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

2013 Statewide Load Impact Evaluation of California Aggregator Demand Response Programs
Volume 2: Baseline Analysis

CALMAC Study ID PGE0341.02

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

April 1, 2014

Table of Contents

Tables

CA Energy Consulting

ABSTRACT

This second volume of the aggregator program evaluation report for program-year 2013 contains an analysis of the performance of the current 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 baseline performance under conditions such as: 1) using aggregations of *individual customer* baselines, compared to baselines for *portfolios* of customers; 2) using a range of caps on day-of percentage *baseline adjustments*; and 3) assuming that all baselines are subject to day-of adjustments, rather than just those for which customers actually selected the adjustment.

Baseline performance is measured by both how *accurate* the method is, regardless of the sign of baseline errors (where baseline errors are defined as the difference between a given calculated baseline and the "true" baseline), and whether the baseline method tends to be *biased* upward or downward (*i.e.*, the baseline method tends to under-state or over-state the true baseline).

Two key conclusions of this baseline study are the following:
- Similarly to previous baseline studies, the accuracy and bias measures for the aggregated customer 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 and adjustment cap, such as the load variability of the particular customers participating in the programs.
- Allowing some day-of adjustment to the 10-in-10 baseline nearly always improves accuracy and reduces bias, but unlimited adjustments often cause baseline errors to increase relative to even a 50 percent cap.

EXECUTIVE SUMMARY

This second volume of the aggregator program evaluation report for program-year 2013 contains an analysis of the performance of the current 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 for *portfolios* of customers (*e.g.,* loads for all of the customer accounts nominated in one aggregator's CBP DO notice portfolio are first summed, and baselines are constructed from the portfolio load);
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 based on customers' actual selection of the day-of baseline adjustment option, to a case where baselines are adjusted for *all* enrolled and nominated customers (*i.e.,* assuming that all customers in a given portfolio selected the adjustment).

The baseline analysis was conducted for both *actual* event days in 2013, as well as a set of event-like non-event days, or *simulated* events. 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 2013 *ex-post* load impact evaluation (*i.e.,* estimated load impacts are added 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, regardless of the sign of baseline errors (where baseline errors are defined as the difference between a given calculated baseline and the "true" baseline), and whether the baseline method tends to be *biased* upward or downward (*i.e.*, the baseline method tends to under-state or over-state the true baseline). The *accuracy* metric used in this study is the median of the absolute values of the percentage baseline errors over all events and customers. To measure *bias*, the study reports values of median percent errors. Percentile values of percent errors, which reflect the distribution of errors around the median, are also shown.

ES.2 Summary of Study Findings

The primary conclusions of this baseline study are the following:

- Similarly to previous baseline studies, the accuracy and bias measures for the aggregated customer 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) and adjustment cap. These factors likely include the characteristics of the particular customers who participate in the programs and the nature of the events included in the study (*e.g.*, whether weather conditions on prior days that make up the baseline calculation were substantially milder or hotter than the event day).

- As a corollary, few common patterns emerge in terms of particular baseline types or cap restrictions producing consistent upward or downward biases.

- Several programs and notice types, such as SDG&E CBP DO and PG&E AMP DA and DO, produced relatively low errors, both in terms of median values and a relatively tight range of errors.

- Allowing some day-of adjustment to the 10-in-10 baseline nearly always improves accuracy and reduces bias (*i.e.*, the un-adjusted option frequently produces the least accurate results within a category, such as aggregated Universal adjustments), 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.

1.   INTRODUCTION AND OBJECTIVES OF THE STUDY

This second volume of the aggregator evaluation report for program-year 2013 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."

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 an aggregator's product portfolio (*e.g.,* DA 1-4), where the baseline uses the 10-in-10 method. 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 hours-ending (HE) 12 to 19.
- AMP – Program baselines are agreed upon within the aggregator contracts. Day-of adjustments are determined monthly by the aggregators at the time of nomination.

The baseline analysis in this study focuses on differences in baseline performance under the following conditions:

- *Level of aggregation* – 1) Baselines are constructed as aggregations of *individual customer* baselines, and 2) baselines are constructed from loads aggregated 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);
- *Selection of adjustment* – Baselines are constructed under two alternative scenarios: 1) using customers' actual choice of selecting or not selecting the day-of baseline adjustment, and 2) applying day-of adjustments to baselines for *all* customers (*i.e.,* assuming that all customers in a portfolio selected the adjustment); and
- *Alternative adjustment caps* – Baselines are constructed using a range of alternative caps on day-of percentage *baseline adjustments* (*e.g.,* cap percentages of 20, 30, 40, and 50, plus unlimited adjustments).

The baseline analysis was conducted for both *actual* event days in 2013, as well as a set of event-like non-event days, or *simulated* events. 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 2013 ex post load impact evaluation (*i.e.,* the estimated reference loads). In the case of *simulated* events, the observed loads on the event-like days serve as "true" baselines, to which all of the relevant alternative baseline methods are compared.

2.   METHODOLOGY

Performance of the alternative baseline methods was measured by statistics that characterize the baselines' *accuracy* and *bias*. The performance measures are calculated on the basis of the average across the event-hours of each event day for each customer service account. That is, the observations used in constructing the performance statistics represent outcomes for a customer's event day.[1] The reported statistics combine information across customers of various types, and events.

Baseline **bias** measures the tendency for a given baseline method to over-state or under-state the "true" baseline. Bias was measured primarily using the *median percentage error ("MPE")*, where the percentage error is defined as the *difference* between the baseline measure in question and the true baseline load (*i.*e., the regression-based baseline for *actual* event days, or the observed load for *simulated* event days), divided by the *level* of the true baseline (see equation below).[2] A positive MPE thus indicates an *upward* bias (or a tendency to *overstate* customers' load reductions), while a negative MPE indicates a *downward* bias (or a tendency to *understate* customers' load reductions). Note that the term "MPE" is typically used to refer to "*mean* percentage error." In this study, we use the median in place of the mean due to occasional large outliers (*e.g.*, percentage errors in excess of 100 percent, due to very low observed loads during the hours in question) that limit the usefulness of mean values.

The percentage error for each customer-event day is calculated as follows:

$$Percentage\ error = (L^P_d - L^A_d) / L^A_d,$$

where in this case,

L^P_d is one of the alternative *predicted* (program) average baseline load for customer-event day *d*; and

---

[1] Baseline errors for hours within an event for a given customer tend to take on quite similar values, so that the average across event hours provides a useful summary of baseline performance for a given customer-event combination.

[2] In response to a request from SDG&E, we also report values of the performance statistics in levels rather than percentages (*e.g.*, median error).

L^A_d is the "true" (based on regression results or observed loads) baseline load on customer-event day *d.*

The median percentage error (MPE) represents the 50th percentile value across the total number (*n*) of observations (*i.e.*, the number of customer-event days) of percent errors. The tables below also report *percentiles* of errors, which provide a more comprehensive picture of the full distribution of baseline errors than median values alone. The percentiles reported are 10, 25, 50, 75, and 90 percentile values, where the 50th percentile value is the median.

Baseline **accuracy** (relative to the true baseline) measures the degree of difference, or error, *regardless of sign*, between two data series, which in this case are the alternative baselines and the true baseline. We use the *median absolute percentage error* (MAPE) statistic to characterize baseline accuracy.[3] It is calculated as the median of the *absolute values* of the percentage errors across customer-event days, where percentage errors are again defined in the formula above. The use of absolute values of the errors means that positive and negative errors do not offset each other as they do in the MPE statistic.

This study also reports measures of the *levels* of baseline accuracy and bias in addition to the *percentage* error versions of the statistics.

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
  - *Aggregated* (*i.e.*, summations of individual customer-level baselines); and
  - *Portfolio* (*i.e.*, 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* (applies day-of adjustments to customer-level or portfolio baselines under the assumption that all baselines are subject to adjustment)
- Event type (actual or simulated)
- Adjustment cap (Unadjusted, 20, 30, 40 and 50 percent, and Unlimited)

---

[3] As with MPE, the term "MAPE" is typically used to refer to mean, rather than median absolute percentage errors. However, the existence of potentially large outlier values affects the ability to usefully interpret mean values for both measures.

For each combination of factors, the following statistics are provided:
- Number of customer accounts (also the percentage of customers who chose the day-of adjustment option)
- Accuracy statistic (MAPE)
- Bias statistic (MPE)
- Percentiles of percent errors (10, 25, 50, 75, and 90 percentiles)

3.1 PG&E CBP

Table 3–1 reports baseline performance statistics for PG&E's CBP DO program.[4] As a guide to the reader, this section contains a detailed description of the three main panels. Later sections describe only highlights of results. The first panel (Aggregated/As Chosen) contains results for baselines that are constructed by summing calculated values of individual customer baselines to an aggregate level, where the individual baselines are adjusted only for those customers who chose the adjustment option during the 2013 program year, and by which cap is applied. Two subpanels report results by event type (Actual and Simulated).

The first column, labeled Cust. Count, identifies the number of customer accounts included in the calculations of that section. Only customers who were nominated for at least one event are included. Due to differences in the dates of the actual and simulated events, the counts of customers may differ between those event types (*e.g.*, for a simulated event in June, the number of nominated customers may be less than for actual events later in the summer). No values are shown in the first line (Unadj.), since at least some customers chose adjustments.[5] All rows in the "As Chosen" panel 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. Nearly all (97 percent) of CBP DO customers selected the baseline adjustment option.

In the first major panel of "As Chosen" results, the "Accuracy" column indicates MAPE values of 2 to 2.7 percent for actual events and 3 to 3.5 percent for simulated events. Both vary somewhat across cap values. The "Bias" column indicates that the alternative baseline methods tend to understate the true baselines by less than one percent on actual event days and overstate the true baseline by 2 percent or less on simulated event days, again varying slightly across adjustment caps. The percentiles of errors indicate that twenty-five percent of the errors fall below -4 percent for the actual events. At the other end of the distribution, 10 percent of customers (90th percentile) have errors greater than 10 percent with a 30 percent adjustment cap.

---

[4] Results are not shown for CBP DA due to the small number (24) of enrolled customer accounts.
[5] The first line of the "Universal" panel shows results for the case in which all customers are assumed to have unadjusted baselines.

The middle panel (Aggregated/Universal) results represent the case in which each customer's baseline receives a day-of adjustment, varying from no adjustment, through the various caps, to unlimited adjustment. Not unexpectedly, given that most customers selected the baseline adjustment, values in this panel are quite similar to those in the "As Chosen" panel. The primary exceptions are 1) the Un-adjusted case shown in the first line, showing that the baseline performance results are considerably worse for these customers without some day-of adjustment, and 2) an increase in the number of understated baselines (see 10th and 25th percentile results) with high adjustment caps on 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. Both the MAPE and MPE values under this baseline method are roughly comparable to the aggregations of individual customer baselines for the actual events. The MPE values show relatively low downward bias for actual events and less than 2 percent upward bias for the simulated events. However, the 10th and 90th percentiles show that ten percent of aggregator/event percent errors can be relatively high for any adjustment cap.

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

Load Type	Adj. Choice	Event Type	Cap	Cust. Count	Accuracy MAPE	Bias MPE	Percentiles of % Errors p10	p25	p50	p75	p90
Aggregated	As Chosen	Actual	Unadj								
			20	482	2.59%	0.25%	-3.12%	-1.46%	0.25%	3.51%	11.41%
			30	482	1.96%	-0.62%	-4.31%	-2.09%	-0.62%	1.42%	9.76%
			40	482	2.59%	-0.71%	-4.26%	-2.88%	-0.71%	0.51%	7.95%
			50	482	2.72%	-0.85%	-4.63%	-2.88%	-0.85%	0.31%	7.55%
			Unlim	482	2.72%	-0.95%	-6.70%	-3.32%	-0.95%	0.03%	7.55%
		Simulated	Unadj								
			20	483	3.48%	2.33%	-5.38%	-0.52%	2.33%	4.77%	8.66%
			30	483	3.23%	1.90%	-5.75%	-1.04%	1.90%	4.77%	8.55%
			40	483	3.12%	1.86%	-7.02%	-1.31%	1.86%	4.44%	8.31%
			50	483	3.06%	1.60%	-7.18%	-1.48%	1.60%	4.32%	8.31%
			Unlim	483	3.33%	1.16%	-7.82%	-1.93%	1.16%	4.32%	8.31%
Aggregated	Universal	Actual	Unadj	482	9.58%	9.43%	-2.89%	2.77%	9.43%	14.57%	18.01%
			20	482	3.34%	-0.14%	-7.99%	-3.33%	-0.14%	3.35%	9.81%
			30	482	2.36%	-0.62%	-11.25%	-4.12%	-0.62%	0.89%	7.19%
			40	482	2.53%	-0.71%	-11.64%	-4.12%	-0.71%	0.28%	5.39%
			50	482	2.54%	-0.86%	-13.43%	-4.12%	-0.86%	-0.03%	4.47%
			Unlim	482	2.64%	-1.72%	-16.28%	-4.94%	-1.72%	-0.54%	3.82%
		Simulated	Unadj	483	5.15%	4.49%	-3.12%	0.43%	4.49%	8.94%	13.42%
			20	483	3.59%	2.11%	-5.57%	-0.79%	2.11%	4.71%	7.93%
			30	483	3.21%	1.85%	-6.66%	-1.22%	1.85%	4.62%	7.68%
			40	483	3.37%	1.81%	-7.01%	-1.54%	1.81%	4.27%	7.65%
			50	483	3.20%	1.60%	-7.18%	-1.69%	1.60%	4.12%	6.89%
			Unlim	483	3.34%	1.02%	-7.98%	-2.24%	1.02%	4.02%	6.67%
Portfolio	Universal	Actual	Unadj	482	10.37%	9.86%	-6.02%	-1.38%	9.86%	17.18%	21.38%
			20	482	2.26%	-0.24%	-9.46%	-3.86%	-0.24%	2.05%	10.92%
			30	482	1.88%	-0.59%	-12.70%	-4.48%	-0.59%	0.88%	9.00%
			40	482	2.12%	-0.70%	-12.70%	-5.48%	-0.70%	0.48%	9.00%
			50	482	2.12%	-0.70%	-12.70%	-5.48%	-0.70%	0.48%	9.00%
			Unlim	482	2.12%	-0.70%	-12.70%	-5.48%	-0.70%	0.48%	9.00%
		Simulated	Unadj	483	5.57%	4.56%	-3.20%	1.39%	4.56%	9.12%	13.91%
			20	483	3.50%	1.91%	-7.57%	-1.81%	1.91%	4.29%	7.93%
			30	483	3.50%	1.76%	-7.58%	-1.85%	1.76%	4.22%	7.93%
			40	483	3.50%	1.76%	-8.03%	-1.85%	1.76%	4.09%	7.40%
			50	483	3.54%	1.56%	-8.03%	-2.01%	1.56%	4.09%	7.40%
			Unlim	483	3.54%	1.56%	-8.03%	-2.01%	1.56%	4.09%	7.40%

3.2 SCE CBP

Table 3–2 shows baseline performance for SCE's CBP DO portfolio.[6] Overall accuracy under the "As Chosen" scenario is relatively poor, with MAPE values of 6 to 7 percent for both actual and simulated events, across all cap levels. The baselines are also biased downward by about 2 to 3 percent for the actual events, and upward by less than 1 percent for the simulated events. The percentile results show that in ten percent of cases for actual events, the baselines are biased downward by 10 percent or more, while in another ten percent of cases, the baselines are biased upward by 6 to 8 percent. There is little pattern suggesting an optimal adjustment cap, although some adjustment nearly always improves performance. Eighty-seven percent of the CBP DO customer accounts selected the day-of adjustment option.

The performance of the portfolio option appears similar to the aggregations of individual baselines.

---

[6] Results are not shown for CBP DA due to the small number (20) of enrolled customer accounts.

Table 3-2: Baseline Performance – SCE CBP DO

Load Type	Adj. Choice	Event Type	Cap	Cust. Count	Accuracy MAPE	Bias MPE	p10	p25	p50	p75	p90
Aggregated	As Chosen	Actual	Unadj								
			20	426	6.1%	-2.1%	-13.0%	-4.0%	-2.1%	-0.4%	6.0%
			30	426	6.7%	-2.8%	-13.0%	-5.8%	-2.8%	-1.2%	7.9%
			40	426	7.0%	-3.3%	-13.0%	-6.6%	-3.3%	-1.3%	8.3%
			50	426	7.2%	-3.3%	-13.0%	-6.7%	-3.3%	-1.3%	8.3%
			Unlim	426	7.3%	-3.3%	-13.0%	-6.7%	-3.3%	-1.4%	8.3%
		Simulated	Unadj								
			20	422	6.1%	0.8%	-5.5%	-1.3%	0.8%	3.7%	9.9%
			30	422	6.2%	0.7%	-6.1%	-2.0%	0.7%	3.9%	9.8%
			40	422	6.2%	0.6%	-6.2%	-2.4%	0.6%	3.9%	9.8%
			50	422	6.2%	0.6%	-6.2%	-2.7%	0.6%	3.9%	9.2%
			Unlim	422	5.8%	0.5%	-7.4%	-3.1%	0.5%	3.4%	8.6%
Aggregated	Universal	Actual	Unadj	426	7.1%	2.2%	-9.5%	-3.6%	2.2%	7.8%	11.1%
			20	426	5.8%	-1.5%	-6.5%	-3.7%	-1.5%	2.4%	6.0%
			30	426	6.4%	-2.3%	-7.7%	-4.6%	-2.3%	1.7%	5.4%
			40	426	6.8%	-2.5%	-9.6%	-6.2%	-2.5%	1.0%	5.4%
			50	426	7.0%	-3.0%	-10.6%	-6.7%	-3.0%	1.0%	5.4%
			Unlim	426	7.3%	-3.4%	-11.3%	-8.5%	-3.4%	1.0%	5.4%
		Simulated	Unadj	422	7.3%	3.0%	-5.2%	-0.5%	3.0%	6.6%	12.0%
			20	422	5.7%	1.0%	-5.3%	-2.0%	1.0%	4.0%	7.7%
			30	422	5.9%	0.8%	-6.1%	-2.4%	0.8%	4.1%	7.8%
			40	422	6.0%	0.6%	-6.4%	-2.6%	0.6%	4.0%	7.9%
			50	422	6.0%	0.6%	-6.5%	-2.9%	0.6%	3.9%	7.9%
			Unlim	422	5.8%	0.0%	-7.9%	-3.9%	0.0%	3.6%	6.9%
Portfolio	Universal	Actual	Unadj	426	7.1%	2.8%	-9.5%	-3.6%	2.8%	7.8%	11.1%
			20	426	6.6%	-2.3%	-9.1%	-5.1%	-2.3%	1.8%	5.7%
			30	426	6.4%	-2.3%	-10.5%	-5.1%	-2.3%	1.0%	5.7%
			40	426	6.5%	-3.1%	-10.5%	-5.1%	-3.1%	1.0%	5.7%
			50	426	6.5%	-3.1%	-10.5%	-5.1%	-3.1%	1.0%	5.7%
			Unlim	426	6.5%	-3.1%	-10.5%	-5.1%	-3.1%	1.0%	5.7%
		Simulated	Unadj	422	9.9%	2.8%	-6.9%	-0.8%	2.8%	7.3%	14.0%
			20	422	7.4%	1.0%	-7.0%	-2.6%	1.0%	4.1%	7.8%
			30	422	7.0%	0.8%	-7.3%	-2.6%	0.8%	4.0%	7.8%
			40	422	6.9%	0.9%	-6.9%	-2.7%	0.9%	4.1%	7.8%
			50	422	6.8%	1.0%	-6.7%	-2.4%	1.0%	4.2%	7.8%
			Unlim	422	5.8%	0.8%	-7.0%	-2.7%	0.8%	4.0%	7.5%

CA Energy Consulting

3.3 SDG&E CBP

Table 3–3 shows baseline performance for SDG&E's CBP DA portfolio. MAPE values of about 5.5 percent for actual events indicate that overall accuracy under the "As Chosen" scenario is somewhat worse than in the Universal scenario. The bias results are somewhat mixed, with upward bias of about 2 to 3 percent for the actual events under As Chosen, and downward bias in the same amounts for the Universal option. At the same time, neither option is biased for the simulated events. The percentile results show that in ten percent of cases, the baselines are biased downward by 10 to 20 percent, while in another ten percent of cases, the baselines are biased upward by at least 9 to 10 percent.

Half of the DA customer accounts selected the day-of adjustment option, while 76 percent of DO customers did so.

The portfolio approach produces similar MAPE values to the aggregated baselines. However, the bias results indicate a greater downward bias for actual events, although similar low bias for simulated events. Like the aggregated baselines, downward biases can be substantial for 10 percent of cases. Results are generally not very sensitive to the alternative adjustment caps, other than the Unadjusted and Unlimited cap cases, which generally fare worst.

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

Load Type	Adj. Choice	Event Type	Cap	Cust. Count	Accuracy MAPE	Bias MPE	Percentiles p10	p25	p50	p75	p90
Aggregate	As Chosen	Actual	Unadj								
			20	140	5.5%	3.7%	-16.4%	-0.3%	3.7%	6.0%	9.1%
			30	140	5.5%	2.8%	-20.1%	-1.9%	2.8%	6.0%	9.1%
			40	140	5.5%	2.8%	-21.7%	-2.8%	2.8%	6.0%	9.1%
			50	140	5.5%	2.7%	-23.0%	-3.0%	2.7%	6.0%	9.1%
			Unlim	140	5.7%	2.5%	-23.7%	-3.0%	2.5%	5.8%	9.1%
		Simulated	Unadj								
			20	135	4.2%	0.3%	-9.5%	-4.0%	0.3%	4.6%	10.5%
			30	135	4.2%	0.3%	-10.3%	-4.0%	0.3%	4.5%	10.5%
			40	135	4.2%	-0.1%	-11.2%	-4.2%	-0.1%	4.5%	10.5%
			50	135	4.2%	-0.1%	-11.5%	-4.2%	-0.1%	4.5%	10.5%
			Unlim	135	4.4%	-0.1%	-11.5%	-4.3%	-0.1%	4.5%	10.5%
Aggregate	Universal	Actual	Unadj	140	8.2%	6.0%	-16.4%	2.1%	6.0%	10.1%	13.0%
			20	140	3.2%	-1.5%	-14.4%	-4.4%	-1.5%	1.9%	4.1%
			30	140	3.3%	-1.9%	-20.1%	-4.4%	-1.9%	0.7%	4.1%
			40	140	3.2%	-2.4%	-21.7%	-4.8%	-2.4%	0.4%	4.1%
			50	140	4.0%	-2.6%	-30.0%	-5.0%	-2.6%	0.4%	5.8%
			Unlim	140	5.3%	-3.7%	-517%	-8.5%	-3.7%	0.4%	5.7%
		Simulated	Unadj	135	4.8%	2.7%	-9.4%	-1.1%	2.7%	6.8%	11.9%
			20	135	4.6%	0.1%	-7.9%	-4.4%	0.1%	4.6%	6.8%
			30	135	4.4%	0.5%	-7.8%	-4.0%	0.5%	4.6%	6.9%
			40	135	4.6%	0.0%	-8.2%	-4.7%	0.0%	4.6%	9.7%
			50	135	4.8%	-0.1%	-8.5%	-4.7%	-0.1%	5.1%	16.0%
			Unlim	135	5.6%	-1.2%	-13.5%	-7.4%	-1.2%	4.6%	26.3%
Portfolio	Universal	Actual	Unadj	140	8.4%	6.5%	-16.4%	2.1%	6.5%	10.2%	14.1%
			20	140	4.8%	-4.2%	-18.2%	-6.7%	-4.2%	1.0%	6.2%
			30	140	5.4%	-4.3%	-27.0%	-7.2%	-4.3%	0.4%	5.6%
			40	140	5.2%	-4.3%	-27.0%	-7.2%	-4.3%	0.3%	5.6%
			50	140	5.6%	-4.3%	-30.5%	-8.5%	-4.3%	0.3%	5.6%
			Unlim	140	5.6%	-4.3%	-213%	-8.5%	-4.3%	0.3%	5.6%
		Simulated	Unadj	135	4.7%	2.4%	-9.4%	-1.1%	2.4%	6.8%	12.0%
			20	135	4.6%	0.0%	-8.2%	-6.0%	0.0%	4.1%	11.3%
			30	135	5.2%	-0.2%	-10.3%	-6.1%	-0.2%	4.1%	11.3%
			40	135	5.2%	-0.2%	-10.3%	-6.1%	-0.2%	4.1%	11.3%
			50	135	5.2%	-0.2%	-10.3%	-6.1%	-0.2%	4.1%	11.3%
			Unlim	135	4.8%	-0.6%	-10.3%	-6.1%	-0.6%	3.7%	8.5%

Table 3–4 reports results for CBP DO. Baseline accuracy for this notice type is generally somewhat better than for the DA case. For the actual events, the aggregations of individual baselines produce MAPE values of around 2 to 3 percent, with little variation across adjustment caps, except for Unadjusted and Unlimited cases. The MPE values indicate less than 1 percent upward bias for As Chosen, and even less downward bias for Universal adjustments. The 10/90 tails of the distribution of errors do not show evidence of extreme errors for these customers.

For the simulated events, MAPE values are fairly similar across the three baseline methods, at around 4 percent. The MPE values indicate modest upward bias of 1 to 2 percent, with no evidence of extreme errors.

The portfolio baselines have essentially no bias for actual events and less than 2 percent upward bias for simulated events. There is no clear trend across adjustment caps, except that no adjustment produces the worst baseline performance.

Table 3–4: Baseline Performance – SDG&E CBP DO

Load Type	Adj. Choice	Event Type	Cap	Cust. Count	Accuracy MAPE	Bias MPE	Percentiles p10	p25	p50	p75	p90
Aggregate	As Chosen	Actual	Unadj								
			20	260	2.88%	0.87%	-3.70%	-0.53%	0.87%	5.26%	9.44%
			30	260	3.29%	0.46%	-6.09%	-1.69%	0.46%	4.63%	9.44%
			40	260	3.45%	0.46%	-6.34%	-1.90%	0.46%	3.99%	9.44%
			50	260	3.23%	0.46%	-6.34%	-1.93%	0.46%	3.91%	9.44%
			Unlim	260	3.03%	0.21%	-6.59%	-2.07%	0.21%	3.91%	9.44%
		Simulated	Unadj								
			20	254	3.90%	2.04%	-4.58%	-0.36%	2.04%	5.75%	9.23%
			30	254	4.16%	2.04%	-5.51%	-1.04%	2.04%	5.69%	9.22%
			40	254	4.35%	2.04%	-5.65%	-1.04%	2.04%	5.58%	9.23%
			50	254	4.49%	2.04%	-5.65%	-1.04%	2.04%	5.58%	9.23%
			Unlim	254	4.64%	2.04%	-5.65%	-1.21%	2.04%	5.58%	9.23%
Aggregate	Universal	Actual	Unadj	260	8.03%	7.78%	0.41%	4.32%	7.78%	9.47%	11.40%
			20	260	1.90%	-0.26%	-3.70%	-1.86%	-0.26%	1.90%	4.98%
			30	260	1.82%	0.11%	-5.59%	-2.09%	0.11%	1.09%	5.14%
			40	260	1.86%	-0.07%	-5.87%	-2.41%	-0.07%	1.78%	5.68%
			50	260	1.93%	-0.40%	-5.99%	-2.37%	-0.40%	1.68%	6.29%
			Unlim	260	2.21%	-0.81%	-6.77%	-3.08%	-0.81%	1.08%	3.03%
		Simulated	Unadj	254	3.53%	2.85%	-2.66%	0.29%	2.85%	7.88%	11.18%
			20	254	3.74%	1.57%	-6.09%	-1.13%	1.57%	4.44%	5.79%
			30	254	3.74%	1.54%	-6.48%	-1.55%	1.54%	4.38%	5.69%
			40	254	3.90%	1.21%	-6.48%	-1.79%	1.21%	4.56%	5.94%
			50	254	3.97%	1.31%	-6.17%	-1.77%	1.31%	4.67%	6.05%
			Unlim	254	4.56%	1.59%	-6.48%	-1.77%	1.59%	5.25%	7.56%
Portfolio	Universal	Actual	Unadj	260	7.22%	6.74%	0.41%	4.14%	6.74%	9.79%	13.97%
			20	260	2.12%	0.00%	-5.28%	-2.12%	0.00%	1.90%	8.14%
			30	260	2.12%	0.00%	-6.41%	-2.12%	0.00%	1.90%	6.60%
			40	260	2.12%	-0.06%	-6.41%	-2.23%	-0.06%	1.86%	6.60%
			50	260	2.12%	-0.06%	-6.41%	-2.35%	-0.06%	1.86%	6.60%
			Unlim	260	2.12%	-0.06%	-6.41%	-2.35%	-0.06%	1.86%	6.60%
		Simulated	Unadj	254	3.52%	2.78%	-2.66%	-0.15%	2.78%	7.88%	11.10%
			20	254	4.46%	1.70%	-6.87%	-1.99%	1.70%	5.01%	6.81%
			30	254	4.46%	1.70%	-6.87%	-1.99%	1.70%	5.01%	6.81%
			40	254	4.46%	1.70%	-6.87%	-1.99%	1.70%	5.01%	6.81%
			50	254	4.34%	1.70%	-6.25%	-1.52%	1.70%	5.01%	6.81%
			Unlim	254	4.46%	1.76%	-6.25%	-1.52%	1.76%	5.16%	7.05%

Tables 3–5 and 3–6 show the same baseline performance information, but in *levels* of MW of baseline errors rather than *percentage* values (*i.e.*, without dividing the baseline errors by the true baseline). As expected, the patterns of results are similar to those in the tables of percentage errors. In Table 3–5, for CBP DA, the median error for the Aggregated/As Chosen scenario for the actual events is 116 kW (0.116 MW) under a 40 percent cap (third row of values in the table), and the 10/90 percentiles show a range from a downward bias of 183 kW to an upward bias of 1.7 MW. Median errors are smaller for the simulated events.

Portfolio baselines show small downward biases of 28 kW for actual events and 5 kW for simulated events. As is typically the case, no adjustment produces the largest errors.

Table 3–5: Baseline Performance (in Levels of MW) – SDG&E CBP DA

Load Type	Adj. Choice	Event Type	Cap	Cust. Count	Accuracy Median Abs. Err.	Bias Median Error	Percentiles of % Errors				
							p10	p25	p50	p75	p90
Aggregated	As Chosen	Actual	Unadj								
			20	142	0.143	0.118	-0.012	0.026	0.118	0.452	1.717
			30	142	0.166	0.116	-0.183	0.019	0.116	0.378	1.717
			40	142	0.171	0.116	-0.183	0.019	0.116	0.337	1.717
			50	142	0.174	0.116	-0.214	0.019	0.116	0.321	1.717
			Unlim	142	0.174	0.116	-0.215	0.019	0.116	0.314	1.717
		Simulated	Unadj								
			20	135	0.068	0.023	-0.276	-0.040	0.023	0.176	0.764
			30	135	0.087	0.020	-0.401	-0.054	0.020	0.151	0.649
			40	135	0.095	0.020	-0.459	-0.065	0.020	0.139	0.551
			50	135	0.091	0.019	-0.594	-0.071	0.019	0.094	0.549
			Unlim	135	0.100	0.019	-0.870	-0.105	0.019	0.094	0.549
Aggregated	Universal	Actual	Unadj	142	0.174	0.143	0.018	0.057	0.143	1.552	1.921
			20	142	0.093	0.007	-0.691	-0.059	0.007	0.107	0.465
			30	142	0.088	-0.011	-0.930	-0.097	-0.011	0.057	0.426
			40	142	0.078	-0.021	-0.497	-0.165	-0.021	0.031	0.409
			50	142	0.078	-0.021	-0.512	-0.157	-0.021	0.012	0.395
			Unlim	142	0.350	-0.078	-1.813	-0.856	-0.078	0.007	0.384
		Simulated	Unadj	135	0.116	0.035	-0.071	-0.001	0.035	0.628	1.312
			20	135	0.088	0.001	-0.587	-0.074	0.001	0.102	0.378
			30	135	0.124	0.000	-0.765	-0.160	0.000	0.069	0.227
			40	135	0.127	0.003	-0.851	-0.330	0.003	0.072	0.159
			50	135	0.122	-0.003	-0.909	-0.401	-0.003	0.069	0.157
			Unlim	135	0.173	-0.060	-1.333	-0.767	-0.060	0.069	0.151
Portfolio	Universal	Actual	Unadj	142	0.174	0.167	0.021	0.071	0.167	1.760	2.097
			20	142	0.115	-0.028	-0.647	-0.247	-0.028	0.086	0.332
			30	142	0.119	-0.028	-0.647	-0.247	-0.028	0.057	0.332
			40	142	0.147	-0.028	-0.647	-0.247	-0.028	0.033	0.332
			50	142	0.169	-0.031	-0.647	-0.254	-0.031	0.014	0.332
			Unlim	142	0.279	-0.078	-1.342	-0.445	-0.078	0.004	0.332
		Simulated	Unadj	135	0.115	0.035	-0.071	-0.001	0.035	0.611	1.312
			20	135	0.098	0.003	-1.111	-0.191	0.003	0.072	0.148
			30	135	0.119	-0.005	-1.131	-0.343	-0.005	0.062	0.130
			40	135	0.119	-0.005	-1.272	-0.343	-0.005	0.056	0.140
			50	135	0.119	-0.005	-1.272	-0.343	-0.005	0.056	0.140
			Unlim	135	0.110	-0.015	-1.272	-0.343	-0.015	0.052	0.130

Table 3–6 shows similar results for CBP DO. For the As Chosen scenario, the median error is near zero for actual events and about 200 kW for simulated events. Results for the portfolio approach are qualitatively similar to the aggregations of individual baselines. No adjustment cap dominates, but some adjustment is always better than no adjustment.

Table 3–6: Baseline Performance (in Levels of MW) – SDG&E CBP DO

Load Type	Adj. Choice	Event Type	Cap	Cust. Count	Accuracy Median Abs. Err.	Bias Median Error	Percentiles of % Errors p10	p25	p50	p75	p90
Aggregated	As Chosen	Actual	Unadj								
			20	260	0.166	0.019	-0.211	-0.111	0.019	0.295	0.572
			30	260	0.259	0.002	-0.352	-0.204	0.002	0.295	0.546
			40	260	0.272	0.002	-0.392	-0.199	0.002	0.295	0.548
			50	260	0.274	0.002	-0.396	-0.195	0.002	0.295	0.551
			Unlim	260	0.223	-0.004	-0.498	-0.195	-0.004	0.249	0.408
		Simulated	Unadj								
			20	255	0.235	0.212	-0.162	0.000	0.212	0.648	1.383
			30	255	0.262	0.206	-0.270	-0.009	0.206	0.648	1.209
			40	255	0.265	0.204	-0.277	-0.009	0.204	0.648	1.147
			50	255	0.265	0.202	-0.277	-0.009	0.202	0.648	1.147
			Unlim	255	0.265	0.166	-0.326	-0.009	0.166	0.489	1.187
Aggregated	Universal	Actual	Unadj	260	0.539	0.539	0.012	0.273	0.539	1.450	2.391
			20	260	0.143	-0.009	-0.259	-0.162	-0.009	0.100	0.414
			30	260	0.191	-0.051	-0.329	-0.219	-0.051	0.064	0.404
			40	260	0.195	-0.058	-0.382	-0.230	-0.058	0.015	0.393
			50	260	0.190	-0.055	-0.378	-0.253	-0.055	0.015	0.393
			Unlim	260	0.199	-0.050	-0.623	-0.275	-0.050	0.015	0.298
		Simulated	Unadj	255	0.243	0.200	-0.106	0.018	0.200	0.688	1.913
			20	255	0.240	0.189	-0.304	-0.038	0.189	0.386	1.155
			30	255	0.249	0.166	-0.326	-0.031	0.166	0.297	0.833
			40	255	0.249	0.174	-0.326	-0.028	0.174	0.308	0.863
			50	255	0.236	0.164	-0.326	-0.027	0.164	0.297	0.900
			Unlim	255	0.265	0.118	-0.370	-0.094	0.118	0.355	1.065
Portfolio	Universal	Actual	Unadj	260	0.550	0.550	0.012	0.281	0.550	1.255	1.950
			20	260	0.181	-0.033	-0.405	-0.227	-0.033	0.028	0.442
			30	260	0.181	-0.033	-0.460	-0.227	-0.033	0.028	0.442
			40	260	0.181	-0.033	-0.460	-0.227	-0.033	0.028	0.442
			50	260	0.181	-0.033	-0.460	-0.227	-0.033	0.028	0.442
			Unlim	260	0.181	-0.033	-0.460	-0.227	-0.033	0.028	0.442
		Simulated	Unadj	255	0.240	0.202	-0.137	0.003	0.202	0.760	1.898
			20	255	0.246	0.208	-0.248	-0.024	0.208	0.409	1.166
			30	255	0.247	0.208	-0.315	-0.023	0.208	0.409	1.166
			40	255	0.247	0.144	-0.375	-0.024	0.144	0.369	1.002
			50	255	0.247	0.144	-0.375	-0.024	0.144	0.369	1.002
			Unlim	255	0.247	0.144	-0.375	-0.024	0.144	0.369	1.002

4. BASELINE PERFORMANCE – PG&E AMP

Table 4–1 reports baseline results for AMP DA. Accuracy and bias are quite low for all three of the baseline methods, with MAPE values of 1 to 2 percent, and positive MPE values of one to two percent in most cases.[7] As usual, the primary exceptions are the unadjusted baseline options.

For AMP DA, 76 percent of customer accounts selected the day-of adjustment option, while 89 percent of DO customers selected the adjustment.

---

[7] 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	Accuracy MAPE	Bias MPE	Percentiles p10	p25	p50	p75	p90
Aggregate	As Chosen	Actual	Unadj								
			20	435	2.57%	2.25%	-1.44%	0.24%	2.25%	5.68%	7.62%
			30	435	2.13%	1.77%	-1.58%	0.12%	1.77%	5.23%	7.13%
			40	435	1.84%	1.49%	-1.59%	-0.04%	1.49%	5.17%	7.08%
			50	435	1.82%	1.25%	-1.59%	-0.13%	1.25%	5.23%	7.32%
			Unlim	435	1.82%	1.20%	-1.59%	-0.76%	1.20%	5.71%	8.47%
		Simulated	Unadj								
			20	449	1.72%	1.37%	-1.64%	-0.22%	1.37%	3.82%	6.34%
			30	449	1.63%	1.13%	-1.68%	-0.28%	1.13%	3.38%	5.69%
			40	449	1.54%	0.96%	-1.49%	-0.28%	0.96%	2.97%	5.72%
			50	449	1.51%	0.86%	-1.31%	-0.28%	0.86%	2.81%	5.33%
			Unlim	449	1.37%	0.80%	-1.46%	-0.49%	0.80%	2.12%	4.62%
Aggregate	Universal	Actual	Unadj	435	6.08%	6.08%	1.79%	3.48%	6.08%	7.50%	10.36%
			20	435	2.65%	2.65%	-1.44%	0.28%	2.65%	5.68%	7.62%
			30	435	2.47%	2.36%	-1.57%	0.12%	2.36%	5.23%	7.13%
			40	435	2.46%	2.10%	-1.59%	-0.04%	2.10%	5.17%	7.08%
			50	435	2.46%	2.10%	-1.59%	-0.13%	2.10%	5.23%	7.32%
			Unlim	435	2.42%	2.06%	-1.59%	-0.79%	2.06%	5.71%	8.47%
		Simulated	Unadj	449	4.93%	4.59%	-1.65%	2.40%	4.59%	6.98%	9.47%
			20	449	1.67%	1.31%	-1.64%	-0.49%	1.31%	3.41%	5.02%
			30	449	1.61%	0.83%	-1.68%	-0.49%	0.83%	2.81%	4.39%
			40	449	1.49%	0.59%	-1.71%	-0.65%	0.59%	2.65%	3.82%
			50	449	1.38%	0.33%	-1.71%	-0.65%	0.33%	2.54%	3.71%
			Unlim	449	1.30%	0.12%	-2.51%	-0.74%	0.12%	1.66%	3.67%
Portfolio	Universal	Actual	Unadj	435	6.08%	6.08%	2.57%	3.36%	6.08%	9.13%	12.22%
			20	435	1.81%	1.16%	-1.59%	-0.64%	1.16%	3.73%	4.99%
			30	435	1.81%	1.16%	-1.59%	-0.64%	1.16%	3.73%	4.99%
			40	435	1.81%	1.16%	-1.59%	-0.64%	1.16%	3.73%	4.99%
			50	435	1.81%	1.16%	-1.59%	-0.64%	1.16%	3.73%	4.99%
			Unlim	435	1.81%	1.16%	-1.59%	-0.64%	1.16%	3.73%	4.99%
		Simulated	Unadj	448	4.74%	4.42%	-3.93%	2.20%	4.42%	7.19%	10.28%
			20	448	1.17%	0.52%	-1.65%	-0.52%	0.52%	1.97%	3.63%
			30	448	1.17%	0.52%	-1.65%	-0.52%	0.52%	1.97%	3.63%
			40	448	1.17%	0.52%	-1.65%	-0.52%	0.52%	1.97%	3.63%
			50	448	1.17%	0.52%	-1.65%	-0.52%	0.52%	1.97%	3.63%
			Unlim	448	1.17%	0.52%	-1.65%	-0.52%	0.52%	1.97%	3.63%

Table 4–2 reports results for AMP DO. Baseline performance values are similar to those for DA notice. MAPE values show median absolute errors of 1 to 3 percent, and biases are generally upward, but by one percent or less. The distributions of errors indicate relatively small errors out to the 10/90th percentiles. The bias results for AMP DO appear

to vary more across adjustment caps than other programs, generally improving across higher caps. For the portfolio option, MPE values indicate a downward bias of one percent for actual events and an upward bias of less than one percent for simulated events.

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

Load Type	Adj. Choice	Event Type	Cap	Cust. Count	Accuracy MAPE	Bias MPE	Percentiles p10	p25	p50	p75	p90
Aggregate	As Chosen	Actual	Unadj								
			20	1364	2.89%	2.59%	-2.40%	-1.08%	2.59%	4.74%	8.13%
			30	1364	2.42%	1.24%	-2.53%	-1.12%	1.24%	4.86%	7.55%
			40	1364	2.49%	0.47%	-2.73%	-1.07%	0.47%	4.83%	8.33%
			50	1364	2.59%	0.22%	-2.89%	-1.00%	0.22%	5.05%	8.67%
			Unlim	1364	2.68%	0.07%	-3.46%	-2.00%	0.07%	4.73%	8.02%
		Simulated	Unadj								
			20	1380	2.30%	1.64%	-1.40%	-0.19%	1.64%	4.09%	5.99%
			30	1380	2.22%	1.46%	-1.59%	-0.14%	1.46%	3.90%	6.22%
			40	1380	2.12%	1.29%	-1.52%	-0.09%	1.29%	3.82%	6.50%
			50	1380	2.18%	1.20%	-1.47%	0.03%	1.20%	3.79%	6.54%
			Unlim	1380	1.91%	1.00%	-1.93%	-0.23%	1.00%	3.54%	5.87%
Aggregate	Universal	Actual	Unadj	1364	7.82%	7.82%	-0.07%	2.15%	7.82%	11.90%	15.11%
			20	1364	2.38%	1.59%	-2.40%	-1.21%	1.59%	2.76%	5.81%
			30	1364	2.15%	0.46%	-2.64%	-1.56%	0.46%	2.55%	4.96%
			40	1364	1.98%	-0.01%	-2.73%	-1.69%	-0.01%	2.60%	4.93%
			50	1364	1.86%	-0.21%	-2.89%	-1.59%	-0.21%	2.55%	5.35%
			Unlim	1364	2.82%	-1.65%	-5.42%	-3.01%	-1.65%	1.11%	4.73%
		Simulated	Unadj	1380	4.58%	4.03%	-0.72%	1.43%	4.03%	6.37%	9.35%
			20	1380	1.74%	1.09%	-1.81%	-0.21%	1.09%	3.40%	4.77%
			30	1380	1.58%	0.61%	-2.00%	-0.57%	0.61%	2.46%	4.53%
			40	1380	1.46%	0.62%	-1.66%	-0.82%	0.62%	2.52%	4.52%
			50	1380	1.44%	0.73%	-1.62%	-0.53%	0.73%	2.58%	4.37%
			Unlim	1380	1.71%	0.52%	-2.11%	-0.59%	0.52%	2.22%	4.11%
Portfolio	Universal	Actual	Unadj	1364	8.75%	8.75%	0.69%	2.92%	8.75%	11.34%	14.92%
			20	1364	1.44%	-1.05%	-2.75%	-2.57%	-1.05%	-0.28%	2.10%
			30	1364	1.44%	-1.05%	-2.75%	-2.57%	-1.05%	-0.52%	2.10%
			40	1364	1.44%	-1.05%	-2.75%	-2.57%	-1.05%	-0.52%	2.10%
			50	1364	1.44%	-1.05%	-2.75%	-2.57%	-1.05%	-0.52%	2.10%
			Unlim	1364	1.44%	-1.05%	-2.75%	-2.57%	-1.05%	-0.52%	2.10%
		Simulated	Unadj	1377	5.19%	4.70%	-0.77%	1.87%	4.70%	7.10%	9.30%
			20	1377	1.58%	0.71%	-1.63%	-0.45%	0.71%	2.57%	4.74%
			30	1377	1.58%	0.71%	-1.63%	-0.45%	0.71%	2.57%	4.74%
			40	1377	1.58%	0.71%	-1.63%	-0.45%	0.71%	2.57%	4.74%
			50	1377	1.58%	0.71%	-1.63%	-0.45%	0.71%	2.57%	4.74%
			Unlim	1377	1.58%	0.71%	-1.63%	-0.45%	0.71%	2.57%	4.74%

5.   BASELINE PERFORMANCE – SCE AMP

Table 5–1 reports results for AMP DA, which provide an interesting case. As indicated in the first panel, the As Chosen baseline is relatively inaccurate, with MAPE values of over 6 percent for actual events, and an upward bias of more than 7 percent. Results are in the same direction, though somewhat smaller, for the simulated events. Now note that the Un-adjusted option for the actual and simulated events in the Universal case, have the same MAPE and MPE values as the As Chosen case. This result arises because all but one of the DA customers declined the baseline adjustment option until September, when more than half changed to the adjustment option. As a result, the As Chosen baseline results are the same as the Universal un-adjusted option.

With Universal adjustments, however, accuracy improves to less than 2 percent median absolute error, and MPE falls to nearly zero, for both actual and simulated events. The range of errors is quite narrow for all adjustment caps (except no adjustment). The portfolio results are similar to the Universal aggregated case.

Table 5-1: Baseline Performance – SCE AMP DA

Load Type	Adj. Choice	Event Type	Cap	Cust. Count	Accuracy MAPE	Bias MPE	p10	p25	p50	p75	p90
Aggregated	As Chosen	Actual	Unadj								
			20	236	6.41%	7.36%	-1.67%	2.39%	7.36%	8.00%	14.68%
			30	236	6.40%	7.36%	-1.59%	2.36%	7.36%	8.00%	14.68%
			40	236	6.40%	7.36%	-1.56%	2.36%	7.36%	8.00%	14.68%
			50	236	6.40%	7.36%	-1.52%	2.38%	7.36%	8.00%	14.68%
			Unlim	236	6.39%	7.36%	-1.42%	2.43%	7.36%	8.00%	14.68%
		Simulated	Unadj								
			20	240	4.46%	3.31%	-3.50%	1.25%	3.31%	6.61%	7.50%
			30	240	4.46%	3.31%	-3.50%	1.38%	3.31%	6.61%	7.50%
			40	240	4.45%	3.31%	-3.50%	1.43%	3.31%	6.61%	7.50%
			50	240	4.44%	3.31%	-3.50%	1.43%	3.31%	6.61%	7.50%
			Unlim	240	4.42%	3.31%	-3.50%	1.43%	3.31%	6.61%	7.50%
Aggregated	Universal	Actual	Unadj	236	6.98%	7.36%	-3.85%	4.29%	7.36%	7.99%	14.68%
			20	236	1.82%	0.15%	-2.03%	-1.10%	0.15%	0.97%	4.85%
			30	236	1.65%	0.02%	-2.91%	-1.49%	0.02%	0.80%	3.68%
			40	236	1.64%	-0.05%	-3.26%	-1.72%	-0.05%	0.87%	3.59%
			50	236	1.63%	-0.13%	-3.31%	-1.76%	-0.13%	0.73%	3.55%
			Unlim	236	1.53%	-0.12%	-3.32%	-1.73%	-0.12%	-0.02%	3.39%
		Simulated	Unadj	240	4.55%	3.31%	-3.50%	2.28%	3.31%	6.61%	7.50%
			20	240	1.62%	0.61%	-1.42%	-0.46%	0.61%	1.81%	3.97%
			30	240	1.57%	0.12%	-1.48%	-0.55%	0.12%	1.98%	3.57%
			40	240	1.58%	0.04%	-1.50%	-0.70%	0.04%	2.05%	3.30%
			50	240	1.59%	0.02%	-1.64%	-0.98%	0.02%	1.96%	3.14%
			Unlim	240	1.62%	0.01%	-1.93%	-1.47%	0.01%	1.93%	2.85%
Portfolio	Universal	Actual	Unadj	236	11.99%	11.31%	6.28%	8.95%	11.31%	14.68%	20.36%
			20	236	1.92%	-0.12%	-2.85%	-1.09%	-0.12%	1.17%	4.44%
			30	236	1.64%	-1.02%	-2.85%	-1.56%	-1.02%	0.70%	3.67%
			40	236	1.64%	-1.02%	-2.85%	-1.56%	-1.02%	0.70%	3.67%
			50	236	1.64%	-1.02%	-2.85%	-1.56%	-1.02%	0.70%	3.67%
			Unlim	236	1.64%	-1.02%	-2.85%	-1.56%	-1.02%	0.70%	3.67%
		Simulated	Unadj	240	7.53%	6.59%	2.15%	3.34%	6.59%	10.13%	14.84%
			20	240	1.99%	-0.08%	-1.41%	-0.96%	-0.08%	3.11%	4.89%
			30	240	2.03%	-0.56%	-2.34%	-1.29%	-0.56%	1.09%	4.89%
			40	240	2.03%	-0.56%	-2.34%	-1.29%	-0.56%	1.09%	4.89%
			50	240	2.03%	-0.56%	-2.34%	-1.29%	-0.56%	1.09%	4.89%
			Unlim	240	2.03%	-0.56%	-2.34%	-1.29%	-0.56%	1.09%	4.89%

Table 5–2 reports results for AMP DO, in which 84 percent of customers selected the day-of adjustment option. For the As Chosen case, accuracy and bias performance are good, with MAPE values of less than 1.5 percent for actual events and about 2.5 percent for simulated events, MPE values of less than 1 percent for both event types. The distribution of errors is also relatively tight.

Accuracy falls under the Universal option, with MAPE values more than doubling for the actual events, though remaining nearly the same for simulated events. MPE values indicate a somewhat larger downward bias for actual events, with some large outliers, but remain nearly the same for simulated events. Portfolio results are similar to the

aggregated Universal case for actual events, but accuracy falls for simulated events, and the MPE values indicate a small upward bias.

Table 5-2: Baseline Performance – SCE AMP DO

Load Type	Adj. Choice	Event Type	Cap	Cust. Count	Accuracy MAPE	Bias MPE	p10	p25	p50	p75	p90
Aggregated	As Chosen	Actual	Unadj								
			20	1535	1.46%	0.65%	-2.01%	-1.02%	0.65%	1.38%	2.07%
			30	1535	1.38%	0.15%	-2.40%	-1.17%	0.15%	1.27%	2.07%
			40	1535	1.36%	-0.19%	-2.87%	-1.14%	-0.19%	1.27%	2.07%
			50	1535	1.37%	-0.39%	-3.05%	-1.12%	-0.39%	1.16%	2.07%
			Unlim	1535	1.78%	-0.74%	-4.22%	-2.21%	-0.74%	0.75%	2.07%
		Simulated	Unadj								
			20	1537	2.53%	0.88%	-3.08%	-0.70%	0.88%	2.58%	4.78%
			30	1537	2.50%	0.63%	-2.92%	-1.19%	0.63%	2.26%	4.48%
			40	1537	2.54%	0.51%	-2.89%	-1.12%	0.51%	2.24%	4.71%
			50	1537	2.60%	0.40%	-2.86%	-1.17%	0.40%	2.23%	4.93%
			Unlim	1537	2.82%	-0.19%	-4.83%	-2.26%	-0.19%	1.83%	4.35%
Aggregated	Universal	Actual	Unadj	1535	4.30%	3.55%	-3.69%	0.65%	3.55%	5.97%	7.33%
			20	1535	3.39%	-0.65%	-14.26%	-2.04%	-0.65%	0.39%	1.39%
			30	1535	3.39%	-1.16%	-14.26%	-3.21%	-1.16%	0.15%	0.47%
			40	1535	3.52%	-1.18%	-14.26%	-3.87%	-1.18%	-0.19%	0.19%
			50	1535	3.64%	-1.60%	-14.26%	-4.25%	-1.60%	-0.39%	0.22%
			Unlim	1535	4.50%	-2.23%	-14.26%	-5.22%	-2.23%	-1.39%	0.73%
		Simulated	Unadj	1537	4.61%	3.19%	-4.20%	1.27%	3.19%	5.63%	8.40%
			20	1537	2.38%	0.15%	-2.96%	-1.25%	0.15%	2.22%	3.62%
			30	1537	2.37%	-0.03%	-3.09%	-1.49%	-0.03%	2.01%	3.53%
			40	1537	2.37%	-0.14%	-3.28%	-1.47%	-0.14%	1.90%	3.86%
			50	1537	2.44%	-0.06%	-3.32%	-1.44%	-0.06%	1.87%	4.16%
			Unlim	1537	3.00%	-0.96%	-4.90%	-2.71%	-0.96%	1.79%	4.26%
Portfolio	Universal	Actual	Unadj	1535	4.31%	3.55%	-3.69%	0.65%	3.55%	6.11%	7.37%
			20	1535	4.04%	-1.64%	-14.26%	-4.52%	-1.64%	-0.22%	0.57%
			30	1535	4.04%	-1.64%	-14.26%	-4.52%	-1.64%	-0.22%	0.57%
			40	1535	4.04%	-1.64%	-14.26%	-4.52%	-1.64%	-0.22%	0.57%
			50	1535	4.04%	-1.64%	-14.26%	-4.52%	-1.64%	-0.22%	0.57%
			Unlim	1535	4.04%	-1.64%	-14.26%	-4.52%	-1.64%	-0.22%	0.57%
		Simulated	Unadj	1537	12.42%	3.88%	-4.07%	0.91%	3.88%	8.52%	14.56%
			20	1537	7.70%	0.93%	-4.17%	-1.65%	0.93%	3.27%	7.35%
			30	1537	6.68%	0.93%	-4.17%	-1.65%	0.93%	3.27%	6.43%
			40	1537	5.81%	0.86%	-4.17%	-1.65%	0.86%	3.19%	6.23%
			50	1537	5.09%	0.86%	-4.17%	-1.65%	0.86%	3.19%	4.97%
			Unlim	1537	3.04%	0.70%	-4.17%	-1.89%	0.70%	2.72%	4.65%

CA Energy Consulting

6. CONCLUSIONS AND RECOMMENDATIONS

The primary conclusions of this baseline study are the following:

- Similarly to previous baseline studies, the accuracy and bias measures for the aggregated customer 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) and adjustment cap. These factors likely include the characteristics of the particular customers who participate in the programs and the nature of the events included in the study (*e.g.*, whether weather conditions on prior days that make up the baseline calculation were substantially milder or hotter than the event day).
- As a corollary, few common patterns emerge in terms of particular baseline types or cap restrictions producing consistent upward or downward biases.
- Several programs and notice types, such as SDG&E CBP DO and PG&E AMP DA and DO, produced relatively low errors, both in terms of median values and a relatively tight range of errors.
- Allowing some day-of adjustment to the 10-in-10 baseline nearly always improves accuracy and reduces bias (*i.e.*, the un-adjusted option frequently produces the least accurate results within a category, such as aggregated Universal adjustments), 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.

While not a conclusion *per se*, we reiterate a point made in previous baseline studies that errors in baseline loads can be magnified into larger errors in estimated load impacts, depending on the relative magnitude of the load impact.

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, and in at least one case, portfolio baselines were outperformed by aggregations of customer baselines.
- Allowing some day-of adjustment nearly always improves baseline performance, which suggests making an adjusted baseline the default settlement baseline.
- We agree with previous recommendations 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.

CA Energy Consulting