FREEMAN, SULLIVAN & CO.

A MEMBER OF THE FSC GROUP

2011 Ex Post and Ex Ante Load Impact Evaluation of San Diego Gas & Electric Company's Summer Saver Program SDG0262

June 1, 2012

Prepared for:
San Diego Gas & Electric Co.

Prepared by:
Peter Malaspina
Michael Perry
Freeman, Sullivan & Co.

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

The FSC Group

Table of Contents

1 Executive Summary

San Diego Gas and Electric Company's (SDG&E) Summer Saver program is a demand response resource based on central air conditioner (CAC) load control. It is implemented through an agreement between SDG&E and Comverge Inc, and is currently scheduled to continue through 2016. This report provides ex post load impact estimates for the Summer Saver program for 2011 and ex ante load impact forecasts for 2012 through 2022.

The Summer Saver program is available to residential customers and commercial facilities with average monthly peak demand up to a maximum of 100 kW over a 12-month period. The Summer Saver season runs from May 1 through October 31 and does not notify participating customers of an event. A Summer Saver event may be triggered if warranted by temperature and system load conditions.

There are four enrollment options each for both residential and commercial customers. Residential customers can choose to be cycled 50% or 100% of the time, and can have cycling occur only on weekdays or on weekends as well. Commercial customers have an option of choosing 30% or 50% cycling, on weekdays only or for seven days a week. The incentive paid for each option varies and is based on the number of CAC tons being controlled at each site.

As of the end of 2011 there were 29,591 premises enrolled in the program, which in aggregate have 152,137 tons of CAC capacity. About 83% of participants were residential customers, who account for 68% of the total tons of cooling that are subject to control under the program. Roughly 53% of residential participants are on the 100% cycling option. Approximately 63% of commercial customers selected the 50% cycling option over the 30% option. Summer Saver enrollment is expected to stay roughly the same for the foreseeable future.

In 2011 the program provided an average of about 18 MW of demand response over six events. Commercial customers provided an average of 3.7 MW, and residential customers provided about 14 MW. Due to weather and seasonal conditions, events in 2011 did not provide nearly the amount of demand response which could be expected under more severe heat. Under 1-in-10 September weather conditions (the hottest conditions currently modeled), it is expected that the program could provide up to 30 MW of demand response on average over a 1-6 PM event.

This is the first Summer Saver evaluation that has been performed using smart meter interval data exclusively. The prevalence of smart meters in the Summer Saver population allows for results to be more representative of the entire Summer Saver population because load data is available for a much greater number of customers. Using smart meter data also reduces the cost of evaluation because they do not require the expensive installation of CAC load loggers. In the future, the implementation of a treatment-control design in conjunction with the use of smart meter data could provide for a highly streamlined evaluation process in which ex post impact estimates are available as soon as the smart meter data becomes available and ex ante estimates become available soon after the end of the summer.

For the future, it is recommended that more data be gathered on the different impacts provided by customers on different cycling strategies. This could best be accomplished using an experimental protocol. The current data suggests that the different cycling options within each customer segment do not provide significantly greater load impacts despite customers on each option having similar overall CAC capacity. If true, this would mean that the annual bill credits paid to participants either over pay for customers on the more severe cycling options or under pay those on the less severe options.

2 Introduction and Program Summary

SDG&E's Summer Saver program is a demand response resource based on CAC load control. It is implemented through an agreement between SDG&E and Alternative Energy Resources (AER), a subsidiary of Comverge Inc,[1] and is currently scheduled to continue through 2016. This report provides ex post load impact estimates for 2011 and ex ante load impact forecasts for 2012 through 2022.

2.1 Program Overview

The Summer Saver program is available to residential customers and commercial facilities with average monthly peak demand up to a maximum of 100 kW over a 12-month period. For both residential and commercial customers enrolled in the program, events may be called between May 1 and October 31. Customers can elect to be eligible for events on weekdays only or on weekdays and weekends. Events must be between 2-hours and 4-hours in duration and cannot be called for more than 40 hours per month or 120 hours per year. Event days cannot include holidays or be called on more than three days in any calendar week.

Summer Saver is classified as a "day-of" demand response program and does not notify participating customers when an event is being called. SDG&E may call an event whenever the utility's electric system supply portfolio reaches resource dispatch equivalence of 15,000 Btu/kWh heat rate, or as utility system conditions warrant. A Summer Saver event may also be triggered as warranted by extreme system conditions, such as: special alerts issued by the California Independent System Operator; SDG&E system emergencies related to grid operations; conditions of high forecasted California spot market prices; or for testing or evaluation purposes.

There are four enrollment options each for residential and commercial customers. Residential customers can choose to be cycled 50% or 100% of the time during an event and can have cycling occur only on weekdays or on both weekdays and weekends. The incentive paid for each option varies; the 50% cycling option pays $11.50/ton of CAC capacity and the 100% cycling option pays $46/ton. The 7-day option pays an extra $10 compared to the weekday-only option. Thus, a residential customer with a 4-ton CAC (which is close to the average) would be paid the following under each option:

- $46 for the summer for the weekday, 50% cycling option;
- $56 for the 7-day, 50% cycling option;
- $184 for the weekday only, 100% cycling option; or
- $194 for the 7-day, 100% cycling option.

Commercial customers have an option of choosing 30% or 50% cycling, on weekdays only or for seven days a week. The incentive payment equals $9/ton for the 30% cycling option and $15/ton for the 50% cycling option. As was true for residential customers, the incremental payment for the 7-day a week option compared with the weekday-only option is $10. The average commercial participant has roughly nine enrolled tons of CAC (although some participants have significantly more). As such,

[1] SDG&E's contract with Comverge Inc was amended in 2007 to reflect that the agreement is thereafter recognized to be between a subsidiary of Comverge Inc, AER, and SDG&E. In this document, the company is referred to as Comverge Inc for convenience.

the incentive payment for the average commercial customer under each enrollment option is as follows:

- $81 for the summer for the weekday, 30% cycling option;
- $91 for the 7-day, 30% cycling option;
- $135 for the weekday only, 50% cycling option; or
- $145 for the 7-day, 50% cycling option.

Enrollment in the Summer Saver program is summarized in Table 2-1. As of November 2011, there are 29,591 customers enrolled in the program, which in aggregate had about 152,137 tons of CAC capacity. About 83% of participants were residential customers who accounted for 68% of the total tons of cooling subject to control under the program. Just over 53% of residential participants were on the 100% cycling option and roughly 63% of commercial customers were on the 50% cycling option. Summer Saver enrollment is expected to remain roughly constant in the immediate future.

Table 2-1: Summer Saver Enrollment, November 2011

Customer Type	Cycling Option	Enrolled Customers	Enrolled Control Devices	Enrolled Tons
Commercial	30%	1,882	4,627	17,447
	50%	3,262	8,134	31,069
	Total	5,144	12,761	48,516
Residential	50%	11,375	13,360	46,456
	100%	13,072	15,961	57,165
	Total	24,447	29,321	103,621
Grand Total		29,591	42,082	152,137

2.2 Ex Post Load Impact Estimates

Six Summer Saver events were called in 2011. The events were each four hours long and began at either 1 PM or 2 PM. Table 2-2 shows the load impacts (averaged across each event hour) for each 2011 event day for residential customers and the ex post impact estimates from 2010 for comparison. In 2011, Summer Saver residential customers delivered an average aggregate load reduction over the six events of 14 MW. Residential impacts ranged from a low of 6 MW on September 9, to a high of 19 MW on September 7 and September 8. A blackout began between 3 and 4 PM on September 8, limiting all load impact estimation for that day to the period 1-3 PM.

Table 2-2: Summer Saver Residential Ex Post Impact Estimates[2]

Year	Date	Impact			Average Temperature[3]	
		Per CAC Unit (kW)	Per Premise (kW)	Aggregate (MW)	Midnight-5 PM	During Event
2010	15-Jul-10	0.43	0.50	12	77	85
	16-Jul-10	0.58	0.67	16	80	88
	17-Aug-10	0.46	0.54	13	77	85
	18-Aug-10	0.58	0.68	17	80	87
	19-Aug-10	0.50	0.58	14	78	85
	23-Aug-10	0.52	0.61	15	77	87
	24-Aug-10	0.53	0.62	15	78	88
	25-Aug-10	0.46	0.54	13	78	85
	27-Sep-10	1.02	1.19	29	87	95
	28-Sep-10	0.52	0.61	15	80	84
	29-Sep-10	0.42	0.49	12	76	82
	Average	0.55	0.64	16	79	86
2011	26-Aug-11	0.34	0.41	10	77	85
	7-Sep-11	0.64	0.77	19	82	90
	8-Sep-11[4]	0.66	0.79	19	81	93
	9-Sep-11	0.20	0.24	6	69	73
	12-Oct-11	0.40	0.49	12	76	93
	13-Oct-11	0.62	0.74	18	78	89
	Average	0.48	0.57	14	78	87

Table 2-3 shows ex post load impact estimates for commercial customers for each 2011 event day and ex post estimates for 2010 events for comparison. Aggregate load impacts varied from a low of 2.1 MW on September 9 to a high of 4.9 MW on September 8. The highest impact for a full event, not interrupted by the blackout, was 4.4 MW on August 26.

[2] Aggregate ex post estimates for 2010 have been revised to reflect two data processing corrections since the report was released. Reported results for 2010 differ from those reported in the 2010 evaluation. See Appendix C for comparison of previously reported values to corrected values.

[3] Average temperatures are calculated as a population weighted average of the temperatures experienced by Summer Saver customers, with temperatures determined by the reading at the customer's nearest weather station.

[4] Ex post estimates for September 8 are only for 1-3 PM, the time before the blackout.

Table 2-3: Summer Saver Commercial Ex Post Impact Estimates[5]

Year	Date	Impact			Average Temperature[6]	
		Per CAC Unit (kW)	Per Premise (kW)	Aggregate (MW)	Midnight-5 PM	Event
2010	15-Jul-10	0.33	0.84	4.4	75	83
	16-Jul-10	0.36	0.93	4.9	77	85
	17-Aug-10	0.32	0.83	4.4	74	82
	18-Aug-10	0.36	0.92	4.9	77	84
	19-Aug-10	0.34	0.87	4.6	76	82
	23-Aug-10	0.32	0.84	4.4	74	84
	24-Aug-10	0.34	0.88	4.7	76	85
	25-Aug-10	0.33	0.85	4.5	75	82
	27-Sep-10	0.47	1.22	6.5	84	92
	28-Sep-10	0.36	0.94	5.0	79	83
	29-Sep-10	0.34	0.88	4.7	76	81
	Average	0.35	0.91	4.8	77	84
2011	26-Aug-11	0.34	0.89	4.4	76	82
	7-Sep-11	0.31	0.79	3.9	81	89
	8-Sep-11[7]	0.38	0.98	4.8	80	91
	9-Sep-11	0.16	0.42	2.1	68	71
	12-Oct-11	0.29	0.75	3.7	75	92
	13-Oct-11	0.26	0.67	3.3	77	86
	Average	0.29	0.75	3.7	76	85

Table 2-4 shows ex post load impact estimates for the whole program for 2011.

[5] Aggregate ex post estimates for 2010 have been revised to reflect two data processing corrections since the report was released. Reported results for 2010 differ from those reported in the 2010 evaluation. See Appendix C for comparison of previously reported values to corrected values.

6 Average temperatures are calculated as a population weighted average of the temperatures experienced by Summer Saver customers, with temperatures determined by the reading at the customer's nearest weather station.

7 Ex post estimates for September 8 are only for 1-3 PM, the time before the blackout.

Table 2-4: Summer Saver Program Ex Post Impact Estimates

Date	Impact			Average	
	Per CAC Unit (kW)	Per Premise (kW)	Aggregate (MW)	Midnight-5 PM	Event
26-Aug-11	0.34	0.49	14.4	77	84
7-Sep-11	0.54	0.77	22.9	82	90
8-Sep-11[8]	0.57	0.81	23.9	81	92
9-Sep-11	0.19	0.27	8.1	69	72
12-Oct-11	0.37	0.53	15.7	76	93
13-Oct-11	0.51	0.72	21.3	78	88
Average	0.42	0.60	17.7	77	86

2.3 Ex Ante Load Impact Estimates

Table 2-5 shows ex ante load impact estimates for residential Summer Saver customers. The values shown are averages over the CPUC Resource Adequacy window of 1-6 PM. Program enrollment is expected to remain stable for the future, so this table applies to the years 2012-2022, under the assumption that the program continues to operate under the same set of rules. The residential Summer Saver program is expected to produce an average of 14 MW of demand response over the course of a 1 PM to 6 PM event on a typical event day in a 1-in-10 weather year. The residential program is expected to produce considerably higher impacts under the much hotter conditions of a 1-in-10 September peak day. Under those conditions, the residential program is expected to produce 24 MW.

[8] Ex post estimates for September 8 are only for 1-3 PM, the time before the blackout.

Table 2-5:
Summer Saver Residential Ex Ante Impact Estimates

Day Type	Per CAC Unit Impact (kW)		Aggregate Impact (MW)	
	Weather Year		Weather Year	
	1-in-10	1-in-2	1-in-10	1-in-2
Typical Event Day	0.48	0.41	14	12
May Monthly Peak	0.39	0.17	11	5
June Monthly Peak	0.34	0.09	10	3
July Monthly Peak	0.52	0.42	15	12
August Monthly Peak	0.48	0.38	14	11
September Monthly Peak	0.83	0.64	24	19
October Monthly Peak	0.47	0.38	14	11

Table 2-6 shows ex ante impact estimates for commercial Summer Saver customers. Again, the values shown are averages over the CPUC Resource Adequacy window of 1-6 PM. Program enrollment for commercial customers is also expected to remain stable for the future, so this table applies to the years 2012-2022, under the assumption that the program continues to operate under the same set of rules. On a typical event day in a 1-in-10 year, the commercial Summer Saver program is expected to produce an average of 5.1 MW of demand response over the course of a 1 PM to 6 PM event. The commercial program is expected to produce considerably higher impacts under the much hotter conditions of a 1-in-10 September peak day. Under those conditions, the commercial program is expected to produce 6.2 MW.

Table 2-6:
Summer Saver Commercial Ex Ante Impact Estimates

Day Type	Per CAC Unit Impact (kW)		Aggregate Impact (MW)	
	Weather Year		Weather Year	
	1-in-10	1-in-2	1-in-10	1-in-2
Typical Event Day	0.40	0.36	5.1	4.6
May Monthly Peak	0.33	0.24	4.3	3.1
June Monthly Peak	0.37	0.24	4.8	3.1
July Monthly Peak	0.39	0.36	5.0	4.7
August Monthly Peak	0.40	0.36	5.1	4.6
September Monthly Peak	0.48	0.42	6.2	5.4
October Monthly Peak	0.34	0.30	4.3	3.9

2.4 Report Structure

The remainder of this report is organized as follows. Section 3 summarizes the data and methodologies that were used to develop the ex post and ex ante load impact estimates and the validation tests that were applied to assess their accuracy. Section 4 contains the ex post load impact estimates, an analysis of control device communication success and an analysis of the distribution of load impacts over customers. Section 5 presents the ex ante estimates. The Appendix contains figures relating to the day-matching strategy used to estimate load impacts.

3 Data and Methodology

This section summarizes the datasets and analysis methods that were used to estimate load impacts for each event in 2011 and for ex ante weather conditions. The choice of ex post model has important implications for ex ante modeling, which means that ex ante modeling is often referred to even though ex ante results are not included in this report. A separate report including ex ante results will be provided in a report to follow. Results from a variety of validation tests are also presented.

3.1 Data

In 2011, six Summer Saver events were called. Table 3-1 shows the date of each event, and the start and stop time of each event. All residential and commercial accounts were called for each event. All events lasted four hours and began at either at 1 PM or 2 PM.

Table 3-1: Summer Saver 2011 Event Summary

Date	Start Time	End Time
8/26/2011	2:00 PM	6:00 PM
9/7/2011	2:00 PM	6:00 PM
9/8/2011	1:00 PM	5:00 PM
9/9/2011	2:00 PM	6:00 PM
10/12/2011	1:00 PM	5:00 PM
10/13/2011	1:00 PM	5:00 PM

SDG&E provided FSC with samples of smart meter interval data for both the residential and commercial populations for the summer of 2011. The sample included data for 762 residential premises and 3,555 commercial premises. The commercial sample encompassed the entire commercial Summer Saver population for which smart meter interval data is available.[9] This is the first time the Summer Saver evaluation is being performed using only smart meter interval data; previous evaluations have relied on CAC logger data. However, in evaluations of the 2009 and 2010 program years, analyses of residential load impacts performed using smart meter interval data produced load impact estimates very close to those estimated using CAC logger data. While these analyses were not performed for commercial customers, FSC does not believe that repeating the same process for the 2011 program year would not also produce similar results as those found using CAC logger data. Additionally, FSC has extensive experience using smart meter data to estimate load impacts for CAC load control programs for other utilities; this method has always been found to produce impact estimates as accurate as those estimated based on CAC logger data.

Tables 3-2 and 3-3 show the distribution of CAC tonnage by cycling option and climate zone for the populations and samples of commercial and residential customers, respectively, as of June, 2011. As the tables show, each sample is representative of the population of participants. The differences between the fraction of customers in each sample cell and each population cell are small; there are effectively no differences across climate zones, while small differences exist across cycling options.

[9] The exact number of premises with data for analysis varied on a day-by-day basis, due to limitations on interval data availability.

Final results are weighted based on cycling option to reflect these slight differences between the sample and the population.

Table 3-2: Distribution of AC Tonnage by Program Option and Climate Zone Residential Population

Cycling and Weekday Options	Group	Climate Zone 1	Climate Zone 2	Climate Zone 4	Total
50%	Population	3%	1%	42%	46%
	Sample	3%	1%	46%	50%
100%	Population	11%	1%	43%	54%
	Sample	11%	1%	39%	50%
Total	**Population**	14%	2%	85%	100%
	Sample	14%	2%	84%	100%

Table 3-3: Distribution of AC Tonnage by Program Option and Climate Zone Commercial Population

Cycling and Weekday Options	Group	Climate Zone 1	Climate Zone 2	Climate Zone 4	Total
30%	Population	14%	0%	23%	37%
	Sample	14%	0%	23%	38%
50%	Population	32%	0%	31%	63%
	Sample	31%	0%	31%	62%
Total	**Population**	**45%**	**1%**	**54%**	**100%**
	Sample	**45%**	**0%**	**55%**	**100%**

3.2 Methodology

The primary task in estimating ex post event impacts is to estimate a reference load for each event. The reference load is a measure of what demand would have been in absence of the demand response event. Although this report focuses on ex post estimation, the ultimate goal of the broader evaluation is to develop both ex post and ex ante load impact estimates. Therefore, ex ante methods are discussed where relevant. The primary task in estimating ex ante event impacts (which are often of more practical concern) is to make the best use of available data on loads and load impacts to predict future program performance. The data and models used to estimate ex post impacts are typically major elements of the ex ante analysis.

The primary source of information used in both the 2009 and 2010 evaluations of Summer Saver for reference load was load observed during non-event times. This was significantly aided by the experimental design put in place for settling the demand response contract with Comverge Inc. Under this contract, a stratified, random load research sample of residential and commercial Summer Saver customers was created. During each event, half of the load research sample would be held back to provide reference load (*i.e.* those CAC units would not be controlled during the event). Individual customer regressions performed well under these conditions because any given customer in the sample had several event periods during which their load could act as reference load because it was not curtailed. Moreover, even if particular events were unique from all other event days (such as September 27, 2010, which was the hottest day of 2010 and the all-time SDG&E system peak), load from one half of the sample could be used to estimate the reference load for the other half in a treatment-control analysis rather than individual customer regressions.

As compared to the two previous program years, the events in 2011 were more complicated to model because several of the event days had unique characteristics and because the experimental design for settlement with Comverge Inc was corrupted. These complications and the modeling decisions that resulted are discussed in Appendix A. The result was that residential ex post impact estimates were developed using individual customer regressions, while commercial ex post impact estimates were developed using a day-matching approach. Each is described below.

3.2.1 Customer Regression Models for Residential Customers

Each customer has a different usage pattern over time, and each customer's usage is likely to respond differently to changes in weather. For this reason, separate regressions were estimated for each premise in the residential sample,[10] but using a common regression specification over all cases. For all premises, the factors used to estimate usage patterns were weather variables interacted with time indicators. These allow the model to take into account different reactions to weather conditions at different times of day, times of week and times of year. For example, a residential customer's energy usage might respond strongly to high temperatures on a Saturday afternoon when they are at home, but it might not respond at all on a Wednesday afternoon when they are at work.

Only non-holiday weekdays were modeled because no events were called on either weekends or holidays, and weekend usage behavior is quite different from weekday usage. Table 3-4 defines the variables and describes the effects they seek to identify. The regression specification was:

$$kWh = a + \sum_{i=1}^{24}\sum_{j=5}^{10} b_{ij} \times lagcdh_i \times hour_i \times month_j + \sum_{i=1}^{24} c_i \times hour_i + \sum_{i=1}^{24} d_i \times wacdh_i \times hour_i$$
$$+ \sum_{i=14}^{20} e_i \times wacdh_i \times earlyevent_i \times hour_i + \sum_{i=15}^{21} f_i \times wacdh_i \times lateevent_i \times hour_i + \varepsilon$$

[10] As discussed in Appendix A, this regression specification was also estimated for commercial units but the results were not ultimately the ones chosen.

Table 3-4: Description of AC Load Regression Variables

Variable	Description
a	Estimated constant
$b-f$	Estimated parameter coefficients
$hour$	Indicator variables representing the hours of the day, designed to estimate the effect of daily schedule on usage behavior and event impacts
$month$	Indicator variable for the month
$earlyevent$	Indicator variable to model the hourly effects of events occurring during 1 PM - 5 PM
$lateevent$	Indicator variable to model the hourly effects of events occurring during 2 PM - 6 PM
$lagcdh$	Weighted average of the previous 24 hours of cooling-degree hours with a base of 70°F
$wacdh$	Weighted average of the previous 3 hours of cooling-degree hours with a base of 75°F. Captures shorter-term effects of high temperatures.
ε	Error term

The conceptual basis for statistical analysis is that with large sample sizes, the effect of unobservable or omitted factors not related to the main effect will disappear due to the power of averaging. Presumably, many factors affect an individual customer's usage other than what can be included in a large-scale model. In a large sample, such as hundreds of customers over three months, it is likely that the effect of these omitted factors is small. However, in smaller samples, such as one or a few customers' regression models, these omitted factors could have an important effect. This means that results for sub-samples of the dataset should be viewed with increasing caution as the sub-samples decrease in size.

A related issue is that any measure of event-impact standard error associated with these individual customer regressions inherently assumes that the model has been fully and correctly specified so that the only remaining unexplained variation is completely random – meaning that it is unrelated to any variables of interest. As noted, this may be untrue at an individual customer level. Moreover, statistical variation can only be calculated based on the observed events during the study period. This means that it cannot take into account the effect of weather patterns or other recurring behavior patterns that are not well-represented in the dataset, but are likely to arise in the future. When the statistical model is asked to provide an extrapolation, there is no procedure for adjusting its uncertainty estimate upward because it is an extrapolation. Both of these issues probably lead to an under-estimation of the true level of variance that should be expected in Summer Saver results – even assuming no operational changes or changes in underlying customer behavior. The degree of this under-estimation is unknown because there is no data to model it.

Given that caveat, standard errors for load impacts are calculated as:

$$se = \sqrt{(stdp^2 + rmse^2)},$$

Where *stdp* is the standard deviation of the prediction, *i.e.*, the standard error associated with the fact that all coefficients are estimated values, and *rmse* is the root-mean-squared-error of the regression, or the error associated with the fact that the model has a baseline of uncertainty in it even if coefficients are estimated perfectly. The *stdp* value is calculated independently for each hourly prediction of each customer's load.

Having calculated the standard error for each hour for each customer, aggregate standard errors are calculated assuming that errors are independent across customers. Therefore, variances can be summed to get aggregate variance.

Having calculated standard errors of predicted load impact, percentiles of load impact are calculated based on a Gaussian (Normal) distribution with standard deviation equal to the calculated standard error and mean equal to the estimated load impact. This calculation is justified by the central limit theorem.

3.2.2 Residential Regression Model Validation

In order for a model to be useful in the context of Summer Saver, it must make accurate predictions of CAC loads, primarily at high temperatures. Three methods of validation are used to assess this capability: in-sample testing, out-of-sample testing and evaluation of general plausibility of predictions.

In-sample Testing

At an individual level and at an aggregate level, the model must explain a large degree of the observed variation in household load during the summer of 2011. This is a test of the in-sample R-squared of the model, which is the simplest test for the model to pass and is a necessary, but not sufficient, condition for the model to be useful. A substantial body of evidence from previous evaluations by FSC and others demonstrates that weather and time variables in a regression model can explain a large amount of the variation in CAC load. Therefore, a model without an aggregate R-squared value of at least 70% would suggest a significant error and would bear significant investigation before being accepted.

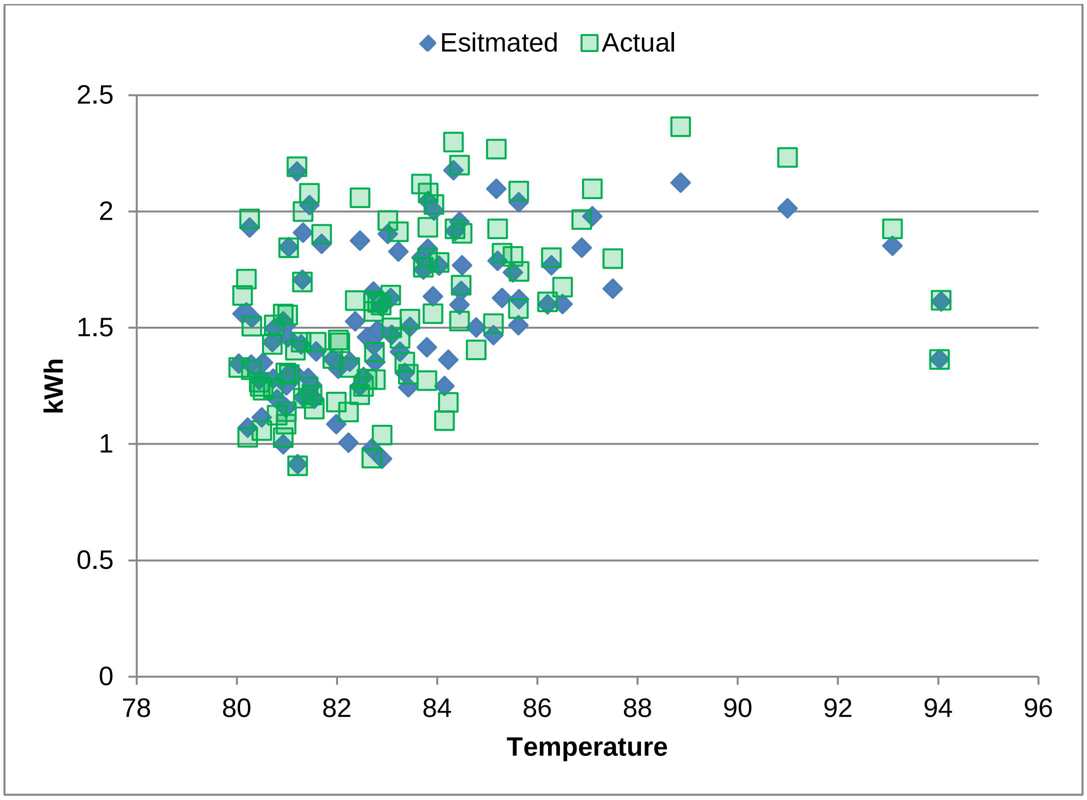
The R-squared of a model can be inflated by including a very large number of variables. In this case, the model will appear to explain a large degree of the variation in load, but it may be highly inaccurate in predicting for conditions outside of the range of values for the data used to estimate the model. This is known as over-fitting. Diagnosing whether a model is over-fit inherently requires judgment. There are several metrics, such as adjusted R-squared, that attempt to penalize models for including many variables, but they are all based on arbitrary weightings of the number of variables as compared to the fit of the model. The method used here to guard against over-fitting is out-of-sample testing, as described below. An over-fit model will not produce accurate out-of-sample predictions.

Although the regressions were performed at the individual premise level, from an evaluation standpoint the focus is less on how the regressions perform for individual premises than on how they perform for the aggregated sample. Therefore, the R-squared (goodness-of-fit) statistic is presented for both the individual regressions and for the aggregate load: the average R-squared among individual residential households is 43% and at an aggregate level the residential R-squared is 87%.

Summer Saver events are only likely to be called at times of very high temperature. Therefore the models must accurately fit load at high temperatures in particular. Figures 3-1 and 3-2 show that the residential models do fit load accurately for the high-temperature periods during the summer of 2011.

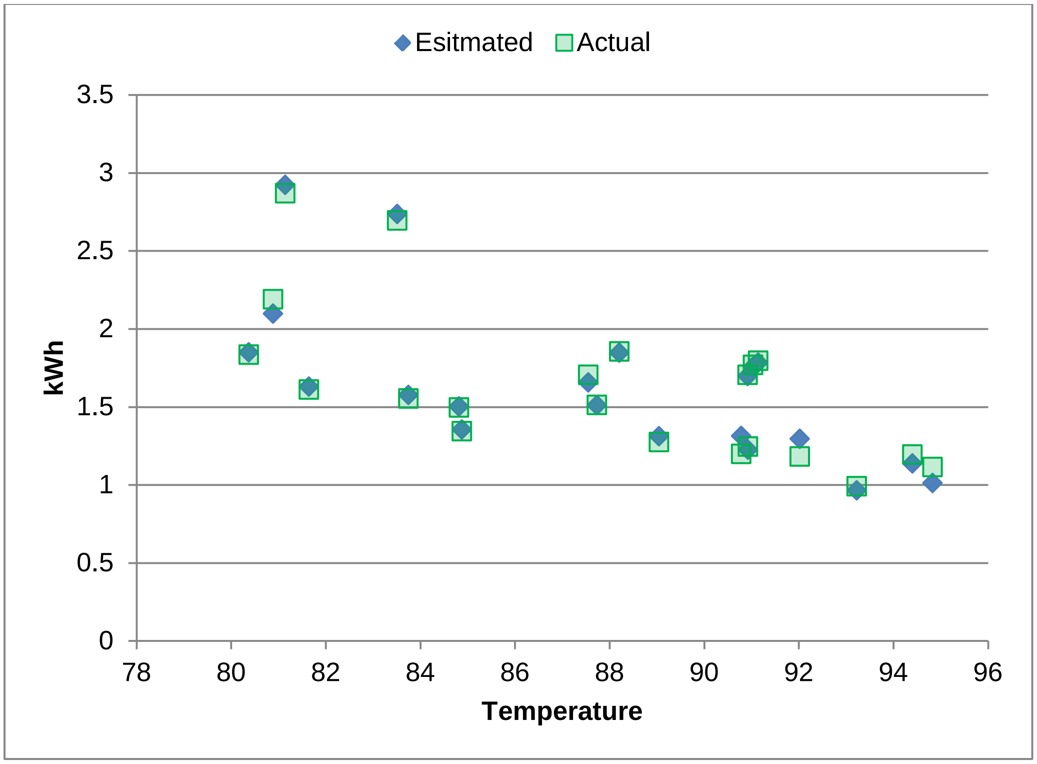
Figure 3-1 shows the average actual hourly load in the residential sample and the predicted hourly load for afternoon non-event hours between 1 PM and 6 PM when the temperature exceeds 80°F. Bias in these figures would show itself as a persistent difference between actual and predicted values in one direction. For example, if the actual values strongly tended to be above the predicted values, then that would indicate that the model under-predicted load at high temperatures. There is little systematic difference between the predicted and actual loads as shown in the figure. On average, residential predicted loads exceed the actual loads by 2%.

Figure 3-1: Actual and Predicted Average Residential Load for 1 PM to 6 PM, Non-event Days When the Temperature Exceeds 80°F



In addition to checking how well the model predicts load at non-event times, it is also important to verify that the model predicts load well during event periods. Figure 3-2 shows the predicted versus actual values during the 2011 events when the temperature exceeds 80°F. This includes all 2011 event hours except those on September 9. For residential households, the actual load exceeds the predicted load by less than 1%.

Figure 3-2: Actual and Predicted Average Residential Load for Event Hours When the Temperature Exceeds 80°F



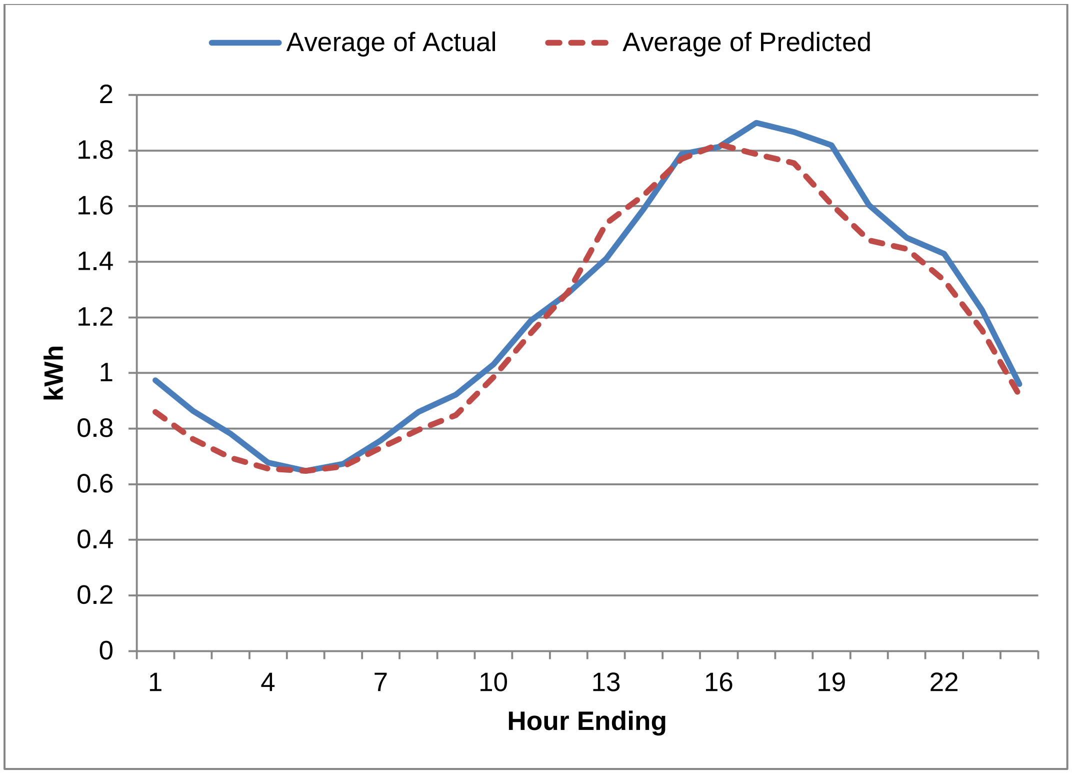
These figures do not necessarily indicate that the model is good at predicting in the ex ante application because these values are predictions for conditions used to fit the model. Instead, these figures show that there is only a small amount of variation in the existing data that the model does not account for at the higher temperature levels.

Out-of-sample Testing

As a second and more stringent test, the model must do well in out-of-sample testing on days included in the 2011 dataset. The procedure for out-of-sample testing consists of re-estimating the model while holding back some of the hot non-event days of the summer from the estimation. Predicted loads were then compared to the actual loads on the days held back. This is a true test of the regression model's predictive power for weather conditions actually observed during the summer of 2011.

Figure 3-3 shows the actual average hourly energy use of residential households for the out-of-sample days, July 6, July 8, and September 6. The close match between predicted values and actual values reflects the ability of the regressions to predict accurately. For residential customers, the average absolute difference between predicted and actual load is approximately 3% during the hours of 1 PM to 6 PM.

Figure 3-3: Average Residential Whole-Building Actual and Predicted Load for Out-of-sample Days



The final test of the model is one of general plausibility in predicting loads during the event periods and for the ex ante weather conditions. This test is less well-specified but consists of producing reasonable household load patterns as a function of weather as compared to results in past years, results from other programs and general knowledge about how the program works. This reality-check test is a crucial way to test the assumptions that go into the model. The ex ante estimates that will be presented in a future report were carefully reviewed and generally display the expected patterns across event conditions and are consistent with other studies after judgmentally accounting for expected differences due to weather conditions and other factors.

3.2.3 Day-matching for Commercial Customers

As noted above, complications arose due to the unique nature of the 2011 event days which led to the use of a day-matching method to produce commercial ex post impact estimates. Under this method, each event day was matched with a non-event day that appeared to provide an accurate reference load based on pre-event, event-period and post-event loads. The underlying concept is that even after accounting for the effects of weather, loads remain highly correlated throughout the day. Observing that loads on an event day and non-event day are very close in the hours before an event and after an event is strongly suggestive that loads during the event would have been similar had the event not occurred. [11]

[11] This is a theoretical argument for using a time-series analysis. However, the data requirements for such an analysis are stringent, the models are much more time consuming to fit and validate, and effectively communicating the methods and

With this conceptual framework in mind, a day's load had to satisfy three basic criteria to be judged to be suitable as a reference load for an event day:

- The event day average loads during the three hours before the event had to be at least as close to the average loads on the reference day during the same hours as they were to the average loads during those hours on any other non-event weekday. In other words, there was no day with pre-event average loads closer to those on the event day than the reference day chosen;
- The event day loads during the event hours had to be below the loads on the reference day during the same hours; and
- The event day loads during the three hours immediately after the event had to be near to or higher than the loads on the reference day during the same hours.

September 7 and August 26 had such high loads that no non-event day had loads that satisfied all the criteria. This was also true for using day-matching to model the impacts of the first two hours of the event on September 8, which was interrupted by the blackout. For these cases, the non-event day with the highest load was chosen and a same-day adjustment was applied. A same-day adjustment is a way to account for known biases in a reference load. In this case, the fact that that load in the hour immediately before the event is much higher than the highest available reference day load indicates the high likelihood of a downward bias in the reference load during the event. To partially correct this bias, the reference load is adjusted by adding to it the difference between event day load and the reference day load during the hour immediately before the event. This adjustment is calculated separately for each cycling option of each customer segment and applied to the day-matching reference load for each event day.

Table 3-5 shows the days that were chosen to provide reference load for each ex post event day. Appendix B shows graphs of the load shapes and adjusted load shapes for each event day load and reference day load.

Table 3-5: Event Days and Matched Reference Load Days for Commercial Customers

Event Day	Matched Days
26-Aug-11	2-Aug-11
7-Sep-11	2-Aug-11
8-Sep-11	2-Aug-11
9-Sep-11	7-Jul-11
12-Oct-11	6-Sep-11
13-Oct-11	25-Aug-11

Based on the figures in Appendix B, the day-matching reference loads for commercial customers appear quite plausible.

results of such a departure from standard load impact evaluation methodologies would be challenging. Additionally, they are of limited to no use in ex ante estimation. For these reasons this simplified approach to addressing autocorrelation is preferred.

Having identified matched days, load impacts for each cycling option within each customer segment were estimated by subtracting average hourly load during each event from average hourly load during the same hours of the matched reference day. Standard errors were calculated at an hourly level as the square root of the sum of squared standard errors of each hourly average load.

3.3 Ex Ante Impact Estimation Methodology

Just as 2011 presented different modeling issues than in previous years for ex post estimation, it also had unique characteristics for ex ante modeling. In the previous two evaluations of this program, ex ante load impact estimates were developed using the same individual customer regressions that were used for ex post. For reasons discussed in Appendix A, ex post estimates from individual customer regressions deserve extra scrutiny this year. Additionally, strictly from the standpoint of ex ante prediction, 2011 was not ideal as a sole source of information for future prediction. Predicting impacts for ex ante weather conditions relies on observing system operation multiple times and under a variety of conditions. For accurate predictions, a larger number of events occurring under different weather conditions are better. Only six events were called in 2011, and several of them are not highly useful for predicting typical system performance.

First, the two October event days had impacts that are lower than what would be expected during similar conditions earlier in the season, or even during October if there was a more extended period of heat.

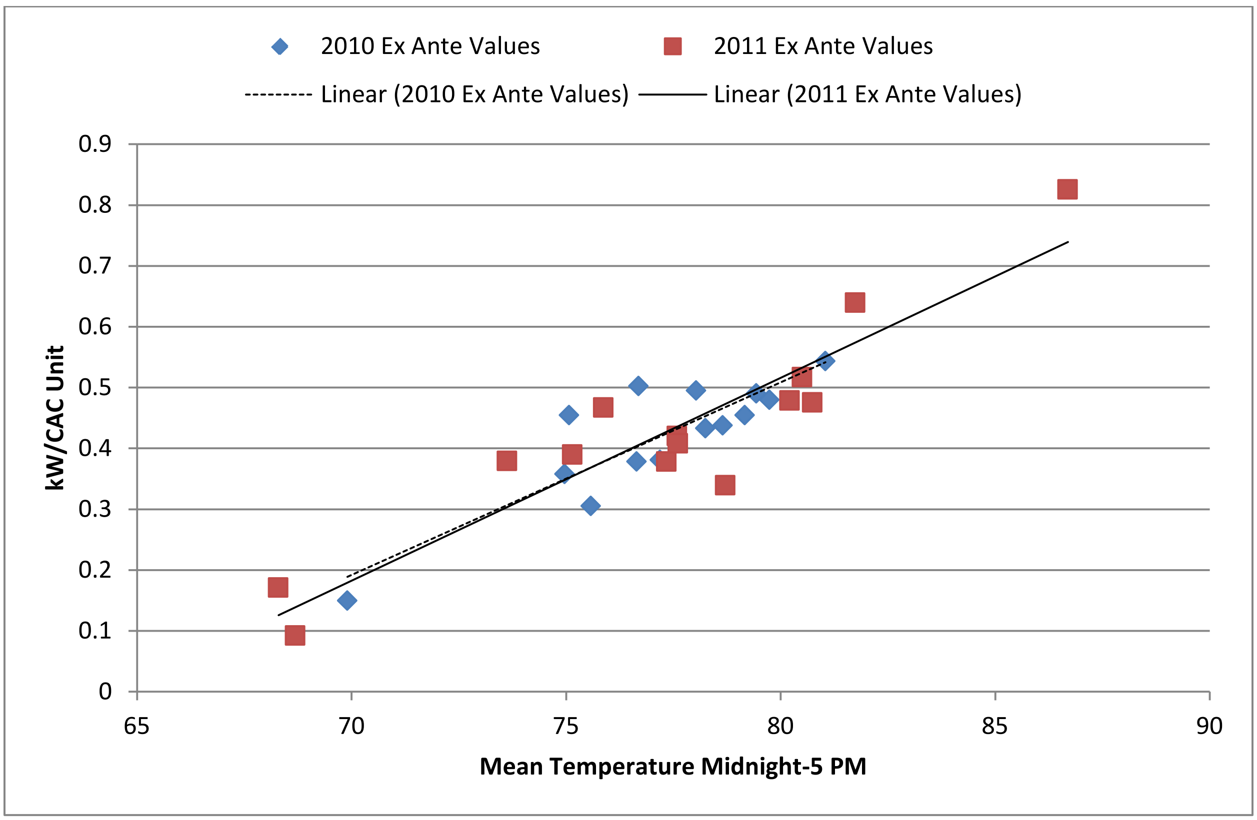
In addition, on one of the event days, September 8, a blackout affected the entire SDG&E territory. This blackout happened in the middle of the event, leaving only two hours in which loads are representative of normal event loads; this limits the usefulness of this event for predictive purposes. The day after the blackout, September 9, was also an event day. This event day provides no particular challenges for ex post modeling; however the temperatures on September 9 are much lower than normal for a Summer Saver event. For this reason, the observed load impact on that day provides little useful input into a predictive ex ante model.

Finally, the day before the blackout, September 7, was the system peak day for 2011 and Summer Saver customers had pre-event loads on that day higher than loads during similar hours on any non-event days. This means that any reference load model for that day will be an extrapolation from loads under observed weather conditions to loads under un-observed weather conditions. Another event day, August 26, presented the same issue for commercial customers but not for residential customers.

For these reasons, it is worth careful consideration whether ex ante results based only on 2011 events are reliable. An alternative option is to use both 2010 and 2011 results to develop ex ante impact estimates. This is reasonable because the program has changed little in the past year, the population of participants has been stable, and in such a short time it is unlikely that underlying customer behavior or CAC operation has changed significantly. Additionally, there were 11 events in 2010 that occurred under conditions more similar to those likely to arise in the future than the 2011 events. The latest event occurred in late September, with the bulk of events occurring in mid-July through early September. The temperatures during events were mainly in the mid-to-high 80s, which is typical of summer weather in San Diego and is also similar to most of the conditions used for ex ante prediction.

For residential customers it makes little difference which strategy is used. Ex ante predictions developed using 2011 data alone are very similar to those developed in 2010. Therefore combining the years' data has no appreciable effect on the ex ante impact estimates. This is shown in Figure 3-7 below. The figure shows point estimates of load impact per CAC unit for each set of ex ante weather conditions as a function of the average temperature from midnight to 5 PM under those conditions. This summary measure of temperature is highly predictive of load impacts. The linear trend-lines between the two sets of data points are practically indistinguishable, which indicates that as a practical matter, predictions using either year's data are essentially identical.[12] Both the average values and the way the predictions vary with temperature are the same. For this reason, ex ante estimates developed using the individual customer regression model are used.

Figure 3-7:
Residential Ex Ante Predictions Developed Using 2011 Data Alone and those Developed in 2010



For commercial customers, due to the variability in impacts induced by the unique 2011 modeling issues and the fairly small number of events, impact estimates developed using both years' data together are likely to be more reliable.

[12] The set of ex ante weather conditions used for modeling changed between the two years. This has no material effect on the figure other than to provide two sets of conditions, with much hotter temperatures in the 2011 ex ante conditions that were not there in the 2010 set of conditions.

Load impact estimates in 2010 were developed using a different sample of customers and a different source of data (CAC load data rather than whole building), which means that combining data from the two years cannot be done by simply developing one large set of individual customer regressions. Instead, a predictive model of load impacts was developed using ex post impact values as a dependent variable.

To determine the best regression to use for ex ante predictions, dozens of models predicting ex post impacts based on different measures of recent temperature were tested. The final regression only includes one explanatory variable because there are only a total of 16 events over the two years – 11 in 2010 and 5 in 2011 (September 8 is excluded due to the blackout). Using more explanatory variables might result in model over-fitting. The model that best predicted average ex post impacts:

$$Impact_c = a + b \cdot mean17_{\mathbf{c}} + \boldsymbol{\varepsilon}_c$$

Table 3-8:
Commercial Ex Ante Regression Variables

Variable	Description
$Impact_c$	Average per CAC unit ex post load impact for each event day from 3 to 5 PM
a	Estimated constant
b	Estimated parameter coefficient
$mean17$	Average temperature over the 17 hours prior to the start of the event
$\boldsymbol{\varepsilon}_t$	The error term, assumed to be a mean zero and uncorrelated with any of the independent variables

The average temperature over the previous 17 hours was chosen as the weather variable for modeling because of its predictive ability and the fact that ex ante impact prediction uses only one day's worth of temperature data for each set of conditions. A model using the average of the previous 24 hours of temperature performed similarly in prediction, but would require additional assumptions about weather in the day prior to each ex ante day. Using the previous 17 hours made full use of the available ex ante weather information without requiring additional assumptions and without sacrificing model accuracy. Models using temperature as far back as 48 hours prior to the event were tested, but were not found to perform better than the model using 17 hours.

It is quite likely that event impacts depend on variables other than this average of recent temperatures, but with 16 points for modeling it is not possible to accurately identify these effects. Ideally, in future years, more data may allow for modeling of impacts using more variables.

The model was estimated separately for customers on the 30% and 50% cycling options.

The regressions were weighted by the inverse of the estimated sampling variance of each ex post data point. The day-matching based estimates of 2011 had significantly larger sampling variances than the regression-based estimates of 2010. Additionally, there were only 5 events that could be used for modeling from 2011, versus 11 from 2010. These two factors meant that the 2010 results had a much larger influence on the ex ante results. Given the noted issues in the 2011 results, this is appropriate.

The dependent variable in each regression was the ex post load impact measured for the window 3 to 5 PM. This variable was chosen because all events covered the hours 3 to 5 PM, and that window does not contain the first hour of any event, which typically has lower impacts due to the gradual event start. Therefore, this dependent variable is a comparable measure of event impact for each test event day and does not introduce confounding factors such as different customer load shapes at different times of day. For example, it would not be as accurate to model total average event impacts using this regression because some events went from 1 to 5 PM, while others went from 2 to 6 PM.

The last step in estimating load impacts was to translate average impacts from 3 to 5 PM to hourly impacts over the entire range of time required for prediction, 1 to 6 PM. Hourly ex post impact estimates for each event in 2010 and 2011 were expressed as a fraction of the average impact from 3 to 5 PM. Table 3-9 gives an example of this process. The first column of Table 3-9 shows how the average event impact for each hour of the five hour events compares to the average impact from 3 to 5 PM. To illustrate, the second column shows the proportions in the first column multiplied by 0.29 kW, the average predicted impact from 3 to 5 PM for commercial customers during a typical event day during a 1-in-2 weather year. To calculate the estimated impact for 1 to 2 PM, for example, 0.50 kW was multiplied by 52% to yield an impact of 0.26 kW. The same strategy is applied for all five hours of the event, as illustrated below in Table 3-9.

Table 3-9:
Hourly Impact Compared to Average Impact from 3–5 PM

Hour of Event	Hourly Impact/ Average 3–5 PM Impact (%)	Hourly Impact for Typical Event Day, 1-in-2 Weather (kW)
1-2 PM	92	0.34
2-3 PM	104	0.38
3-4 PM	99	0.36
4-5 PM	101	0.37
5-6 PM	97	0.35

This method constrains the relative size of event impacts across different hours to be the same for each event. Event impacts vary with weather, as usual, but in this model the ratio of the impact at 4 PM to the impact at 5 PM, for example, is always the same. A separate ex ante model could be used for each event hour separately. Such a strategy would have the virtue of independently identifying the effect of weather on event impacts at different times of day. That is not done here because there are not enough data points per hour to meaningfully identify differences in the effect of temperature on event impact at different hours. Given the highly auto-correlated nature of the data, the differential impact of weather on different event hours is likely to be difficult to measure as compared to the primary effect of temperature on average event impact. This might be a worthwhile effort after several years of data collection or if the data started implying that such effects were more important than they currently appear.

4 Ex Post Load Impact Estimates

This section contains the ex post load impact estimates for program year 2011. Residential estimates are provided first, followed by commercial estimates. The section also contains an analysis on control device communication failure and an analysis of the distribution of impacts across customers.

4.1 Residential Ex Post Load Impact Estimates

Table 4-1 shows the ex post load impact estimates for residential Summer Saver customers for 2011. Summer Saver residential customers delivered an average aggregate load reduction over the six events of 14 MW. Residential impacts ranged from a low of 6 MW on September 9, to a high of 19 MW on September 7 and September 8. Due to the modeling issues discussed in Appendix A, these results contain a higher than usual level of uncertainty, but they provide no evidence that program performance in 2011 deviated significantly from 2010.

Table 4-1: Residential Ex Post Load Impact Estimates

Date	Impact			Temperature	
	Per CAC Unit (kW)	Per Premise (kW)	Aggregate (MW)	Midnight-5 PM	During Event
26-Aug-11	0.34	0.41	10	77	85
7-Sep-11	0.64	0.77	19	82	90
8-Sep-11[13]	0.66	0.79	19	81	93
9-Sep-11	0.20	0.24	6	69	73
12-Oct-11	0.40	0.49	12	76	93
13-Oct-11	0.62	0.74	18	78	89
Average	0.48	0.57	14	78	87

4.2 Commercial Ex Post Load Impact Results

Table 4-2 shows the ex post load impact estimates for commercial Summer Saver customers for 2011. Summer Saver commercial customers delivered an average aggregate load reduction over the six events of 3.7 MW. Commercial impacts ranged from a low of 2.1 MW on September 9, to a high of 4.9 MW on September 8. The highest average impact for a full event, unaffected by the blackout, was 4.4 MW on August 26. Again, these results contain a higher than usual level of uncertainty, but they provide no evidence that program performance in 2011 deviated significantly from 2010.

[13] Results only include the first two hours of the event. The second two hours were affected by the blackout.

Table 4-2: Commercial Ex Post Load Impact Estimates

Date	Impact			Average Temperature	
	Per CAC Unit (kW)	Per Premise (kW)	Aggregate (MW)	Midnight-5 PM	Event
26-Aug-11	0.34	0.89	4.4	76	82
7-Sep-11	0.31	0.79	3.9	81	89
8-Sep-11[14]	0.38	0.98	4.8	80	91
9-Sep-11	0.16	0.42	2.1	68	71
12-Oct-11	0.29	0.75	3.7	75	92
13-Oct-11	0.26	0.67	3.3	77	86
Average	0.29	0.75	3.7	76	85

4.3 Load Impacts by Cycling Option

Table 4-3 shows load impacts per CAC unit and in aggregate by cycling option for residential and commercial customers. Within each segment, the average impact per unit is very close. This suggests a selection bias on the part of customers, with those who are more likely to have large CAC loads being more likely to choose the less intensive option. This selection bias has been noted in previous evaluations, although its effect is particularly stark here. Direct measurement of CAC load was only taken for a small sample of customers for contract settlement, so it is not possible to determine whether load impacts as a percentage of CAC load are significantly greater for the higher cycling options. It is worth noting that for residential customers, whole-building reference loads are significantly higher for customers on the 50% cycling option. Residential customers on the 50% option cycling had average whole-building reference loads of 2.24 kW over all six events in 2011, whereas those on 100% cycling had reference loads of 1.66 kW. This is despite the fact that those on 100% cycling have slightly higher CAC tons per premise.

For commercial customers, those on 50% cycling tend to have much lower whole-building loads, but this is less informative than for residential customers. CAC load is typically a large percentage of whole-building loads for residential customers, while for commercial customers this is less consistently true.

[14] Results only include the first two hours of the event. The second two hours were affected by the blackout.

Table 4-3: Per CAC Unit Load Reductions by Cycling Option (kW)

Date	Per CAC (kW)				Aggregate (MW)			
	Cycling Option				Cycling Option			
	Residential		Commercial		Residential		Commercial	
	100	50	50	30	100	50	50	30
26-Aug-11	0.37	0.31	0.34	0.35	5.8	4.3	2.8	1.6
7-Sep-11	0.67	0.62	0.31	0.30	10.5	8.5	2.5	1.4
8-Sep-11	0.64	0.67	0.41	0.32	10.0	9.2	3.3	1.5
9-Sep-11	0.2	0.20	0.18	0.13	3.1	2.7	1.5	0.6
12-Oct-11	0.41	0.40	0.27	0.33	6.5	5.4	2.2	1.5
13-Oct-11	0.61	0.63	0.27	0.24	9.5	8.7	2.2	1.1
Average	0.48	0.47	0.30	0.28	7.6	6.3	2.4	1.3

In light of these findings, and the fact that the residential 100% cycling group is paid four times as much to participate as the 50% cycling group, it may be possible to improve program cost effectiveness by increasing the share of program participants on the lower cost 50% cycling option and/or by reducing the incentive paid for 100% cycling while increasing the incentive paid for 50% cycling. The same may be true for the commercial cycling options.

4.4 Control Device Communications Failure

The load-control switches that trigger events to happen at the customer level rely on radio signals for event activation. If the switch is broken, if the signal is blocked or if the signal is sent on a frequency that the device is not set up to receive, then the event will not occur for that device. This is referred to as control device communication failure.

Direct measurement of control device communication was not done for the 2011 evaluation. However, a load research sample of CAC load was collected for the sake of contract settlement with Comverge Inc. This sample contained 177 customers on the 100% cycling option. Customers on 100% cycling that do not have event load reductions of very close to 100% can be presumed to be affected by communication failure. Also, there is no obvious reason why customers on 100% cycling should have different communication failure rates from residential customers on other cycling options, so this analysis probably reflects communication across the residential Summer Saver population. Commercial Summer Saver customers may have different rates of communication failure due to differing building types and switch locations.

As shown in Table 4-4, an analysis of the number of customers in the 100% cycling group that had load above 0.02 kW during each event hour of 2011 revealed that communication failure was variable, but tended to be about 15% during the middle hours of most events. The higher percentage of non-zero loads in the first hour can be attributed to the fact that for each customer, events actually begin sometime in the first half-hour of the event, rather than immediately at the top of the hour. It should be noted that the samples underlying the values for the two October events are smaller, with data from only 65 customers used to calculate the failure rate for October 12 and only 50 customers used to calculate the failure rate for October 13.

Table 4-4: Percentage of Premises on 100% Cycling with Non-zero[15] Load during Each Event Hour

Event Date	Event Hour			
	1	2	3	4
26-Aug	33%	9%	13%	14%
7-Sep	43%	15%	16%	18%
8-Sep	33%	13%	NA[16]	NA[17]
9-Sep	16%	10%	9%	9%
12-Oct	17%	12%	12%	13%
13-Oct	31%	33%	37%	38%
Average	29%	15%	16%	16%

Communications failure did not affect the same customers for each event; only 3% of sampled customers showed failure for all of the events for which they were called. Almost 13% of sampled customers showed failure for more than 50% of the event hours for which they were called, and 49% showed failure for more than 10% of their event hours.

The overall distribution of control device communication failure in this sample, including the average level of failure is quite similar to what was observed in 2010.

4.5 The Distribution of Impacts across Customers

In previous evaluations, the distribution of event impacts across customers was estimated based on the distribution of average estimates from individual customer regressions. Recent internal analysis has shown that this method contains too much noise to be useful as an indicator of the real distribution of event impacts at the customer level.

As an alternative, Table 4-5 shows estimated event impacts for customers segmented into deciles of average load on hot, non-event days. In this procedure, each customer was placed into a decile category based on their average usage during the hours 12-6 PM on the days used for day-matching (listed in Table A-1 in Appendix A). Impact estimates were calculated separately for each decile using day-matching plus a same-day adjustment, with reference loads provided by the days listed in Table A-1. The same-day adjustment procedure was applied in the same manner as the adjustment used to produce the primary impact estimates for commercial participants (described above in section 3.2.3). This is a different procedure than the one used to estimate ex post impacts, which is why the overall average values in the table differ from the overall average ex post event impact.

[15] The rule actually used was greater than 0.02 kW of CAC load.

[16] No useful data due to the blackout.

[17] No useful data due to the blackout.

As the table shows, non-event day loads are highly predictive of average impacts. The table indicates that the top 30% of customers provide 67% and 60% of residential and commercial aggregate load impacts, respectively.

Table 4-5 also reports the standard errors of the estimates for each decile. It is important to note that while the overall trends in the table are consistent and likely reflect a true underlying pattern, the estimates at the decile level have fairly large standard errors. For example, the impact estimate for the highest decile for residential customers is statistically significantly different at the 5% level from the impact in the 5th decile, but not from 6th, 7th, 8th or 9th deciles. For commercial customers, none of the impact estimates are statistically significantly different from each other. When the data is divided into quartiles rather than deciles (not shown) some statistically significant differences appear for commercial customers.

Table 4-5: Average Estimated Impacts within Deciles of Usage

Decile	Residential Customers			Commercial Customers		
	Average Impact (kW)	% of Total	Impact Standard Error (kW)	Average Impact (kW)	% of Total	Impact Standard Error (kW)
1	0.03	1	0.09	0.03	1	0.06
2	0.11	2	0.15	0.08	2	0.12
3	0.06	1	0.17	0.16	4	0.17
4	0.26	5	0.19	0.23	5	0.19
5	0.31	5	0.23	0.37	9	0.21
6	0.43	8	0.24	0.36	8	0.25
7	0.69	12	0.28	0.51	12	0.26
8	1.02	18	0.33	0.53	12	0.31
9	1.26	22	0.34	0.77	18	0.40
10	1.55	27	0.50	1.27	29	1.13

5 Ex Ante Load Impact Estimates

The models described above were used to estimate load impacts based on ex ante event conditions and enrollment projections for the years 2012 through 2022. Enrollment is not expected to change in the future, so the tables below represent predictions for the whole period 2012 through 2022. FSC was provided with data by SDG&E that represents weather under 1-in-2 and 1-in-10 year conditions for each monthly system peak day.[18] The ex ante event window is from 1 to 6 PM, which is the CPUC resource adequacy window.

Tables 5-1 and 5-2 summarize the average and aggregate load impact estimates for residential and commercial customers, respectively. Aggregate impacts are based on steady enrollment levels equal to those as of fall 2011. Load impact estimates are presented for the average AC unit and for each customer segment as a whole.

For a typical event with 1-in-2 year weather conditions, the average impact per AC unit is 0.41 kW for residential customers. The 1-in-10 year typical event day estimate is 17% higher at 0.48 kW. The aggregate program load reduction potential for residential customers is 12 MW for a typical event day under 1-in-2 year weather conditions and 14 MW under 1-in-10 year weather conditions. September ex ante conditions are much hotter than typical conditions. The residential program is estimated to provide an average impact of 24 MW over a five-hour event on a 1-in-10 September event day.

There is significant variation in load impacts across months and weather conditions. Based on 1-in-2 year weather, the low temperatures in June, reflecting the well known "June Gloom" typically experienced in San Diego, result in small average and aggregate load impact estimates. The June 1-in-2 impact for residential customers is only 14% of the September estimate, which is the highest of any month in 1-in-2 year weather conditions. For residential customers the June 1-in-10 year estimate is almost 4 times higher than the 1-in-2 year estimate, which is a result of the average temperature being 11 degrees warmer than the 1-in-10 weather for June.

Commercial customers are estimated to provide similar, but slightly lower per CAC unit impacts than residential customers. Due to the smaller number of commercial installations in the program, aggregate impacts for the commercial segment are much smaller than for residential customers. The commercial program is expected to provide the highest impact under 1-in-10 conditions in September, when its expected impact is 6.2 MW.

Tables 5-3 and 5-4 provide ex ante estimates on an hourly basis for residential and commercial customers, respectively. Residential impacts peak in the hours 3-5 PM, while commercial impacts are relatively flat over the event hours.

Table 5-5 provides program-level ex ante aggregate estimates for each hour. The program is expected to provide its highest impact under 1-in-10 conditions in September. Under those conditions, the average impact over the event window is expected to be 30 MW, with an hourly peak of 34 MW from 3 to 5 PM.

[18] The typical event day is an hourly average of the weather during the top 9 system load days in a 1-in-2 year and in a 1-in-10 year.

Table 5-1:
Summer Saver Residential Ex Ante Impact Estimates

Day Type	Per CAC Unit (kW)		Aggregate (MW)	
	Weather Year		Weather Year	
	1-in-10	1-in-2	1-in-10	1-in-2
Typical Event Day	0.48	0.41	14	12
May Monthly Peak	0.39	0.17	11	5
June Monthly Peak	0.34	0.09	10	3
July Monthly Peak	0.52	0.42	15	12
August Monthly Peak	0.48	0.38	14	11
September Monthly Peak	0.83	0.64	24	19
October Monthly Peak	0.47	0.38	14	11

Table 5-2:

Summer Saver Commercial Ex Ante Impact Estimates

Day Type	Per CAC Unit (kW)		Aggregate (MW)	
	Weather Year		Weather Year	
	1-in-10	1-in-2	1-in-10	1-in-2
Typical Event Day	0.40	0.36	5.1	4.6
May Monthly Peak	0.33	0.24	4.3	3.1
June Monthly Peak	0.37	0.24	4.8	3.1
July Monthly Peak	0.39	0.36	5.0	4.7
August Monthly Peak	0.40	0.36	5.1	4.6
September Monthly Peak	0.48	0.42	6.2	5.4
October Monthly Peak	0.34	0.30	4.3	3.9

Table 5-3:
Aggregate Load Reductions by Day Type, Weather Year and Hour
All Residential Customers

Weather Year	Day Type	Hour of Day					Average
		1 to 2 PM	2 to 3 PM	3 to 4 PM	4 to 5 PM	5 to 6 PM	
1-in-2	Typical Event Day	10	10	14	14	12	12
	May Monthly Peak	4	4	6	6	5	5
	June Monthly Peak	2	2	3	3	3	3
	July Monthly Peak	10	10	14	14	12	12
	August Monthly Peak	9	9	13	13	11	11
	September Monthly Peak	15	16	22	21	19	19
	October Monthly Peak	9	9	13	13	11	11
1-in-10	Typical Event Day	11	11	17	16	14	14
	May Monthly Peak	8	9	13	14	12	11
	June Monthly Peak	6	7	11	13	12	10
	July Monthly Peak	12	13	18	18	15	15
	August Monthly Peak	12	11	16	16	14	14
	September Monthly Peak	19	20	28	28	25	24
	October Monthly Peak	10	11	15	16	15	14

Table 5-4:
Aggregate Load Reductions by Day Type, Weather Year and Hour
All Commercial Customers

Weather Year	Day Type	Hour of Day					Average
		1 to 2 PM	2 to 3 PM	3 to 4 PM	4 to 5 PM	5 to 6 PM	
1-in-2	Typical Event Day	4.4	4.9	4.6	4.7	4.5	4.6
	May Monthly Peak	2.8	3.2	3.1	3.1	2.9	3.1
	June Monthly Peak	2.9	3.3	3.1	3.2	3.1	3.1
	July Monthly Peak	4.4	4.9	4.6	4.7	4.6	4.7
	August Monthly Peak	4.4	4.9	4.6	4.7	4.5	4.6
	September Monthly Peak	5.0	5.6	5.4	5.5	5.3	5.4
	October Monthly Peak	3.6	4.1	3.8	4.0	3.8	3.9
1-in-10	Typical Event Day	4.7	5.4	5.1	5.3	5.0	5.1
	May Monthly Peak	4.0	4.5	4.2	4.4	4.2	4.3
	June Monthly Peak	4.5	5.0	4.7	4.9	4.7	4.8
	July Monthly Peak	4.7	5.3	5.0	5.1	4.9	5.0
	August Monthly Peak	4.9	5.4	5.1	5.3	5.0	5.1
	September Monthly Peak	5.8	6.5	6.1	6.4	6.1	6.2
	October Monthly Peak	4.1	4.6	4.4	4.5	4.2	4.3

Table 5-5:
Aggregate Load Reductions by Day Type, Weather Year and Hour
All Customers

Weather Year	Day Type	Hour of Day					Average
		1 to 2 PM	2 to 3 PM	3 to 4 PM	4 to 5 PM	5 to 6 PM	
1-in-2	Typical Event Day	14	15	19	19	17	17
	May Monthly Peak	7	7	9	9	8	8
	June Monthly Peak	5	5	6	6	6	6
	July Monthly Peak	14	15	19	19	17	17
	August Monthly Peak	13	14	18	18	16	16
	September Monthly Peak	20	22	27	27	24	24
	October Monthly Peak	13	13	17	17	15	15
1-in-10	Typical Event Day	16	16	22	21	19	19
	May Monthly Peak	12	14	17	18	16	15
	June Monthly Peak	11	12	16	18	17	15
	July Monthly Peak	17	18	23	23	20	20
	August Monthly Peak	17	16	21	21	19	19
	September Monthly Peak	25	27	34	34	31	30
	October Monthly Peak	14	16	19	21	19	18

Appendix A. Discussion of Modeling Choices

As compared to the two previous years, Summer Saver events in 2011 did not lend themselves well to modeling by observing loads at non-event times with similar temperatures. This was true for two reasons. First, the load research sample for contract settlement was corrupted, leaving only a small sample of customers with unperturbed load on any given event day. Second, four of the six events were not ideal from a modeling perspective in that they had certain unusual aspects that made them different from all non-event days during 2011. This is especially true if the goal is to use the 2011 events alone as input into a predictive model of event impact as a function of event day temperatures.

For these reasons, it was determined that reference load estimation should not be limited to being based on loads observed during similar weather conditions when other sources of reference load may be more accurate. With this guideline in mind, two different methodologies were used to estimate load impacts for both customer segments – individual customer regression based on weather, as has been used previously, and day-matching based on load shapes and magnitudes. Each method is described above in section 3 in the context of either residential or commercial customers. In fact, both methods were used for both customer segments and the results are compared in this Appendix. The main conclusion from using these two methods is that from a practical standpoint, the two methods are each adequate for residential customers. For commercial customers, only the day-matching method produced reliable estimates.

The initial reason for looking to alternatives to individual customer regressions came from the fairly poor performance of the method for commercial customers on the October 12 and 13 event days[19]. This is shown in Figures A-1 and A-2, which compare predicted reference load on those days to actual load.

As shown in Figure A-1, the model under-predicts loads in the time leading up to the event on October 12. This is due to the lack of other comparable days in the summer that are so hot following a very cool period. The model then predicts an implausible spike in reference load during the 4-5 PM hour because the load and temperature information from the rest of the summer indicate that a day with such a high temperature must have a large event impact. This spike is not observed during any non-event day; commercial loads tend to peak during the 3-4 PM hour. This suggests that the spike is an artifact of the model trying to fit a large event impact. Moreover, examination of the load data itself indicates that it is much more likely that load impacts for that day are simply lower than would be expected on a day with that temperature. This makes sense given that it was an unusually hot day in mid-October and the days leading up to it were significantly cooler.

[19] It also performed badly on September 9, but this could have been more easily fixed had the decision been made to use the regression model as the primary source of impact estimates. This is discussed below.

Figure A-1: Average Commercial Actual Load and Predicted Reference Load for October 12

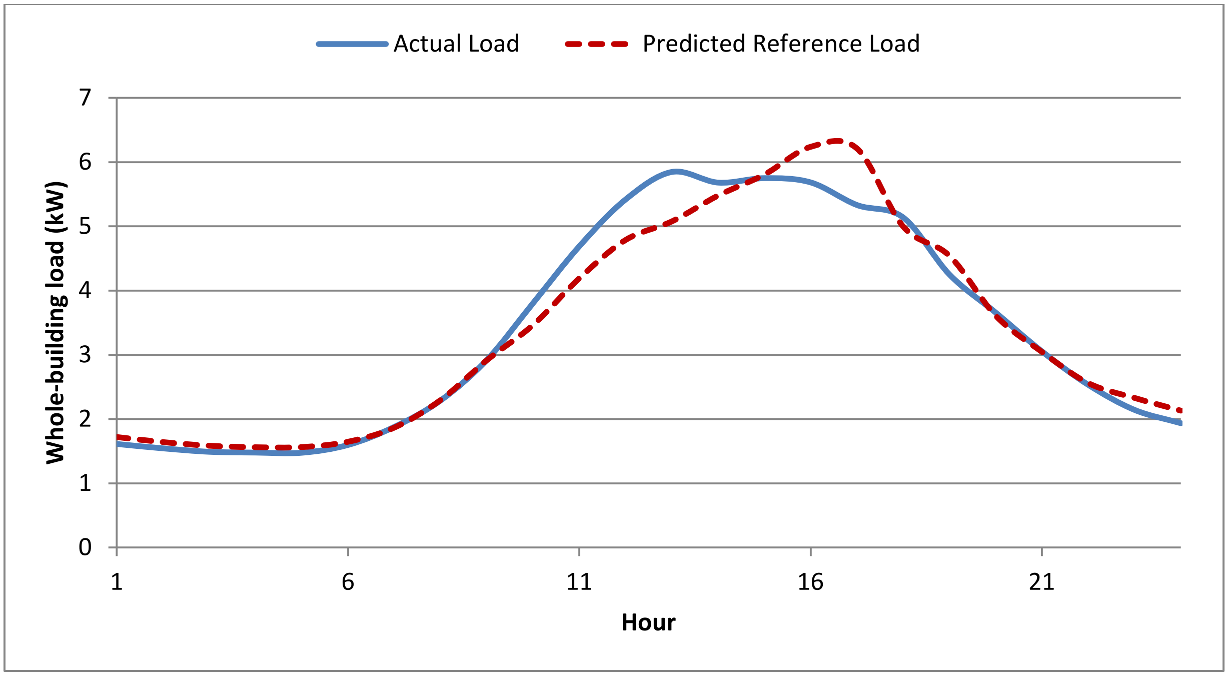
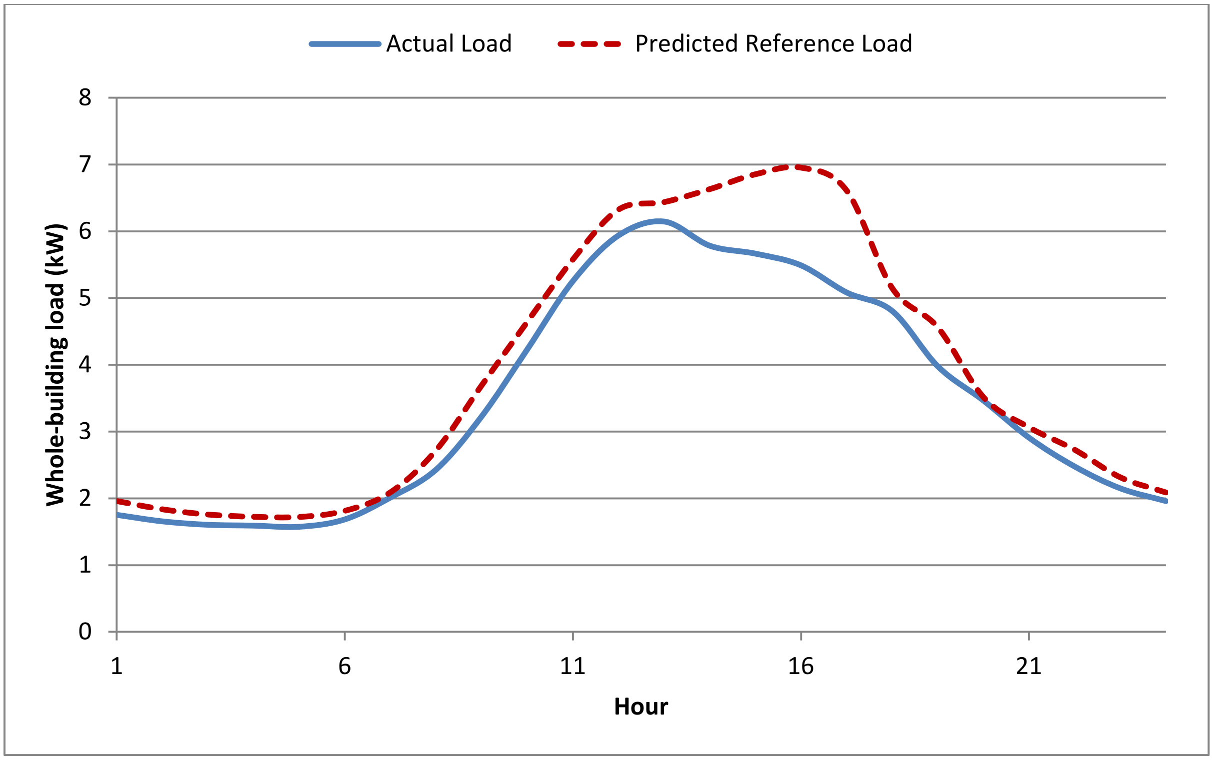


Figure A-2 shows that the model forces the reference load to be implausibly high on October 13 during the 4-5 PM hour. Again, it does this because the other information that the model is based on indicates that the event impact should be higher on such a hot day. That the model, and all other plausible regression models based on weather, produced such clear inaccuracies for 2 out of 6 event days prompted the use of day-matching to estimate commercial ex post load impacts. As shown in the figures in Appendix B, day-matching produces more plausible reference loads for these October event days.

Figure A-2: Average Commercial Actual Load and Predicted Reference Load for October 13



A.1. Day Matching

Under certain conditions, individual customer regressions do not necessarily provide the most accurate reference load estimates. This occurs when there is reason to believe that the loads on an event day are not accurately predicted by a simple function of the temperature on that day. Two factors arose in 2011 that call into question estimates based on individual customer regressions. First, two of the event days occurred on days of unseasonable warmth in mid-fall, leading to smaller loads than when similar temperatures occurred earlier in the season. Second, the only heat wave of the summer took place from September 6-8. The last two days were both event days and each had higher loads during the pre-event hours than any other day of the summer, including the only other heat wave of the summer. This means that the only source of reference load is an extrapolation from loads observed during cooler conditions. In this situation, linear regression has no particular advantage over simpler methods, such as the day-matching method used here.

There were a total of six event days in 2011; two of them occurred in mid-October. Figure A-3 shows that for residential customers the loads on those days were much lower than on the only non-event day with comparable temperature and were similar to loads observed on days with lower temperatures. Figure A-4 shows a side by side comparison for each day's average whole building load and average temperature.

For residential customers, average temperatures peaked at 95°F on October 12 and 13. Temperatures peaked at 94°F on September 6 and at 87°F on August 25, both non-event days. As Figure A-3 shows, between the two, August 25 provides a much more plausible reference load for the

October event days even though the temperature on September 6, as indicated by Figure A-4 is closer to that on the October event days. There are several other non-event days with higher loads than August 25. The important point is that those days also have much higher loads than the October event days, despite being substantially cooler. In other words, merely warm days in mid-summer tend to have higher loads than hot days in October. This means that a temperature-based model may produce inaccurate estimates for the October event days.

Figure A-3: Residential Whole-building Load on October Event Days, August 25 and September 6

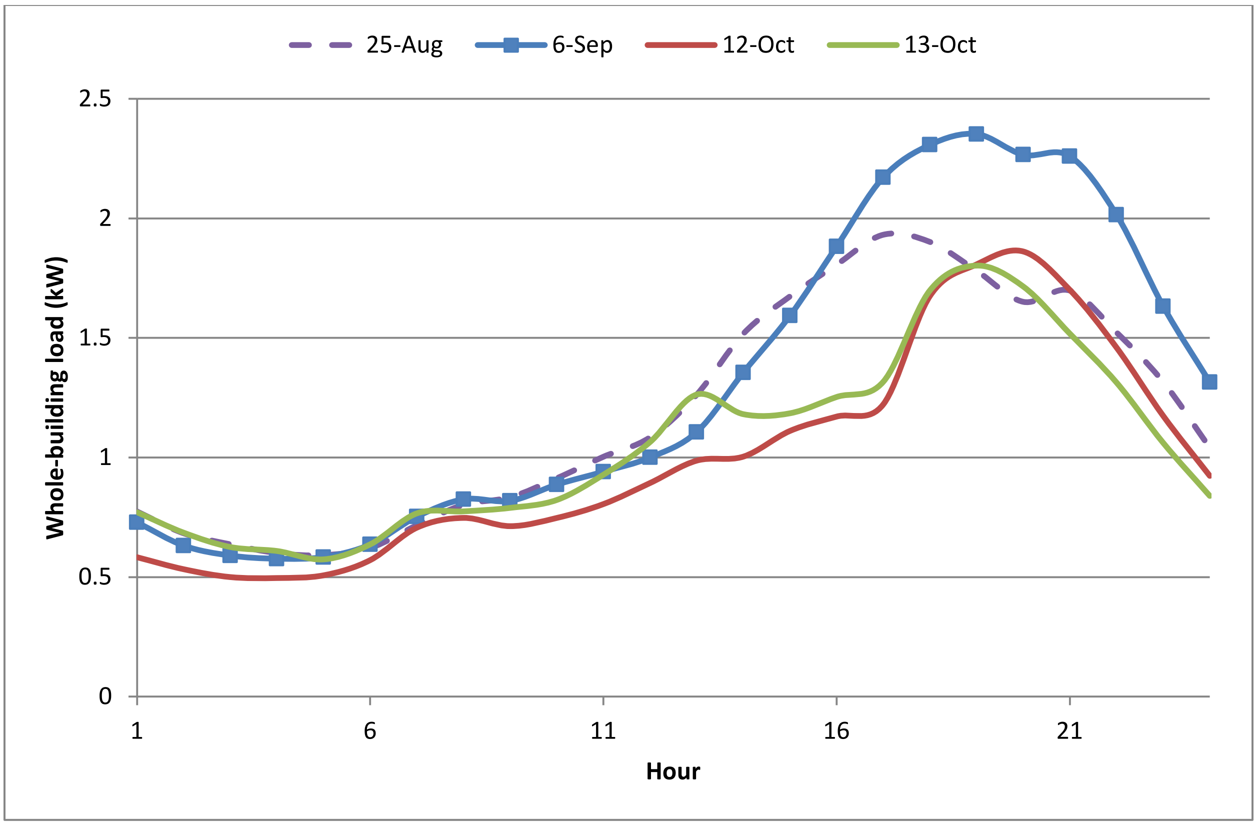
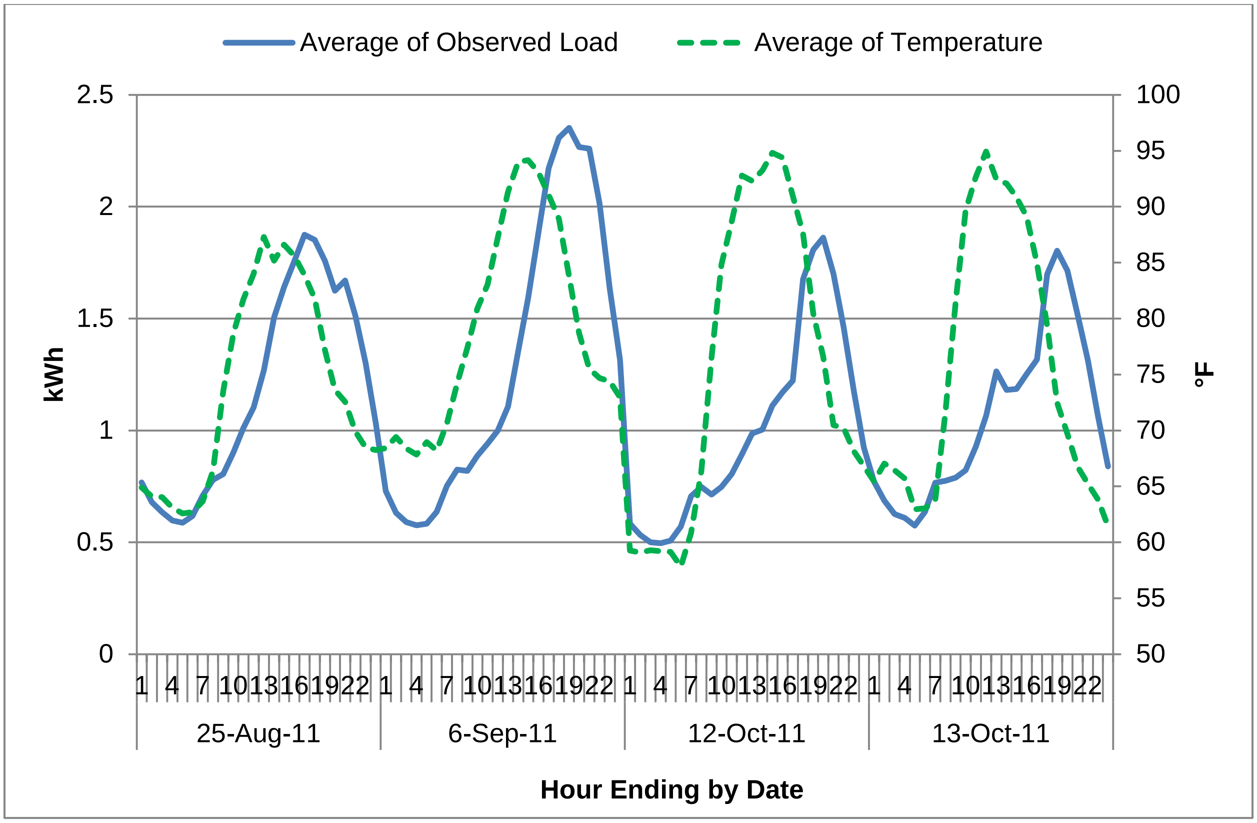


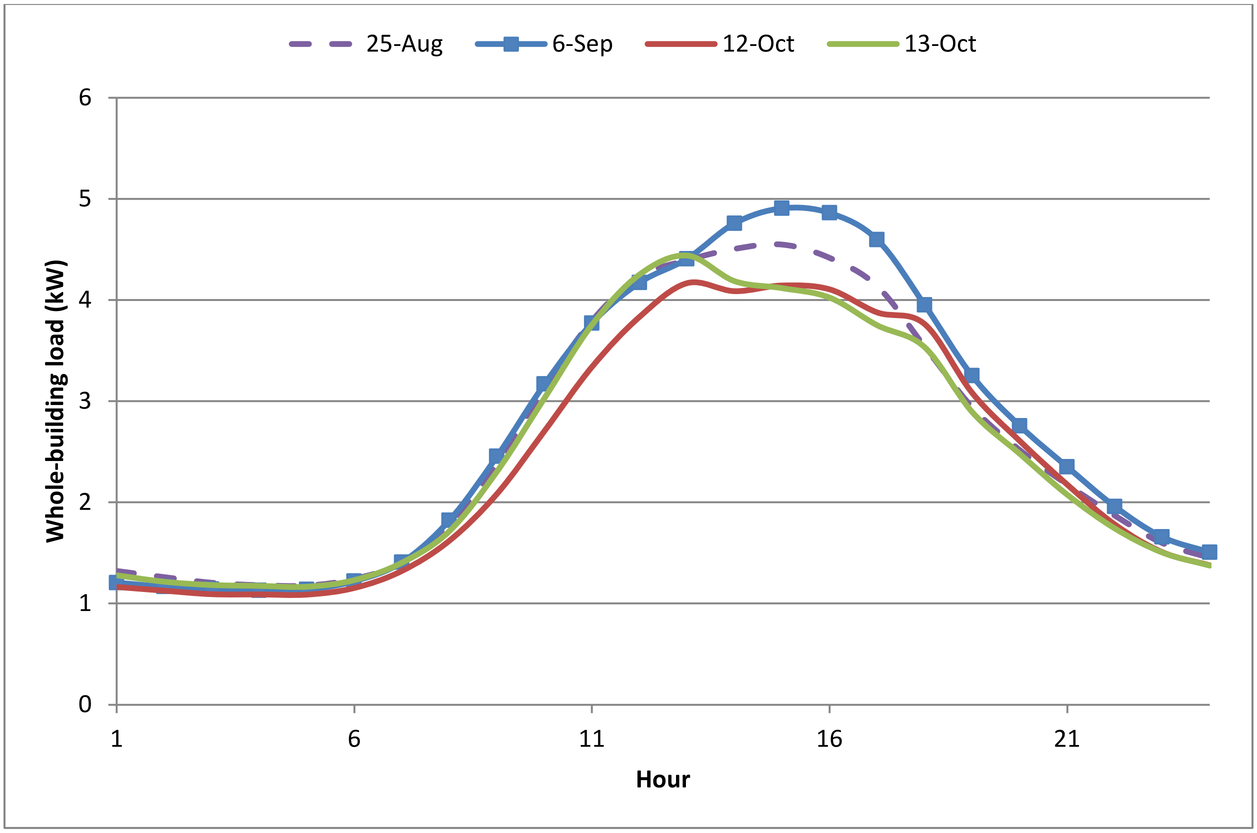
Figure A-3 illustrates a point made in the 2010 evaluation as well. Loads vary for many unobservable reasons, which can lead temperature-based estimates to be inaccurate in certain circumstances. In the 2010 evaluation, however, there was a treatment-control design that automatically provided good reference load estimates during all event hours. Although such a design was in place for 2011, the design was corrupted and cannot be used for this analysis, as mentioned above.

Figure A-4: Residential Whole-building Load and Average Temperature on October Event Days, August 25 and September 6



For commercial customers the situation is more complex; but the basic conclusion is similar, as is shown in Figure A-5. The two October event days have loads in the pre-event hours that are similar to the only other day of comparable heat, September 6. However, they also have similar loads during those hours to August 25, a much cooler day. The September 6 load during the afternoon and evening is significantly higher than that on August 25, which makes sense given the higher temperatures. The September 6 load remains significantly higher than the October event day loads in the post-event hours. This suggests that the October event day loads, in the absence of an event, would have behaved more similarly to the load on August 25, which is slightly lower than the October event day loads during the post-event hours. A regression based on weather, however, does not yield this result, as shown above in Figures A-1 and A-2.

Figure A-5: Commercial Whole-building Load on October Event Days, August 25 and September 6



Given these complications, the day-matching procedure that is described in section 3.2.3 was applied to both commercial and residential customers. Table A-1 shows the days that were chosen to provide reference load for each ex post event day. Appendix B shows graphs of the load shapes and adjusted load shapes for each event day load and reference day load.

Table A-1: Event Days and Matched Reference Load Days

Event Day	Matched Days	
	Residential	Commercial
26-Aug-11	29-Aug-11	2-Aug-11
7-Sep-11	6-Sep-11	2-Aug-11
8-Sep-11	9-Sep-11	2-Aug-11
9-Sep-11	31-Aug-11	7-Jul-11
12-Oct-11	24-Aug-11	6-Sep-11
13-Oct-11	25-Aug-11	25-Aug-11

Based on the figures in Appendix B, the day-matching reference loads for commercial customers appear quite plausible. The day-matching reference loads for residential customers appear less accurate, but still fairly plausible in most cases.

Having identified matched days, load impacts for each cycling option within each customer segment were estimated by subtracting average hourly load during each event from average hourly load during the same hours of the matched reference day. Standard errors were calculated at an hourly level as the square root of the sum of squared standard errors of each hourly average load.

A.2. Results Comparison

Table A-2 shows a comparison of residential ex post estimates developed using day matching and individual customer regressions. The table shows values for each residential cycling option separately and for all customers. The average estimates from day-matching are lower, due primarily to the October event days where the regression function produces a larger impact based on the high temperatures on those days.

Table A-2: Ex Post Load Impact Estimates for Residential Customers Developed Using Two Methods (kW/CAC unit)

Date	50		100		All	
	Day Matching	Regression	Day Matching	Regression	Day Matching	Regression
26-Aug-11	0.42	0.44	0.31	0.38	0.36	0.41
7-Sep-11	0.94	0.80	1.00	0.74	0.97	0.77
8-Sep-11[20]	0.64	0.77	0.48	0.81	0.55	0.79
9-Sep-11	0.18	0.24	0.08	0.24	0.13	0.24
12-Oct-11	0.16	0.50	0.25	0.48	0.21	0.49
13-Oct-11	0.45	0.73	0.36	0.76	0.40	0.74
Average	0.47	0.58	0.41	0.57	0.44	0.57

While there are some appreciable differences in the estimates developed using each method for residential customers, these differences are of secondary importance to the issue of whether either set of estimates leads to different conclusions about expected future program performance. To this end, both sets of estimates are consistent with the ex ante estimates developed in 2010, and either set of ex post estimates leads to nearly identical ex ante estimates for 2012 and beyond. This will be documented in the ex ante report to follow. In the end the regression model was chosen on pragmatic grounds. Both the ex post and ex ante regression models were already fully built and their output documented by the time the day-matching results were being produced. It took substantially less work to verify that using the day-matching model would not materially change ex ante results than it would take to fully produce and document those results.

[20] Result is only calculated over the first two hours of the event.

Table A-3 shows a comparison of ex post estimates for commercial customers developed using day matching and individual customer regressions. The table shows values for each commercial cycling option separately and for all commercial customers together. The estimates vary across methods substantially. In one case the estimated event impact is negative for customers on 30% cycling. This case is less important than it appears because it takes place under unusually cool event conditions. In this case, the model fits a general trend to event impact as a function of temperature and the best fit happens to be negative at such a low temperature. This would not occur if there were many observable events at temperatures in the mid-70s. Moreover, if the regression results were being used as the final commercial ex ante estimates, then that day could have been modeled separately, leading to a more reasonable, but still quite low impact estimate.

More important is the general implausibility of the regression results, as displayed in Figures A-1 and A-2. The figures in Appendix B show that, at the least, the day-matching procedure produces plausible reference loads in almost all cases. This is not true for the regression model. Additionally, unlike in the residential case, the regression model produces ex ante results different enough from previous results to be questionable given the amount of useful information they are based on. For these reasons, it was decided to use the day-matching results to produce the commercial ex post results. Additionally, it was decided to use the day-matching ex post results in conjunction with 2010 ex post results to develop an ex ante model for commercial customers. This will be documented in the ex ante report to follow.

Table A-3: Ex Post Load Impact Estimates for Commercial Customers Developed Using Two Methods (kW/CAC unit)

Date	30		50		All	
	Day Matching	Regression	Day Matching	Regression	Day Matching	Regression
26-Aug-11	0.35	0.28	0.34	0.28	0.34	0.28
7-Sep-11	0.30	0.63	0.31	0.57	0.31	0.59
8-Sep-11[21]	0.32	0.58	0.41	0.75	0.38	0.67
9-Sep-11	0.13	-0.24	0.18	0.16	0.16	0.01
12-Oct-11	0.33	0.45	0.27	0.26	0.29	0.34
13-Oct-11	0.24	0.61	0.27	0.45	0.26	0.52
Average	0.28	0.39	0.30	0.41	0.29	0.40

[21] Result is only calculated over the first two hours of the event.

Appendix B. Day-matching Load Shapes

This appendix provides information on the plausibility of the reference loads obtained through day-matching. Table A-1, above, shows the list of days used as matches for each event day for each customer segment. The two sections that follow show the whole-building load of each event day for each cycling option within each customer segment as compared to the whole-building load on the matched reference day. The adjusted reference day load is also shown.

B.1. Residential Day-Matching Figures (Event Window Shaded)

Figure B-1: Residential Load on August 26 and Matched Unadjusted Reference Load

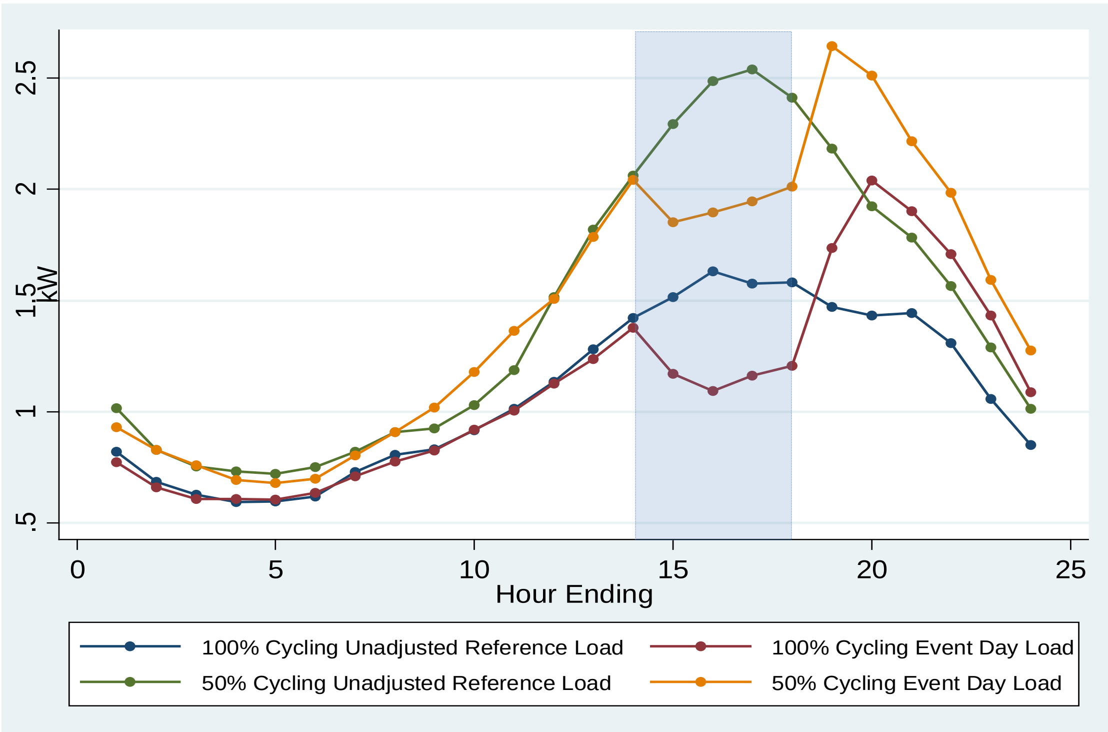


Figure B-2: Residential Load on August 26 and Matched Adjusted Reference Load

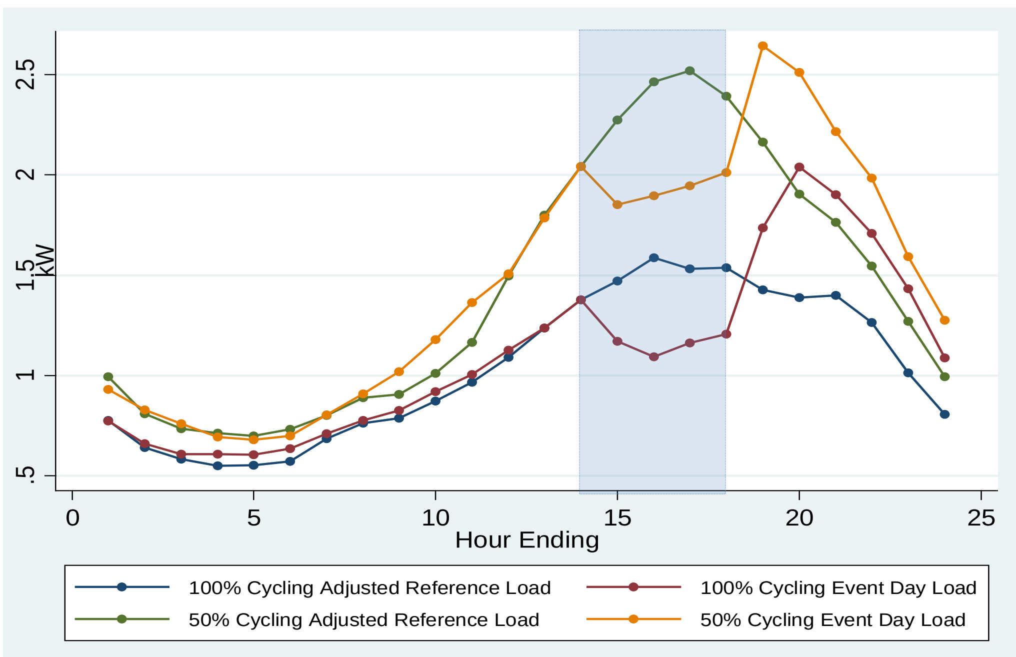


Figure B-3: Residential Load on September 7 and Matched Unadjusted Reference Load

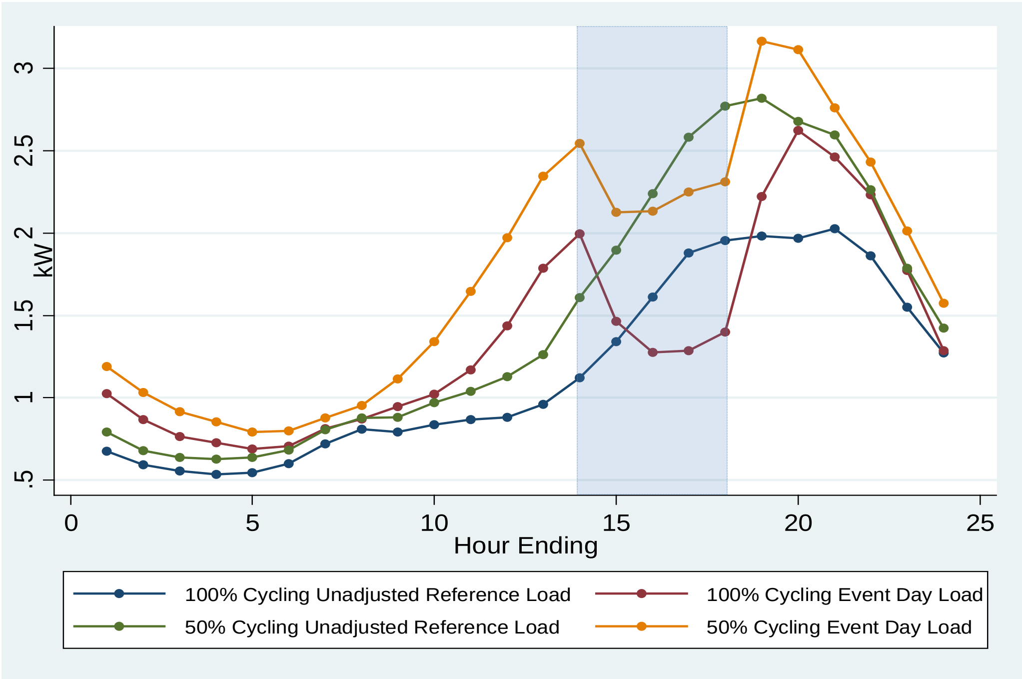


Figure B-4: Residential Load on September 7 and Matched Adjusted Reference Load

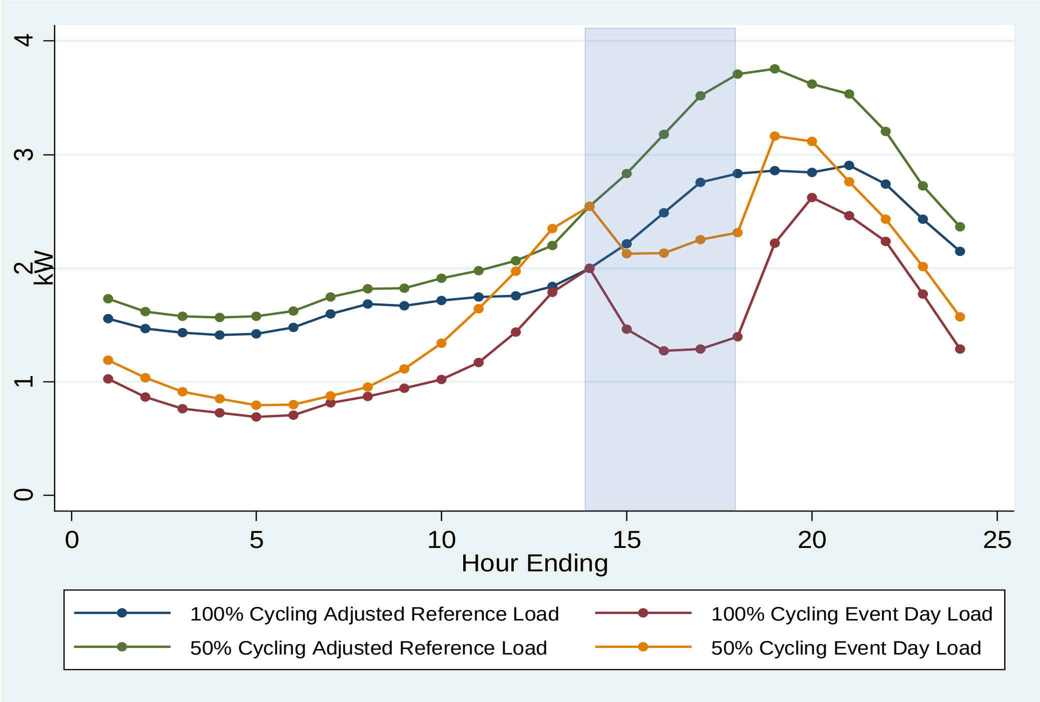


Figure B-5: Residential Load on September 8 and Matched Unadjusted Reference Load

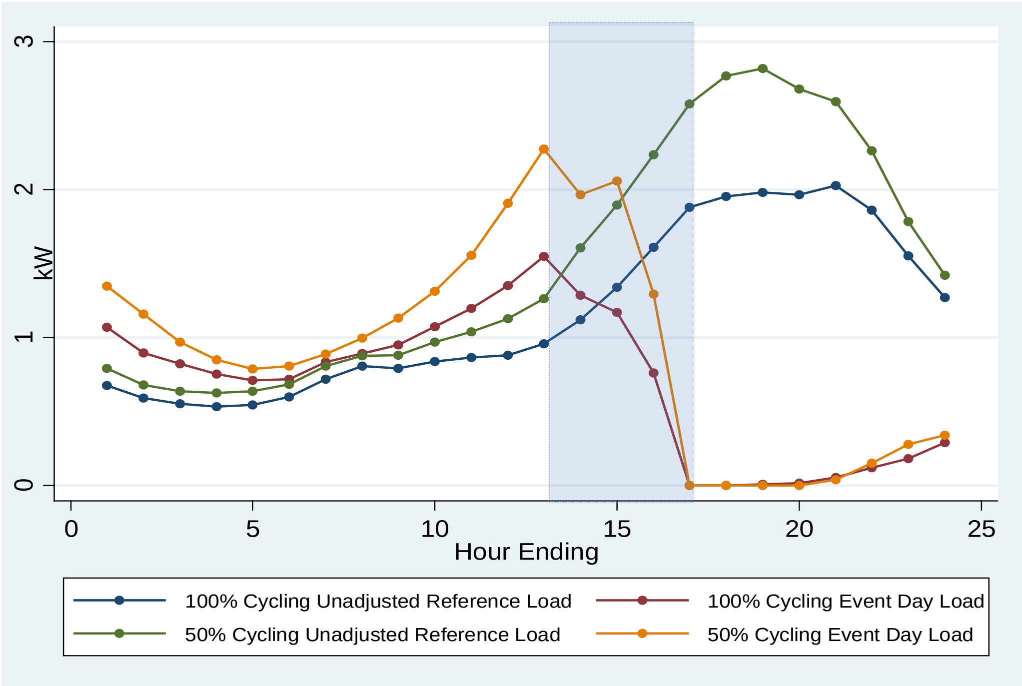


Figure B-6: Residential Load on September 8 and Matched Adjusted Reference Load

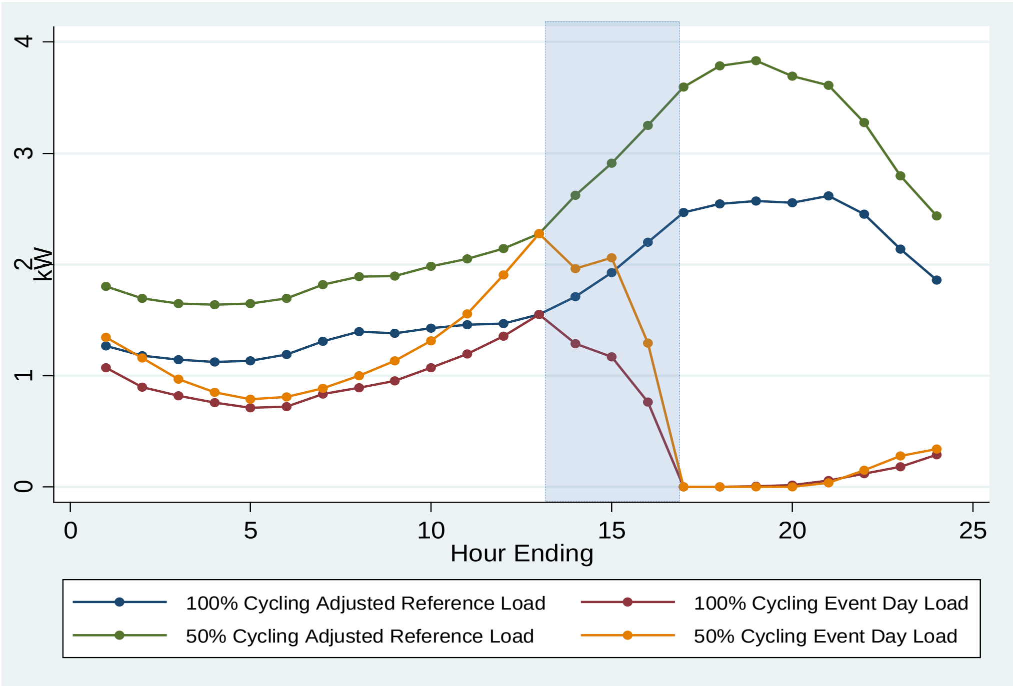


Figure B-7: Residential Load on September 9 and Matched Unadjusted Reference Load

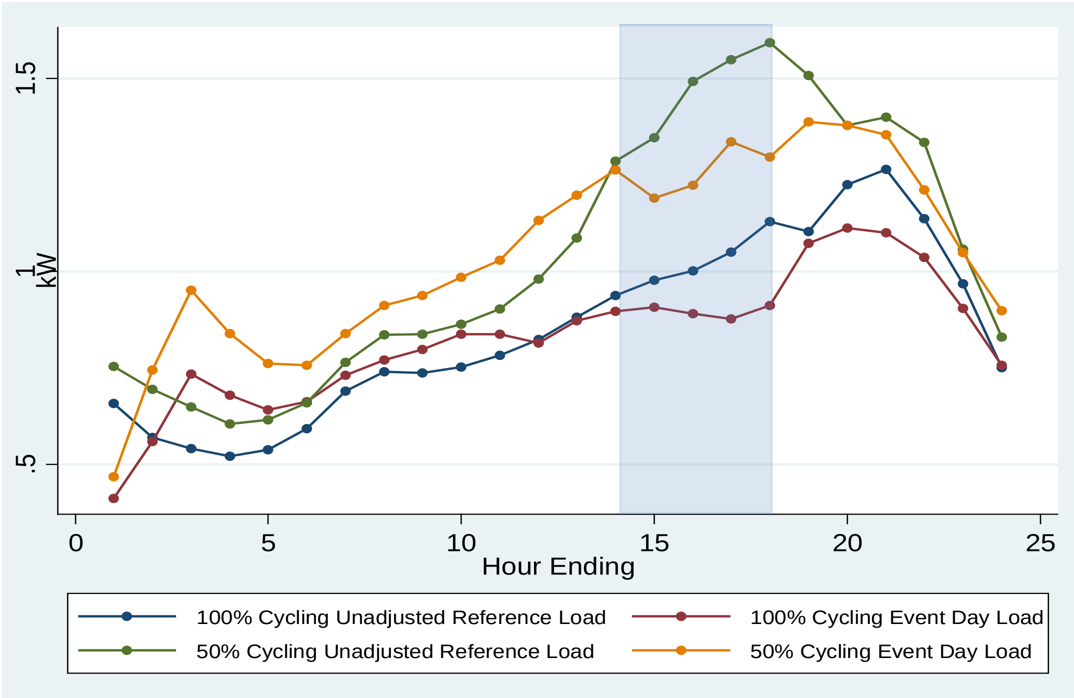


Figure B-8: Residential Load on September 9 and Matched Adjusted Reference Load

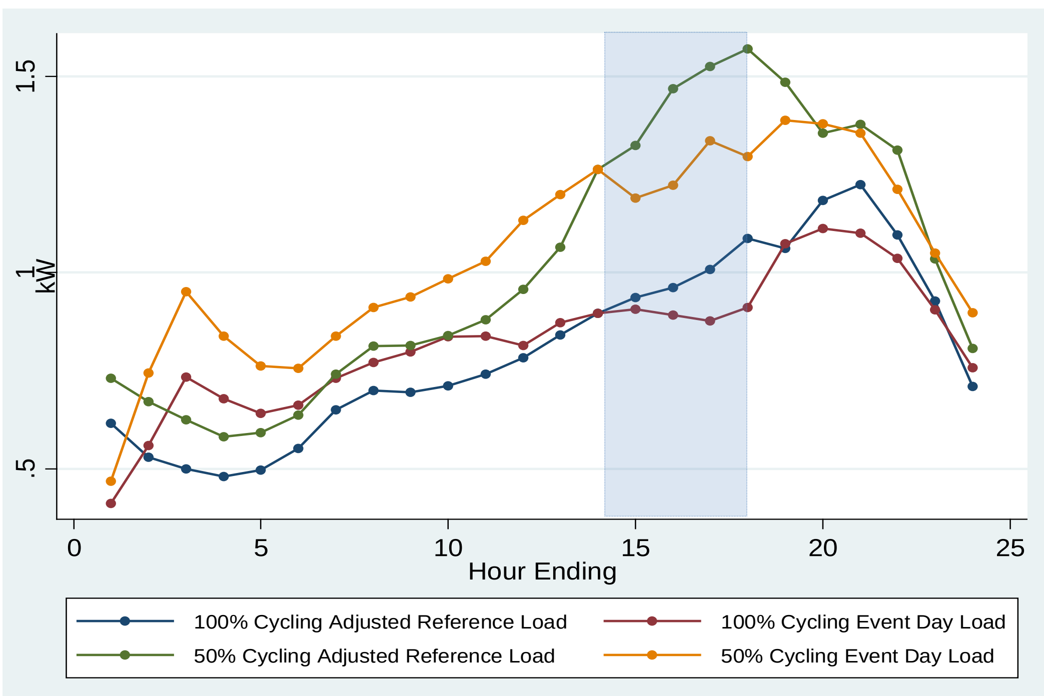


Figure B-9: Residential Load on October 12 and Matched Unadjusted Reference Load

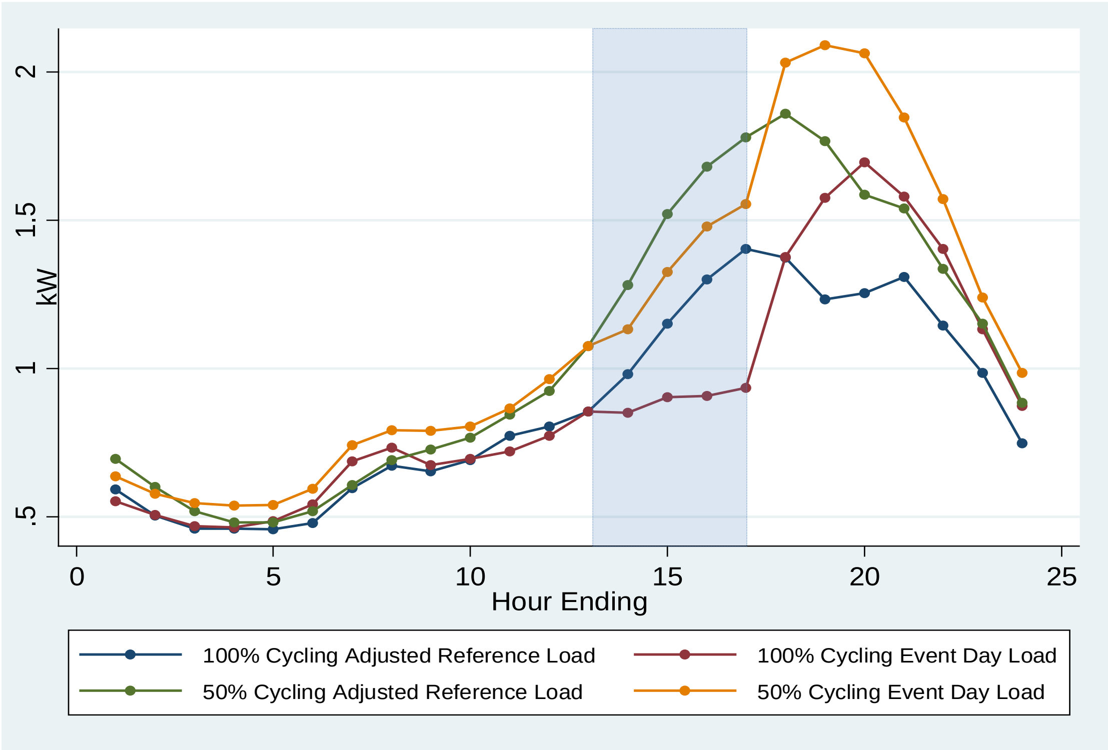


Figure B-10: Residential Load on October 12 and Matched Adjusted Reference Load

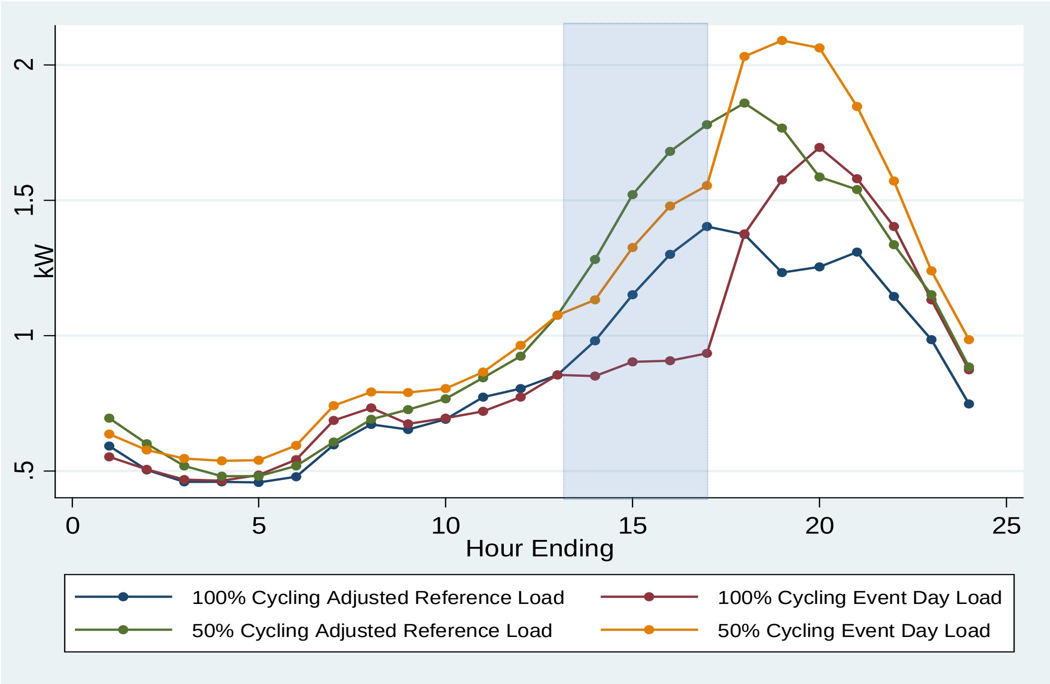


Figure B-11: Residential Load on October 13 and Matched Unadjusted Reference Load

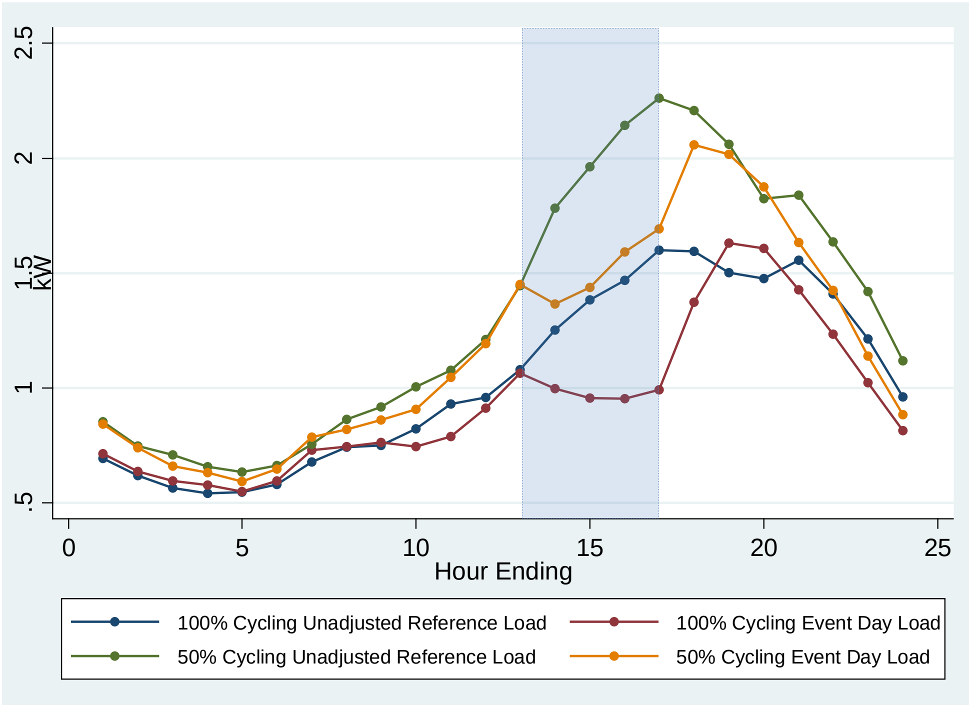
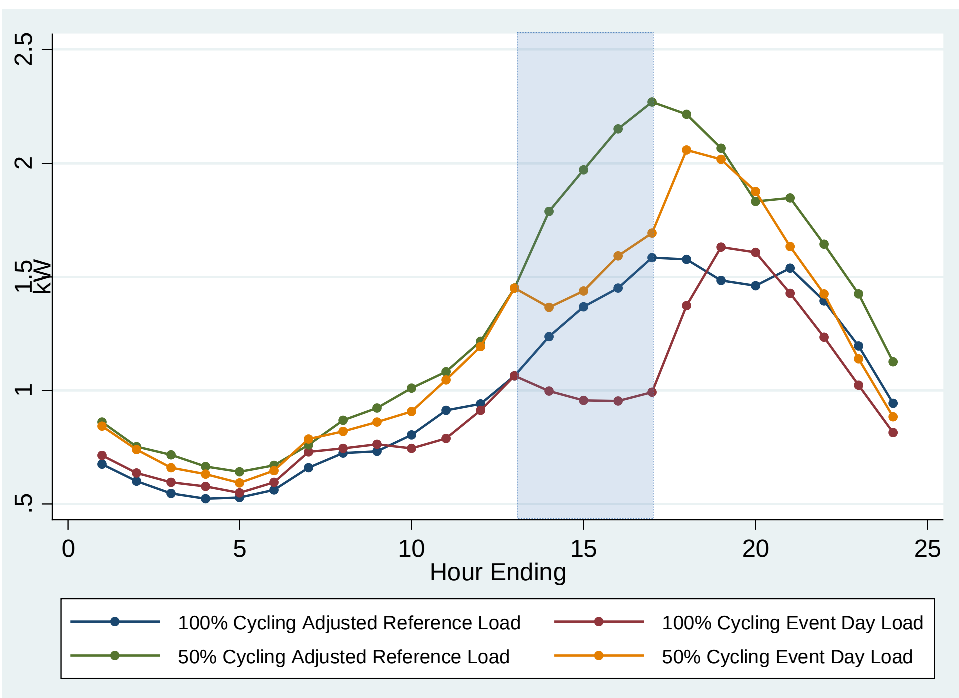


Figure B-12: Residential Load on October 13 and Matched Adjusted Reference Load



B.2. Commercial Day-Matching Figures (Event Window Shaded)

Figure B-13: Commercial Load on August 26 and Matched Unadjusted Reference Load

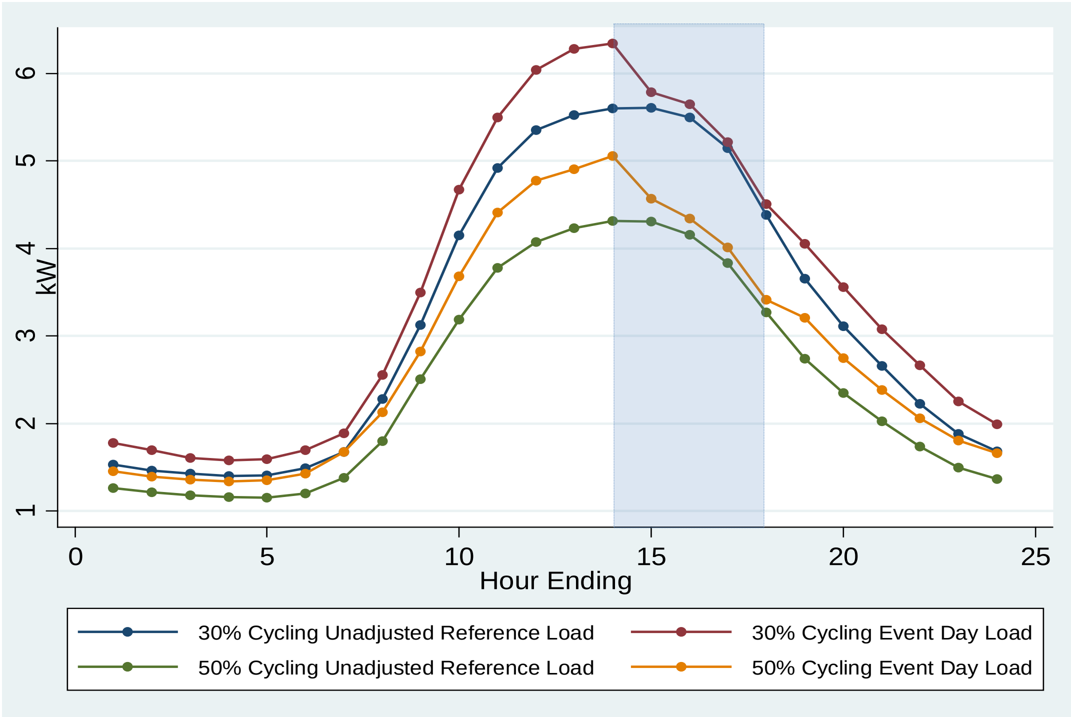


Figure B-14: Commercial Load on August 26 and Matched Adjusted Reference Load

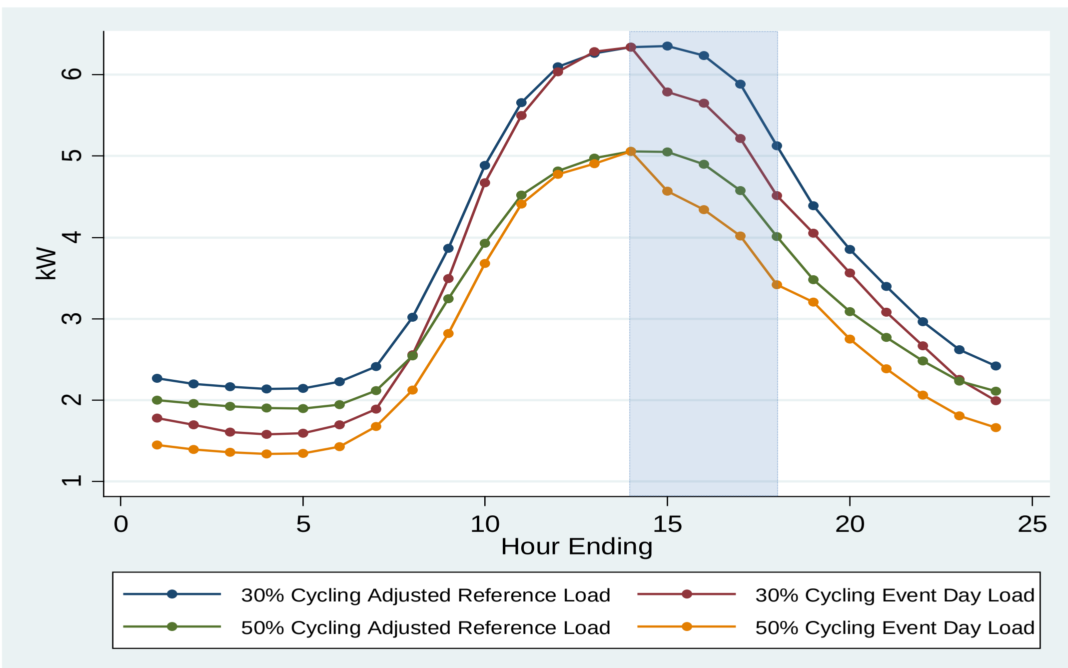


Figure B-15: Commercial Load on September 7 and Matched Unadjusted Reference Load

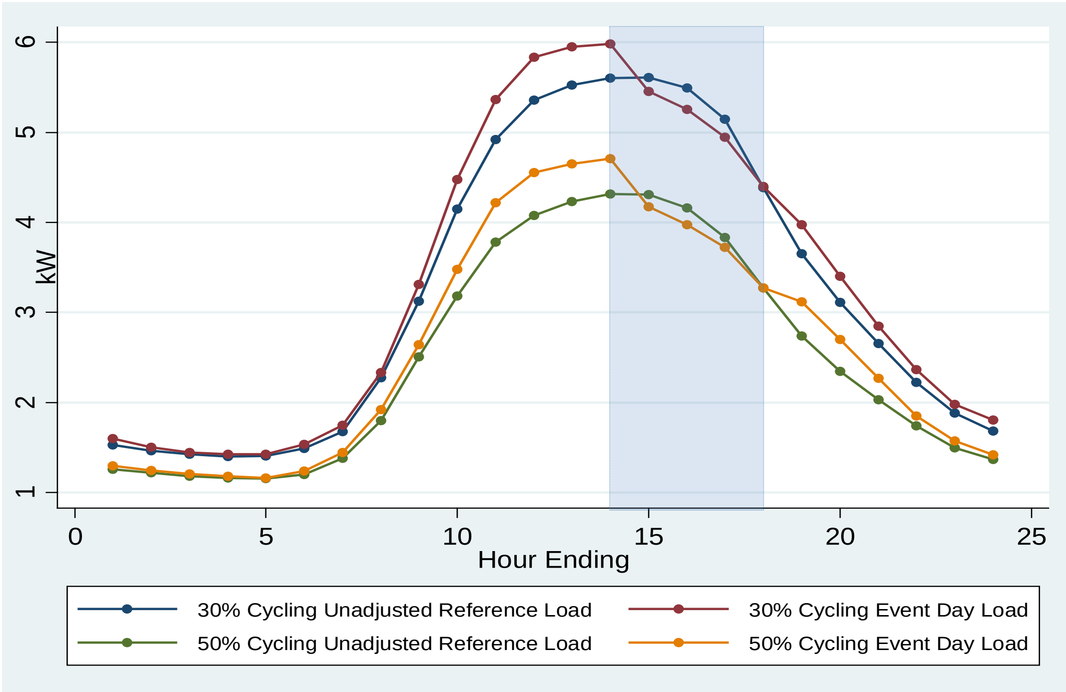


Figure B-16: Commercial Load on September 7 and Matched Adjusted Reference Load

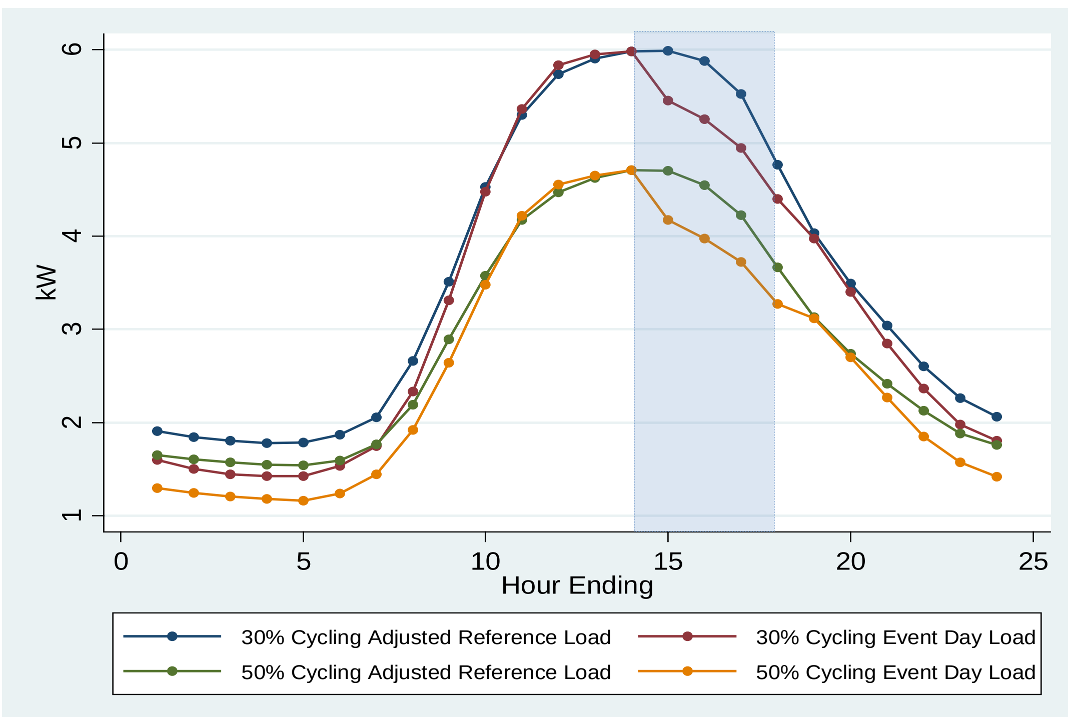


Figure B-17: Commercial Load on September 8 and Matched Unadjusted Reference Load

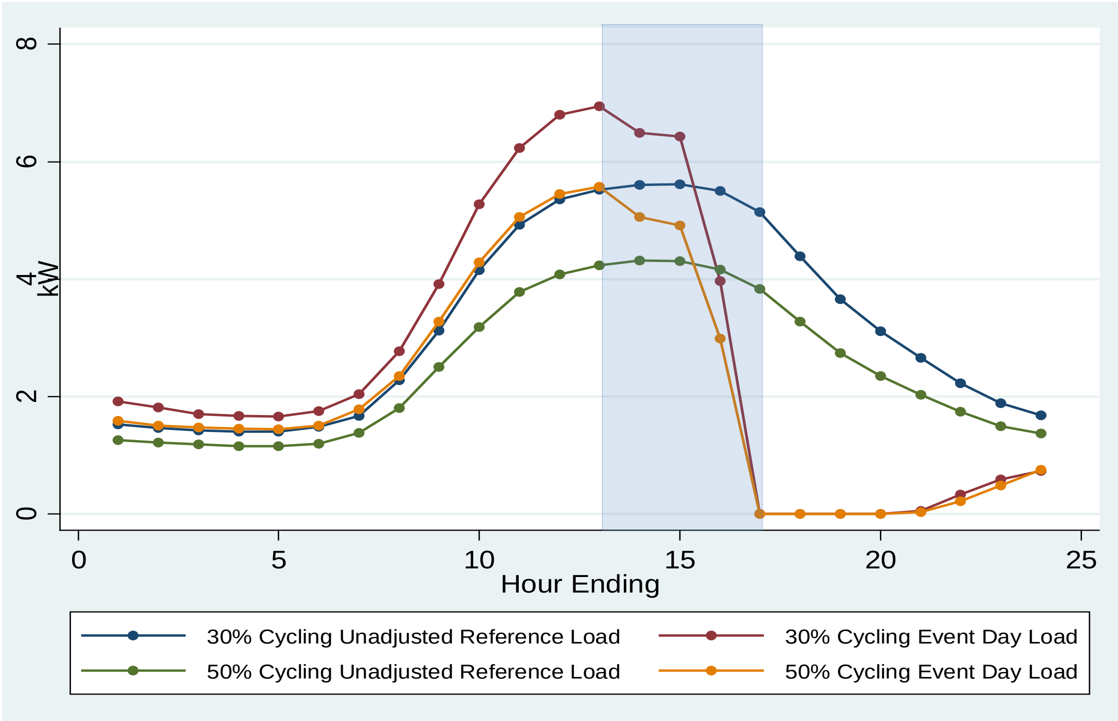


Figure B-18: Commercial Load on September 8 and Matched Adjusted Reference Load

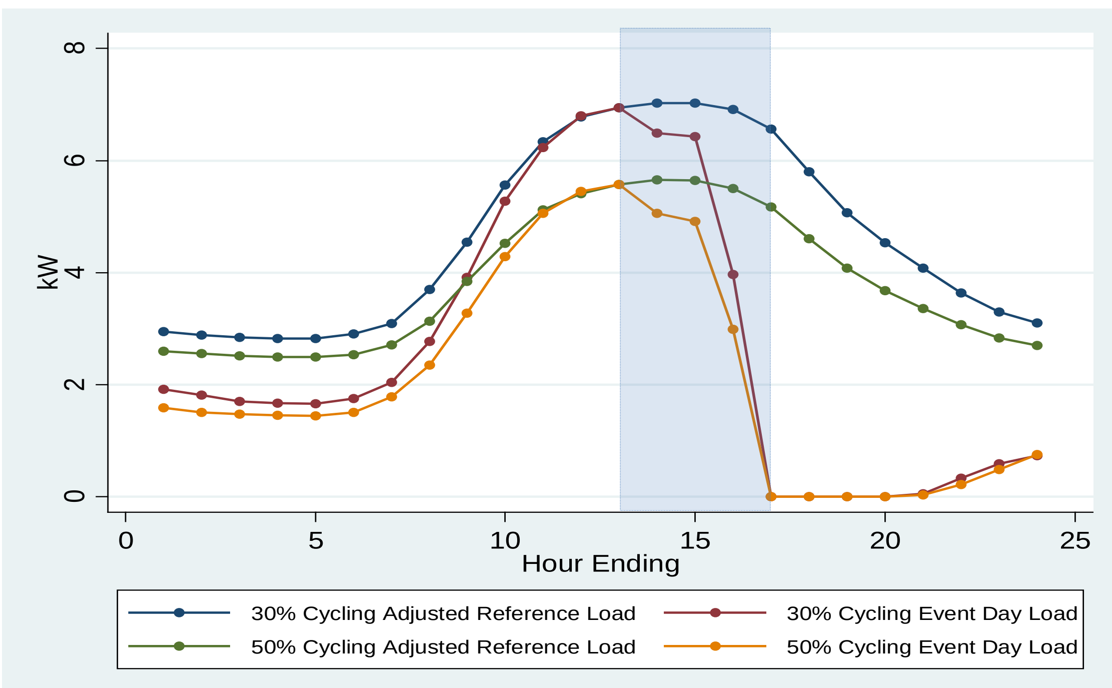


Figure B-19: Commercial Load on September 9 and Matched Unadjusted Reference Load

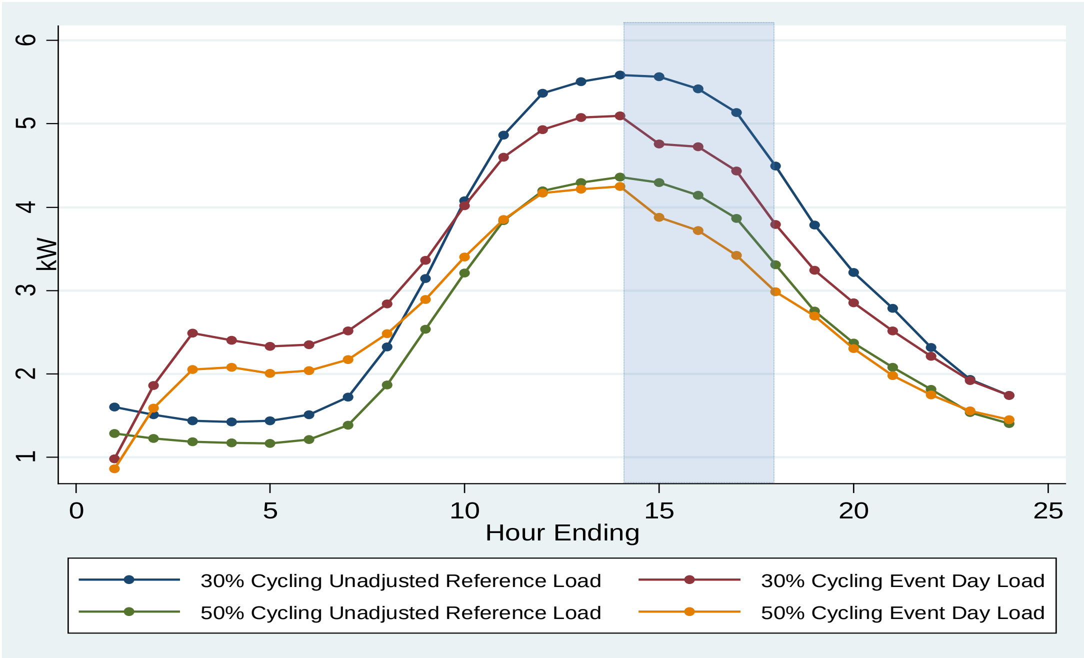


Figure B-20: Commercial Load on September 9 and Matched Adjusted Reference Load

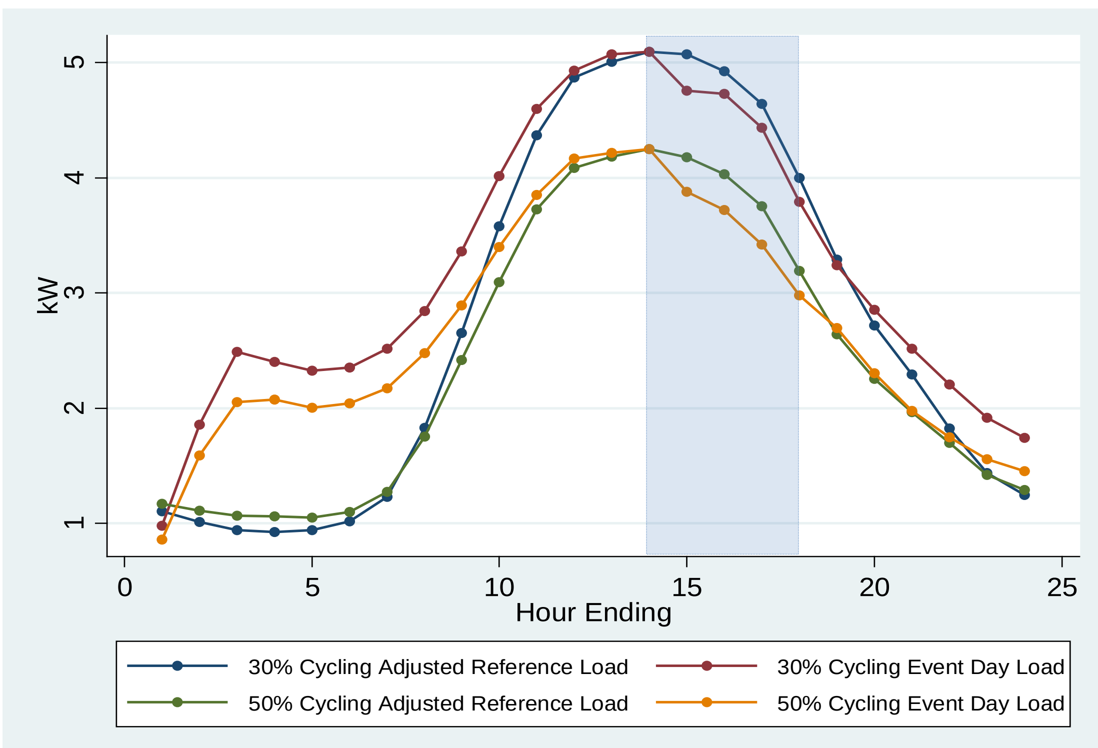


Figure B-21: Commercial Load on October 12 and Matched Unadjusted Reference Load

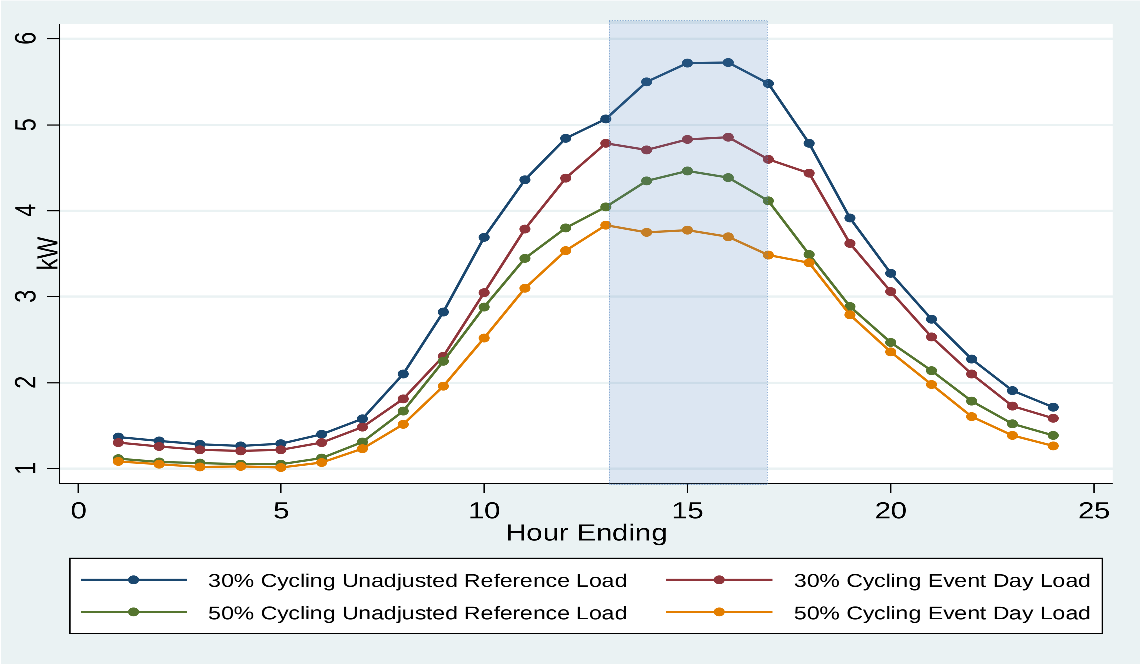


Figure B-22: Commercial Load on October 12 and Matched Adjusted Reference Load

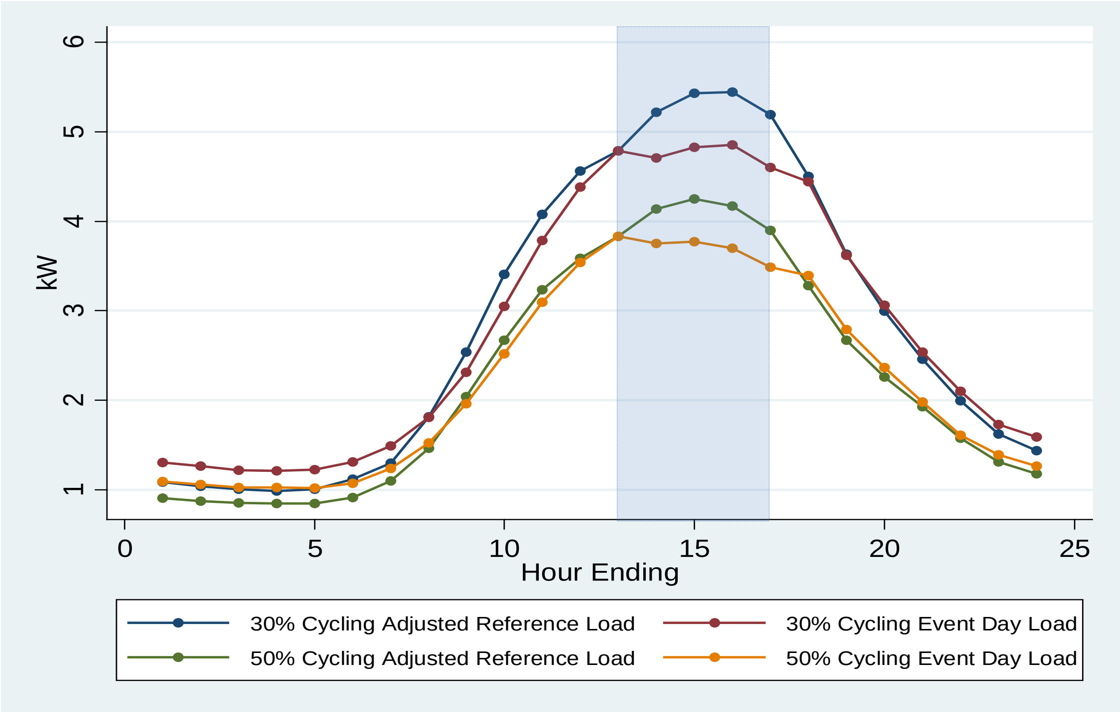


Figure B-23: Commercial Load on October 13 and Matched Unadjusted Reference Load

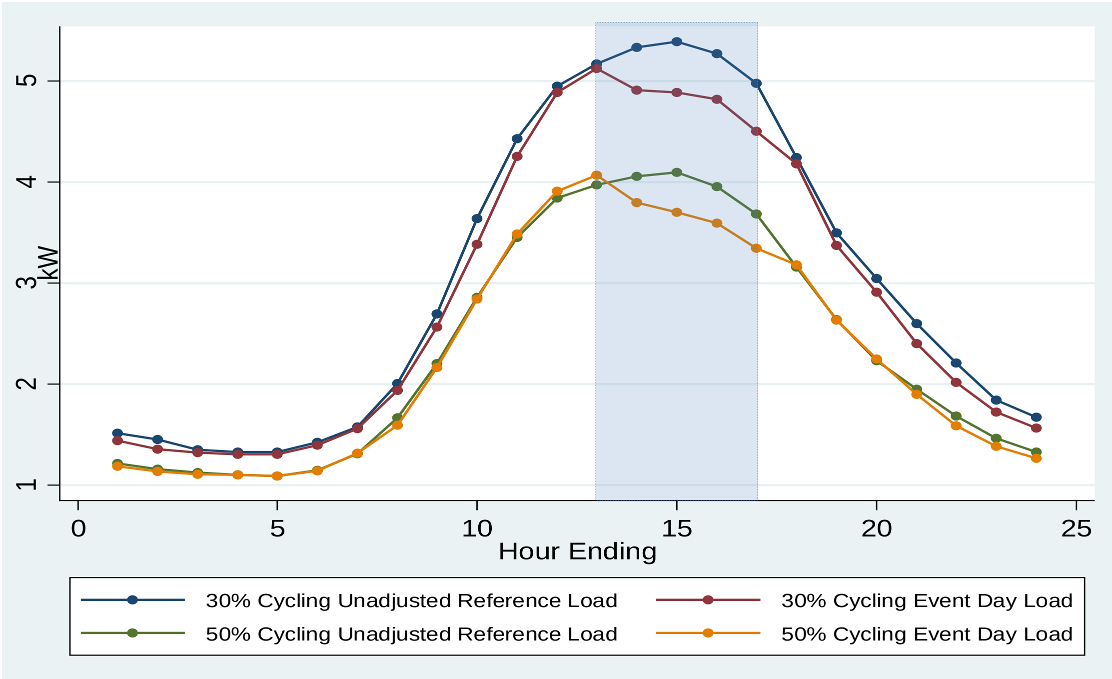
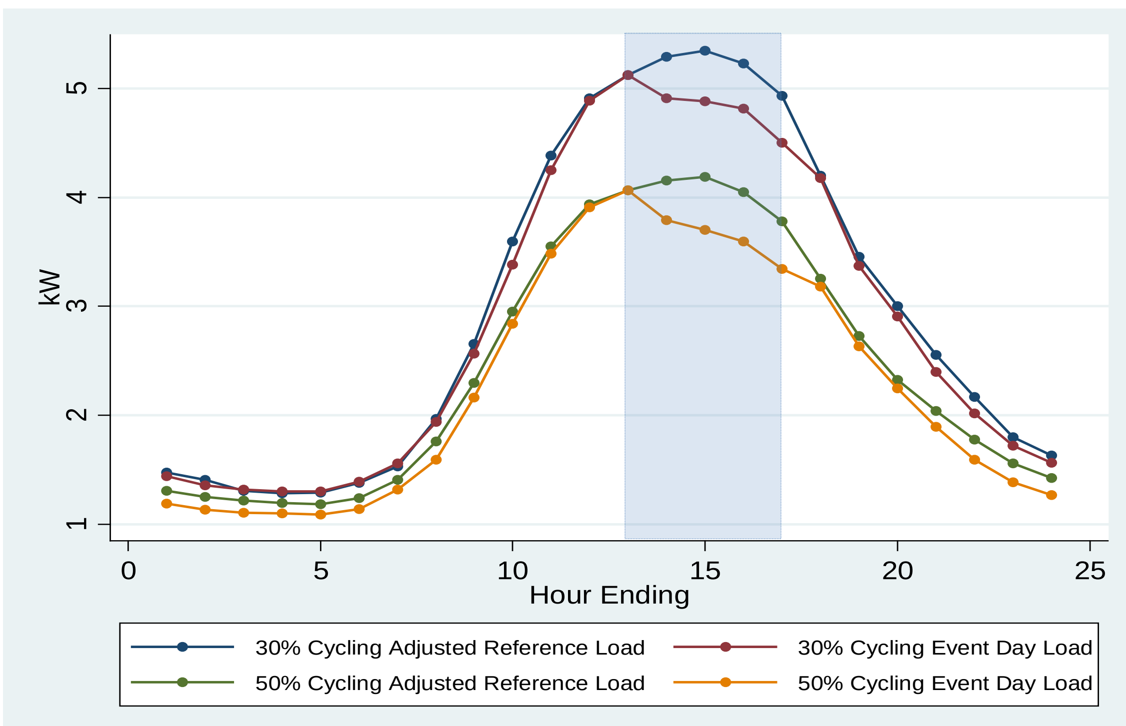


Figure B-24: Commercial Load on October 13 and Matched Adjusted Reference Load



Appendix C. Revised 2010 Ex Post Values

Two data processing errors were discovered that affect the aggregate residential and commercial ex post values from 2010. Table C-1 shows the originally reported and revised values for each customer segment and for all customers. The largest change is for the system peak day, September 27, 2010, where the revised value is 4 MW above the originally reported value. The other changes range from 0 to 2 MW, all in the positive direction. All values in each column are reported to two significant digits, as was done for the 2010 evaluation. This leads some of the values in the “All” columns to appear too large or too small due to rounding, although they are not.

Table C-1: Originally Reported and Revised 2010 Ex Post Aggregate Impact Estimates (MW)

Date	Residential		Commercial		All	
	Originally Reported	Revised	Originally Reported	Revised	Originally Reported	Revised
15-Jul-10	11	12	4.7	4.4	16	16
16-Jul-10	15	16	5.2	4.9	21	21
17-Aug-10	12	13	4.7	4.4	16	17
18-Aug-10	15	17	5.2	4.9	20	22
19-Aug-10	13	14	4.9	4.6	17	19
23-Aug-10	13	15	4.7	4.4	18	19
24-Aug-10	13	15	4.9	4.7	18	20
25-Aug-10	11	13	4.8	4.5	16	18
27-Sep-10	26	29	6.8	6.5	32	36
28-Sep-10	13	15	5.3	5.0	18	20
29-Sep-10	10	12	4.9	4.7	15	17
Average	14	16	5.0	4.8	19	21