# 2008 Load Impact Evaluation of California Statewide Aggregator Demand Response Programs

## Volume 2: Baseline Analysis of AMP Aggregator Demand Response Program

## CALMAC Study ID PGE0274.02

Steven D. Braithwait and David Armstrong

**Christensen Associates Energy Consulting, LLC**

May 1, 2009

# Table of Contents


**ABSTRACT** ..... **1**

**EXECUTIVE SUMMARY** ..... **2**

PROJECT OBJECTIVES ..... 2
APPROACH ..... 2
RESULTS ..... 3

**1. INTRODUCTION** ..... **6**

**2. DATA** ..... **7**

2.1 CUSTOMERS ..... 7
2.2 EVENTS ..... 7

**3. APPROACH** ..... **9**

3.1 BASELINE PERFORMANCE STATISTICS ..... 9
3.2 ADJUSTED BASELINES ..... 10

**4. RESULTS FOR *EVENT-TYPE DAYS*** ..... **11**

4.1 UNADJUSTED BASELINES ..... 11
4.2 ADJUSTED BASELINES ..... 16
4.3 DISTRIBUTIONS OF RELATIVE ERRORS ..... 21

**5. RESULTS FOR *EVENT DAYS*** ..... **27**

5.1 UNADJUSTED BASELINES ..... 27
5.2 ADJUSTED BASELINES ..... 29
5.3 CONCLUSIONS FOR EVENT DAYS ..... 32

**6. GAMING** ..... **32**

6.1 COMPARISON OF LOADS ON EVENT-DAYS AND EVENT-TYPE DAYS ..... 33
6.2 ANALYSIS OF PRE-EVENT USAGE ..... 38
6.3 CONCLUSIONS ABOUT GAMING ..... 38

**7. CONCLUSIONS** ..... **39**

7.1 BASELINE PERFORMANCE – EVENT-TYPE DAYS ..... 39
7.2 BASELINE PERFORMANCE – EVENT DAYS ..... 39
7.3 WAS GAMING SUCCESSFULLY AVOIDED? ..... 39

**APPENDIX: TECHNICAL BACKGROUND** ..... **41**

BASELINE DEFINITIONS ..... 41
BASELINE ERRORS ..... 41
DIFFERENCES IN LOAD IMPACTS AND BASELINE ERRORS ..... 42

## List of Tables


TABLE ES.1 ACCURACY OF *UNADJUSTED* BASELINES – EVENT-TYPE DAYS ............................. 4
TABLE ES.2 ACCURACY OF *ADJUSTED* BASELINES – EVENT-TYPE DAYS.................................... 4
TABLE ES.3 BIAS OF *UNADJUSTED* BASELINES ................................................................................... 4
TABLE ES.4 BIAS OF *ADJUSTED* BASELINES........................................................................................... 5
TABLE 1. CHARACTERISTICS OF AMP BASELINE CUSTOMERS .................................................... 7
TABLE 2. EVENT-TYPE DAYS ................................................................................................................... 8
TABLE 3. ACCURACY OF UNADJUSTED BASELINES |
*(RELATIVE ROOT MEAN SQUARE ERROR, OR THEIL U-STATISTIC)* ................................ 12
TABLE 4. BIAS OF UNADJUSTED BASELINES *(MEDIAN PERCENT ERRORS)* ............................. 14
TABLE 5. ACCURACY OF ADJUSTED BASELINES
*(RELATIVE ROOT MEAN SQUARE ERROR, OR THEIL U-STATISTIC)* ................................ 16
TABLE 6. BIAS OF ADJUSTED BASELINES *(MEDIAN PERCENT ERRORS)* ................................... 19
TABLE 7. ACCURACY OF UNADJUSTED BASELINES – EVENT DAYS
*(RELATIVE ROOT MEAN SQUARE ERROR, OR THEIL U-STATISTIC)* ................................ 28
TABLE 8. BIAS OF *UNADJUSTED* BASELINES – EVENT DAYS *(MEDIAN PERCENT ERRORS)*... 29
TABLE 9. ACCURACY OF ADJUSTED BASELINES – EVENT DAYS
*(RELATIVE ROOT MEAN SQUARE ERROR, OR THEIL U-STATISTIC)* ................................ 30
TABLE 10. BIAS OF *ADJUSTED* BASELINES – EVENT DAYS *(MEDIAN PERCENT ERRORS)* ...... 31
TABLE 11. RATIOS OF AVERAGE MORNING USAGE – EVENT AND NON-EVENT DAYS
(*BY CHOICE OF ADJUSTED BASELINE*) ................................................................................ 38
TABLE 12. QUARTILES OF RATIOS OF AVERAGE MORNING USAGE ......................................... 38

## List of Figures


FIGURE 1. ACCURACY OF *UNADJUSTED* BASELINES – *AGGREGATOR* ........ 13
FIGURE 2. ACCURACY OF *UNADJUSTED BASELINES – SUM-OF-CUSTOMERS* ........ 13
FIGURE 3. BIAS OF *UNADJUSTED* BASELINES – *AGGREGATOR* ........ 15
FIGURE 4. BIAS OF *UNADJUSTED* BASELINES – *SUM-OF-CUSTOMERS* ........ 15
FIGURE 5. ACCURACY OF *ADJUSTED* BASELINES – *AGGREGATOR* ........ 17
FIGURE 6. ACCURACY OF *ADJUSTED* BASELINES – *SUM OF CUSTOMER* ........ 17
FIGURE 7. BIAS OF *ADJUSTED* BASELINES – *AGGREGATOR* ........ 20
FIGURE 8. BIAS OF *ADJUSTED* BASELINES – *SUM OF CUSTOMER* ........ 20
FIGURE 9. AVERAGE EVENT-TYPE DAY % ERRORS FOR *ADJUSTED 10-IN-10* AGGREGATOR AND SUM-OF-CUSTOMER BASELINES – *ALL INDUSTRY TYPES* ........ 22
FIGURE 10. AVERAGE EVENT-DAY % ERRORS FOR *ADJUSTED 10-IN-10* AGGREGATOR AND SUM-OF-CUSTOMER BASELINES – *COMMERCIAL* ........ 23
FIGURE 11. AVERAGE EVENT-DAY % ERRORS FOR *ADJUSTED 10-IN-10* AGGREGATOR AND SUM-OF-CUSTOMER BASELINES – *INDUSTRIAL* ........ 23
FIGURE 12. AVERAGE EVENT-DAY % ERRORS FOR *ADJUSTED 10-IN-10* AGGREGATOR AND SUM-OF-CUSTOMER BASELINES – *SCHOOLS* ........ 24
FIGURE 13. DISTRIBUTIONS OF AVERAGE EVENT-DAY % ERRORS FOR *UNADJUSTED* AND *ADJUSTED 10-IN-10* BASELINES – *INDIVIDUAL CUSTOMER ACCOUNTS* ........ 25
FIGURE 14. AVERAGE EVENT-DAY % ERRORS FOR *ADJUSTED 10-IN-10* BASELINES, BY CUSTOMER SIZE (AVERAGE PEAK KW) – *INDUSTRIAL CUSTOMERS* ........ 26
FIGURE 15. AVERAGE EVENT-DAY % ERRORS FOR *ADJUSTED 10-IN-10* BASELINES, BY CUSTOMER SIZE (AVERAGE PEAK KW) – *COMMERCIAL CUSTOMERS* ........ 26
FIGURE 16. COMPARISON OF BIAS OF *ADJUSTED* AND *UNADJUSTED* BASELINES – *AGGREGATOR METHOD* ........ 32
FIGURE 17. AGGREGATOR 2; INDUSTRIAL; *NO ADJUSTED BL* ........ 34
FIGURE 18. AGGREGATOR 2; INDUSTRIAL; *ADJUSTED BL OPTION* ........ 34
FIGURE 19. AGGREGATOR 2; COMMERCIAL; *ADJUSTED BL OPTION* ........ 35
FIGURE 20. AGGREGATOR 3; INDUSTRIAL; *ADJUSTED BL OPTION* ........ 35
FIGURE 21. AGGREGATOR 4; INDUSTRIAL; *NO ADJUSTED BL* ........ 36
FIGURE 22. AGGREGATOR 4; INDUSTRIAL; *ADJUSTED BL OPTION* ........ 36
FIGURE 23. AGGREGATOR 4; COMMERCIAL; *ADJUSTED BL OPTION* ........ 37
FIGURE 24. AGGREGATOR 4; SCHOOLS; *ADJUSTED BL OPTION* ........ 37

## Abstract

This report summarizes the results of a baseline analysis that was undertaken as part of the statewide ex-post evaluation of the Aggregator demand response programs. The objective of the baseline analysis was to assess the relative accuracy and bias of several alternative methods for calculating baselines for measuring load impacts for settlement. Data for some 600 customers enrolled in PG&E's Aggregator Managed Portfolio (AMP) program were used to assess the performance of a range of unadjusted and adjusted baseline methods. Of particular interest were differences in performance between baselines calculated for the aggregation of load data across customers enrolled for a particular aggregator, and a baseline calculated as the sum of individual customers' baselines.

## Executive Summary

This volume documents the results of a baseline analysis study undertaken in the context of a load impact evaluation of aggregated demand response ("DR") programs operated by the three California investor-owned utilities (IOUs), Pacific Gas and Electric ("PG&E"), Southern California Edison ("SCE"), and San Diego Gas and Electric ("SDG&E") for Program Year 2008. In these programs, aggregators contract with commercial and industrial customers to act on their behalf to provide demand response load reductions to the utility, arrange load curtailments, receive incentive payments, and pay penalties (if warranted) to the utility. Each aggregator forms a "portfolio" of individual customers such that their aggregated load participates in the DR programs. This baseline analysis used data only for PG&E's Aggregator Managed Portfolio ("AMP") program.

### *Project Objectives*

This study addressed a continuing issue in the design of such DR programs, which is the accuracy and bias of various alternative baseline calculation methods that might be used to calculate the *baseline load* that is used to measure load reductions during events, and ultimately provide the settlement for incentive payments. Of particular interest for the aggregator programs are four issues:

1. Whether the baseline for the settlement payment for an aggregated portfolio should be constructed using the coincident aggregated load of all enrolled customers who are nominated by the aggregator for the month in which an event is called, or by calculating the individual baselines for each such customer, and summing the results to develop the aggregate baseline.
2. How many representative non-event days prior to an event day should be included in the baseline calculation (*e.g.*, the three, five, or 10 days with the highest event-period consumption in the previous 10 days).
3. Should the baseline calculation for the event day be adjusted using pre-event usage data on the event day in an attempt to compensate for the fact that consumption during event days may differ from consumption on the days prior to the event due to anomalous effects such as extreme temperatures?
4. Was gaming avoided for the customers and aggregators who selected an adjusted baseline option in PG&E's AMP program in 2008?

### *Approach*

The study used hourly interval energy usage data for all AMP customers who were nominated during the summer months. The performance of a range of alternative baseline methods was examined in terms of both accuracy and bias (*e.g.,* the tendency of a baseline method to under-state or over-state the "true" baseline). Data were used for both the five actual AMP events in 2008, as well as a selection of ten event-type days of similar high temperatures and PG&E system load. The use of event-type days was designed to expand the number of observations on which the baseline analysis was undertaken, as well as to provide cases in which the "true" baseline is represented by customers' or aggregators' observed usage during a "pseudo-event" period.

For the event days, the estimated hourly load reductions from the ex post load impact evaluation (as described in Volume 1 of this report) for each customer and each event were added to their observed load on event days to create an estimate of their true baseline.

To examine potential differences in baseline performance by type of customers, each aggregator's customers were classified into three categories – industrial-type customers, which are not particularly weather sensitive; commercial-type customers, which are presumed to be weather sensitive; and schools, whose load patterns often vary during summer months due to vacation schedules. Aggregated loads for those sub-groups were calculated for each aggregator.

Baseline performance (relative to the true baseline) was measured primarily by two statistics that have been used in previous baseline studies. Accuracy was measured using the *relative root mean square error* statistic (RRMSE, sometimes referred to as the Theil U-statistic). Bias was measured using the *median percent error*, where positive errors indicate *downward bias* (*i.e.*, the true baseline exceeds the estimated baseline), and negative errors indicate *upward bias* (*i.e.*, the estimated baseline exceeds the true baseline).

### *Results*

The results of the event-type day baseline analysis provide evidence on the relative accuracies and biases of the aggregator (coincident portfolio) baseline approach and the sum-of-customer (individual) baseline approach for the 3-, 5-, and 10-in-10 representative days. Also observed are the effects of morning adjustments to 3-, 5-, and 10-in-10 baselines, using the symmetric and upward-only adjustment methods.

The results are somewhat mixed, suggesting that baseline performance depends to some extent on the characteristics of particular customers' load patterns, and on the nature of event days and the days that make up the representative days for constructing the baselines. However, certain patterns were evident. For this analysis of the PG&E AMP program customers in 2008, the most accurate methods (that is, those with the lowest relative errors) were the aggregator methods with symmetric adjustment. The same methods generally exhibited the smallest biases as well, with some indications that the adjusted 10-in-10 method performed best overall. In particular, the aggregator methods with symmetric adjustment reduced a downward bias that was typically found in unadjusted versions of the methods (that is, the unadjusted baseline was systematically less than the "true" baseline). Some details on the baseline performance findings are presented below.

### Accuracy

Tables ES.1 and ES.2 summarize *accuracy* results for event-type days, for unadjusted and adjusted baselines respectively, showing overall relative errors for each AMP aggregator and in total. Accuracies of the unadjusted baselines for both aggregator and sum-of-customer aggregation methods were similar in magnitude, with relative errors ranging from approximately 5 to 9 percent. Adjustments to the baseline generally improved accuracy, typically cutting relative errors approximately in half, especially for the aggregator method.

**Table ES.1 Accuracy of *Unadjusted* Baselines – Event-Type Days**

| | | *Aggregator* | | | *Sum of Customers* | | |
|---|---|---|---|---|---|---|---|
| | | Unadjusted | | | Unadjusted | | |
| Agg. | Level | 3-in-10 | 5-in-10 | 10-in-10 | 3-in-10 | 5-in-10 | 10-in-10 |
| 1 | Total | 0.057 | 0.069 | 0.092 | 0.054 | 0.057 | 0.091 |
| 2 | Total | 0.065 | 0.074 | 0.102 | 0.055 | 0.065 | 0.102 |
| 3 | Total | 0.049 | 0.056 | 0.080 | 0.068 | 0.052 | 0.080 |
| 4 | Total | 0.061 | 0.053 | 0.049 | 0.120 | 0.093 | 0.049 |
| All | TOTAL | 0.056 | 0.062 | 0.083 | 0.075 | 0.062 | 0.083 |

**Table ES.2 Accuracy of *Adjusted* Baselines – Event-Type Days**

| | | *Aggregator* | | | | | *Sum of Customers* | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Symmetric Adjustment | | | *Upward-only* | | Symmetric Adjustment | | | *Upward-only* | |
| Agg. | Level | 3-in-10 | 5-in-10 | 10-in-10 | 5-in-10 | 10-in-10 | 3-in-10 | 5-in-10 | 10-in-10 | 5-in-10 | 10-in-10 |
| 1 | Total | 0.022 | 0.023 | 0.022 | 0.022 | 0.022 | 0.034 | 0.025 | 0.027 | 0.044 | 0.024 |
| 2 | Total | 0.025 | 0.028 | 0.027 | 0.034 | 0.030 | 0.033 | 0.030 | 0.026 | 0.039 | 0.029 |
| 3 | Total | 0.022 | 0.021 | 0.020 | 0.025 | 0.020 | 0.043 | 0.037 | 0.034 | 0.071 | 0.033 |
| 4 | Total | 0.044 | 0.039 | 0.037 | 0.053 | 0.037 | 0.087 | 0.071 | 0.041 | 0.118 | 0.063 |
| All | TOTAL | 0.029 | 0.028 | 0.027 | 0.034 | 0.028 | 0.051 | 0.043 | 0.036 | 0.074 | 0.039 |

Similar accuracy results were obtained in the analysis of actual AMP event-day data.

## Bias

Tables ES.3 and ES.4 summarize *bias* results for unadjusted and adjusted baselines for the event-type days.[1] The overall results for the *unadjusted* baseline vary considerably, though certain patterns are discernable. With a few exceptions (in particular, the 3-in-10 baseline for the sum-of-customers method), the unadjusted baselines tended to under-state the actual baseline (*i.e.*, median percent error values are positive), and the under-statement usually increased with the number of days included in the baseline calculation (*e.g.*, from 3 to 10). The *adjusted* baselines typically reduced the under-statement bias for the aggregator method, in many cases shifting the distribution of errors such that the median percent error was negative, indicating a small over-statement bias. A similar effect occurred with the sum-of-customers method.

**Table ES.3 Bias of *Unadjusted* Baselines**

| | | *Aggregator* | | | *Sum of Customers* | | |
|---|---|---|---|---|---|---|---|
| | | Unadjusted | | | Unadjusted | | |
| Agg. | Level | 3-in-10 | 5-in-10 | 10-in-10 | 3-in-10 | 5-in-10 | 10-in-10 |
| 1 | Total | 4.42% | 5.59% | 8.45% | -0.37% | 2.57% | 8.28% |
| 2 | Total | 1.39% | 3.23% | 7.76% | -2.75% | 0.75% | 7.68% |
| 3 | Total | 3.51% | 4.82% | 8.60% | 0.89% | 3.09% | 8.55% |
| 4 | Total | 0.01% | 1.07% | 4.14% | -4.70% | -2.71% | 4.14% |
| All | TOTAL | 2.47% | 3.75% | 7.24% | -0.90% | 1.55% | 7.15% |

[1] Note again that by the definition of baseline error used in this study, *positive* errors represent downward biases (*i.e.*, the baseline being tested under-states the true baseline), while *negative* errors represent upward biases (*i.e.*, the baseline being tested over-states the true baseline).

**Table ES.4 Bias of *Adjusted* Baselines**

| | | Aggregator | | | | | Sum of Customers | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Symmetric Adjustment | | | Upward-only Adjustment | | Symmetric Adjustment | | | Upward-only Adjustment | |
| Agg. | Level | 3-in-10 | 5-in-10 | 10-in-10 | 5-in-10 | 10-in-10 | 3-in-10 | 5-in-10 | 10-in-10 | 5-in-10 | 10-in-10 |
| 1 | Total | -0.03% | 0.72% | 0.97% | 0.72% | 0.97% | -2.12% | -0.76% | 1.51% | -2.81% | 0.64% |
| 2 | Total | -1.59% | -1.13% | -0.12% | -2.41% | -1.17% | -3.63% | -2.33% | 0.56% | -4.49% | -0.51% |
| 3 | Total | -0.98% | -0.52% | 0.22% | -0.92% | -0.05% | -1.72% | -1.29% | 1.37% | -2.75% | 0.33% |
| 4 | Total | -0.70% | -0.59% | -0.05% | -2.29% | -0.80% | -3.03% | -2.79% | -0.48% | -5.31% | -2.14% |
| All | TOTAL | -0.71% | -0.36% | 0.26% | -1.29% | -0.38% | -2.25% | -1.52% | 0.70% | -3.76% | -0.40% |

Similar patterns of results were obtained using data for AMP event days.

Tables in the body of the report contain results by aggregator and industry type, and for event days as well as event-type days. Major findings of the baseline study include the following:

1. Regarding the accuracy of the *aggregator* method of calculating baselines compared to the *sum-of-customer* method of using the individual customer baselines and summing them, the results suggest that the aggregator method is more accurate than the sum-of-customer method, though not in all cases and not by a wide margin (*e.g.,* compare the two sets of columns in Tables ES.1 for unadjusted baselines and ES.2 for adjusted baselines). The accuracies of both methods are improved when morning adjustments are applied, using either the symmetric or the upward-only adjustment approaches.
2. Regarding the effect of *morning adjustments* to the 3-in-10 baseline on *bias*, the results suggest that the adjustments reduce the bias of the unadjusted baseline (which is typically a downward bias) relative to the "true" baseline (*e.g.*, compare the first columns between Tables ES.3 and ES.4).
3. Expanding the analysis to consider adjusted 5-in-10 and 10-in-10 baselines produced results suggesting that the *adjusted 10-in-10 method* may produce both the greatest accuracy and the smallest bias (*e.g.*, see the third columns in Tables ES.2 and ES.4).
4. The performance of the alternative baseline methods on actual *event days* in terms of accuracy and bias was qualitatively similar to their performance on the *event-type* days summarized above.
5. Regarding the question of the extent to which gaming was avoided for the adjusted baseline option, analysis in this study revealed no evidence of systematic increases in pre-event consumption on event days for those aggregators that offered an adjusted baseline option. Evidence of such increases could be indicative of attempts to *game* the adjusted baseline. Only one case was found, for just one industrial customer of one aggregator, in which hourly usage rose by an unusual amount in the four hours prior to one event.
6. Baseline performance in terms of accuracy and bias varied by the broad industry types considered in this analysis (*e.g.*, industrial, commercial and schools). However, differences in results were not consistent, and often varied more across aggregators than across industry types. Results on the distributions of baseline errors across customers in those industry types were presented in the report. However, additional research would be needed to identify features of customers that might produce more accurate baseline calculations.

## 1. Introduction

This report summarizes the results of a baseline analysis that was undertaken as part of the statewide ex-post evaluation of the Aggregator demand response programs. The objective of the baseline analysis was to assess the relative accuracy and bias of several alternative methods for calculating baselines for measuring load impacts for settlement.

The original scope of work involved analysis to address three baseline issues regarding PG&E's Aggregator Managed Portfolio ("AMP") program:

1. Compare the accuracy in measuring load reductions of two alternative methods—Test whether estimating load impacts by comparing actual *aggregate* program loads during an event to an *aggregator* baseline, or to a baseline constructed as the sum of individual *customer-specific* baselines, is more accurate in measuring load reductions.
2. Evaluate whether *morning adjustments* to the 3-in-10 baseline actually improve the bias of the baseline (*i.e.*, the tendency of the calculated baseline to understate or overstate the "true" baseline).
3. Test whether *gaming* is successfully avoided.

The scope was expanded to consider a number of additional baselines and adjustment mechanisms, including the following:

4. Evaluate and compare the accuracy of the following baselines using *day-of adjustment*:
    a. aggregated 3-in-10,
    b. individual 3-in-10,
    c. aggregated 5-in-10,
    d. individual 5-in-10,
    e. aggregated 10-in-10, and
    f. individual 10-in-10.

    The adjustment to be used should be the one AMP currently uses in 2008; that is, the ratio of a) the average load of the 4 hours preceding the event to b) the average load of the same 4 hours of the baseline days.

5. Evaluate the effects of *upward-only* day-of adjustment vs. symmetric day-of adjustment on a baseline. The baseline models to be studied include:
    a. 5-in-10 with a *symmetric* adjustment vs. 5-in-10 with an *upward-only* adjustment, and
    b. 10-in-10 with a *symmetric* adjustment vs. 10-in-10 with an *upward-only* adjustment.
6. Evaluate the effects of allowing the option of symmetric adjustment on:
    a. 10-in-10 (*i.e.*, 10-in-10 unadjusted vs 10-in-10 with a symmetric adjustment), and
    b. 5-in-10 (*i.e.*, 5-in-10 unadjusted vs 5-in-10 with a symmetric adjustment).

## 2. Data

### *2.1 Customers*

We used data for nearly all of the customers that were nominated by each of four of the five AMP aggregators (one aggregator had only one customer) for the relevant months during 2008. Given the interest in adjusted baselines and gaming for those customers who selected the adjusted baseline option in 2008, some portions of the analysis were conducted separately by customers' choice of adjusted baseline. In addition, to examine potential differences in baseline performance between weather-sensitive and non-weather sensitive customers, we constructed sub-groups of customer types based on their categorization within the standard eight industry groups used in load impact evaluations. The customer types were designed to differentiate between "Industrial-type" customers that are likely to be relatively non-weather sensitive (Industry types 1-3, which include manufacturing, construction, wholesale trade and other utilities), and "Commercial-type" customers, which are likely to be relatively more weather sensitive (Industry types 4, 5, and 7, which include retail stores; offices, hotels and services; and government and institutions). Schools (6) were treated separately due to their unique scheduling differences during the summer period.

The number of customer accounts included in the analysis, and their industry type and usage characteristics are shown in Table 1. Aggregators 1 and 2 have relatively large shares of commercial customers, while aggregators 3 and 4 have large shares of industrial customers, and some schools. Aggregators 2 and 4 had a substantial share of customers accept the adjusted baseline option. The last two columns suggest no large systematic differences in average size across aggregators, or by industry type or by choice of adjusted baseline.

**Table 1. Characteristics of AMP Baseline Customers**

| | Industry | Count | | | Max Demand (MW) | | | Ind. Type | Average Max kW | |
|---|---|---|---|---|---|---|---|---|---|---|
| Agg | Group | Adj. BL | No Adj | Total | Adj. BL | No Adj | Total | % of Total | Adj. BL | No Adj |
| 1 | 1 | | 84 | 84 | | 53.2 | 53.2 | 56% | | 634 |
| | 2 | | 70 | 70 | | 41.0 | 41.0 | 44% | | 586 |
| | Total | | 154 | 154 | | 94.2 | 94.2 | | | 612 |
| 2 | 1 | 22 | 13 | 35 | 12.2 | 9.3 | 21.6 | 26% | 557 | 719 |
| | 2 | 81 | 2 | 83 | 61.8 | 0.9 | 62.6 | 74% | 762 | 436 |
| | Total | 103 | 15 | 118 | 74.0 | 10.2 | 84.2 | | 719 | 681 |
| 3 | 1 | 10 | 118 | 128 | 5.7 | 137.6 | 143.3 | 81% | 574 | 1,166 |
| | 2 | 2 | 28 | 30 | 2.8 | 23.3 | 26.0 | 15% | 1,376 | 831 |
| | 3 | 3 | 11 | 14 | 0.6 | 6.7 | 7.3 | 4% | 194 | 611 |
| | Total | 15 | 157 | 172 | 9.1 | 167.5 | 176.6 | | 605 | 1,067 |
| 4 | 1 | 26 | 42 | 68 | 26.3 | 57.7 | 84.0 | 72% | 1,010 | 1,374 |
| | 2 | 30 | 3 | 33 | 12.6 | 6.3 | 18.9 | 16% | 419 | 2,110 |
| | 3 | 3 | | 3 | 14.3 | 0.0 | 14.3 | 12% | 4,761 | |
| | Total | 59 | 45 | 104 | 53.1 | 64.1 | 117.2 | | 900 | 1,423 |
| ALL | 1 | 58 | 257 | 315 | 44.3 | 257.9 | 302.1 | 64% | 763 | 1,003 |
| | 2 | 113 | 103 | 216 | 77.1 | 71.5 | 148.5 | 31% | 682 | 694 |
| | 3 | 6 | 11 | 17 | 14.9 | 6.7 | 21.6 | 5% | 2,477 | 611 |
| | Total | 177 | 371 | 548 | 136.2 | 336.1 | 472.2 | | 769 | 906 |

### *2.2 Events*

Given the relatively small number of AMP events (one actual event for one of the aggregators, and four test events for a mix of aggregators), and the availability of a number of days of relatively high PG&E system load and temperatures, we conducted much of the baseline analysis for ten *event-type* days during the May to September period. These are shown in Table 2 (actual and test events are shown in highlight).[2] The simulated events were assumed to be five hours in length, from hours ending 14 through 18. Morning adjustments were made using consumption in the four hours prior to the "event." In addition to the event-type days, we also examined baseline performance on the actual event days, as well as the analysis of possible gaming.

**Table 2. Event-type Days**

| | AMP Ld (HE 14) | wCDD | DOW | Max temp | Event-like day |
|---|---|---|---|---|---|
| 12-May-08 | 479,979 | 1.5 | 1 | | |
| 13-May-08 | 498,251 | 4.2 | 2 | | |
| 14-May-08 | 534,785 | 9.8 | 3 | | |
| 15-May-08 | 554,240 | 17.6 | 4 | 99.6 | |
| 16-May-08 | 495,091 | 18.8 | 5 | 101.4 | |
| 19-May-08 | 516,584 | 9.8 | 1 | | |
| 20-May-08 | 512,840 | 6.6 | 2 | | |
| 21-May-08 | 490,420 | 1.3 | 3 | | |
| 22-May-08 | 483,425 | 1.6 | 4 | | |
| 23-May-08 | 453,046 | 1.0 | 5 | | |
| 26-May-08 | 327,109 | - | 1 | | |
| 27-May-08 | 464,516 | 0.2 | 2 | | |
| 28-May-08 | 480,879 | 0.1 | 3 | | |
| 29-May-08 | 481,636 | 0.7 | 4 | | |
| 30-May-08 | 470,445 | 1.5 | 5 | | |
| 2-Jun-08 | 486,074 | 2.8 | 1 | | |
| 3-Jun-08 | 501,945 | 3.2 | 2 | | |
| 4-Jun-08 | 497,220 | 1.9 | 3 | | |
| 5-Jun-08 | 514,825 | 4.2 | 4 | | |
| 6-Jun-08 | 479,156 | 3.3 | 5 | | |
| 9-Jun-08 | 529,353 | 11.8 | 1 | 97 | |
| 10-Jun-08 | 519,318 | 9.6 | 2 | | |
| 11-Jun-08 | 508,973 | 8.7 | 3 | | |
| 12-Jun-08 | 534,801 | 11.0 | 4 | | |
| 13-Jun-08 | 515,111 | 11.6 | 5 | 96 | |
| 16-Jun-08 | 506,289 | 8.0 | 1 | | |
| 17-Jun-08 | 528,354 | 9.2 | 2 | 95 | |
| 18-Jun-08 | 534,248 | 12.1 | 3 | | |
| 19-Jun-08 | 550,430 | 14.5 | 4 | 98 | |
| **20-Jun-08** | **544,268** | **19.5** | **5** | **103.2** | **X** |
| 23-Jun-08 | 511,975 | 8.1 | 1 | | |
| 24-Jun-08 | 530,964 | 8.3 | 2 | | |
| 25-Jun-08 | 522,482 | 7.1 | 3 | | |
| 26-Jun-08 | 528,335 | 7.4 | 4 | | |
| 27-Jun-08 | 503,760 | 10.8 | 5 | | |
| 30-Jun-08 | 505,490 | 8.0 | 1 | | |
| 1-Jul-08 | 510,351 | 7.7 | 2 | | |
| 2-Jul-08 | 518,861 | 8.5 | 3 | | |
| 3-Jul-08 | 514,146 | 9.7 | 4 | | |
| 4-Jul-08 | 376,261 | 7.4 | 5 | | |
| **7-Jul-08** | **566,085** | **17.2** | **1** | **104.2** | **X** |
| **8-Jul-08** | **583,342** | **20.9** | **2** | **107.2** | **X** |
| 9-Jul-08 | 572,732 | 21.0 | 3 | 105.6 | |
| **10-Jul-08** | **574,970** | **19.2** | **4** | **102.2** | **X** |
| 11-Jul-08 | 542,147 | 12.2 | 5 | | |

| | AMP Ld (HE 14) | wCDD | DOW | Max temp | Event-like day |
|---|---|---|---|---|---|
| 14-Jul-08 | 548,959 | 12.0 | 1 | | |
| 15-Jul-08 | 541,164 | 9.6 | 2 | | |
| 16-Jul-08 | 544,699 | 9.9 | 3 | | |
| 17-Jul-08 | 540,822 | 10.2 | 4 | | |
| 18-Jul-08 | 516,688 | 9.5 | 5 | | |
| 21-Jul-08 | 508,257 | 5.1 | 1 | | |
| 22-Jul-08 | 531,218 | 7.3 | 2 | | |
| 23-Jul-08 | 542,347 | 11.0 | 3 | | |
| 24-Jul-08 | 541,589 | 10.6 | 4 | | |
| 25-Jul-08 | 520,680 | 11.4 | 5 | | |
| 28-Jul-08 | 514,269 | 7.7 | 1 | | |
| 29-Jul-08 | 522,534 | 8.4 | 2 | | |
| 30-Jul-08 | 526,255 | 9.3 | 3 | | |
| 31-Jul-08 | 529,726 | 10.0 | 4 | | |
| 1-Aug-08 | 509,300 | 9.6 | 5 | | |
| 4-Aug-08 | 518,059 | 9.4 | 1 | | |
| 5-Aug-08 | 522,927 | 9.3 | 2 | | |
| 6-Aug-08 | 522,683 | 10.7 | 3 | | |
| 7-Aug-08 | 521,209 | 10.4 | 4 | | |
| 8-Aug-08 | 504,878 | 8.6 | 5 | | |
| 11-Aug-08 | 555,597 | 13.1 | 1 | 97.2 | |
| 12-Aug-08 | 548,655 | 12.9 | 2 | 97.8 | |
| **13-Aug-08** | **572,702** | **15.7** | **3** | **101.6** | **X** |
| 14-Aug-08 | 555,178 | 15.4 | 4 | 101.2 | |
| **15-Aug-08** | **554,229** | **16.2** | **5** | **101.8** | **X** |
| 18-Aug-08 | 547,062 | 7.2 | 1 | | |
| 19-Aug-08 | 546,683 | 6.1 | 2 | | |
| 20-Aug-08 | 567,727 | 8.1 | 3 | | |
| 21-Aug-08 | 574,845 | 10.6 | 4 | | |
| 22-Aug-08 | 558,487 | 11.5 | 5 | | |
| 25-Aug-08 | 566,229 | 12.7 | 1 | | |
| 26-Aug-08 | 574,105 | 11.7 | 2 | | |
| **27-Aug-08** | **585,571** | **16.1** | **3** | **101** | **X** |
| **28-Aug-08** | **596,961** | **19.3** | **4** | **104.6** | **X** |
| **29-Aug-08** | **568,149** | **19.3** | **5** | **103.8** | **X** |
| 1-Sep-08 | 388,091 | 7.0 | 1 | | |
| 2-Sep-08 | 553,231 | 10.8 | 2 | | |
| 3-Sep-08 | 569,000 | 13.5 | 3 | 98.2 | |
| **4-Sep-08** | **590,351** | **15.8** | **4** | **101.2** | **X** |
| 5-Sep-08 | 578,943 | 16.8 | 5 | 102.4 | |

[2] Note that three of the selected event-type days are in the last week of August, just prior to Labor Day. Detailed investigation indicated that a dozen or so manufacturing customers had unusually low loads on some of those days, which produced very large over-stated baseline estimates. These observations were left in the analysis on the basis that previous years have seen events called on days like that, just prior to or following summer holidays, which creates difficulty in modeling both usage and baselines on those days for a number of customers.

## 3. Approach

Two general alternative methods for constructing the baseline load for aggregators were examined, in both unadjusted and adjusted forms. These are the following:

1. *Aggregator-level baseline* – In this method, the hourly loads for all of an aggregator's nominated customers are summed, and the resulting aggregator loads are used to identify the highest 3-in-10 (as well as 5-in-10 and 10-in-10) days for each event-type day, and the average loads over the selected days are calculated. The resulting aggregator baselines are then compared to the *actual* aggregator load for each of the event-type days. This is the current baseline approach used for AMP, with the 3-in-10 averaging method.[3]
2. *Sum-of-customer baseline* – In this method, the hourly loads for each of an aggregator's customers are used separately to identify their highest 3-in-10 (or 5-in-10 and 10-in-10) days for each event-type day, the average loads over those three days are calculated, and then the individual customer baseline loads are summed up to produce a (different) aggregator baseline load for each event-type day. The resulting sum-of-customer baselines are then compared to the *actual* aggregator load for each of the event-type days.[4]

Two different methods were used for developing the "true" baselines to which the alternative baseline methods were compared, depending on whether the events being analyzed were *actual* event days or *event-type* days.[5] An advantage of using event-type days that were not actual event days is that consumers' *actual* loads on those days may be used as the *true* baseline for purposes of comparing alternative baselines which are estimated as averages of previous days' loads. In the case of actual events, the true baselines must be estimated, typically using information from regression analyses of customers' loads. For the actual events in this study, we constructed the "true" baseline for each customer as the sum of their observed load and our estimated load impact coefficients from the individual customer regressions described in Volume 1 of the report.[6] The true baseline for each aggregator and sub-group of customers, for each event, was then calculated as the sum of the individual baselines for the relevant customers.

### *3.1 Baseline performance statistics*

For each of the baseline methods, two statistics are calculated to compare the performance of estimated baselines to the true baselines (*e.g.*, the actual load on the event-type day). One statistic measures *accuracy*, while the other measures *bias*, or the tendency of a particular baseline method to under-state or over-state the true baseline.

---

[3] Three of the aggregators offered their customers a choice of an adjusted 3-in-10 baseline for 2008. Otherwise, the program baseline was an unadjusted aggregator-level 3-in-10 baseline.

[4] The primary difference between the two baselines is analogous to the difference between coincident and non-coincident demands. The sum-of-customers baseline adds together each customer's (non-coincident) average of highest three loads in the past ten days, while the aggregator baseline averages each customer's loads over the three (coincident) days that represent the *aggregator's* highest load. It is generally acknowledged that summing each individual customers' highest three loads will tend to produce a higher baseline than if the baseline is based on the highest (diversified) load of the aggregator.

[5] Days on which events were called for only some aggregators were included as event-type days for the aggregators who were not called.

[6] This method is analogous to the approach used to construct program reference loads in the ex post and ex ante load impact evaluations from the observed loads on event days and the estimated program load impacts.

### 3.1.1 Accuracy

Accuracy is measured using the *relative root mean square error* statistic (RRMSE, sometimes referred to as the Theil U-statistic). The formula for this statistic is the following:

$$U\text{-statistic} = [(1/n) \sum (e_h)^2]^{1/2} \,/\, [(1/n) \sum (L^A_h)^2]^{1/2} ,$$

where

| | |
|---|---|
| $e_h$ | $= (L^A_h - L^P_h)$, |
| $L^A_h$ | is *actual* load, |
| $L^P_h$ | is *predicted* (baseline) load, |
| $n$ | is the total number of event days and hours, and |
| the sum | is across event days and hours, for each aggregator, or sub-group by industry type. |

This statistic measures the degree of difference, or error, between the two data series, $L^P_h$ and $L^A_h$. It is nominally bounded by 0 and 1, with values closer to 0 indicating greater accuracy. Since the root-mean squared *errors* are normalized by the root-mean squared *load* levels, the resulting statistic is a normalized, or percentage measure of accuracy relative to the true baseline. For example, a value of 0.05 indicates an average 5 percent error in the baseline relative to its mean value.

### 3.1.2 Bias

The other statistic, which is used to measure the typical *direction* of error, is the *median % error*:

*Median percentage error* = Median of ($e_h$ / $L^A_h$), across event days and hours, for each aggregator, or sub-group by industry type.

This statistic has been used to measure the *bias* in the baseline load, indicating the extent to which a given baseline method tends to *over-state* or *under-state* the true baseline. While the median statistic serves to indicate the typical bias tendency, examining the *distribution* of percent errors provides insight into the full range of baseline errors. Finally, it is important to note that the convention of defining errors, $e_h$, as the difference between actual and estimated baseline values ($L^A_h - L^P_h$), implies that *positive* errors represent downward bias, or under-stated baselines, while *negative* errors represent upward bias, or over-stated baselines.

## *3.2 Adjusted baselines*

Two sets of *adjusted* versions of each of the baseline methods have also been assessed—a *symmetric* adjustment, and an *upward-only* adjustment. In both cases, the adjustments take the form of the ratio of the average load on the event day in the four hours prior to the event, to the average load in the same four hours of the unadjusted baseline, based on the highest three, five, or ten day approaches. The adjustment involves multiplying the unadjusted baseline times the adjustment ratio. The objective of the adjustment is to take advantage of information on customers' usage in the pre-event hours of an event day to improve the accuracy of an unadjusted baseline, which otherwise represents customers' usage on days that may be less extreme in terms of weather conditions than the event day. The symmetric adjustments were limited to no more than a 20 percent increase or decrease from the unadjusted baseline. For the upward-only adjustment, only positive adjustments were made.

## 4. Results for *Event-type Days*

### *4.1 Unadjusted baselines*

We begin by establishing a reference point of performance results for the alternative *unadjusted* baselines.

### 4.1.1 Accuracy

Table 3 shows *accuracy* results for unadjusted versions of the three alternative baseline methods that use different numbers of days in the baseline calculation (*e.g.*, 3, 5, or 10), and for the two different methods for calculating aggregate baselines—aggregator and sum-of-customers. Figures 1 and 2 plot the values in Table 3, providing a visual characterization of the pattern of results. The bars in the figures are grouped first by aggregator (see indications in Figure 1), then by industry type. The three bars in each panel show results for the different baseline calculation methods (*e.g.*, 3, 5, or 10-in-10). The following observations characterize some of the notable results:

- For the unadjusted 3-in-10 *aggregator* baseline, shown in the first column, and focusing first on the rows labeled TOTAL for each aggregator, relative errors range from about 5 to 6.5 percent across the aggregators, with an overall relative error of 5.6 percent.
- For the comparable unadjusted *sum-of-customer* baseline, shown in the first column of the second group of columns, the relative errors are generally somewhat larger, with an overall relative error of 7.5 percent.
- Moving across to the 5-in-10 and 10-in-10 columns, the overall relative errors for the aggregator baselines are generally higher, while the sum-of-customer values fall somewhat in the middle case and then increase to nearly the same values as the aggregator method. Patterns across aggregators and industry types vary.
- Comparing results by industry type, the findings suggest that the relative errors in unadjusted baselines for *commercial* customers are generally larger than those for *industrial* customers, and frequently are higher when more days are included in the baseline calculation.
- For both methods, *schools* often have among the highest relative errors.

**Table 3. Accuracy of Unadjusted Baselines**
***(Relative root mean square error, or Theil U-statistic)***

| | | *Aggregator* | | | *Sum of Customers* | | |
|---|---|---|---|---|---|---|---|
| | | **Unadjusted** | | | **Unadjusted** | | |
| **Agg.** | **Industry** | **3-in-10** | **5-in-10** | **10-in-10** | **3-in-10** | **5-in-10** | **10-in-10** |
| 1 | Industry | 0.050 | 0.055 | 0.071 | 0.060 | 0.042 | 0.070 |
| | Commercial | 0.062 | 0.077 | 0.104 | 0.050 | 0.065 | 0.103 |
| | TOTAL | 0.057 | 0.069 | 0.092 | 0.054 | 0.057 | 0.091 |
| 2 | Industry | 0.128 | 0.122 | 0.131 | 0.160 | 0.138 | 0.125 |
| | Commercial | 0.061 | 0.072 | 0.101 | 0.046 | 0.060 | 0.101 |
| | TOTAL | 0.065 | 0.074 | 0.102 | 0.055 | 0.065 | 0.102 |
| 3 | Industry | 0.048 | 0.055 | 0.079 | 0.069 | 0.051 | 0.079 |
| | Commercial | 0.067 | 0.074 | 0.098 | 0.044 | 0.057 | 0.097 |
| | Schools | 0.107 | 0.120 | 0.155 | 0.094 | 0.110 | 0.155 |
| | TOTAL | 0.049 | 0.056 | 0.080 | 0.068 | 0.052 | 0.080 |
| 4 | Industry | 0.059 | 0.049 | 0.040 | 0.125 | 0.096 | 0.040 |
| | Commercial | 0.045 | 0.052 | 0.069 | 0.030 | 0.041 | 0.069 |
| | Schools | 0.097 | 0.094 | 0.119 | 0.097 | 0.093 | 0.116 |
| | TOTAL | 0.061 | 0.053 | 0.049 | 0.120 | 0.093 | 0.049 |
| All | Industry | 0.053 | 0.057 | 0.076 | 0.082 | 0.062 | 0.075 |
| | Commercial | 0.061 | 0.073 | 0.101 | 0.047 | 0.061 | 0.100 |
| | Schools | 0.099 | 0.101 | 0.128 | 0.096 | 0.097 | 0.126 |
| | TOTAL | 0.056 | 0.062 | 0.083 | 0.075 | 0.062 | 0.083 |

**Figure 1. Accuracy of *Unadjusted* Baselines – *Aggregator***

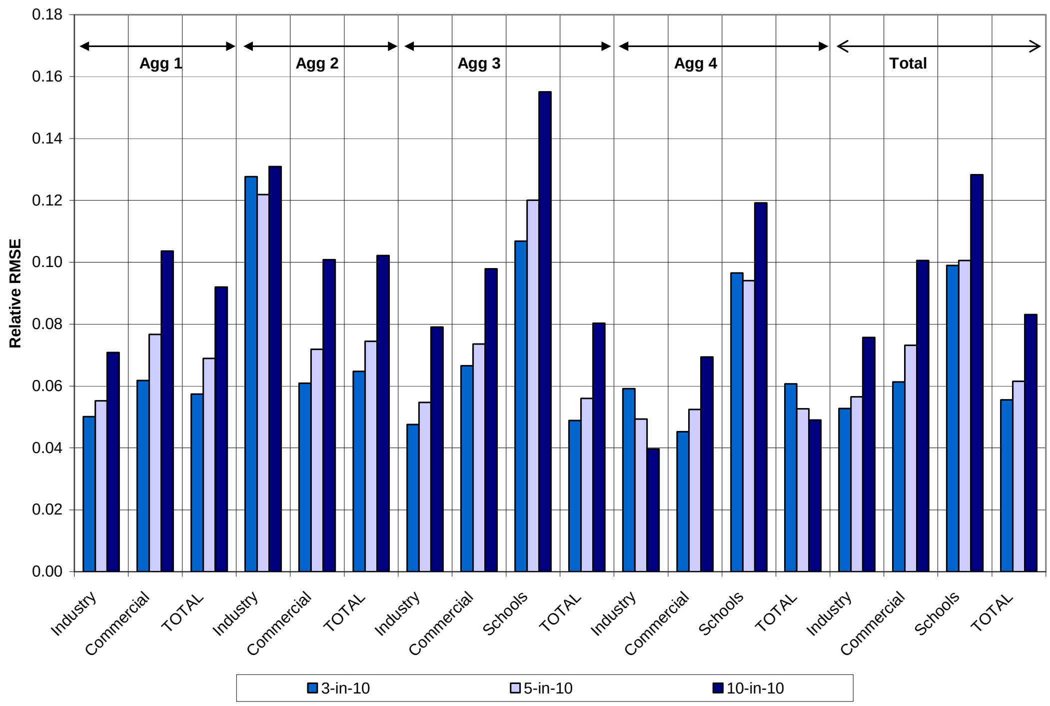
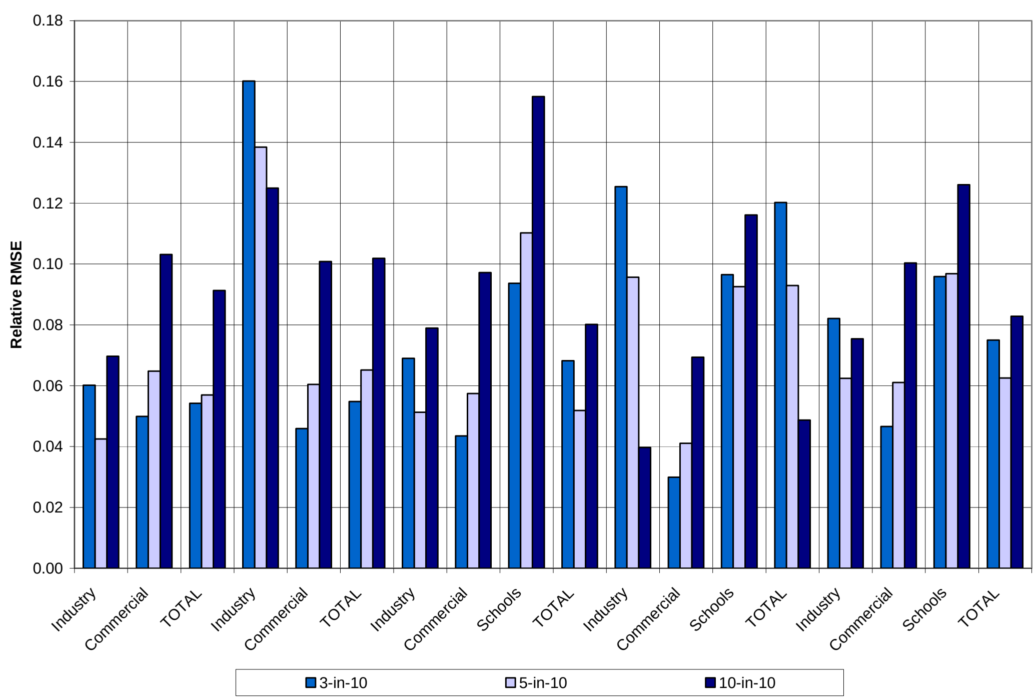



**Figure 2. Accuracy of *Unadjusted Baselines – Sum-of-Customers***



### 4.1.2 Bias

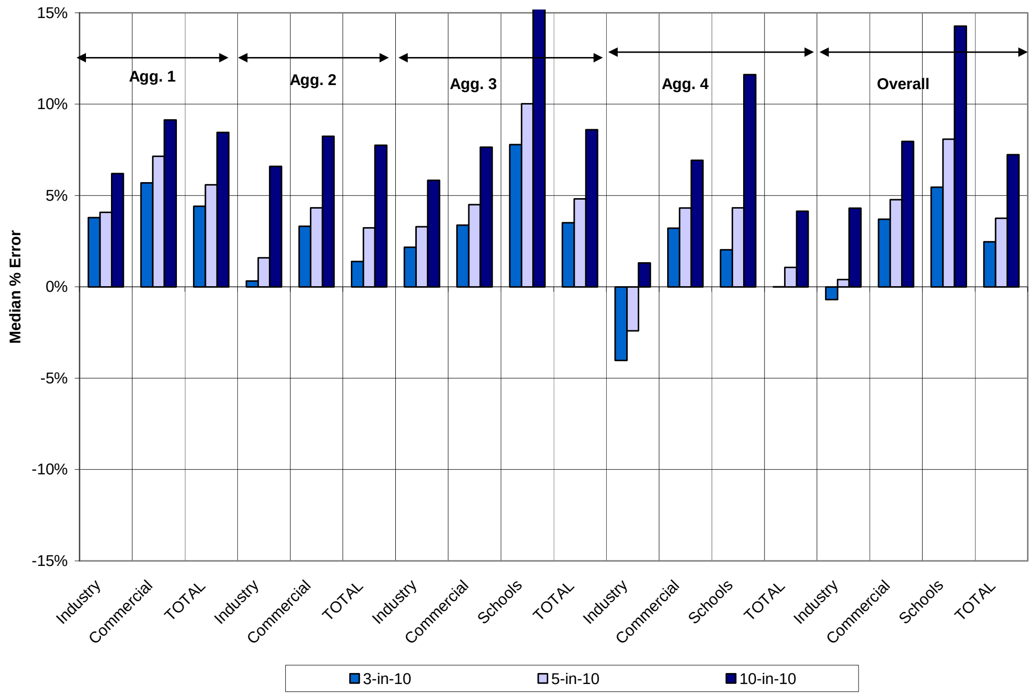
Table 4, and Figures 3 and 4 present comparable results for unadjusted baseline *bias*, showing the median percent errors across event-type days and hours, both by aggregator and overall. As noted above, *positive* errors (*i.e.*, estimated baseline is less than actual) indicate *under-stated* baselines, or downward bias, and *negative* errors indicate *over-stated* baselines, or upward bias. Observations include the following:

- The values in the TOTAL rows in the first column for the aggregator method are positive for each aggregator and overall, indicating that the unadjusted 3-in-10 aggregator baseline is biased downward (*i.e.*, typically *under-states* the true baseline), by about 2 to 3 percent.
- In contrast, the sum-of-customer method produces *over-stated* baselines ranging from less than one to nearly 5 percent.
- Looking across both aggregation methods, the overall downward bias of the unadjusted baseline tends to grow larger as the number of days included in the baseline average increases. This is not unexpected, particularly for weather-sensitive customers, as the included days may be increasingly milder than the event-type days.
- Looking at industry types, the downward bias of the unadjusted baselines is generally larger for *commercial* than for *industrial* customers, and the difference is greater with more days included in the baseline.

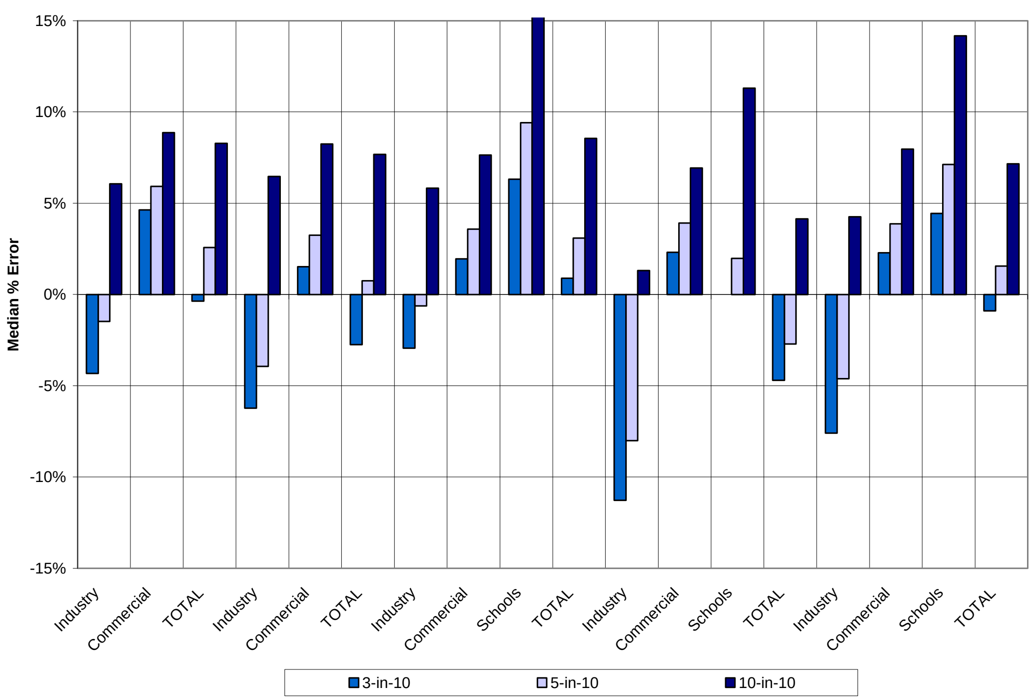
**Table 4. Bias of Unadjusted Baselines**
***(Median percent errors)***

| | | Aggregator | | | Sum of Customers | | |
|---|---|---|---|---|---|---|---|
| | | Unadjusted | | | Unadjusted | | |
| **Agg.** | **Industry** | **3-in-10** | **5-in-10** | **10-in-10** | **3-in-10** | **5-in-10** | **10-in-10** |
| 1 | Industry | 3.79% | 4.08% | 6.21% | -4.33% | -1.48% | 6.07% |
| | Commercial | 5.70% | 7.15% | 9.13% | 4.63% | 5.92% | 8.87% |
| | TOTAL | 4.42% | 5.59% | 8.45% | -0.37% | 2.57% | 8.28% |
| 2 | Industry | 0.33% | 1.59% | 6.59% | -6.23% | -3.94% | 6.47% |
| | Commercial | 3.32% | 4.33% | 8.25% | 1.52% | 3.26% | 8.25% |
| | TOTAL | 1.39% | 3.23% | 7.76% | -2.75% | 0.75% | 7.68% |
| 3 | Industry | 2.17% | 3.29% | 5.84% | -2.94% | -0.62% | 5.83% |
| | Commercial | 3.37% | 4.50% | 7.64% | 1.96% | 3.58% | 7.64% |
| | Schools | 7.79% | 10.03% | 15.55% | 6.32% | 9.41% | 15.55% |
| | TOTAL | 3.51% | 4.82% | 8.60% | 0.89% | 3.09% | 8.55% |
| 4 | Industry | -4.03% | -2.40% | 1.31% | -11.27% | -8.01% | 1.31% |
| | Commercial | 3.21% | 4.31% | 6.93% | 2.31% | 3.91% | 6.93% |
| | Schools | 2.02% | 4.32% | 11.62% | -0.01% | 1.98% | 11.30% |
| | TOTAL | 0.01% | 1.07% | 4.14% | -4.70% | -2.71% | 4.14% |
| All | Industry | -0.70% | 0.40% | 4.30% | -7.60% | -4.61% | 4.25% |
| | Commercial | 3.71% | 4.78% | 7.96% | 2.28% | 3.88% | 7.96% |
| | Schools | 5.46% | 8.08% | 14.28% | 4.43% | 7.13% | 14.17% |
| | TOTAL | 2.47% | 3.75% | 7.24% | -0.90% | 1.55% | 7.15% |

## Figure 3. Bias of *Unadjusted* Baselines – *Aggregator*




## Figure 4. Bias of *Unadjusted* Baselines – *Sum-of-Customers*



### 4.2 Adjusted baselines

## 4.2.1 Accuracy

Table 5, and Figures 5 and 6 report accuracy results for the various alternative adjustment methods, for the aggregator and sum-of-customer baselines. Key findings include the following:

- Focusing first on the TOTAL rows, the symmetric morning adjustment generally improves baseline accuracy substantially, reducing relative errors by half or more in many cases compared to the unadjusted baselines.
- For the *aggregator* baseline in particular, the relative errors of the adjusted baselines are very similar across the number of days included in the baseline, even for the upward-only adjustment method.
- Differences in relative accuracy for the sum-of-customer baseline are greater, with the adjusted 10-in-10 baseline generally showing the greatest accuracy, and the upward-only adjustment alternative generally producing somewhat larger relative errors than the corresponding symmetric adjustment.
- The adjusted 5-in-10 and 10-in-10 baselines are substantially more accurate than the unadjusted, with relative errors approximately half that of unadjusted versions.
- Looking across industry types, the adjusted baselines for *commercial* customers are generally more accurate than those for *industrial* customers, though in many cases the differences are not great, and the adjusted baselines for *schools* are the least accurate.
- The adjusted *aggregator* baselines are generally somewhat more accurate than the *sum-of-customer* baselines, especially so for industrial customers and for the upward-only adjustments, where the relative errors of the sum-of-customer baselines are often substantially larger than for the aggregator baselines.
- The two upward-only adjustments reduce the accuracy of the aggregator baseline only slightly compared to the symmetric adjustments, but reduce the accuracy of the sum-of-customers baseline more substantially for some aggregators and industry types.

**Table 5. Accuracy of Adjusted Baselines**
***(Relative root mean square error, or Theil U-statistic)***

| | | *Aggregator* | | | | | *Sum of Customers* | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Symmetric Adjustment | | | *Upward-only* | | Symmetric Adjustment | | | *Upward-only* | |
| Agg. | Industry | 3-in-10 | 5-in-10 | 10-in-10 | 5-in-10 | 10-in-10 | 3-in-10 | 5-in-10 | 10-in-10 | 5-in-10 | 10-in-10 |
| 1 | Industry | 0.028 | 0.025 | 0.023 | 0.025 | 0.023 | 0.050 | 0.036 | 0.035 | 0.067 | 0.033 |
| | Commercial | 0.017 | 0.021 | 0.022 | 0.020 | 0.022 | 0.018 | 0.015 | 0.019 | 0.018 | 0.015 |
| | TOTAL | 0.022 | 0.023 | 0.022 | 0.022 | 0.022 | 0.034 | 0.025 | 0.027 | 0.044 | 0.024 |
| 2 | Industry | 0.061 | 0.054 | 0.054 | 0.116 | 0.087 | 0.106 | 0.090 | 0.062 | 0.149 | 0.094 |
| | Commercial | 0.022 | 0.026 | 0.025 | 0.026 | 0.025 | 0.026 | 0.025 | 0.024 | 0.026 | 0.023 |
| | TOTAL | 0.025 | 0.028 | 0.027 | 0.034 | 0.030 | 0.033 | 0.030 | 0.026 | 0.039 | 0.029 |
| 3 | Industry | 0.022 | 0.021 | 0.020 | 0.025 | 0.020 | 0.044 | 0.037 | 0.034 | 0.073 | 0.034 |
| | Commercial | 0.021 | 0.019 | 0.021 | 0.019 | 0.021 | 0.024 | 0.022 | 0.024 | 0.025 | 0.022 |
| | Schools | 0.050 | 0.043 | 0.049 | 0.051 | 0.047 | 0.039 | 0.040 | 0.050 | 0.055 | 0.049 |
| | TOTAL | 0.022 | 0.021 | 0.020 | 0.025 | 0.020 | 0.043 | 0.037 | 0.034 | 0.071 | 0.033 |
| 4 | Industry | 0.045 | 0.040 | 0.037 | 0.055 | 0.039 | 0.091 | 0.075 | 0.043 | 0.124 | 0.065 |
| | Commercial | 0.017 | 0.017 | 0.020 | 0.018 | 0.020 | 0.014 | 0.017 | 0.023 | 0.019 | 0.020 |
| | Schools | 0.051 | 0.052 | 0.061 | 0.062 | 0.030 | 0.054 | 0.049 | 0.036 | 0.087 | 0.065 |
| | TOTAL | 0.044 | 0.039 | 0.037 | 0.053 | 0.037 | 0.087 | 0.071 | 0.041 | 0.118 | 0.063 |
| All | Industry | 0.031 | 0.029 | 0.028 | 0.036 | 0.029 | 0.057 | 0.048 | 0.039 | 0.084 | 0.043 |
| | Commercial | 0.021 | 0.024 | 0.023 | 0.023 | 0.023 | 0.023 | 0.021 | 0.022 | 0.023 | 0.020 |
| | Schools | 0.051 | 0.050 | 0.058 | 0.060 | 0.035 | 0.051 | 0.047 | 0.040 | 0.081 | 0.061 |
| | TOTAL | 0.029 | 0.028 | 0.027 | 0.034 | 0.028 | 0.051 | 0.043 | 0.036 | 0.074 | 0.039 |

**Figure 5. Accuracy of *Adjusted* Baselines – *Aggregator***

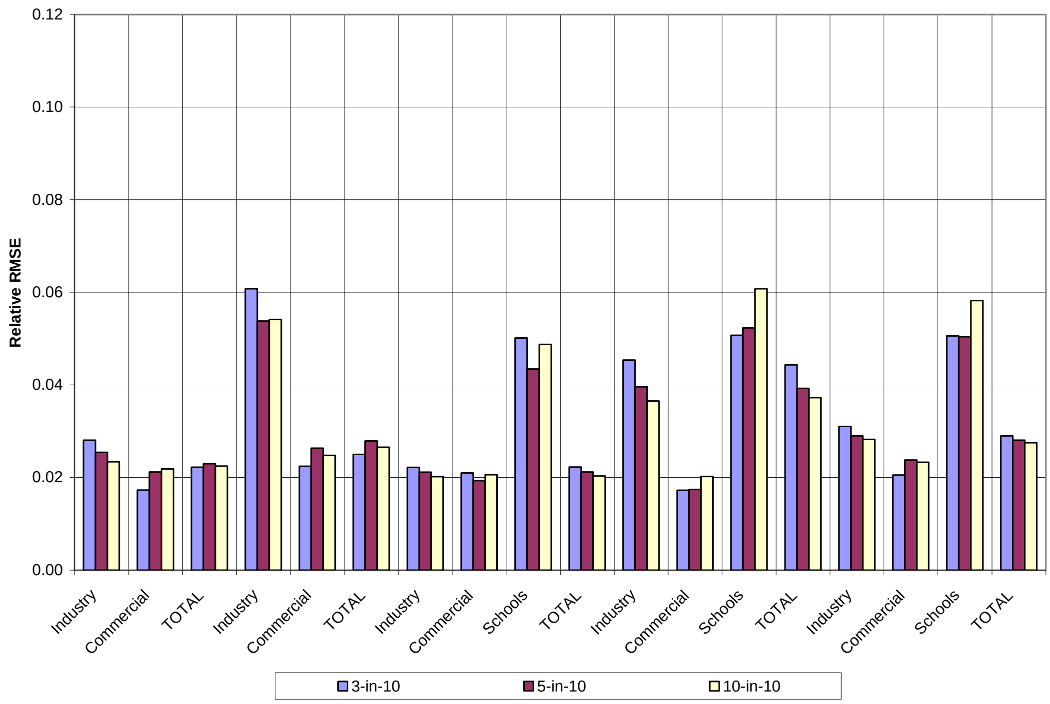
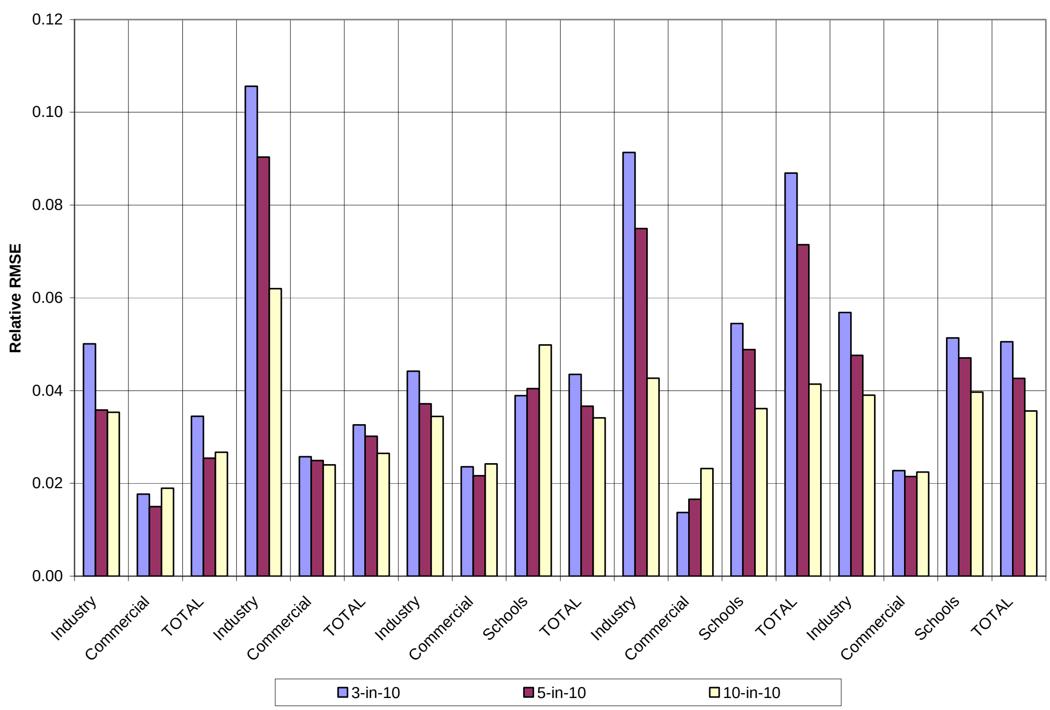



**Figure 6. Accuracy of *Adjusted* Baselines – *Sum of Customer***



### 4.2.2 Bias

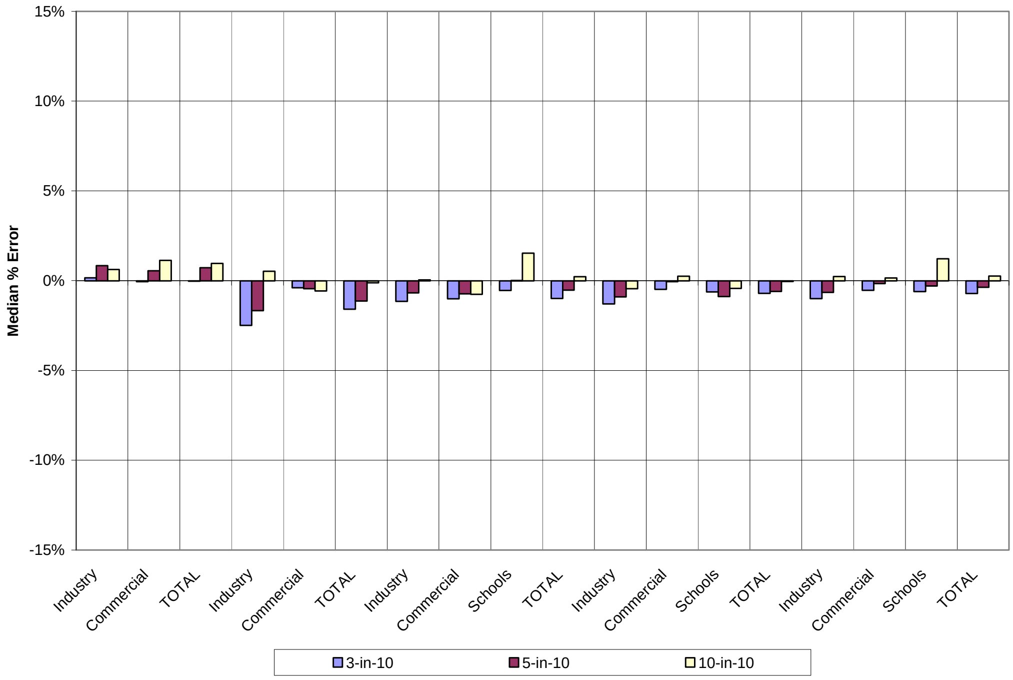
Table 6, and Figures 7 and 8 report bias results for the alternative adjustment methods. Key results are the following:

- At the TOTAL level for the aggregator method, and looking first at the first column, the morning adjustments generally reduce the magnitude of the bias, typically converting a downward bias (under-statement) of the unadjusted 3-in-10 baselines to a small upward bias (*e.g.*, a negative value of less than one percent for three of the four aggregators). (Compare Figure 7 to Figure 3.)
- In contrast, the morning adjustments to the 3-in-10 sum-of-customer baselines generally increase a small upward bias. (Compare Figure 8 to Figure 4.)
- Looking across columns as the number of days included in the *aggregator* baseline increases, the extent of the small upward bias appears to decrease, to the point that the biases for the adjusted 10-in-10 baseline are very small under-statements or over-statements. Across all customers, the median % error is less than one percent. (See the value of 0.26% in the last row of the third column.)
- For the *sum-of-customers* baseline, the median percent error across all customers changes somewhat more than for the aggregator baseline, from an upward bias of 2.25% for the adjusted 3-in-10, to a downward bias of 0.70% for the adjusted 10-in-10.
- Looking *across industry types*, there are few consistent patterns for the aggregator baselines, though the biases for commercial types are generally smaller than industrial, and the adjusted 10-in-10 baseline usually shows the smallest bias.
- For the sum-of-customers method, the adjusted baselines for *industrial* customers are generally biased upward (*i.e.*, the median % errors take on negative values) by more than those for commercial customers, but this feature is reduced by moving from the 3-in-10 to 10-in-10 baseline.
- The upward bias and difference between industrial and commercial customer types is particularly evident for the *upward-only adjustment* for the 5-in-10 baseline (*e.g.*, see the second to last columns in each group.
- The upward-only adjustments to the 5-in-10 and 10-in-10 baselines *increase* the bias of the aggregator baseline modestly, particularly for the fourth aggregator, but increase the bias more substantially for the sum-of-customers baseline, not unexpectedly producing greater *upward bias*, which for the 5-in-10 is around 4 percent.

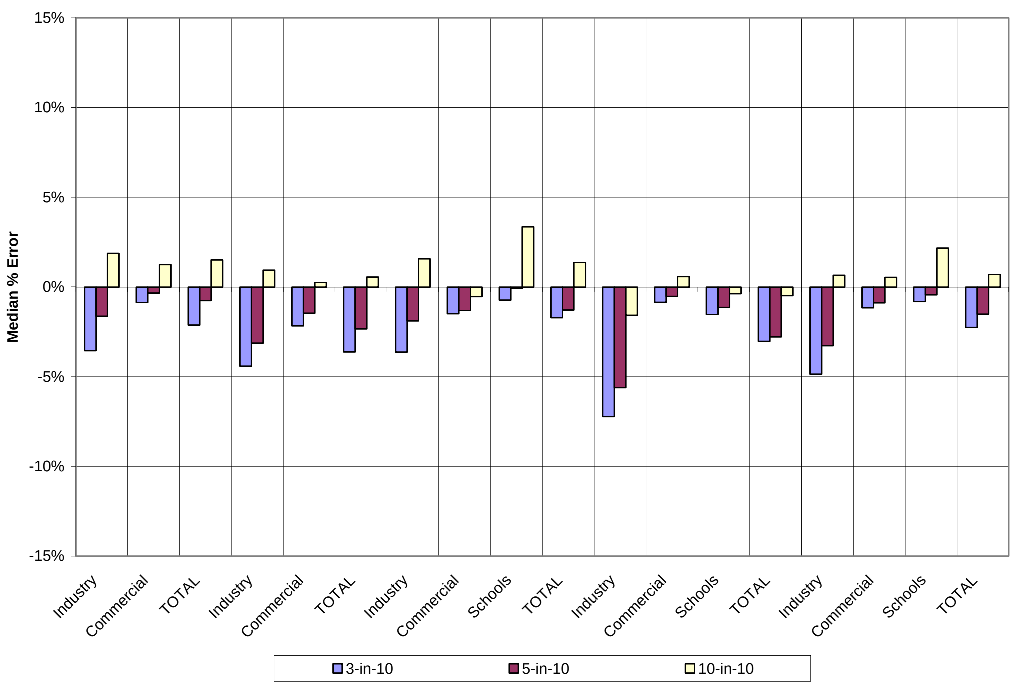
## Table 6. Bias of Adjusted Baselines
### *(Median percent errors)*

| | | *Aggregator* | | | | | *Sum of Customers* | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Symmetric Adjustment | | | *Upward-only Adjustment* | | Symmetric Adjustment | | | *Upward-only Adjustment* | |
| **Agg.** | **Industry** | **3-in-10** | **5-in-10** | **10-in-10** | **5-in-10** | **10-in-10** | **3-in-10** | **5-in-10** | **10-in-10** | **5-in-10** | **10-in-10** |
| 1 | Industry | 0.16% | 0.84% | 0.63% | 0.84% | 0.63% | -3.55% | -1.63% | 1.87% | -4.95% | -0.12% |
| | Commercial | -0.06% | 0.56% | 1.12% | 0.56% | 1.12% | -0.86% | -0.34% | 1.25% | -1.19% | 0.98% |
| | TOTAL | -0.03% | 0.72% | 0.97% | 0.72% | 0.97% | -2.12% | -0.76% | 1.51% | -2.81% | 0.64% |
| 2 | Industry | -2.49% | -1.67% | 0.52% | -3.19% | -1.86% | -4.42% | -3.13% | 0.94% | -7.10% | -0.80% |
| | Commercial | -0.39% | -0.45% | -0.57% | -0.97% | -0.57% | -2.17% | -1.47% | 0.25% | -2.65% | -0.24% |
| | TOTAL | -1.59% | -1.13% | -0.12% | -2.41% | -1.17% | -3.63% | -2.33% | 0.56% | -4.49% | -0.51% |
| 3 | Industry | -1.15% | -0.68% | 0.05% | -1.38% | -0.43% | -3.63% | -1.90% | 1.57% | -5.58% | -0.04% |
| | Commercial | -1.00% | -0.73% | -0.76% | -0.76% | -0.76% | -1.49% | -1.31% | -0.53% | -1.75% | -0.89% |
| | Schools | -0.55% | 0.00% | 1.53% | -0.40% | 1.37% | -0.73% | -0.07% | 3.36% | -1.21% | 2.89% |
| | TOTAL | -0.98% | -0.52% | 0.22% | -0.92% | -0.05% | -1.72% | -1.29% | 1.37% | -2.75% | 0.33% |
| 4 | Industry | -1.29% | -0.91% | -0.45% | -3.69% | -1.63% | -7.23% | -5.61% | -1.58% | -11.44% | -4.63% |
| | Commercial | -0.49% | -0.06% | 0.25% | -0.28% | 0.25% | -0.85% | -0.53% | 0.58% | -1.37% | 0.22% |
| | Schools | -0.62% | -0.88% | -0.43% | -4.15% | -1.00% | -1.53% | -1.14% | -0.38% | -4.87% | -0.78% |
| | TOTAL | -0.70% | -0.59% | -0.05% | -2.29% | -0.80% | -3.03% | -2.79% | -0.48% | -5.31% | -2.14% |
| All | Industry | -1.00% | -0.65% | 0.23% | -2.29% | -1.04% | -4.86% | -3.27% | 0.65% | -7.80% | -2.10% |
| | Commercial | -0.54% | -0.16% | 0.15% | -0.25% | 0.15% | -1.16% | -0.88% | 0.53% | -1.68% | 0.11% |
| | Schools | -0.61% | -0.29% | 1.22% | -1.57% | 0.34% | -0.81% | -0.43% | 2.17% | -2.05% | 0.88% |
| | TOTAL | -0.71% | -0.36% | 0.26% | -1.29% | -0.38% | -2.25% | -1.52% | 0.70% | -3.76% | -0.40% |

## Figure 7. Bias of *Adjusted* Baselines – *Aggregator*




## Figure 8. Bias of *Adjusted* Baselines – *Sum of Customer*



### 4.2.3 Conclusions – Event-type days

The variability of the above results across aggregators and customer types suggests that baseline performance depends on a number of factors, and that conclusions regarding the performance of particular baseline methods are not definitive in all cases.[7] Nevertheless, some reasonably consistent findings may be reported on the key issues of interest to the utilities. These include the following:

1. An aggregator baseline approach appears to generally provide a more accurate estimate of the true baseline than a sum-of-customer baseline that is constructed as the sum of individual customer baselines. For unadjusted baselines, the difference in accuracy is modest, particularly as the number of days included in the baseline increases; for adjusted baselines, the difference is somewhat greater, but declines with the number of days included in the baseline.
2. For the AMP data, an unadjusted 3-in-10 aggregator baseline approach *under-stated* the true baseline by amounts ranging from less than 1 to somewhat more than 4 percent across aggregators. In contrast, the unadjusted 3-in-10 sum-of-customer baseline was biased upward, by about 1 percent overall. As the number of days included in the baseline increased, both methods produced larger downward biases, converging to a median percent error of around 7 percent for the unadjusted 10-in-10 baseline.
3. *Morning adjustments* to the 3-in-10 baseline improved both the accuracy and bias of the unadjusted version, particularly for the aggregator method. The biases of both methods were smallest for the adjusted 10-in-10 baseline.
4. The *accuracies* of adjusted versions of the 3-in-10, 5-in-10, and 10-in-10 aggregator baselines were quite similar, and somewhat more accurate than the comparable sum-of-customer baselines. The accuracies of the two methods appeared to converge somewhat as the number of days included in the baseline increased to 10-in-10.
5. The *biases* of adjusted versions of the 3-in-10, 5-in-10, and 10-in-10 aggregator baselines were also more similar than the unadjusted versions, and considerably smaller than the comparable sum-of-customer baselines. However, the biases of the two methods also appeared to converge somewhat as the number of days included in the baseline increased to 10-in-10.
6. The *upward-only* adjustments to the 5-in-10 and 10-in-10 baselines *increased* the bias of the aggregator baseline modestly, but increased the bias more substantially for the sum-of-customers baseline, not unexpectedly producing greater *upward bias*.

## *4.3 Distributions of relative errors*

While the median percent error provides a useful indicator of the tendency of a particular baseline method to under-state or over-state the true baseline, the single median value can mask a potentially wide range of relative (percent) errors across event days and hours. This section illustrates several features of the range of estimated baseline errors. The first part of the section focuses on results at the aggregator/industry-type level. The second part shows underlying results at the individual customer-account level. Given the relatively strong performance of the

[7] Additional calculations made but not reported here suggest that baseline performance can also depend on the nature and timing of events, such as whether they are isolated events that follow several days of non-event days, or are events that occur following one or more events, thus pushing back the days included in the baseline calculation farther away from the event day.

adjusted 10-in-10 baseline report above, the results in this section use percent errors calculated for that baseline method.

### 4.3.1 Distributions by aggregator and industry type

The following figures show the relationship between the relative (percentage) errors of the adjusted *aggregator* and *sum of customer* baselines (using the adjusted 10-in-10 baseline), where each point represents the average percent error across event hours for an aggregator, industry type and event-type day.[8] The values are sorted according to the value for the aggregator baseline. Figure 9 shows values across all industry types. For the most part, those errors range from –5 percent (indicating a five percent over-statement) to +5 percent, with a handful of outliers, and a small positive median value. The percent errors of the sum-of-customer baseline appear on average to lie above the values for the aggregator baseline (thus indicating a somewhat *higher* baseline), which is consistent with the difference in overall medians (0.70 percent for the sum-of-customers, versus 0.26 percent for the aggregator method).

**Figure 9. Average Event-Type Day % Errors for *Adjusted 10-in-10* Aggregator and Sum-of-Customer Baselines – *All Industry Types***

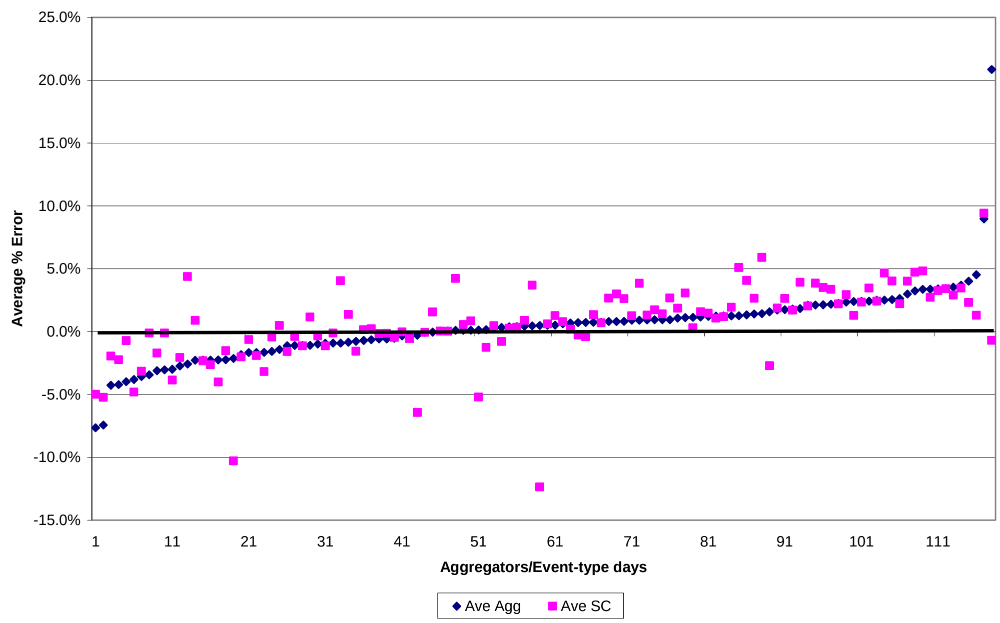



Figure 10 shows comparable values for *commercial* customer types only. In this case, the range of values is tighter and the differences between baseline-type are for the most part relatively small. Figure 11 shows values for *industrial* customer types. Here the underlying range of values is somewhat greater than for the commercial customers, and the differences between the aggregator and sum-of-customer baselines are greater. Figure 12 shows values for *schools*, which include several outliers with large errors.

[8] The percent error values across hours for a given event and aggregator tend to be quite similar, so that averaging errors across hours in an event simplifies the charts without discarding too much information.

**Figure 10. Average Event-Type Day % Errors for *Adjusted 10-in-10* Aggregator and Sum-of-Customer Baselines – *Commercial***

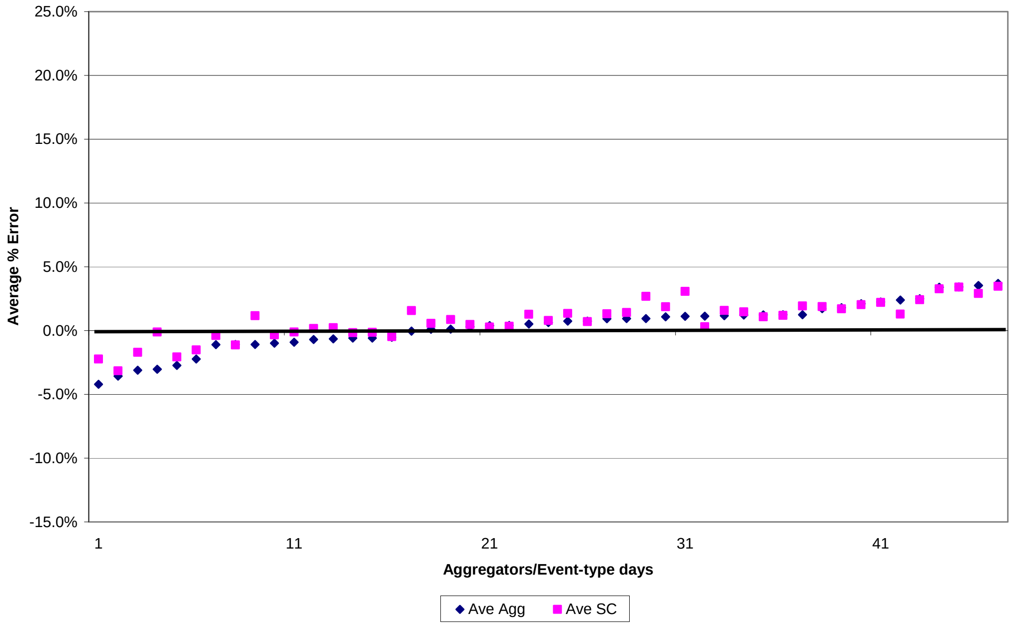



**Figure 11. Average Event-Type Day % Errors for *Adjusted 10-in-10* Aggregator and Sum-of-Customer Baselines – *Industrial***

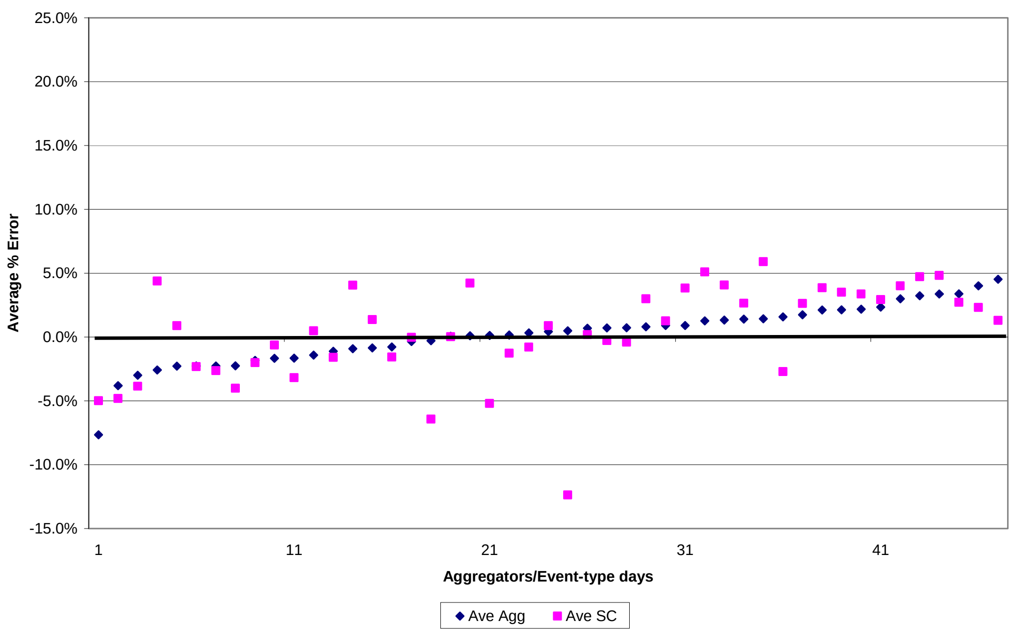
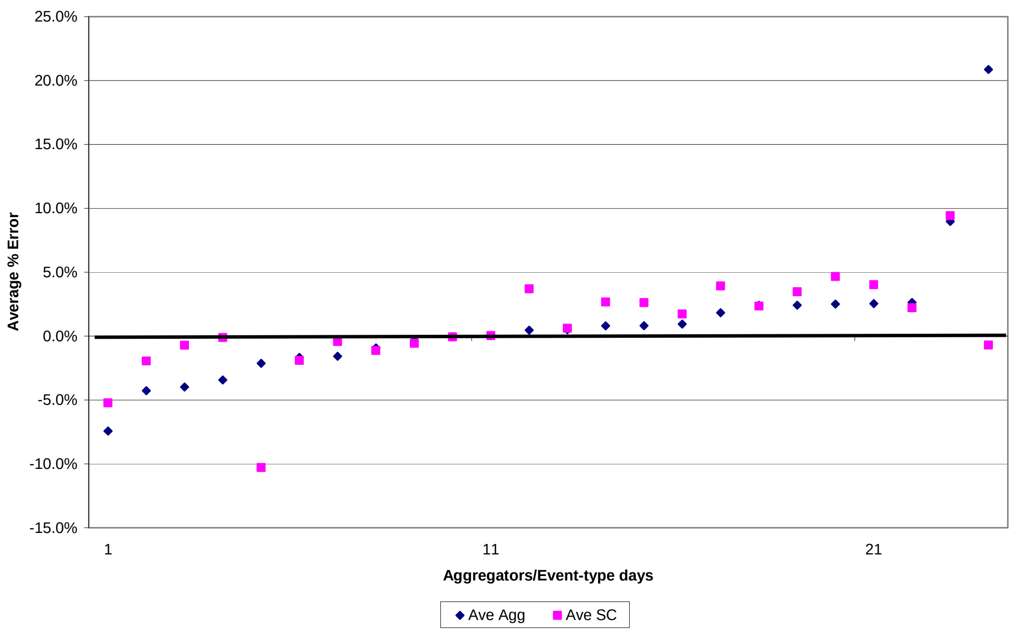


**Figure 12. Average Event-Type Day % Errors for *Adjusted 10-in-10* Aggregator and Sum-of-Customer Baselines – *Schools***

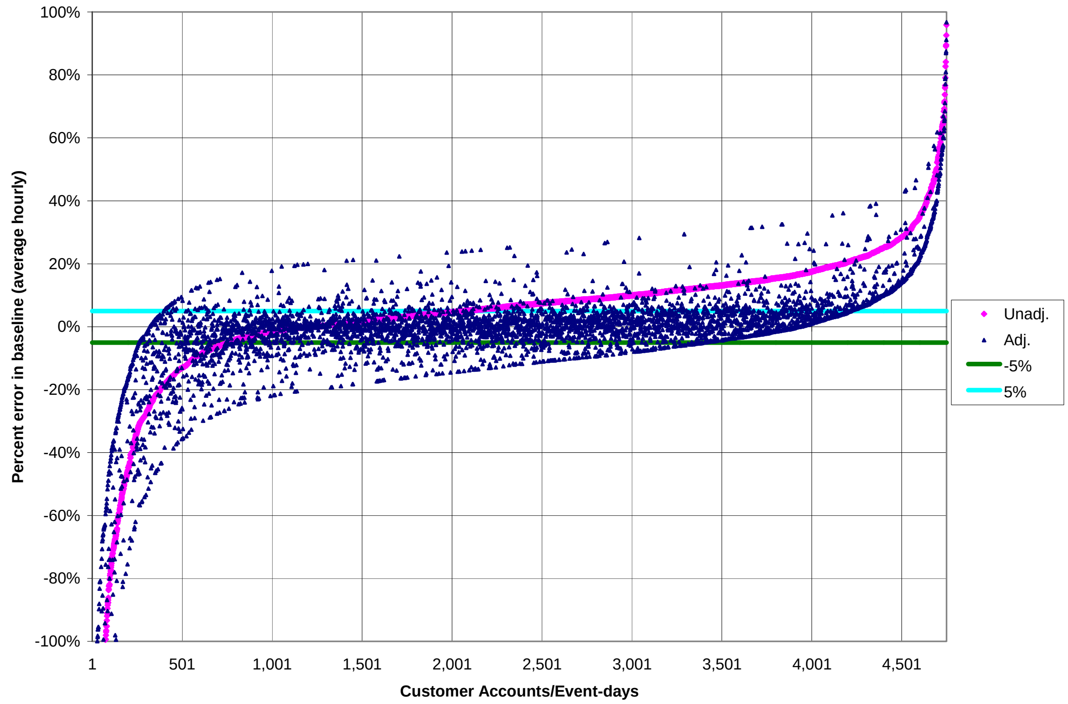



## 4.3.2. Customer-level distributions of baseline errors

The figures in this section are designed to illustrate the variability in relative errors at the customer level, which underlie the distributions shown in the above figures. Figure 13 shows the distributions of percent errors for unadjusted and adjusted 10-in-10 aggregator baselines. The points, which represent the average percent error for a customer and event-type day, are sorted by the values for the *unadjusted* baselines, thus providing an indication of the improvements in the percent errors due to the morning adjustments, as well as the breadth of the distributions across customers and event-type days. The unadjusted baseline *under-states* the true baseline in more than two-thirds of the cases (*i.e.*, the curve crosses the horizontal axis less than a third of the way from the origin), which is consistent with an estimated median percent error of positive 7.2 percent.[9] The *adjusted* baseline points show a relatively high density within about 5 percent on either side of the horizontal axis (see 5% lines in the figure), thus indicating the extent to which the adjustments reduce the baseline errors. The median percent error for the adjusted baseline across all customers and event-type days is 0.81 percent. The bounds on the distribution of errors for the adjusted baseline are due to the 80% limits set on the morning adjustment.

[9] Very large baseline over-statements (the initial tail of the distribution) occur when a customer's actual load during the event period on an event-type day is quite low relative to a baseline calculated by averaging usage across several previous days of irregular loads (*e.g.*, 100 kW actual load compared to a baseline load of 500 kW), resulting in a large negative error divided by a small actual baseline, thus producing a very large negative value (*e.g.*, (100 – 500) = –400, divided by 100, which implies a relative error of –400 percent). Recall that this baseline analysis used event-type days on which the customers did not actually face an event, and thus had no incentive (other than the existing peak demand charge) to reduce load.

**Figure 13. Distributions of Average Event-Type Day % Errors for *Unadjusted* and *Adjusted 10-in-10* Baselines – *Individual Customer Accounts***




Figures 14 and 15 are designed to investigate the potential importance of the large percent errors that occur in at least 5 percent of cases (around 250 customer/event-type days) at both ends of the distribution shown in Figure 13 (*e.g.*, whether large errors tend to be associated with small or large customers). Figures 14 and 15 plot average event-type day percent errors for the adjusted 10-in-10 baseline against customer size, measured by customers' average hourly usage during event periods on non-event days, for *industrial* and *commercial* customers respectively. Figure 14 illustrates a relatively wide range of percent errors (across the truncated horizontal axis) for industrial customers, but also demonstrates that most of the largest errors are associated with the smallest customer accounts. The errors are also distributed reasonably symmetrically around the origin (the median of the percent errors across all customer/events is 0.4%). Figure 15 shows that the range of percent errors for commercial customers is tighter, with fewer extremely large errors, and the bulk of the errors are grouped fairly tightly around the origin (the median of the percent errors across all customer/events is 1.1%, indicating a small under-statement in the true baseline). The largest errors are again associated with smaller customers.

**Figure 14. Average Event-Type Day % Errors for *Adjusted 10-in-10* Baselines, by Customer Size (Average Peak kW) – *Industrial Customers***

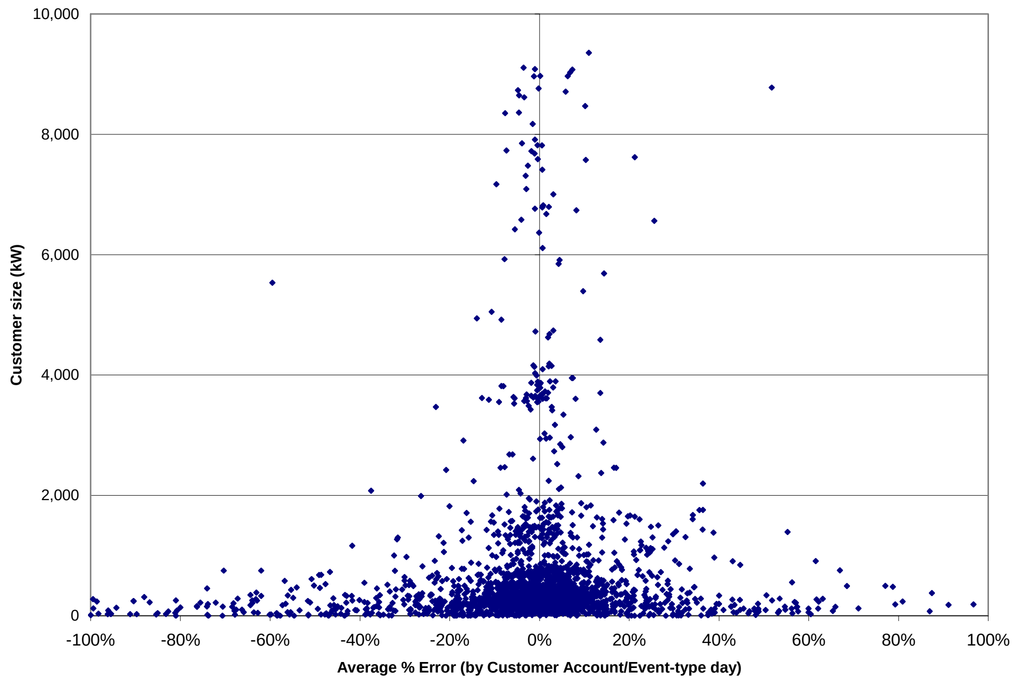
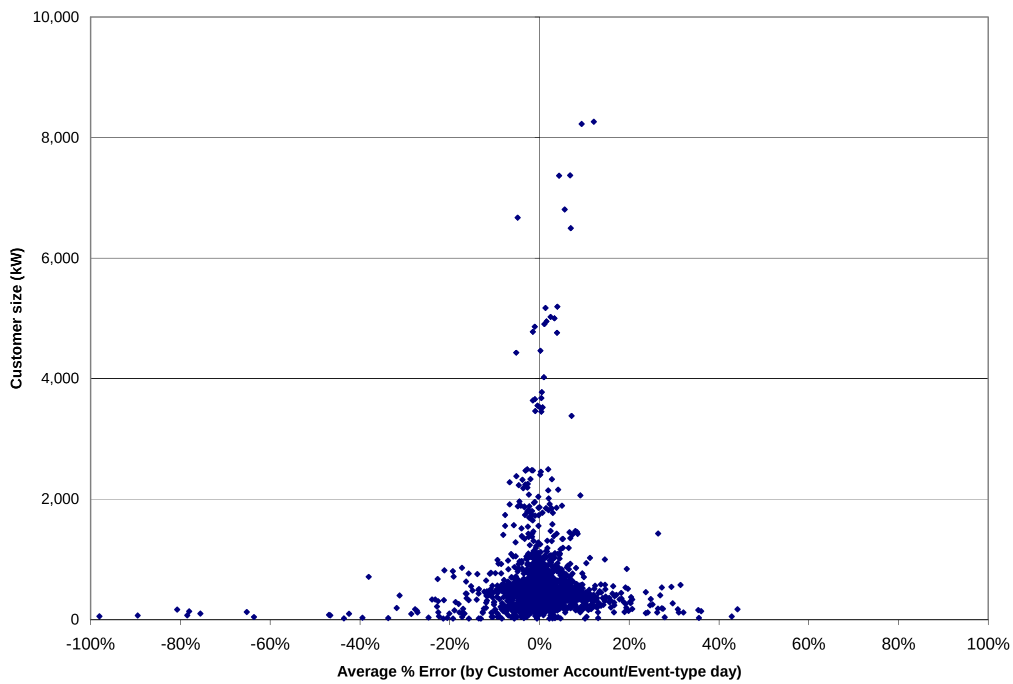



**Figure 15. Average Event-Type Day % Errors for *Adjusted 10-in-10* Baselines, by Customer Size (Average Peak kW) – *Commercial Customers***



# 5. Results for *Event Days*

This section presents baseline performance statistics for alternative baseline methods for event days, where baseline calculations are included for each aggregator that was called for each event. In this analysis, results are differentiated by both industry type and customers' choice of adjusted baseline.

## *5.1 Unadjusted baselines*

We begin by establishing a reference point of performance results for the alternative *unadjusted* baselines on AMP event days.

### 5.1.1 Accuracy

Table 7 shows *accuracy* results for unadjusted versions of the three different methods based on the number of days selected for inclusion in the baseline calculation (*e.g.*, 3, 5, or 10), and for the two different methods for calculating aggregate baselines—aggregator and sum-of-customers. Results are shown separately for those who accepted the adjusted baseline option and those who did not (see Table 1 for a summary of the numbers of customers at each aggregator who selected the adjusted baseline option). Portions of the table that are left un-shaded apply to the baseline choice for that group of customers. For example, the blocks of unadjusted baseline values for Aggregator 2's group of customers who selected the *adjusted* baseline are shaded, while the block for those that did not select it is not shaded. In the tables showing results for adjusted baselines, the pattern of shading is reversed, where the areas for those who actually accepted the adjusted baseline option are not shaded.

The following observations characterize some of the notable results:

- The accuracy results for the event days are qualitatively similar to those for event-type days presented in Section 4. For the *unadjusted 3-in-10 aggregator* baseline, shown in the first column, and focusing on the last group of "Total" rows, the relative errors averaged about 7 percent, for both those who selected the adjusted baseline option and those that did not.
- For the unadjusted sum-of-customer baseline, the relative errors were generally comparable, to somewhat larger than the aggregator results.
- The relative errors for the aggregator baseline generally increased with the number of days included in the baseline average.

**Table 7. Accuracy of Unadjusted Baselines – Event Days**
***(Relative root mean square error, or Theil U-statistic)***

| | | | *Aggregator* | | | *Sum of Customers* | | |
|---|---|---|---|---|---|---|---|---|
| | | | Unadjusted | | | Unadjusted | | |
| **Agg.** | **Adj. BL?** | **Industry** | **3-in-10** | **5-in-10** | **10-in-10** | **3-in-10** | **5-in-10** | **10-in-10** |
| | | Industry | 0.081 | 0.091 | 0.111 | 0.031 | 0.051 | 0.109 |
| | | Commercial | 0.107 | 0.139 | 0.175 | 0.100 | 0.131 | 0.174 |
| 1 | No | TOTAL | 0.100 | 0.127 | 0.158 | 0.085 | 0.113 | 0.158 |
| | | Industry | 0.084 | 0.081 | 0.139 | 0.100 | 0.082 | 0.126 |
| | | Commercial | 0.079 | 0.076 | 0.078 | 0.078 | 0.076 | 0.078 |
| | No | TOTAL | 0.084 | 0.081 | 0.139 | 0.100 | 0.082 | 0.126 |
| | | Industry | 0.149 | 0.144 | 0.131 | 0.187 | 0.165 | 0.131 |
| | | Commercial | 0.048 | 0.068 | 0.098 | 0.036 | 0.057 | 0.098 |
| 2 | Yes | TOTAL | 0.053 | 0.071 | 0.099 | 0.046 | 0.062 | 0.099 |
| | | Industry | 0.046 | 0.056 | 0.079 | 0.015 | 0.018 | 0.078 |
| | | Commercial | 0.045 | 0.066 | 0.093 | 0.022 | 0.042 | 0.090 |
| | | Schools | 0.072 | 0.076 | 0.111 | 0.075 | 0.072 | 0.111 |
| | No | TOTAL | 0.046 | 0.057 | 0.079 | 0.016 | 0.020 | 0.079 |
| | | Industry | 0.046 | 0.069 | 0.097 | 0.053 | 0.061 | 0.097 |
| | | Commercial | 0.035 | 0.069 | 0.116 | 0.037 | 0.068 | 0.116 |
| | | Schools | 0.156 | 0.158 | 0.211 | 0.153 | 0.155 | 0.211 |
| 3 | Yes | TOTAL | 0.046 | 0.070 | 0.102 | 0.051 | 0.064 | 0.102 |
| | | Industry | 0.099 | 0.089 | 0.051 | 0.161 | 0.132 | 0.051 |
| | | Commercial | 0.014 | 0.010 | 0.040 | 0.016 | 0.010 | 0.040 |
| | | Schools | n/a | n/a | n/a | n/a | n/a | n/a |
| | No | TOTAL | 0.099 | 0.088 | 0.051 | 0.160 | 0.131 | 0.051 |
| | | Industry | 0.110 | 0.106 | 0.110 | 0.162 | 0.138 | 0.110 |
| | | Commercial | 0.050 | 0.063 | 0.081 | 0.039 | 0.053 | 0.081 |
| | | Schools | 0.137 | 0.129 | 0.146 | 0.135 | 0.125 | 0.143 |
| 4 | Yes | TOTAL | 0.108 | 0.105 | 0.115 | 0.137 | 0.120 | 0.114 |
| | | Industry | 0.070 | 0.074 | 0.083 | 0.090 | 0.078 | 0.082 |
| | | Commercial | 0.100 | 0.131 | 0.165 | 0.093 | 0.122 | 0.164 |
| | | Schools | 0.072 | 0.076 | 0.111 | 0.075 | 0.072 | 0.111 |
| | No | TOTAL | 0.077 | 0.088 | 0.104 | 0.090 | 0.088 | 0.103 |
| | | Industry | 0.115 | 0.112 | 0.114 | 0.164 | 0.140 | 0.114 |
| | | Commercial | 0.048 | 0.068 | 0.097 | 0.036 | 0.056 | 0.097 |
| | | Schools | 0.137 | 0.129 | 0.146 | 0.135 | 0.125 | 0.143 |
| Total | Yes | TOTAL | 0.067 | 0.078 | 0.102 | 0.072 | 0.076 | 0.102 |
| | All | TOTAL | 0.074 | 0.085 | 0.103 | 0.086 | 0.085 | 0.103 |

### 5.1.2 Bias

Table 8 presents results for the bias in the unadjusted baselines. Observations include the following:

- All but one of the values in the "Total" group of rows at the bottom of the table in the first column for the aggregator method are positive, indicating the common result that the unadjusted 3-in-10 baseline is typically *biased downward*. The overall value was nearly 3 percent, with values typically ranging from near zero to 4 percent for different industry types.
- The sum-of-customer method produced more variable results, with a small upward bias overall, and a range of upward and downward biases for various aggregators and industry types.
- Moving across the number of days included in the baseline, both the aggregator and sum-of-customer methods showed increased downward biases, averaging 8 percent for the 10-in-10 method in both cases.

- The downward bias was generally somewhat larger for the commercial customer type than for the industrial type.

**Table 8. Bias of *Unadjusted* Baselines – Event Days**
***(Median percent errors)***

| | | | Aggregator | | | Sum of Customers | | |
|---|---|---|---|---|---|---|---|---|
| | | | | Unadjusted | | | Unadjusted | |
| Agg. | Adj. BL? | Industry | 3-in-10 Un | 5-in-10 Un | 10-in-10 Un | 3-in-10 Un | 5-in-10 Un | 10-in-10 Un |
| | | Industry | 10.43% | 12.17% | 14.44% | 2.78% | 7.21% | 14.10% |
| | | Commercial | 10.78% | 13.37% | 16.47% | 9.99% | 12.51% | 16.47% |
| 1 | No | TOTAL | 10.53% | 12.61% | 15.33% | 3.51% | 7.93% | 15.15% |
| | | Industry | -1.15% | -0.91% | 6.41% | -14.24% | -10.89% | 6.76% |
| | | Commercial | -0.97% | 1.17% | 4.85% | -1.50% | 0.85% | 4.85% |
| | No | TOTAL | -1.15% | 0.96% | 5.60% | -5.63% | -1.94% | 6.06% |
| | | Industry | -0.09% | 2.10% | 8.43% | -3.70% | 0.50% | 8.43% |
| | | Commercial | 5.41% | 8.71% | 11.33% | 3.45% | 7.26% | 11.33% |
| 2 | Yes | TOTAL | 2.54% | 4.50% | 8.45% | 0.97% | 3.43% | 8.45% |
| | | Industry | 4.02% | 5.23% | 8.02% | -1.80% | 0.94% | 8.00% |
| | | Commercial | 4.53% | 6.58% | 9.36% | 1.35% | 4.14% | 8.81% |
| | | Schools | 1.90% | 4.26% | 10.40% | -0.53% | 3.09% | 10.40% |
| | No | TOTAL | 4.24% | 5.73% | 8.07% | 0.05% | 2.66% | 8.02% |
| | | Industry | 3.34% | 3.87% | 7.90% | -2.20% | 1.78% | 7.90% |
| | | Commercial | 1.58% | 4.63% | 8.58% | 1.41% | 4.30% | 8.58% |
| | | Schools | 3.01% | 5.61% | 16.36% | 2.20% | 4.65% | 16.36% |
| 3 | Yes | TOTAL | 3.04% | 3.87% | 9.41% | -0.77% | 1.78% | 9.41% |
| | | Industry | -8.87% | -7.49% | -4.49% | -14.24% | -11.30% | -4.52% |
| | | Commercial | -0.86% | 0.45% | 3.61% | -1.45% | 0.47% | 3.61% |
| | | Schools | n/a | n/a | n/a | n/a | n/a | n/a |
| | No | TOTAL | -4.34% | -2.97% | 0.60% | -12.98% | -9.87% | 0.56% |
| | | Industry | -0.91% | 0.68% | 4.76% | -6.81% | -3.06% | 4.76% |
| | | Commercial | 4.93% | 5.91% | 7.56% | 3.61% | 5.04% | 7.56% |
| | | Schools | -0.44% | 1.50% | 7.47% | -1.79% | 0.34% | 6.84% |
| 4 | Yes | TOTAL | 2.86% | 3.68% | 6.92% | 0.00% | 1.76% | 6.92% |
| | | Industry | 3.07% | 4.91% | 8.04% | -3.22% | 0.39% | 8.02% |
| | | Commercial | 3.28% | 5.56% | 7.66% | 1.30% | 3.35% | 7.62% |
| | | Schools | 1.90% | 4.26% | 10.40% | -0.53% | 3.09% | 10.40% |
| | No | TOTAL | 3.14% | 5.27% | 8.00% | -1.31% | 1.78% | 7.95% |
| | | Industry | 0.27% | 2.41% | 8.23% | -4.82% | -1.16% | 8.23% |
| | | Commercial | 4.00% | 6.61% | 8.89% | 2.77% | 5.52% | 8.89% |
| | | Schools | -0.04% | 1.72% | 11.55% | -1.79% | 0.34% | 10.92% |
| Total | Yes | TOTAL | 2.82% | 3.85% | 8.25% | 0.93% | 2.98% | 8.25% |
| | All | TOTAL | 2.91% | 4.77% | 8.06% | -0.44% | 2.55% | 8.03% |

### *5.2 Adjusted baselines*

This section shows accuracy and bias results for *adjusted* versions of each of the alternative baseline methods for the event days. Note that the customers who did *not* select the adjusted baseline option actually faced the unadjusted 3-in-10 baseline in the AMP events, while those who did select the adjusted baseline faced the adjusted 3-in-10 baseline.

### 5.2.1 Accuracy

Table 9 shows *accuracy* results for the various adjusted versions of the two methods for aggregating customers. Key findings include the following:

- Focusing first on the adjusted aggregator 3-in-10 baseline results for all customers (the bottom sets of rows), the adjusted baseline showed substantially smaller relative errors than the corresponding unadjusted baseline in nearly every case, with an overall relative error of 2.7 percent compared to 7.4 percent for the unadjusted version.

- Similar results were obtained for the sum-of-customers method, though the relative errors were somewhat larger.
- With regard to choice of adjusted baseline, accuracy was improved with adjustment for both those who selected the option and those that did not.

**Table 9. Accuracy of Adjusted Baselines – Event Days (Relative root mean square error, or Theil U-statistic)**

| | | | *Aggregator* | | | *Sum of Customers* | | |
|---|---|---|---|---|---|---|---|---|
| | | | Symmetric Adjustment | | | Symmetric Adjustment | | |
| Agg. | Adj. BL? | Industry | 3-in-10 | 5-in-10 | 10-in-10 | 3-in-10 | 5-in-10 | 10-in-10 |
| | | Industry | 0.017 | 0.018 | 0.023 | 0.026 | 0.016 | 0.048 |
| | | Commercial | 0.018 | 0.022 | 0.054 | 0.020 | 0.030 | 0.060 |
| 1 | No | TOTAL | 0.017 | 0.021 | 0.047 | 0.022 | 0.027 | 0.057 |
| | | Industry | 0.060 | 0.046 | 0.043 | 0.086 | 0.063 | 0.066 |
| | | Commercial | 0.125 | 0.115 | 0.068 | 0.096 | 0.118 | 0.066 |
| | No | TOTAL | 0.061 | 0.046 | 0.043 | 0.086 | 0.063 | 0.066 |
| | | Industry | 0.036 | 0.038 | 0.022 | 0.115 | 0.100 | 0.057 |
| | | Commercial | 0.028 | 0.021 | 0.014 | 0.026 | 0.021 | 0.020 |
| 2 | Yes | TOTAL | 0.028 | 0.022 | 0.014 | 0.032 | 0.026 | 0.022 |
| | | Industry | 0.022 | 0.016 | 0.018 | 0.009 | 0.010 | 0.032 |
| | | Commercial | 0.015 | 0.015 | 0.013 | 0.037 | 0.025 | 0.011 |
| | | Schools | 0.050 | 0.037 | 0.025 | 0.049 | 0.036 | 0.020 |
| | No | TOTAL | 0.022 | 0.016 | 0.018 | 0.012 | 0.012 | 0.031 |
| | | Industry | 0.028 | 0.031 | 0.021 | 0.054 | 0.038 | 0.025 |
| | | Commercial | 0.023 | 0.032 | 0.017 | 0.024 | 0.033 | 0.014 |
| | | Schools | 0.100 | 0.075 | 0.098 | 0.108 | 0.078 | 0.095 |
| 3 | Yes | TOTAL | 0.028 | 0.031 | 0.021 | 0.050 | 0.037 | 0.024 |
| | | Industry | 0.030 | 0.028 | 0.025 | 0.113 | 0.100 | 0.051 |
| | | Commercial | 0.050 | 0.040 | 0.040 | 0.044 | 0.042 | 0.040 |
| | | Schools | n/a | n/a | n/a | n/a | n/a | n/a |
| | No | TOTAL | 0.031 | 0.028 | 0.025 | 0.112 | 0.099 | 0.051 |
| | | Industry | 0.046 | 0.052 | 0.059 | 0.092 | 0.079 | 0.073 |
| | | Commercial | 0.016 | 0.019 | 0.022 | 0.008 | 0.014 | 0.028 |
| | | Schools | 0.066 | 0.081 | 0.089 | 0.067 | 0.081 | 0.087 |
| 4 | Yes | TOTAL | 0.048 | 0.056 | 0.063 | 0.074 | 0.071 | 0.070 |
| | | Industry | 0.026 | 0.027 | 0.037 | 0.063 | 0.058 | 0.050 |
| | | Commercial | 0.019 | 0.022 | 0.050 | 0.024 | 0.030 | 0.056 |
| | | Schools | 0.050 | 0.037 | 0.025 | 0.049 | 0.036 | 0.020 |
| | No | TOTAL | 0.024 | 0.026 | 0.040 | 0.058 | 0.054 | 0.051 |
| | | Industry | 0.044 | 0.049 | 0.053 | 0.095 | 0.082 | 0.069 |
| | | Commercial | 0.028 | 0.021 | 0.015 | 0.026 | 0.020 | 0.020 |
| | | Schools | 0.066 | 0.081 | 0.089 | 0.067 | 0.081 | 0.087 |
| Total | Yes | TOTAL | 0.033 | 0.031 | 0.030 | 0.043 | 0.038 | 0.036 |
| | All | TOTAL | 0.027 | 0.028 | 0.037 | 0.054 | 0.050 | 0.047 |

### 5.2.2 Bias

Table 10 shows *bias* results for the various adjusted versions of the two methods for aggregating customers. Key findings include the following:

- In contrast to the typical downward bias of the unadjusted baselines, the adjusted baselines generally showed small upward biases. The overall median percent error indicated a 1 percent upward bias for the aggregator method, compared to nearly 3 percent downward bias for the unadjusted case.

- The 3-in-10 sum-of-customers baseline showed similar patterns, with somewhat greater upward biases than for the aggregator method.
- Adding more days to the baseline calculation reduced the upward bias somewhat.

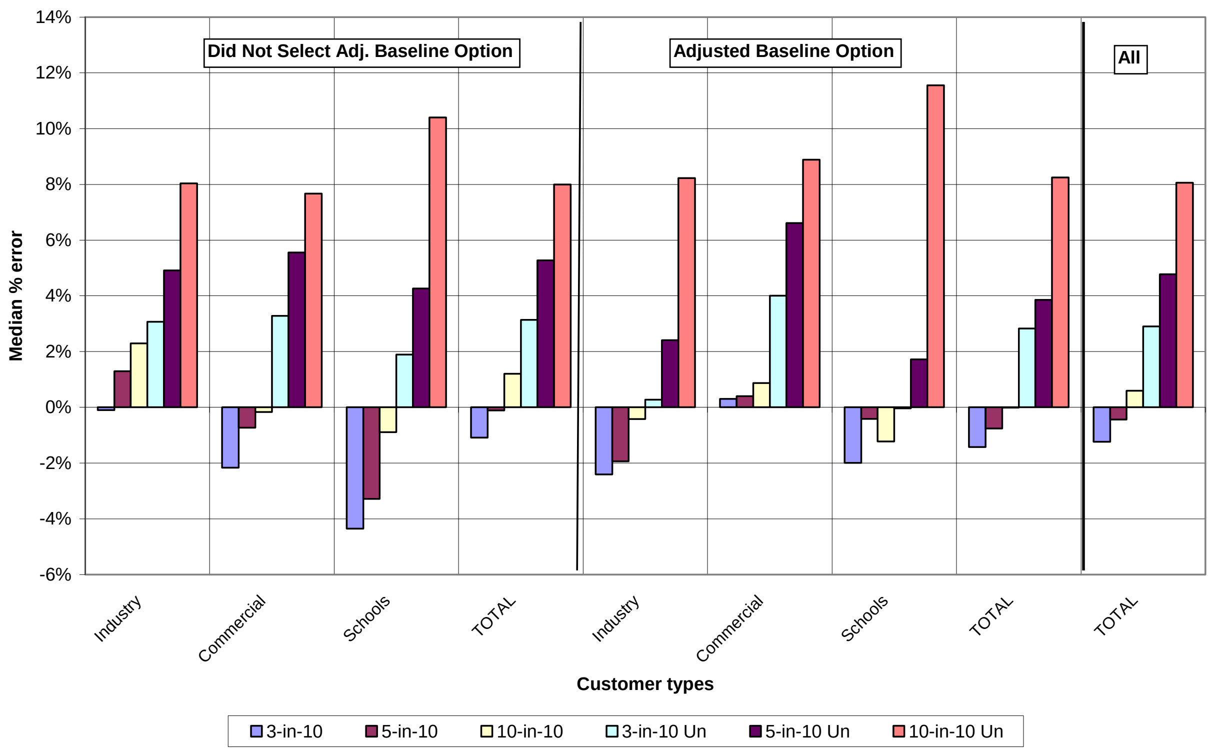
**Table 10. Bias of *Adjusted* Baselines – Event Days**
***(Median percent errors)***

| | | | Aggregator | | | Sum of Customers | | |
|---|---|---|---|---|---|---|---|---|
| | | | Symmetric Adjustment | | | Symmetric Adjustment | | |
| Agg. | Adj. BL? | Industry | 3-in-10 | 5-in-10 | 10-in-10 | 3-in-10 | 5-in-10 | 10-in-10 |
| 1 | No | Industry | 1.27% | 1.93% | 2.74% | -1.52% | 0.97% | 6.01% |
| | | Commercial | 1.23% | 1.72% | 2.00% | 1.27% | 1.71% | 3.62% |
| | | TOTAL | 1.25% | 1.92% | 2.39% | -0.63% | 1.33% | 4.00% |
| 2 | No | Industry | -1.84% | -3.55% | 1.96% | -10.17% | -4.43% | 1.95% |
| | | Commercial | -5.84% | -5.57% | -1.22% | -4.78% | -5.71% | -1.27% |
| | | TOTAL | -4.81% | -4.11% | 1.18% | -7.57% | -4.56% | 1.21% |
| | Yes | Industry | -3.93% | -3.70% | -1.21% | -3.60% | -2.54% | -0.25% |
| | | Commercial | -1.70% | -1.42% | 0.33% | -1.40% | -0.23% | 2.01% |
| | | TOTAL | -2.77% | -2.25% | -0.44% | -2.98% | -1.85% | 0.84% |
| 3 | No | Industry | 1.79% | 1.63% | 2.06% | 0.02% | 0.28% | 3.37% |
| | | Commercial | -1.17% | -0.91% | -0.77% | -3.28% | -2.16% | 0.28% |
| | | Schools | -4.36% | -3.29% | -0.89% | -4.18% | -2.66% | 1.76% |
| | | TOTAL | -0.79% | -0.67% | 0.61% | -2.39% | -1.42% | 2.18% |
| | Yes | Industry | -2.21% | -1.63% | -0.27% | -5.22% | -3.52% | 0.23% |
| | | Commercial | -2.37% | -2.69% | -1.69% | -2.50% | -2.54% | -1.32% |
| | | Schools | -7.23% | -4.09% | 3.65% | -7.13% | -4.75% | 4.82% |
| | | TOTAL | -2.65% | -1.85% | -1.10% | -3.39% | -3.52% | -0.21% |
| 4 | No | Industry | -2.67% | -2.03% | -2.23% | -11.32% | -9.25% | -4.30% |
| | | Commercial | -4.40% | -2.67% | -3.23% | -3.17% | -3.38% | -3.24% |
| | | Schools | n/a | n/a | n/a | n/a | n/a | n/a |
| | | TOTAL | -3.20% | -2.03% | -2.43% | -7.86% | -6.45% | -3.33% |
| | Yes | Industry | -0.80% | -1.02% | 0.16% | -2.75% | -1.10% | 1.99% |
| | | Commercial | 1.07% | 1.61% | 2.11% | 0.26% | 1.31% | 2.66% |
| | | Schools | -1.34% | -0.09% | -1.22% | -2.88% | -1.24% | 0.53% |
| | | TOTAL | 0.42% | 0.47% | 1.06% | -0.52% | 0.28% | 1.78% |
| Total | No | Industry | -0.10% | 1.30% | 2.29% | -2.60% | -0.29% | 3.38% |
| | | Commercial | -2.16% | -0.73% | -0.17% | -2.37% | -2.04% | 0.02% |
| | | Schools | -4.36% | -3.29% | -0.89% | -4.18% | -2.66% | 1.76% |
| | | TOTAL | -1.08% | -0.11% | 1.21% | -2.53% | -1.49% | 1.95% |
| | Yes | Industry | -2.40% | -1.94% | -0.42% | -4.03% | -2.69% | 0.04% |
| | | Commercial | 0.30% | 0.40% | 0.88% | -0.37% | 0.24% | 1.57% |
| | | Schools | -1.99% | -0.41% | -1.22% | -3.41% | -1.51% | 0.53% |
| | | TOTAL | -1.42% | -0.76% | -0.01% | -2.43% | -0.97% | 0.84% |
| | All | TOTAL | -1.24% | -0.44% | 0.60% | -2.46% | -1.34% | 1.15% |

Figure 16 provides a comparison of the bias results in the lower "Total" panel of Table 10 for the Aggregator method to the comparable results in Table 8 for unadjusted baseline. Each set of bars shows the median % errors for the three *adjusted* baselines and then the three *unadjusted* baselines. The first set of panels presents results for those customers who did *not* select the adjusted baseline option, while the second set of panels shows results for those who did select the option. The final set of bars shows results for all customers.

The figure clearly shows the typical result that the downward bias (positive median % error) of the unadjusted baseline becomes greater as the number of days included in the baseline average expands, with the largest bias for the 10-in-10 baseline. The figure also clearly shows the smaller biases of the adjusted baselines. As the number of days included in the baseline calculation increases from 3 to 10, the median percent errors appear to move in a positive direction, either reducing an upward bias, or moving from an upward bias to a downward bias.

**Figure 16. Comparison of Bias of *Adjusted* and *Unadjusted* Baselines – *Aggregator Method***




### *5.3 Conclusions for event days*

The performance of the alternative baseline methods on event days, in terms of accuracy and bias, was qualitatively similar to their performance on the *event-type* days presented in Section 4. In particular, adjusting the baseline for morning usage generally improves the accuracy and reduces the bias of the unadjusted baselines. Performance results varied considerably across aggregators and industry types. The adjusted 10-in-10 did not dominate the other methods as it appeared to do for the event-type days. However, it performed at least as well and often better than the other adjusted baselines.

## 6. Gaming

An issue of concern for adopting the adjusted baseline method is whether customers and aggregators would try or succeed in "gaming" the baseline by artificially increasing usage in the morning hours that are used to construct the adjustment factor. Such an increase could have the effect of increasing the baseline used for settlement, and hence the achieved load impacts on event days.

We looked for evidence of gaming among the aggregators who offered an adjusted baseline option and the customers who accepted it. We examined the issue from two directions. First we constructed aggregate load profiles for all of the AMP event and event-type days for each aggregator, by industry type and choice of adjusted baseline. We then examined the event-day loads for evidence of increases in usage prior to the events compared to typical usage patterns in the same hours on event-type days. Second, for all AMP customers we calculated average usage in pre-event hours on both event days and event-type days, and examined the ratios of the two values for evidence of significantly higher values on event days. Results of both methods are summarized below.

### *6.1 Comparison of loads on event-days and event-type days*

The following charts show aggregated loads for sub-groups of customers (*e.g.*, by industry type and choice of adjusted baseline) for aggregators that offered an adjusted baseline option, and for which a reasonable number of customers selected the option. The load profiles are shown for the five event days (only some aggregators were called for some of the events) and the ten event-type days that were used in the baseline analysis. Bold lines indicate days on which events were called for that aggregator. For some aggregators and sub-groups, the loads appear to be grouped at two different usage levels. This typically occurs due to customers being added to or removed from the group during the summer period, through changes in monthly nominations. Graphs are shown primarily for sub-groups that selected the adjusted baseline. In a few cases, graphs are also shown for groups that did not select the adjusted baselines.

Looking across the figures, the load reductions on event days are usually quite evident, with the industrial customers typically showing the largest reductions during the event hours. Across the aggregators and sub-groups, there is only one instance of an event-day load that takes on a shape potentially indicative of gaming. That instance is shown in Figure 18, for the industrial sub-group of the second aggregator. This is a relatively small group, with peak-period demand of about 10 MW. On further investigation, the group is dominated by one large customer who joined mid-way through the summer, which explains the two different typical load profile levels for the group prior to and after August 1. Examination of that customer's loads indicates somewhat variable loads, like many industrial customers, on some days operating at levels that are half that on other days. On the event day in question, September 5, the customer's load began at a level suggestive of a lower level of operations, particularly following a similar pattern as on the previous day. However, around noon the load increased by nearly 2 MW and stayed there until the hour prior to the event (hour 15), at which time it dropped by about 2 MW.

It is not possible to know whether this load profile is indicative of actual gaming behavior, or an attempt, for example, to increase needed production temporarily prior to the event. However, the load increase is at least suggestive of how such gaming behavior could be conducted.

**Figure 17. Aggregator 2; Industrial; *No Adjusted BL***

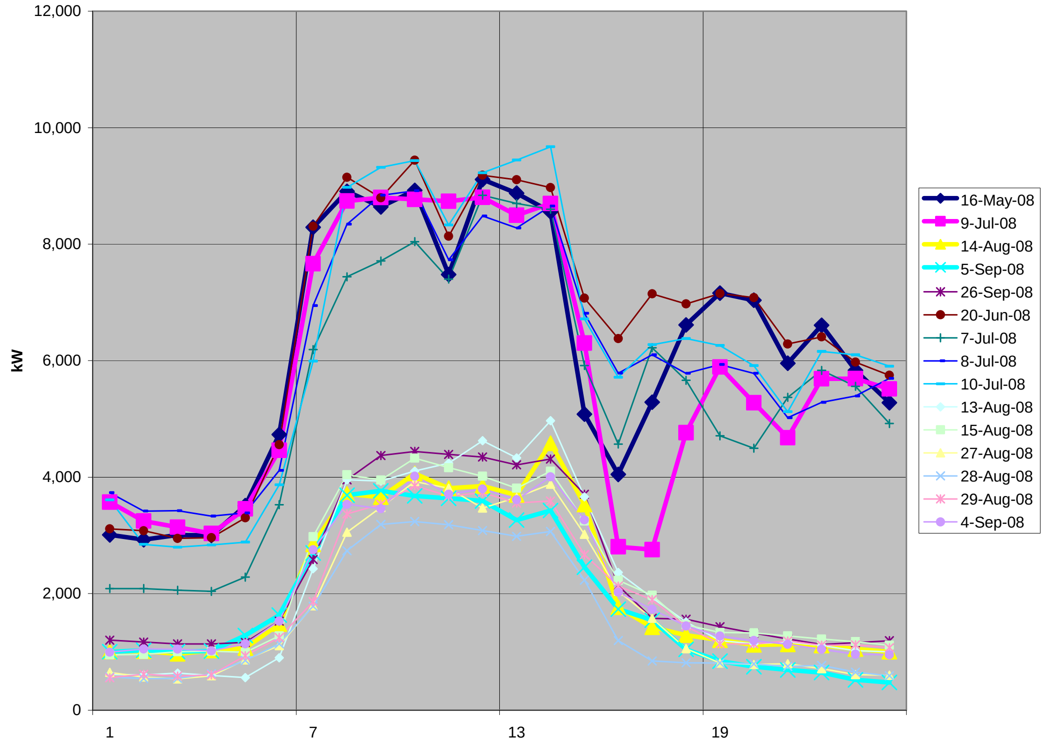
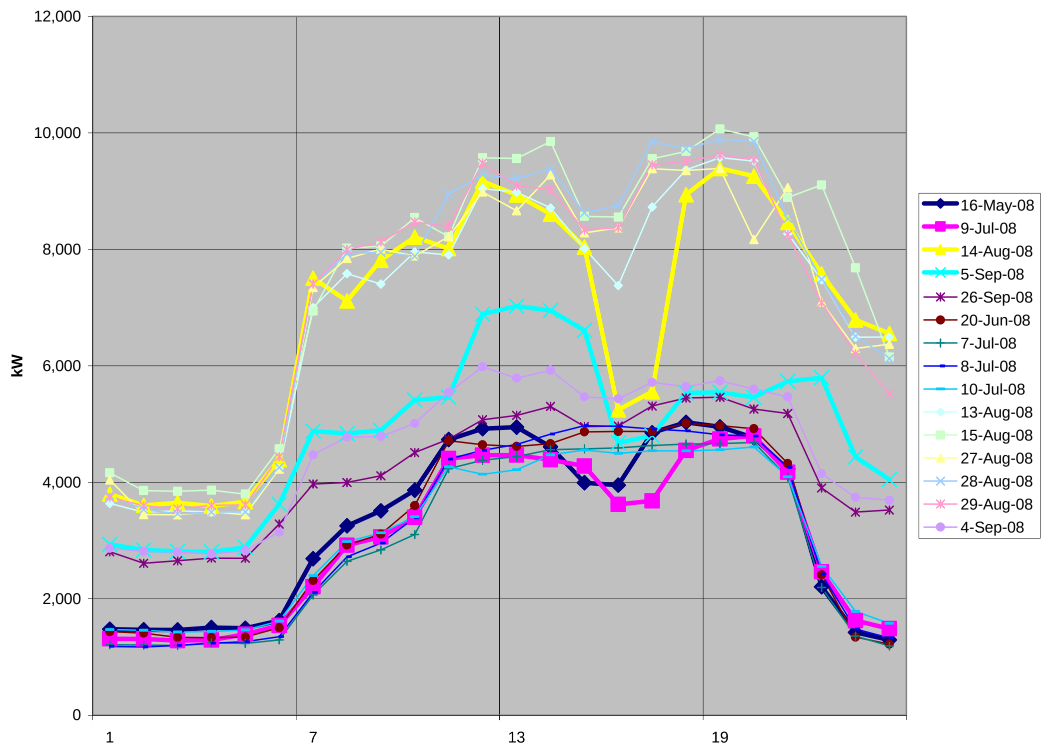
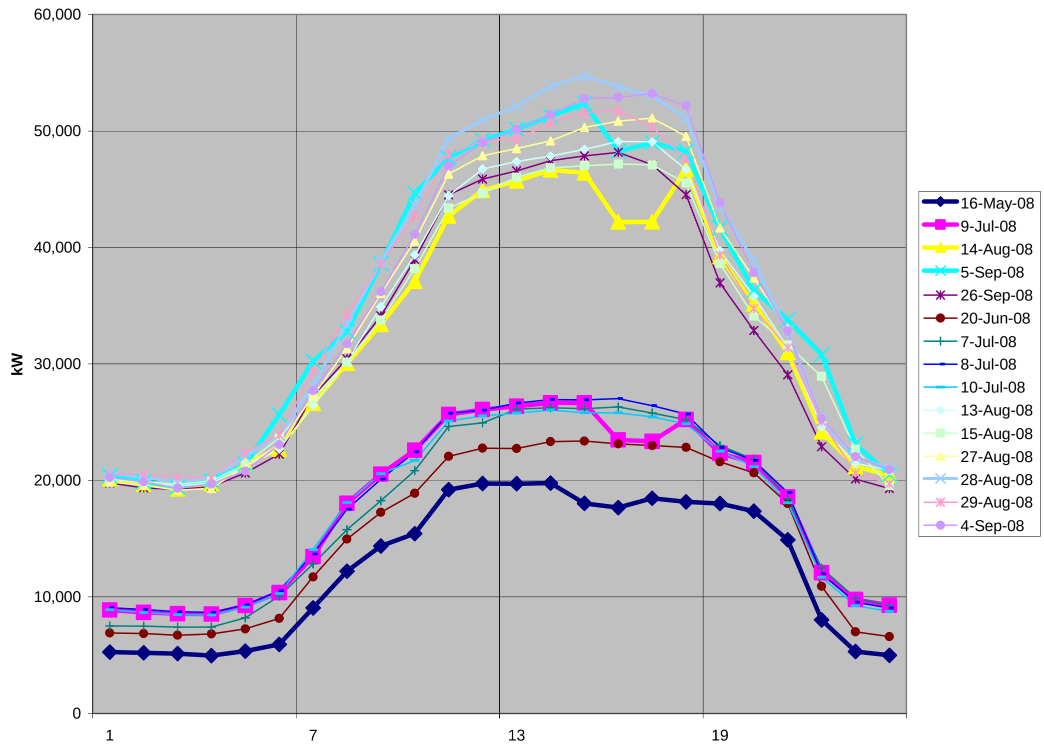



**Figure 18. Aggregator 2; Industrial; *Adjusted BL Option***



## Figure 19. Aggregator 2; Commercial; *Adjusted BL Option*




## Figure 20. Aggregator 3; Industrial; *Adjusted BL Option*

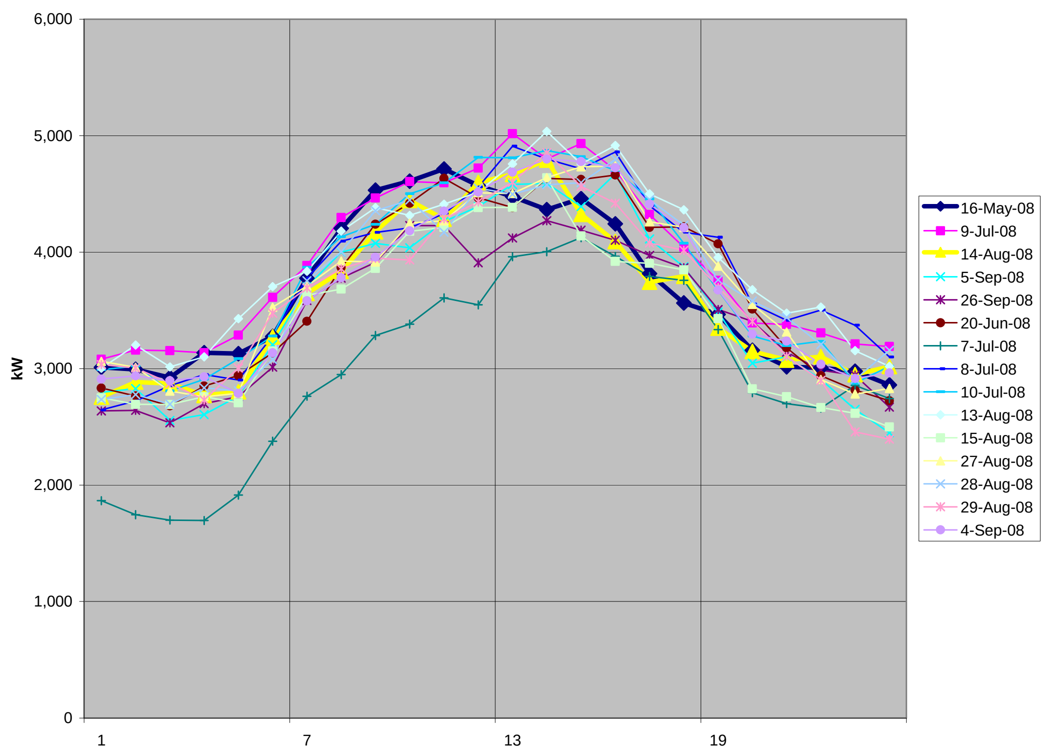


**Figure 21. Aggregator 4; Industrial; *No Adjusted BL***

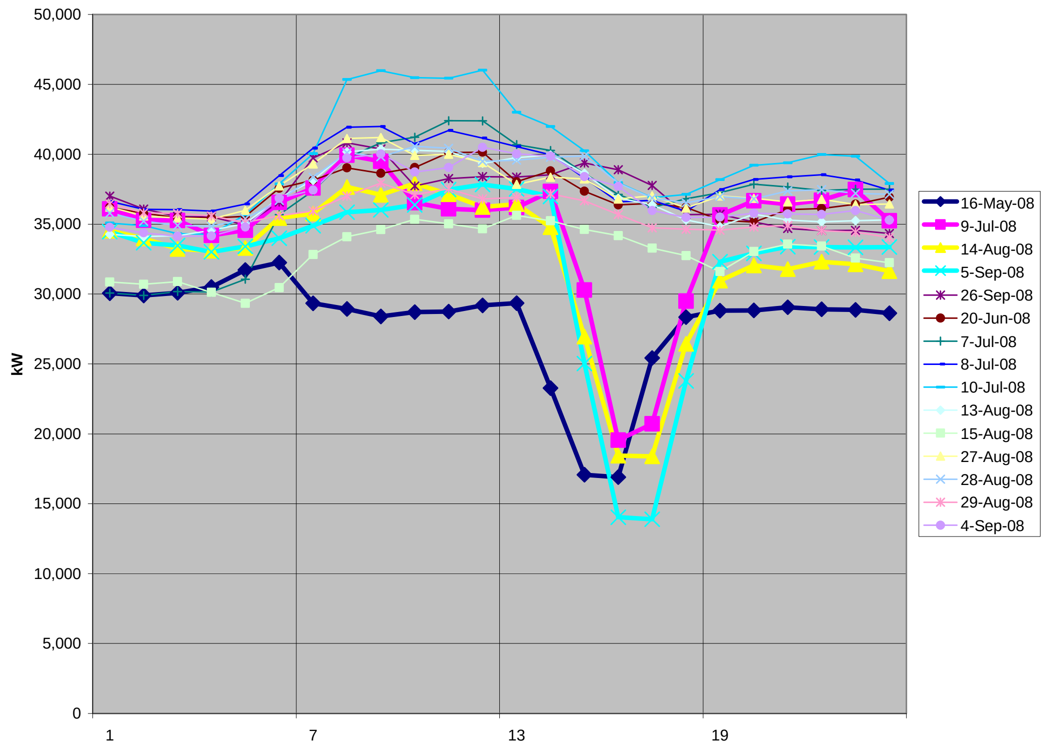



**Figure 22. Aggregator 4; Industrial; *Adjusted BL Option***

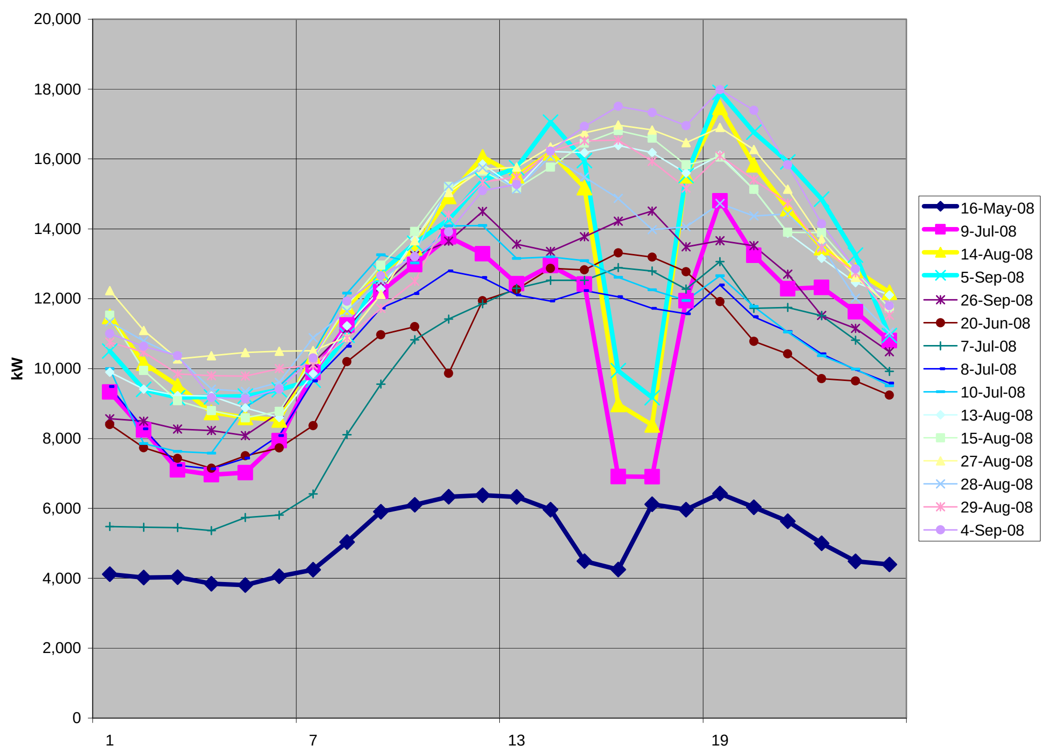


**Figure 23. Aggregator 4; Commercial; *Adjusted BL Option***

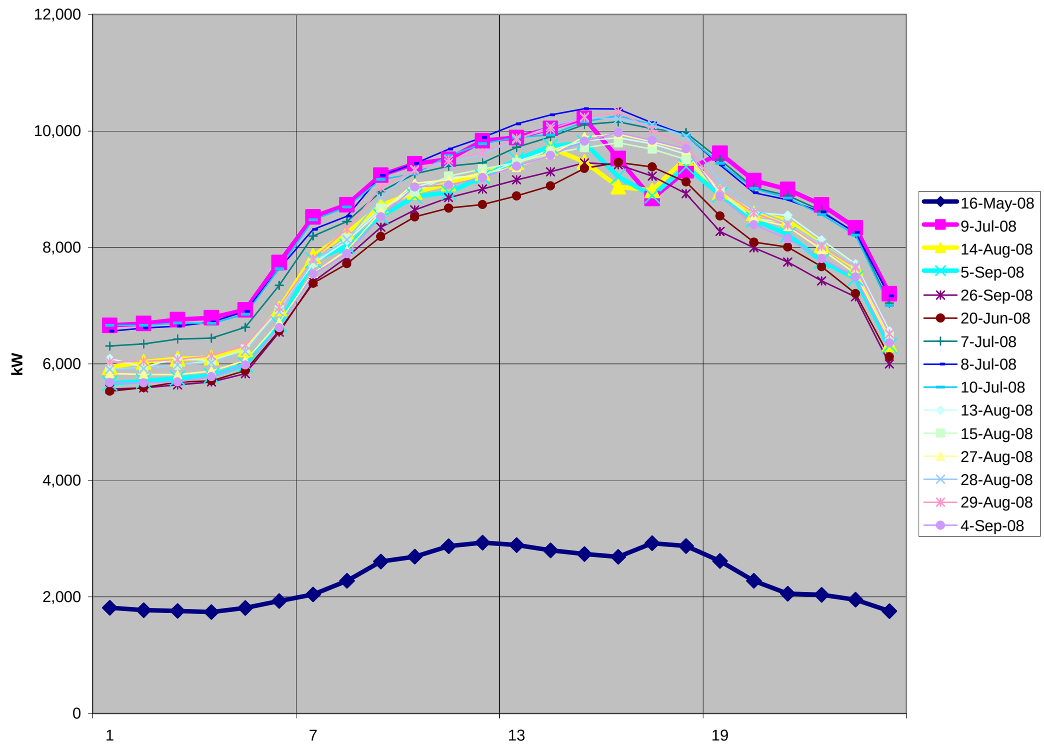
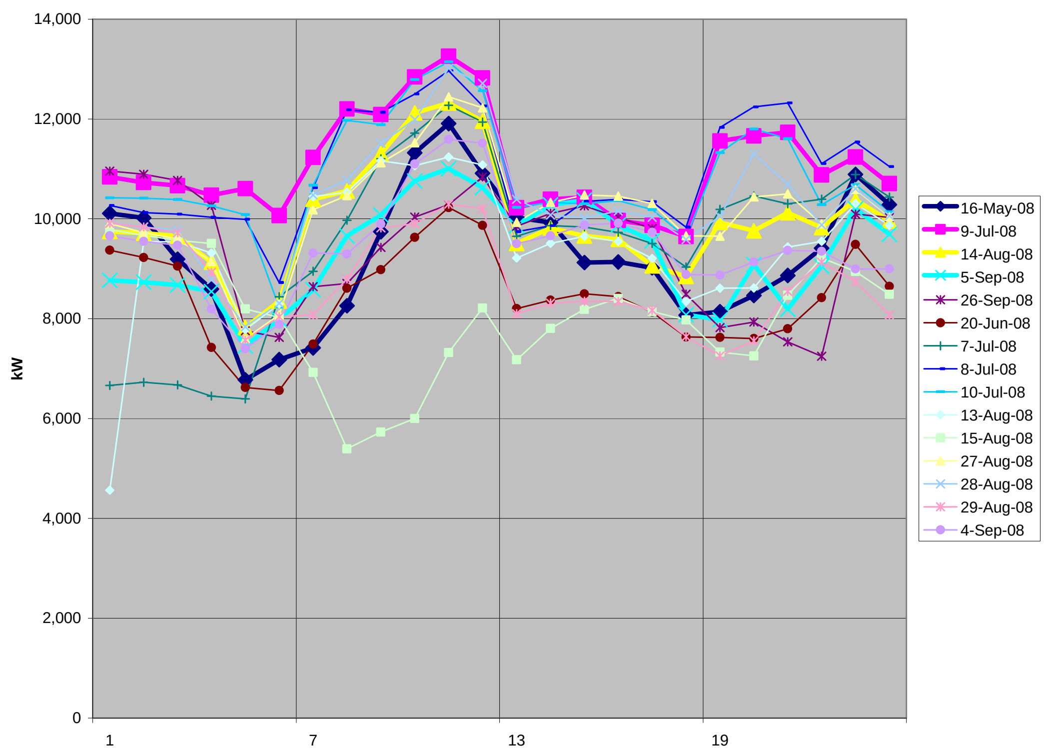



**Figure 24. Aggregator 4; Schools; *Adjusted BL Option***



### *6.2 Analysis of pre-event usage*

This section presents results of an analysis of AMP customers' typical pre-event usage levels on event days compared to that usage level on event-type days that were not called as events. Table 11 summarizes the number of customers, and the averages, standard deviations and coefficients of variation of ratios of their pre-event usage levels across customers in the three customer types, and by their choice of the adjusted baseline option. As seen in the second set of columns, the average values of the event-day to event-type day pre-event usage ratio are near 1.0, and differ only negligibly between those that accepted the adjusted baseline option and those that did not. Variability of the ratio is greater for the industrial customers than for the commercial customers.

**Table 11. Ratios of Average Morning Usage – Event and Non-Event Days (*By Choice of Adjusted Baseline*)**

| Customer type | Count | | Ave. AM kWh - Event/ Non-event | | Standard Deviation | | Coeff. of Variation | |
|---|---|---|---|---|---|---|---|---|
| | No | Adj. BL | No | Adj. BL | No | Adj. BL | No | Adj. BL |
| 1. Ind | 193 | 56 | 0.98 | 0.98 | 0.39 | 0.38 | 0.39 | 0.39 |
| 2. Comm'l | 94 | 109 | 0.99 | 0.99 | 0.05 | 0.18 | 0.05 | 0.18 |
| 3. Schools | 9 | 6 | 1.01 | 1.00 | 0.18 | 0.11 | 0.18 | 0.11 |
| Grand Total | 296 | 171 | 0.99 | 0.98 | 0.31 | 0.26 | 0.32 | 0.26 |

Table 12 provides additional detail on the distributions of the morning-usage ratios, showing quartile values. The median values, like the averages in Table 11, are close to one, for both those who selected adjusted baselines and those who didn't. Differences in the distributions of values are also modest, with the spread somewhat wider for the industrial groups than for the commercial groups. None of these values suggest concern about systematic gaming efforts.

**Table 12. Quartiles of Ratios of Average Morning Usage**

| Quartiles | Not Adjusted | | Adjusted BL | |
|---|---|---|---|---|
| | Ind | Comm'l | Ind | Comm'l |
| Min | 0.61 | 0.85 | 0.55 | 0.59 |
| First quartile | 0.88 | 0.91 | 0.97 | 0.94 |
| Median | 0.99 | 0.94 | 1.01 | 0.99 |
| Third quartile | 1.08 | 0.96 | 1.10 | 1.02 |
| Max | 3.17 | 0.97 | 2.09 | 2.61 |

### *6.3 Conclusions about gaming*

The analysis of sub-group level aggregated load data and individual customer pre-event usage on event days and event-type days finds little if any evidence of artificial increases in pre-event usage in an attempt to "game" the adjusted baseline. The sub-group load profiles show little difference between groups that faced adjusted baselines and those that did not, and the load profiles for adjusted-baseline groups show little difference between event days and event-type non-event days. Only one case was found, for one industrial customer of one aggregator, in which the load rose unusually in the four hours prior to one event, possibly indicating an attempt to increase the baseline from which the load impact would be measured.

Analysis of the distribution of ratios of pre-event usage on event and event-type days confirmed the findings from the aggregated load data, revealing no evidence of systematic increases in pre-event consumption on event days.

## 7. Conclusions

### *7.1 Baseline performance – event-type days*

The results of this baseline analysis using event-type days provide a reasonably consistent story regarding the baseline issues of the relative accuracy of aggregator and sum-of-customer baselines, and the effect of morning adjustments to 3-, 5-, and 10-in-10 baselines on the bias of unadjusted baselines. Some results are mixed, suggesting that baseline performance depends on the characteristics of customers and event days. Major findings include the following:

1. Regarding the accuracy of the *aggregator* method of calculating baselines compared to the *sum-of-customer* method, the results suggest that the aggregator method is more accurate, but not by a wide margin.
2. Regarding the effect of *morning adjustments* to the 3-in-10 baseline on *bias*, the results suggest that the adjustments do improve the bias of the unadjusted baseline relative to the "true" baseline.
3. Expanding the analysis to consider adjusted 5-in-10 and 10-in-10 baselines produced results suggesting that the *adjusted 10-in-10 method* may produce both the greatest accuracy and the smallest bias, although this result did not hold in the case of event days.
4. Examination of the performance of *upward-only* adjustments to the 5-in-10 and 10-in-10 baseline methods suggests that they reduce baseline performance, but not dramatically.
5. Comparing unadjusted 5-in-10 and 10-in-10 baselines to comparable symmetric adjusted versions illustrates the improved performance of the adjusted versions, which should be taken into account in any decision to allow a choice among those options.
6. Examination of the variability of percent errors of 10-in-10 baselines for *individual customers* illustrates the likely source of greater baseline errors in sum-of-customer baselines compared to aggregator baselines.

### *7.2 Baseline performance – event days*

The performance of the alternative baseline methods on event days, in terms of accuracy and bias, was qualitatively similar to their performance on the *event-type* days presented in Section 4. In particular, adjusting the baseline for morning usage generally improves the accuracy and reduces the bias of the unadjusted baselines. Performance results varied considerably across aggregators and industry types. The adjusted 10-in-10 did not dominate the other methods as it appeared to do for the event-type days. However, it performed at least as well and often better than the other adjusted baselines.

### *7.3 Was gaming successfully avoided?*

Analysis in this study revealed no evidence of systematic increases in pre-event consumption on event days that would be indicative of attempts to game the adjusted baseline. Only one case was found, for one industrial customer of one aggregator, in which hourly usage rose unusually in the four hours prior to one event, possibly indicating an attempt to increase the baseline from which the load impact would be measured.

The evidence in this baseline analysis suggests that adjusted baselines are more accurate and less biased than unadjusted baselines. However, widespread adoption of adjusted baselines would seem to call for monitoring, possibly during the event season, to check for unusual load changes that could indicate gaming behavior. Creation and examination of aggregator load profiles like those examined in Section 6 could serve as an example.

## Appendix: Technical Background

It may be useful to point out several related features of any baseline analysis that involves comparisons of alternative baseline methods. These include the relationships between baseline definitions, baseline errors, and implied differences in estimated load impacts. The present baseline analysis differs from previous analyses due to the additional objective of measuring baselines for *aggregated* groups of customers.

### *Baseline definitions*

Consider the following definitions:

Individual baseline: $IBL^i_d = f(E^i_{d-t})$,

Aggregator baseline: $ABL_d = f(\sum E^i_{d-t})$,

Sum of customer baselines: $SBL_d = \sum IBL^i_d = \sum f(E^i_{d-t})$.

For simplicity, assume that the baselines are calculated as the average across hours in an event. Thus, the value *E* represent average hourly load during the event period, the superscript, *i,* refers to an individual customer, *d* refers to the event day, and the function *f* refers to a rule for calculating the baseline across previous days, (*d–t*) (*e.g.*, average of highest 3 days in previous 10 eligible days). The *aggregator* baseline applies the baseline definition to the aggregated load of customers in the group, while the *sum of customer* baseline adds up the calculated baselines of each individual customer in the group.

### *Baseline errors*

Baseline analyses typically calculate and compare different measures of baseline errors, defined as the difference between the *true* baseline (TBL) and the estimated baseline, as defined above. For example, baseline errors for an individual customer and an aggregated group of customers may be written as:

$ERRI^i_d = TBL^i_d - IBL^i_d$ , and

$ERRA_d = \sum TBL^i_d - ABL_d$.

When dealing with event-type days on which events were not actually called, the true baseline equals actual consumption during the "event" period. Given the interest in comparing the performance of the aggregator and sum-of-customer baselines, we can define the difference in errors for those two baselines as:

$$
\begin{aligned}
DiffERR &= ERRA_d - \sum ERRI^i_d \,. \\
&= \sum TBL^i_d - ABL_d - (\sum TBL^i_d - \sum IBL^i_d) \\
&= - ABL_d + \sum IBL^i_d.
\end{aligned}
$$

That is, differences in the *errors* of the two baselines are equal to the differences between the two baselines (*i.e.*, for purposes of comparing the errors of two alternative baselines, the true baselines drop out of the consideration). In the current baseline analysis, the primary interest is in differences in the accuracy and bias of different baseline methods, both of which statistics are functions of baseline errors across a number of events, and customers or aggregators. However, the performance statistics for each baseline method are of interest in themselves, so that we calculate the baseline errors relative to the true baseline and then compare results.

### *Differences in load impacts and baseline errors*

Load impacts (*i.e.*, differences between the baseline and actual load) corresponding to the alternative baseline methods may be written as follows:

Individual load impact (ILI): $IBL^i_d - E^i_d$,

Aggregate load impact (ALI): $ABL_d - \sum E^i_d$,

Sum of customer load impact (SLI): $SBL_d - \sum E^i_d$.

The difference between the aggregator load impact and the sum of customer load impacts may be written as:

$$\begin{aligned} \text{DiffLI} &= \text{ALI} - \text{SLI} \\ &= (ABL_d - \sum E^i_d) - (SBL_d - \sum E^i_d) \\ &= ABL_d - SBL_d \\ &= ABL_d - \sum IBL^i_d \\ &= -\text{DiffERR}. \end{aligned}$$

That is, the difference between the estimated *load impacts* relative to two alternative baselines is the same as the negative of the difference between the *baseline errors*. This result points to the importance of baseline performance in calculating accurate load impacts.