer Study," CA Energy Consulting, October 27, 2010; "2011 Load Impact Evaluation of California Statewide Demand Bidding Programs (DBP) for Non-Residential Customers: Ex Post and Ex Ante Report," CA Energy Consulting, CALMAC Study ID SCE0317.01, May 29, 2012.

3.   BASELINE PERFORMANCE – CAPACITY BIDDING PROGRAMS

This section reports baseline performance results for the three utilities' CBP programs. Results for portfolios of DA and DO notice are shown separately. The accuracy and bias performance statistics are reported by four primary factors of interest in the study. These are the following:

- Load type
    - Aggregate (summations of customer-level baselines); and
    - Portfolio (baselines based on the sum of nominated customer loads within a portfolio; *e.g.,* CBP DO 1-4)
- Adjustment choice
    - As chosen (aggregates individual customer-level baselines applying the day-of adjustment only if it was chosen by the customer)
    - Universal (aggregates customer-level baselines, all of which are adjusted regardless of customer choice)
- Event type (actual or simulated)
- Adjustment cap (Unadjusted, 20, 30, 40 and 50 percent, and Unlimited)

For each combination of factors, the following statistics are provided:
- Number of customer accounts
- Bias statistics (MPE and Median % Error)
- Accuracy statistics (MAPE and RRMSE)
- Percentiles of percent errors (10, 25, 50, 75, and 90)

3.1 PG&E CBP

Table 3–1 reports baseline performance statistics for PG&E's CBP DA program. As a guide to the reader, we begin this section by describing the three main panels in some detail. The first panel contains results for baselines that are constructed by summing individual customer baselines to an aggregate level, where the individual baselines are adjusted only where the customer chose the adjustment during the 2012 program year, and by which cap is applied. Two sub-panels report results by event type (actual and simulated). No values are shown in the first line (Unadj.), since at least some customers chose adjustments.[3] All rows in these sub-panels reflect aggregations of unadjusted baselines for those who chose no adjustment and adjusted baselines for those who chose adjustment, with values in each row reflecting the indicated caps.

In the first major panel of "As Chosen" results, the "Bias" columns indicate that the alternative baseline methods tend to understate the true baselines by about 13 percent on actual event days and 6 percent on simulated event days, according to MPE, but by less than 1 percent according to the median % error. The "Accuracy" columns indicate typical MAPE values (average errors) of about 16 percent for actual events and 11

---

[3] The first line of the "Universal" panel shows results for the case in which all customers are assumed to have unadjusted baselines.

percent for simulated events.[4] Bias and accuracy results vary by relatively small amounts across adjustment caps. The percentiles of errors illustrate the reason for the difference between mean and median percent errors in the earlier columns. That is, twenty-five percent of the errors fall below – 10 percent, and some are quite large. At the other end of the distribution, 25 percent of customers have errors greater than one percent, and the 90 percentile values are only 3 to 4 percent. Thus, while the errors are centered just below zero, the negative errors (under-stated baselines) tend to be much larger than the positive errors.

The middle panel of "universal" results represents the case in which each customer's baseline receives a day-of adjustment, varying from no adjustment, through the various caps, to unlimited adjustment. For the actual events, the MAPE values are relatively large, with average errors declining from about 19 percent with no adjustment, to about 10 percent with unlimited adjustment. The median percent errors generally indicate about 1.5 to 2 percent downward bias. The mean percent errors are somewhat smaller than in the "as-chosen" case, declining from about a 10 percent downward bias to a 1.4 percent upward bias over the range of caps. The cause, as seen in the percentiles, is a greater number of larger upward biases than in the "as chosen" case, which offset some of the large downward biases. Results for the simulated events have similar patterns, but the errors are generally smaller than for the actual events.

The third panel shows the case of portfolio loads, in which baselines are calculated after summing each customer's load within a portfolio. The average errors (MAPE) under this baseline method are larger than either of the aggregations of individual customer baselines for the unadjusted case and restrictive cap cases, though they fall with larger adjustments. However, the RRMSE values are more in line with the aggregated baseline cases, suggesting that the largest percent errors are associated with relatively small absolute errors. The median percent errors show a relatively small downward bias for the actual events, declining with the size of the adjustment cap, and zero for the simulated events for caps of 30 percent and above. The percentile results show relatively large percent errors at both ends of the distribution, including up to 25 percent of the observations having negative percent errors of more than 10 or 15 percent for the actual events.

---

[4] The discussion in the report focuses on the more straightforward MAPE values. However, as a guide to the reader, the results in the first panel, in which RRMSE values are generally less than MAPE values, indicate that smaller *percent* errors across customers are associated with larger *absolute* errors (since RRMSE weights larger errors more heavily), or, equivalently, that the larger percentage values indicated by MAPE are associated with smaller absolute errors.

Table 3-1: Baseline Performance – PG&E CBP DA

Load Type	Adj. Choice	Event Type	Cap	Cust. Count	Bias		Accuracy		Percentiles				
					MPE	Median % Error	MAPE	RRMSE	p10	p25	p50	p75	p90
Aggregate	As Chosen	Actual	Unadj										
			20	167	-13.1%	-0.5%	16.4%	4.2%	-21.9%	-8.9%	-0.5%	2.5%	5.3%
			30	167	-13.3%	-0.5%	15.9%	3.8%	-20.2%	-8.9%	-0.5%	0.9%	3.5%
			40	167	-13.2%	-0.4%	15.7%	3.6%	-18.7%	-8.9%	-0.4%	0.9%	3.3%
			50	167	-13.1%	-0.5%	15.6%	3.3%	-17.0%	-8.9%	-0.5%	0.9%	3.1%
			Unlim	167	-13.0%	-1.7%	15.1%	2.7%	-15.8%	-8.5%	-1.7%	0.8%	1.6%
		Simulated	Unadj										
			20	165	-6.2%	-1.1%	11.4%	2.2%	-10.6%	-4.5%	-1.1%	1.6%	10.4%
			30	165	-6.1%	-0.2%	11.6%	2.7%	-8.8%	-5.7%	-0.2%	1.8%	10.4%
			40	165	-5.9%	-0.4%	11.6%	2.7%	-8.3%	-5.9%	-0.4%	3.2%	10.4%
			50	165	-5.7%	-0.4%	11.5%	2.7%	-8.3%	-4.5%	-0.4%	3.2%	10.4%
			Unlim	165	-5.2%	0.0%	11.2%	3.3%	-8.3%	-4.3%	0.0%	1.8%	10.9%
Aggregate	Universal	Actual	Unadj	167	-10.3%	0.8%	18.7%	6.2%	-24.3%	-7.9%	0.8%	6.8%	18.7%
			20	167	-11.1%	-1.0%	15.7%	4.1%	-21.9%	-14.5%	-1.0%	3.4%	11.4%
			30	167	-10.1%	-1.0%	13.9%	3.7%	-19.9%	-16.5%	-1.0%	1.1%	11.4%
			40	167	-8.8%	-1.5%	12.3%	3.4%	-18.8%	-14.8%	-1.5%	0.9%	11.4%
			50	167	-7.1%	-1.3%	11.6%	3.1%	-18.5%	-14.8%	-1.3%	1.5%	11.9%
			Unlim	167	1.3%	-1.3%	9.8%	2.5%	-14.8%	-8.5%	-1.3%	1.6%	11.9%
		Simulated	Unadj	165	-5.6%	0.6%	11.8%	2.9%	-13.0%	-4.4%	0.6%	3.4%	10.5%
			20	165	-4.9%	-1.2%	9.2%	2.1%	-10.6%	-5.9%	-1.2%	0.8%	4.3%
			30	165	-3.8%	-1.4%	8.5%	2.5%	-8.8%	-6.2%	-1.4%	1.0%	3.6%
			40	165	-2.6%	-1.4%	7.5%	2.5%	-6.7%	-5.9%	-1.4%	1.1%	3.7%
			50	165	-1.4%	-0.8%	6.5%	2.5%	-6.5%	-4.5%	-0.8%	1.4%	6.9%
			Unlim	165	2.6%	0.0%	6.7%	3.1%	-6.2%	-3.1%	0.0%	2.7%	16.4%
Portfolio	Universal	Actual	Unadj	166	-28.7%	-1.8%	37.1%	10.6%	-96.6%	-42.6%	-1.8%	9.6%	18.7%
			20	166	-19.2%	-1.7%	22.7%	3.3%	-57.3%	-15.1%	-1.7%	0.3%	11.5%
			30	166	-14.5%	-1.7%	17.5%	2.8%	-37.6%	-15.0%	-1.7%	-0.2%	11.5%
			40	166	-10.0%	-1.6%	13.1%	2.6%	-18.7%	-11.4%	-1.6%	-0.2%	11.5%
			50	166	-6.5%	-1.6%	10.4%	2.6%	-15.0%	-9.5%	-1.6%	0.9%	11.5%
			Unlim	166	1.7%	-1.1%	9.6%	2.6%	-14.0%	-5.3%	-1.1%	3.0%	11.8%
		Simulated	Unadj	164	-12.9%	0.00%	18.5%	4.6%	-31.3%	-17.0%	0.00%	3.2%	10.5%
			20	164	-6.0%	-1.13%	10.5%	2.5%	-12.0%	-5.1%	-1.13%	1.2%	6.4%
			30	164	-3.9%	0.00%	9.1%	2.6%	-13.9%	-5.7%	0.00%	1.6%	12.2%
			40	164	-2.5%	0.00%	7.5%	2.6%	-7.7%	-3.8%	0.00%	2.4%	6.6%
			50	164	-0.9%	0.00%	7.0%	2.6%	-6.6%	-3.3%	0.00%	2.7%	14.6%
			Unlim	164	2.0%	0.00%	8.3%	2.6%	-7.7%	-3.8%	0.00%	2.7%	17.1%

The results for PG&E CBP DO, shown in Table 3–2, have similar patterns to those for CBP DA. Average percent errors as indicated by MAPE are quite large for both "as-chosen" and "universal" aggregations of customer-level baselines, and even larger for the portfolio baselines. Median percent errors generally indicate either downward or upward biases of less than one to 1.4 percent. The median percent errors for the portfolio method are relatively small, especially for the simulated events, but mask a relatively small number of large negative errors, as shown in the percentiles.

CA Energy Consulting

Table 3-2: Baseline Performance – PG&E CBP DO

Load Type	Adj. Choice	Event Type	Cap	Cust. Count	Bias		Accuracy		Percentiles				
					MPE	Median % Error	MAPE	RRMSE	p10	p25	p50	p75	p90
Aggregate	As Chosen	Actual	Unadj										
			20	370	-5.6%	0.4%	20.5%	2.4%	-52.9%	-4.5%	0.4%	8.2%	22.1%
			30	370	-6.1%	-0.3%	20.3%	2.3%	-52.9%	-4.6%	-0.3%	8.2%	22.1%
			40	370	-6.5%	-0.5%	20.2%	2.4%	-52.9%	-4.7%	-0.5%	8.2%	22.1%
			50	370	-6.7%	-0.5%	20.2%	2.5%	-52.9%	-4.7%	-0.5%	8.2%	22.1%
			Unlim	370	-8.3%	-1.1%	21.2%	7.9%	-52.9%	-12.8%	-1.1%	8.2%	22.1%
		Simulated	Unadj										
			20	364	0.0%	1.3%	7.3%	2.8%	-6.9%	-1.0%	1.3%	5.2%	9.9%
			30	364	-0.1%	1.1%	7.1%	2.7%	-6.9%	-0.5%	1.1%	4.8%	8.9%
			40	364	-0.1%	1.4%	7.1%	2.7%	-6.9%	-0.7%	1.4%	4.7%	8.9%
			50	364	-0.1%	1.4%	7.2%	2.8%	-6.9%	-0.8%	1.4%	4.7%	8.9%
			Unlim	364	-0.3%	1.2%	7.4%	3.2%	-6.9%	-2.4%	1.2%	5.4%	8.9%
Aggregate	Universal	Actual	Unadj	370	-2.7%	5.4%	22.6%	6.7%	-52.9%	-4.4%	5.4%	9.2%	22.1%
			20	370	-4.3%	0.1%	14.9%	2.3%	-25.4%	-4.0%	0.1%	1.5%	6.5%
			30	370	-4.8%	-0.8%	14.0%	2.2%	-25.4%	-4.0%	-0.8%	1.2%	4.2%
			40	370	-4.7%	-1.2%	13.4%	2.3%	-25.4%	-3.9%	-1.2%	0.8%	3.4%
			50	370	-4.3%	-1.5%	12.9%	2.4%	-21.7%	-3.9%	-1.5%	0.6%	3.2%
			Unlim	370	-5.4%	-2.2%	13.5%	7.8%	-21.7%	-9.3%	-2.2%	0.2%	3.5%
		Simulated	Unadj	364	1.30%	2.83%	8.1%	4.3%	-6.7%	0.1%	2.83%	6.5%	10.6%
			20	364	0.25%	1.19%	5.4%	2.8%	-6.9%	-1.4%	1.19%	3.7%	9.5%
			30	364	-0.36%	0.84%	5.5%	2.7%	-6.9%	-1.0%	0.84%	3.8%	8.1%
			40	364	-0.73%	0.74%	5.8%	2.7%	-8.0%	-1.0%	0.74%	3.5%	5.9%
			50	364	-0.89%	0.55%	6.0%	2.8%	-8.0%	-1.8%	0.55%	3.6%	5.9%
			Unlim	364	-1.16%	0.55%	6.2%	3.1%	-8.0%	-2.4%	0.55%	3.4%	6.2%
Portfolio	Universal	Actual	Unadj	370	-13.2%	4.9%	32.5%	7.5%	-96.1%	-10.3%	4.9%	10.2%	22.1%
			20	370	-11.7%	-2.1%	21.7%	2.5%	-90.4%	-11.9%	-2.1%	0.6%	6.5%
			30	370	-10.1%	-2.1%	19.0%	2.4%	-68.1%	-11.9%	-2.1%	0.3%	3.2%
			40	370	-8.1%	-2.2%	16.5%	2.3%	-46.2%	-11.9%	-2.2%	0.3%	3.2%
			50	370	-5.8%	-2.2%	14.1%	2.2%	-23.1%	-11.9%	-2.2%	0.3%	3.2%
			Unlim	370	-4.0%	-1.1%	12.1%	2.2%	-21.7%	-3.8%	-1.1%	0.8%	3.3%
		Simulated	Unadj	364	-13.2%	1.19%	21.9%	8.9%	-39.9%	-4.6%	1.19%	6.4%	13.7%
			20	364	-9.3%	0.00%	14.9%	3.9%	-29.2%	-2.1%	0.00%	4.1%	9.4%
			30	364	-7.9%	0.00%	12.9%	3.5%	-27.6%	-2.1%	0.00%	3.7%	8.1%
			40	364	-6.3%	0.00%	11.2%	3.6%	-29.2%	-3.2%	0.00%	3.3%	6.1%
			50	364	-4.4%	0.00%	9.5%	3.5%	-21.0%	-3.2%	0.00%	3.7%	7.0%
			Unlim	364	-0.4%	0.62%	6.8%	3.5%	-8.0%	-2.1%	0.62%	4.4%	11.7%

3.2 SCE CBP

No results are provided for CBP DA due to the very small number of customer accounts. The results for CBP DO are provided in Table 3–3. Overall accuracy is somewhat greater than for the two previous programs, at 4 to 6 percent MAPE values for both types of aggregate loads and events. The median % errors for the "as chosen" case show an upward bias of 2.4 percent for actual events and a downward bias of 2.7 percent for simulated events. For the "universal" case, the median percent errors are smaller, showing less than one percent downward bias for actual events, and less than one percent upward bias for simulated events. There is little consistent pattern in MAPE or median percent errors across the alternative caps, except that unadjusted baselines have larger errors than any adjustment, and unlimited adjustments produce slightly higher errors.

The average errors for the portfolio baselines are similar to the aggregate baselines, with MAPE values of 4 to 5 percent. Median percent errors are similar to the universal aggregate baselines, with small downward biases for the actual events and upward biases for the simulated events. The median errors generally fall considerably from the unadjusted case to at least some adjustment, but the errors are generally not improved with less restrictive caps.

Table 3-3: Baseline Performance – SCE CBP DO

Load Type	Adj. Choice	Event Type	Cap	Cust. Count	Bias MPE	Bias Median % Error	Accuracy MAPE	Accuracy RRMSE	Percentiles p10	p25	p50	p75	p90
Aggregate	As Chosen	Actual	Unadj										
			20	362	3.19%	2.75%	4.84%	3.79%	-3.49%	0.76%	2.75%	6.42%	9.90%
			30	362	3.19%	2.69%	4.83%	3.79%	-3.44%	0.76%	2.69%	6.42%	9.90%
			40	362	3.17%	2.69%	4.84%	3.79%	-3.49%	0.38%	2.69%	6.42%	9.90%
			50	362	3.15%	2.69%	4.88%	3.79%	-3.49%	0.38%	2.69%	6.42%	9.90%
			Unlim	362	3.05%	2.69%	4.98%	3.79%	-3.49%	0.38%	2.69%	6.42%	9.90%
		Simulated	Unadj										
			20	362	-1.30%	-2.12%	6.11%	4.41%	-9.75%	-5.74%	-2.12%	2.65%	10.30%
			30	362	-1.24%	-2.12%	6.05%	4.41%	-9.29%	-5.53%	-2.12%	2.36%	10.30%
			40	362	-1.17%	-2.12%	6.00%	4.41%	-9.29%	-5.13%	-2.12%	2.36%	11.30%
			50	362	-1.09%	-2.12%	5.97%	4.41%	-9.29%	-5.01%	-2.12%	2.36%	11.37%
			Unlim	362	-0.94%	-1.91%	5.95%	4.41%	-9.29%	-4.78%	-1.91%	2.36%	11.37%
Aggregate	Universal	Actual	Unadj	362	3.20%	2.75%	4.97%	3.79%	-3.5%	0.8%	2.75%	5.9%	9.9%
			20	362	0.07%	-0.18%	3.58%	1.99%	-5.7%	-2.1%	-0.18%	2.1%	5.4%
			30	362	-0.23%	-0.34%	4.00%	2.02%	-5.9%	-2.5%	-0.34%	2.1%	6.9%
			40	362	-0.32%	-0.37%	4.16%	2.03%	-6.0%	-2.5%	-0.37%	2.1%	6.8%
			50	362	-0.39%	-0.37%	4.25%	2.03%	-6.6%	-2.5%	-0.37%	2.1%	6.7%
			Unlim	362	-0.53%	-0.37%	4.36%	2.03%	-6.8%	-2.5%	-0.37%	2.1%	6.7%
		Simulated	Unadj	362	-1.43%	-2.12%	6.24%	4.41%	-10.8%	-6.4%	-2.12%	2.7%	10.3%
			20	362	0.60%	0.98%	3.96%	2.21%	-6.0%	-2.0%	0.98%	3.0%	7.2%
			30	362	0.93%	0.86%	4.34%	2.26%	-5.0%	-2.2%	0.86%	3.5%	10.1%
			40	362	1.09%	0.86%	4.62%	2.28%	-6.0%	-2.4%	0.86%	3.5%	11.0%
			50	362	1.14%	0.86%	4.76%	2.28%	-6.0%	-2.5%	0.86%	3.4%	11.0%
			Unlim	362	1.27%	0.97%	4.85%	2.29%	-6.0%	-2.4%	0.97%	3.4%	11.0%
Portfolio	Universal	Actual	Unadj	368	0.91%	2.18%	5.91%	3.90%	-11.29%	-1.90%	2.18%	5.07%	9.58%
			20	368	-0.51%	-0.36%	4.11%	2.02%	-6.10%	-2.07%	-0.36%	2.04%	6.40%
			30	368	-0.62%	-0.36%	4.30%	2.02%	-7.55%	-2.07%	-0.36%	1.78%	6.40%
			40	368	-0.62%	-0.36%	4.30%	2.02%	-7.55%	-2.07%	-0.36%	1.78%	6.40%
			50	368	-0.62%	-0.36%	4.30%	2.02%	-7.55%	-2.07%	-0.36%	1.78%	6.40%
			Unlim	368	-0.62%	-0.36%	4.30%	2.02%	-7.55%	-2.07%	-0.36%	1.78%	6.40%
		Simulated	Unadj	366	-4.95%	-3.88%	9.00%	5.06%	-20.22%	-8.99%	-3.88%	0.84%	7.02%
			20	366	1.13%	0.98%	5.23%	2.41%	-5.96%	-2.11%	0.98%	4.03%	10.96%
			30	366	1.94%	1.35%	5.23%	2.29%	-5.61%	-1.78%	1.35%	4.61%	13.04%
			40	366	2.15%	1.45%	5.14%	2.29%	-4.39%	-1.65%	1.45%	4.82%	13.04%
			50	366	2.07%	1.45%	5.06%	2.29%	-4.39%	-1.65%	1.45%	4.82%	13.04%
			Unlim	366	1.90%	1.45%	4.89%	2.29%	-4.39%	-1.65%	1.45%	4.61%	11.68%

3.3 SDG&E CBP

Table 3–4 shows baseline performance for SDG&E's CBP DA portfolio. Overall accuracy is poor for both the "as-chosen" and "universal" cases, with MAPE errors of more than 40 percent for actual events, and 20 percent for simulated events. The baselines are generally biased downward, with median percent errors of about –3 percent for actual events and –3.7 percent for simulated events for the as chosen adjustments. The biases are smaller, less than 1 percent, plus or minus, in most cases for universal adjustments.

For the as-chosen case, the results generally do not vary by adjustment cap. For the universal case, the bias generally becomes smaller with some adjustment, but becomes larger for unlimited adjustments.

The portfolio baselines have similar patterns of errors to the aggregate baselines. MAPE values are nearly 50% for the actual events, and do not improve much with greater adjustments. Median percent errors are one percent or smaller, requiring some adjustment for the simulated events. The percentile results show that the small median errors mask a large number of large downward and upward biases at the tails of the distributions of percent errors.

Table 3-4: Baseline Performance – SDG&E CBP DA

Load Type	Adj. Choice	Event Type	Cap	Cust. Count	Bias MPE	Bias Median % Error	Accuracy MAPE	Accuracy RRMSE	p10	p25	p50	p75	p90
Aggregate	As Chosen	Actual	Unadj 20	79	-37.9%	-2.2%	42.5%	12.3%	-82.6%	-14.5%	-2.2%	3.1%	5.1%
			30	79	-38.2%	-2.8%	42.4%	12.2%	-82.6%	-14.5%	-2.8%	1.6%	4.9%
			40	79	-38.3%	-3.1%	42.3%	12.2%	-82.6%	-14.5%	-3.1%	1.4%	4.9%
			50	79	-38.4%	-3.2%	42.3%	12.2%	-82.6%	-14.5%	-3.2%	1.3%	4.9%
			Unlim	79	-38.6%	-4.8%	42.5%	12.3%	-82.6%	-14.5%	-4.8%	1.1%	4.9%
		Simulated	Unadj 20	78	-12.5%	-3.7%	20.0%	10.1%	-47.5%	-26.9%	-3.7%	3.5%	17.1%
			30	78	-12.5%	-3.7%	20.0%	10.1%	-47.5%	-26.9%	-3.7%	3.3%	17.1%
			40	78	-12.5%	-3.7%	19.9%	10.1%	-47.5%	-26.9%	-3.7%	3.1%	17.1%
			50	78	-12.5%	-3.7%	19.9%	10.1%	-47.5%	-26.9%	-3.7%	3.1%	17.1%
			Unlim	78	-12.5%	-3.7%	19.9%	10.1%	-47.5%	-26.9%	-3.7%	3.1%	17.1%
Aggregate	Universal	Actual	Unadj	79	-36.0%	-4.1%	44.6%	14.9%	-82.6%	-14.5%	-4.1%	6.9%	13.9%
			20	79	-41.0%	-0.2%	45.7%	8.8%	-81.4%	-21.3%	-0.2%	3.8%	9.0%
			30	79	-41.2%	-0.6%	47.1%	8.4%	-79.7%	-15.1%	-0.6%	3.7%	10.7%
			40	79	-40.9%	0.1%	48.7%	8.5%	-78.0%	-15.3%	0.1%	6.5%	10.6%
			50	79	-40.8%	0.6%	51.3%	9.0%	-76.3%	-15.4%	0.6%	8.4%	17.1%
			Unlim	79	-80.6%	-2.3%	93.3%	14.2%	-282.4%	-63.7%	-2.3%	4.8%	29.3%
		Simulated	Unadj	78	-12.1%	-4.1%	20.6%	10.8%	-47.5%	-26.9%	-4.1%	6.5%	17.1%
			20	78	-12.6%	-2.3%	19.2%	8.1%	-39.1%	-28.4%	-2.3%	4.2%	10.1%
			30	78	-12.1%	-0.1%	19.6%	8.2%	-47.5%	-31.2%	-0.1%	4.2%	16.4%
			40	78	-11.5%	-0.6%	20.2%	8.6%	-52.8%	-32.9%	-0.6%	4.2%	18.0%
			50	78	-11.0%	-0.1%	21.2%	9.4%	-57.0%	-33.2%	-0.1%	4.2%	26.5%
			Unlim	78	-16.4%	-0.1%	33.6%	13.5%	-65.4%	-30.2%	-0.1%	6.1%	33.6%
Portfolio	Universal	Actual	Unadj	79	-34.5%	-1.0%	44.2%	15.0%	-82.9%	-14.1%	-1.0%	7.1%	16.2%
			20	79	-39.5%	-1.0%	46.6%	8.1%	-65.4%	-16.3%	-1.0%	2.2%	15.9%
			30	79	-38.4%	-1.0%	47.5%	8.4%	-55.0%	-13.9%	-1.0%	4.5%	23.7%
			40	79	-36.1%	-0.4%	48.3%	9.9%	-45.4%	-13.9%	-0.4%	4.5%	27.0%
			50	79	-35.4%	-1.0%	49.2%	11.2%	-45.4%	-13.9%	-1.0%	4.5%	30.7%
			Unlim	79	-34.7%	-1.0%	49.9%	12.4%	-45.4%	-13.9%	-1.0%	4.5%	33.7%
		Simulated	Unadj	78	-14.1%	-6.8%	22.9%	10.6%	-49.3%	-32.3%	-6.8%	6.5%	17.4%
			20	78	-11.2%	-0.8%	18.4%	10.0%	-41.6%	-34.1%	-0.8%	2.9%	12.6%
			30	78	-10.3%	-0.1%	19.9%	11.2%	-50.8%	-29.8%	-0.1%	3.9%	15.3%
			40	78	-9.3%	-0.1%	22.0%	12.3%	-56.4%	-24.6%	-0.1%	6.6%	24.6%
			50	78	-8.9%	-0.1%	23.9%	13.2%	-56.4%	-28.5%	-0.1%	6.6%	37.2%
			Unlim	78	-7.8%	-0.1%	26.2%	14.2%	-56.4%	-28.5%	-0.1%	6.6%	37.4%

Table 3–5 reports results for CBP DO. Baseline accuracy for this program and notice is generally good, regardless of type and adjustment. For the aggregations of individual baselines, MAPE values generally show average errors of less than 5 percent. Biases are

less than one percent downward for actual events and slightly more than one percent upward for simulated events. Biases vary only slightly with alternative adjustment caps.

With at least some adjustment, the portfolio baselines produce average errors of 4 to 6 percent. Biases measured by the median percent error are less than one percent for both actual and simulated events, and do not improve past a 30 percent cap. The range of percent errors is also relatively narrow, with relatively few errors greater than 5 percent.

Table 3-5: Baseline Performance – SDG&E CBP DO

Load Type	Adj. Choice	Event Type	Cap	Cust. Count	Bias		Accuracy		Percentiles				
					MPE	Median % Error	MAPE	RRMSE	p10	p25	p50	p75	p90
Aggregate	As Chosen	Actual	Unadj										
			20	321	-0.31%	0.43%	4.80%	2.81%	-9.1%	-1.6%	0.43%	2.7%	9.4%
			30	321	-0.45%	0.05%	4.81%	2.86%	-9.1%	-2.0%	0.05%	2.5%	9.4%
			40	321	-0.47%	-0.14%	4.83%	2.87%	-9.1%	-2.1%	-0.14%	2.5%	9.4%
			50	321	-0.47%	-0.12%	4.84%	2.89%	-9.1%	-2.1%	-0.12%	2.5%	9.4%
			Unlim	321	-0.79%	0.07%	5.21%	3.39%	-10.4%	-3.1%	0.07%	2.5%	9.4%
		Simulated	Unadj										
			20	320	0.27%	1.16%	4.49%	3.85%	-4.0%	-1.1%	1.16%	4.1%	6.7%
			30	320	0.28%	1.18%	4.52%	3.93%	-4.1%	-1.1%	1.18%	4.1%	6.7%
			40	320	0.29%	1.14%	4.55%	3.99%	-4.3%	-1.2%	1.14%	4.1%	6.9%
			50	320	0.30%	1.16%	4.58%	4.05%	-4.5%	-1.3%	1.16%	4.1%	7.1%
			Unlim	320	0.29%	1.12%	4.59%	4.09%	-4.5%	-1.3%	1.12%	4.1%	7.3%
Aggregate	Universal	Actual	Unadj	321	2.25%	4.13%	7.18%	6.13%	-9.1%	-1.9%	4.13%	7.6%	11.5%
			20	321	-2.17%	-0.43%	3.96%	2.94%	-11.3%	-1.6%	-0.43%	1.1%	3.3%
			30	321	-2.28%	-0.64%	3.89%	2.93%	-9.4%	-2.0%	-0.64%	0.8%	3.4%
			40	321	-2.18%	-0.56%	3.78%	2.91%	-7.7%	-2.3%	-0.56%	0.9%	3.3%
			50	321	-2.04%	-0.56%	3.65%	2.89%	-6.0%	-2.3%	-0.56%	0.9%	3.2%
			Unlim	321	-1.84%	-0.43%	3.91%	3.38%	-6.0%	-2.7%	-0.43%	1.0%	4.0%
		Simulated	Unadj	320	0.34%	0.50%	4.44%	4.05%	-4.5%	-0.7%	0.50%	4.8%	8.3%
			20	320	0.01%	1.16%	4.50%	3.91%	-4.0%	-1.7%	1.16%	3.6%	6.5%
			30	320	-0.01%	1.12%	4.55%	3.99%	-4.1%	-1.7%	1.12%	3.6%	6.7%
			40	320	-0.03%	1.05%	4.58%	4.05%	-4.3%	-1.6%	1.05%	3.5%	6.9%
			50	320	-0.02%	1.03%	4.61%	4.11%	-4.5%	-1.7%	1.03%	3.4%	7.0%
			Unlim	320	-0.03%	0.89%	4.61%	4.15%	-4.5%	-1.7%	0.89%	3.5%	7.2%
Portfolio	Universal	Actual	Unadj	321	-1.52%	1.30%	8.90%	6.09%	-22.99%	-4.04%	1.30%	5.91%	9.73%
			20	321	-1.33%	-0.12%	3.81%	3.19%	-4.39%	-2.54%	-0.12%	1.36%	4.00%
			30	321	-1.36%	-0.03%	3.56%	3.14%	-4.39%	-1.45%	-0.03%	1.36%	3.69%
			40	321	-1.36%	-0.03%	3.56%	3.14%	-4.39%	-1.45%	-0.03%	1.36%	3.69%
			50	321	-1.36%	-0.03%	3.56%	3.14%	-4.39%	-1.45%	-0.03%	1.36%	3.69%
			Unlim	321	-1.36%	-0.03%	3.56%	3.14%	-4.39%	-1.45%	-0.03%	1.36%	3.69%
		Simulated	Unadj	320	0.76%	0.56%	6.13%	4.67%	-6.56%	-1.28%	0.56%	4.83%	8.27%
			20	320	0.99%	0.81%	5.56%	4.75%	-4.84%	-1.08%	0.81%	3.76%	7.72%
			30	320	1.00%	0.81%	5.56%	4.63%	-4.84%	-1.08%	0.81%	4.26%	8.38%
			40	320	0.88%	0.81%	5.44%	4.51%	-4.84%	-1.08%	0.81%	4.26%	8.38%
			50	320	0.75%	0.81%	5.32%	4.40%	-4.84%	-1.08%	0.81%	4.26%	8.38%
			Unlim	320	-0.42%	0.48%	5.15%	4.28%	-5.54%	-1.50%	0.48%	3.75%	7.72%

4. BASELINE PERFORMANCE – PG&E AMP

Table 4–1 reports results for AMP DA. Accuracy and bias are relatively poor for the as-chosen aggregations of baselines, with average errors of 9 percent and median percent errors of 10 percent for actual events, and 6 percent for both for simulated events.

Results for universal adjustments are somewhat better, with average errors falling from 9 percent for the unadjusted case to 2.6 percent for a 50 percent cap for actual events, and from 6 percent for no adjustment to 2.1 percent with 50 percent adjustment for simulated events. Median percent errors indicate upward biases ranging from 10 percent for no adjustment to 2.4 for unlimited adjustment for actual events, and 6 percent for no adjustment to 1.4 for unlimited adjustment for simulated events.

The portfolio baselines are somewhat more accurate, with MAPE values falling from 9 percent to 2.5 percent with some adjustment for actual events, and from less than 6 percent to 1.9 percent for simulated events. Portfolio baselines are still biased upward, but only by 2.5 percent for actual events, and 1.6 percent for simulated events (both with some adjustment).[5]

---

[5] Values that are constant across alternative adjustment caps indicate that the caps are not binding. In this case, an adjustment of less than 20 percent improves accuracy and bias relative to the no-adjustment case, but no larger adjustment is needed.

Table 4-1: Baseline Performance – PG&E AMP DA

Load Type	Adj. Choice	Event Type	Cap	Cust. Count	Bias MPE	Bias Median % Error	Accuracy MAPE	Accuracy RRMSE	p10	p25	p50	p75	p90
Aggregate	As Chosen	Actual	Unadj										
			20	233	9.17%	10.05%	9.17%	9.54%	5.75%	5.75%	10.1%	11.7%	11.7%
			30	233	9.17%	10.05%	9.17%	9.54%	5.75%	5.75%	10.1%	11.7%	11.7%
			40	233	9.17%	10.05%	9.17%	9.54%	5.75%	5.75%	10.1%	11.7%	11.7%
			50	233	9.17%	10.05%	9.17%	9.54%	5.75%	5.75%	10.1%	11.7%	11.7%
			Unlim	233	9.17%	10.05%	9.17%	9.54%	5.75%	5.75%	10.1%	11.7%	11.7%
		Simulated	Unadj	231									
			20	231	5.93%	5.96%	5.93%	6.71%	1.60%	4.00%	5.96%	7.8%	10.3%
			30	231	5.93%	5.96%	5.93%	6.71%	1.60%	4.00%	5.96%	7.8%	10.3%
			40	231	5.93%	5.96%	5.93%	6.71%	1.60%	4.00%	5.96%	7.8%	10.3%
			50	231	5.93%	5.96%	5.93%	6.71%	1.60%	4.00%	5.96%	7.8%	10.3%
			Unlim	233	5.93%	5.96%	5.93%	6.71%	1.60%	4.00%	5.96%	7.8%	10.3%
Aggregate	Universal	Actual	Unadj	233	9.17%	10.05%	9.17%	9.54%	5.75%	5.75%	10.05%	11.7%	11.7%
			20	233	4.42%	5.41%	4.42%	4.98%	1.44%	1.44%	5.41%	6.4%	6.4%
			30	233	3.20%	4.51%	3.20%	3.96%	0.03%	0.03%	4.51%	5.1%	5.1%
			40	233	2.65%	4.11%	2.86%	3.41%	-0.31%	-0.31%	4.11%	4.1%	4.1%
			50	233	2.33%	3.45%	2.60%	3.05%	-0.41%	-0.41%	3.45%	3.9%	3.9%
			Unlim	231	1.34%	2.37%	2.70%	2.79%	-2.04%	-2.04%	2.37%	3.7%	3.7%
		Simulated	Unadj	231	5.93%	5.96%	5.93%	6.71%	1.60%	4.00%	5.96%	7.8%	10.3%
			20	231	3.04%	3.58%	3.04%	3.43%	0.28%	2.26%	3.58%	4.0%	4.3%
			30	231	2.52%	2.68%	2.52%	2.92%	0.36%	1.22%	2.68%	3.8%	4.3%
			40	231	2.25%	2.41%	2.25%	2.75%	0.24%	0.53%	2.41%	3.7%	4.4%
			50	231	2.04%	2.19%	2.07%	2.66%	-0.10%	0.19%	2.19%	3.6%	4.5%
			Unlim	233	1.55%	1.43%	1.90%	2.44%	-1.15%	-0.07%	1.43%	3.2%	4.4%
Portfolio	Universal	Actual	Unadj	233	9.09%	9.87%	9.09%	9.50%	5.55%	5.55%	9.87%	11.9%	11.9%
			20	233	1.63%	2.48%	2.49%	2.67%	-1.29%	-1.29%	2.48%	3.7%	3.7%
			30	233	1.63%	2.48%	2.49%	2.67%	-1.29%	-1.29%	2.48%	3.7%	3.7%
			40	233	1.63%	2.48%	2.49%	2.67%	-1.29%	-1.29%	2.48%	3.7%	3.7%
			50	233	1.63%	2.48%	2.49%	2.67%	-1.29%	-1.29%	2.48%	3.7%	3.7%
			Unlim	231	1.63%	2.48%	2.49%	2.67%	-1.29%	-1.29%	2.48%	3.7%	3.7%
		Simulated	Unadj	231	5.66%	5.65%	5.66%	6.49%	1.54%	3.60%	5.65%	7.6%	10.1%
			20	231	1.76%	1.58%	1.87%	2.43%	-0.40%	0.24%	1.58%	3.1%	4.6%
			30	231	1.76%	1.58%	1.87%	2.43%	-0.40%	0.24%	1.58%	3.1%	4.6%
			40	231	1.76%	1.58%	1.87%	2.43%	-0.40%	0.24%	1.58%	3.1%	4.6%
			50	231	1.76%	1.58%	1.87%	2.43%	-0.40%	0.24%	1.58%	3.1%	4.6%
			Unlim	164	1.76%	1.58%	1.87%	2.43%	-0.40%	0.24%	1.58%	3.1%	4.6%

Table 4–2 reports results for AMP DO. The average error for the as-chosen case is relatively constant at about 3 percent for actual and simulated events, and at or below 2 percent for the universal case. Median percent errors show slight downward biases of about 1 percent or less for actual events, and upward biases of 2 percent or less for simulated events. For the universal case, the small downward bias is relatively constant for any adjustment, and smaller than with no adjustment, for actual events. For simulated events, the small upward bias becomes even smaller as adjustment caps are relaxed.

For the portfolio baseline, the average error is comparable to the universal aggregate baseline. For actual events, the median percent error indicates a 1.5 percent downward bias, and not varying across caps. The bias is less than 1 percent upward for simulated events. The range of percent errors is also quite narrow.

Table 4-2: Baseline Performance – PG&E AMP DO

Load Type	Adj. Choice	Event Type	Cap	Cust. Count	Bias MPE	Bias Median % Error	Accuracy MAPE	Accuracy RRMSE	p10	p25	p50	p75	p90
Aggregate	As Chosen	Actual	Unadj										
			20	1,125	0.91%	-0.05%	2.78%	2.86%	-4.88%	-1.38%	-0.05%	1.17%	10.5%
			30	1,125	0.63%	-0.55%	2.91%	3.08%	-4.85%	-2.08%	-0.55%	1.17%	10.5%
			40	1,125	0.50%	-0.71%	2.99%	3.24%	-4.66%	-2.41%	-0.71%	1.17%	10.5%
			50	1,125	0.43%	-0.74%	3.05%	3.37%	-4.43%	-2.64%	-0.74%	1.17%	10.5%
			Unlim	1,125	0.21%	-0.76%	3.25%	3.78%	-4.34%	-3.44%	-0.76%	1.17%	10.5%
		Simulated	Unadj										
			20	1,121	2.42%	2.11%	2.83%	2.67%	-0.60%	0.46%	2.11%	4.08%	6.90%
			30	1,121	2.37%	2.11%	2.89%	2.73%	-0.79%	0.14%	2.11%	4.08%	6.90%
			40	1,121	2.33%	2.11%	2.93%	2.80%	-1.18%	0.04%	2.11%	4.08%	6.90%
			50	1,121	2.31%	2.11%	2.95%	2.86%	-1.51%	0.01%	2.11%	4.08%	6.90%
			Unlim	1,121	2.24%	2.02%	2.99%	2.97%	-2.28%	-0.36%	2.02%	4.26%	6.90%
Aggregate	Universal	Actual	Unadj	1,125	3.21%	2.86%	3.83%	3.47%	-2.77%	1.17%	2.86%	3.88%	10.48%
			20	1,125	-0.71%	-0.79%	1.79%	1.60%	-4.93%	-1.47%	-0.79%	-0.05%	3.81%
			30	1,125	-1.24%	-0.92%	1.89%	1.91%	-4.90%	-2.12%	-0.92%	-0.44%	2.61%
			40	1,125	-1.51%	-0.99%	1.96%	2.15%	-4.71%	-2.57%	-0.99%	-0.66%	1.91%
			50	1,125	-1.68%	-1.09%	2.02%	2.33%	-4.46%	-2.86%	-1.09%	-0.83%	1.51%
			Unlim	1,125	-2.27%	-1.47%	2.27%	2.92%	-4.29%	-3.68%	-1.47%	-1.11%	-0.09%
		Simulated	Unadj	1,121	3.32%	2.77%	3.51%	3.49%	0.06%	1.20%	2.77%	6.67%	7.03%
			20	1,121	1.51%	1.67%	1.93%	1.95%	-0.60%	0.43%	1.67%	2.30%	3.17%
			30	1,121	1.28%	1.40%	1.82%	1.96%	-0.79%	0.12%	1.40%	2.20%	2.85%
			40	1,121	1.15%	1.19%	1.77%	2.03%	-1.18%	0.02%	1.19%	2.20%	3.02%
			50	1,121	1.08%	0.93%	1.74%	2.10%	-1.51%	0.00%	0.93%	2.21%	3.32%
			Unlim	1,121	0.82%	0.43%	1.66%	2.22%	-2.28%	-0.44%	0.43%	1.96%	3.91%
Portfolio	Universal	Actual	Unadj	1,125	2.93%	1.22%	3.79%	3.67%	-2.21%	0.20%	1.22%	4.32%	10.45%
			20	1,125	-1.78%	-1.49%	1.90%	3.28%	-4.36%	-2.51%	-1.49%	-0.51%	0.41%
			30	1,125	-1.78%	-1.49%	1.90%	3.28%	-4.36%	-2.51%	-1.49%	-0.51%	0.41%
			40	1,125	-1.78%	-1.49%	1.90%	3.28%	-4.36%	-2.51%	-1.49%	-0.51%	0.41%
			50	1,125	-1.78%	-1.49%	1.90%	3.28%	-4.36%	-2.51%	-1.49%	-0.51%	0.41%
			Unlim	1,125	-1.78%	-1.49%	1.90%	3.28%	-4.36%	-2.51%	-1.49%	-0.51%	0.41%
		Simulated	Unadj	1,121	2.07%	2.00%	3.37%	4.02%	-1.49%	0.27%	2.00%	4.01%	6.78%
			20	1,121	1.04%	0.69%	1.90%	2.46%	-2.18%	-0.42%	0.69%	2.54%	4.32%
			30	1,121	1.04%	0.69%	1.90%	2.46%	-2.18%	-0.42%	0.69%	2.54%	4.32%
			40	1,121	1.04%	0.69%	1.90%	2.46%	-2.18%	-0.42%	0.69%	2.54%	4.32%
			50	1,121	1.04%	0.69%	1.90%	2.46%	-2.18%	-0.42%	0.69%	2.54%	4.32%
			Unlim	1,121	1.04%	0.69%	1.90%	2.46%	-2.18%	-0.42%	0.69%	2.54%	4.32%

5. BASELINE PERFORMANCE – SCE DRC

Table 5–1 reports results for DRC DA. As for some of the other programs, measures of accuracy and bias for the aggregated baselines indicate relatively high average errors of about 9 percent and 8 percent for the actual events, and 18 to 20 percent or more for the simulated events. Median percent errors vary considerably, showing 11 percent *upward* bias for the actual events for the as-chosen adjustments, and 6 percent *downward* bias (for all except the unadjusted option) for the universal adjustment case. For simulated events, the as-chosen adjustment shows *downward* bias of 14 percent for the as-chosen case and 10 to 40 percent *upward* bias for the universal case. For the latter case, the upward bias increases uniformly as the adjustment cap is relaxed.

The portfolio baseline has considerably smaller average errors (2 to 3 percent) and bias (1 to 3 percent downward bias) than the aggregations of individual baselines. For the most part, those values are constant across adjustment caps after some adjustment is allowed. The range of percent errors for the portfolio baselines is quite narrow.

Table 5-1: Baseline Performance – SCE DRC DA

Load Type	Adj. Choice	Event Type	Cap	Cust. Count	Bias MPE	Bias Median % Error	Accuracy MAPE	Accuracy RRMSE	p10	p25	p50	p75	p90
Aggregate	As Chosen	Actual	Unadj 20	79	8.9%	10.9%	8.9%	10.4%	0.43%	0.43%	10.9%	15.5%	15.5%
			30	79	8.9%	10.9%	8.9%	10.4%	0.43%	0.43%	10.9%	15.5%	15.5%
			40	79	8.9%	10.9%	8.9%	10.4%	0.43%	0.43%	10.9%	15.5%	15.5%
			50	79	8.9%	10.9%	8.9%	10.4%	0.43%	0.43%	10.9%	15.5%	15.5%
			Unlim	79	8.9%	10.9%	8.9%	10.4%	0.43%	0.43%	10.9%	15.5%	15.5%
		Simulated	Unadj 20	74	-9.3%	-14.1%	18.1%	18.5%	-26.6%	-23.9%	-14.1%	5.4%	17.6%
			30	74	-9.3%	-14.1%	18.1%	18.5%	-26.6%	-23.9%	-14.1%	5.4%	17.6%
			40	74	-9.3%	-14.1%	18.1%	18.5%	-26.6%	-23.9%	-14.1%	5.4%	17.6%
			50	74	-9.3%	-14.1%	18.1%	18.5%	-26.6%	-23.9%	-14.1%	5.4%	17.6%
			Unlim	74	-9.3%	-14.1%	18.1%	18.5%	-26.6%	-23.9%	-14.1%	5.4%	17.6%
Aggregate	Universal	Actual	Unadj	79	8.9%	10.9%	8.9%	10.4%	0.4%	0.4%	10.9%	15.5%	15.5%
			20	79	-7.6%	-5.7%	7.6%	10.9%	-16.0%	-16.0%	-5.7%	-1.1%	-1.1%
			30	79	-7.6%	-5.7%	7.6%	10.9%	-16.0%	-16.0%	-5.7%	-1.1%	-1.1%
			40	79	-7.6%	-5.7%	7.6%	10.9%	-16.0%	-16.0%	-5.7%	-1.1%	-1.1%
			50	79	-7.6%	-5.7%	7.6%	10.9%	-16.0%	-16.0%	-5.7%	-1.1%	-1.1%
			Unlim	79	-7.6%	-5.7%	7.6%	10.9%	-16.0%	-16.0%	-5.7%	-1.1%	-1.1%
		Simulated	Unadj	74	-9.3%	-14.1%	18.1%	18.5%	-26.6%	-23.9%	-14.1%	5.4%	17.6%
			20	74	11.7%	8.7%	12.4%	24.6%	-1.3%	0.9%	8.7%	22.6%	30.7%
			30	74	20.1%	19.1%	20.1%	26.3%	11.4%	13.2%	19.1%	26.9%	30.7%
			40	74	26.3%	25.6%	26.3%	28.6%	23.1%	23.6%	25.6%	29.0%	30.7%
			50	74	32.5%	33.7%	32.5%	32.0%	23.1%	26.9%	33.7%	38.0%	39.4%
			Unlim	74	40.0%	39.4%	40.0%	37.4%	23.1%	26.9%	39.4%	53.1%	58.2%
Portfolio	Universal	Actual	Unadj	79	0.57%	1.11%	6.59%	7.76%	-9.03%	-9.03%	1.11%	9.62%	9.62%
			20	79	-1.65%	-2.65%	3.11%	3.23%	-4.49%	-4.49%	-2.65%	2.20%	2.20%
			30	79	-1.65%	-2.65%	3.11%	3.23%	-4.49%	-4.49%	-2.65%	2.20%	2.20%
			40	79	-1.65%	-2.65%	3.11%	3.23%	-4.49%	-4.49%	-2.65%	2.20%	2.20%
			50	79	-1.65%	-2.65%	3.11%	3.23%	-4.49%	-4.49%	-2.65%	2.20%	2.20%
			Unlim	79	-1.65%	-2.65%	3.11%	3.23%	-4.49%	-4.49%	-2.65%	2.20%	2.20%
		Simulated	Unadj	76	-1.26%	-0.86%	2.72%	3.45%	-6.85%	-3.37%	-0.86%	0.42%	3.99%
			20	76	-0.74%	-1.20%	2.45%	2.69%	-4.02%	-3.14%	-1.20%	2.36%	2.76%
			30	76	-0.74%	-1.20%	2.45%	2.69%	-4.02%	-3.14%	-1.20%	2.36%	2.76%
			40	76	-0.74%	-1.20%	2.45%	2.69%	-4.02%	-3.14%	-1.20%	2.36%	2.76%
			50	76	-0.74%	-1.20%	2.45%	2.69%	-4.02%	-3.14%	-1.20%	2.36%	2.76%
			Unlim	76	-0.74%	-1.20%	2.45%	2.69%	-4.02%	-3.14%	-1.20%	2.36%	2.76%

Table 5–2 reports results for DRC DO. In this case, average errors are relatively constant for the aggregations of individual baselines, across the adjustment choice, event type, and adjustment caps, with values of 6.5 percent to 7.5 percent. The median percent errors indicate almost zero bias for the as-chosen adjustments for the actual events (*i.e.*, positive and negative errors roughly cancel each other out). However, there is downward bias of 5 percent for the actual events for the universal case. For the simulated events, the downward bias is about 2 percent for both as-chosen and

CA Energy Consulting

universal. Once any adjustment is allowed, the errors are relatively flat across adjustment caps, except for the unlimited adjustment.

The average errors for the portfolio baselines are smaller than for the aggregations of individual baselines. Average errors for both actual and simulated events are about 3 percent, except for the unadjusted case. The median percent errors show a 1.3 percent downward bias for the actual events and 0.7 percent upward bias for the simulated events, except for the unadjusted cases. For the simulated events, an adjustment of 20 to 30 percent is needed to reduce the upward bias from 2 percent to 0.7 percent.

Table 5-2: Baseline Performance – SCE DRC DO

Load Type	Adj. Choice	Event Type	Cap	Cust. Count	Bias		Accuracy		Percentiles				
					MPE	Median % Error	MAPE	RRMSE	p10	p25	p50	p75	p90
Aggregate	As Chosen	Actual	Unadj										
			20	829	-0.27%	-0.03%	6.87%	7.09%	-9.94%	-8.74%	-0.03%	6.92%	10.2%
			30	829	-0.27%	-0.03%	6.87%	7.09%	-9.94%	-8.74%	-0.03%	6.92%	10.2%
			40	829	-0.27%	-0.03%	6.87%	7.09%	-9.94%	-8.74%	-0.03%	6.92%	10.2%
			50	829	-0.27%	-0.03%	6.87%	7.09%	-9.94%	-8.74%	-0.03%	6.92%	10.2%
			Unlim	829	-0.27%	-0.03%	6.87%	7.09%	-9.94%	-8.74%	-0.03%	6.92%	10.2%
		Simulated	Unadj										
			20	822	-2.07%	-1.84%	6.51%	10.5%	-8.27%	-4.13%	-1.84%	1.39%	6.97%
			30	822	-2.07%	-1.84%	6.51%	10.5%	-8.27%	-4.13%	-1.84%	1.39%	6.97%
			40	822	-2.07%	-1.84%	6.51%	10.5%	-8.27%	-4.13%	-1.84%	1.39%	6.97%
			50	822	-2.07%	-1.84%	6.51%	10.5%	-8.27%	-4.13%	-1.84%	1.39%	6.97%
			Unlim	822	-2.07%	-1.84%	6.51%	10.5%	-8.27%	-4.13%	-1.84%	1.39%	6.97%
Aggregate	Universal	Actual	Unadj	829	-0.27%	-0.03%	6.87%	7.09%	-9.94%	-8.75%	-0.03%	6.92%	10.2%
			20	829	-2.53%	-4.97%	5.32%	5.33%	-6.96%	-6.65%	-4.97%	1.07%	7.30%
			30	829	-3.46%	-4.97%	5.87%	5.84%	-9.47%	-8.60%	-4.97%	1.07%	6.16%
			40	829	-4.08%	-4.97%	5.81%	5.87%	-10.2%	-9.47%	-4.97%	1.07%	4.13%
			50	829	-4.90%	-4.97%	5.53%	5.98%	-11.9%	-9.47%	-4.97%	0.84%	1.07%
			Unlim	829	-7.32%	-8.01%	7.68%	7.57%	-13.1%	-12.5%	-8.01%	-3.39%	1.07%
		Simulated	Unadj	822	-2.07%	-1.83%	6.51%	10.5%	-8.26%	-4.12%	-1.83%	1.39%	6.97%
			20	822	-2.26%	-1.08%	7.18%	11.2%	-9.75%	-6.29%	-1.08%	3.31%	4.01%
			30	822	-1.97%	-1.98%	7.62%	11.4%	-9.75%	-5.76%	-1.98%	3.86%	4.67%
			40	822	-2.00%	-2.03%	7.66%	11.2%	-9.75%	-5.97%	-2.03%	3.86%	5.98%
			50	822	-2.19%	-2.03%	7.48%	10.8%	-9.75%	-5.97%	-2.03%	3.86%	5.98%
			Unlim	822	-2.24%	-2.03%	7.42%	10.7%	-9.75%	-5.97%	-2.03%	3.86%	5.98%
Portfolio	Universal	Actual	Unadj	829	-3.15%	0.59%	7.74%	7.26%	-13.9%	-12.0%	0.59%	3.61%	7.69%
			20	829	-2.10%	-1.34%	3.03%	4.21%	-9.82%	-2.73%	-1.34%	0.87%	1.79%
			30	829	-2.10%	-1.34%	3.03%	4.21%	-9.82%	-2.73%	-1.34%	0.87%	1.79%
			40	829	-2.10%	-1.34%	3.03%	4.21%	-9.82%	-2.73%	-1.34%	0.87%	1.79%
			50	829	-2.10%	-1.34%	3.03%	4.21%	-9.82%	-2.73%	-1.34%	0.87%	1.79%
			Unlim	829	-2.10%	-1.34%	3.03%	4.21%	-9.82%	-2.73%	-1.34%	0.87%	1.79%
		Simulated	Unadj	822	2.05%	1.76%	6.67%	9.02%	-8.33%	-2.46%	1.76%	6.66%	18.9%
			20	822	0.77%	1.78%	3.29%	3.24%	-3.80%	-2.64%	1.78%	3.75%	4.90%
			30	822	0.51%	0.66%	3.12%	3.09%	-3.80%	-2.64%	0.66%	3.75%	4.90%
			40	822	0.51%	0.66%	3.12%	3.09%	-3.80%	-2.64%	0.66%	3.75%	4.90%
			50	822	0.51%	0.66%	3.12%	3.09%	-3.80%	-2.64%	0.66%	3.75%	4.90%
			Unlim	822	0.51%	0.66%	3.12%	3.09%	-3.80%	-2.64%	0.66%	3.75%	4.90%

6.   CONCLUSIONS AND RECOMMENDATIONS

The primary conclusions of this baseline study are the following:

- The accuracy and bias measures for the aggregated baseline methods can vary substantially across utilities, programs, and notice types, suggesting that baseline errors depend on a number of factors other than baseline type (*e.g.*, adjusted 10-in-10). These likely include the characteristics of the particular customers who participate in the programs and the nature of the events.
- The portfolio baselines, which are calculated on the basis of aggregations of the loads of individual customer accounts into their portfolio category (*e.g.*, aggregator and notice type) are often somewhat more accurate and have smaller biases than do aggregations of individual customer baselines. However, in cases where the aggregated baseline errors are relatively large and numerous, such as PG&E and SDG&E CBP DA, the portfolio baselines do not improve the results much, at least without baseline adjustments. Of some note, even in cases of large average percent errors, the typical bias as measured by the median percent error was quite small; however, the range of errors could be quite large.
- Allowing some day-of adjustment to the 10-in-10 baseline often improves accuracy and reduces bias somewhat, but unlimited adjustments often cause baseline errors to increase relative to even a 50 percent cap. However, in this study, there are few clear patterns of the degree of improvement in performance under different adjustment cap restrictions. For many of the programs, caps above 20 or 30 percent were not binding.

Based on the above findings, we can offer the following recommendations, some of which reiterate the recommendations in the baseline study conducted in conjunction with the 2011 aggregator evaluation:

- The results of this study suggest that using *portfolio baselines* rather than aggregations of *individual customer baselines* could improve the accuracy and reduce the bias of aggregator settlement baselines. However, the improvements are typically not dramatic, particularly in cases where aggregated baseline errors are large.
- Allowing some day-of adjustment nearly always improves baseline performance, which suggests making an adjusted baseline the default settlement baseline.
- With few exceptions, baseline caps of greater than 20 or 30 percent were generally not binding. We agree with last year's recommendation that higher caps are likely only needed in cases where customers' loads are quite variable, which suggests that no baseline method can accurately reflect their load.