FREEMAN, SULLIVAN & CO.

A MEMBER OF THE FSC GROUP

Assessment of Settlement Baseline Methods for Ontario Power Authority's Commercial & Industrial Event Based Demand Response Programs

September 28, 2010

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

Prepared for:

Nik Schruder
Ontario Power Authority
120 Adelaide Street West, Ste. 1600
Toronto, Ontario

Prepared by:

Josh Bode, M.P.P.
Michael Perry, Ph.D.
Seth Morgan, M.A.
Freeman, Sullivan & Co.

Table of Contents

1 EXECUTIVE SUMMARY

This report evaluates the accuracy of current and alternate baselines used for settlement of Ontario Power Authority's (OPA) large commercial and industrial Demand Response Program (DR). In total, 48 baseline methods were tested using a year of unperturbed pre-enrolment data from 95 customers. The evaluations presented here were designed to identify a baseline methodology that:

- Is accurate for both small and large customers;
- Is fair across settlement accounts and customers;
- Avoids extreme errors that could negatively affect individual settlement payments; and
- Is accurate not only for the most common event window but across all event windows.

In other words, the optimal baseline must provide OPA and program participants with small load-impact errors and therefore small settlement payment errors on average (accuracy), must have a narrow distribution of errors across accounts (fairness), and must lack extreme errors at the settlement account or individual customer level.

We separately analyzed accuracy by customer size, by groupings into settlement accounts, and by individual customer to ensure the findings are fair and applicable for both DR-1 and DR-3 programs that settle at an aggregate or individual level.

Identifying the most accurate baseline is crucial as baseline error is unfair to both the customer and the OPA. In comparing accuracy, we calculate baseline, load-impact and settlement payment errors. Baseline errors are the difference between what the calculated baseline usage is for a customer and what that customer's usage actually would have been in the absence of an event. Normally, what the customer's usage actually would have been is not observable, but in our analysis of pre-enrolment data it is. This means we can directly calculate baseline errors. Load-impact errors are the difference between the demand reduction the customer gets credit for and the demand reduction the customer actually provided. Again, normally this is not observable because only the customer's actual load in the presence of the event is measured. Using pre-enrolment data allows us to quantify these errors as well. Baseline errors are magnified into proportionally larger load-impact errors and thus can generate large settlement-payment losses to OPA. Small baseline errors can produce very large load-impact and settlement-payment errors when the share of the load contracted for load reduction is relatively small.

Of 48 baselines initially analyzed, 6 produced average load impact errors within +/-2%, reducing the current baseline error by a factor of 12. These errors were calculated for an event window of 3 P.M. to 7 P.M., and were also averaged separately for customers above one MW of contracted load reduction and below one MW of contracted load reduction. These 6 baselines included the Top 7, 9 and 10 of 10 Hourly baselines each with a 4-hour and 6-hour same-day adjustment. All were compared to OPA's current method of Top 15 of 20 Hourly (with and without same-day adjustments) to highlight the improvements that can be realized with these alternate baseline methods.

For the group of methods within the +/-2% average error threshold, it became more important to narrow the list of potential baselines based on the distribution of accuracy across accounts and their propensity of producing extreme errors. Baselines 10 of 10[1] and Top 9 of 10 Hourly each with a 6-hour adjustment exhibited the narrowest normalized error distributions and relatively few extreme values across settlement accounts. Both also perform well across different event window periods, though the 10 of 10 is the most robust over time. Furthermore, the 10 of 10 baseline with a 6-hour adjustment slightly outperforms the top 9 of 10 method with 6-hour adjustment by providing availability and utilization payment errors closer to zero during the typical event hour for the average customer and average settlement account.

We recommend a 10 of 10 baseline with a 6-hour adjustment for OPA's DR-3 and DR-1 programs. This method averages a very low overall load-impact error (-0.5%) during the most common event period lasting from 3 P.M. to 7 P.M. It is accurate for customers both above and below one MW of contracted load reduction (load impact errors of -0.7% and 0.5%, respectively). This means that it will be appropriate for the range of customers in DR-3, as well as the smaller customers in DR-1. Among the other relatively accurate baselines, it produces the narrowest distribution of errors and generates few extreme error values whether error distributions are calculated at the customer level (as would apply to DR-1) or at the settlement account level (as would apply to DR-3). Finally, this baseline remains on average the most accurate baseline across event windows starting as early as 12 P.M. and as late as 5 P.M.

An important caveat to this conclusion is that customers may react to the addition of a same-day adjustment in ways that could lead the adjustment to overestimate the baseline. This is a bigger risk for very large customers for whom electricity is a substantial expense and who may have managers tasked with managing power costs. Also, there is a small amount of evidence that some large customers shift load to the hours before an event. If this practice occurred regularly after the adoption of a same-day adjustment, then it would also cause the adjustment to overestimate the baseline. Whether either of these issues will actually arise is unknown at this time. We recommend that if a same-day adjustment is adopted, that the method be reassessed after several events have passed to determine whether there is evidence that customers have reacted to the adjustment in ways that lead to baseline inaccuracy. We recommend particular

[1] The 10 of 10 Hourly and 10 of 10 Daily baselines are always equivalent as both methods use all hours of all days. Therefore, no distinction will be made in terms of results and a single 10 of 10 baseline will be shown.

attention be given to the largest customers. We also recommend considering custom baseline methods for the largest customers.

2 INTRODUCTION

This report evaluates the accuracy of current and alternate baselines used for settlement of Ontario Power Authority's (OPA) large commercial and industrial Demand Response Program (DR). In total, 48 baseline methods were tested using a year of unperturbed pre-enrolment data from 95 individual customers. The baseline assessment was designed to identify a baseline methodology that:

- Is accurate for both small and large customers;
- Is fair across settlement accounts and customers;
- Avoids extreme errors that could negatively affect individual settlement payments; and
- Is accurate not only for the most common event window but across all event windows.

Identifying the most accurate baseline is crucial as baseline error is unfair to both the customer and the OPA. In comparing accuracy, we calculate baseline, load-impact and settlement payment errors. Baseline errors are the difference between what the calculated baseline usage is for a customer and what that customer's usage actually would have been in the absence of an event. For customers in the DR-3 or DR-1 program, on an event day we cannot observe what usage would have been in the absence of an event because an event actually occurred. However, when we use data for these customers from before the time they enrolled in DR-3 or DR-1, then we can observe actual usage on days that are very similar to event days—we discuss how we choose these days below. This means we can directly calculate baseline errors by comparing actual loads on event-like days to calculated baselines for those days. Load-impact errors are the difference between the demand reduction the customer gets credit for and the demand reduction the customer actually provided. Again, normally this is not observable because only the customer's actual load in the presence of the event is measured. Using pre-enrolment data allows us to quantify these errors as well. Baseline errors are magnified into larger load-impact errors and thus can generate large settlement-payment losses to OPA. Small baseline errors can produce very large load-impact and settlement-payment errors when the share of the load contracted for load reduction is relatively small.

The remainder of this section:

- Provides an overview of publicly available baseline accuracy studies;
- Presents background on DR-3 and the new DR-1, the two event-based OPA DR programs that rely on settlement payments;
- Details the relationship between baseline and load impact error;
- Describes the 48 baseline methods assessed; and
- Details the decision criteria employed to recommend a baseline method for DR-1 and DR-3.

2.1 Prior Baseline Accuracy Studies

The topic of Demand Response Program settlement baseline accuracy and precision has been studied in the past, primarily for demand buy-back programs.

KEMA (2003)[2] compared the accuracy of six settlement baselines in 2003 using 646 accounts from multiple regions across the U.S. In total, 206 of the study sites were participants in a DR program, while 440 were not. Baseline error was assessed for non-participants. The non-participant baseline estimates were compared with actual loads for each hour of simulated curtailment periods.

Quantum Consulting (2004)[3] estimated the accuracy of four settlement baselines using data from 450 accounts in California, none of which were enrolled in DR programs. The study compared baseline predicted loads to actual loads on three types of days – high load days, low load days, and consecutive days. It selected 3 to 7 proxy event days based on system load. The number of event days selected varied by utility.

Lawrence Berkeley National Lab (2008)[4] also compared accuracy of 7 alternate settlement baselines using data from 32 sites in California, all of which were enrolled in an Auto-DR program. The study compared baseline predicted loads to actual loads on 60 days per site. The proxy event days were selected based on weather.

As part of the California Statewide Demand Bidding Program, Christensen Associates Energy Consulting compared the settlement baselines. They compared baseline predictions to an implied baseline – that is, the actual load plus a regression based estimate of load impacts – for 592 participants that submitted curtailment bids.

This study differs from prior studies in several ways:

- 48 baselines were tested—more than any prior study.
- Baseline error was calculated by comparing actual and predicted baseline load for DR participants. With the exception of the LBNL study, all studies have draw conclusions based on either non-participants or compared one estimate to another estimate. Using non-participants can produce selection issues. On the other hand, errors can only be assessed by comparing actual loads to estimates.

---

[2] "*Protocol Development For Demand Response Calculation—Findings and Recommendations*;" Consultant Report for the California Energy Commission, February 2003.

[3] *"Working Group 2 Demand Response Program Evaluation – Program Year 2004 Final Report;"* Quantum Consulting, December 2004.

[4] *"Estimating Demand Response Load Impacts: Evaluation of Baseline Load Models for Non-Residential Buildings in California;"* Demand Response Research Center, Ernest Orlando Lawrence Berkeley National Laboratory, January 2008**.**

- Both the baseline error and load impact error are calculated, assuming full compliance. Baseline errors are an indirect measure of performance. We quantify the effect of the baseline error on the metrics that affect program performance and cost-effectiveness. None of the prior studies quantified the extent to which the baseline error biased load impacts and settlement payments.[5]

- Baselines robustness to varying event windows is tested.

- We quantify the distribution of errors both across individual accounts and across aggregated settlement accounts. All prior studies have focused on the distribution across individual accounts.

This study has limitations as well. It is Ontario-specific and based on the current mix of customers in the OPA C&I programs. The most accurate baseline for the OPA programs may not be the most accurate baseline elsewhere. It is also possible, though unlikely, that new participants in the OPA DR program could differ substantially from the current mix of participants.

2.2 Background on DR-1 and DR-3

DR-3 allows participants and aggregators to enter into contractual agreements for load reductions with OPA. Participants can choose to enrol directly with OPA, provided they meet minimum load-reduction criteria, or can participate through an aggregator. Under DR-3, an aggregator or direct participant must commit to a specific load-reduction amount for either 100 or 200 hours per year. The OPA has discretion regarding the timing of events and currently determines event days based on IESO day-of supply-cushion estimates. In exchange for load reductions, the DR-3 Program makes both availability (capacity) and energy payments. The OPA can reduce payments if participants fail to provide the contractual load reduction or are unavailable to provide load reductions (the equivalent of a scheduled outage). Settlement for the DR-3 program takes place at the level of settlement accounts, which can be made up of multiple customers. Both demand-reduction assessment and payment by OPA takes place at the level of the settlement account. For settlement purposes, compliance with the contracted load reductions is determined through day-matching baseline methods.

DR-1 is a voluntary demand response (DR) program offered by the OPA to pilot the concept of demand response (DR) among industrial and commercial customers. Under DR-1, OPA makes payments for actual load reduction only. There are no payments or set-offs associated with a customer deciding not to participate, or where a customer has indicated willingness to perform and then not followed through. DR-1 is explicitly intended as a feeder program for DR-3.

[5] The LBNL study compared load impact estimates from alternate baselines to the most accurate baseline. It did not, however, assess accuracy of load impact estimates.

2.3 Relationships Between Baseline, Load Impact and Settlement Error

In evaluating OPA's baseline calculations, it is important not only to assess the accuracy of the baseline (or reference load), but the effect on estimated load impacts and settlement payments. Although accuracy of reference loads, load impacts and settlement payments are closely related, they are not one and the same. For baseline methods, over- and under-predictions of the reference load can lead to magnified errors in the estimated load impacts used for settlement payments. Ultimately, the effect of baseline error on over- or under-payments depends on the reference-load bias, percent load reductions and settlement rules. The effect of baseline error on the accuracy of load impacts and settlement calculations is more important than the baseline accuracy per se. With smaller percent load reductions, small baseline errors translate into far larger errors in the estimated load reductions. For example, suppose a customer has agreed to provide one MW of load reduction during an event. Suppose that during a certain event the baseline overestimates the actual load by 0.5 MW (the baseline error). This results in a load-impact percentage error of 50% (baseline error divided by contracted MW). However, if the contracted reduction load was less, say 0.75 MW, the load-impact error would increase to 66% (0.5 MW divided by 0.75 MW).

Table 2-1 summarizes the relationship between the percent load reduction error as a function of the baseline error and the true percent load reduction.

Table 2-1
Impact of Reference Load Baseline Error on Load Impact Error as a Function of Actual Load Reductions

		Actual Load Reduction									
		5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
Reference Load Baseline Error	-10.0%	-200.0%	-100.0%	-66.7%	-50.0%	-40.0%	-33.3%	-28.6%	-25.0%	-22.2%	-20.0%
	-8.0%	-160.0%	-80.0%	-53.3%	-40.0%	-32.0%	-26.7%	-22.9%	-20.0%	-17.8%	-16.0%
	-6.0%	-120.0%	-60.0%	-40.0%	-30.0%	-24.0%	-20.0%	-17.1%	-15.0%	-13.3%	-12.0%
	-4.0%	-80.0%	-40.0%	-26.7%	-20.0%	-16.0%	-13.3%	-11.4%	-10.0%	-8.9%	-8.0%
	-2.0%	-40.0%	-20.0%	-13.3%	-10.0%	-8.0%	-6.7%	-5.7%	-5.0%	-4.4%	-4.0%
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	2.0%	40.0%	20.0%	13.3%	10.0%	8.0%	6.7%	5.7%	5.0%	4.4%	4.0%
	4.0%	80.0%	40.0%	26.7%	20.0%	16.0%	13.3%	11.4%	10.0%	8.9%	8.0%
	8.0%	160.0%	80.0%	53.3%	40.0%	32.0%	26.7%	22.9%	20.0%	17.8%	16.0%
	10.0%	200.0%	100.0%	66.7%	50.0%	40.0%	33.3%	28.6%	25.0%	22.2%	20.0%

2.4 Baselines Tested

In total, 48 baseline methods were tested for each customer and settlement account. For each baseline method, the baseline, estimated load impact and settlement payment error were calculated. The settlement baseline methods varied across three dimensions: days used to calculate the baseline, same day adjustments, and "hourly" versus "daily" baseline calculations. The following sets of days were chosen for calculating a potentially optimal baseline:

- Top 3, 7 and 9 out of the last 10 non-event days;
- Bottom 3 and 7 out of the last 10 non-event days;
- All 10 of the last 10 non-event days; and
- Top and Bottom 15 out of the last 20 non-event days.

Each of the above baselines was also calculated using two types of same-day adjustment. These same-day (or in-day) adjustments were applied to the baseline day-selection methods. Both four- and six-hour adjustments were tested. All adjustments included a two-hour buffer between the event period and the period used to calculate the adjustment. Details on the calculations of the same-day adjustments are presented in the Methodology section below and in Appendix G.

Finally, each baseline and its adjustments were calculated using an "hourly" and "daily" approach with respect to selection of the top or bottom days. In the first, each hour of each day included in the baseline calculation receives its own ranking. This means that the days included in the calculation of the baseline for the hour ending 4 P.M. might be different than the days included in the calculation of the baseline for the hour ending 5 P.M. In the second method, each potential baseline day is ranked according to the average usage during the same hours as the event. In this method, the same set of days is used for all hours of the baseline calculation. We refer to this as the “daily” method. The OPA uses the hourly method for the Top 15 of 20 non-event days.

OPA’s current approach in which the baseline is determined by choosing the top “x” days out of the last “y” days is very common in the industry for similar programs. There are numerous examples across North America of DR programs that employ baseline calculations without same day adjustments, similar to the current DR-1 and DR-3 method. The prior baseline studies have found baseline errors in the range of 6-8% averaged across many customers, with much higher values for particular customers. The 6.7% baseline error in the current DR-3 baseline is not uncommon.[6] A common next step is to introduce a same-day adjustment, which can markedly increase accuracy, as discussed below.

2.5 Decision Criteria

Three main criteria were used to determine the optimal baseline for OPA's DR-3 and DR-1 programs: average load-impact percentage error, the width of the distribution of load-impact errors and the minimization of extreme errors at the customer and settlement account levels.

[6] We use baseline error to compare OPA baselines to other jurisdiction because those values are published. The more appropriate metric, load impact error, has not been published in other studies.

First, any baseline with an average load-impact error greater than 2% was disqualified. This was done if the average error was larger than 2% either for the whole set of customers or for either of the two subpopulations of customers with contracted demand reduction above and below one MW. Of the baseline methods remaining, those with narrower distributions of errors based on the 90% range criterion were selected for further consideration. This criterion was considered both at the settlement account level and at the individual customer level because DR-3 settlement takes place at the settlement account level, while DR-1 settlement takes place at the customer level. Finally, the optimal baseline was selected based upon the minimization of squared errors to eliminate the possibility of certain customers with large errors dramatically increasing settlement payment losses despite a small program average error. This criterion was also calculated at the settlement account level and the customer level. Baseline accuracy was also assessed during multiple event windows to evaluate robustness across time. The *Analysis of Baseline* section will expand on these decision criteria.

3 METHODOLOGY

The calling of an event reduces the actual load to a level lower than what would be observed in the absence of the event. Because of this, it is not possible to evaluate baseline methods on true event days as the counterfactual reference load is not known. However, baseline accuracy can be assessed by comparing the known estimates to the known actual load on event-like, or proxy event days. This analysis therefore relies on unperturbed pre-enrolment data from which naturally occurring participant behaviour under event-like conditions can be directly observed.

3.1 Baseline Calculation

3.1.1 Selection of Proxy Events

The OPA selects actual event days based on IESO day-of supply cushion estimates. A low supply cushion is typically correlated with high system load. Proxy event days were therefore chosen by selecting those days above the 90th percentile in terms of system load for each customer. Ten of those days were then selected randomly and separately for each customer as proxy event days.[7] We initially assume an event window of 3 P.M. to 7 P.M. on those selected days as 12 of the 34 events called in 2008 and 2009 were during that time. This was more than any other event window. For each event, we simulate full compliance with the calculated baseline based on the contracted MW.

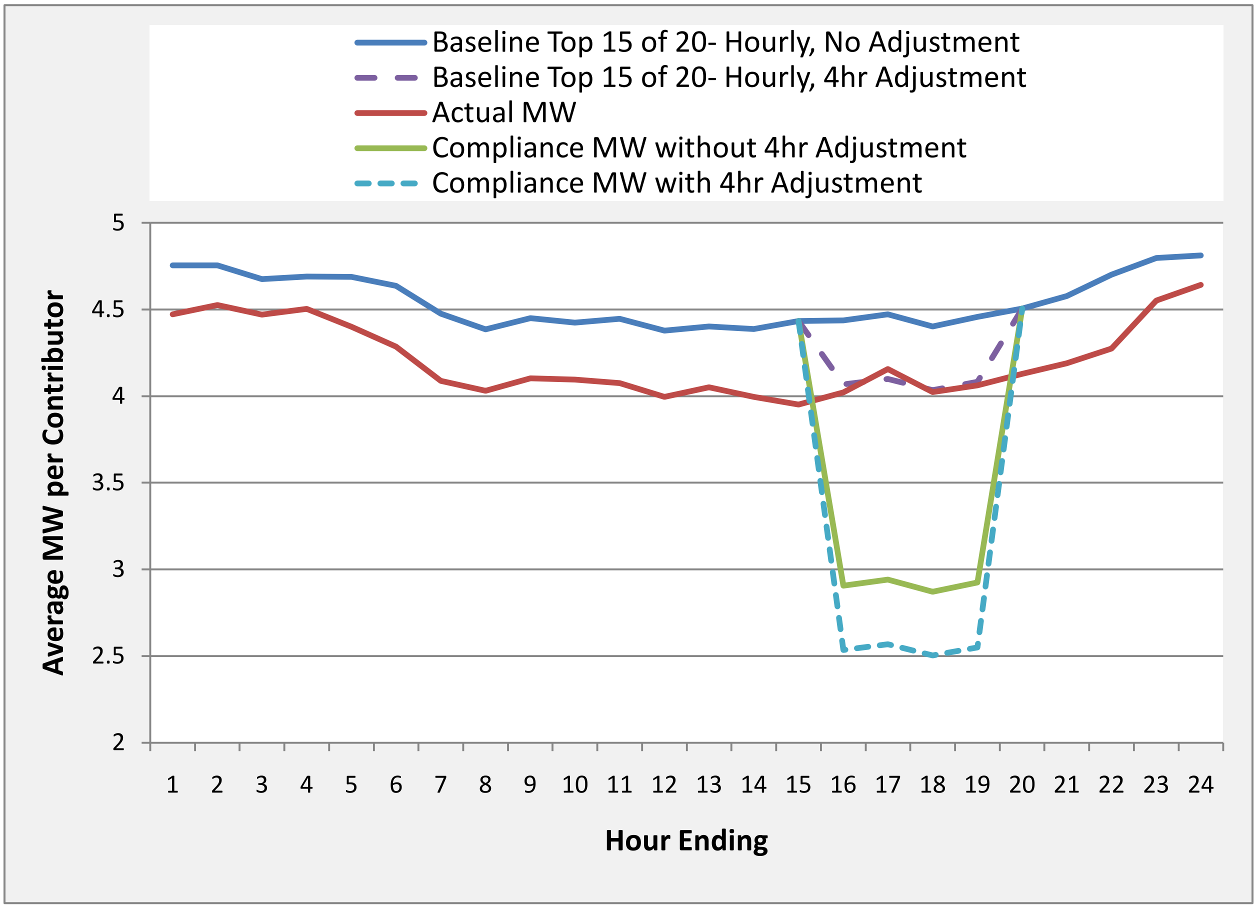
3.1.2 Calculation of Same-day Adjustments

The same-day adjustment calibrates the baseline to pre-event load levels. As previously mentioned, the same-day adjustments used in this analysis include a four- and six-hour adjustment with a two-hour buffer. To calculate these adjustments, the event-period baseline is multiplied by the ratio of the averages of actual and baseline loads during the four or six hours preceding a two-hour buffer immediately prior to the event window. This calculation is done separately for each customer for each event in the case of DR-1 and for each settlement account for each event in the case of DR-3. For example, if a customer had average usage of 400 kW during the 4 hours before the 2-hour buffer immediately preceding the event and an average baseline of 300 kW during that same time, then the same-day adjustment would consist of multiplying the baseline during the event by 4/3. The idea is that because the baseline is biased down before the event, it is likely to be biased down during the event by a similar proportion. Alternatively, if the baseline shows an upward bias, the adjustment reduces the baseline load so it better reflects load patterns before the event. Figure 3-1 shows the example of an unadjusted and adjusted Top 15 of 20 Hourly baseline with compliance loads for the average customer. In the figure the adjustment only applies to the hours of the event. Compliance loads are shown both with and without the adjustment. The important aspect to notice is that without the adjustment, the customer needs to provide a noticeably smaller

[7] Using the 80th percentile as the cut-off for selection had no major effect on the calculated results or the ranking of baseline methods.

amount of load reduction than with the adjustment. See Appendix G for more detail on the calculation of the same-day adjustment and the rationale behind the two-hour buffer.

Figure 3-1:
Baseline Top 15 of 20- Hourly With and Without a 4hr Same-day Adjustment



3.1.3 Hourly Versus Daily Baselines

For each baseline method, we considered two possible ways to calculate the baseline – an "hourly" and a "daily" approach. In the first, each hour of each day included in the baseline calculation receives its own ranking. This means that the days included in the calculation of the baseline for the hour ending 4 P.M. might be different than the days included in the calculation of the baseline for the hour ending 5 P.M. This is currently how OPA calculates baselines using its Top 15 of 20 method. We refer to this as the "hourly" method.

In the second method, each potential baseline day is ranked according to the average usage during the same hours as the event. In this method, the same set of days is used for all hours of the baseline calculation. We refer to this as the "daily" method. It was our hypothesis that the daily method could lead to smaller upward biases because some customers' usage profiles have periodic peaks that occur at different hours on different days. The hourly method would, on average, pick up more of these peaks than typically occur in any given day.

4 ANALYSIS OF BASELINES

4.1 Average Load Impact Errors

In order for a baseline method to be viable it had to be accurate on average across all customers and for customers both above and below one MW of contracted load reduction. Baselines that were not accurate for both of these customer groups were not given further consideration. Out of the 48 baselines assessed, 12 baselines produced impacts within a+/- 5% margin of error and 6 baselines produced estimates within a +/- 2% margin of error. The pool was narrowed down to the six baselines that provided +/- 2% error for customers both above and below one MW of contracted load reduction.

Table 4-1 lists average load-impact errors for the six most accurate baselines and compares them to the current Top 15 of 20 Hourly (adjusted and unadjusted) baseline.[8] The average errors presented in the table are averages of individual customer error weighted by average usage. Note that all the best performing baseline methods include a same-day adjustment, which is a more recent development among best practices in the industry and which significantly reduces average errors. Among the 16 methods tested without a same-day adjustment, OPA's current method performs better than nine. Among the full set of 48 methods tested, OPA's current method performs better than 10 of them. This means that the current method is better than one of the methods that includes the same-day adjustment—the bottom 3 of 10 method. Appendix A shows average errors for the full set of 48 methods tested. The strongest conclusion from Appendix A is that methods with a same-day adjustment are much more accurate than methods without. This can also be seen in Table 4-1 by comparing OPA's current method to OPA's current method plus a four- or six-hour same-day adjustment. The average error decreases dramatically with the adjustment.

The current baseline overestimates load impacts by roughly 25%. The error is larger for customers over one MW (27.4%) than for customers below one MW (8.5%). The same-day adjustments reduce the average load-impact percentage error for the current Top 15 of 20 Hourly to less than 1%. However, the load-impact percentage error for customers below one MW of contracted load reduction exceeds the threshold of +/- 2% error.

[8] Appendix A contains average load-impact error tables for all forty-eight baselines.

Table 4-1:
Baseline Load Impact Error by Customer Contracted MW Size

Eligible Baselines	Program	>1MW/hour	<1MW/hour
Top7 of 10 Hourly- 4hr Adjustment	-0.9%	-1.4%	1.8%
Top7 of 10 Hourly- 6hr Adjustment	-0.1%	-0.4%	1.8%
10 of 10- 4hr Adjustment	-0.5%	-0.7%	0.5%
10 of 10- 6hr Adjustment	-0.5%	-0.7%	0.5%
Top 9 of 10 Hourly- 4hr Adjustment	-1.2%	-1.5%	0.6%
Top 9 of 10 Hourly- 6hr Adjustment	-1.0%	-1.2%	0.6%
Top 15 of 20 Hourly- No Adjustment	24.6%	27.4%	8.5%
Top 15 of 20 Hourly- 4hr Adjustment	0.3%	0.0%	2.3%
Top 15 of 20 Hourly- 6hr Adjustment	0.6%	0.3%	2.3%

4.2 The Distribution of Load Impact Errors

To compare the effect of baseline methods on the distribution of load impacts, and thus asses the fairness of each of the six selected baselines across settlement accounts and across individual customers, focus was given to what the normalized error rather than the load-impact percentage error arising from baseline. The normalized error is calculated by subtracting the actual load from the baseline and dividing by the actual load. It provides even grounds for comparison by weighting settlement accounts and customers equally regardless of their overall size or amount of contracted load reduction. There are two main concerns in assessing the distribution of errors: the width of the distribution for the large majority of customers and the size of very large outliers. First, narrower distributions of errors are better. A narrow distribution means that errors for individual customers are almost all very small. Second, baselines that do not produce extreme errors are recommended. A distribution could be quite narrow for the vast majority of cases, but could have a few very large outliers.

To evaluate the width of the distribution of errors for each method, the lower and upper error bounds for the middle 80 and 90% of settlement accounts and customers were calculated. We assessed the distribution at both the settlement account and customers level because the selected baseline will ultimately apply to DR-3 and DR-1, respectively. Table 4-2 summarizes those results.[9]

[9] Appendix F provides histograms of these distributions. Histograms allow the viewing of the entire range of the distribution without excluding any data.

Table 4-2:
Distribution of Normalized Baseline Errors (All Customers Included)

Comparison Level	Baseline	80% of Distribution			90% of Distribution		
		Range	Lower	Upper	Range	Lower	Upper
Individual Customer	Top7 of 10 Hourly- 4hr Adjustment	13.6%	-6.0%	7.6%	17.6%	-7.7%	9.9%
	Top7 of 10 Hourly- 6hr Adjustment	11.6%	-4.7%	6.9%	16.7%	-7.0%	9.7%
	10 of 10- 4hr Adjustment	10.8%	-4.7%	6.1%	17.5%	-7.9%	9.6%
	10 of 10- 6hr Adjustment	10.4%	-4.6%	5.8%	16.1%	-7.5%	8.6%
	Top 9 of 10 Hourly- 4hr Adjustment	11.2%	-4.8%	6.4%	17.6%	-8.2%	9.4%
	Top 9 of 10 Hourly- 6hr Adjustment	9.4%	-4.0%	5.4%	14.2%	-6.6%	7.6%
	Top 15 of 20 Hourly- 4hr Adjustment	13.2%	-6.2%	7.0%	19.0%	-8.1%	10.9 %
	Top 15 of 20 Hourly- 6hr Adjustment	12.3%	-5.3%	7.0%	16.6%	-7.1%	9.5%
Settlement Account	Top7 of 10 Hourly- 4hr Adjustment	20.1%	-11.5%	8.6%	22.2%	-12.3%	9.9%
	Top7 of 10 Hourly- 6hr Adjustment	14.7%	-5.8%	8.9%	20.4%	-11.0%	9.4%
	10 of 10- 4hr Adjustment	17.1%	-9.1%	8.0%	21.9%	-12.3%	9.6%
	10 of 10- 6hr Adjustment	14.1%	-7.5%	6.6%	17.0%	-8.4%	8.6%
	Top 9 of 10 Hourly- 4hr Adjustment	16.6%	-8.5%	8.1%	22.0%	-12.6%	9.4%
	Top 9 of 10 Hourly- 6hr Adjustment	13.3%	-6.6%	6.7%	17.1%	-8.5%	8.6%
	Top 15 of 20 Hourly- 4hr Adjustment	19.7%	-11.8%	7.9%	22.1%	-14.1%	8.0%
	Top 15 of 20 Hourly- 6hr Adjustment	19.0%	-11.6%	7.4%	20.8%	-12.6%	8.2%

At the customer and settlement account level, the Top 9 of 10 Hourly and 10 of 10 baselines with 6-hour same-day adjustments produced the narrowest distribution of errors. For the 10 of 10 baseline with a 6-hour adjustment, 90% of settlement accounts have baseline errors between -8.4% and 8.6%. The results for the 9 of 10 Hourly baseline with a 6-hour adjustment are similar.

4.3 Minimizing Extreme Errors

The last criterion was to determine whether a distribution was prone to produce extreme errors. To make the comparison, the sum of the squared normalized errors was calculated. This gives more weight to large errors as compared to small errors. Weighting larger errors helps weed out the most extreme errors that could lead to individual customers receiving highly inaccurate payments. Thus, the baseline with the smallest sum of squared errors is less prone to extreme errors and, usually, has a smaller distribution of errors. Two baselines have both narrow distributions of errors and are not prone to extreme errors. These are the Top 9 of 10 Hourly and 10 of 10 baselines, both with 6-hour adjustments. Their sum of squared normalized errors equal 0.101 and 0.104, respectively. Table 4-3 shows the comparison of extreme errors of all shortlisted baselines as well as the OPA's current baseline with and without adjustment.

Table 4-3:
Comparison of Extreme Errors Arising from Baselines (All Customers Included)

Eligible Baselines	Sum of Squared Normalized Errors by Customer	Sum of Squared Normalized Errors by Settlement Account
	∑(normalized error)2	∑(normalized error)2
Top7 of 10 Hourly- 4hr Adjustment	0.551	0.138
Top7 of 10 Hourly- 6hr Adjustment	0.559	0.148
10 of 10- 4hr Adjustment	0.331	0.104
10 of 10- 6hr Adjustment	0.334	0.104
Top 9 of 10 Hourly- 4hr Adjustment	0.347	0.099
Top 9 of 10 Hourly- 6hr Adjustment	0.350	0.101
Top 15 of 20 Hourly- 4hr Adjustment	0.712	0.143
Top 15 of 20 Hourly- 6hr Adjustment	0.716	0.148

Both the 10 of 10 and the Top 9 of 10 baselines, each with 6-hour adjustments, do well at minimizing extreme errors.

The analysis of distributions could be expanded even further to consider small customers separately at the individual customer level to produce results even more “specific” to DR-1. This has not been done, and is not likely to produce any novel results. Because the error distributions are calculated for normalized errors, the large customers are not over-weighted in these analyses. For that reason, their inclusion in the analyses of error distributions at the individual customer level is useful because it provides more data on how the baseline methods work across different types of customer behaviour. It is shown that the full error distributions at the individual customer level are quite narrow for the recommended baseline method, and that they do not contain large outliers. That means that the distributions are narrow and do not contain large outliers for the set of small customers separately.

4.4 Baseline Robustness Across Different Event Windows

The results of the baseline calculations presented above were based upon an event window lasting from 3 P.M. to 7 P.M. However, it is important to demonstrate that the recommended 10 of 10 baseline with a 6-hour adjustment performs well across other event windows as well. This will demonstrate that the baseline is robust across time and works well despite weather-sensitivity issues. For example, a same-day adjustment for an event starting at noon would be calculated using early morning hours. A customer whose load is highly weather-sensitive might object that such an adjustment would not be accurate for them. This analysis shows that such concerns do not change the basic fact of which methods produce the lowest average errors and which methods produce the narrowest error distributions across customers and across settlement accounts.

The same 6 baselines and same-day adjustments (plus OPA's current baseline) were evaluated using the event windows presented in Table 4-4. Each adjustment included a two-hour buffer between the adjustment period and the event. Window 4 is the original event window and produces the same outcomes as already shown.

Table 4-4:
Event Windows

Event Window	Event Hours	4hr Adj. Hours	6hr Adj. Hours
Window 1	12pm- 4pm	6am- 10am	4am- 10am
Window 2	1pm- 5pm	7am-11am	5am-11am
Window 3	2pm- 6pm	8am- 12pm	6am- 12pm
Window 4	3pm- 7pm	9am- 1pm	7am- 1pm
Window 5	4pm- 8pm	10am- 2pm	8am- 2pm
Window 6	5pm- 9pm	11am- 3pm	9am- 3pm

The table above shows that a same-day adjustment for an event starting at noon, for example, would be calculated using early morning hours. Table 4-5 displays the average load-impact percentage errors across all event windows for each baseline for the program as a whole and within the subpopulations above and below one contracted MW.

The results show that the 10 of 10 baseline with a 6-hour same-day adjustment is the most accurate baseline across multiple event windows. In fact, its error percentage stays below the +/-2% threshold on average for the program and for both customer size categories. The Top 9 of 10 Hourly baseline with a 6-hour adjustment continues to perform well, though the 10 of 10 method with a 4-hour adjustment is actually the second-best baseline in terms of average accuracy across time. In fact, it is worth noting that for Window 1 and Window 2 (corresponding to events early in the afternoon) this baseline slightly outperforms the same method with a

6-hour adjustment.[10] The difference, though, is small and most events do not start at those times.

In examining the width of the distributions of errors, the same basic result holds, although the 10 of 10 with a 4-hour adjustment and Top 9 of 10 with a 4-hour adjustment also do well. As before, narrow error distributions are preferred as they imply that few customers will be faced with large errors in their baseline. Table 4-5 below shows the average range containing the middle 90% of customers for each baseline method across the six event windows.[11] By this metric, the 10 of 10 with a 6-hour adjustment and the Top 9 of 10 with a 6-hour adjustment have slightly wider distributions than the same methods with four-hour adjustments. However, the normalized baseline error distributions for these methods with 6-hour adjustments have smaller error distributions for the middle event windows, which is about when most events have been called (Appendix D).

Table 4-5:
Average Distribution Range of 90% of Normalized Baseline Errors For All Customers

Baseline Method	Settlement Acct	Individual Customer
Top 7 of 10, Hourly- 4hr Adj.	20.2%	15.8%
Top 7 of 10, Hourly- 6hr Adj.	21.6%	17.7%
10 of 10- 4hr Adj.	19.4%	15.0%
10 of 10- 6hr Adj.	20.8%	15.9%
Top 9 of 10, Hourly- 4hr Adj.	19.3%	14.2%
Top 9 of 10 Hourly- 6hr Adj.	20.5%	15.2%
Top 15 of 20, Hourly- 4hr Adj.	19.9%	16.2%
Top 15 of 20, Hourly- 6hr Adj.	22.5%	17.7%

Finally, the possibility that a few customers could have large errors depending on the baseline and event time is given consideration by examining the sum of squared normalized errors across the six event windows. The event window averages show that the 10 of 10 with a 6-hour adjustment and the Top 9 of 10 with a 6-hour adjustment once again do very well. Once again the same two methods with four-hour adjustments slightly outperform their six-hour adjustment counterparts. This is not surprising since this is another way of measuring the width of the distribution, but including all outlying data points. Table 4-6 displays the sum of squares results averaged across all event windows. Specific event window outcomes by baseline can be found in Appendix D.

[10] Tables detailing the load-impact percentage error per event window for each baseline can be found in Appendix B.

[11] See Appendix C for results for each event window by baseline.

Table 4-6:
Average Sum of Squared Normalized Baseline Errors For All Customers

Baseline Method	Settlement Acct.	Individual Customer
Top 7 of 10, Hourly- 4hr Adj.	0.200	0.497
Top 7 of 10, Hourly- 6hr Adj.	0.230	0.543
10 of 10- 4hr Adj.	0.149	0.314
10 of 10- 6hr Adj.	0.163	0.337
Top 9 of 10, Hourly- 4hr Adj.	0.151	0.331
Top 9 of 10 Hourly- 6hr Adj.	0.171	0.361
Top 15 of 20, Hourly- 4hr Adj.	0.195	0.590
Top 15 of 20, Hourly- 6hr Adj.	0.220	0.628

5 RECOMMENDATIONS

Many of the 48 baseline methods improve upon OPA's current baseline method of Top 15 of 20 Hourly without adjustment. The bulk of improvement is achieved through the use of a same-day adjustment. 6 of the 48 baselines (Top 7, 9 and 10 of 10 Hourly each with 4-hour and 6-hour adjustments) provide average load-impact errors within +/- 2%. These average errors were calculated both for customers above one MW of contracted load reduction and for customers below one MW of contracted load reduction so that recommendations could be made for both DR-1 and DR-3. Below an average error value of 2% in absolute value, it becomes more important to look at the distribution of baseline errors (normalized by customer size) to address issues of fairness across customers (for DR-1) and settlement accounts (for DR-3). Also, calculating the sum of squared normalized errors and eliminating baselines with large sum-of-squares values ensures that no baseline with extreme errors for any customers or accounts will be selected. Baselines 10 of 10 and Top 9 of 10 Hourly each with 6-hour adjustments were shortlisted for their accuracy, fairness across settlement accounts and across individual customers (narrow error distributions) and lack of extreme errors. These two baselines are also accurate across multiple event windows and are thus robust across time. However, the 10 of 10 baseline with a 6-hour adjustment is the most accurate of the 2 baselines, is highly accurate for large and small customers, and has a narrow error distribution whether calculated at the settlement account level or at the individual customer level. The 10 of 10 baseline with a 6-hour adjustment is therefore the baseline we recommend to OPA for the DR-1 and DR-3 programs.

As mentioned in section 1, when a same-day adjustment is introduced, customers can behave in ways that undermine the accuracy of the adjustment. This is primarily a concern with very large customers who actively manage electricity costs. We cannot yet say whether this issue will actually arise. We recommend that if a same-day adjustment is adopted, that the method be reassessed after several events have passed to determine whether there is evidence that customers have reacted to the adjustment in ways that lead to inaccuracy. We recommend particular attention be given to the largest customers. We also recommend considering custom baseline methods for the largest customers. For example, if there is a customer that consistently shifts load to pre-event hours when an event is coming, then the 10 of 10 method without adjustment would likely be more accurate for that customer than the 10 of 10 method with the same-day adjustment.

APPENDIX A LOAD-IMPACT PERCENTAGE ERRORS BY BASELINE

	Program Average					
Baseline Method	Daily- No Adj	Daily- 4hr Adj	Daily- 6hr Adj	Hourly- No Adj	Hourly- 4h Adj	Hourly- 6h Adj
Top 3 of 10	46.7%	11.9%	14.6%	51.2%	-3.8%	-2.9%
Top 7 of 10	25.0%	7.6%	8.4%	27.6%	-0.9%	-0.1%
10 of 10	-2.4%	-0.5%	-0.5%	-2.4%	-0.5%	-0.5%
Top 15 of 20	22.7%	8.3%	7.9%	24.6%	0.3%	0.6%
Bottom 3 of 10	-66.4%	-23.8%	-25.4%	-72.4%	-0.4%	-2.3%
Bottom 7 of 10	-23.4%	-7.1%	-8.2%	-25.4%	1.7%	1.6%
Top 9 of 10	8.1%	2.2%	2.5%	9.0%	-1.2%	-1.0%
Bottom 15 of 20	-29.4%	-6.4%	-7.9%	-31.0%	1.3%	-0.2%

	Above 1MW Contracted					
Baseline Method	Daily- No Adj	Daily- 4hr Adj	Daily- 6hr Adj	Hourly- No Adj	Hourly- 4h Adj	Hourly- 6h Adj
Top 3 of 10	47.1%	10.9%	13.8%	51.8%	-5.0%	-3.9%
Top 7 of 10	26.6%	7.5%	8.3%	29.4%	-1.4%	-0.4%
10 of 10	-0.4%	-0.7%	-0.7%	-0.4%	-0.7%	-0.7%
Top 15 of 20	25.3%	8.5%	7.9%	27.4%	0.0%	0.3%
Bottom 3 of 10	-63.4%	-24.4%	-26.0%	-69.9%	0.0%	-2.3%
Bottom 7 of 10	-20.8%	-7.0%	-8.1%	-22.8%	2.1%	2.1%
Top 9 of 10	9.9%	2.0%	2.4%	10.8%	-1.5%	-1.2%
Bottom 15 of 20	-24.2%	-5.9%	-7.6%	-25.8%	2.0%	0.3%

	Below 1MW Contracted					
Baseline Method	Daily- No Adj	Daily- 4hr Adj	Daily- 6hr Adj	Hourly- No Adj	Hourly- 4h Adj	Hourly- 6h Adj
Top 3 of 10	44.1%	17.7%	19.0%	48.0%	3.1%	3.2%
Top 7 of 10	16.0%	8.2%	9.1%	17.4%	1.8%	1.8%
10 of 10	-13.9%	0.5%	0.5%	-13.9%	0.5%	0.5%
Top 15 of 20	7.5%	7.2%	7.9%	8.5%	2.3%	2.3%
Bottom 3 of 10	-83.7%	-20.0%	-22.2%	-86.9%	-2.4%	-2.3%
Bottom 7 of 10	-38.7%	-7.7%	-8.3%	-40.4%	-0.6%	-0.7%
Top 9 of 10	-1.9%	3.0%	3.3%	-1.5%	0.6%	0.6%
Bottom 15 of 20	-59.2%	-8.9%	-9.7%	-60.7%	-3.1%	-3.4%

APPENDIX B PROGRAM LOAD-IMPACT PERCENTAGE ERRORS BY BASELINE ACROSS EVENT WINDOWS

	Average Load Impact Percentage Error- Program						
Baseline Method	Window 1	Window 2	Window 3	Window 4	Window 5	Window 6	Average
Top 7 of 10, Hourly- 4hr Adj.	5.1%	3.4%	1.9%	-0.9%	-3.4%	-5.3%	3.3%
Top 7 of 10, Hourly- 6hr Adj.	6.7%	4.1%	2.4%	-0.1%	-2.0%	-3.6%	3.2%
10 of 10, Hourly- 4hr Adj.	2.6%	2.0%	1.5%	-0.5%	-2.2%	-3.9%	2.1%
10 of 10, Hourly- 6hr Adj.	3.2%	1.6%	0.9%	-0.5%	-1.2%	-2.2%	1.6%
Top 9 of 10, Hourly- 4hr Adj.	3.1%	2.0%	1.1%	-1.2%	-3.2%	-4.9%	2.6%
Top 9 of 10, Hourly- 6hr Adj.	4.3%	2.2%	0.8%	-1.0%	-2.2%	-3.4%	2.3%
Top 15 of 20, Hourly- No Adj.	26.8%	26.2%	26.0%	24.6%	24.0%	25.1%	25.5%
Top 15 of 20, Hourly- 4hr Adj.	3.6%	2.9%	2.0%	0.3%	-1.8%	-3.5%	2.4%
Top 15 of 20, Hourly- 6hr Adj.	4.9%	3.0%	1.8%	0.6%	-0.7%	-1.7%	2.1%

	Average Load Impact Percentage Error- Above 1MW Contracted						
Baseline Method	Window 1	Window 2	Window 3	Window 4	Window 5	Window 6	Average
Top 7 of 10, Hourly- 4hr Adj.	6.1%	3.9%	2.1%	-1.4%	-4.5%	-6.8%	4.1%
Top 7 of 10, Hourly- 6hr Adj.	8.1%	4.9%	2.6%	-0.4%	-2.7%	-4.8%	3.9%
10 of 10, Hourly- 4hr Adj.	3.1%	2.3%	1.7%	-0.7%	-2.7%	-4.7%	2.5%
10 of 10, Hourly- 6hr Adj.	3.9%	1.9%	0.9%	-0.7%	-1.5%	-2.7%	1.9%
Top 9 of 10, Hourly- 4hr Adj.	3.8%	2.4%	1.2%	-1.5%	-4.0%	-6.0%	3.2%
Top 9 of 10, Hourly- 6hr Adj.	5.3%	2.7%	0.9%	-1.2%	-2.7%	-4.2%	2.8%
Top 15 of 20, Hourly- No Adj.	30.3%	29.4%	29.0%	27.4%	26.7%	28.1%	28.5%
Top 15 of 20, Hourly- 4hr Adj.	4.2%	3.2%	2.0%	0.0%	-2.6%	-4.5%	2.8%
Top 15 of 20, Hourly- 6hr Adj.	6.0%	3.5%	1.8%	0.3%	-1.2%	-2.4%	2.5%

	Average Load Impact Percentage Error- Below 1MW Contracted						
Baseline Method	Window 1	Window 2	Window 3	Window 4	Window 5	Window 6	Average
Top 7 of 10, Hourly- 4hr Adj.	-0.4%	0.3%	0.9%	1.8%	2.8%	3.4%	1.6%
Top 7 of 10, Hourly- 6hr Adj.	-1.6%	-0.5%	1.1%	1.8%	2.3%	3.0%	1.7%
10 of 10, Hourly- 4hr Adj.	-0.3%	0.2%	0.4%	0.5%	0.9%	1.0%	0.6%
10 of 10, Hourly- 6hr Adj.	-0.8%	-0.3%	0.5%	0.5%	0.4%	0.7%	0.5%
Top 9 of 10, Hourly- 4hr Adj.	-0.6%	-0.2%	0.2%	0.6%	1.3%	1.6%	0.8%
Top 9 of 10, Hourly- 6hr Adj.	-1.1%	-0.6%	0.4%	0.6%	0.7%	1.2%	0.8%
Top 15 of 20, Hourly- No Adj.	6.9%	7.5%	8.3%	8.5%	8.3%	8.3%	8.0%
Top 15 of 20, Hourly- 4hr Adj.	0.1%	1.2%	1.8%	2.3%	2.7%	2.5%	1.8%
Top 15 of 20, Hourly- 6hr Adj.	-1.2%	0.3%	1.9%	2.3%	2.3%	2.4%	1.7%

APPENDIX C DISTRIBUTION RANGE OF 90% OF NORMALIZED BASELINE ERRORS BY BASELINE ACROSS EVENT WINDOWS

	Distribution Range of 90% of Normalized Baseline Errors by Individual Customer						
Baseline Method	Window 1	Window 2	Window 3	Window 4	Window 5	Window 6	Average
Top 7 of 10, Hourly- 4hr Adj.	17.3%	15.0%	15.0%	17.6%	16.2%	13.4%	15.8%
Top 7 of 10, Hourly- 6hr Adj.	21.1%	18.4%	14.2%	16.7%	20.0%	15.8%	17.7%
10 of 10, Hourly- 4hr Adj.	13.0%	12.6%	14.2%	17.5%	15.5%	17.3%	15.0%
10 of 10, Hourly- 6hr Adj.	16.1%	15.9%	12.9%	16.1%	18.6%	15.9%	15.9%
Top 9 of 10, Hourly- 4hr Adj.	13.0%	12.4%	14.2%	17.6%	13.4%	14.8%	14.2%
Top 9 of 10, Hourly- 6hr Adj.	16.6%	15.1%	13.9%	14.2%	16.5%	14.9%	15.2%
Top 15 of 20, Hourly- No Adj.	56.7%	54.9%	58.1%	58.2%	51.3%	63.4%	57.1%
Top 15 of 20, Hourly- 4hr Adj.	14.5%	13.9%	16.3%	19.0%	16.8%	16.4%	16.2%
Top 15 of 20, Hourly- 6hr Adj.	21.6%	17.6%	15.8%	16.6%	18.8%	15.7%	17.7%

	Distribution Range of 90% of Normalized Baseline Errors by Settlement Account						
Baseline Method	Window 1	Window 2	Window 3	Window 4	Window 5	Window 6	Average
Top 7 of 10, Hourly- 4hr Adj.	25.5%	12.8%	16.3%	22.2%	24.0%	20.2%	20.2%
Top 7 of 10, Hourly- 6hr Adj.	31.1%	18.7%	13.7%	20.4%	23.8%	22.1%	21.6%
10 of 10, Hourly- 4hr Adj.	25.2%	12.8%	13.2%	21.9%	23.5%	19.5%	19.4%
10 of 10, Hourly- 6hr Adj.	29.1%	19.7%	12.2%	17.0%	25.2%	21.8%	20.8%
Top 9 of 10, Hourly- 4hr Adj.	25.5%	13.4%	13.4%	22.0%	22.3%	19.2%	19.3%
Top 9 of 10, Hourly- 6hr Adj.	29.9%	20.5%	12.2%	17.1%	21.9%	21.5%	20.5%
Top 15 of 20, Hourly- No Adj.	61.1%	56.5%	54.9%	45.6%	40.8%	42.1%	50.2%
Top 15 of 20, Hourly- 4hr Adj.	26.3%	12.4%	15.1%	22.1%	22.9%	20.4%	19.9%
Top 15 of 20, Hourly- 6hr Adj.	31.6%	21.5%	14.4%	20.8%	24.2%	22.4%	22.5%

APPENDIX D DISTRIBUTION RANGES WITH LOWER AND UPPER LIMITS FOR EACH BASELINE ACROSS EVENT WINDOWS

Comparison Level	Event Window	80% of Accounts			90% of Accounts		
		Range	Lower	Upper	Range	Lower	Upper
Individual Customer	1	9.2%	-5.1%	4.1%	17.3%	-7.0%	10.3%
	2	9.7%	-4.7%	5.0%	15.0%	-7.6%	7.4%
	3	12.0%	-5.2%	6.8%	15.0%	-7.1%	7.9%
	4	13.6%	-6.0%	7.6%	17.6%	-7.7%	9.9%
	5	12.0%	-5.3%	6.7%	16.2%	-6.0%	10.2%
	6	9.9%	-4.3%	5.6%	13.4%	-5.7%	7.7%
Settlement Account	1	14.9%	-4.6%	10.3%	25.5%	-7.0%	18.5%
	2	11.8%	-5.1%	6.7%	12.8%	-5.2%	7.6%
	3	13.4%	-5.6%	7.8%	16.3%	-6.9%	9.4%
	4	20.1%	-11.5%	8.6%	22.2%	-12.3%	9.9%
	5	20.6%	-12.4%	8.2%	24.0%	-15.5%	8.5%
	6	18.6%	-12.3%	6.3%	20.2%	-13.2%	7.0%

Top 7 of 10- 6hr Adjustment							
Comparison Level	Event Window	80% of Accounts			90% of Accounts		
		Range	Lower	Upper	Range	Lower	Upper
Individual Customer	1	10.9%	-7.3%	3.6%	21.1%	-8.3%	12.8%
	2	10.1%	-5.4%	4.7%	18.4%	-7.8%	10.6%
	3	10.9%	-4.9%	6.0%	14.2%	-6.8%	7.4%
	4	11.6%	-4.7%	6.9%	16.7%	-7.0%	9.7%
	5	11.1%	-4.8%	6.3%	20.0%	-9.4%	10.6%
	6	11.3%	-4.4%	6.9%	15.8%	-6.8%	9.0%
Settlement Account	1	25.3%	-7.4%	17.9%	31.1%	-7.9%	23.2%
	2	15.9%	-5.3%	10.6%	18.7%	-7.8%	10.9%
	3	11.5%	-4.9%	6.6%	13.7%	-5.8%	7.9%
	4	14.7%	-5.8%	8.9%	20.4%	-11.0%	9.4%
	5	18.5%	-10.5%	8.0%	23.8%	-13.8%	10.0%
	6	15.5%	-8.6%	6.9%	22.1%	-14.2%	7.9%

10 of 10- 4hr Adjustment							
		80% of Accounts			90% of Accounts		
Comparison Level	Event Window	Range	Lower	Upper	Range	Lower	Upper
Individual Customer	1	7.6%	-4.0%	3.6%	13.0%	-6.1%	6.9%
	2	8.6%	-3.6%	5.0%	12.6%	-6.3%	6.3%
	3	9.0%	-3.9%	5.1%	14.2%	-6.0%	8.2%
	4	10.8%	-4.7%	6.1%	17.5%	-7.9%	9.6%
	5	10.4%	-4.7%	5.7%	15.5%	-6.5%	9.0%
	6	8.7%	-4.1%	4.6%	17.3%	-9.0%	8.3%
Settlement Account	1	14.9%	-5.4%	9.5%	25.2%	-7.8%	17.4%
	2	11.4%	-5.9%	5.5%	12.8%	-6.8%	6.0%
	3	10.6%	-2.8%	7.8%	13.2%	-5.0%	8.2%
	4	17.1%	-9.1%	8.0%	21.9%	-12.3%	9.6%
	5	19.2%	-12.4%	6.8%	23.5%	-14.5%	9.0%
	6	18.0%	-11.2%	6.8%	19.5%	-12.6%	6.9%

10 of 10- 6hr Adjustment							
		80% of Accounts			90% of Accounts		
Comparison Level	Event Window	Range	Lower	Upper	Range	Lower	Upper
Individual Customer	1	8.5%	-4.3%	4.2%	16.1%	-7.2%	8.9%
	2	8.5%	-3.8%	4.7%	15.9%	-6.5%	9.4%
	3	8.6%	-3.7%	4.9%	12.9%	-5.6%	7.3%
	4	10.4%	-4.6%	5.8%	16.1%	-7.5%	8.6%
	5	9.7%	-4.5%	5.2%	18.6%	-8.6%	10.0%
	6	9.0%	-4.3%	4.7%	15.9%	-7.0%	8.9%
Settlement Account	1	24.8%	-5.6%	19.2%	29.1%	-9.2%	19.9%
	2	15.5%	-6.1%	9.4%	19.7%	-9.5%	10.2%
	3	8.9%	-3.2%	5.7%	12.2%	-4.9%	7.3%
	4	14.1%	-7.5%	6.6%	17.0%	-8.4%	8.6%
	5	19.7%	-12.9%	6.8%	25.2%	-15.2%	10.0%
	6	14.8%	-7.5%	7.3%	21.8%	-12.9%	8.9%

Top 9 of 10- 4hr Adjustment							
		80% of Accounts			90% of Accounts		
Comparison Level	Event Window	Range	Lower	Upper	Range	Lower	Upper
Individual Customer	1	8.5%	-4.2%	4.3%	13.0%	-6.0%	7.0%
	2	8.1%	-3.4%	4.7%	12.4%	-6.4%	6.0%
	3	8.7%	-4.0%	4.7%	14.2%	-5.9%	8.3%
	4	11.2%	-4.8%	6.4%	17.6%	-8.2%	9.4%
	5	10.1%	-4.4%	5.7%	13.4%	-5.7%	7.7%
	6	8.6%	-4.0%	4.6%	14.8%	-7.0%	7.8%
Settlement Account	1	15.1%	-5.3%	9.8%	25.5%	-8.1%	17.4%
	2	11.7%	-5.7%	6.0%	13.4%	-7.4%	6.0%
	3	12.1%	-4.0%	8.1%	13.4%	-5.1%	8.3%
	4	16.6%	-8.5%	8.1%	22.0%	-12.6%	9.4%
	5	19.5%	-12.8%	6.7%	22.3%	-14.9%	7.4%
	6	17.5%	-11.8%	5.7%	19.2%	-12.8%	6.4%

Top 9 of 10- 6hr Adjustment							
		80% of Accounts			90% of Accounts		
Comparison Level	Event Window	Range	Lower	Upper	Range	Lower	Upper
Individual Customer	1	8.7%	-4.8%	3.9%	16.6%	-7.2%	9.4%
	2	9.0%	-4.0%	5.0%	15.1%	-6.3%	8.8%
	3	8.6%	-3.5%	5.1%	13.9%	-6.6%	7.3%
	4	9.4%	-4.0%	5.4%	14.2%	-6.6%	7.6%
	5	8.5%	-3.9%	4.6%	16.5%	-8.5%	8.0%
	6	9.0%	-4.1%	4.9%	14.9%	-6.4%	8.5%
Settlement Account	1	24.7%	-5.6%	19.1%	29.9%	-9.4%	20.5%
	2	14.7%	-5.9%	8.8%	20.5%	-10.0%	10.5%
	3	10.3%	-4.1%	6.2%	12.2%	-4.9%	7.3%
	4	13.3%	-6.6%	6.7%	17.1%	-8.5%	8.6%
	5	19.3%	-13.3%	6.0%	21.9%	-14.2%	7.7%
	6	12.2%	-6.4%	5.8%	21.5%	-13.3%	8.2%

Top 15 of 20- No Adjustment							
		80% of Accounts			90% of Accounts		
Comparison Level	Event Window	Range	Lower	Upper	Range	Lower	Upper
Individual Customer	1	29.0%	-8.8%	20.2%	56.7%	-16.3%	40.4%
	2	32.5%	-8.4%	24.1%	54.9%	-17.2%	37.7%
	3	33.9%	-8.4%	25.5%	58.1%	-17.8%	40.3%
	4	33.1%	-8.3%	24.8%	58.2%	-17.6%	40.6%
	5	31.1%	-8.8%	22.3%	51.3%	-17.5%	33.8%
	6	31.1%	-8.6%	22.5%	63.4%	-17.7%	45.7%
Settlement Account	1	47.3%	-3.1%	44.2%	61.1%	-3.6%	57.5%
	2	41.7%	-4.0%	37.7%	56.5%	-4.1%	52.4%
	3	38.1%	-2.7%	35.4%	54.9%	-4.6%	50.3%
	4	31.6%	-2.8%	28.8%	45.6%	-5.0%	40.6%
	5	36.3%	-4.8%	31.5%	40.8%	-7.0%	33.8%
	6	35.6%	-4.4%	31.2%	42.1%	-7.1%	35.0%

Top 15 of 20- 4hr Adjustment							
		80% of Accounts			90% of Accounts		
Comparison Level	Event Window	Range	Lower	Upper	Range	Lower	Upper
Individual Customer	1	8.9%	-4.8%	4.1%	14.5%	-6.6%	7.9%
	2	8.9%	-3.9%	5.0%	13.9%	-6.6%	7.3%
	3	11.5%	-5.9%	5.6%	16.3%	-7.3%	9.0%
	4	13.2%	-6.2%	7.0%	19.0%	-8.1%	10.9%
	5	11.7%	-5.2%	6.5%	16.8%	-7.0%	9.8%
	6	10.5%	-4.5%	6.0%	16.4%	-6.5%	9.9%
Settlement Account	1	14.2%	-6.3%	7.9%	26.3%	-7.4%	18.9%
	2	9.4%	-4.4%	5.0%	12.4%	-6.8%	5.6%
	3	13.4%	-7.0%	6.4%	15.1%	-8.4%	6.7%
	4	19.7%	-11.8%	7.9%	22.1%	-14.1%	8.0%
	5	20.0%	-13.5%	6.5%	22.9%	-15.3%	7.6%
	6	16.5%	-10.5%	6.0%	20.4%	-13.1%	7.3%

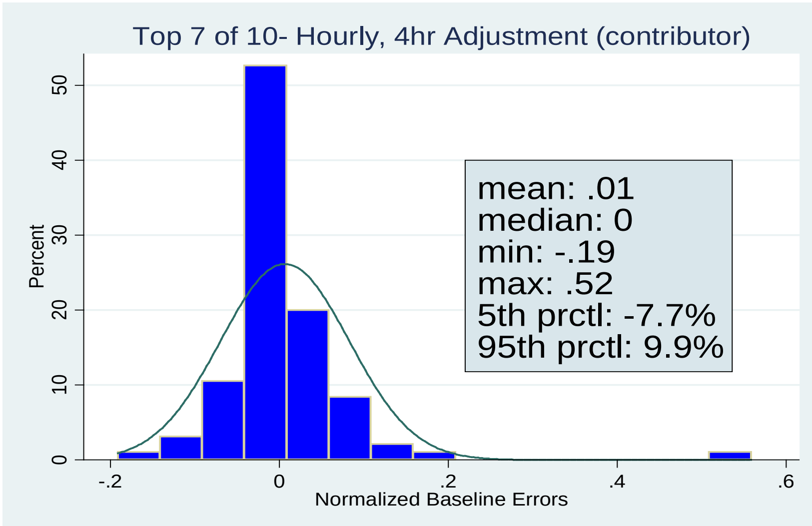
Top 15 of 20- 6hr Adjustment							
		80% of Accounts			90% of Accounts		
Comparison Level	Event Window	Range	Lower	Upper	Range	Lower	Upper
Individual Customer	1	10.3%	-6.3%	4.0%	21.6%	-8.5%	13.1%
	2	10.6%	-5.5%	5.1%	17.6%	-7.6%	10.0%
	3	10.3%	-4.1%	6.2%	15.8%	-7.3%	8.5%
	4	12.3%	-5.3%	7.0%	16.6%	-7.1%	9.5%
	5	12.4%	-5.1%	7.3%	18.8%	-8.0%	10.8%
	6	12.0%	-5.4%	6.6%	15.7%	-6.1%	9.6%
Settlement Account	1	25.6%	-7.3%	18.3%	31.6%	-8.5%	23.1%
	2	11.7%	-4.6%	7.1%	21.5%	-9.5%	12.0%
	3	10.5%	-4.3%	6.2%	14.4%	-8.1%	6.3%
	4	19.0%	-11.6%	7.4%	20.8%	-12.6%	8.2%
	5	21.5%	-13.6%	7.9%	24.2%	-14.2%	10.0%
	6	13.3%	-6.7%	6.6%	22.4%	-15.3%	7.1%

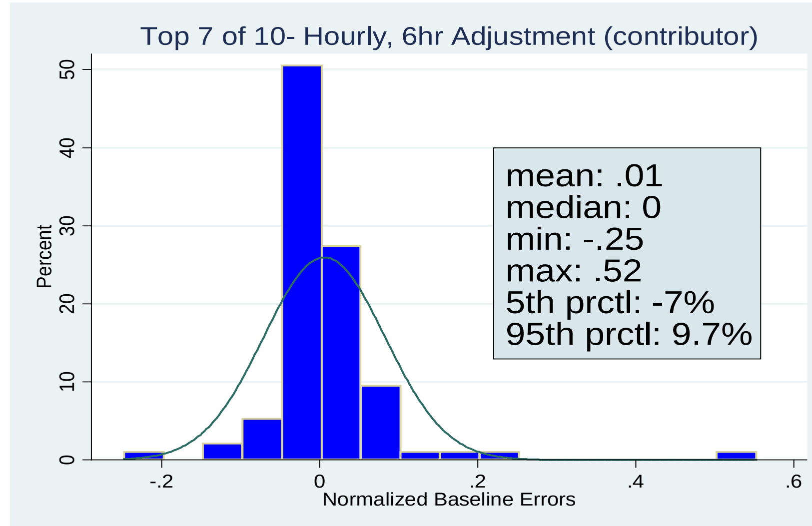
APPENDIX E SUM OF SQUARED NORMALIZED ERRORS BY BASELINE ACROSS EVENT WINDOWS

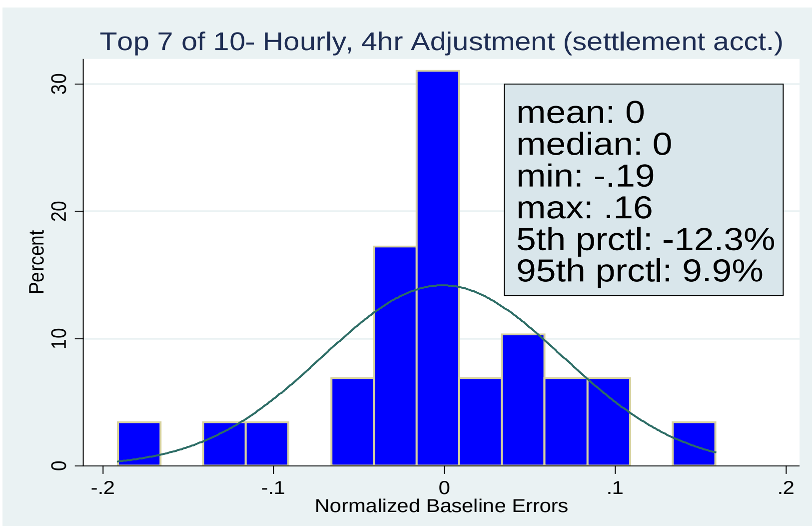
	Sum of Squared Normalized Baseline Errors by Individual Customer						
Baseline Method	Window 1	Window 2	Window 3	Window 4	Window 5	Window 6	Average
Top 7 of 10, Hourly- 4hr Adj.	0.465	0.430	0.501	0.551	0.496	0.541	0.497
Top 7 of 10, Hourly- 6hr Adj.	0.605	0.494	0.515	0.559	0.499	0.583	0.543
10 of 10, Hourly- 4hr Adj.	0.327	0.298	0.321	0.331	0.289	0.316	0.314
10 of 10, Hourly- 6hr Adj.	0.397	0.337	0.333	0.334	0.287	0.332	0.337
Top 9 of 10, Hourly- 4hr Adj.	0.350	0.313	0.332	0.347	0.306	0.340	0.331
Top 9 of 10, Hourly- 6hr Adj.	0.447	0.363	0.348	0.350	0.302	0.358	0.361
Top 15 of 20, Hourly- No Adj.	3.568	3.796	4.038	4.084	3.755	3.856	3.850
Top 15 of 20, Hourly- 4hr Adj.	0.410	0.421	0.552	0.712	0.684	0.758	0.590
Top 15 of 20, Hourly- 6hr Adj.	0.532	0.463	0.559	0.716	0.684	0.815	0.628

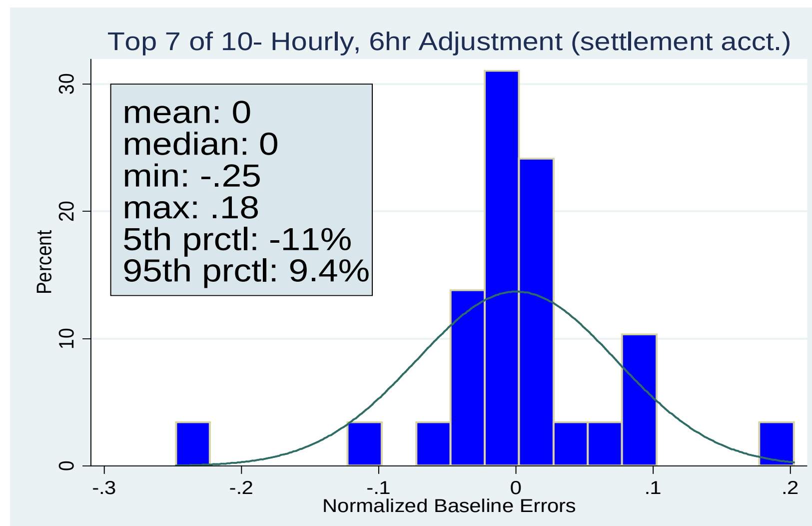
	Sum of Squared Normalized Baseline Errors by Settlement Account						
Baseline Method	Window 1	Window 2	Window 3	Window 4	Window 5	Window 6	Average
Top 7 of 10, Hourly- 4hr Adj.	0.365	0.253	0.217	0.138	0.117	0.107	0.200
Top 7 of 10, Hourly- 6hr Adj.	0.474	0.307	0.237	0.148	0.116	0.099	0.230
10 of 10, Hourly- 4hr Adj.	0.246	0.174	0.139	0.104	0.108	0.121	0.149
10 of 10, Hourly- 6hr Adj.	0.309	0.205	0.149	0.104	0.104	0.106	0.163
Top 9 of 10, Hourly- 4hr Adj.	0.270	0.185	0.142	0.099	0.099	0.112	0.151
Top 9 of 10, Hourly- 6hr Adj.	0.355	0.226	0.156	0.101	0.093	0.096	0.171
Top 15 of 20, Hourly- No Adj.	2.451	2.471	2.427	2.131	1.791	1.727	2.166
Top 15 of 20, Hourly- 4hr Adj.	0.312	0.225	0.199	0.143	0.140	0.152	0.195
Top 15 of 20, Hourly- 6hr Adj.	0.404	0.262	0.211	0.148	0.141	0.153	0.220

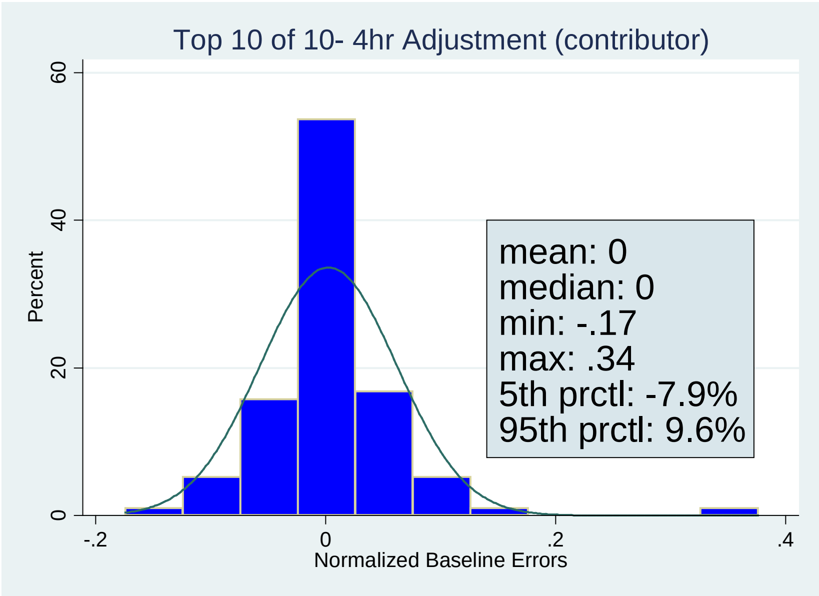
APPENDIX F HISTOGRAM DISTRIBUTIONS OF NORMALIZED ERROR BY BASELINE AND TYPE (EVENT WINDOW 3 P.M. TO 7 P.M.)

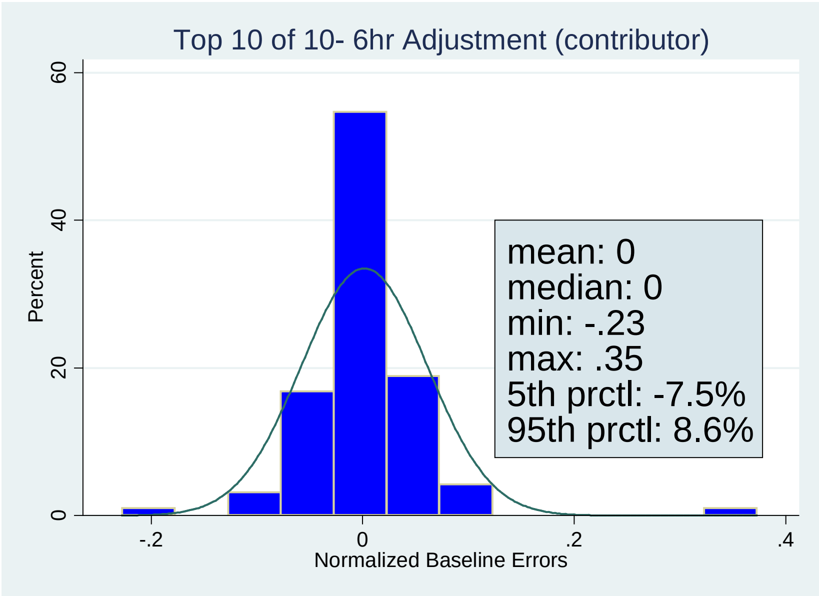


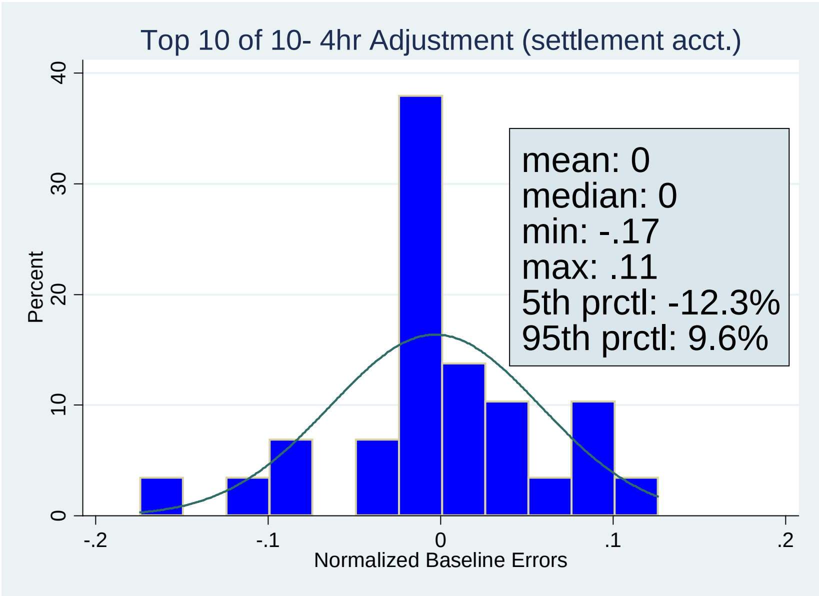


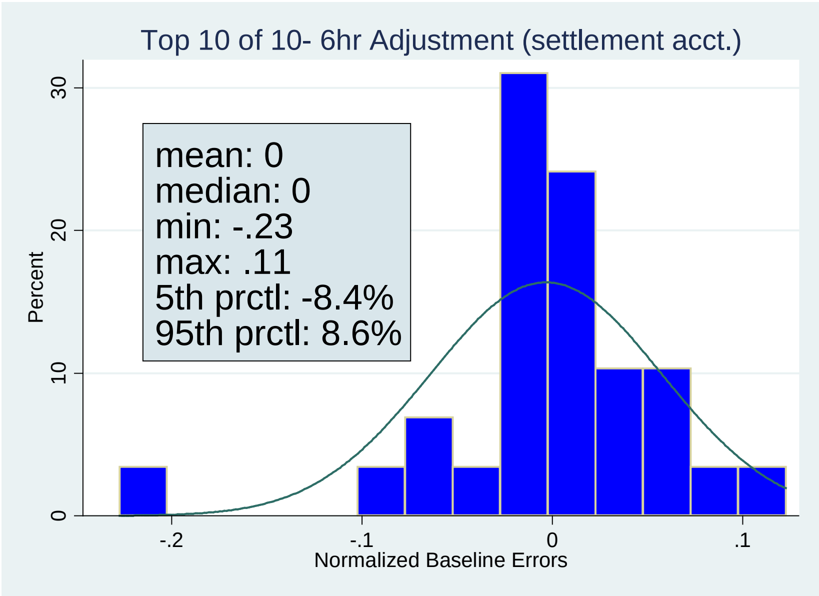


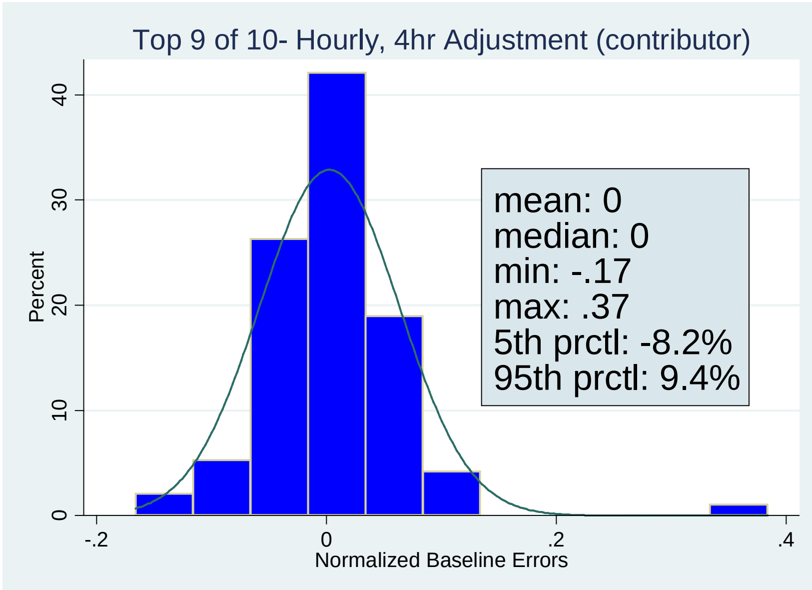


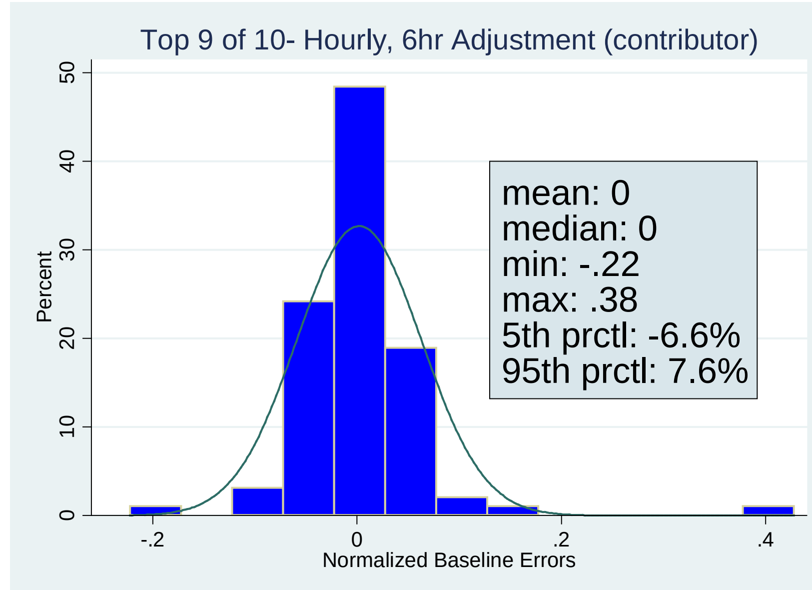

Top 10 of 10- 4hr Adjustment (contributor)
Percent
0
20
40
60
-.2
0
.2
.4
Normalized Baseline Errors
mean: 0
median: 0
min: -.17
max: .34
5th prctl: -7.9%
95th prctl: 9.6%

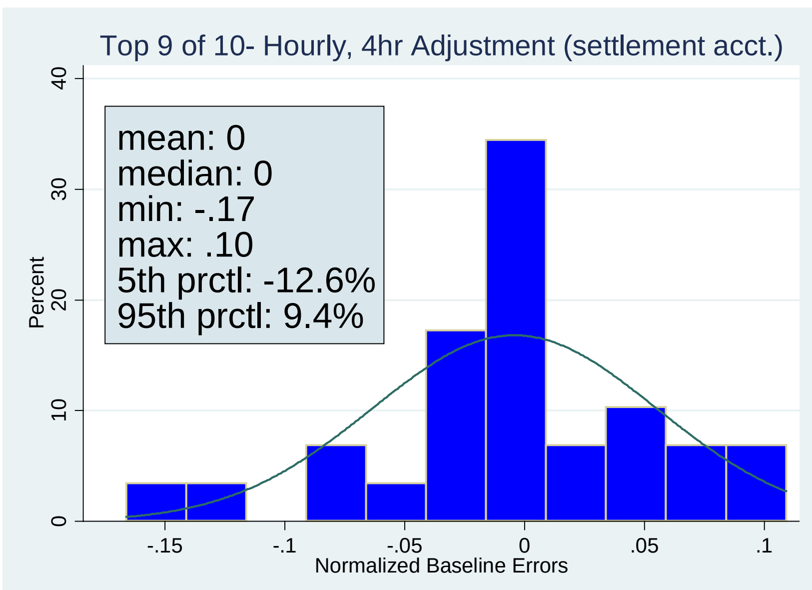

Top 10 of 10- 6hr Adjustment (contributor)
Percent
0
20
40
60
-.2
0
.2
.4
Normalized Baseline Errors
mean: 0
median: 0
min: -.23
max: .35
5th prctl: -7.5%
95th prctl: 8.6%

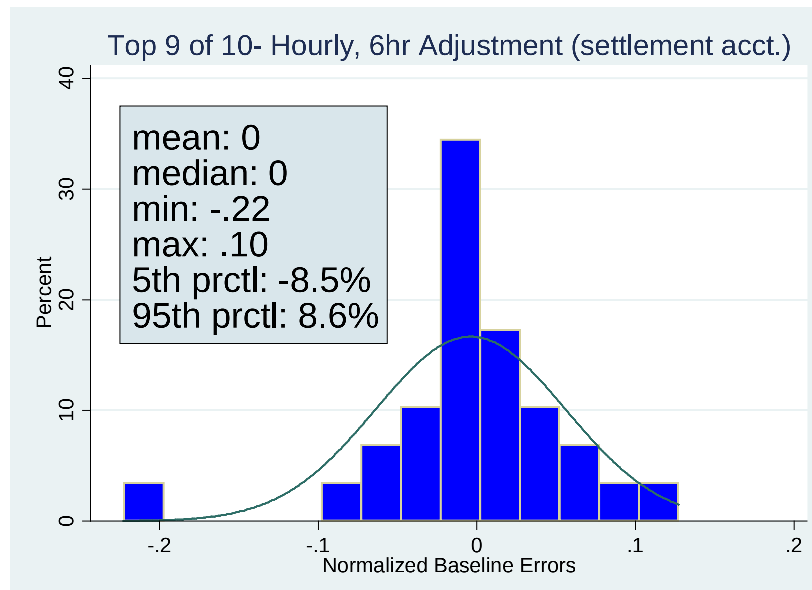

Top 10 of 10- 4hr Adjustment (settlement acct.)
Percent
0
10
20
30
40
-.2
-.1
0
.1
.2
Normalized Baseline Errors
mean: 0
median: 0
min: -.17
max: .11
5th prctl: -12.3%
95th prctl: 9.6%

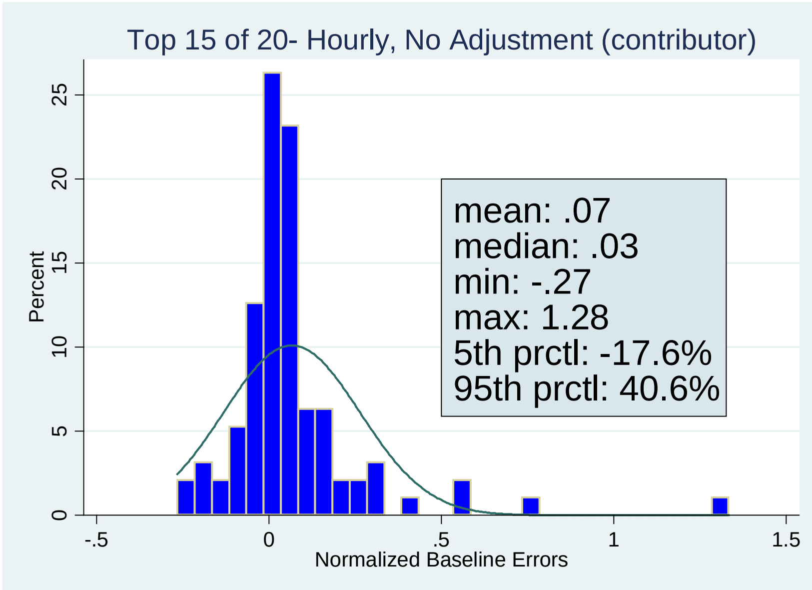

Top 10 of 10- 6hr Adjustment (settlement acct.)
Percent
0
10
20
30
-.2
-.1
0
.1
Normalized Baseline Errors
mean: 0
median: 0
min: -.23
max: .11
5th prctl: -8.4%
95th prctl: 8.6%

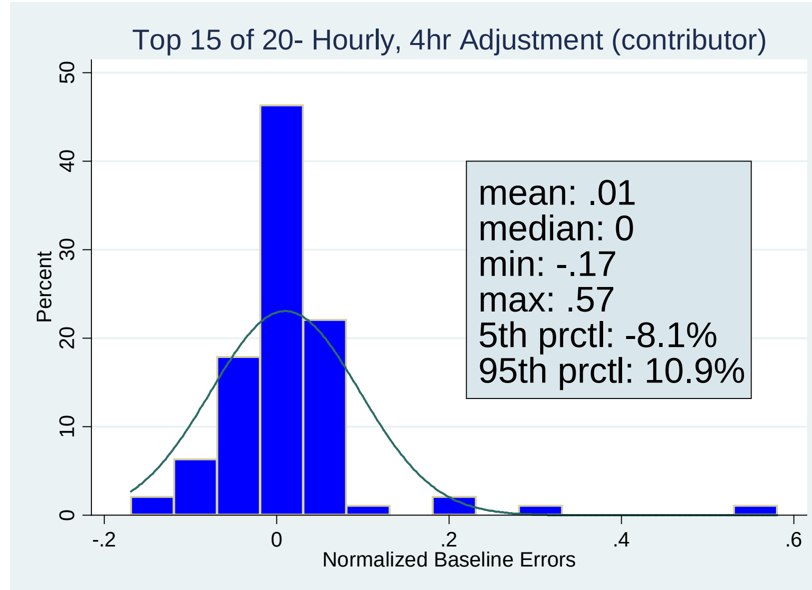

Top 9 of 10- Hourly, 4hr Adjustment (contributor)
Percent
0
10
20
30
40
mean: 0
median: 0
min: -.17
max: .37
5th prctl: -8.2%
95th prctl: 9.4%
-.2
0
.2
.4
Normalized Baseline Errors

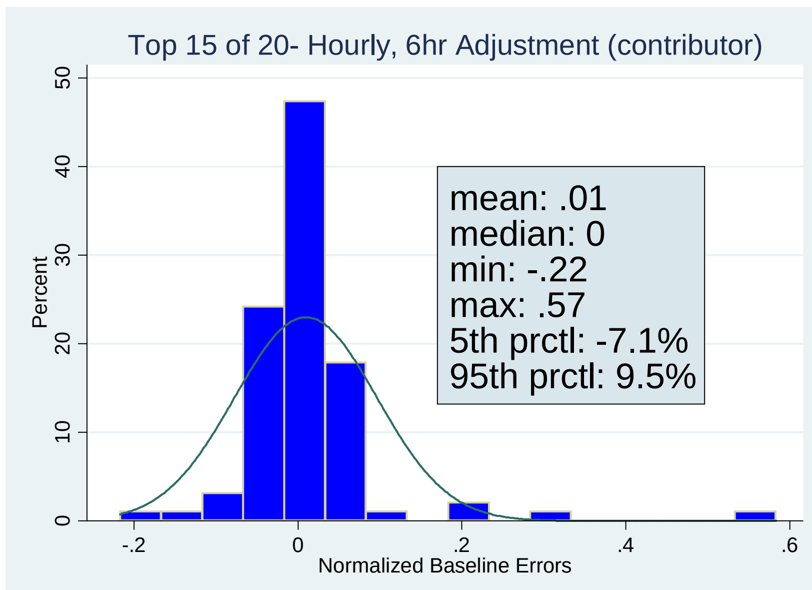

Top 9 of 10- Hourly, 6hr Adjustment (contributor)
Percent
0
10
20
30
40
50
mean: 0
median: 0
min: -.22
max: .38
5th prctl: -6.6%
95th prctl: 7.6%
-.2
0
.2
.4
Normalized Baseline Errors

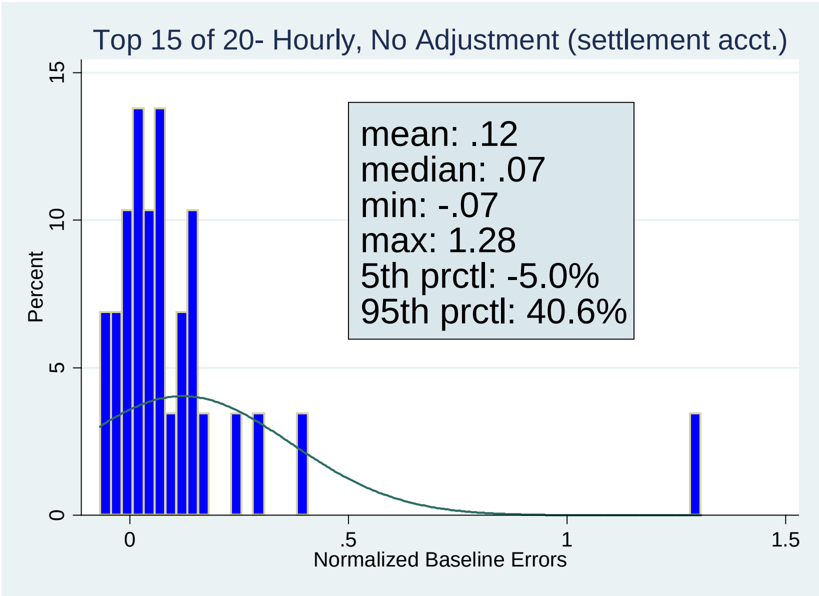

Top 9 of 10- Hourly, 4hr Adjustment (settlement acct.)
Percent
0
10
20
30
40
mean: 0
median: 0
min: -.17
max: .10
5th prctl: -12.6%
95th prctl: 9.4%
-.15
-.1
-.05
0
.05
.1
Normalized Baseline Errors

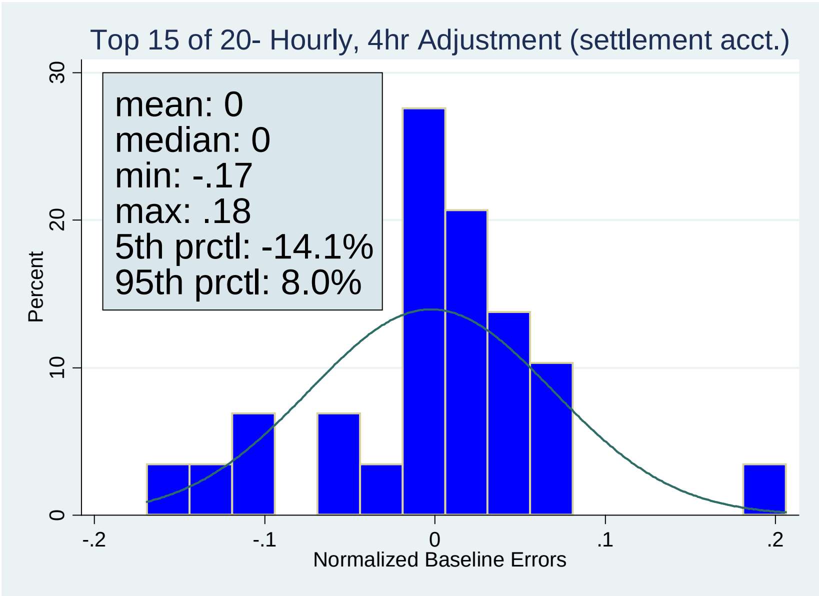

Top 9 of 10- Hourly, 6hr Adjustment (settlement acct.)
Percent
0
10
20
30
40
mean: 0
median: 0
min: -.22
max: .10
5th prctl: -8.5%
95th prctl: 8.6%
-.2
-.1
0
.1
.2
Normalized Baseline Errors

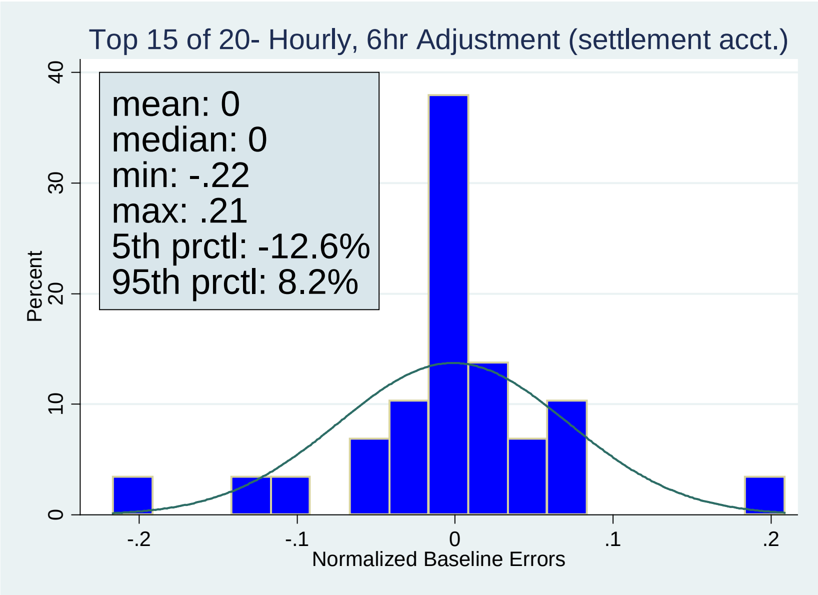

Top 15 of 20- Hourly, No Adjustment (contributor)
mean: .07
median: .03
min: -.27
max: 1.28
5th prctl: -17.6%
95th prctl: 40.6%
Percent
0
5
10
15
20
25
-.5
0
.5
1
1.5
Normalized Baseline Errors


Top 15 of 20- Hourly, 4hr Adjustment (contributor)
mean: .01
median: 0
min: -.17
max: .57
5th prctl: -8.1%
95th prctl: 10.9%
Percent
0
10
20
30
40
50
-.2
0
.2
.4
.6
Normalized Baseline Errors


Top 15 of 20- Hourly, 6hr Adjustment (contributor)
mean: .01
median: 0
min: -.22
max: .57
5th prctl: -7.1%
95th prctl: 9.5%
Percent
0
10
20
30
40
50
-.2
0
.2
.4
.6
Normalized Baseline Errors


Top 15 of 20- Hourly, No Adjustment (settlement acct.)
mean: .12
median: .07
min: -.07
max: 1.28
5th prctl: -5.0%
95th prctl: 40.6%
Percent
0
5
10
15
0
.5
1
1.5
Normalized Baseline Errors


Top 15 of 20- Hourly, 4hr Adjustment (settlement acct.)
mean: 0
median: 0
min: -.17
max: .18
5th prctl: -14.1%
95th prctl: 8.0%
Percent
0
10
20
30
-.2
-.1
0
.1
.2
Normalized Baseline Errors


Top 15 of 20- Hourly, 6hr Adjustment (settlement acct.)
mean: 0
median: 0
min: -.22
max: .21
5th prctl: -12.6%
95th prctl: 8.2%
Percent
0
10
20
30
40
-.2
-.1
0
.1
.2
Normalized Baseline Errors

APPENDIX G EXAMPLE OF SAME-DAY ADJUSTMENT CALCULATION

A same-day adjustment is a way to reduce the error between the baseline and actual loads during an event period. The idea behind the adjustment is that the period before the event—when customers' loads are unperturbed—can be used to check the accuracy of the baseline calculation for that day and adjust the baseline up or down to make it more accurate. If the baseline calculation is below the actual load before the event, then it is assumed that the baseline would also be below what the unperturbed load would have been during the event period. And similarly, if the baseline is above the actual load before the event, then it is also assumed to be above what the unperturbed load would have been during the event period. This relationship between the two loads is used to determine the same-day adjustment for that customer. The calculation is illustrated using a concrete example:

Suppose a 4-hour same-day adjustment is to be applied to a customer during an event window of 3 P.M. to 7 P.M.. The hours 9 A.M. to 1 P.M. are used to determine the adjustment (this includes the 2-hour buffer between the adjustment period and the event period—this buffer is explained below). A 6-hour adjustment would use the hours 7 A.M. to 1 P.M..

Suppose a customer has the actual hourly load shown in Table G-1 below. For simplicity of the example, only unperturbed load is considered—that is, the customer's use is known in the absence of the event. This allows the focus to be on the effect of the adjustment and how the adjusted baseline compares to the number it is supposed to predict. Suppose the 10 of 10 baseline method with no adjustment is used for this customer giving the baseline also shown in Table G-1. Comparing across loads during the same hour, it is obvious that the unadjusted baseline is higher than the actual—both before the event and during the event. The baseline therefore over-predicts the actual load.

Table G-1:
Actual Use and Unadjusted Top 10 of 10 Baseline for a Particular Customer

	10:00 AM	11:00 AM	12:00 PM	1:00 PM	2:00 PM	3:00 PM	4:00 PM	5:00 PM	6:00 PM	7:00 PM
Actual Unperturbed Use (kW)	4.2	4	3.7	4.1	4	4.1	4	4.1	3.9	3.9
10 of 10 Baseline with No Adjustment (kW)	5.1	5	4.8	5.1	4.9	5	5.1	5.3	4.9	5

The same-day adjustment consists of applying a multiplier to the baseline during the event period. To calculate that multiplier two numbers must be calculated: the average actual use during the pre-event period 9 A.M. to 1 P.M. and the average baseline with no adjustment during the same period. In this example, the average actual use during the period is:

$$avguse_{pre-event} = \frac{4.2 + 4.0 + 3.7 + 4.1}{4} = 4\ kW$$

The average baseline with no adjustment is:

$$avgbaseline_{pre-event} = \frac{5.1 + 5.0 + 4.8 + 5.1}{4} = 5\ kW$$

The multiplier to be applied is the ratio of the average actual use to the average unadjusted baseline:

$$actual - to - baselineloadratio = \frac{avguse_{pre-event}}{avgbaseline_{pre-event}} = \frac{4kW}{5kW} = 0.8$$

The rationale behind applying this multiplier to the baseline during the event period is that given the proximity of the pre-event period to the event window, it is reasonable to assume that the baseline percentage error would be roughly the same during the event window as it was during the pre-event period. To adjust the baseline closer to what the actual load would be without an event, the baseline of each event hour is multiplied by the actual-to-baseline load ratio. In this case, that ratio is 0.8.

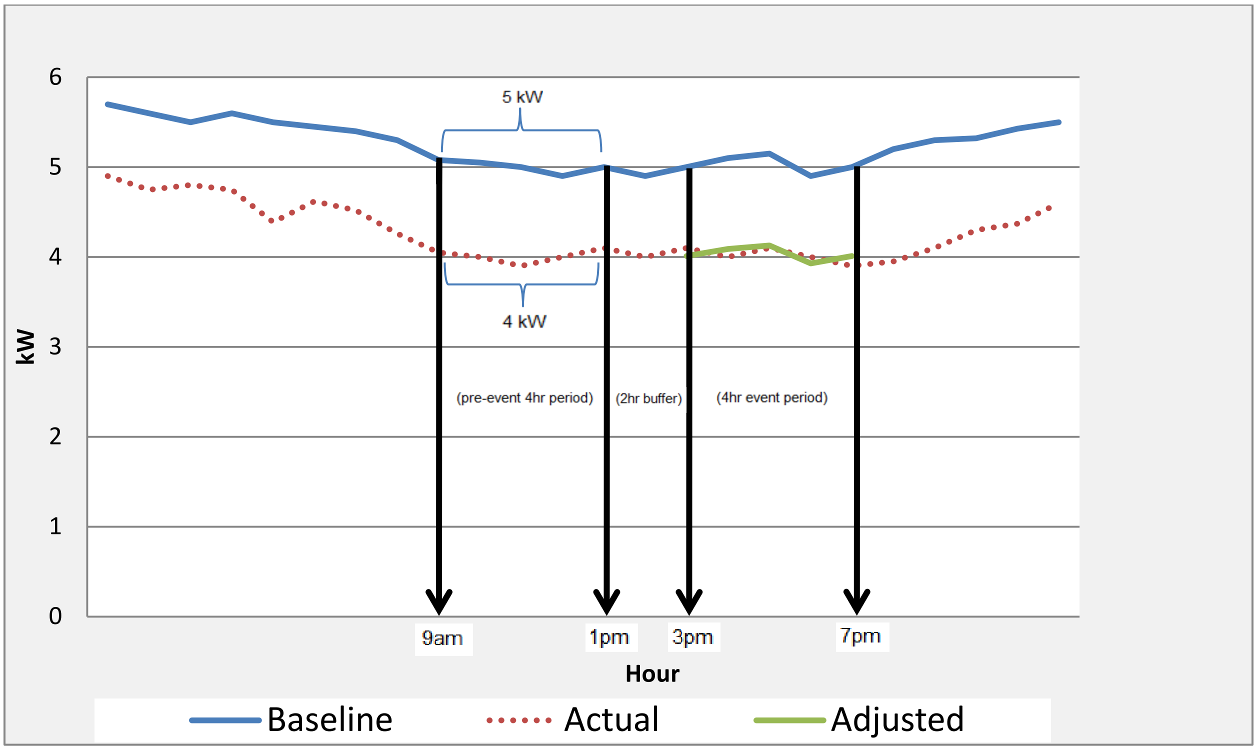
The result of this calculation for this example is shown in Table G-2. Note that the adjustment only applies to the hours during the event window. The numbers in the adjusted row equal the numbers in the unadjusted row, multiplied by 0.8. The adjusted baseline numbers are much closer to the actual us.

Table G-2:
Unadjusted 10 of 10 Baseline and 10 of 10 Baseline with Four-Hour Same-Day Adjustment for a Particular Customer

	4:00 PM	5:00 PM	6:00 PM	7:00 PM
Top 10 of 10 baseline with no adjustment (kW)	5.1	5.3	4.9	5
Top 10 of 10 baseline with same-day adjustment (kW)	4.1	4.2	3.9	4
Actual Unperturbed Use (kW)	4	4.1	3.9	3.9

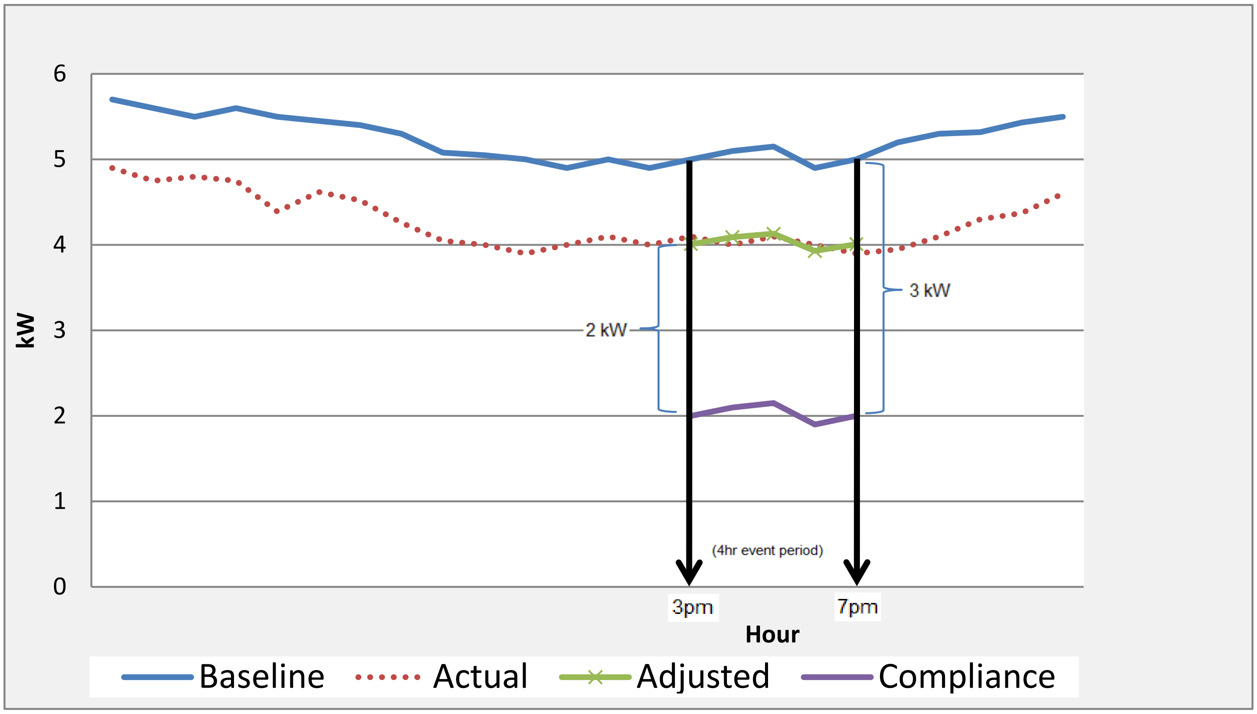
Figure G-1 shows the same example graphically. The black arrows identify the 4-hour event period lasting from 3 P.M. to 7 P.M. and the 4-hour pre-event period from 9 A.M. to 1 P.M. (with a buffer from 1 P.M. to 3 P.M.). The pre-event period baseline average is 5 kW, while the pre-event actual use is 4 kW. This results in an actual-to-baseline load ratio of 0.8. Multiplying the baseline load during each event hour by the ratio gives the adjusted baseline. Notice how the adjusted load is much closer to the unperturbed actual load during the event hours. The error between the baseline and actual loads has been reduced dramatically. This reduces load impact errors and generates more accurate settlement payments. Figure G-2 shows the effect baseline adjustment has on load impact errors.

Figure G-1:
Same-day Adjustment



Assume the contracted reduction for the customer in this example is 3 kW. With an unadjusted baseline of around 5 kW during the event period, OPA would expect this customer to reduce load to around 2 kW during the event. However, the unperturbed actual load of the customer is only about 4 kW. Therefore, this customer needs to reduce load by only 2 kW to achieve full compliance. This equates to a load impact error percentage of 33%. OPA would only receive 66% of the reduction for which it has paid. Same-day adjustments, by bringing the baseline closer to the unperturbed actual load, can dramatically reduce this error and ensure OPA receives a greater share of the expected reduction.

Figure G-2:
Compliance kW & Same-Day Adjustment



Improvements in accuracy from applying same-day adjustments can also be observed by looking at the distribution of load-impact percentage errors. With adjusted loads closer to actual loads for each customer, adjusted baselines should exhibit tighter distributions of errors, and the distributions should be centered closer to zero. Figure G-3 demonstrates this effect. The histogram on the left shows the distribution of errors during the event hours of the average proxy event day using OPA's current baseline without any adjustments. The histogram on the right shows the distribution of load impact errors for the same baseline but with a four-hour same-day adjustment. The distribution of the adjusted baseline is narrower and centered closer to zero. A narrower distribution of baseline errors means greater fairness across DR-3 participants.

A two-hour buffer between the event period and the period used to calculate the same-day adjustment was used for the entire baseline analysis. Participants receive notice two hours prior to an event. That means that if the same-day adjustment period includes the hours immediately before the event, when participants are aware of the event, then they can induce a larger same-day adjustment for themselves by deliberately increasing their usage during that period. This would have the effect of increasing their adjusted baseline during the event, which would make it possible for them to receive credit for full-compliance with a smaller actual load reduction.

Figure G-3:
Distribution of Load Impact Percentage Error & Same-day Adjustment