FREEMAN, SULLIVAN & CO.

A MEMBER OF THE FSC GROUP

2011 Load Impact Evaluation for Pacific Gas and Electric Company's SmartAC Program

June 1, 2012

Prepared for:
Pacific Gas and Electric Company
245 Market Street, 330 E
San Francisco, CA 94105

Prepared by:
Stephen George
Liz Hartmann
Mike 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

This evaluation documents the ex post and ex ante load impact analysis, methodology and results for PG&E's SmartAC™ program for residential and small/medium business (SMB) customers. The SmartAC™ program involves the installation of programmable communicating thermostats (PCTs) and/or switches in households and small/medium businesses with central air conditioning (CAC). When a SmartAC event is called, the control devices limit the duty cycles of CAC units or adjust thermostat temperature settings, thereby reducing demand.

There are three device types currently used by PG&E to control air conditioners and each has different functional capabilities. Understanding the differences in device functionality and operating modes is important because PG&E has taken a number of steps to increase load impacts from SmartAC by changing cycling strategies where possible.

LCR5000 and LCR5200 switches are control devices that attach directly to central air conditioning units. They control the duty cycle directly using one of several different algorithms.[1] The two algorithms relevant to this report are simple and TrueCycle2 cycling. A device controlled using 50% simple cycling will have its duty cycle limited to run no more than 50% of the time. It will behave normally any time when its duty cycle would run less than 50% of the time. A device controlled under 50% TrueCycle2 is limited to run no more than 50% of a baseline value. The baseline value is an adjusted average of the device's load during a number of pre-selected "learning days." The adjustment is done by scaling the baseline by the ratio of the load during the hour prior to the event and the average load during the same hour on learning days. The TrueCycle2 algorithm is therefore an attempt to limit units to use 50% of the load they would have otherwise used. Therefore, devices controlled using 50% TrueCycle2 should provide larger average load impacts under any situation where the devices would have been running at less than 100% duty cycle during the event. UtilityPro PCTs can be used to remotely adjust temperatures or can be used in the same manner as the switches described above. UtilityPro PCTs can operate under either simple or TrueCycle2 cycling. ExpressStat PCTs can adjust temperatures or implement simple cycling, but do not have an adaptive cycling capability.

Duty cycle control, not temperature control, was used exclusively during the 2011 season for all control devices. The standard cycling strategy for 2011 was 50% TrueCycle2 for residential customers with switches or UtilityPro PCTs and 33% for SMB participants. Residential customers with the older ExpressStat PCTs were subject to 50% simple cycling.

SmartAC events can be called for testing purposes or under emergency or in anticipation of emergency conditions between May 1st and October 31st, up to 6 hours or less in each event, for a maximum of 100 hours per season. No system-wide events were called in 2011. Several test events were called for subsets of the population as discussed in detail throughout this report. Residential customer enrollment at the end of summer 2011 was roughly 146,600 accounts and SMB customer enrollment was roughly 6,200 accounts. There were 161,484 active installed devices among residential accounts and 11,533 devices for SMB accounts.

[1] Duty cycle is the fraction of time that an air conditioning compressor is active. Duty cycles vary significantly with temperature. The hotter the temperature across the hours of a day, the longer the duty cycle.

Load impact estimates for both residential and SMB customers increased substantially in 2011 compared with 2010. Residential ex post impacts increased an average of 70% and SMB ex post impacts increased an average of 120%.[2] These increases are primarily due to the adoption of different and improved control strategies that PG&E implemented in 2011 in response to findings in the 2010 SmartAC evaluation. The effects of differences in weather, program population and customer behavior are small, as discussed in Appendix E. The operational changes were the adoption of different and improved control strategies for all devices in 2011.

It would have been difficult to claim that the significant improvements in load impacts were the result of the operational changes had PG&E not been committed to fully understanding the performance of the program. PG&E's willingness to fund the collection of control device internal logger data, to implement multiple test events based on sound experimental designs in both 2010 and 2011, and to undertake side-by-side testing of different control strategies collectively made it possible to determine with near certainty that the improvements are the result of operational changes. These efforts far exceed the minimum evaluation requirements for the program. PG&E's commitment to evaluation is paying dividends in the form of a large and well-understood increase in program performance.

1.1 Residential SmartAC Ex Post Load Impact Summary

Table 1-1 shows the average impact per device for each load research event along with average temperature over the event period for the residential SmartAC population. The first three events began at 1 PM while the final four began at 4 PM. All seven events ended at 6 PM. Table 1-1 only shows impacts for the last two hours of the five-hour events so that the seven events are comparable. The largest impact occurred on June 21, which had an estimated impact of 0.77 kW per device. Not coincidentally, June 21 was the hottest event day of the summer. The average impact of 0.50 kW represents a 22% reduction in whole-house load. The percent reduction across event days ranged from a low of 17% to a high of 26%.

The overall average event effect of 0.50 kW with an average event temperature of 94 degrees is much greater than the average effects seen in the 2008, 2009 and 2010 ex post evaluations. The average event temperature for 2008, 2009 and 2010 was 93 degrees. The average event impacts were 0.19 kW, 0.26 kW and 0.22 kW, respectively.[3] The evidence strongly suggests that the substantial increases in load reduction observed in 2011 are primarily due to changes in the control device strategy in 2011. More detail on the source of the difference between 2011 and other years is provided in Appendix E.

[2] These increases were calculated by comparing average impacts only for event hours that overlapped between the two years. For residential, this was 4 to 6 PM and for non-residential, it was 3 to 5 PM.

[3] The impact estimates for prior years represent the average across an entire event period and the hours for each event varied. The 2011 value of 0.5 kW represents the same two hours of the day, from 4 PM to 6 PM. As such, these yearly averages are not completely comparable.

Table 1-1: Average Residential per Device Reference Loads, Impacts and Temperatures from 4 to 6 PM on 2011 Event Days

Event Date	Event Hours	Average Reference Load[2] (kW)	Average Event Impact (kW)	Percent Impact[4]	Average Temperature 4–6 PM (°F)
6/15/2011	4-6 PM	1.97	0.34	17%	92
6/21/2011	4-6 PM	2.95	0.77	26%	99
6/22/2011	4-6 PM	2.66	0.57	21%	92
8/24/2011	4-6 PM	2.34	0.59	25%	92
9/6/2011	4-6 PM	1.99	0.39	19%	94
9/7/2011[5]	4-6 PM	2.18	0.48	22%	95
9/8/2011	4-6 PM	1.98	0.39	20%	90
Average	**4-6 PM**	**2.30**	**0.50**	**22%**	**94**

1.2 Residential SmartAC and Dually-Enrolled Customer Ex Ante Load Impact Summary

Table 1-2 shows the average ex ante impact estimates for the residential SmartAC population over the resource adequacy window of 1-6 PM. For the 1-in-2 weather year, the highest estimated impact is on the July peak day, with an average impact of 100 MW and a peak hourly impact of 126 MW. The July peak day also shows the highest impacts for the 1-in-10 weather year. The mean impact over the five-hour event is almost 128 MW and the peak hourly impact is 160 MW.

[4] Impacts are a percentage of whole-building load for residential customers. This must be kept in mind when comparing percent impacts to those in previous evaluations, which were calculated as a percentage of CAC load.

[5] Excludes about 10,000 customers called for a substation test event. Results for those customers have been reported in a separate document.

Table 1-2:
2012 Residential SmartAC Load Impact Estimates
By Weather Year and Day Type
(Event Period 1-6 PM)

Weather Year	Day Type	Mean Hourly Per Customer Impact (kW)	Max. Hourly Per Customer Impact (kW)	Aggregate Mean Hourly Impact (MW)	Aggregate Max Hourly Impact (MW)
1-in-2	Typical Event Day	0.46	0.58	71.9	90.1
	May Peak Day	0.29	0.37	45.9	57.5
	June Peak Day	0.36	0.45	56.5	70.9
	July Peak Day	0.65	0.81	100.7	126.3
	August Peak Day	0.46	0.59	72.2	90.5
	September Peak Day	0.48	0.61	74.3	93.1
	October Peak Day	0.16	0.2	24.5	30.7
1-in-10	Typical Event Day	0.68	0.86	105.8	132.6
	May Peak Day	0.57	0.71	89.0	111.5
	June Peak Day	0.62	0.78	97.0	121.6
	July Peak Day	0.82	1.02	127.5	159.8
	August Peak Day	0.73	0.92	113.6	142.4
	September Peak Day	0.56	0.7	85.8	107.5
	October Peak Day	0.46	0.58	70.7	88.5

Per customer ex ante load impacts are shown for dually-enrolled customers in **Error! Reference source not found.**3. For the 1-in-2 weather year, the highest mean aggregate impact is on the July peak day, with an estimated aggregate impact of 3.9 MW. For the 1-in-10 weather year, August provides the largest estimated impact – 5.6 MW. Although the per customer average is slightly lower in August than in July for the 1-in-10 weather year, the projected enrollment increases between July and August enough to make up for that difference.

Table 1-3:
2012 Dually-Enrolled Load Impact Estimates
By Weather Year and Day Type
(Event Period 1-6 PM)

Weather Year	Day Type	Mean Hourly Per Customer Impact (kW)	Max. Hourly Per Customer Impact (kW)	Aggregate Mean Hourly Impact (MW)	Aggregate Max Hourly Impact (MW)
1-in-2	Typical Event Day	0.56	0.69	3.4	4.2
	May Peak Day	0.44	0.53	1.8	2.1
	June Peak Day	0.43	0.52	2.0	2.4
	July Peak Day	0.72	0.88	3.9	4.8
	August Peak Day	0.54	0.66	3.3	4.0
	September Peak Day	0.56	0.69	3.8	4.7
	October Peak Day	0.17	0.21	1.3	1.6
1-in-10	Typical Event Day	0.78	0.96	4.8	5.9
	May Peak Day	0.67	0.82	2.7	3.3
	June Peak Day	0.67	0.82	3.2	3.9
	July Peak Day	0.94	1.15	5.1	6.2
	August Peak Day	0.91	1.11	5.6	6.8
	September Peak Day	0.6	0.74	4.1	5.0
	October Peak Day	0.5	0.62	3.8	4.7

1.3 SMB SmartAC Ex Post Load Impact Summary

Table 1-4 shows the average impact per device for each load research event along with average temperature from 3 to 5 PM for the SMB SmartAC population. The first three events ran from 1 to 6 PM while the final six occurred between 3 and 5 PM. Table 1-2 only shows impacts from 3 to 5 PM for each event day so that the nine events are comparable. The largest impact occurred on June 22, which had an estimated impact of 0.44 kW per CAC unit.[6] The average impact across all events is 0.29 kW per CAC unit, or roughly 17% of CAC load. This is almost three times greater than the average impact observed during 2010. These results are not completely comparable because 2010 events were cooler on average and lasted from 2 to 6 PM. However, despite these differences, it is clear that program performance improved substantially in 2011. This is discussed further in the SMB Ex Post section below and in Appendix E.

[6] It should be kept in the mind that SMB cycling is 33% whereas residential cycling is 50%, which partially explains why average SMB impacts are somewhat less than residential impacts.

Table 1-4: Average SMB per CAC Unit Reference Loads, Impacts and Temperatures During Event Hours

Event Date	Event Hours	Average Reference Load (kW)	Average Event Impact (kW)	Percent of CAC Load	Average Temperature (°F)
6/21/2011	3-5 PM	2.06	0.27	14%	96
6/22/2011	3-5 PM	1.92	0.44	23%	89
6/23/2011	3-5 PM	1.41	0.20	14%	83
8/24/2011	3-5 PM	1.60	0.23	14%	89
9/6/2011	3-5 PM	1.67	0.25	15%	90
9/7/2011	3-5 PM	1.69	0.28	16%	94
9/8/2011	3-5 PM	1.44	0.20	14%	87
9/20/2011	3-5 PM	1.87	0.43	23%	94
9/21/2011	3-5 PM	1.72	0.34	19%	92
Average	**3-5 PM**	**1.71**	**0.29**	**17%**	**90**

1.4 SMB SmartAC Ex Ante Load Impact Summary

Table 1-5 shows the average ex ante impact estimates for the non-residential SmartAC population for the resource adequacy window of 1-6 PM. For the 1-in-2 weather year, the largest impacts are projected to occur in July with a mean aggregate impact of 4.2 MW. The maximum aggregate impact in any hour for the July peak is 4.9 MW. For the 1-in-10 weather year, average aggregate impacts reach 4.8 MW for the July peak day and the aggregate maximum hourly impact is 5.6 MW.

Table 1-5:
2012 SMB SmartAC Load Impact Estimates
By Weather Year and Day Type
(Event Period 1 to 6 PM)

Weather Year	Day Type	Mean Hourly Per Customer Impact	Max. Hourly Per Customer Impact	Aggregate Mean Hourly Impact	Aggregate Max Hourly Impact
		(kW)	(kW)	(MW)	(MW)
1-in-2	Typical Event Day	0.56	0.65	3.3	3.9
	May Peak Day	0.37	0.44	2.3	2.8
	June Peak Day	0.47	0.56	2.9	3.4
	July Peak Day	0.69	0.81	4.2	4.9
	August Peak Day	0.55	0.65	3.2	3.8
	September Peak Day	0.51	0.61	3.0	3.5
	October Peak Day	0.32	0.38	1.8	2.2
1-in-10	Typical Event Day	0.72	0.85	4.3	5.0
	May Peak Day	0.63	0.74	3.9	4.6
	June Peak Day	0.66	0.78	4.0	4.8
	July Peak Day	0.79	0.93	4.8	5.6
	August Peak Day	0.76	0.88	4.5	5.2
	September Peak Day	0.61	0.72	3.5	4.2
	October Peak Day	0.50	0.59	2.8	3.3

1.5 Recommendations

Between the 2010 and 2011 seasons, PG&E implemented several changes in the program that led to substantial increases in load impacts. There are still opportunities for further increases in load impacts and cost-effectiveness within the constraints of the SmartAC tariff, primarily through further improvements in device communication.

The findings regarding the distribution of event impacts over residential customers suggest that there is also substantial room to improve the average load impacts of new customers (and thereby the program cost-effectiveness) through targeting based on average summer usage or a similar metric. Concentrating recruitment on high users would likely lead to a lower number of new recruits since the target population would be smaller, but these recruits would provide much higher average impacts per customer. One way this might be done is by focusing recruitment efforts on customers in the top 30% of average summer usage. FSC's analysis suggests that this would improve average residential per customer load impacts by about 50%.

Findings on multi-device customers suggest different strategies for residential and SMB recruiting. For residential customers, multi-device premises appear to provide no more load impact per premise than single-device premises. This suggests that they are less cost-effective because purchasing and installing more than one control device costs more but results in no more impacts. For SMB customers, multi-device premises appear to provide roughly the same amount of impact per device as single device premises. For example, customers with two devices show impacts about twice as large as customers with one device. This suggests that for the SMB segment, multi-device premises are more cost-effective because they require only one trip by a technician for installation and for any future maintenance. A possible explanation for the different finding for residential and SMB customers is that SMB customers are more likely to run multiple CAC units at once than are residential customers. If a large store or warehouse has multiple CAC units, they probably need to run all units at the same time to keep the space within the desired temperature range. For residential customers, however, it is more likely that the second or third CAC units are auxiliary coolers – either for parts of the house that are seldom used or for separate buildings, such as guest houses.

Exclusively using the TrueCycle2 control strategy appears to have paid off in higher load impacts, apparently due to improved signal reception for PCTs and increased cycling time for switches. Additionally, a 50% TrueCycle2 control strategy tends to provide larger impacts for PCTs, assuming perfect signal reception, than does a 2-1-1 temperature setback strategy. The same seems to hold true for a 33% TrueCycle2 strategy versus a 1-1-1 temperature setback strategy. These latter conclusions are based only on the fact that program performance appears to have improved by more than can be explained by improved communication and the improvement in switch performance alone. Given that, in a future evaluation it may be worth PG&E's effort to directly test temperature setback strategies against TrueCycle2 in the same way that TrueCycle2 was tested against simple cycling in 2011.

The control group-based experimental design using SmartMeter data for residential customers led to a streamlined evaluation process and more verifiably accurate results than in previous evaluations. We recommend that this design should be retained for future evaluations. There are several variations that could be added to the design, including:

- Testing the system more locally, such as at the substation level;
- Avoiding potential customer discomfort by using test events of one or two hours duration (although the post-event survey done in 2011 showed that customer discomfort is minimal during two-hour events);
- Testing load impacts for multi-device households; and
- Testing load impacts for TrueCycle2 side-by-side with a temperature ramping strategy.

For SMB customers, the 2012 evaluation should also be conducted using a control group-based experimental design and SmartMeter data rather than end-use logger data. With careful implementation and the use of the entire SMB population, this strategy should produce highly accurate impact estimates both cheaper and faster than the current logger-based method.

Although it is somewhat expensive, PG&E should continue to monitor the communication success of its control devices by accessing device log data for a sample of customers. It appears that improvements in device communication success rates are still possible and there is no more effective way to confirm control signal receipt and addressing than through field visits to obtain device logs.

2 Overview of SmartAC Program and Evaluation Plan

PG&E's SmartAC™ program involves the installation of programmable communicating thermostats (PCTs) and/or direct load control switches (switches) in households and small/medium businesses with central (or packaged) air conditioning (CAC). The control devices allow CAC equipment to be cycled or thermostats to be adjusted when an event is triggered, thereby reducing energy demand associated with AC load. The standard cycling strategy for switches is 50% for residential customers and 33% for non-residential participants. The standard temperature adjustment for customers with PCTs is 2 degrees in the first event hour and 1 degree in each of the following two event hours for residential customers. For non-residential customers, the standard operational strategy involves a 1-degree adjustment in each of the first three event hours. SmartAC events can only be called under emergency or in anticipation of emergency conditions, or for testing purposes, between May 1 and October 31 and for an event period of 6 hours or less for no more than 100 hours per season.

Table 2-1 shows the number of enrolled control devices as of September 8, 2011 by customer type, device type and local capacity area (LCA). It is important to distinguish between enrolled customers and enrolled devices, as many customers, especially SMB customers, have multiple CAC units and, therefore, multiple control devices. Some accounts even have both kinds of control device associated with separate CAC units. Residential customer enrollment at the end of summer 2011 was roughly 146,600 accounts and SMB customer enrollment was roughly 6,200 accounts. There were 161,484 active installed devices among residential accounts and 11,533 devices for SMB accounts.

As seen in Table 2-1, the majority of SmartAC devices are associated with residential households. Indeed, the residential segment comprises 93% of all SmartAC devices, 99% of switches and 71% of PCTs. The distribution of device types is almost completely reversed for SMB accounts, where 90% of all devices are PCTS. SMB accounts have roughly 1.9 devices per customer, whereas residential accounts average 1.1 devices per customer.

Since the 2010 program year evaluation, the number of residential devices has grown by 11%. The number of devices per customer has remained stable for both residential and SMB customers. Although the number of SMB customers is still much smaller than the number of residential customers, it has grown substantially since 2009 due to increased marketing in that segment. In September 2009, there were approximately 1,000 SMB SmartAC customers and in July 2010, there were roughly 3,400 accounts. By September 2011, there were over 6,000 SMB SmartAC accounts and more than 11,000 active, installed control devices.

As was true the last two years, the Greater Bay Area LCA had the largest share of SmartAC devices followed by Greater Fresno and "Other."

Table 2-1: SmartAC Active Control Devices September 8, 2011 Event Day

Customer Class	Local Capacity Area	PCTs	Switches	Total Devices
SMB	Greater Bay Area	3,633	269	3,902
	Greater Fresno	1632	232	1,864
	Kern	410	21	431
	Northern Coast	966	120	1,086
	Other	2,219	249	2,468
	Sierra	784	93	877
	Stockton	766	139	905
	Total	**10,410**	**1,123**	**11,533**
Residential	Greater Bay Area	7,939	48,959	56,898
	Greater Fresno	6,086	23,338	29,424
	Kern	1,761	3,978	5,739
	Northern Coast	1,483	8,218	9,701
	Other	4,255	22,016	26,271
	Sierra	2,199	17,252	19,451
	Stockton	2,250	11,750	14,000
	Total	**25,973**	**135,511**	**161,484**

2.1 SmartAC Analytical Overview

The basic analytical requirement for SmartAC impact evaluation is the estimation of a reference load during event periods. This requires some sort of experimental design. The experimental design for measuring load control impacts differed substantially in 2011 from the design in previous evaluations. Previous evaluations used a design in which test events were called for a sample of customers and reference loads were estimated for those customers based on loads observed for the same customers on non-event days. The 2011 design is based on the use of control groups for every test event. This is an improvement in design because it does not require the use of a statistical model to predict usage during event periods based on loads observed during non-event periods, which are often cooler. This issue was noted in the 2010 evaluation. There are typically not many days in a summer with temperatures high enough to test the SmartAC system under conditions similar to those when PG&E would call the program to ease a capacity constraint. In 2010, test events were called on every day of the summer in which system-wide average maximum temperature was above the low 90s when weighted over all SmartAC customers. This left little relevant load data observed under appropriate conditions in the absence of events, which meant that reference load models required unverifiable extrapolation.

In 2011, test events were also called on every day in which the system-wide average maximum temperature exceeded the low 90s. However, there was a control group for every test event. This allows for much more precise and verifiable reference load estimates.

The residential experimental design was based on the use of SmartMeter interval data to measure loads. The entire residential SmartAC population was divided into 10 groups. For each test event, between one and three groups would receive the event signal and the remaining groups were used as a control group for estimating reference loads. The large size of the residential population meant that each group consisted of roughly 14,000 households. With groups this large, reference loads and event impacts contain only a trivial amount of sampling variance. This produced the most accurate and precise ex post load impacts in any evaluation since the inception of the SmartAC program. This design also allowed for direct testing of control strategies and different possible event timing options under identical event conditions.

Many of the advantages of this design would not have been realized without PG&E's ability to extract large volumes of SmartMeter interval data on short notice. Due to the ease of data acquisition, residential ex post impact estimates were available within 48 hours after each event of the summer. This would not have been possible at most utilities where the data acquisition process is much longer and more difficult.

The SMB experimental design was based on CAC logger data for a sample of customers because the SmartMeter-based evaluation technique had not been well-validated for SMB customers. SmartAC has a substantially smaller SMB population and CAC load is a smaller fraction of whole-building load than it is for residential customers. Moreover, in the 2010 evaluation, SMB load impact estimates were quite low due to operational problems in the program. This could have been problematic for isolating load impacts using SmartMeter data. Therefore, for the sake of retaining impact estimation accuracy, CAC logger data was used. The control group strategy was implemented in the SMB population by dividing the sample of customers with installed loggers into two equal sized, random groups. For each test event, one group was called and the other was held back as a control group. The groups alternated between receiving an event signal and being used as a control group.

The choice of control strategies has been an important issue over the past two program years. Major progress has been made in determining which type of strategy works best for providing reliable load impacts while remaining within the rules set by the SmartAC tariff and not subjecting customers to significant discomfort. Major improvements in load impacts were achieved in 2011 by using direct duty cycle control rather than temperature setback for PCTs. With temperature setback control, only one control signal is sent to each device whereas other cycling strategies allow multiple signaling during the event period. In this way, even if one communication is missed, the next one may still get through. Further improvements were made in 2011 through the use of a more sophisticated duty cycle control algorithm than was used previously. These two changes account for a 70% increase in residential load impacts and a 120% increase in SMB load impacts. The control strategies used are shown in Table 2-2 along with associated problems and improvements. These issues are discussed extensively throughout the evaluation.

Table 2-2: SmartAC 2010 and 2011 Control Strategies

Segment	Device	2010 Control Strategy	Problems in 2010	2011 Control Strategy	2011 Result
Residential[7]	Switches	50% True Cycle	Defaulted to simple cycling	50% TrueCycle2	No default to simple cycling
	Utility Pros	2-1-1 temperature setback	Poor communication	50% TrueCycle2	37% improvement in communication
Commercial	Switches	33% True Cycle	Poor communication and defaulted to simple cycling	33% TrueCycle2	22% improvement in communication
	Utility Pros	1-1-1 temperature setback	Poor communication	33% TrueCycle2	52% improvement in communication

In addition to estimating load impacts, FSC also analyzed the degree to which SmartAC control devices received the signal to reduce load. This was identified as a major problem in the 2010 SmartAC evaluation, which led to the change in control strategies noted above. Communication in the SMB population improved substantially from an average success rate of 45% in 2010 to 64% in 2011. Communication changes in the residential population are more difficult to assess because a somewhat different population was sampled in 2011 than in 2010. This is discussed further in Section 4.3.

FSC also conducted a survey of SmartAC customers following a test event to assess the degree to which customers felt discomfort due to the event. A control survey on SmartAC customers who did not experience the event was also conducted. Results from this survey showed that customers who were called for a two-hour event did not report more discomfort than customers who were not called for an event.

2.2 Report Organization

The remainder of this evaluation is organized as follows. Section 3 describes the load research sample design and the experimental operation of the sample that generates the end use load data used in the SMB load impact analysis. It also summarizes some of the control failures and other issues that were identified during the research. Section 4 covers the residential portion of the evaluation, discussing both methodology and results. Section 5 mirrors Section 4, but covers the SMB sample. Following the main body of the paper are four appendices that go into greater technical detail on a number of topics.

[7] The table does not include a fairly small number of residential Express Stats, which are an older type of PCT that are no longer installed.

3 M&E Sample and Experimental Design

This section details the recruitment, addressing verification, logger installation and retrieval effort for the M&E sample of SMB customers. It also discusses the experimental designs implemented for residential and SMB customers.

3.1 M&E Recruitment and Logistics

This section describes the process of customer recruitment, logger installation and logger removal. Section 3.1.1 describes how and from where customers were recruited and the process of installing loggers. Section 3.1.2 covers the downloading of data from the loggers and their removal.

3.1.1 Customer Recruitment and Logger Installation

The experimental design for SMB customers required a recruited M&E sample. This sample had CAC load loggers installed on up to five CAC units per premise for the period June through September 2011. Phone recruitment of SMB customers took place in April 2011.

It was decided to install 700 CAC load loggers. Premises were eligible to have between one and five loggers installed per site. Both of these decisions were made with budget constraints in mind. The total number of SMB loggers was much higher than in the previous evaluation. This was done in order to develop a strong view of whether the changes implemented since the 2010 evaluation were effective. The increase in sample size was possible because there was no need for a residential logger sample. This freed up budget originally planned for residential loggers. The limit of five loggers per site was implemented in order to ensure that the sample covered a sufficiently broad range of sites. With up to 50 control devices at some SMB premises, with no limit on loggers per site, the sample could end up consisting of only a small number of sites.

Recruiting proceeded with an introduction letter sent to approximately 3,000 SMB SmartAC customers followed by a phone call. Phone calls were made until the target number of loggers was installed.

In order for the M&E sample to represent the SMB SmartAC population, the goal was for the distribution of SMB CAC units in the M&E sample to approximate the distribution of CAC units in the SmartAC population. To achieve this, the PG&E service territory was divided into four areas based on a combination of PG&E's climate zones and FSC's judgment about which areas experience weather similar enough to be aggregated. FSC divided the territory into four regions: climate zones R and S, the hot parts of zone X, and zone T plus the cool parts of zone X. The definition of the cool part of zone X is determined by a customer's assigned local capacity area and their nearest weather station. This definition is shown in Table 3-1. The way to read the table is that a customer is assigned to the cool parts of Zone X only if their premise is assigned to a combination of a given capacity area and one of the weather stations associated with that capacity area.

Table 3-1: The Cool Parts of Climate Zone X
(Defined as the Intersection of Given Capacity Areas and Weather Station Assignments)

Capacity Area	Weather Stations for that Capacity Area
Greater Bay Area	Concord, San Ramon
Other	Angels Camp, Auburn, Bakersfield, Chico, Marysville, Paso Robles, Red Bluff, Sacramento, Stockton

A quota was assigned to each region for phone recruiting. Once the quota was filled for a region, no more customers were recruited from that zone. Phone recruiters were provided a list of 3,600 SMB SmartAC customers to call.

The outcome of recruiting was that 42% of the devices were in zone T and the cooler parts of zone X, 28% were in zone S, 21% were in zone R and 9% were in the hot parts of zone X. This fairly closely matched the proportions in the SMB SmartAC population as of March 2011, as shown in Table 3-2.

Table 3-2: SMB SmartAC Devices by Climate Zone

Climate Zone	SmartAC Population		Logger Sample	
	# of Devices	% of Devices	# of Devices	% of Devices
Cool X and T	4,317	38%	289	42%
Hot X	995	9%	60	9%
R	2,734	24%	145	21%
S	3,410	30%	190	28%
Total	11,456	-	684	-

Following recruitment, the next step was to communicate with each of the M&E sample participants' control devices, instructing it to recognize signals aimed specifically at the M&E sample. In the jargon of the current load control contractor, this is known as setting the splinter for the device. This was important because the M&E sample would be called for numerous test events over the course of the summer, while the rest of the SMB SmartAC Program would not.

Logger installations began on April 11 and finished on June 7, 2011. An important step in the process was that after an initial round of recruiting, signals were sent out to set the splinter for all the control devices at the premises of recruited customers. Devices at these premises that were on CAC units and were not chosen to be logged were also re-addressed. This was to ensure that test events accurately simulated the load response of the entire premise under the conditions of a system-wide event. If non-logged units were not re-addressed, then there would be the possibility that those units would compensate during test events, affecting the observed loads of the controlled units. This was also necessary so that discomfort survey results would reflect the effect of a full SmartAC event.

Upon installation of data loggers, installers checked to see if the splinter had received the signal. For premises where more than a certain number of devices failed to receive the splinter signal, that premise was abandoned and a replacement premise was recruited for the M&E sample. If a site was to have a logger installed on one or two units, then the installer walked away if all the devices had the wrong splinter. If the site was to have three loggers installed, then the installer walked away if two units had the wrong splinter. For sites where four or five loggers were to be installed, installers walked away if three units had the wrong splinter. Loggers were not installed on units with bad splinters. This rule meant that some premises were included in the M&E sample even though not all the devices at that premise would be activated for all test events. This was a pragmatic decision made based on keeping installation costs and timing manageable.

Installers encountered several other situations where pursuing the logger installation was not appropriate and abandoned. In all these situations, the installer would contact FSC and request a replacement recruit from the same stratification cell as the abandoned premise. Table 3-3 shows a tally of the reasons for site abandonment.

Table 3-3: Summary of Reasons Why Initial Sample Recruits Were Abandoned

Reason for Abandonment	Number of Devices
Control device missing or broken	34
With a broken or missing CAC (no probability of responding to a DLC radio signal)	10
Actually a residential premise	22
Customer changed their mind at installation	6
Technician could not access to the unit either due to logistical or physical impediments	59
Wrong splinter	90

Because technicians did not install loggers on sites with broken or missing control devices or CAC units, the load impacts observed in the SMB sample are larger on a per device basis than those in the SMB population. As shown in Table 3-3, there were a total of 44 such cases (shown in the first two rows), out of a total of 854 units visited where the technicians could verify operation. Based on these values, a correction factor of 5% (44/854) is subtracted from load impacts calculated from the SMB sample. This correction factor is substantially lower than in 2010, which makes sense because the SMB SmartAC population has more than doubled since the previous evaluation. This means that there are many more newly recruited sites in the population. At these sites, there has been less time for the CAC or control device to break or be removed.

For each logger installation, the field technicians also collected a significant amount of information that was recorded on an On-Site Verification Form (OVF), as well as through a series of digital photographs. The OVF included:

- Customer name, address and phone number;
- DLC control device serial number and operational status;
- Control signal strength being received by the control device;
- Information derived from interrogating the control device’s memory, such as all addressing information, most importantly, the “splinter” address assigned to that device;
- AC logger serial number;

- Make, model and serial number of the CAC unit itself; and
- The CAC unit's kW and kVar values as measured by the technician.

3.1.2 Logger Retrieval and Data Downloading

Logger retrieval took place in mid-October. At that time, technicians confirmed that the logger was still functioning. Technicians also retrieved internal data from the control devices themselves, including information on whether the device was controlled during any particular hour of the summer. The total number of devices installed initially was 720, and the total number of devices used in analysis was 679. Of the 720 devices that were successfully installed, 28 devices did not have data at the time of retrieval due to a variety of reasons, including:

- Logger was missing;
- Thermostat was removed/broken/replaced;
- CAC unit was not working; or
- Technician was unable to read logger.

Of the 692 loggers that had recorded CAC load data, 6 had to be excluded because the CACs they were installed on were not included in PG&E's SmartAC database,[8] and 7 loggers did not have any data on event days due to logger malfunction (even though they did have data at other times). The total number of premises initially installed was 374 and the total number used in analysis was 367.

3.2 SMB Experimental Design and Operations

The SMB test event protocol called for test events to always be called when the temperature was expected to be above 95°F based on a combined temperature forecast for Fresno, Concord and Sacramento and to be considered for test events if the temperature was expected to be above 85°F. The protocol initially called for each test event to be from 1 PM to 6 PM and to use 33% TrueCycle2 cycling for all devices in the SMB sample. For each event day, devices at half the premises in the sample would be activated for an event while the CAC units at the other half of the premises would provide reference load. The only change to these protocols that took place was that event length was shortened starting at the end of June over concerns about customer comfort.

3.3 Residential Experimental Design and Operations

The residential test event protocol called for test events to always be called when the temperature was expected to be above 95°F based on the same average forecast for SMB customers and to be considered for a test event if the temperature was expected to be above 85°F. The protocol initially called for test events lasting from 1 PM to 6 PM. This was shortened to 4 PM to 6 PM for events after late June due to concerns about customer discomfort. On each event day, certain groups of customers would be activated for events, while the remaining groups would provide reference load.

---

8 In almost all cases, the technician would install loggers on a specific, pre-determined CAC unit. In some cases, however, they installed on other units at the same site due to access issues. That some of these units are not included in PG&E's database indicates that either the customer had new CACs installed since joining the program or that they had CACs that they did not include when they signed up.

Customers were divided into 10 random groups based on the last digit in the serial number of their control device. Each control device was subject to exactly one test event during the summer. These occurred over the course of seven days. On each of those days, one group was activated using the standard SmartAC test event control strategies for 2011. These were 50% TrueCycle2 for switches and UtilityPro PCTs and 50% simple cycling for ExpressStat PCTs. On two of the days – June 21 and June 22 – more than one group of customers was called on the same day for side-by-side testing. On June 21, one randomized group of customers with switches was operated using 50% simple cycling and another group of customers including all three device types was operated using the normal control strategies but only for the hours 4 to 6 PM. On June 22, the same simple cycle test was repeated using a different group of customers with switches. Results from these tests are discussed in the residential ex post results section.

3.4 Control Device Success Rates

The load control switches and PCTs used to activate events have internal data loggers that keep track of when the device received an event signal and record load shed minutes (how many minutes per hour the device operated to curtail load). As part of logger retrieval for the SMB sample, technicians downloaded the internal data logs from each device.

For residential customers, communication verification was performed by visiting and downloading data logs from roughly 300 control devices at the premises of customers who were dually-enrolled in PG&E's SmartAC and SmartRate program. The SmartRate program is a critical peak price program that subjects customers to roughly 15, 5-hour events each summer during which the price of electricity increases by about $0.60/kWh. Dually-enrolled customers have their CAC units controlled during these events, just as SmartAC customers do under SmartAC events. This group of customers was chosen as opposed to SmartAC customers not on SmartRate because the regular SmartAC customers were only subject to one event each during 2011, which would provide little data for evaluating communication.

Ideally, each device would receive each event signal. For switches and PCTs operating under the TrueCycle2 algorithm, shed minutes should be greater than or equal to 30 minutes per hour for residential and 20 minutes per hour for SMB devices.

A variety of issues can lead to a device not receiving a signal, but the reasons can be divided into two main categories. First, a device might not receive a signal because the device itself has the incorrect address, perhaps because it had difficulty in receiving the address at the time of installation due to paging system communication issues. Second, a device might not receive a signal because something blocked the signal from getting through, such as a thick wall. The second issue is thought to affect PCTs more frequently because they are located indoors while switches are located outdoors.

Table 3-4 shows the number of device loggers that were successfully downloaded and that had control data for at least one event. The sample of device loggers is smaller than the M&E sample count because control data does not appear if a customer does not have their thermostat in cool mode or if it was powered off entirely. This issue was also noted in the 2010 evaluation. Additionally, this year 26 devices in the residential group were listed in the PG&E database of dually-enrolled customers, but not in Cooper's database, indicating that these customers were not sent event signals. Finally, another 47 residential devices of the dually-enrolled SmartAC/SmartRate customers are listed in both PG&E's and Cooper's database but somehow responded to the SmartAC test event control command instead of the SmartRate event days. Both of these groups were excluded from the analysis.

Table 3-4: Number of Devices with Valid Internal Data during at Least One Event

Device Type	Residential		SMB	
	Total Downloaded	Total with Control Data	Total Downloaded	Total with Control Data
Switch	168	100	53	51
Utility Pro	117	83	598	555
Express Stat	0	0	0	0
Total	285	183	651	606

The remainder of the section is divided between switches and PCTs due to an important difference between the two: for switches, the data logs used to analyze the effectiveness of control signals record shed minutes whether the CAC unit is off or on, but for PCTs, this is not true.

If the PCT is not in cooling mode, then it will record no evidence of the event. However, there is a way to determine whether the PCT was in cooling mode during each event. The PCT records the temperature, the temperature set-point it is programmed to, and the number of minutes the CAC runs each hour. If the PCT shows that the temperature is above the set point but the CAC does not run during a given hour, then it can be inferred that the PCT is not in cooling mode.[9] Therefore, for PCTs, a distinction is made between devices that are not in cooling mode and devices that are in cooling mode and fail to receive a signal. Both of these situations lead to zero load impact for the device, but for different reasons and only the second situation reflects a communication failure.

A PCT that is not in cooling mode even on hot days suggests that the CAC has little potential to provide load reduction in the program. On the other hand, a PCT in cooling mode that fails to respond to an event is only an operational problem that can be addressed. Understanding the number of customers, and what types of customers that are not running CAC on hot days may have an impact on program recruitment.

Table 3-5 shows the number of switches with data for each event and the percentage of switches successfully controlled for residential and SMB devices. On average, residential switches among dually-enrolled SmartRate SmartAC customers successfully received event signals 80% of the time. This is less than was found in the 2010 evaluation for residential SmartAC switches, where a 97% success rate was observed. It is not currently clear why the rate is lower in this group.

On average, SMB switches successfully received event signals 60% of the time. This is a significant improvement over the 50% success rate for SMB switches in the 2010 evaluation. It is likely due to

[9]This issue can also be addressed to some degree by examining CAC logger data, which was collected for the SMB sample. This is an imprecise exercise though because the CAC loggers pick up all load to the unit, including the fan. These loads can be present even when the device is not in cooling mode and therefore not receiving event signals.

the change in control strategy from temperature setback to TrueCycle2. The signal for TrueCycle2 is sent every half hour during an event, while the temperature setback signal is sent only once. This means that devices are more likely to receive the TrueCycle2 signal.

The number of devices with data varies over events. On the commercial side, this is primarily due to the fact that the experimental design entailed controlling half the group for each event, which is why the number of devices with data alternates between about 19 and 31 for each event. For residential switches, it is not clear why the number of devices with data varies over events. The data logs have gaps for unknown reasons.

Table 3-5: Percentage of Switches Controlled and the Number of Devices with Data for Each Event

Event Date	Residential		Commercial	
	Percentage Controlled	N	Percentage Controlled	N
21-Jun-11	87	98	41	19
22-Jun-11	81	97	77	31
23-Jun-11	N/A	N/A	42	19
5-Jul-11	77	99	N/A	N/A
6-Jul-11	78	99	N/A	N/A
28-Jul-11	78	99	N/A	N/A
29-Jul-11	79	97	N/A	N/A
17-Aug-11	79	98	N/A	N/A
18-Aug-11	74	99	N/A	N/A
23-Aug-11	78	100	N/A	N/A
24-Aug-11	N/A	N/A	77	32
29-Aug-11	78	100	N/A	N/A
2-Sep-11	78	100	N/A	N/A
6-Sep-11	80	100	42	19
7-Sep-11	97	100	75	32
8-Sep-11	82	100	42	19
Average	80	99	60	24

Table 3-6 shows the percentage of PCTs that were in cooling mode during each event, as determined by the temperature, set-point and run time. It also shows the percentage of the PCTs in cooling mode that received an event signal for each event. The table shows that only 61% of residential CAC units were cooling during the event, while 76% of SMB CAC units were cooling. Of units that were cooling, 58% of residential units received event signals, on average, while 80% of SMB units did. Both of these values are significant improvements over the values observed for PCTs in 2010.

Investigation of device performance and communication system issues is ongoing for the 2012 season. Despite some remaining uncertainty about data interpretation, it is clear that the 2011 program achieved a significant improvement in communication.

Table 3-6: Percentage of PCTs Cooling, Percentage Controlled and the Number of Devices with Data for Each Event

Event Date	Residential			Commercial		
	N	Percentage Cooling	Percentage Controlled	N	Percentage Cooling	Percentage Controlled
21-Jun-11	57	57	76	174	82	83
22-Jun-11	45	66	80	176	76	85
23-Jun-11	N/A	N/A	N/A	182	72	82
5-Jul-11	52	63	0	N/A	N/A	N/A
6-Jul-11	53	60	24	N/A	N/A	N/A
28-Jul-11	64	59	79	N/A	N/A	N/A
29-Jul-11	59	65	73	N/A	N/A	N/A
17-Aug-11	60	51	13	N/A	N/A	N/A
18-Aug-11	12	60	0	N/A	N/A	N/A
23-Aug-11	59	57	80	N/A	N/A	N/A
24-Aug-11	N/A	N/A	N/A	223	70	81
29-Aug-11	58	63	58	N/A	N/A	N/A
2-Sep-11	59	64	83	N/A	N/A	N/A
6-Sep-11	66	61	87	200	77	69
7-Sep-11	62	62	68	198	75	81
8-Sep-11	57	64	77	192	80	66
20-Sep-11	58	58	77	237	74	82
21-Sep-11	N/A	N/A	N/A	202	80	82
Average	55	61	58	198	76	80

4 Residential Impact Analysis

The 2011 residential SmartAC program produced average ex post impacts per device of 0.50 kW during the hours 4 to 6 PM.[10] This represents a substantial improvement over the impacts observed in 2010.

The source of this improvement is the change in control strategies that occurred for both major control device types between 2010 and 2011. Switch control devices were shifted from the 50% True Cycle algorithm in 2010 to the 50% TrueCycle2 algorithm in 2011. PCTs were shifted from a 2-1-1 temperature setback strategy in 2010 to 50% simple cycling for Express Stats and 50% TrueCycle2 for Utility Pro PCTs in 2011. The 2-1-1 temperature set back is a temperature ramping strategy that increases the thermostat 2 degrees in the first event hour followed by an additional 1 degree in the next event hour and 1 more degree in the next hour (2, 1, 1). The improvement in switch performance may result primarily from actually implementing the 50% True Cycle algorithm in 2011, as opposed to the simple cycling that most, if not, all switches in 2010 defaulted to. . The improvement in PCT performance is due to differences between TrueCycle2 and the 2-1-1 temperature setback strategy used in 2010. TrueCycle2 sends repeated control messages (every half hour) versus just one control message when temperature set back was used. This meant that control devices were more likely to receive the control signal since it was repeated multiple times throughout the event.

This chapter is divided into six main sections. Section 4.1 describes the approach used in the ex post impact analysis, including a description of the experimental design used. The section is fairly brief due to the nature of the experiment and the simple ex post methodology that it allows. Section 4.2 provides an explanation of how the methods and results were validated.

Section 4.3 describes the ex post impact results for all seven 2011 event days. This section also includes results from the experimental testing of different load control strategies and event timing (side-by-side test events) and a discussion of the differences between 2010 and 2011 ex post impacts.

Section 4.4 explains the methodology used in the ex ante impact analysis for both residential SmartAC and dually-enrolled SmartAC and SmartRate customers. Sections 4.5 and 4.6 present the ex ante results for residential SmartAC customers and dually-enrolled customers, respectively.

There are several technical appendices to this document that are relevant to the ex post evaluation. These are:

- Appendix A discusses the methodology for predicting residential whole-building loads and snap-backs that were used in the load impact tables that accompany this evaluation;
- Appendix B has the same discussion as Appendix A but for SMB loads and snap-backs;
- Appendix C discusses the results of the side-by-side testing of different control strategies and event timing;
- Appendix D provides an ex post analysis of Target Cycle, a control device algorithm that allows the program operator to select a numerical demand reduction target, which the algorithm aims for;

10 These are the only hours that all events covered.

- Appendix E provides a more detailed discussion of the evidence regarding the change in program performance between 2010 and 2011. It also provides a detailed discussion of the differences in methodology between the 2010 and 2011 evaluations; and
- Appendix F discusses the change in temperature experienced by customers with PCTs during events.

4.1 Residential SmartAC Ex Post Methodology

The methods used in the 2011 SmartAC residential evaluation greatly differ from prior years. A joint effort among FSC, PG&E and Cooper (the SmartAC implementation contractor) led to the execution of a large-scale experimental design for estimating residential ex post load impacts. Because SmartMeter data was available for the vast majority of the residential SmartAC population, randomized experiments could be conducted on each event day.

For each of the seven test events during the summer of 2011, a random 10% sample of the residential SmartAC population was called for the event while the rest of the customers served as the control group. Ex post impacts were calculated by comparing the average event day usage in the control group to the average event day usage in the treated group. Sample sizes of several thousand customers for each event combined with an experimental design that removed all self-selection bias eliminated the need for more complex regression methods, as were used in previous evaluations. This design has the following advantages:

- Eliminating the significant expense and logistical problems associated with recruiting a load research sample, installing data loggers and retrieving data loggers;
- Highlighting a very important benefit of SmartMeters: simpler and more accurate measurement and evaluation;
- Substantially reducing sampling error and selection bias from ex post estimates because treatment and control groups are large (several thousand customers) and randomly drawn from the population, with no self-selection;[11]
- Eliminating the need for complicated, weather-based regression functions for estimating ex post load impacts. As compared to previous years, ex post load impact estimates are no longer dependent on functional form assumptions. Instead, impacts can be estimated using simple averages of loads;
- High accurate ex post event impacts available , within just a couple of days after an event occurs. This allows for immediate identification of operational issues that need to be addressed. It also enables the evaluation process to progress more quickly because there is no delay associated with getting end use data loggers out of the field.
- The large size of the treatment groups allow for highly accurate results to be calculated at the LCA level or even for an individual city. This could not be done using data loggers due to the high cost of each sample point; and
- Making possible direct side-by-side testing of the effects of different control strategies and different event timing, without confounding weather, time or population effects. That is, it allows for complete isolation of the difference in DR event effects due to different treatments, without other confounding variables.

[11] Previously, load could only be gathered from customers who volunteered to be part of a research sample, inducing possible selection bias.

The basic format of the experimental design is that of a randomized controlled trial (RCT). To perform an RCT with SmartAC events required the ability to activate control devices for large, randomly selected groups of customers while not activating the remainder of the SmartAC population. Cooper, the SmartAC implementer, relies on control device addressing to call events on different subsets of SmartAC customers. For this evaluation, device addressing was in place prior to the study because each control device receives an address based on the last digit of its serial number when it is manufactured.[12] This resulted in 10 random groups of devices – each group corresponding to a different final digit of the serial number (0-9). FSC's testing showed that this addressing scheme resulted in a valid randomization with no significant differences in population characteristics between any of the 10 random groups of devices. Table 4-1 shows that each randomized group has approximately the same proportion of customers in each LCA and that each group has roughly the same average usage.

Table 4-1: Comparison of Randomized Groups

Randomized Group	Greater Bay Area	Greater Fresno	Kern	Northern Coast	Other	Sierra	Stockton	Mean Hourly Usage (kW)	Mean Daily Usage (kW)
0	34%	19%	4%	6%	17%	11%	9%	1.3	31.9
1	33%	19%	4%	6%	17%	11%	9%	1.2	29.7
2	34%	19%	4%	6%	17%	11%	9%	1.3	31.9
3	33%	19%	4%	6%	17%	12%	9%	1.3	30.1
4	34%	19%	3%	6%	17%	12%	9%	1.3	30.1
5	33%	19%	4%	6%	17%	11%	9%	1.3	30.6
6	33%	19%	4%	6%	17%	11%	9%	1.3	30.6
7	33%	19%	4%	6%	17%	11%	10%	1.3	31.2
8	33%	19%	4%	6%	17%	11%	10%	1.3	32.0
9	33%	19%	4%	6%	17%	11%	9%	1.3	31.6

The use of pre-existing addressing was the result of a pragmatic decision. Ideally, the SmartAC population would have been divided into a greater number of smaller groups. This would have allowed for greater flexibility to call more events without affecting customer comfort too much. Moreover, randomized groups that were even one-third the size of the groups that were used would have still been valid to produce highly accurate impact estimates. Finally, if random groups had been assigned as part of the experiment then multi-device houses would have been more easily dealt with (see below). However, re-addressing the entire SmartAC population into 20-30 randomized groups would have been a time-intensive and logistically difficult process with the potential for many problems. It was decided that avoiding re-addressing was worth a small compromise in experimental design.

12 The exception to this was the group of Express Stat PCTs, which comprise about 6% of all customer control devices. This set of devices was re-addressed for the study based on the last digit of the device serial number.

On September 8, the final event day of 2011, there were approximately 163,000 installed residential SmartAC devices. These devices were installed at about 148,000 customers' homes. Due to the pre-existing addressing and the need for SmartMeter data from each participating customer, not all SmartAC customers are included in the design. There are almost 14,000 SmartAC customers with more than 1 control device in their homes (just under 10% of the population). These houses were omitted from the primary analysis. Over 90% of customers with multiple CAC units had control devices in different randomized groups, meaning that one control device could be called for an event while another device in the house would not. In a situation like this, the whole-house load impact would not represent the true effect of a SmartAC event on that household. For this reason, all customers with two or more devices were not included in the primary ex post analysis. Instead, a secondary analysis of these premises was undertaken separately, as described in Section 4.2.3. This analysis showed that multi-device premises provide the same average impact as single device premises. Therefore, the primary results, which are based only on single-device premises, can be applied to the entire population. To estimate impacts for the entire SmartAC population, the impact estimates per device are multiplied by the total number of devices in the population.

After excluding customers with multiple CAC units from the primary analysis, there were slightly fewer than 134,000 residential SmartAC customers by the last event of the summer. For 6% of these customers, interval data was not available. However, an analysis of customers with no interval data showed that they are distributed similarly throughout PG&E's territory as the full SmartAC population and, therefore, their exclusion would not bias the results.

On average, the interval data for about 125,000 customers was used to calculate ex post impacts on each event day. In the case of the first event and each of the final five events, one group of customers, based on device serial number, was called for the event and the remaining set of customers served as a control group. For the second and third test events, multiple groups of customers were called for events for the sake of side-by-side testing of different control strategies and timing. Side-by-side testing is discussed in Appendix C.

Ex post event impacts for each LCA are estimated for each hour of each event by taking the average load in the group that received the event and subtracting it from the average load in the larger group that did not receive the event. Impact estimates for the entire SmartAC population for each hour of each event are calculated by taking a weighted average of the impact estimates for each LCA, with weights determined by the number of devices in each LCA.

4.2 Validity of Residential Load Impact Estimates

The values in Table 4-1 above, in addition to the method for selecting test event groups and their large size make a strong prima facie case for the validity of the load impact estimates for residential customers. Substantially more evidence is available in the Excel-based load impact tables that accompany this document. As is typical for this evaluation, that file contains too many different individual tables to be reproduced in the primary document, but they constitute a necessary accompaniment in order to fully understand the results. In this case, the tables shows raw average loads for customers involved in each test event, both those whose loads were controlled and those whose loads were not. Based on examining those loads in the pre- and post-event hours for each event, it is clear that the experimental method produced nearly perfect reference loads for each test event. Even at the level of individual LCAs, reference loads are highly accurate in most cases.

4.3 Residential SmartAC Ex Post Load Impact Results

This section is broken into four parts. The first section presents the ex post load impacts for the seven primary test events called over the summer. The second part presents impacts by customer usage decile. The third section provides an explanation of methodology and presentation of impacts for customers with exactly two AC units. Finally, the fourth section discusses load impacts for customers dually-enrolled on SmartRate and SmartAC as compared to customers enrolled only on SmartAC and customers enrolled only on SmartRate.

4.3.1 SmartAC Primary Test Event Results

Table 4-2 shows the average impact per customer for each load research event along with average temperature over the event period for the residential SmartAC population. The first three events began at 1 PM and lasted until 6 PM. All subsequent events lasted from 4 to 6 PM. To make the results comparable, the table only includes data on hours 4 to 6 PM for the first three events. Table 4-3 shows the event averages for the first three events over the entire event period. The largest impact occurred on June 21, which had an estimated impact of 0.77 kW per customer. Not coincidentally, June 21 was the hottest event day. The average impact of 0.50 kW represents an average of 22% in whole house load reduction. The percent reduction across event days ranged from a low of 17% to a high of 26%.

The overall average event effect of 0.50 kW, with an average event temperature of 94 degrees, is much greater than the average effects seen in the 2008, 2009 and 2010 ex post results. The average event temperature for 2008, 2009 and 2010 was 93 degrees. The average event impacts were 0.19 kW, 0.26 kW and 0.22 kW, respectively. The evidence suggests that the substantial increases in load reduction observed in 2011 are due to changes in the control device strategy in 2011 (see Appendix E).

Table 4-2 also includes the aggregate event impacts from 4 to 6 PM on each event day. This number is the estimated impact actually seen on those days, based on the number of customers called for each event. June 21 and June 22 show much greater aggregate impacts than the other event days because side-by-side test events were called on those days, which means more than one group of customers had their AC's controlled. On the event days when only one group was called, aggregate impacts range from 5.2 MW to 9.6 MW.

Table 4-2: Average Residential per Device Reference Loads,[13] Impacts and Temperatures from 4 to 6 PM on 2011 Event Days

Event Date	Event Hours	Average Reference Load (kW)	Average Event Impact (kW)	Percent Impact	Aggregate Event Impact (MW)	Average Temperature (°F)
6/15/2011	4-6 PM	1.97	0.34	17%	5.2	92
6/21/2011	4-6 PM	2.95	0.77	26%	29.4	99
6/22/2011	4-6 PM	2.66	0.57	21%	17.3	92
8/24/2011	4-6 PM	2.34	0.59	25%	9.6	92
9/6/2011	4-6 PM	1.99	0.39	19%	6.3	94
9/7/2011*	4-6 PM	2.18	0.48	22%	7.8	95
9/8/2011	4-6 PM	1.98	0.39	20%	6.3	90
Average	**4-6 PM**	**2.30**	**0.50**	**22%**	**11.7**	**94**

*excludes about 10k customers called for the substation event
†Includes impacts from side-by-side tests (see Appendix C)

Table 4-3 shows the average event impacts for the five-hour events called in June 2011. Average impacts over the five-hour period were slightly lower than the average impacts over the last two hours of the event, because residential loads tend to be lower in the early afternoon and peak from 5 to 6 PM. The average impact over the three five-hour events was 0.46 kW while the average impact from 4 to 6 PM on those same three days was 0.56 kW. The average aggregate impact over the three five-hour events was 11.6 MW.

Table 4-3: Average Residential per Account Reference Loads, Impacts and Temperatures during Event Hours on June 2011 Event Days

Event Date	Event Hours	Average Reference Load (kW)	Average Event Impact (kW)	Percent Impact	Aggregate Event Impact (MW)	Average Temperature (°F)
6/15/2011	1-6 PM	1.71	0.27	16%	4.0	91
6/21/2011	1-6 PM	2.60	0.63	24%	20.2†	98
6/22/2011	1-6 PM	2.40	0.47	20%	10.5†	93
Average	**1-6 PM**	**2.24**	**0.46**	**20%**	**11.6**	**94**

†Includes impacts from side-by-side tests (see Appendix C)

Impacts can be broken down by type of control device. Table 4-4 shows the per premise impacts by device type for residential SmartAC customers. On average, switch customers provide nearly twice the impacts of PCT customers. Table 4-4 only shows impacts for the hours from 4 to 6 PM because all events covered those hours, and the result holds when comparing two-hour and five-hour events as well. Additionally, this difference is not due to systematic temperature or building size differences between device types. In fact, premises with PCTs tend to have somewhat higher reference loads than those with switches, indicating that the performance gap is even larger than the table indicates.

13 Reference loads are whole-building loads.

As is shown in the section on communication success rates above, PCTs have worse signal reception than switches. This probably accounts for much of the performance gap.

Table 4-4: Average Residential Impacts per Account by Device Type

Date	Event Hours	PCT	Switch
15-Jun-11	4-6 PM	0.27	0.36
21-Jun-11	4-6 PM	0.46	0.83
22-Jun-11	4-6 PM	0.39	0.61
24-Aug-11	4-6 PM	0.37	0.63
6-Sep-11	4-6 PM	0.26	0.41
7-Sep-11	4-6 PM	0.33	0.53
8-Sep-11	4-6 PM	0.18	0.43
Average	**4-6 PM**	**0.32**	**0.54**

Ex post load impacts in 2011 were about 73% higher than those in 2010. The evidence in Appendix E shows that this change is due to improvements in the control strategies for both types of control device. In 2010, switches were controlled using 50% True Cycle algorithmic cycling. However, further analysis showed that for all switches observed, the True Cycle algorithm defaulted to using simple cycling. This can occur for a several reasons which are not discussed here. PCTs in 2010 were controlled using a 2-1-1 temperature setback strategy which was subject to significant signal failure. In 2011, switches and Utility Pro PCTs were controlled using the 50% TrueCycle2 algorithm and Express Stat PCTs were controlled using 50% simple cycling.

The size of the improvement in the per-device load impacts from 2010 to 2011 is almost identical to the per-device difference between TrueCycle2 and simple cycling during the side-by-side test. This suggests that the performance improvement is attributed to the change in cycling algorithms. Additionally, it is quite plausible that a similar improvement in PCT performance would have resulted from a combination of the 25% improved signal reception due to the improved addressing efforts plus an additional improvement due to the performance of cycling over setback for devices that receive signals. A more in-depth discussion of these issues, along with evidence about alternative possible explanations of the change in performance, is presented in Appendix E. The concordance among the evidence presented is strong and consistent.

4.3.2 Distribution of Impacts Across Customers

Table 4-5 shows the average impact from 4 to 6 PM averaged across all seven event days by customer usage deciles. Customers were divided into deciles based on average monthly usage in June 2011. Customers in the lowest decile had an average monthly usage of 184 kW compared to 1,286 kW for customers in the highest decile of usage.[14] As would be expected, customers with higher average usage show greater absolute impacts. Customers in the greatest decile of usage provided impacts over four times greater than customers in the lowest decile of usage.

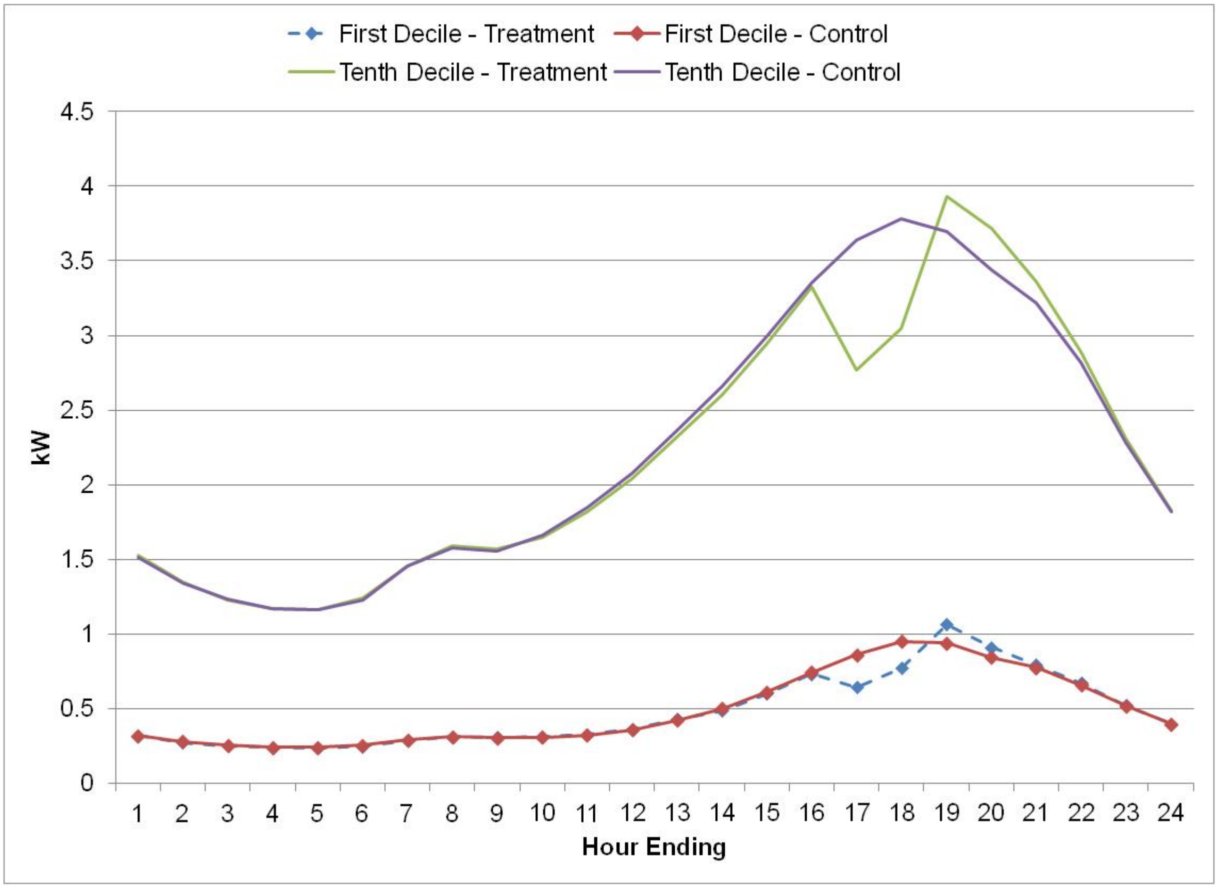
[14] Usage deciles could also be calculated using daily SmartMeter data instead of monthly billing data.

Table 4-5: Average Event Impacts by Usage Decile

Monthly Usage Decile	Average Monthly Usage (kW)	Average Impact from 4-6 PM (kW)
1	184	**0.19**
2	305	**0.31**
3	376	**0.37**
4	441	**0.43**
5	505	**0.48**
6	573	**0.53**
7	654	**0.61**
8	755	**0.66**
9	900	**0.72**
10	1,286	**0.85**

Figure 4-1 further illustrates how different impacts and usage are from the first to the tenth decile. The solid purple and solid green line represent control and treatment load across the four, two-hour event days for customers in the tenth percentile. The red line with diamond markers and the blue dotted line with diamond markers shows the control and treatment usage for customers in the first decile. These findings suggest that PG&E could increase program impacts by focusing marketing efforts on customers with higher-than-average monthly usage.

Figure 4-1: Event Impacts for Two-hour Events – 1st and 10th Usage Deciles



4.3.3 Customers with Two SmartAC Devices

Although customers with two or more SmartAC devices were excluded from the main ex post analysis, it is still possible to explore the impacts these customers provide. Of customers with more than 1 CAC unit, over 90% (about 13,000 customers) have exactly 2 CAC units. For customers with two CAC units, it is possible to calculate impacts for circumstances when both devices in a household are controlled and also when only one of two devices is controlled. Just under 5% of customers with two CACs have devices that both fall in the same randomized group used for calling events. If the group number was truly random, it would be expected that 10% of customers with 2 devices would have both devices fall within the same group. However, it appears that for customers with two devices, the most common occurrence is to have devices in sequential groups (e.g., first device is in group 2 and second device is in group 3). As one of 10 groups is called for each event, customers with one device will be called for 10% of events. Customers with two devices (as long as the devices are not in the same control group), however, will be called for two of ten events, or 20% of the time.

Ex Post Methods & Results for Customers with Both Devices Controlled at the Same Time

To estimate ex post results for customers with both devices treated, the treatment group for any given day is defined to be customers with both devices in the group called for an event that day. Customers with no devices being tested serve as the control group; customers with only one of two devices tested are dropped from the analysis, as are customers with more than two devices. For example, on June 15, group 2 was called for an event. For the analysis of customers with 2 devices, the treatment group consisted of about 40 customers who had 2 devices and were both in group 2. Customers with two devices and exactly one device in group 2 were dropped from the analysis for that day. The remaining customers with two devices and no devices in group 2 served as the control group.

For any given 2011 event day, there were about 8,500 customers with 2 devices in the control group and around 50 customers with both devices called in the treatment group. Impacts were calculated by comparing treatment and control group usage during the event hours with a same-day adjustment. When analyzing the full residential SmartAC population, same-day adjustments are not needed between the treatment and control group due to the large number of customers. However, when looking only at customers with two devices, the adjustment increases the accuracy of impact estimates. The same-day adjustment is based on the ratio of usage between the treatment and control groups for the four hours prior to the event start. For example, for an event that runs from 4–6 PM, the average ratio of treatment usage to control usage over the hours of 12–4 PM is used to adjust the control usage for the entire day. In addition to the same-day adjustment, event days with the same event hours were analyzed together to compensate for small sample sizes. There were five-hour events on June 15, 21 and 22, which were analyzed together, and two-hour events on August 24, September 6, 7 and 8, which were also analyzed together.

Table 4-6 shows the ex post event impacts for customers with two devices compared to the main results of customers with one device. The first two columns show the results for the five-hour event days and the last two columns have results for the two-hour event days. For five-hour events, customers with two devices showed slightly lower impacts, with an average of 0.37 kW compared to 0.42 kW. For two-hour events, however, customers with two devices had slightly higher impacts than customers with one device. Although standard errors are not shown in Table 4-6, differences in load impacts between customers with one device and customers with two devices have quite high standard errors as compared to their magnitude, suggesting that the difference in the average impact between the two groups is not statistically significant.

Table 4-6: Ex Post Event Impacts for Customers with Two Devices Treated

Hour	Five-Hour Events		Two-hour Events	
	Customers with 2 Devices	Customers with 1 Device	Customers with 2 Devices	Customers with 1 Device
1-2 PM	0.15	0.27	n/a	n/a
2-3 PM	0.29	0.36	n/a	n/a
3-4 PM	0.48	0.44	n/a	n/a
4-5 PM	0.53	0.52	0.54	0.53
5-6 PM	0.40	0.53	0.47	0.42
Average	0.37	0.42	0.51	0.47

Figure 4-2 depicts the load impacts on five-hour event days for customers with one device and customers with two devices. What is striking about the figure is that although the average customer with two devices has a much higher average usage, the event impacts are very similar. To illustrate, the difference between the two groups in terms of average hourly usage is 0.57 kW, with customers with two devices having greater usage.[15] However, in terms of impacts, customers with one device actually give average hourly impacts that are 0.05 kW greater than customers with two devices. This shows that despite having greater household loads, customers with two devices do not supply significantly greater impacts for the SmartAC program; load impacts are similar between the two groups.

These results suggest that installing control devices on houses with more than one CAC may not be cost effective. Customers with two devices, regardless of whether one or both devices are called, show equal or lesser impacts than households with only one CAC. As hardware and installation costs are higher for two devices than for one device, households with one CAC provide much greater impacts for the cost.

[15] This number is calculated by comparing average hourly control group usage for both sets of customers.

Figure 4-2: Event Impacts for Five-hour Events Customers with Two Control Devices Treated vs. Customers with One Control Device

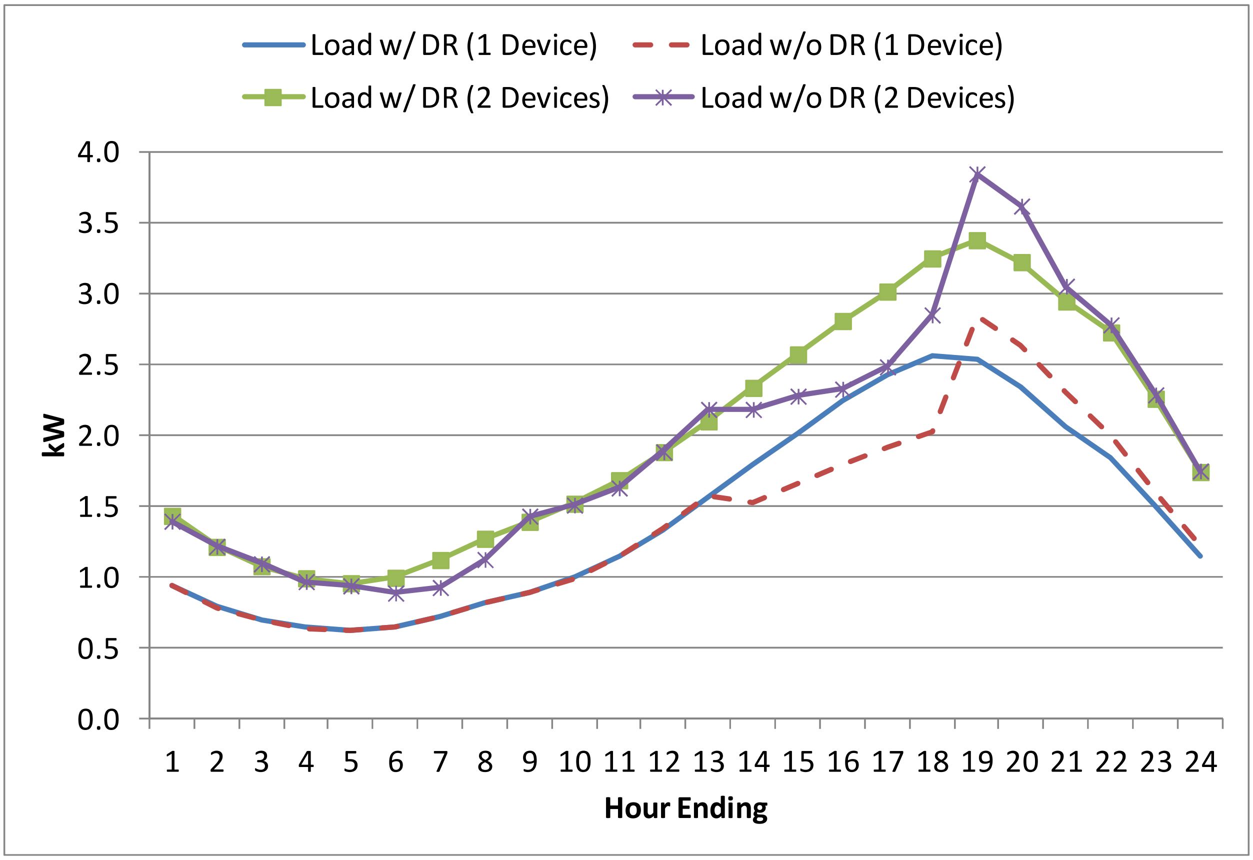
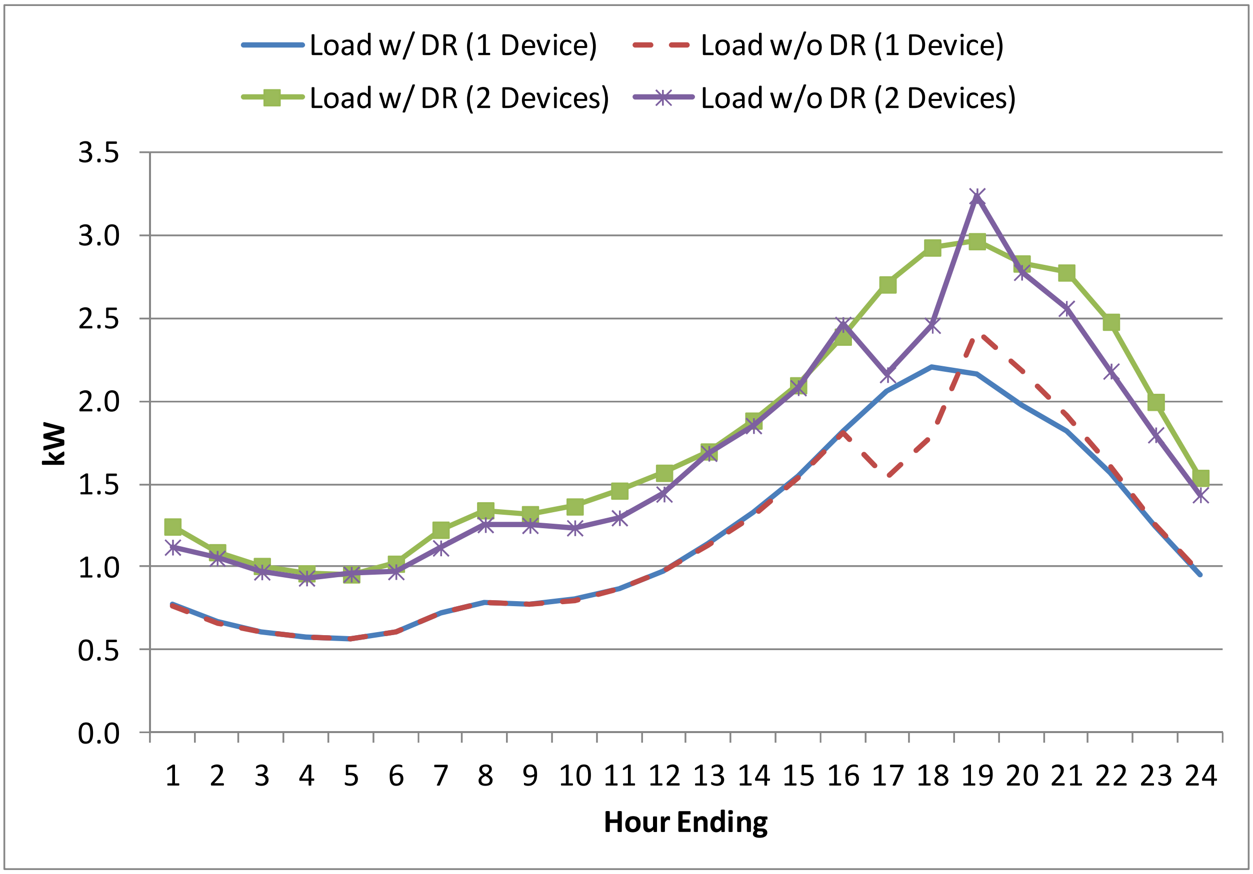


Figure 4-3 depicts the load impacts on two-hour event days for customers with one device and customers with two devices. The graph shows that the patterns noted on five-hour event days also hold true on two-hour event days. Customers with two devices gave average hourly impacts that are 0.04 kW greater than customers with one device. However, customers with two devices also used an average of 0.59 kW more per hour than customers with one device. As with the five-hour event results, this shows that customers with two devices do not supply significantly greater load reductions that customers with one device.

Figure 4-3: Event Impacts for Two-Hour Events Customers with Two Control Devices Treated vs. Customers with One Control Device



Ex Post Methods & Results for Customers with One of Two Devices Controlled

To estimate load impact for households where one of the two devices was controlled, the treatment group consisted of customers for whom one device was called for the event and the other was not. The control group consisted of customers for whom neither device was called. Customers with both devices treated on the event day were dropped from the analysis. For each event day during the summer of 2011, there were about 8,000 customers in the control group and just over 1,900 in the treatment group. Unlike the analysis of households where both devices were controlled, a same-day adjustment is not needed here due to sufficient sample sizes.

Table 4-7 shows the ex post event impacts for customers with two devices for events when one device was called compared to the main results of customers with one device. The first two columns show the results for the five-hour event days and the last two columns have results for the two-hour event days. For five-hour events, customers with two devices who had one device called showed impacts less than half the size of those seen for customers with one device. Similarly, for two-hour events, customers with two devices who had one device called had average hourly impacts of 0.23 kW compared to 0.47 kW for customers with only one device.

These results suggest that at premises with multiple CAC units, when one unit is controlled while others are not, that the uncontrolled units compensate for the controlled unit by running more. This appears to contradict the straightforward interpretation of the result for customers with two devices in one random group, which is that residential customers with more than one CAC tend to have only one running at a time. There is not sufficient data currently available to decide between these possibilities, but there may be in future evaluations.

Table 4-7: Ex Post Event Impacts for Customers with One of Two Devices Treated

Hour	Five-Hour Events		Two-Hour Events	
	Customers with 1 of 2 Devices Treated	Customers with 1 Device	Customers with 1 of 2 Devices Treated	Customers with 1 Device
1-2 PM	0.12	0.27	n/a	n/a
2-3 PM	0.16	0.36	n/a	n/a
3-4 PM	0.17	0.44	n/a	n/a
4-5 PM	0.20	0.52	0.27	0.53
5-6 PM	0.24	0.53	0.20	0.42
Average	0.18	0.42	0.23	0.47

Figure 4-4 depicts the load impacts on five-hour event days for customers with one device and customers with one of two devices treated. As noted in Table 4-7 above, the impacts for customers where one of two devices were treated are less than half the size of impacts for customers with just one device. Additionally, because reference loads are also higher for customers with two devices, the difference in percent impacts between the two groups is even greater. Households with just one device showed 20% impacts on average for five-hour event days while households with two devices where one was controlled had an average of 7% impacts over the five hours.

Figure 4-4: Event Impacts for Five-hour Events Customers with Two Control Devices, One Treated vs. Customers with One Control Device

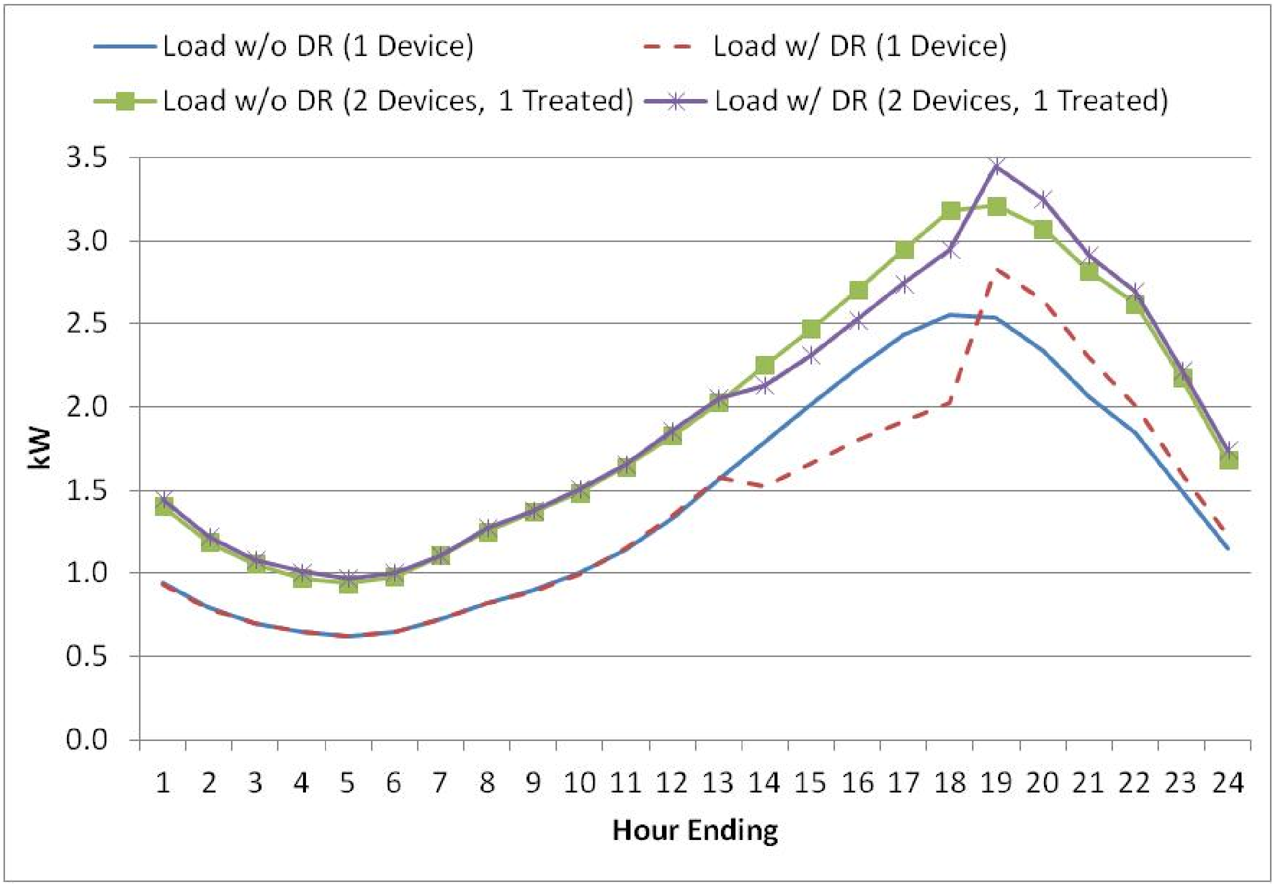
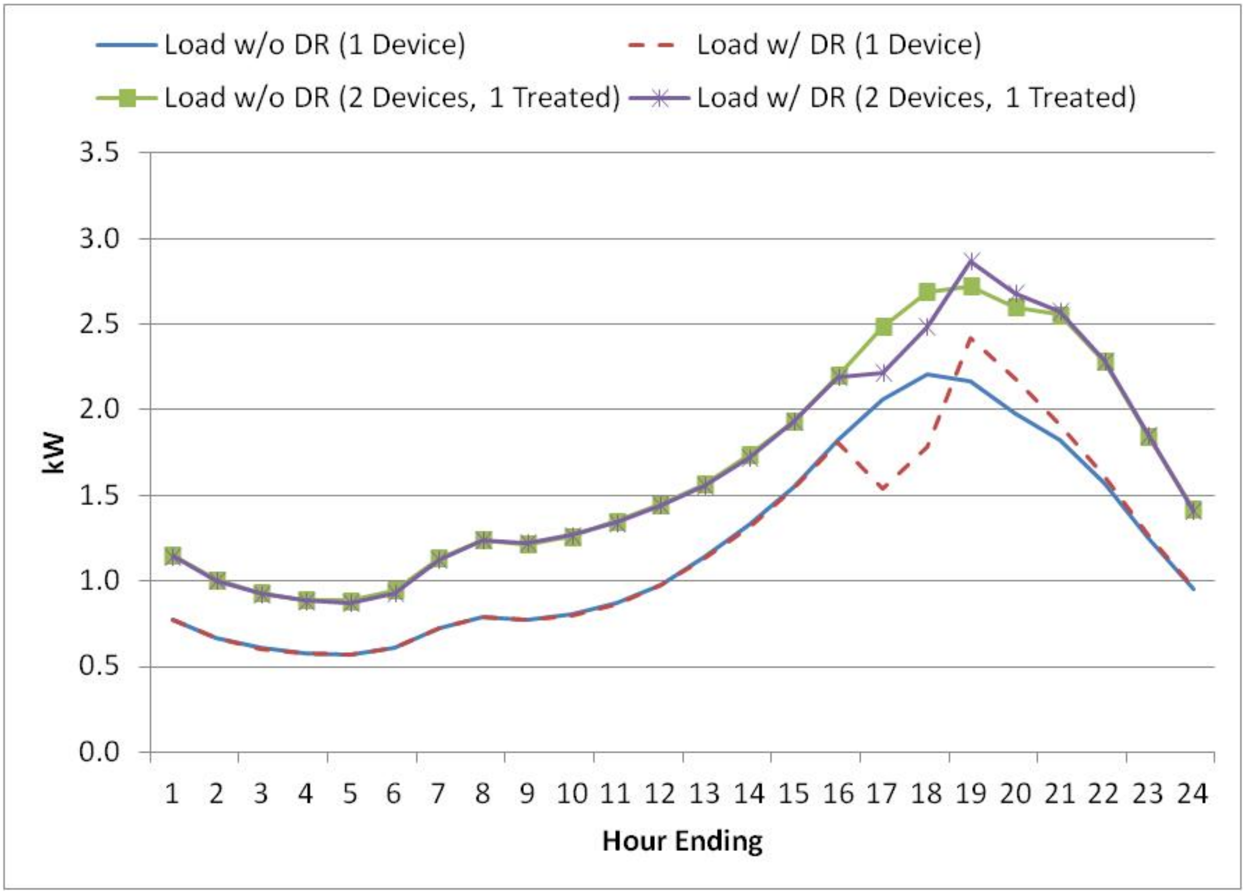


Figure 4-5 depicts the load impacts on two-hour event days for customers with one device and customers with one of two devices treated. Just as with the five-hour events, customers with one device showed impacts twice as large as customers who had one of two devices treated. Customers with one of two devices treated showed average impacts of 11% of whole house load over the two hours while customers with one device has average impacts of over 25%.

Figure 4-5: Event Impacts for Two-hour Events Customers with Two Control Devices, One Treated vs. Customers with One Control Device



4.3.4 Dually-enrolled SmartAC/SmartRate Customers

In addition to being on SmartAC, some customers are also enrolled on SmartRate. About 4,700 customers are enrolled on both SmartAC and SmartRate. SmartRate is a dynamic rate that overlays other available tariffs. SmartRate has a high price during the peak period on up to 15 event days per year, referred to as Smart Days, and slightly lower prices at all other times during the summer. For non-CARE customers, the peak price adder is $0.60 per kWh. For customers enrolled in both SmartAC and SmartRate, CAC control devices activate for both SmartRate and SmartAC event days. During the summer of 2011, 5 of the 15 SmartRate days were also SmartAC test event days.

For those five days, estimates of event impacts for dually-enrolled customers are compared to event impacts of matched groups of customers only on SmartRate and only on SmartAC. This produces estimates of the incremental impact of SmartAC on SmartRate customers and vice versa. Matching is necessary because the group of dually-enrolled SmartAC/SmartRate customers is different than the group of SmartAC customers and the group of SmartRate customers. The matching process finds customers within the SmartAC and SmartRate populations that have characteristics similar to the dually-enrolled population. This ensures that the comparison of load impacts isolates the differences in impacts due to the programs rather than due to underlying differences between the three groups.

First, dually-enrolled customers were matched to SmartRate only customers using propensity score matching based on hourly usage on hot non-event days and local capacity area. Next, the same set of dually-enrolled customers was matched to SmartAC-only customers. In order to keep data requests manageable, SmartMeter data was only obtained on event days for the SmartAC population. Therefore, the propensity score match that produced a set of SmartAC-only customers with characteristics comparable to dually-enrolled customers was based on hourly usage on the two SmartAC event days that did not overlap with SmartRate events. Only SmartAC customers who were not treated on June 15 and August 24 were included in the match.

Figure 4-6 provides evidence that the matching procedure produced three sets of customers (dually-enrolled, SmartRate-only and SmartAC-only) that have quite similar loads on hot non-event days. The figure shows the average usage of each of the three groups on June 15, 2011 and August 24, 2011. These are the only two days that were included in both propensity score matches. All three groups have very similar usage patterns throughout the day. More importantly, they have nearly identical usage from 4 to 6 PM, the hours over which all events overlap. SmartAC only customers used about 2% more than dually-enrolled customers from 4 to 6 PM and SmartRate only customers used about 3% more than dually-enrolled customers.

Figure 4-6: Average Usage Among Matched Groups on June 15 & August 24, 2011

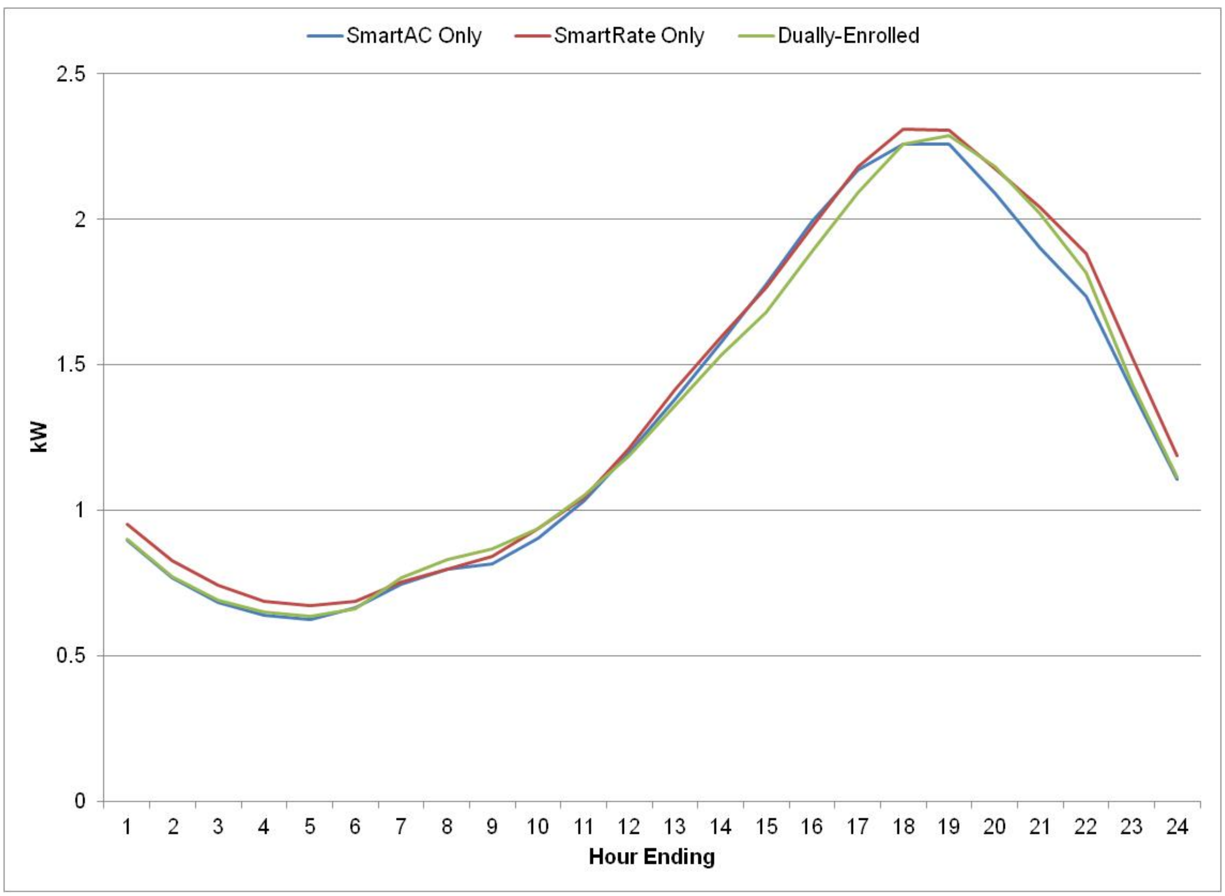


Table 4-8 shows impacts for each of the three matched groups on the five overlapping SmartAC/ SmartRate days. Event impacts are shown only for the hours 4 to 6 PM because those are the hours over which the events always overlapped. Not surprisingly, dually-enrolled customers showed the greatest load impacts of the three groups. SmartAC-only customers showed lower impacts than the dually-enrolled group but only by about 15%. SmartRate-only customers, however, showed less than half the impacts of dually-enrolled customers. This highlights the importance of the CAC control device for achieving greater load impacts.

From the perspective of the SmartAC program, having a customer on SmartRate as well appears to provide a modest incremental impact. On the other hand, it appears that for the SmartRate program, having customers also on SmartAC provides much greater impact. More detail about the effect of SmartAC on SmartRate customers is provided in FSC's evaluation of PG&E's residential pricing programs for 2011.[16]

Table 4-8: Load Impacts for SmartAC, SmartRate and Dually-Enrolled Customers

Date	Program Called	Event Hours	Avg. Hourly Impacts		
			SmartAC Customers	SmartRate Customers	Dually-Enrolled Customers
6/21/2011	SMR/SMAC	4-6 PM	0.65	0.54	0.93
6/22/2011	SMR/SMAC	4-6 PM	0.71	0.43	0.86
9/6/2011	SMR/SMAC	4-6 PM	0.47	0.24	0.53
9/7/2011	SMR/SMAC	4-6 PM	0.58	0.22	0.56
9/8/2011	SMR/SMAC	4-6 PM	0.47	0.23	0.53
Average Impact when All Groups Called			0.58	0.33	0.68

4.4 Residential SmartAC and Dually-Enrolled Customer Ex Ante Methodology

This section explains the steps used to predict ex ante load impacts for residential SmartAC customers as well as residential customer who are enrolled on both SmartAC and SmartRate (dually-enrolled customers). The methods in this section differ substantially from previous evaluations due to the different set of data provided by the evaluation design. This design provides highly precise impact estimates at the aggregate level for each LCA for each event. Because the experimental design that estimates the ex post impacts uses a different treatment group each time, it does not provide a set of CAC load data for the same set of customers over an entire summer. This means that different modeling steps are necessary. Appendix E discusses the conceptual and practical differences between this year's approach and the approach used in previous evaluations. Results for residential SmartAC customers and dually-enrolled customers were calculated separately using similar methods. Section 4.1.1 describes how ex ante load impacts were calculated for the residential SmartAC and Section 4.1.2 explains the methodology for dually-enrolled customers by discussing the difference in methods.

16 See "2011 Ex Post Load Impact Evaluation of Pacific Gas and Electric Company's Residential Time-Based Pricing" prepared for PG&E by the FSC Group.

There are two issues that must be dealt with in this modeling. First, the weather observed during events in 2011 is different than the ex ante weather conditions of interest. Second, there are only seven test events for each LCA to use for modeling. Both of these issues should improve in future years as more data is gathered using these methods. In two years' time, there could be 20-30 events per LCA to use for modeling impacts, including hotter ones than those used here. As it is, the modeling procedure outlined here makes the most of the data that does exist, including using the observed performance of the program across other LCAs to inform the predictions for LCAs that did not experience high temperature events in 2011.

At a high level, the modeling steps consist of the following analysis:

- First, a regression model was developed to explain average ex post impacts from 4 to 6 PM as a function of recent temperatures. This model was estimated at the level of each LCA separately. This model used a combination of LCA-specific and territory-wide event impact information. The model was used to estimate average impacts from 4 to 6 PM for the set of ex ante weather conditions. The estimates of the average impact from 4 to 6 PM were then converted to hourly impacts from 1 to 6 PM using a scaling factor based on load impacts observed during longer events.
- Next, whole-house reference loads from 4 to 6 PM were predicted for each set of ex ante weather conditions based on the loads observed over the summer of 2011. Load shapes were estimated by taking the average load for each hour of the day, by LCA. The same load shapes calculated based on the entire residential SmartAC population were used in the dually-enrolled analysis as well.
- Finally, a similar regression model was applied to snapback as was applied to event impacts.

The first step, which provides estimated load impacts, is described in detail below. The steps used to predict whole-house loads and snap-back are described in Appendix A.

4.4.1 Estimating Ex Ante Load Impacts for Residential SmartAC

Ex ante impact estimates were calculated by making predictions for ex ante weather conditions using a regression model of ex post impacts. The ex ante weather conditions are the same that were used for the 2010 SmartAC evaluation and have been chosen to be representative of 1-in-2 and 1-in-10 monthly peak days and 1-in-2 and 1-in-10 typical event days.

For those familiar with the regression modeling approach used to estimate load impacts in previous SmartAC evaluations and in other load impact evaluations such as the statewide Aggregator Program, statewide CPP or SmartRate, note that this methodology is quite different. Under the experimental design for test events in this evaluation, virtually no modeling is required to produce ex post impacts. In previous years, both ex post and ex ante impacts were an output of the same regression model. Here, ex post impact estimates are taken as the dependent variable in an ex ante regression model. This is a substantial step forward in modeling impacts because it has taken virtually all the uncertainty out of the ex post step of the process. Previously, to argue about ex ante impacts was equivalent to arguing about ex post impacts because the same model produced both sets of impacts. In this evaluation, there is very little to argue about regarding ex post impacts. The only serious uncertainty that remains is how to best use ex post impacts to predict ex ante impacts.

To determine the best regression to use for ex ante predictions, FSC tested dozens of models predicting ex post impacts based on different measures of recent temperature. The final regression only includes one explanatory variable because there were only seven test events for each LCA. Using more explanatory variables in such a case would result in model over-fitting. The model that best predicted average ex post impacts by LCA was:

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

Table 4-9:
Description of CAC Load Regression Variables

Variable	Description
$Impact_c$	Average per device ex post load impact for each event day from 4 to 6 PM (2 to 7 PM for dually-enrolled customers)
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 based both on its predictive ability and on the fact that ex ante impact prediction is based on 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 seven points per LCA for modeling, it is not possible to identify these effects. Ideally, in future years, results from 2011 will be combined with future results to allow for modeling of impacts using more variables.

The choice of using load impact measured for the window 4 to 6 PM for the dependent variable was made because all test events covered the hours 4 to 6 PM. Moreover, side-by-side testing showed that the impact from 4 to 6 PM barely changed whether the event started at 4 PM or 1 PM. 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 6 PM, while others went from 4 to 6 PM. In such a case, the longer events might have lower average impacts due to the fact that CAC loads are lower earlier in the day rather than because the event was less effective.

The results of these regressions for the residential SmartAC population are shown in Figure 4-7. The graph for each LCA includes seven points, one for each event day, and the fitted line created by the regression. Although the relationship between temperature and impacts varies for each LCA, all LCAs show that the hotter the day was prior to the event start, the greater the impacts.

Figure 4-7:
Average Event Impacts Versus *Mean17* by LCA

Predictions from the LCA-level regressions were used when the ex ante weather temperatures fell within ranges that were experienced over the summer of 2011. However, in several cases, the ex ante weather scenarios for an LCA were warmer – as determined by the 17-hour average temperature – than any conditions experienced within the LCA during 2011. For example, the mean temperature over the 17 hours prior to the event reached a maximum of 76.5°F on June 21 for the Greater Bay Area. There are eight ex ante weather days for that LCA that have a value for *mean17* that is greater than 76.5°F.

In these cases, the LCA-level predictions must be extrapolated. There are many possible ways to do this, and in this case a hybrid method of two possibilities was used.

One possibility is to make the assumption that the relationship between impacts and *mean17* remains constant even under more extreme weather conditions. These predictions take into account trends specific to each LCA, but they also involve making important assumptions about impacts at extreme temperatures that may not hold.

Another option for these cases is to use predictions developed from the same regression specification applied to a pooled dataset containing average ex post impacts from 4 to 6 PM for every LCA and every test event. This allows for the relationship between impact and temperature seen in other LCAs that may have experienced higher temperatures to inform the predictions for ex ante days that would be extrapolated under LCA-level regressions. However, this approach also has its flaws as population differences between different LCAs are not taken into account under the pooled regression.

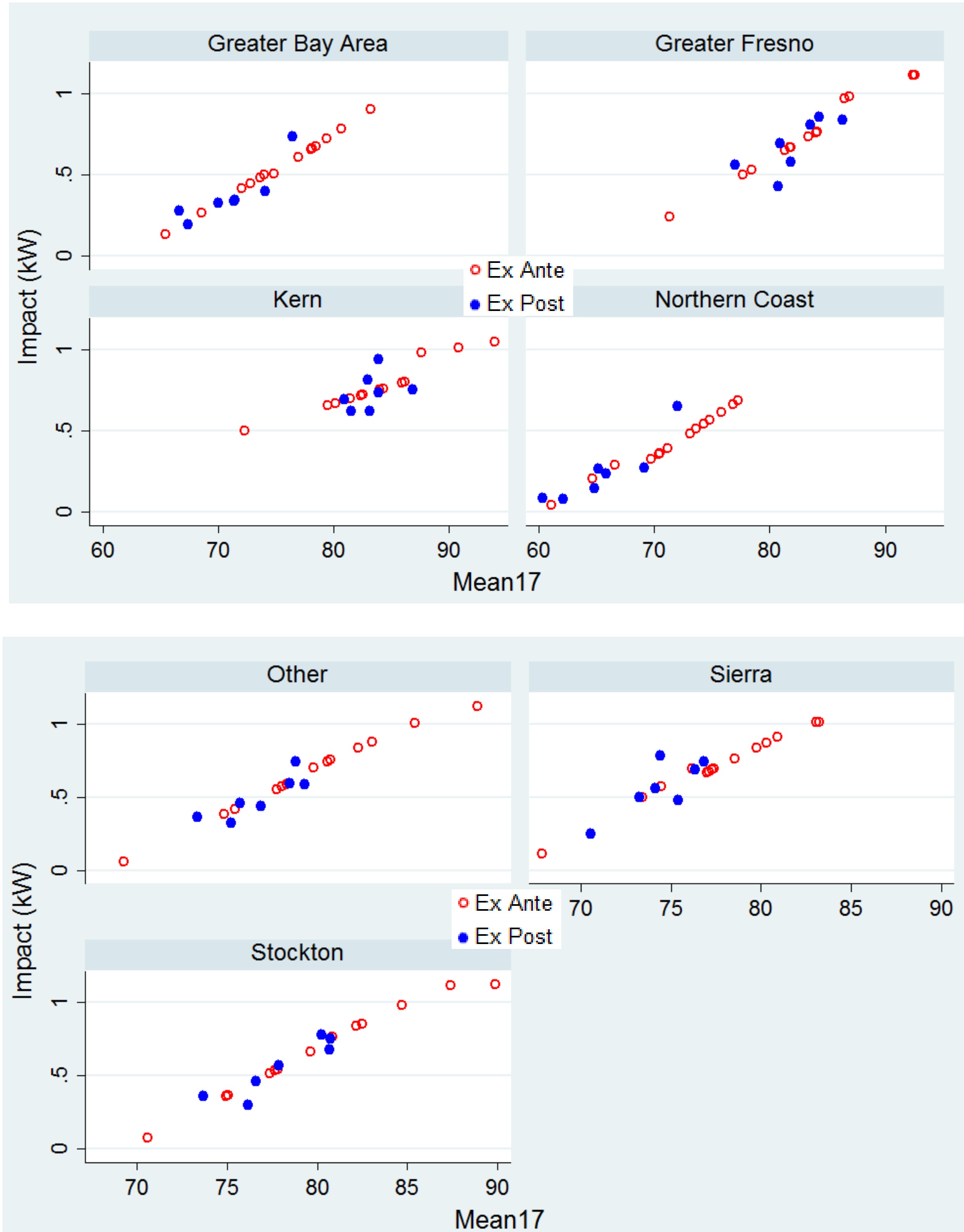
For any ex ante weather conditions hotter than six of the seven observed events, FSC used the average of the predictions from the LCA-level regression and the pooled regression. This allowed the regression to be informed both by real world ex post data and LCA-specific relationships between impacts and temperature.

Finally, all impacts were capped at 1.125 kW per device. This is 45% of the maximum average CAC load seen for any well-represented LCA in the CAC logger data collected during the summer of 2009 for the SmartAC evaluation. This dataset is appropriate to use here because there were no test events in 2009 and because there were several days for which temperatures in particular LCAs significantly exceeded 105 degrees – reaching 112 at one point in one LCA. This cap is based on the assumption that under real world conditions, the greatest load reduction likely to be achieved through 50% cycling is 45%. The basis for this assumption is that communication with control devices will never be perfect and the TrueCycle2 algorithm will never perform perfectly. This cap affected 12 out of 98 LCA-level ex ante predictions. For example, the LCA-level regression predicted an average hourly impact from 4 to 6 PM of 1.33 for the July system peak day in a 1-in-10 weather year for Stockton. The average of *mean17* for that day was 90°F – which is more than 9 degrees greater than the maximum *mean17* seen for the Stockton LCA during 2011. In cases like this, capping keeps the impacts within a realistic range.

Figure 4-8 displays the final ex ante and ex post estimates graphed against *mean17*. The solid blue circles represent ex post results and the hollow red circles are ex ante results. By graphing both ex ante and ex post results on the same plot, it is obvious that the ex ante results fall in line with the ex post results for observed temperatures. Additionally, the graphs show that each LCA has a slightly different pattern of ex ante impacts outside of observed temperature ranges, when the average of the LCA-level and whole-territory impacts is used.

The effects of both the hybrid approach to extrapolation and the cap on impacts are visible in several of the graphs. The effect of the hybrid approach can be seen in the discontinuities in the ex ante trends for Greater Fresno, Kern, Northern Coast and Sierra near the upper extreme of ex post temperatures for those areas. Those predictions are a combination of the predictions based on the LCA-level regression and the regression pooled over all LCAs. The result of the cap on impacts is visible for the highest impacts for Greater Fresno, Stockton and Other.

Figure 4-8:
Ex Post and Ex Ante Impacts Versus *Mean17* by LCA



The last step in estimating load impacts was to translate average impacts from 4 to 6 PM to hourly impacts over the entire range of time required for prediction, 1 to 6 PM. Using ex post impact estimates for each LCA from the first three test events, which had event hours from 1 to 6 PM, the average impact for each hour from 1 to 6 PM was expressed as a fraction of the average impact from 4 to 6 PM. The averages were calculated over all three events for each LCA separately. Table 4-10 gives an example of this process. The first column of Table 4-10 shows how the average event impact for each hour of the five hour events compares to the average impact from 4 to 6 PM, using the Greater Bay Area as an example. To illustrate, the second column shows the proportions in the first column multiplied by 0.50 kW, the average predicted impact from 4 to 6 PM for the Greater Bay Area 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 shown in Table 4-10.

Table 4-10:
Hourly Impact Compared to Average Impact from 4–6 PM

Hour of Event	Hourly Impact/ Average 4–6 PM Impact (%)	Hourly Impact for Greater Bay Area Typical Event Day, 1-in-2 Weather (kW)
1-2 PM	52	0.26
2-3 PM	68	0.34
3-4 PM	84	0.42
4-5 PM	99	0.49
5-6 PM	101	0.51

This approach was used for each LCA for each set of ex ante event conditions. The implication of this strategy is that the ratio between any two hours of predicted event impacts is constant across all ex ante conditions for a given LCA. Examination of the hourly ex post event impacts for the three events from 1 to 6 PM shows that this relationship is roughly true for these events. This assumption is discussed in Appendix E. Impacts for the overall SmartAC population were calculated by taking a weighted average of the LCA results.

4.4.2 Estimating Ex Ante Load Impacts for Dually-Enrolled Customers

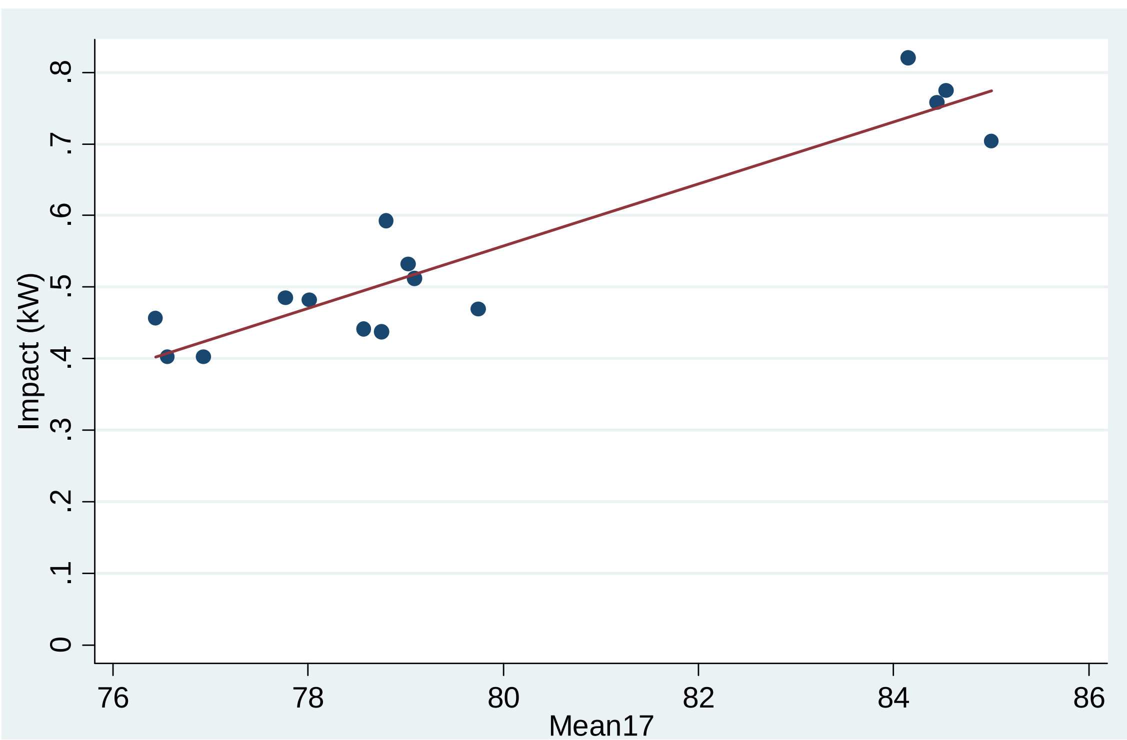
Ex ante load impacts for dually-enrolled customers were calculated using essentially same methodology as that described above for residential SmartAC customers with a few minor differences. This separate analysis involved the same major steps and some small modifications:

- First, a regression model was developed to explain average ex post impacts from 2 to 7 PM as a function of recent temperatures. There were 15 SmartRate events (dually-enrolled customers have their ACs controlled for both SmartAC and SmartRate events) which the regression was based on. Because there are less than 5,000 dually-enrolled customers, ex ante estimates were calculated at the whole territory level. The model was used to estimate average impacts from 2 to 7 PM for the set of ex ante weather conditions for each LCA;

- The estimates of the average impact from 2 to 7 PM were then converted to hourly impacts from 2 to 7 PM using a scaling factor in the same way described above for SmartAC customers.;[17]
- Estimates from 2 to 7 PM were adjusted to cover the resource adequacy hours of 1 to 6 PM;
- Because residential SmartAC customers and dually-enrolled customers show similar usage, whole-house load for dually-enrolled customers was taken directly from the SmartAC ex ante analysis. The process used to calculate those reference loads is described in Appendix A; and
- As with whole house loads, snapback was also taken directly from the residential SmartAC analysis. The methods used to calculate snapback can also be found in Appendix A.

The same regression model as described above was used to calculated dually-enrolled customer ex ante impacts. The only difference is that average impacts from 2 to 7 PM were used instead of from 4 to 6 PM. Figure 4-8 shows the average impact from 2 to 7 PM for the dually-enrolled population for each SmartRate event day in 2011 versus the *mean17*.

Figure 4-9:
Average Event Impacts Versus *Mean17* for Dually-Enrolled Customers

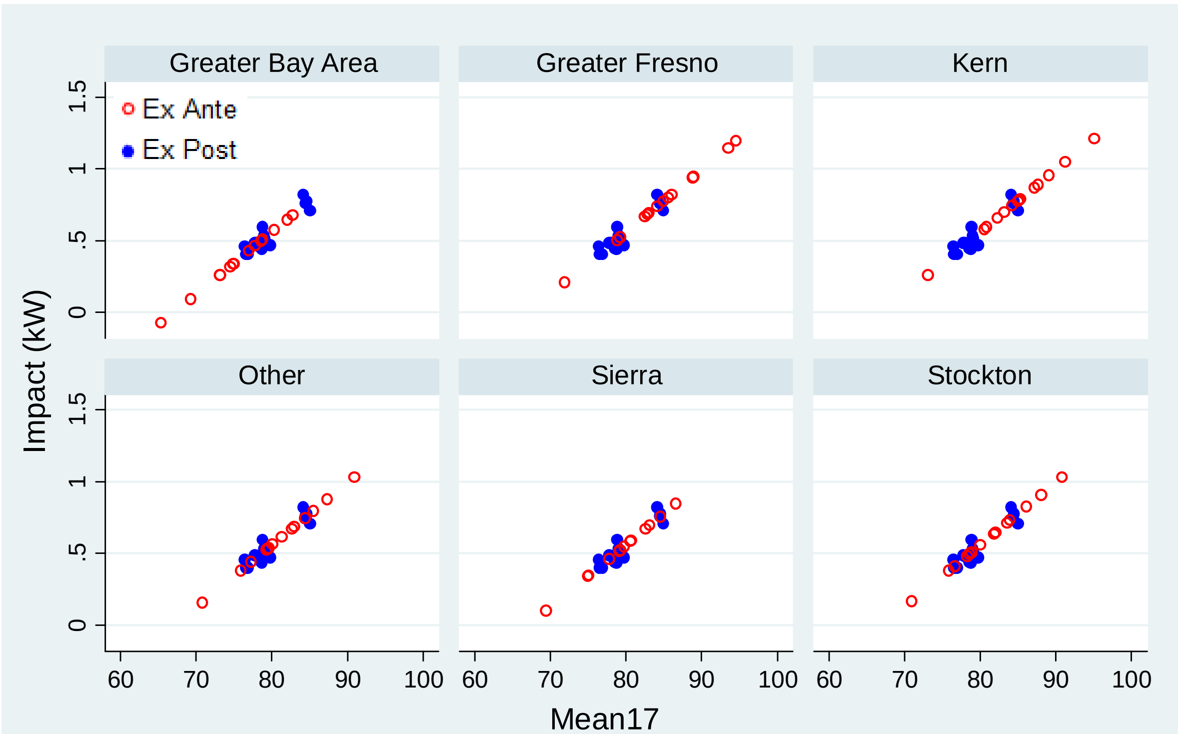


Estimates from the entire dually-enrolled population where then applied to ex ante weather scenarios calculated for each LCA. This allowed each LCA to have different impact estimates for the same ex ante scenarios because temperatures vary across LCAs. However, because the regression was based on whole territory results, predictions from each LCA all fall on the same line. Figure 4-10 displays the final ex ante and ex post estimates graphed against *mean17*. The solid blue circles represent ex post

[17] For dually-enrolled customers, average impacts were estimated from 2 to 7 PM. Average impacts were then converted to hourly impacts from 2 to 7 PM using a scaling factor based on ex post impacts.

results and the hollow red circles are ex ante results. The ex post impacts are the same for each LCA since estimates based on the entire dually-enrolled population were used.

Figure 4-10:
Ex Post and Ex Ante Impacts Versus *Mean17* by LCA



Unlike impact estimates for the residential SmartAC population, dually-enrolled customer impacts were not capped. Because customers who are dually-enrolled are notified about the SmartRate events, these customers may be shifting load in addition to their AC units being controlled.

All SmartRate events in 2011 were called from 2 to 7 PM. For future years, however, events will be called from 1 to 6 PM, to match the resource adequacy window. In order to incorporate this change into the ex ante results, event impacts had to be adjusted. Of the five-hour event, four of the hours stay the same; events in 2011 and in future years cover the hours from 2 to 6 PM. For those hours, the event impact estimates were not changed. However, from 1 to 2 PM, the model described so far provides no event impact estimates. In order to fill that gap, the percentage impact estimated for the hour from 2 to 3 PM was applied to the reference load from 1 to 2 PM. This means the percentage impact for hours 1 to 2 PM is always the same as the percent impact for hours 2 to 3 PM in the ex ante results. The level of inaccuracy for the overall average predicted impact due to this assumption is likely to be quite small.

4.5 Residential SmartAC Ex Ante Load Impact Results

The SmartAC program is intended to alleviate system stress during times of very high demand. The primary purpose of this evaluation is to predict load impacts during such conditions. These ex ante predictions cover a pre-chosen set of temperature profiles meant to mimic what could be expected for monthly system peak days that might occur every other year and every tenth year. Aggregate estimates of load impacts combine estimates of per customer load impacts developed in this report with estimates of program enrollment, developed in a separate effort by PG&E.

As mentioned previously, the ex ante weather conditions are mainly outside of the range of weather that was observed in 2011. This means that the model's predictions are extrapolations outside of the range of available data, which adds uncertainty. One way this has been dealt with in this evaluation is to combine LCA-specific and whole-territory ex ante estimates for predictions outside of the temperatures range seen during 2011 event days, as described in Section 4.4.1 above.

Enrollment projections for residential customers by local capacity area as of August of each year are presented in Table 4-11. The source for these projections is PG&E's enrollment projections for 2012-2022, developed in conjunction with the Brattle Group. Residential enrollment is projected to remain fairly stable in all LCAs for the foreseeable future. It is expected to increase by about 3% in 2013 and then decrease somewhat thereafter.

Table 4-11:
Projected Residential Enrollment for August of Each Year (1000s of customers)

LCA	2012	2013	2014	2015-2022
Greater Bay Area	54.1	52.8	51.8	51.6
Greater Fresno	27.8	27.1	26.6	26.5
Kern	5.8	5.6	5.5	5.5
Northern Coast	9.5	9.3	9.1	9.1
Other	26.5	25.9	25.4	25.3
Sierra	18.6	18.1	17.8	17.7
Stockton	12.9	12.6	12.4	12.3
Total	155.2	151.4	148.6	148.0

Ex ante load impact estimates are shown for residential customers in **Error! Reference source not found.**12. The first column shows the average hourly per customer ex ante load impact estimate over the event period from 1 to 6 PM and the second column shows the maximum per customer hourly impact estimate during the event. The third column shows the estimated aggregate load impact over the period 1 to 6 PM. The first set of rows corresponds to 1-in-2 weather conditions while the second set covers 1-in-10 weather conditions. For the 1-in-2 weather year, the highest estimated impact is on the July peak day, with an average impact of 100 MW and a peak hourly impact of 126 MW. The July peak day also shows the highest impacts for the 1-in-10 weather year. The mean impact over the five-hour event is almost 128 MW and the peak hourly impact is 160 MW.

Table 4-12:
2012 Residential SmartAC Load Impact Estimates
By Weather Year and Day Type
(Event Period 1-6 PM)

Weather Year	Day Type	Mean Hourly Per Customer Impact	Max. Hourly Per Customer Impact	Aggregate Mean Hourly Impact	Aggregate Max Hourly Impact
		(kW)	(kW)	(MW)	(MW)
1-in-2	Typical Event Day	0.46	0.58	71.9	90.1
	May Peak Day	0.29	0.37	45.9	57.5
	June Peak Day	0.36	0.45	56.5	70.9
	July Peak Day	0.65	0.81	100.7	126.3
	August Peak Day	0.46	0.59	72.2	90.5
	September Peak Day	0.48	0.61	74.3	93.1
	October Peak Day	0.16	0.2	24.5	30.7
1-in-10	Typical Event Day	0.68	0.86	105.8	132.6
	May Peak Day	0.57	0.71	89.0	111.5
	June Peak Day	0.62	0.78	97.0	121.6
	July Peak Day	0.82	1.02	127.5	159.8
	August Peak Day	0.73	0.92	113.6	142.4
	September Peak Day	0.56	0.7	85.8	107.5
	October Peak Day	0.46	0.58	70.7	88.5

The Typical Event Day estimates are based on projected August 2012 enrollment. The excel-based load impact tables that accompany this report provide substantially more information about the ex ante conditions, load shapes and impact estimates than is practical to include here. In interpreting those results it is important to keep in mind that loads and load impacts depend not only on the temperature during the event, but also on the temperature before the event. Two days with the same high temperature can provide very different load impacts if one of them has much higher overnight low due to heat retention in buildings.

4.6 Dually-Enrolled Customer Ex Ante Load Impact Results

Enrollment estimates for dually-enrolled SmartRate-SmartAC customers are shown in Table 4-13. The number of dually-enrolled customers is projected to increase by a factor of about 3 over the next two years and then stabilize.

Table 4-13:
Projected Residential Enrollment for August of Each Year (1000s of customers)

LCA	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Greater Bay Area	0.6	1.3	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6
Greater Fresno	1.2	2.9	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5
Kern	1.4	3.3	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9
Other	1.2	2.8	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3
Sierra	0.6	1.5	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8
Stockton	1.2	2.8	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3
Total	6.2	14.6	17.4	17.4	17.4	17.4	17.4	17.4	17.4	17.4	17.4

Per customer ex ante load impacts are shown for dually-enrolled customers in **Error! Reference source not found.**14. The first column shows the average hourly per customer ex ante load impact estimate over the event period from 1 to 6 PM and the second column shows the maximum hourly per customer impact estimate during the event. The third row shows the estimated aggregate load impact. The first set of rows corresponds to 1-in-2 weather conditions while the second set covers 1-in-10 weather conditions. For the 1-in-2 weather year, the highest mean aggregate impact is on the July peak day, with an estimated aggregate impact of 3.9 MW. For the 1-in-10 weather year, August provides the largest estimated impact – 5.6 MW. Although the per customer average is slightly lower in August than in July for the 1-in-10 weather year, the projected enrollment increases between July and August enough to make up for that difference.

Due to the fast pace of projected enrollment, aggregate impact estimates vary both due to changes in enrollment and due to differences in per customer impacts. The Typical Event Day estimates are based on projected August 2012 enrollment.

Table 4-14:
2012 Dually-Enrolled Load Impact Estimates
By Weather Year and Day Type
(Event Period 1-6 PM)

Weather Year	Day Type	Mean Hourly Per Customer Impact (kW)	Max. Hourly Per Customer Impact (kW)	Aggregate Mean Hourly Impact (MW)	Aggregate Max Hourly Impact (MW)
1-in-2	Typical Event Day	0.56	0.69	3.4	4.2
	May Peak Day	0.44	0.53	1.8	2.1
	June Peak Day	0.43	0.52	2.0	2.4
	July Peak Day	0.72	0.88	3.9	4.8
	August Peak Day	0.54	0.66	3.3	4.0
	September Peak Day	0.56	0.69	3.8	4.7
	October Peak Day	0.17	0.21	1.3	1.6
1-in-10	Typical Event Day	0.78	0.96	4.8	5.9
	May Peak Day	0.67	0.82	2.7	3.3
	June Peak Day	0.67	0.82	3.2	3.9
	July Peak Day	0.94	1.15	5.1	6.2
	August Peak Day	0.91	1.11	5.6	6.8
	September Peak Day	0.60	0.74	4.1	5.0
	October Peak Day	0.50	0.62	3.8	4.7

5 SMB Impact Analysis

The 2011 SMB SmartAC program produced average ex post impacts per device of 0.29 kW during the hours 3 to 5 PM.[18] This represents a substantial improvement over the impacts observed in 2010.

This chapter parallels the organization of the residential impact analysis section. It is divided into five main sections. Section 5.1 describes the approach used in the ex post impact analysis, including a description of the experimental design used. Validation for the SMB load impacts is presented in Section 5.2. Section 5.3 describes the ex post impact results for all nine 2011 event days. It also includes two alternate approaches for calculating SMB ex post impacts: logger data at the industry level and SmartMeter data. Section 5.4 describes the methodology used in the ex ante impact analysis and Section 5.5 presents the results of that analysis.

5.1 SMB SmartAC Ex Post Methodology

As with the residential methodology, the methods used in the 2011 SMB SmartAC evaluation greatly differ from prior years. Because this year's events utilized random assignment to treatment or control groups, impacts could be estimated without the use of individual customer regressions. This provided impacts that were simple to calculate, but it also meant developing new methods for LCA-, industry- and demand-category-level impact estimates. Appendix E discusses the conceptual and practical differences between this year's approach and the approach used in previous evaluations.

For each of the nine test events during the summer of 2011, a randomly-chosen 50% of the SMB M&E sample was called for the event while the rest of the customers served as the control group. These groups were fixed at the beginning of the season and the groups were alternated between being controlled or not on each test event day. Whole-territory ex post impacts were calculated by comparing the average event day usage in the control group to the average event day usage in the treated group, with a same-day adjustment applied to the control load. The same-day adjustment was based on the same process described in section 4.3.3 above. There were approximately 340 loggers for treated customers and 340 loggers for control customers each event day.

For customers with multiple CAC units, all units were included in the same group and called for test events, even if the unit did not have a CAC logger. This accurately simulated the experience of a SmartAC event at the level of each premise.

An adjustment factor was applied to the data based on the fact that there was working device and CAC unit verification in the installation process. In the SmartAC population, more control failure and broken control devices should be expected than in the M&E sample. This correction factor was calculated as 5% for SMB CAC units. This means that for SMB customers impacts were adjusted downward by 5%. See Section 3.1 above for more discussion of this correction factor.

Ex post event impacts were calculated differently for smaller segments of the population than they were for the whole SMB SmartAC population. For average impacts across the entire territory, no regressions were needed. First, the average usage was calculated in the treatment and control groups for each hour of each event day. Next, a same-day adjustment was applied to the control load in

[18] These are the only hours that all events covered.

order to reduce bias. This adjustment consisted of multiplying control group usage by the ratio of the average treatment group load to the average control group load over the four hours before the event. The ratio ranged from 0.96 to 1.05, implying that the unadjusted load itself was already a good reference load. The adjustment made the reference load slightly more accurate. Treatment group usage during the event hours was subtracted from adjusted control group usage to calculate impacts.

Regressions were used to estimate ex post impacts at the LCA, industry and demand category levels. First, a regression model was developed to explain the average ex post percent impacts from 3 to 5 PM on each event day for the whole territory. The regression consisted of a linear regression of average impact over a two-hour period on one explanatory variable. The same explanatory variable was chosen for this regression as was chosen for the residential regression; mean temperature over the first 17 hours of each event day (*mean17)*. As in the residential case, only one variable was used in the regression because of the small number of test events. Percent impact was the dependent variable in the regression because event impacts vary more consistently with temperature as a percentage of CAC load. The model was:

$$PercentImpact_c = a + b \cdot mean17_{\mathbf{c}} + \boldsymbol{\varepsilon}_{\boldsymbol{c}}$$

Table 5-1: Description of CAC Load Regression Variables

Variable	Description
$PercentImpact_c$	Average per device ex post load percent 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
$\boldsymbol{\varepsilon}_{\boldsymbol{t}}$	The error term, assumed to be a mean zero and uncorrelated with any of the independent variables

Results from the model were used to predict percent impacts from 3 to 5 PM for each LCA, industry and demand category based on the weather specific to each category for each time period. These estimates were converted to hourly percent impacts (either from 1 to 6 PM or 3 to 5 PM) using a scaling factor based on load impacts observed for the entire territory. Table 5-2 shows the scaling factor for each hour of each type of event day experienced in 2011. For example, if the regression predicted impact from 3 to 5 PM is 0.47 kW for a five-hour event, then the impact in the first hour would be 0.51 kW. Additionally, if the prediction is 0.47 kW for a two-hour event, then the predicted impacts for both hours of the event is also 0.47 kW.

Table 5-2: Hourly Impact Compared to Average Impact from 3 to 5 PM

Hour of Event	Hourly Impact/ Average 3-5 PM Impact		Example Impacts	
	5-hour event	2-hour event	5-hour event	2-hour event
1-2 PM	107.9%	-	0.51	-
2-3 PM	106.1%	-	0.50	-
3-4 PM	106.0%	99.7%	0.50	0.47
4-5 PM	94.0%	100.3%	0.44	0.47
5-6 PM	63.3%	-	0.30	-

This scaling-factor approach produces impact estimates under the assumption that the impacts do not vary across hours within a given event. This is not strictly true. However, event impacts over hours due tend to follow strong patterns, with commercial event impacts peaking in mid-afternoon and falling off significantly in the early evening as businesses close. Simultaneously estimating average event impacts and the change in shape of event impacts due to observable factors provides noisy estimates with the amount of data available. The approach here emphasizes accurately measuring the relationship between average event impact and temperature, which is the primary driver of event impacts. This issue is discussed more in Appendix E.

Finally, predicted percent load impacts must be converted to absolute ex post load impacts. To get absolute ex post load impacts, predicted percent load impacts were applied to estimated reference loads. Reference loads were calculated for each LCA, industry and demand category based on observed load on non-event days in the SMB load research samples from 2009, 2010 and 2011. Average load was calculated for bins of *mean17* temperature. For example, if the event day in Greater Fresno had a *mean17* of 91°F, then the reference load for the ex post results would come from the 90-92°F bin. Table 5-3 shows how predicted impacts were applied to estimated reference loads. For example, the estimated reference load for the Greater Bay Area on June 21 is 1.85 kW from 1 to 2 PM. With a 20% predicted load impact, the absolute impact for that hour comes out to 0.37 kW.

Table 5-3: Percent Impact Converted to Absolute Impact Greater Bay Area - June 21, 2011

Hour	Load w/o DR	Predicted % Impact	Absolute Impact
1-2 PM	1.85	0.20	0.37
2-3 PM	1.97	0.20	0.38
3-4 PM	2.06	0.20	0.40
4-5 PM	2.05	0.17	0.35
5-6 PM	1.84	0.12	0.21

5.2 Validity of SMB Load Impact Estimates

As in the case of residential customers, the experimental method used to estimate impacts makes a strong prima facie case for the estimates' validity. Again, further evidence is available in the Excel-based load impact tables that accompany this report. The tables show raw average loads for customers whose loads were controlled and adjusted average loads for customers whose loads were not controlled for each test event day. Based on examining those loads in the pre- and post-event hours for each event, it is clear that the method used produced quite accurate reference loads for each test event.

Unlike the residential case, estimates at the LCA level were developed based on a regression model. The validity of that approach can be assessed only by taking a stand on the assumptions that it is based on. Those assumptions are plausible as an approximation, but will likely be significantly improved through broader data collection in the future. Currently, there is little direct evidence about LCA-specific load impacts for SMB customers due to the cost of data collection.

5.3 SMB SmartAC Ex Post Load Impact Results

This section is broken into three parts. The first section presents the ex post load impacts for the nine primary test events called over the summer. The second part discusses how logger data could be used to estimate ex post impacts for different customer segments if sample sizes were larger. Finally, the third section provides a comparison of the ex post load impacts calculated using logger and SmartMeter data, which allows for an analysis of how impacts increase as the number of devices per premise increases.

5.3.1 SmartAC Primary Test Event Results

Table 5-4 shows the average impact per device for each load research event along with average temperature over the event period for the SMB SmartAC population. The first three events ran from 1 to 6 PM and the final six occurred between 3 and 5 PM. Table 5-4 only shows impacts from 3 to 5 PM of each event day so that the nine events are comparable. The largest impact occurred on June 22, which had an estimated impact of 0.44 kW per CAC unit. The average impact across all events is 0.29 kW per CAC unit, or roughly 17% of CAC load. This is a large improvement over last year when impacts averaged 0.11 kW per device, about 7% of CAC load.

Table 5-4 also includes the aggregate event impacts from 3 to 5 PM on each event day. This number is the estimated impact actually seen on those days, based on the number of customers called for each event. Because the number of the devices in the logger sample is relatively small (<350 devices called for an event each event day), aggregate impacts are also low compared to residential aggregate impacts. The average aggregate impact from 3 to 5 PM across all event days was one-tenth of a MW.

Table 5-4: Average SMB per Device Reference Loads,[19] Impacts and Temperatures from 3 to 5 PM on 2011 Event Days

Event Date	Event Hours	Average Reference Load (kW)	Average Event Impact (kW)	Percent Impact	Aggregate Event Impact (MW)	Average Temperature (°F)
6/21/2011	3-5 PM	2.06	0.27	14%	0.09	96
6/22/2011	3-5 PM	1.92	0.44	23%	0.15	89
6/23/2011	3-5 PM	1.41	0.20	14%	0.07	83
8/24/2011	3-5 PM	1.60	0.23	14%	0.08	89
9/6/2011	3-5 PM	1.67	0.25	15%	0.08	90
9/7/2011	3-5 PM	1.69	0.28	16%	0.10	94
9/8/2011	3-5 PM	1.44	0.20	14%	0.07	87
9/20/2011	3-5 PM	1.87	0.43	23%	0.15	94
9/21/2011	3-5 PM	1.72	0.34	19%	0.11	92
Average	**3-5 PM**	**1.71**	**0.29**	**17%**	**0.10**	**90**

Table 5-5 shows the average event impacts for the five-hour events called in June 2011. Average and aggregate impacts over the five-hour period were almost exactly the same as over the last two hours of the event.

Table 5-5: Average SMB per Device Reference Loads, Impacts and Temperatures During Event Hours on June 2011 Event Days

Event Date	Event Hours	Average Reference Load (kW)	Average Event Impact (kW)	Percent Impact	Aggregate Event Impact (MW)	Average Temperature (°F)
6/21/2011	1-6 PM	1.99	0.30	15%	0.10	95
6/22/2011	1-6 PM	1.80	0.39	21%	0.13	88
6/23/2011	1-6 PM	1.33	0.18	14%	0.06	82
Average	**1-6 PM**	**1.71**	**0.29**	**17%**	**0.10**	**88**

SMB load impact estimates can also be broken down by device type. Table 5-6 shows the load impact estimates for SmartAC SMB customers by event length (there was not sufficient data to calculate daily impacts by device type). Unlike for residential customers, device type did not definitively affect impacts for SMB customers. For two-hour events, PCTs showed slightly higher impacts than switches, but the opposite was true for five-hour events.

19 Reference loads are CAC-only loads.

Table 5-6: Average SMB per Account Impacts by Device Type

Event Type	PCT		Switch	
	Impact	% Impact	Impact	% Impact
2 hour events	0.31	18%	0.21	14%
5 hour events	0.29	16%	0.35	21%

For residential customers, a breakdown of impacts by usage decile was also included. However, this was not feasible for the SMB population. Dividing residential customers into deciles based on usage in June 2011 led to meaningful groups where usage on event days increased, on average, as usage deciles increased. This was not true for SMB customers. Because electricity usage varies much more for non-industrial customers and many other processes come in to play, whole-building usage in June 2011 was not a good predictor of average usage on event days.

5.3.2 Ex Post Impacts at the Industry Level

As mentioned in the methodology section, ex post percent impacts for each customer segment were based on impacts calculated at the territory level. This is due to the fact that when the M&E sample is broken down to the LCA or industry level, sample sizes are too small for accurate estimation. However, there are a few industry categories for which logger data can be used to examine ex post impacts. Which industry categories produced usable event impacts was determined visually based on how well treatment and control loads matched in pre-event hours and on whether there were any patterns during or immediately after the events that implied significant random noise affecting the results.

Table 5-7 shows the ex post impacts calculated for four industry types on two-hour event days using logger data.[20] The results show that percent impacts do vary across industries. Due to the partially subjective methodology and the small samples underlying the table, it is not prudent to base final estimates on these values. However, because the results are suggestive of important differences, they are interesting in light of what would be possible in the future with an evaluation based on SmartMeter data and larger sample sizes.

[20] Table 5-7 only shows two-hour event days because there were six event days on which to base the results. There were only three, five-hour event days so those results are less reliable.

Table 5-7: Industry Impacts on Two-hour Event Days Using Logger Data

Industry	Load w/o DR	Impact	% Impact	Avg. Event Temperature	# of Loggers (Treatment & Control)
Manufacturing	1.75	0.40	23%	91	22
Retail stores	2.50	0.54	22%	93	62
Offices, Hotels, Finance, Services	1.75	0.30	17%	91	329
Other or unknown	1.41	0.25	18%	89	131

5.3.3 SmartMeter Ex Post Impacts and the Impacts of Multi-device Premises

Although the main ex post analysis was conducted using logger data, SmartMeter data was also available for the majority of customers with loggers. Of the 367 customers with logger data, 274 (75%) also had SmartMeter data available. Analyzing SmartMeter data allows for an examination of aggregate impacts of multiple devices at one premise. Although logger results from CAC units at the same premise could also be aggregated, a maximum of five loggers were allowed per site in order to ensure that the logger sample included customers representing the whole service territory. This meant that sample customers with multiple CAC units did not have loggers on every unit. With not every unit logged, an analysis of logger data alone cannot reveal exactly what happens at premises with many devices. To begin with, the analysis of communication problems suggests that some units will not receive any given event signal. This effect probably differs across different sites according to where each control device is located. Even if every unit has perfect communication, it is likely that the cycling algorithm affects each unit somewhat differently which could have unpredictable results at the premise level. Analyzing SmartMeter data, and comparing the results to those from the logger analysis, allows for a judgment about the degree to which multi-device premises provide load in proportion to the number of devices.

Using logger data, per device impacts can be calculated and multiplied by the average number of devices per sample customer to get an estimate of per customer impacts. With SmartMeter data, per customer impacts are estimated and per device estimates can be calculated by dividing the per customer impacts by the average number of devices per sample customer. Since per customer and per device impacts can be calculated with both types of data, the results can be compared.

Table 5-8 presents per device and per customer impacts for both logger and SmartMeter data for the same sample of customers (the 274 customers that had both logger and SmartMeter data). This sample consists of all customers with valid logger data who also had SmartMeter data available for the summer of 2011. Impacts are presented as averages across five-hour and two-hour event days. Although impacts using logger data were calculated for individual event days, this is less feasible when using whole-building load data. SMB customers have widely varying usage patterns so even though the sample is randomly divided into two groups, average daily usage does not match well between the two groups on an event-by-event basis. By averaging usage across event days of the same length with alternating treatment groups, these differences tend to cancel each other out. It is important to keep in mind that this analysis was performed using a small sample of SmartMeter data. The larger

SMB SmartAC population could potentially provide enough data that these concerns would be minimized and SmartMeter-based impacts could be developed for individual event days.

SmartMeter and logger data both produce similar per device impact estimates. Looking at per customer impacts, the SmartMeter and logger impacts are also similar, which is necessarily true because the per customer values are equal to the per device values multiplied by the number of devices per customer. In both cases, impacts for two-hour events are larger than for five-hour events.

Table 5-8: Comparison of Ex Post Impacts from SmartMeter and Logger Data

Event Type	Average Per Customer Impacts		Average Per Device Impacts	
	SmartMeter Data	Logger Data	SmartMeter Data	Logger Data
5 hour	0.36	0.52	0.20	0.30
2 hour	0.58	0.58	0.33	0.33

Table 5-9 shows the results of testing whether the differences in impact estimates from the two methods are significantly different. Because standard errors for the SmartMeter estimates are so high, the power of the test is not very strong. However, the results do show that per device impacts for five-hour events were not statistically different between the two types of data.

Table 5-9: Test of Significant Difference between Per Device Impacts from SmartMeter and Logger Data

Event Type	SmartMeter Data (1)		Logger Data (2)		Difference between (1) and (2)	
	Impact	SE	Impact	SE	Difference	p-value
5 hour	0.20	0.65	0.30	0.06	-0.09	0.88
2 hour	0.33	0.42	0.33	0.04	0.00	1.00

These findings are interesting in light of what was found in the assessment of multi-device customers in the residential population. As discussed in Section 4.2.3, residential customers with two devices show similar impacts to customers with one device. This suggests that residential households with more than one CAC unit may not be cost-effective for the program because it costs more to install additional control devices but does not seem to provide greater impacts. This is not the case with SMB customers. These results show that SMB customers with multiple devices are providing approximately the same impacts per device as customers with one device. Unlike with residential customers, this suggests that multi-device SMB premises may be more cost-effective than single device premises. Each additional unit produces, on average, similar impacts as the first unit, with small additional installation costs.

5.4 SMB Ex Ante Methodology

This section explains how ex ante event impacts were calculated.[21] Impacts were calculated by making predictions for ex ante weather conditions using a regression model of ex post percentage impacts. Unlike previous years, no individual customer regressions were used. The ex ante weather conditions were the same that were used for the 2010 SmartAC evaluation and have been chosen to be representative of 1-in-2 and 1-in-10 monthly peak days and 1-in-2 and 1-in-10 typical event days.

Ex ante impacts were predicted using the same regression model that was used to calculate ex post percentage impacts for segments of the population below the whole service territory as described in Section 4.4 above. Again, ex post percentage impact estimates were taken as the dependent variable and *mean17* was used as the explanatory variable.

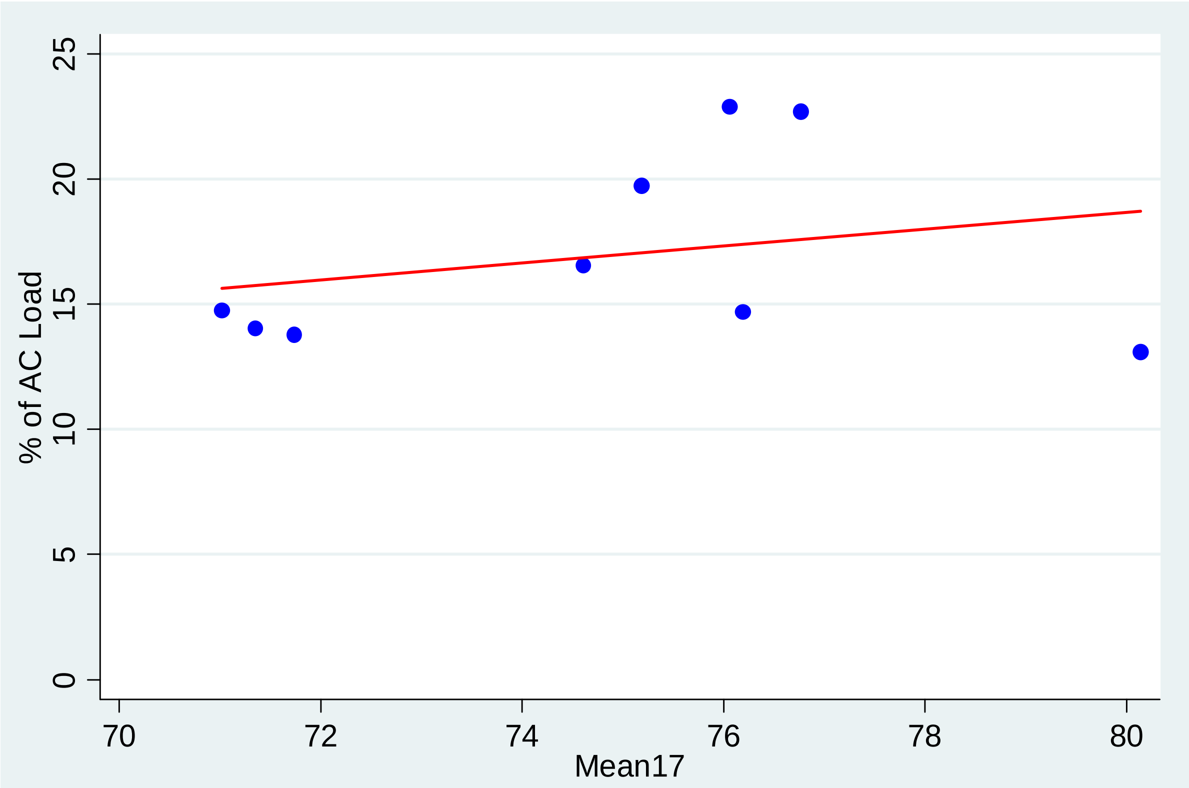
The results of the regression are shown in Figure 4-1. There are nine points in the graph, one for each event day of the summer. The general trend shows that as temperatures increase, so do impacts. However, there is a significant amount of variability around the trend line. For example, at *mean17* values near 76, there is one event with an average impact of 14% and 2 more with averages of about 23%. That is a 64% increase from the lower impact to the higher impacts, for effectively the same weather. In looking into these cases further, it was found that the weather the day before these events did not differ greatly. Also, the distribution of hot temperatures across different parts of the service territory (which could cause different event impacts for the same overall average weather) did not differ greatly.

Similarly, identical percentage impacts were observed on the hottest event day (June 21, with an average event temperature of 96°F and one of the coolest event days (June 23, with an average event temperature of 83°F). It is important to keep in mind that event impacts can vary significantly due to factors other than weather that may not be observable. This is also evident in the residential results. Moreover, with only a small number of events, it is likely that these results do not show the full degree to which individual event results can deviate from averages based on weather.

Over several years of data collection, the degree of variability in program impacts should become better understood. However, it is unlikely that there will ever be a way to drastically reduce this variability.

21 Explanations of how estimated CAC reference load and snapback were calculated are in Appendix B.

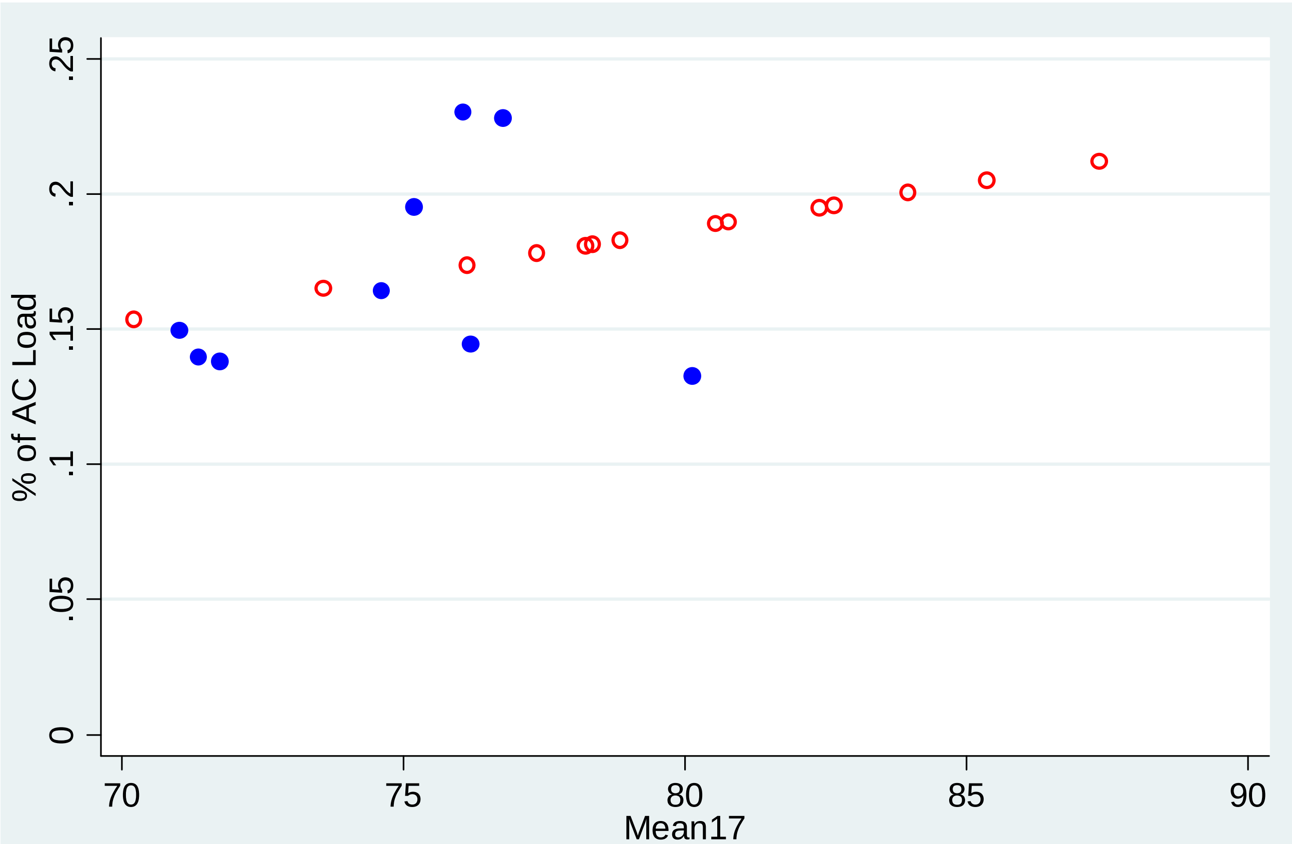
Figure 5-1:
Actual and Predicted SMB Ex Post Event Impacts as Percent of CAC Load



The results of this regression were used to predict all ex ante impacts – including those for LCA, industry and demand categories. Ex ante weather was calculated for the whole territory and for each LCA. Territory-wide regression estimates were then used to predict ex ante load impacts. Due to the small sample size, regressions could not be tailored to subsets of the SMB population.

Figure 4-2 displays the final territory-wide ex ante and ex post estimates graphed against *mean17*. The solid blue circles represent ex post results and the hollow red circles are ex ante results. Graphing both ex ante and ex post results on the same plot shows that the ex ante estimates follow the same trend as the ex post estimates.

Figure 5-2:
Ex Post and Ex Ante Impacts Versus *mean17*



Just as with the ex post results, the predicted percentage impact from 3 to 5 PM must be translated into hourly percentage impacts over the entire range of time required for prediction, 1 to 6 PM. The process is exactly the same as it was for ex post impacts, as discussed in Section 5.1.

Finally, predicted percentage load impacts must be converted to absolute load impacts. To get absolute load impacts, predicted percentage load impacts were applied to estimated reference loads. Reference loads were calculated for the ex ante analysis the same way they were calculated for the ex post analysis, as discussed in Section 5.1.

5.5 SMB SmartAC Ex Ante Load Impact Results

This section presents the ex ante impacts for SMB customers in the SmartAC program. As was the case for residential participants, aggregate estimates of load impacts combine estimates of per customer load impacts developed in this report with estimates of program enrollment, developed in a separate effort by PG&E.

Enrollment projections for SMB customers by local capacity area as of August of each year are presented in Table 5-1. The source for these projections is PG&E's enrollment projections for 2012-2022, developed in conjunction with the Brattle Group. SMB enrollment is projected to decline steadily until 2015 at which point it is projected to remain fairly stable, since SMB SmartAC is closed to new customers pursuant to CPUC Decision 12-04-045

Table 5-10:
Projected SMB Enrollment for August of Each Year

LCA	2012	2013	2014	2015-2022
Greater Bay Area	2,050	1,821	1,623	1,561
Greater Fresno	838	744	663	638
Kern	248	220	196	189
Northern Coast	688	611	545	524
Other	1,196	1,062	947	910
Sierra	482	428	381	367
Stockton	408	362	323	311
Total	5,910	5,248	4,678	4,500

Table 5-2 shows the per-customer and aggregate ex ante impact estimates. For the 1-in-2 weather year, the highest aggregate mean hourly impact occurs on the July peak day, with an impact of 4.2 MW. The highest individual hourly impact during a 1-in-2 year is also the July value – 4.9 MW. The July peak day also shows the highest impacts for the 1-in-10 weather year. The largest aggregate impact over the five-hour event is 4.8 MW and highest individual hour provides an estimated impact of 5.6 MW.

Table 5-11:
2012 SMB SmartAC Load Impact Estimates
By Weather Year and Day Type
(Event Period 1 to 6 PM)

Weather Year	Day Type	Mean Hourly Per Customer Impact	Max. Hourly Per Customer Impact	Aggregate Mean Hourly Impact	Aggregate Max Hourly Impact
		(kW)	(kW)	(MW)	(MW)
1-in-2	Typical Event Day	0.55	0.65	3.3	3.9
	May Peak Day	0.37	0.45	2.3	2.8
	June Peak Day	0.47	0.56	2.9	3.4
	July Peak Day	0.69	0.81	4.2	4.9
	August Peak Day	0.55	0.65	3.2	3.8
	September Peak Day	0.51	0.61	3.0	3.5
	October Peak Day	0.32	0.39	1.8	2.2
1-in-10	Typical Event Day	0.72	0.85	4.3	5.0
	May Peak Day	0.63	0.74	3.9	4.6
	June Peak Day	0.66	0.78	4.0	4.8
	July Peak Day	0.79	0.93	4.8	5.6
	August Peak Day	0.76	0.88	4.5	5.2
	September Peak Day	0.61	0.72	3.5	4.2
	October Peak Day	0.50	0.59	2.8	3.3

The Typical Event Day estimates are based on projected August 2012 enrollment. The excel-based load impact tables that accompany this report provide substantially more information about the ex ante conditions, load shapes and impact estimates than is practical to include here. In interpreting those results it is important to keep in mind that loads and load impacts depend not only on the temperature during the event, but also on the temperature before the event. Two days with the same high temperature can provide very different load impacts if one of them has much higher overnight low due to heat retention in buildings.

The improvement in load impact estimates since the 2010 evaluation is discussed extensively elsewhere in this report. It is worth mentioning, however, that ex ante estimates for SMB customers did not show as much improvement since the 2010 evaluation as did ex post estimates because some improvement was assumed for ex ante estimates in the 2010 evaluation. See that evaluation for more detail. Nevertheless, improvement in ex ante estimates still occurred because the actual level of improvement exceeded the assumed amount in the 2010 evaluation.

6 Conclusions

The 2011 evaluation of the SmartAC program has revealed large improvements in program performance. As evidenced by the section on device communication, there is still the potential for improvements in the future. The SMB segment of the program has substantial room for operational improvement through better communication. The residential segment of the program could probably be improved through targeting high energy users in future marketing.

This evaluation has also provided a substantially greater amount of data on program performance than has ever been possible previously. The multi-device premise results, side-by-side testing results and the increased certainty about ex post results are all elements of this report that were never feasible previously when SmartMeter interval data was not yet available.

It has already been stated that PG&E's commitment to fully understanding this program has led to the program performance gains so far and to the possibility of future gains. It is also worth noting that this commitment to understanding also lays bare problems with the program that likely exist in other similar load control programs but that are never understood or addressed.

Appendix A. Residential Ex Post Load Impact Tables Methodology

A.1. Estimating Ex Ante Load Without DR

Although estimating impacts is the most important part of the ex ante analysis, whole-building reference load data is needed to illustrate the magnitude of impacts. This estimation took place in three steps:

1. The average hourly usage for each LCA was calculated based on control group load for all seven event days. This provides an average hot-day load shape, but does not account for temperature variation;
2. Next, a regression model, which was similar to the one used to predict load impacts, was also used to model average whole-building loads from 4 to 6 PM. The regression had the same form and the same independent variable as the load impact regression. Only the dependent variable was different. Also, each regression was estimated only at the LCA level – no pooled estimates were used – and the values for whole-building load were not capped. This model was used to predict average loads without demand response from 4 to 6 PM for each set of ex ante weather conditions; and
3. Finally, each LCA's control load during each hour for each set of ex ante conditions was adjusted up or down by the ratio of the load predictions from step 2 by the average building load from 4 to 6 PM in step 1.

Figure A-1 depicts the process used to calculate the load shapes for ex ante results. As an illustrative example, the figure shows the ex ante scenario for the typical event day for the Greater Bay Area during a 1-in-2 weather year. The solid purple line shows the average load shape for all Greater Bay Area control group customers over the seven events during the summer of 2011. The purple circle shows the average usage from 4 to 6 PM over all 2011 event days while the green square shows the predicted average usage from 4 to 6 PM for the typical event day in a 1-in-2 weather year for the Greater Bay Area. Finally, the dotted green line shows the average control usage adjusted upwards using the ratio between the green square and the purple circle (represented by the black bracket). The values represented by the dotted green line are the load without demand response.

Figure A-1:
Graphic Depiction of Control Load Calculations
Greater Bay Area, 1-in-2 Weather Year, Typical Event Day

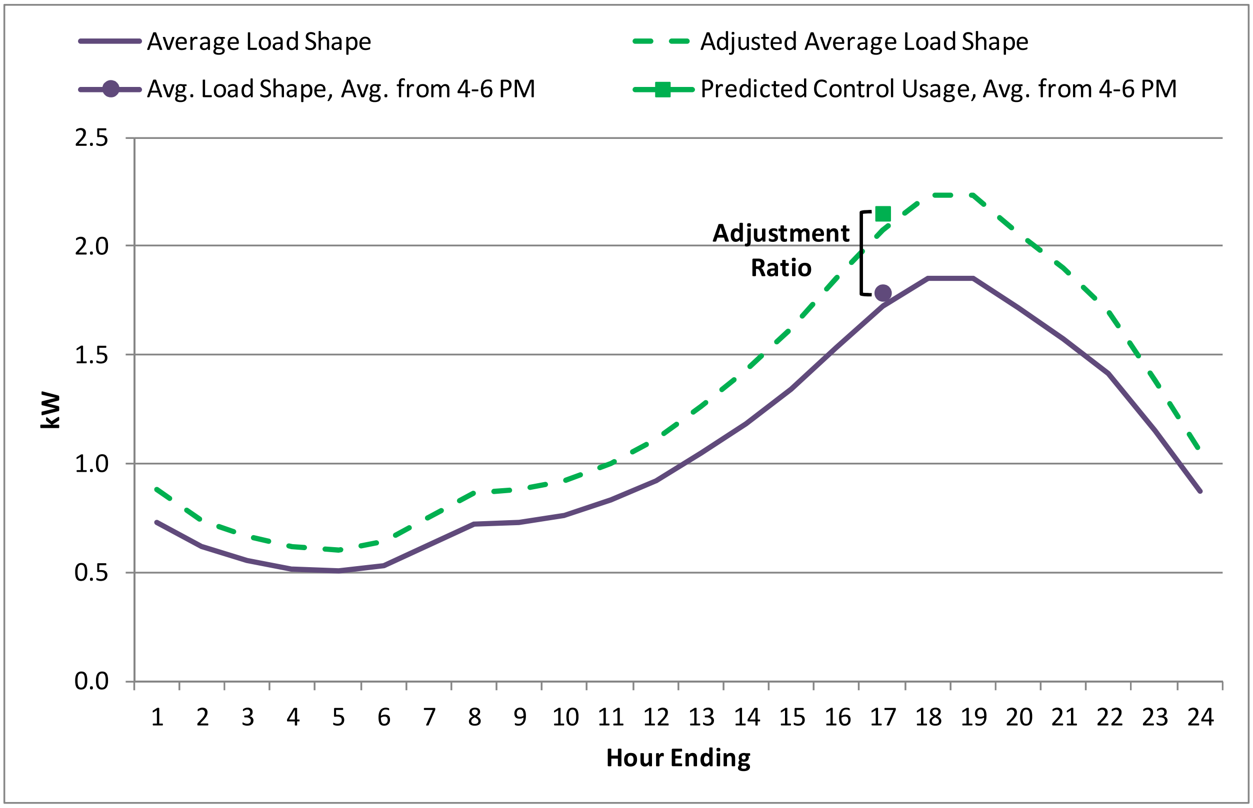
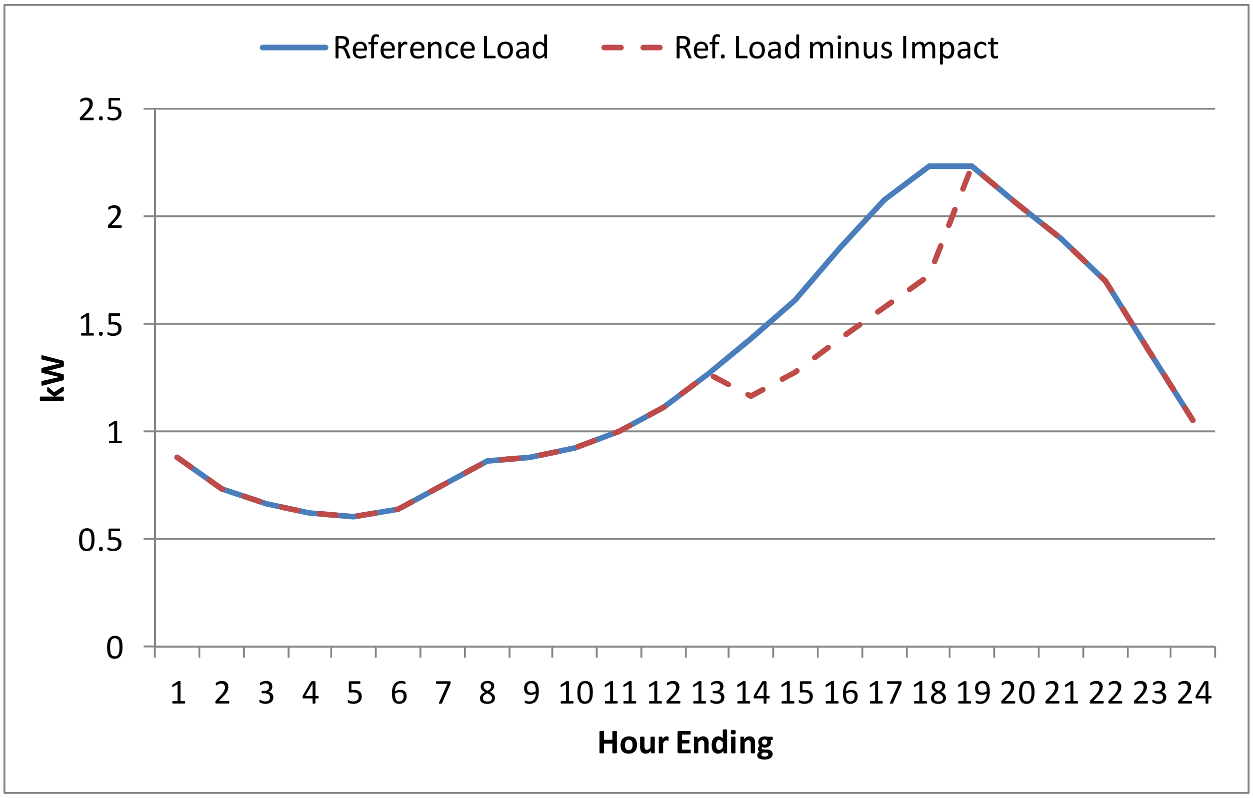


Figure A-2 shows the next step in creating the ex ante tables. As an example, it shows the Greater Bay Area under 1-in-2 weather conditions for the typical event day. The figure shows the loads as exactly the same for all hours except during the event, where the magnitude of the impact has been subtracted from the reference load to create the event load.

Figure A-2:
Graphic Depiction of Ex Ante Impact Calculations
Greater Bay Area, 1-in-2 Weather Year, Typical Event Day

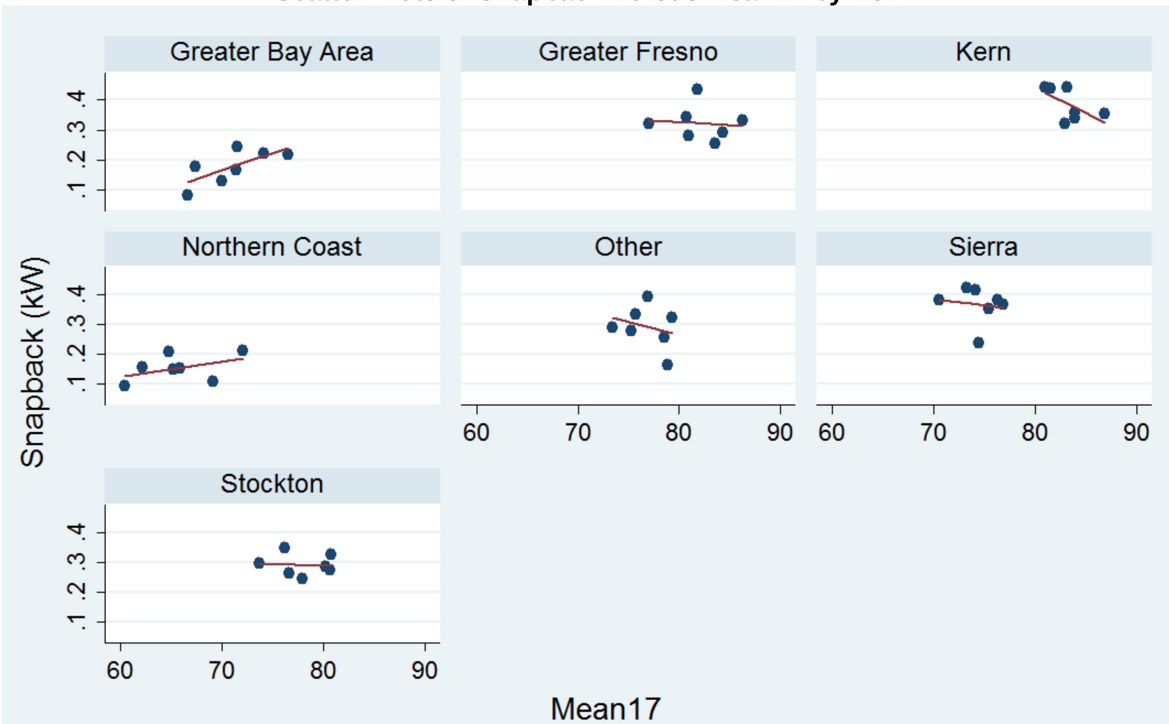


A.2. Estimating Ex Ante Snapback

As the final step in the ex ante analysis, snapback loads are predicted for all hours after the event ends. Snapback was not found to be a consistent function of temperature. Figure A-3 shows the scatter plot of snapback – measured as the average difference between reference load and event-day load during the first post-event hour – versus *mean17* for each LCA. The figure shows that the relationship varies across LCAs. For example, in the cooler LCAs (Greater Bay Area and Northern Coast) higher temperatures over the 17 hours before the event are associated with larger snapback. For the other five LCAs, where temperatures were warmer, snapback is fairly consistent across temperatures or even tends to be lower at higher temperatures. It is likely that when a CAC is controlled for an event, the building becomes hot enough that the CAC turns on full blast during the hour after the event is over. Regardless of whether it is 95°F or 105°F, the CAC will work at its maximum capacity for the hour after the event.[22]

[22] This statement is a hypothesis based on the data currently available. In future evaluations, more data will be available to better test this idea.

Figure A-3:
Scatter Plots of Snapback Versus *Mean17* by LCA



Perhaps with more data in future years, a regression would be able to accurately model snapback over the full spectrum of temperatures for each LCA. However, for this year's analysis, the average snapback across all event days for each LCA was used for ex ante prediction.[23] Table A-1 shows the average snapback in the first hour after the event for each LCA.

Table A-1:
Average Snapback From 6 to 7 PM by LCA

LCA	Average Snapback From 6–7 PM (kW)
Greater Bay Area	0.18
Greater Fresno	0.32
Kern	0.39
Northern Coast	0.15
Other	0.29
Sierra	0.36
Stockton	0.29

[23] Although the first three events are five hours long and the last four are only two hours, a side-by-side experiment was conducted on June 21, 2011 that showed the snapback for five hour and two hour events was nearly identical.

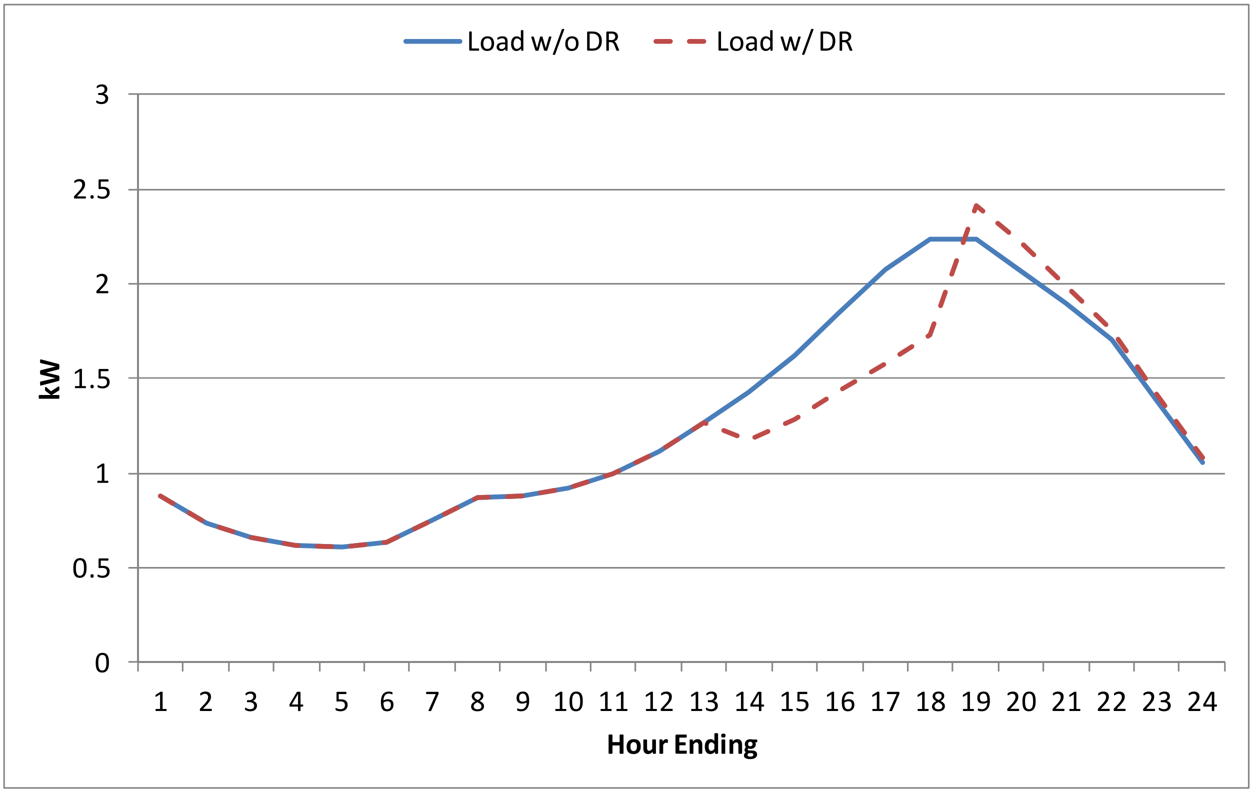
Just as with event load impacts, the average snapback for 6 to 7 PM was translated to hourly snapback using the ratio of average snapback in each hour to average snapback from 6 to 7 PM. Table A-2 shows these ratios for each LCA. For the Greater Bay Area, for example, the table shows that the snapback from 7 to 8 PM is 95% of the snapback from 6 to 7 PM. Multiplying this ratio by the value in Table A-1, the snapback from 7 to 8 PM is 0.171 kW.

Table A-2:
Hourly Snapback Compared to Average Snapback from 6 to 7 PM

Hour	Greater Bay Area	Greater Fresno	Kern	Northern Coast	Other	Sierra	Stockton
6- 7 PM	1.00	1.00	1.00	1.00	1.00	1.00	1.00
7- 8 PM	0.88	0.95	0.98	0.84	0.88	0.82	0.95
8- 9 PM	0.50	0.66	0.67	0.53	0.52	0.52	0.63
9- 10 PM	0.30	0.40	0.46	0.31	0.33	0.31	0.45
10- 11 PM	0.19	0.23	0.34	0.24	0.21	0.21	0.24
11 PM- 12 AM	0.15	0.17	0.27	0.23	0.13	0.11	0.19

Figure A-4 shows the final ex ante results for the Greater Bay Area typical event day during a 1-in-2 weather year. All hours leading up to the event have exactly the same load with and without demand response. For the event hours, impacts are subtracted from the reference load as described above. For hours after the event, the snapback is added to the reference load based on the calculations also described above. This produces the estimates of load with DR for the post-event hours.

Figure A-4:
Ex Ante Results Example
Greater Bay Area, 1-in-2 Weather Year, Typical Event Day



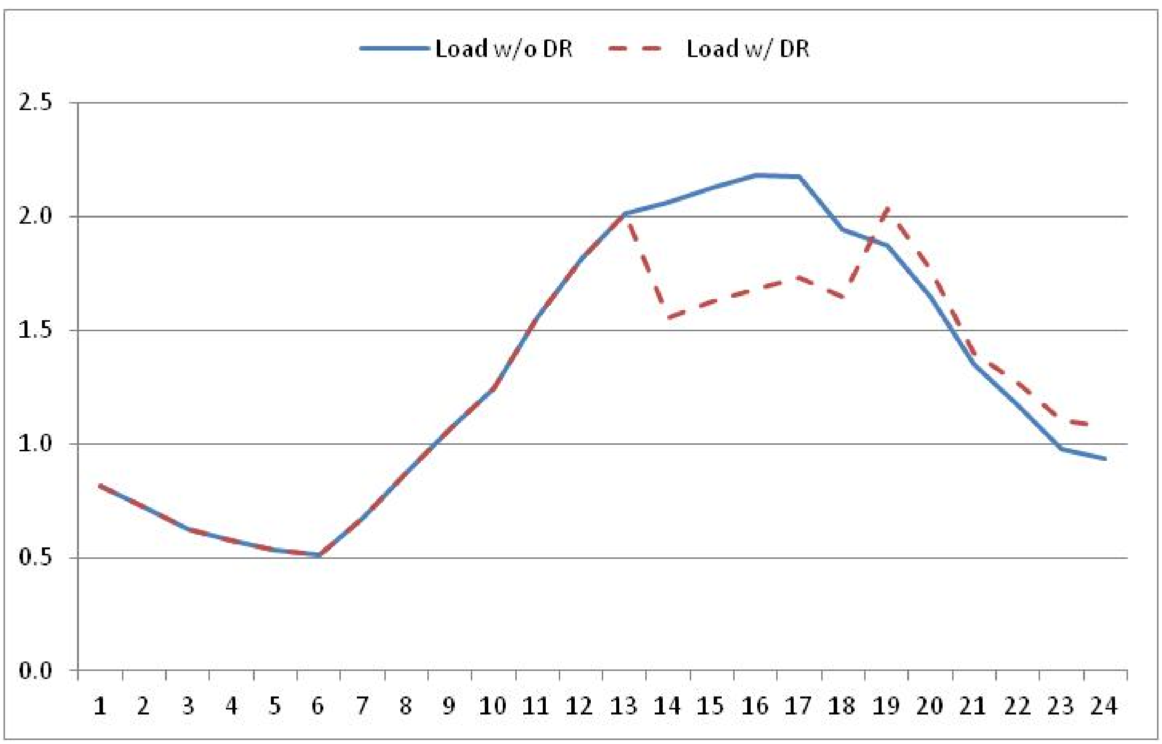
Appendix B. SMB Ex Post Load Impact Tables Methodology

For SMB customers, reference loads were based on observed load on non-event days from the load research samples in 2009, 2010 and 2011. Average load was calculated for bins of *mean17* temperature. For example, if the event day in Greater Fresno had a *mean17* of 91°F, then the reference load for the ex post results would come from the 90-92°F bin. Additionally, snapback for all hours after the event was calculated based on the average observed snapback for five-hour and two-hour events. Snapback was calculated as a percentage of reference load.

Finally, the predicted impacts and snapback were combined with estimated reference loads to create the input for ex post tables. First, impacts were subtracted from the reference load to calculate load with DR. Next, the reference load was multiplied by the average snapback as a percentage of reference load to get the magnitude of predicted snapback. Finally, snapback was added to the reference load in all post-event hours to get load with DR.

Figure B-1 illustrates an example of the final outcome of this process. It shows the ex post results for Kern on June 21, 2011. The average load for both the treatment and control group is the same for all hours leading up to the event. Over the five-hour event period, the distance between the solid blue line and the dotted red line represents the impacts for each hour. For the last six hours of the day, the difference between the two lines shows the snapback.

Figure B-1: Ex Post Figure for Kern on June 21, 2011



Appendix C. SmartAC Side-by-side Testing Results

This section presents the ex post load impacts of the side-by-side test events for residential customers during the summer of 2011. During these tests, two random groups of customers, based on the serial number of the control device as described above, were called for events with different control strategies or timing between the groups. The average loads for the customers in each group during and after the event can be directly compared to each other and to the control group loads to measure the effect of the different cycling strategies or event timing. The virtue of this approach is that the only difference between the groups is the control strategy or timing. There are no confounding effects, such as differing populations or weather, which could confound the interpretation of differences between the groups' load impacts. The procedure constitutes a true, randomized experiment on the variables of interest.

Two different types of side-by-side tests were performed in 2011. First, on both June 21 and June 22, a five-hour event from 1 to 6 PM was tested alongside a two-hour event from 4 to 6 PM. The purpose of this test was to find out whether load impacts during the residential peak hours from 4 to 6 PM varied substantially based on whether the event had already been ongoing for several hours prior to that time. This is an important question because the time at which the program is most likely to be needed is in the late afternoon or early evening, but it could also be important for the program to be able to supply load reductions earlier in the day as well. The results of this testing show that the load reduction from 4 to 6 PM is largely unaffected by whether the event began at 1 PM or 4 PM. A secondary finding from this testing is that the snapback effect from a two-hour event is roughly as large in the first post-event hour whether the event lasted five hours or two hours. This is useful information for the program operator if there is any concern about post-event snapback putting strain on the distribution system.

Second, on June 21, the 50% TrueCycle2 control algorithm was tested against 50% simple cycling during a five-hour event. The TrueCycle2 algorithm was found to provide, on average, 70% greater load reduction over the entire event. This result is important because it shows the value of using the TrueCycle2 algorithm, which has costs in terms of implementation complexity, as compared to simple cycling. Again, despite larger load reductions, the first-hour snapback effect was similar between the two strategies.

C.1. Five-hour Versus Two-hour Event Impacts

FSC examined how impacts varied from 4 to 6 PM for five-hour and two-hour events. Two side-by-side tests of five-hour versus two-hour events were called – one on June 21 and one on June 22. The interval data for the customers involved in the test events was pooled for both days so the impacts reported are the average over two days. Examining the data for each day separately does not alter any of the conclusions. Table C-1 shows average event impacts for each hour of the events, mean and maximum impact for the period from 4 to 6 PM and snapback after the event ended. Impacts in the table are per device impacts for single-device customers. For the two hours that the events overlapped, there is a 7% difference in impacts. The five-hour event had slightly higher impacts during this period than the two-hour event, at 0.65 kW and 0.61 kW, respectively. Both events had nearly the same maximum impact, around 0.65 kW. Finally, snapback for the five-hour event was also about 7% higher than snapback for the two-hour event.

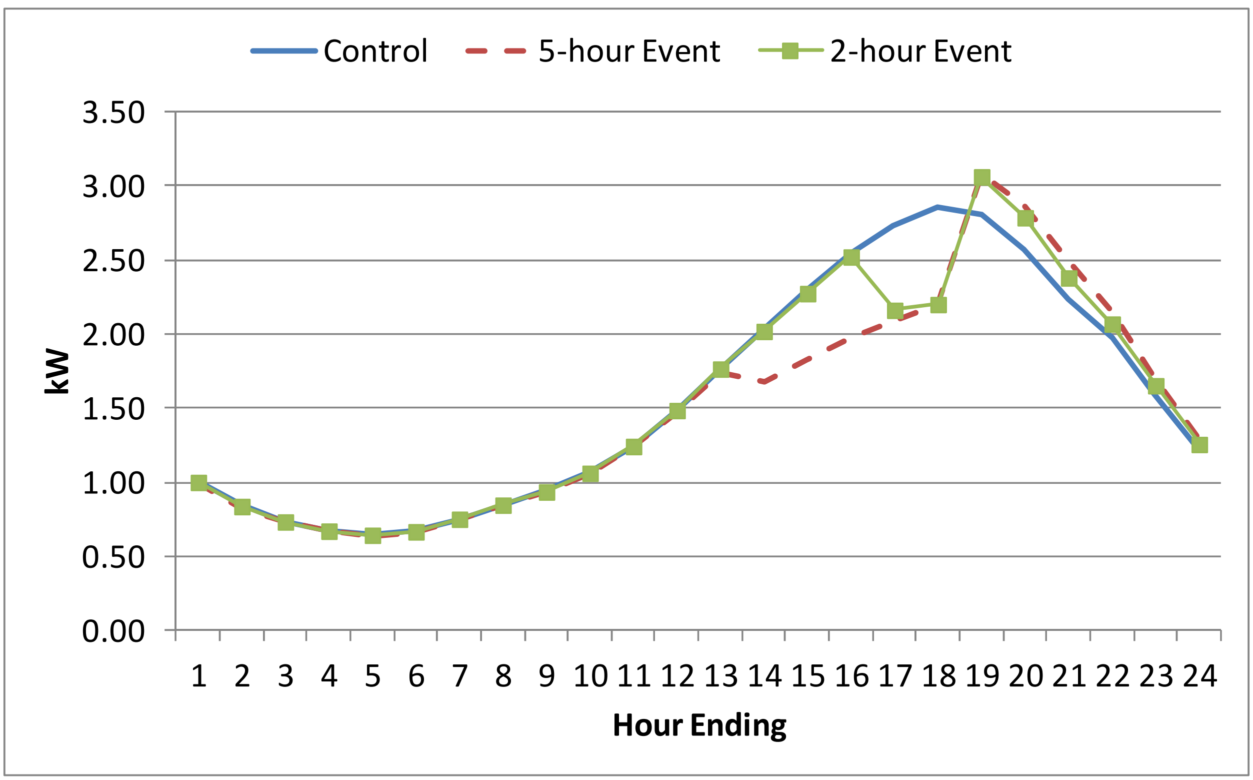
Table C-1: Side-by-side Event Impacts for Five-Hour Versus Two-Hour Events (kW)

Hour	Five-hour True Cycle	Two-hour True Cycle	% Difference
1-2 PM	0.36	n/a	n/a
2-3 PM	0.47	n/a	n/a
3-4 PM	0.56	n/a	n/a
4-5 PM	0.65	0.57	13%
5-6 PM	0.66	0.65	1%
Mean Impact*	0.65	0.61	7%
Maximum Impact*	0.66	0.65	1%
Snapback	0.28	0.26	7%

*Calculated from 4- 6 PM for five-hour event

Figure C-1 shows the results of this side-by-side test graphically. The green dotted line and the red line with markers show that the impacts from 4 to 6 PM are very similar for both event groups. The results show that whether a two-hour or a five-hour event is called, PG&E will likely see comparable impacts during the 4 to 6 PM time period.

Figure C-1: Side-by-side Event Impacts for Five-hour Versus Two-hour Events June 21 & 22



C.2. TrueCycle2 Versus Simple Cycling Event Impacts

The side-by-side test of the 50% TrueCycle2 cycling algorithm with 50% simple cycling was performed for residential customers from 1 to 6 PM on June 21. This analysis was only performed on switches.[24] Impacts in the table are per device impacts for single-device customers. Table C-2 shows the event impacts for the side-by-side events. On average over the five hours, the TrueCycle2 algorithm produced 70% higher impacts in switches than simple cycling. Additionally, the maximum impact for TrueCycle2 was 0.74 kW compared to 0.41 kW for simple cycling, a difference of 72%.

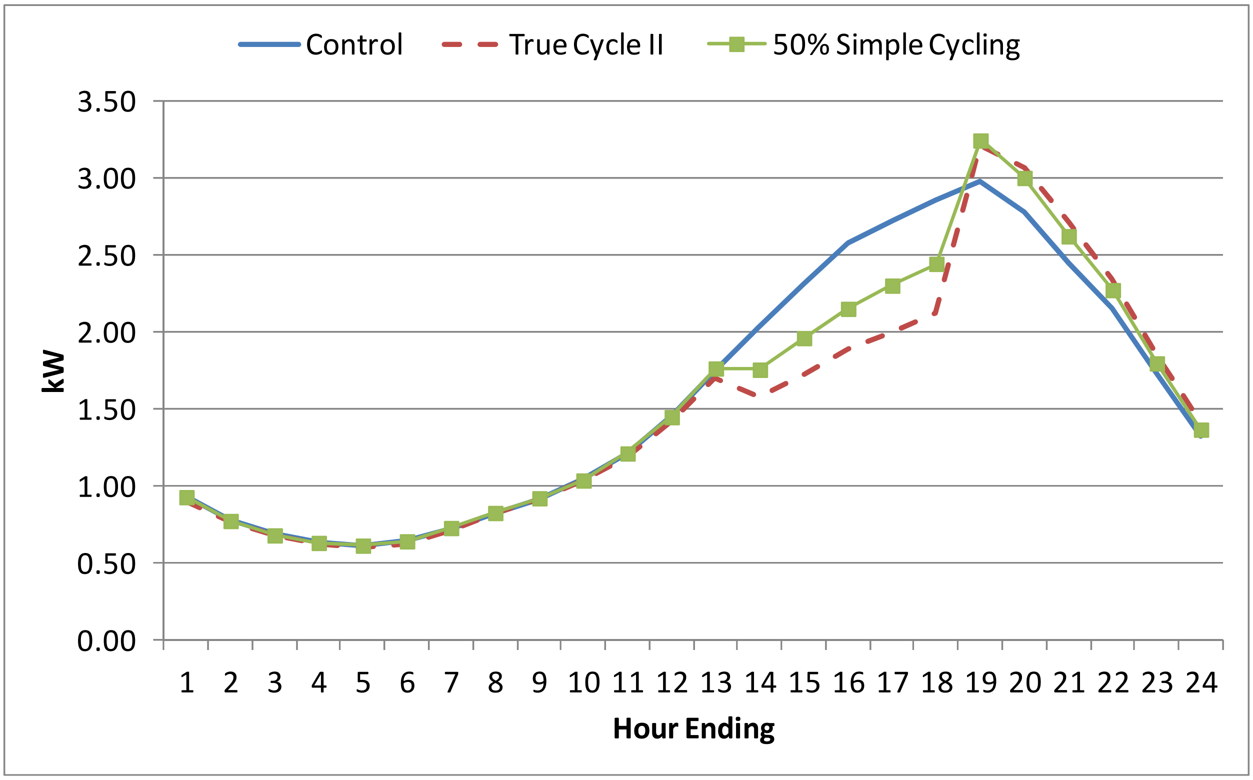
Table C-2: Side-by-side Event Impacts for TrueCycle2 Versus Simple Cycling (kW)

Hour	TrueCycle2	Simple Cycling	% Difference
1-2 PM	0.45	0.27	66%
2-3 PM	0.59	0.35	70%
3-4 PM	0.70	0.43	62%
4-5 PM	0.72	0.42	72%
5-6 PM	0.74	0.41	80%
Mean Impact	0.64	0.38	70%
Maximum Impact	0.74	0.43	72%
Snapback	0.23	0.27	-14%

Figure C-2 illustrates the average event impacts on June 21 for the five-hour TrueCycle2 event and five-hour simple cycle event. The figure shows that the impacts under TrueCycle2 are much greater during all five hours of the event. Snapback for all hours after the event, however, is nearly the same for both strategies. The results of this test show that TrueCycle2 provides substantial additional load drop and is a highly valuable tool. As shown elsewhere in this evaluation, it appears that the transition to TrueCycle2 is the cause of a large increase in SmartAC load reductions in 2011.

[24] In the early stages of evaluation planning, there was a decision to test some form of cycling against 2-1-1 temperature setback for PCTs. This plan was abandoned when it was decided to run 50% TrueCycle2 for Utility Pro PCTs and 50% simple cycling for Express Stat PCTs for every event. Due to this change in plans, the experimental apparatus to test TrueCycle2 against simple cycling for Utility Pros was not in place for 2011. Such a test could be run in 2012. Express Stat PCTs cannot run TrueCycle2.

Figure C-2: Side-by-side Event Impacts for TrueCycle2 Versus Simple Cycling June 21



Appendix D. Analysis of the Target Cycle Control Strategy

Target Cycle is an advanced cycling technique that allows a utility to specify the amount of load reduction desired from the individual CAC unit. For example, a unit can be directed to provide 1 kW of load reduction. The control algorithm then uses a baseline methodology to predict what the unit's usage would have been in the absence of control. The algorithm then reduces the unit's usage to 1 kW less than the predicted usage. The algorithm is meant to provide the chosen amount of load reduction on average each hour.

The algorithm can work in two ways: with and without learning days. If learning days are employed, the control device records the CAC unit's usage on hot non-event days (chosen by the program operator) and uses this information to predict usage during event periods. In the absence of learning days, the control device assumes a 100% duty cycle for a baseline, and then adjusts that prediction based on a multiplicative same-day adjustment of the load using the hour before the event as a reference. For example, suppose no learning days have taken place, an event occurs from 4-6 PM and the control device is directed to provide 1 kW of demand response. The algorithm will adjust usage downward to 1 kW less than the load observed during the hour 3-4 PM.

The following analysis uses CAC logger data to estimate Target Cycle impacts on a sample of 9 CAC units over 15 events during the summer of 2011. Each event was two hours long and Target Cycle was directed to provide 0.5 kW of load impact during each event. The small sample of loggers complicates the analysis, because the error inherent in the estimation methods can be indistinguishable from the actual variance of impacts across customers and event days. For this reason, the conclusion in this section must be considered tentative.

D.1.1. Analysis Method

Of the nine loggers for which data was retrieved: two were excluded from the analysis; one CAC unit was idle for the duration of the summer; and the other CAC was generally not in use in the hours leading up to the event, and in most cases, not in use during the event. Hourly usage data for the remaining seven loggers were then analyzed to estimate hourly impacts for each customer during each event.

Events were called over a range of temperature conditions. However, in order to provide 0.5 kW during an event, the temperature must be hot enough for the CAC to be at least using 0.5 kW. This constraint led to the exclusion of 7 of the 15 event days because CAC loads were too low on those days for Target Cycle to work.

The primary analysis was conducted using individual customer regressions. The regression models used take into effect variations in temperature, hourly and monthly effects. The specification used and variable descriptions are provided below:

$$kW_t = A + \sum_{i=1}^{24}\sum_{j=1}^{12} B_{ij} \times Hour_i \times Month_j \times CDH_t + \sum_{i=1}^{24} C_i \times Hour_i \times LagCDH_t + \sum_{i=17}^{20}\sum_{j=1}^{15} D_{ij} \times Hour_i \times Event_j + e_t$$

Variable	Description
kW_t	Hourly kW measurement at time t
A	Estimated constant term
B through D_{ij}	Estimated parameters
$Hour_i$	Indicator variables for each hour
$Month_j$	Indicator variables for each month
CDH_t	Cooling degree hours (base 75) per day at time t
$LagCDH_t$	Weighted average of cooling degree hours (base 90) prior to time t
$Event_j$	Indicator variables for each event.
e_t	error term

In addition to regression models, impacts were also estimated by comparing the average usage during events to baselines. To calculate the baselines, each customer's usage values on all non-event weekdays with maximum temperature greater than 90°F are averaged together. An adjustment is determined by calculating the difference between the baseline and the average usage in the two hours preceding the events. The adjustment is then added to the baseline so that the baseline usage matches the average pre-event usage in the hours leading up to the event. Baseline impacts are then calculated by taking the difference between the baseline and the average usage during event hours.

D.1.2. Results

Individual customer regressions were used to estimate event impacts. Figure D-1 shows the distribution of estimated hourly impacts for each qualified event day, for each customer. The distribution's mean is 0.71 kW, above the target of 0.5 kW, with a standard deviation of 0.86 kW. The standard deviation is large relative to the size of the estimated impacts, which is to be expected. Any baseline methodology, such as the one that Target Cycle uses, relies on averaging over many customers to be accurate. Such a method should not be expected to be accurate at the level of the individual customer. The observations above 2 kW are the result of a single unit, and negative observations are probably a result of model variance, which is to be expected when trying to predict a customer's usage for a single day.

Figure D-1: Histogram of Event Impacts

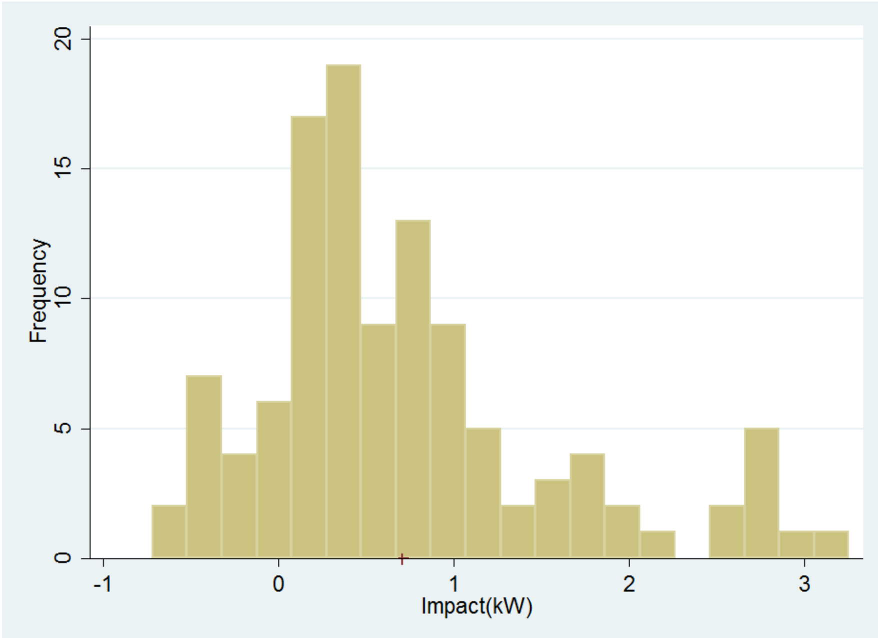


Figure D-2 shows predicted load with and without the Target Cycle intervention averaged across all units and all events. The load profiles are typical for smart CAC customers; however, the impact achieved is significantly higher for Target Cycle on comparable event days (September 6 and 7) when compared to a sample of SmartAC customers from the same weather station.

Figure D-2: Average Load With and Without DR

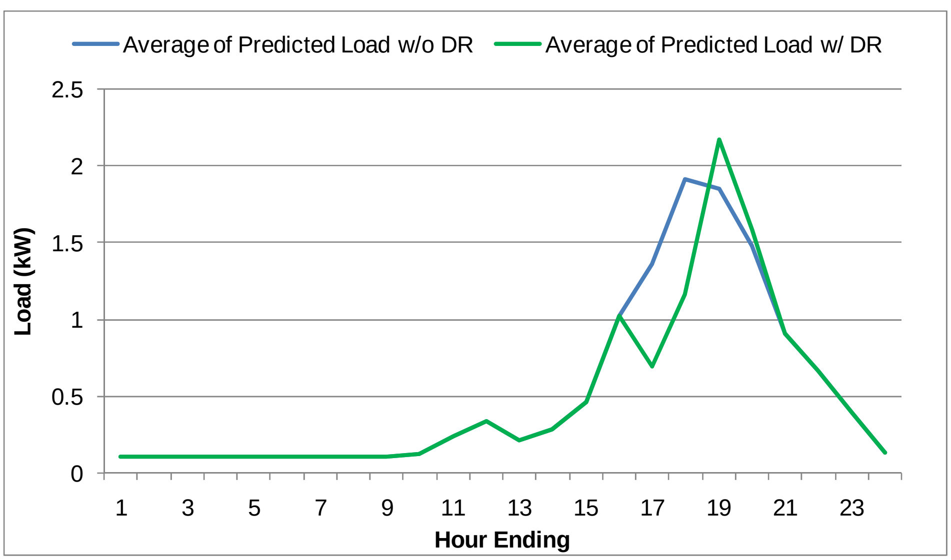


Table D-1 shows the hourly impacts over the 8 event days with temperatures above 90°F. Average impacts are frequently above the target; however, this is the result of a single customer with estimated impacts between 2-3 kW.

Table D-1: Hourly Impacts for Qualifying Events

Date	Hourly Impacts (kW)		Maximum Temp
	4-5 PM	5-6 PM	
23-Aug-11	0.52	0.78	94.0
2-Sep-11	0.81	0.58	94.5
6-Sep-11	0.92	0.68	95.5
7-Sep-11	0.82	0.74	94.0
19-Sep-11	0.44	0.15	94.0
20-Sep-11	0.93	1.68	97.5
22-Sep-11	0.74	0.81	93.5
28-Sep-11	0.14	0.59	95.0
Average	**0.66**	**0.75**	**94.8**

Table D-2 shows the average impact estimate for each logger, under both the regression and baseline estimation methods. In this case, the baseline estimation method produced smaller estimates; however, the directions and magnitude of the differences in estimates were relatively consistent across all loggers. Note that the baseline results are not calculated for individual event days because the baselines are only accurate when taking averages across many days.

Table D-2: Comparison of Estimated Impacts to Baseline Estimates

Logger ID	Average Hourly Impacts (kWh)	
	Regressions	Baseline
2323693	1.97	1.82
2323708	0.91	0.69
2323788	0.44	0.19
2323808	0.38	0.19
2323823	0.40	0.29
2324412	0.27	0.20
2325110	0.57	0.38
Average	**0.71**	**0.54**

D.1.3. Conclusion

The small sample used in the Target Cycle analysis could only provide definitive answers about the efficacy of the Target Cycle program in an extreme case of failure. This analysis would never be able to provide strong evidence of success due to the underlying variability of the baseline method the algorithm uses and the variability in CAC usage. This second source of variability ensures that models with a small number of data points will have substantial unexplainable variation, limiting the conclusions that can be drawn.

Target Cycle did not produce consistent 0.5 kW reductions across events or customers. Averaged across customers, the sample produced impacts that exceeded the target by more than 20%. One customer in particular produced impacts far in excess of the target, with average impacts of 1.97 kW. However, given the size of the sample, and the inherent error in the analysis, it is virtually impossible to reject the hypothesis that Target Cycle *could* provide an average 0.5 kW reduction across customers on a much larger sample of customers. If PG&E has an interest in pursuing the use of Target Cycle, a larger test is warranted.

Appendix E. Differences in Impacts and Methods Between 2011 and Previous Evaluations

The impact values estimated in this evaluation are substantially larger than in 2010 for both residential and SMB customers. The methods used for impact estimation also differed. This Appendix discusses the most likely reasons for the increase in impacts and also discusses the differences in methods used between this evaluation and previous evaluations. The two main conclusions of this section are that the increase in load impact estimates for both customer segments was mainly due to changes in control strategies, and that the new methods used in this evaluation are an unambiguous improvement over previous methods, enabled by the availability of much greater amounts of data.

First the changes in program impacts are discussed and then the differences in methods are discussed.

E.1. Source of Increase in Impacts from 2010 to 2011

An important difference between the 2010 and 2011 SmartAC load impact estimates is that the 2011 values are substantially higher under comparable conditions. This is true both for residential and SMB customers. Figure E-1 shows 2010 and 2011 residential ex post impacts as a function of daily high temperature. The values in 2011 are about 73% higher, while the average maximum temperatures on event days in the two years are almost identical at 94°F. To account for differing event timing during the day, average impacts are only calculated for the hours 4 to 6 PM. Figure E-1 also shows trend-lines through each year's impacts.

Figure E-1: 2010 and 2011 Residential Ex Post Impact Estimates Versus Daily Maximum Temperature

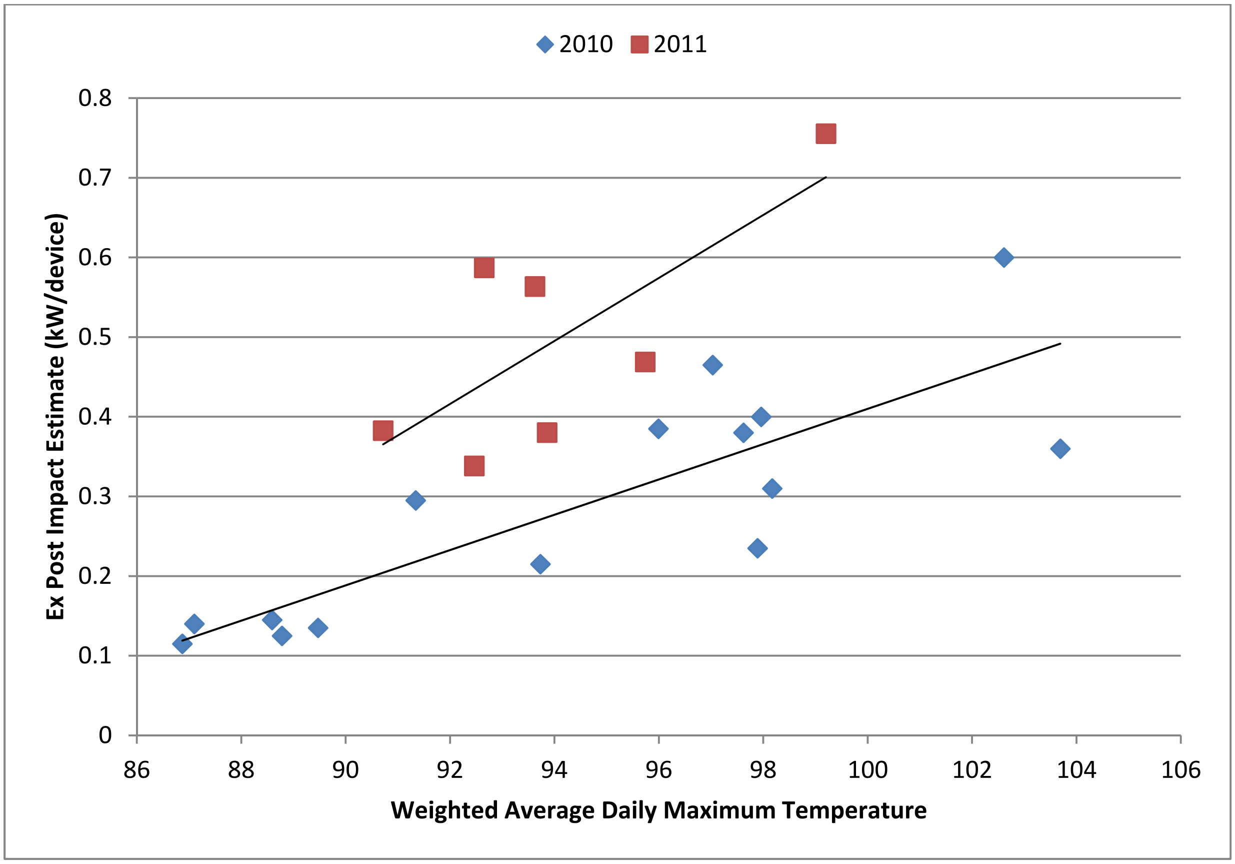
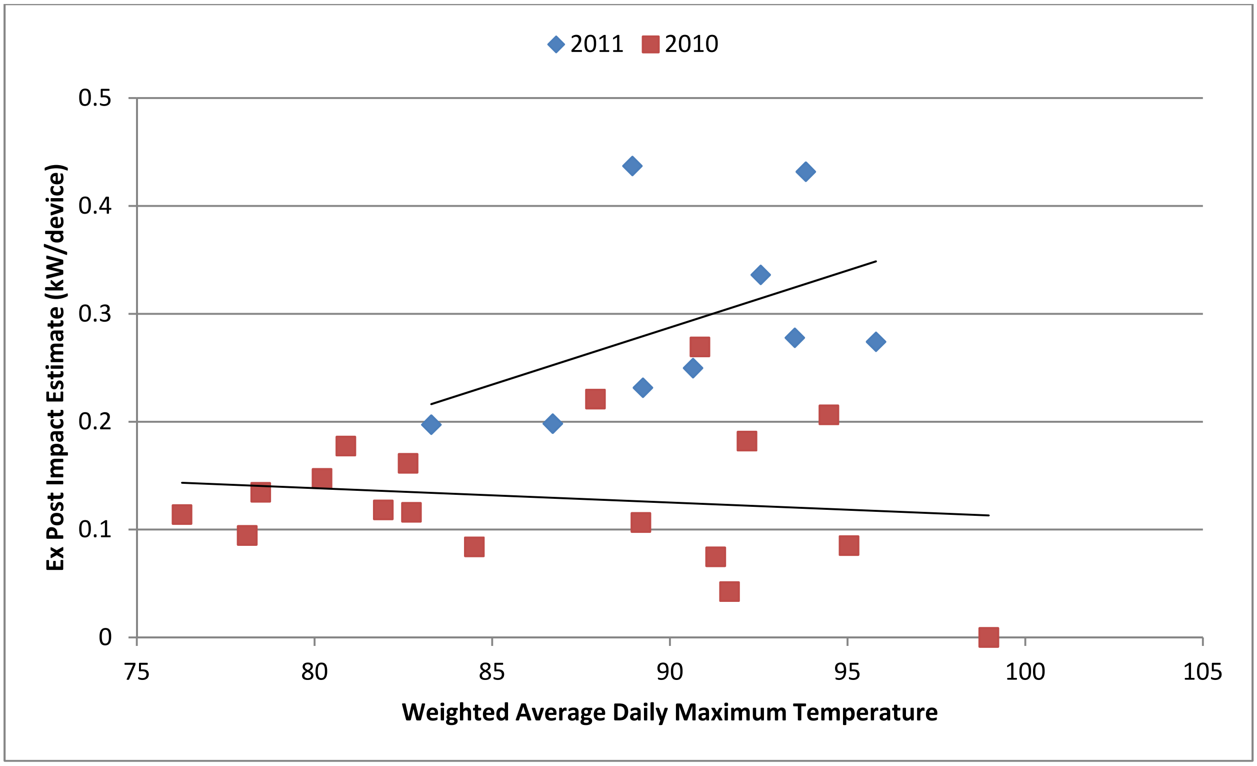


Figure E-2 shows the same set of relationships for SMB customers in 2010 and 2011. Here, only the hours from 3-5 PM are used for averaging because each event included those hours. The SMB values in 2011 are about 120% greater with event temperatures that are about 3 degrees higher, on average.

Figure E-2: 2010 and 2011 SMB Ex Post Impact Estimates Versus Daily Maximum Temperature



The investigation of this issue focuses on ex post impact estimates rather than ex ante estimates. Ex post estimates are a more direct measure of program performance because they refer to real events. Ex ante impact estimates are necessarily a function of the observed ex post estimates. Therefore, it makes sense to focus on the ex post estimates as the source of the difference.

Table E-1 summarizes FSC's conclusions about the sources of improvement in impacts for each customer segment. The far right column shows the total amount of improvement that must be explained. For residential customers, there is a 70% improvement to be explained, which means that the 2011 impacts are approximately 1.7 times as large as those in 2010 during the same hours (4-6 PM). For SMB customers, the multiplier is 2.2. This is because there is a 120% improvement in load impacts for SMB customers from 2010 to 2011 during the same hours (3-5 PM).

The next column to the left is the weight of each device type; this is the fraction of each device type in the population. For example, residential devices are 84% switches.

To the left of that is the temperature multiplier. This is a rough estimate of the degree to which impacts changed due to temperature differences between the two years. For residential customers, this value is 1 because average temperatures during events were almost identical between the two years. For SMB customers it is 1.1 because temperatures were about 3 degrees warmer, on average, which yields about a 10% increase in impacts.[25]

[25] This is a very rough estimate based on Figure E-2.

The next column is the impact multiplier. This is an estimated amount by which impacts increased due to control strategy changes, for devices that received signals. That is, this multiplier does not include the effect of improved signal reception. This value is 1.7 for residential switches, which was established in side-by-side testing. For residential PCTs, it is calculated to be 1.3. This calculation is done taking all the other residential values as given and solving for the value that produces an overall improvement of 70% or a multiplier of 1.7. Because there is a certain amount of uncertainty in all the values underlying this result, the result itself should be taken to include a fairly large amount of uncertainty.

The impact multiplier for SMB switches is assumed to be the same as it was for residential switches. This is probably roughly correct. SMB switches also shifted from True Cycle to TrueCycle2 in 2011, although the level of cycling differed. The impact multiplier for SMB PCTs is solved for in the same way that it was for residential PCTs. Again, this value should be viewed as including significant uncertainty.

The communication multiplier is the increase in impact values due to improved signal reception alone. These values are derived from the analyses of communication in the 2010 and 2011 evaluations.

The final conclusion for the overall increase in impacts for segment and device type can read off the table as, for example, SMB PCTs had a 52% increase in communication success, a 30% increase in load impacts for devices receiving signals and a 10% increase in impacts due to slightly higher temperatures.

Table E-1: Sources of Load Impact Increase from 2010 to 2011

Segment	Device Type	Communication multiplier (A)	Impact multiplier (B)	Temperature Multiplier (C)	Weight (D)	Total Multiplier Switch(AxBxCxD)+ PCT(AxBxCxD)
Residential	Switches	1.00	1.70	1.00	0.84	1.7
	PCTs	1.37	1.30[26]	1.00	0.16	
SMB	Switches	1.22	1.70[27]	1.10	0.10	2.2
	PCTs	1.52	1.30[28]	1.10	0.90	

There are several possible sources of the change between the years in addition to those in Table E-1. It does not appear that these other sources account for much, if any, of the change in observed impacts. These other possibilities deserve discussion though.

First, there are several differences in how the estimates for each year were derived. The 2011 residential values were estimated using large random samples of whole-building SmartMeter data

26 Value is solved for using other values in the table.

27 Value is assumed to be the same for SMB PCTs using 33% cycling as for residential PCTs using 50% cycling.

28 Value is solved for using other values in the table.

while the 2010 residential values were estimated primarily using a much smaller sample of CAC logger data.[29]

The 2011 and 2010 SMB values were each estimated based on a sample of CAC load data.

Different modeling techniques were also employed. The 2010 values were estimated using customer-level regression models of CAC logger data. This produced both ex post and ex ante impact estimates. The 2011 ex post estimates were obtained through averaging and subtracting of loads of treatment and control groups. The 2011 ex ante estimates were calculated from a regression model of the ex post estimates as a function of weather.

Finally, it is possible that SmartAC customer CAC usage patterns changed between the two years. If, for example, customers in the 2011 population were much higher users of CAC than customers in the 2010 population, then this could explain the differences between the years. This could happen because the same set of customers had different behavior in the two years, because new customers joined the program and existing customers left, or a combination of the two. Our investigation strongly suggests that this is not a major source of change. We do not document the evidence for this conclusion here, but will provide it by request.

The issues of changes in control strategy, differing data collection and differing modeling assumptions are each addressed below. Prior to those discussions is a brief discussion on the differences between the residential load impact results observed in 2008 and in 2010,[30] specifically addressing why program performance seemed so much worse in 2010 than 2008.

E.1.1. Differences Between 2008 and 2010

To a casual observer, the larger residential impacts observed in 2011 may appear to be a return to form for the SmartAC program. The primary purpose of this section is not to fully re-analyze the differences between previous years of program performance. However, to understand the situation in 2011, it is important to recognize that this impression is not accurate. It is true that ex ante load impact projections for 2010 were substantially lower in 2008. However, actual observed program performance during the two years was similar, with the exception of program performance during the small number of very hot events in each year. This is shown in Figure E-3. The figure shows ex post load impact estimates for 2008 and 2010 as a function of weighted average daily maximum temperatures for the SmartAC population for each year. Again, to account for differing event timing during the day, average impacts are only calculated for the hours 4 PM to 6 PM.

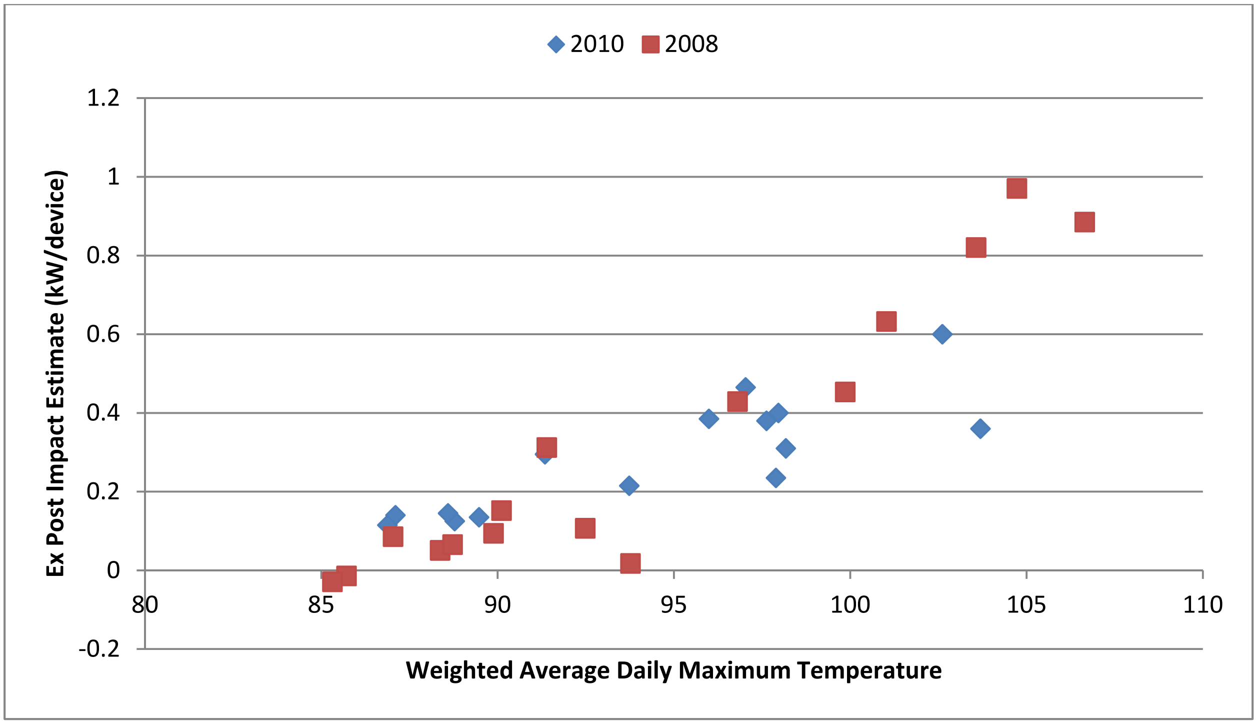
There are two important aspects of the Figure E-3. First, taken as a whole, program performance in the two years is quite comparable. The average high temperature for the test events in the two years is almost identical, as is the average load impact. Additionally, the overall variance in impacts at similar temperature levels is similar between the two years. If the two sets of points were not shaped and colored differently, it would be difficult to correctly categorize each point as belonging to 2008 or 2010 results.

[29] Some analysis of SmartMeter data informed the 2010 analysis as well. See the 2010 SmartAC evaluation.

[30] No AC logger-based ex post estimates were developed for 2009. Ex ante results in 2009 were calculated based on data collected in 2008 and so did not differ substantially from the 2008 estimates.

Second, at high temperatures, 2010 had smaller impacts on average than 2008. This tendency shows up fairly strongly for all events with high temperatures above 95°F. This is important because the event impacts at high temperatures are what primarily determine ex ante impact predictions. This is because ex ante weather conditions have quite high temperatures. It is important to recognize that this fact is based on a small sample of data. Between the two years, only 15 events occurred above 95°F and only 5 occurred with temperatures substantially above 100°F.

Figure E-3: 2008 and 2010 Ex Post Impact Estimates Versus Daily Maximum Temperature



There is no strong evidence as to what caused lower impacts during the hot event days in 2010. Several plausible explanations remain. First, the differences in performance could have been due to random variation. In this case, the variation is unexplainable based on observable factors. Even in 2008 alone, performance over events varied significantly at similar temperatures.

Second, customers could have used less CAC load at hot times in 2010 than in 2008, for at least two reasons. First, 2010 was, on average, a cooler summer and the heat that did occur was very late in the season. This could have led customers to use less CAC through habit formation during 2010. Second, the state of the economy in the summer of 2010 was significantly worse than in summer 2008, which could affect CAC loads in several ways. During the non-event times of the summer, CAC loads were about 16% less in 2010 than in 2008. Unfortunately, both summers only provide a small sample of hot days that does not allow for distinguishing between these hypotheses.

The primary conclusion of this subsection remains that overall program performance was similar in 2008 and 2010, but that load impact predictions for times of very high temperature differed due to unexplained differences observed among a small number of hot days in each year.

E.1.2. Differences in Program Operation Between 2010 and 2011

Two major changes in SmartAC operating strategy took place between 2010 and 2011. First, all switches were shifted from the True Cycle algorithm to the TrueCycle2 algorithm. Second, all PCTs were shifted from a temperature setback strategy to cycling. ExpressStat PCTs used simple cycling, while UtilityPro PCTs used TrueCycle2. These changes affected both residential and SMB customers.

The change for switches took place because Cooper, the SmartAC implementer, believed that TrueCycle2 would perform better because it did not require learning days and was less likely to shift to simple cycling (the default cycling strategy when the algorithm encountered problems in implementing its algorithm). The change for PCTs took place because it was expected that this would help improve the rate at which PCTs received event signals. This would occur because temperature setback commands are sent as a one-time signal for each event while cycling commands are sent every half hour. Commands sent multiple times have better odds of getting through in situations of poor reception.

Table E-2 (which is a repetition of Table 2-2) summarizes the changes in control strategy, the issues those changes were meant to address and the results of the changes.

Table E-2: SmartAC 2010 and 2011 Control Strategies

Segment	Device	2010 Control Strategy	Problems in 2010	2011 Control Strategy	Result
Residential[31]	Switches	50% True Cycle	Defaulted to simple cycling	50% TrueCycle2	No default to simple cycling
	Utility Pros	2-1-1 temperature setback	Poor communication	50% TrueCycle2	37% improvement in communication
Commercial	Switches	33% True Cycle	Poor communication[32] and defaulted to simple cycling	33% TrueCycle2	22% improvement in communication; no default to simple cycling
	Utility Pros	1-1-1 temperature setback	Poor communication	33% TrueCycle2	52% improvement in communication

The current evidence regarding a temperature setback strategy versus simple cycling or TrueCycle2 is not highly developed for residential or SMB customers. In this case, though, that evidence is of secondary importance. Of greater importance is that cycling strategies significantly improved performance among PCTs through improved signal reception. Based on improved signal strength alone, there would be an expected 37% increase in performance among residential PCTs and a 52% improvement among SMB PCTs.

31 The table does not include a fairly small number of residential Express Stats, which are an older type of PCT that are no longer installed.

32 Poor communication to SMB switches in 2010 was thought to be partially due to a malfunction in control devices installed on dual-stage CAC units. This problem was also addressed before the 2011 season.

The 2011 evaluation produced direct evidence about the performance of TrueCycle2 as compared to simple cycling for switches. This evidence is relevant because the 2010 SmartAC evaluation documented the fact that almost every switch during every event defaulted to simple cycling during 2010. Therefore, True Cycle in 2010 behaved the same as simple cycling. This means that the side-by side testing of TrueCycle2 versus simple cycling in 2011 provides relevant evidence about whether the changes in impacts observed are due to the changes in program operation. As documented elsewhere in this evaluation, during the side-by-side test of 50% TrueCycle2 against 50% simple cycling, TrueCycle2 provided about 70% greater load impacts than simple cycling. This occurred on a day with a maximum average temperature of 99°F, which is a bit higher than the overall average high temperatures observed during the events in 2010 and 2011, but is still within the normal range of SmartAC test event conditions. It is plausible that there would be a similarly large difference in performance between 33% TrueCycle2 and 33% simple cycling, although this was not tested.

That this outperformance by TrueCycle2 so closely matches the differences in residential load impacts observed between the two years is the strongest evidence that operational changes account for most or all of the observed differences in load impacts. As shown in Table E-1, this load impact improvement combined with significant improvements in other aspects of operations leaves only a fairly mild residual increase in performance to be explained. To do so requires only fairly mild assumptions about impact increases due to the change from True Cycle to TrueCycle2 for SMB switches and the change from temperature setback to TrueCycle2 for all PCTs.

The remainder of this section presents evidence that the increase in load impacts is real and is not due to differences in load measurement techniques or differences in modeling assumptions.

E.1.3. Differences in Load Measurement

Assessing the possibility that the difference is due to different data measurement methods (e.g., loggers versus whole building) is inherently difficult as long as the other possibilities remain. However, in other evaluations using the same CAC logger technology and measurement protocol, FSC has observed that load impacts measured using CAC loggers and using whole-building interval data are quite comparable. Two recent examples can be found in FSC's 2010 evaluation of the Ontario Power Authority peaksaver program and FSC's 2009 evaluation of San Diego Gas and Electric's Summer Saver program. There is no reason to expect that the situation would be different here.

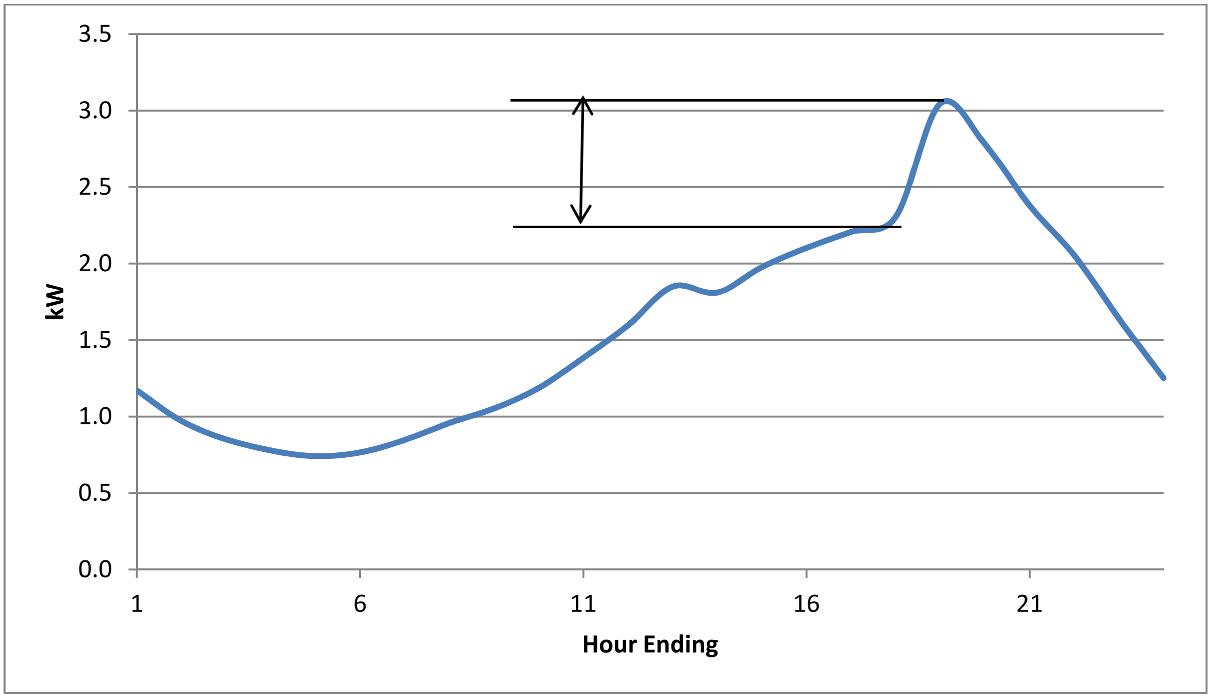
For the SMB population, the load measurement method was the same in 2010 and 2011.

E.1.4. Differences in Modeling Assumptions

To address the possibility that the differences in impact estimates arose due to differences in the models used in 2010 and 2011, it is useful to have a measurement of load impact or a proxy for such a measurement that is independent of modeling assumptions. If the 2010 or 2011 modeling assumptions themselves are in question, then it is better if there is evidence regarding this issue that does not rely on the modeling assumptions for either year. This is the same as saying that it would be useful to have some indicator of event impact that does not require estimating a reference load.

The proxy for event impact used here is the difference in load between the last hour of each event and the first post-event hour for all customers exposed to each event (referred to in the remainder of this section as "proxy impact"). This value, which is explained graphically in Figure E-4, requires no model of a reference load and is measured only based on average loads within the group of customers that were called for an event. Additionally, its value should be the same whether whole-house load or CAC load is being measured.

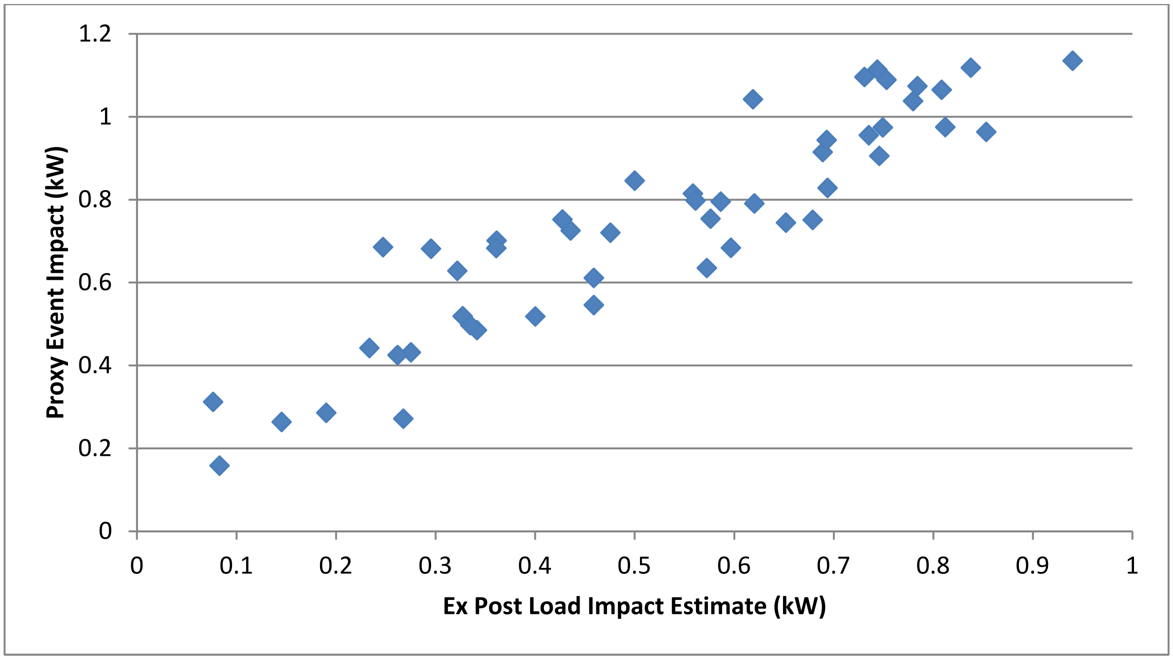
Figure E-4: Proxy Event Impact Illustration



This use of proxy impact as a direct proxy for load impact is a simplification, but a fairly accurate and very useful one. The proxy impact, as defined, is a combination of true load impact during the final event hour and the snapback effect. Snapback itself is generally correlated with load impact as well, but the relationship is complicated (as demonstrated elsewhere in this evaluation) by the fact that CAC units often hit their maximum duty cycle during the snapback period, which means that the snapback effect is capped.

Figure E-5 presents evidence that proxy event impact is a useful proxy for true event impact. It shows a scatter plot of proxy impact and ex post impact estimates for all local capacity areas over all residential test events in 2011. The year 2011 is selected because the evaluation design in 2011 produced highly precise ex post load impact estimates with almost no modeling assumptions. The relationship between the two measures is quite close – the correlation is above 90%. Also, the slope of a trend line between the two is very close to 1 – meaning that proxy impact moves in a one-to-one fashion with real impact.

Figure E-5: Proxy Impacts Versus Ex Post Impact Estimates for Each LCA for Each Event in 2011

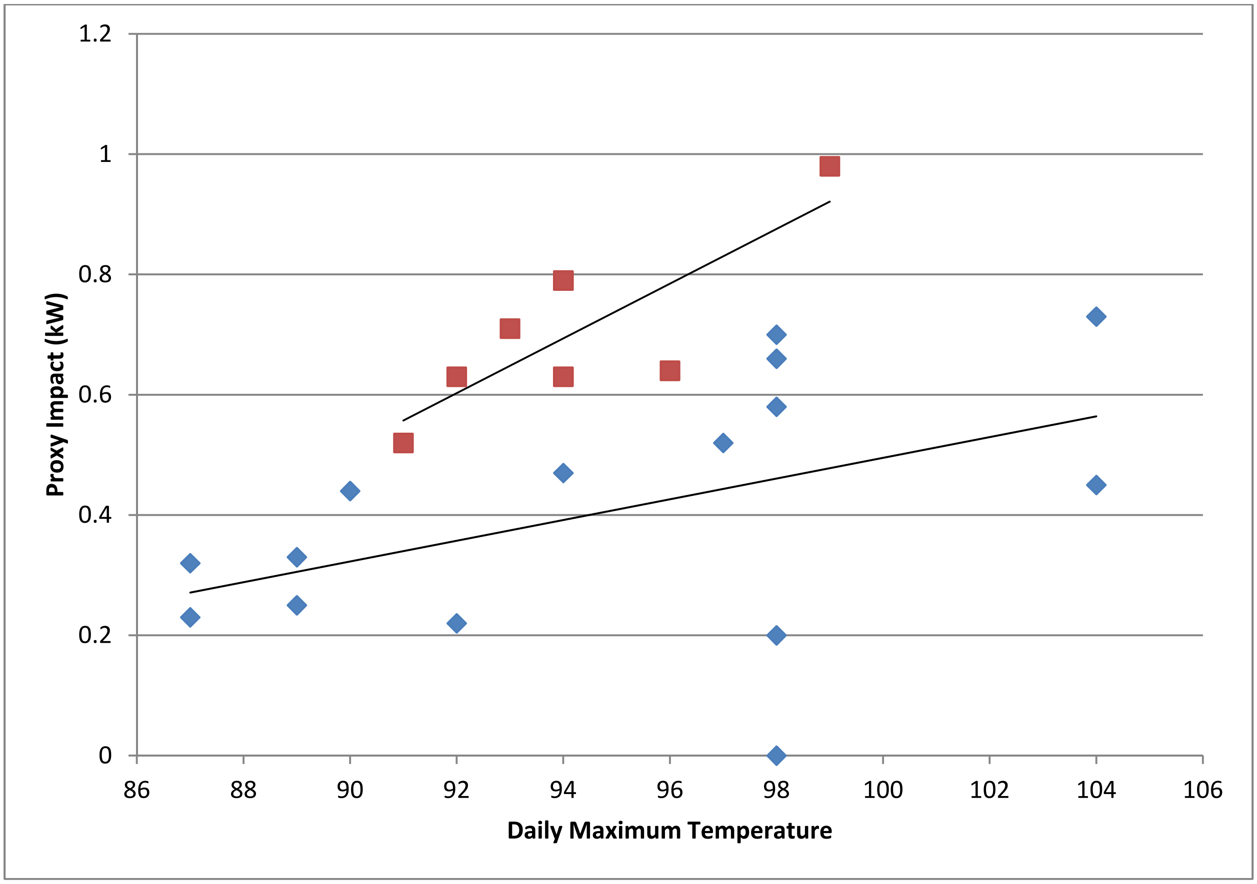


The evidence that modeling assumptions are not the primary source of the discrepancy between the 2010 and 2011 residential impacts is shown in Figure E-6. The figure shows a scatter plot of proxy impacts as a function of weighted average daily high temperature for the SmartAC population. It shows this for the 2010 and 2011 event days, calculated for the full SmartAC population. Figure E-6 also shows trend lines for 2010 and 2011. The 2011 values are red squares and the 2010 values are blue diamonds.

The main notable aspect of Figure E-6 is that the average proxy impact in 2011 is about 70% larger than the average for 2010, while the average maximum temperature is almost identical. This relationship, which is based only on averages of collected data with no other modeling assumptions, is virtually identical to the relationship shown in Figure E-1, which is based on the modeling assumptions of each year's evaluation. If, for example, the 2010 ex post impact estimates were systematically too low due to incorrect modeling assumptions, then Figure E-6 would look different from Figure E-1.

Figure E-6 would show that proxy impact values for 2010, which require no modeling assumptions, were of comparable magnitude to proxy impact values for 2011. Instead, the graph shows that even the raw load data provides strong evidence that 2011 load impacts were about 70% higher than 2010 load impacts.

Figure E-6: Proxy Impact Versus Weighted Average Daily Maximum Temperature for the Residential SmartAC Population in 2010 and 2011



The proxy event impact is not an accurate indicator of SMB event impacts because in 2010 almost every event ended at 6 PM when many businesses are closing down for the day. This makes many of the proxy event impacts negative. This analysis presumably could work well for SMB events that ended in the middle of the afternoon.

E.2. Comparison of Methods Between the 2011 Evaluation and Previous Evaluations

The statistical methods used to develop ex post and ex ante impact estimates in this evaluation differ from those used in previous evaluations. The primary reason for this is that more data has become available through the installation of SmartMeters. The main conclusions of this section are that:

- The methods used in this evaluation are an improvement over previous methods. This is primarily because the use of large, random control groups eliminates virtually all uncertainty about ex post impacts, which then provide a stronger basis for ex ante estimation; and
- There was no better alternative to the methods used previously (and still used in many other evaluations) due to data limitations.

This section begins with a high level comparison of methods. A more detailed discussion follows, which includes responses to objections that have arisen to the control-group method.

E.2.1. Overview of 2011 Methods Versus Previous Methods

Both the ex ante and ex post methods differ substantially between the 2011 evaluation and previous evaluations. In previous evaluations, for both residential and SMB customers, the primary method used to develop load impact estimates was to use individual customer regressions performed using CAC logger data from samples containing roughly 300 customers each. This strategy relied on loads observed on non-event days combined with an assumed (and validated) functional form to inform the model about what loads would look like during events. The same regression model could be used to estimate ex post and ex ante impacts because it could be used to estimate loads with and without DR.

In this evaluation, ex post estimates are developed by comparing loads between customers who had their loads controlled and customers who did not. The key is that these sets of customers are randomly-chosen so that the groups are statistically very similar. This strategy was enabled by the availability of SmartMeter interval data for almost the entire SmartAC residential population. Although CAC loggers were still used for measuring SMB loads in this evaluation, the elimination of loggers from the residential segment freed up the evaluation budget so that a larger sample of loggers could be used on the SMB segment. This allowed a similar treatment-control method to be used for SMB customers, because the sample was sufficiently large.

The ex post estimates for 2011 are more reliable on an event-by-event basis than those in previous evaluations. The main reason for this is that no two days are ever alike and CAC loads vary substantially even between days that appear similar in their basic weather and seasonal characteristics. This makes the regression modeling approach inherently more uncertain than the control group method because it relies for reference load on non-event days with different characteristics, plus a regression functional form that has to be invented to measure ex post impacts. Given the amount of data previously available, there was no way around these problems.

By contrast, with the control group method, there is no concern about differing CAC loads on different days. Nor is there any concern about regression functional forms. The control group reveals an unbiased reference load which can then be used to estimate impacts through subtraction.

The limitations of the regression method in ex post carry through directly to ex ante estimation because any inaccuracies in the ex post impacts due to random differences in CAC loads between similar days or due to functional form problems will automatically affect ex ante estimates. On average, this appears to be a fairly small problem. Impact estimates are mainly stable across years for most programs evaluated using this method. This would not be true if this method was not mainly reliable. Even in cases of large changes, such as between SmartAC in 2008 and SmartAC in 2010, the culprit is usually lack of data (in this case, few hot days).

In contrast, with a near-perfect set of ex post estimates, ex ante modeling can be done more transparently and with the primary variable of interest (load impacts) as the dependent variable. This is an improvement because there is no uncertainty about the input to the ex ante model, in contrast to the prior method. This allows for stronger conclusions about which variations in impacts are related to observable variables and which are random. As an additional benefit, it would be quite simple for an outsider to take the ex post estimates reported and to attempt to find a better ex ante model.

E.2.2. More Detailed Description of Methods and Discussion of Objections

This section contains a more detailed description of the two contrasting methods. It also directly addresses some of the objections to the newer methods that have arisen. The description of methods below uses the SmartAC program as the primary example, but the main points of discussion apply to the evaluations of virtually every event-based DR program.

Each set of methods begins with collecting hourly or sub-hourly load data for a sample of customers in SmartAC for the summer of interest. The method used in prior SmartAC evaluations (hereafter the prior method), as well as in many of the event-based load impact evaluations performed by FSC and others in the past several years then consists of the following broadly defined steps:

- Fit regression models to this load data, usually for each customer individually. The models typically express load as a function of interactions of temperature variables and time variables. In most cases, all time and temperature effects are allowed to vary at the hourly level. Frequently, all effects will also be interacted with month and day of the week effects as well.
- Event effects can enter the model in several possible ways; usually events are modeled as a function of temperature-based variables, time of day and/or the elapsed time of the event at the time in question. Event effects and weather effects typically are modeled to vary independently at the hourly level, which means that the model might predict a relationship between load and temperature at 1 PM that varies dramatically from that at 2 PM. Typically, several different functions of recent weather are included in the model. For example, a specification might include CDH,[33] CDH-squared and CDH interacted with the morning low temperature; in this case, all of these weather effects would be interacted with each hour of the day and possibly month of the year and day of the week.
- In this step, many possible specifications of load as a function of time and temperature variables are tested in order to find a model that predicts load accurately during the observed time periods and that predicts plausible event impacts. Fit is determined by in-sample R-squared calculations, or, preferably, through an out-of-sample testing protocol.
- Use the selected model to estimate ex post load impacts for all sampled customers for all events during the time period of interest.
- Use the same regression model to predict ex ante load impacts for all sampled customers by applying the model to the chosen ex ante weather conditions.

In contrast, the method used in this evaluation (hereafter the control group method) consists of the following steps:

- Develop ex post impacts at an hourly level for each event by directly comparing observed load to reference loads. Reference loads are developed here through the use of a control group, but the same basic method could be applied in the absence of a control group through weather-based day-matching;
- Develop a regression model of ex post impacts as a function of weather. Ex post impacts for this step are average impacts over a given event for the whole sample of customers (possibly divided into LCAs). This aggregation over time and customers means that there are exactly as many data points in these regressions as there are events. The small number of data points limits these regressions to one explanatory variable plus a constant; and
- Use the regression model to predict ex ante impacts by applying the regression model to chosen ex ante weather conditions. This produces an average event impact for each set of conditions for the whole sample of customers. Event impacts are then modeled at an hourly level by assuming that each event has the same basic shape in its effect. For example, if on average the effect from 1 PM to 2 PM is 75% of the average event effect over the entire time

[33] CDH stands for cooling degree hours, a standard measure of how hot a particular day was, used for modeling climate control loads.

period, then the predicted ex ante event impact for 1 PM to 2 PM is 75% of the regression-predicted average event impact over the entire event under the given weather conditions.

The remainder of this section addresses possible objections to the control group method in comparison to the old method.

Objection: the control group method is rudimentary in the treatment of relationships between event impacts and weather that may differ across hours of the day

This is a valid objection, but the prior method is not practically better on this point. To illustrate the issue, suppose it was the case that on days above 105°F event impacts at 1 PM tended to be 120% of impacts at 5 PM, while on days between 90°F and 95°F event impacts at 1 PM tended to be 60% of impacts at 5 PM. In that case, with sufficient data, the prior method would provide an estimate of that relationship, while the control group method would wrongly attribute the same basic event impact shape to both sets of weather conditions. However, the actual difference in relative event impacts across hours appears to be much smaller than the example, and sufficient data is not yet available any may or may not become available.

In the control group method, ex post impact estimates are free to vary independently across hours, but not ex ante estimates. 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. With only seven residential or nine SMB ex post impacts to model, the relationships between temperature and impact are identified fairly weakly, with individual observations quite influential. 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.

Objection: the control group method does not account for more subtle effects because it is limited to one temperature-based independent variable in its predictive model of load impacts

The current control group method includes only one temperature-based independent variable in its predictive model of load impacts. In contrast, because the prior method models the load data at the customer level, including a full season of hourly load data for each customer, it can contain literally hundreds of temperature and time-based independent variables. For example, in the bulleted case above where the stipulated model includes CDH, CDH-squared and CDH interacted with the morning low temperature, if each is interacted with 24 hourly dummy variables and three monthly dummy variables, then the model has 3 x 24 x 3 = 216 temperature effects in it.

The basic problem is that the subtler impacts are not well-identified in the prior method. The method does produce accurate estimates on average because effects that are not well-identified tend only to add small amounts of random noise to predictions. The reason that the subtler impacts are not well identified is limited data.

The data is limited in that the number of time periods providing relevant non-event reference load for a given summer for any program is necessarily small. In the best cases there may be 10-20 non-event days that can be considered comparable to load control days. In the case of recent SmartAC evaluations, the number of relevant non-event days has been much smaller.

To illustrate this point it helps to consider the amount of data that would be necessary to identify the complex set of weather interactions typically included in models based on the prior method. Consider the simple example of identifying the effect of hot mornings on load later in the day. First, only days where the ultimate temperature hits above roughly 90°F are of any interest. That automatically excludes all but perhaps 20 non-event days per customer during a summer under the best modeling conditions. It would not be unusual to have only three or four such days. Next, the load will clearly vary with respect to high temperature itself in a way that is not pre-defined for the modeler. This means that the effect of hot mornings must be identified simultaneously with the effect of daily high temperature. The implication of this is that to identify the effect of hot mornings, the modeler must observe a wide variety of morning temperatures for each relevant daily high temperature. That is, there must be a variety of cooler and warmer mornings on days when the temperature reaches 90°F and also a variety of cooler and warmer mornings on days when the temperature reaches 95°F. This level of temperature variation rarely happens in one summer. Even over many years it is unlikely that the modeler will observe a broad range of morning temperatures for each given daily high temperature. More generally, only a small subset of the possible variations in conditions that affect load will be observed during any given season. This forces the modeler using the prior method to measure event impacts using a model that is customized for the patterns that happened to be observed during one season.

An objection to the above points might be that there is more variation than is being acknowledged. The dataset consists of thousands of hours of load data for hundreds of customers over a large geographical territory. This provides more variation than what is observed at the level of the aggregate sample. While this is true, the load data at the customer or device level are so noisy as to be virtually meaningless except when averaged over a minimum of several dozen customers. This drastically reduces the effective sample size. Additionally, the many hours of data are reduced in practical usefulness by the fact that load data is highly correlated over time at the hourly level. The amount of useful variation in the data does not extend much beyond what is observed on average over all sample customers at the daily level. This means that the effective situation is much closer to one where there are at most 10-20 relevant data points for modeling in one summer.

Objection: the control group method does not allow for the estimation of individual customer impacts

The prior method includes an estimated regression model for each customer. This allows for an examination of the distribution of impacts over customers. FSC's investigation into this issue through the estimation of "false-event" effects on customers who never receive events shows that the prior method's outputs at the individual customer level are much more random noise than actual event impact. When averaged over many customers, the results are accurate, but not at the customer level. That is why the distribution of customer impacts section in this report focuses on impacts aggregated over deciles of usage.

Appendix F. Event Discomfort Results

Following the test event on Tuesday, September 6, 2010, the Population Research Systems (PRS) arm of The FSC Group conducted a post-event survey on a random sample of SmartAC customers. High temperatures for the day were 98°F in Fresno, 98°F in Sacramento and 94°F in Concord.

The residential survey sample consisted of 1,000 customers who received the event (referred to as the treatment group or the treated) and 1,000 customers that did not. For the SMB survey, all customers in the M&E sample were contacted. The survey is given to both treatment and control customers in order to distinguish between event-induced discomfort and discomfort that customers would report anyway.

The survey of treatment and control groups started immediately after the event. In total, 877 surveys were completed. The survey was similar for residential and SMB customers. There were a total of 654 residential surveys completed and 223 SMB surveys completed. In the residential group, there were 323 surveys completed by customers who had the event and 331 surveys by control customers. In the SMB group, there were 111 surveys by customers who had the event and 112 by control customers.

The key discomfort questions were:

- Was there any time on Tuesday when the temperature in your home or place of business was uncomfortable; and
- If so, during which hours were you uncomfortable?

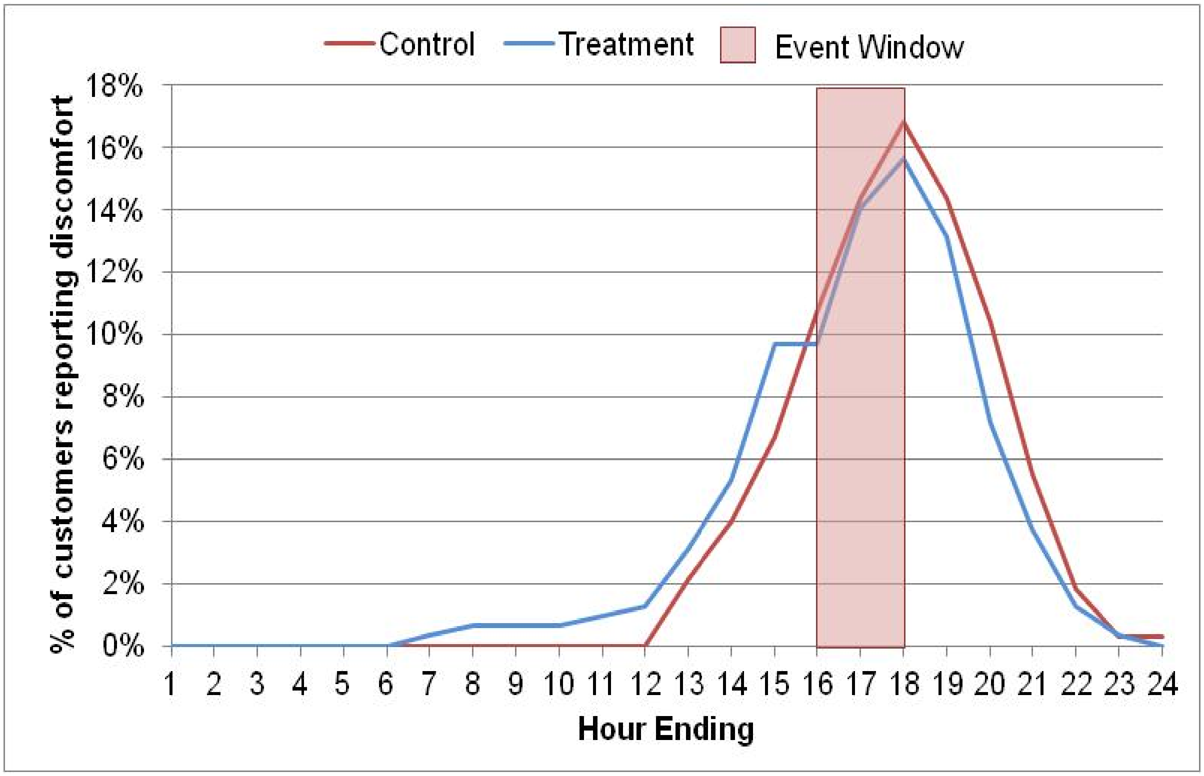
Key findings were that:

- Residential customers in the treatment group show a slight decrease in discomfort relative to the control group during event hours, but the difference is not statistically significant; and
- SMB customers in the treatment group show an increase in discomfort relative to the control group during event hours, but the difference is not statistically significant.

F.1. Residential Customers

For residential customers, an average of 12% of treatment customers reported being uncomfortable during event hours compared to 13% of the control customers. This difference is not statistically significant. Over the entire event day, an average of 3.65% of treated customers was uncomfortable in each hour. For the control group, the average was 3.64%. Figure F-1 shows the levels of discomfort for each hour of the day for both groups. The peak values in Figure F-1 are below the percentages of people who reported discomfort at any point because people could report discomfort at different times of day. Reported discomfort in the treatment and controls groups peaked at the very end of the event.

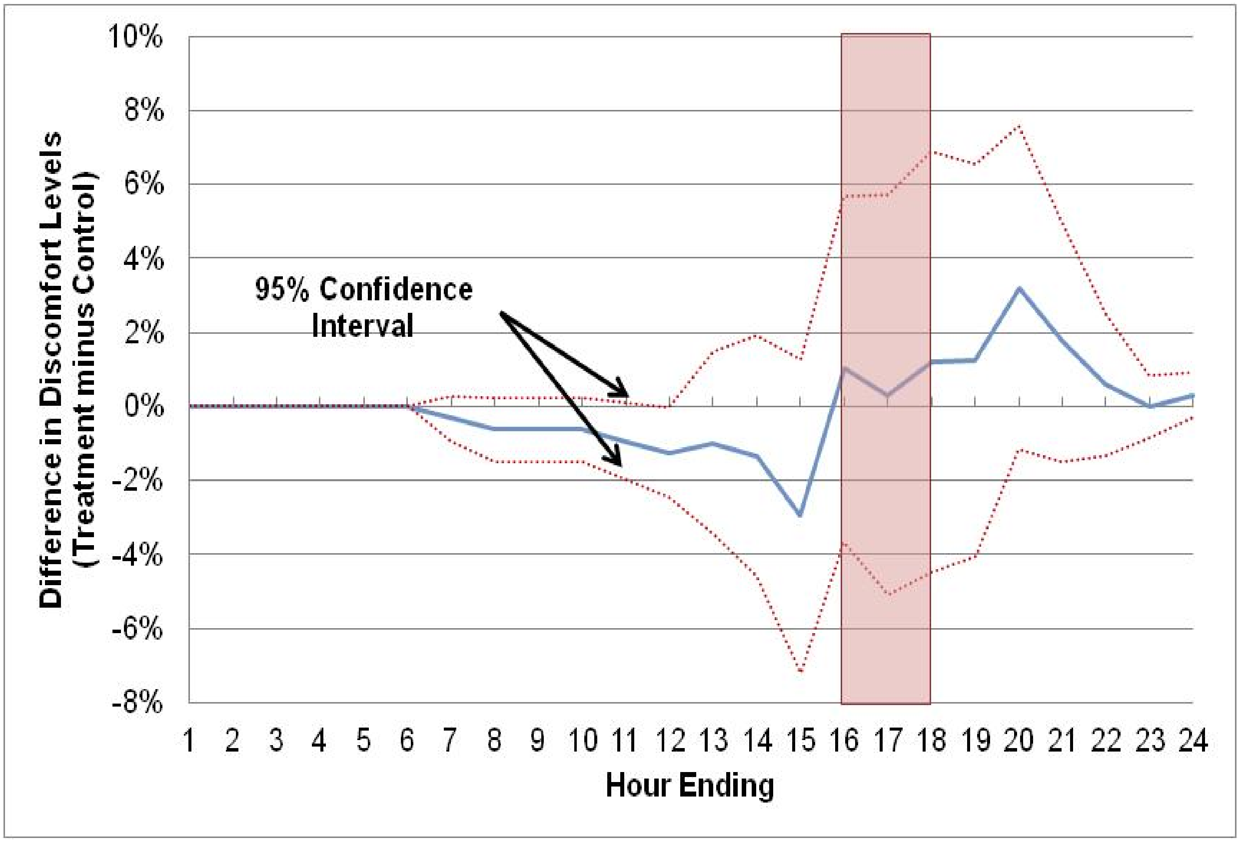
Figure F-1: Reported Discomfort Levels in the Residential Treatment and Control Groups



Treated and control customers show similar levels of discomfort throughout the day. The largest difference between the two groups occurred from 7 to 8 PM when 10% of control customers reported being uncomfortable compared to 7% of the treated population. To test whether the difference in discomfort between groups is statistically significant or simply random variation, a difference-in-means test for statistical significance was used. A difference-in-means test determines whether or not there is a statistically-significant difference in discomfort levels between the treatment and control groups. If the 95% confidence interval around the difference does not include zero, the difference is not likely due to random variation in response.

Figure F-2 shows the difference in means plotted for each hour of the day, along with the lower and upper bounds of the 95% confidence interval for that difference during each hour. As seen, the confidence interval always includes zero. This means that the difference in reported discomfort levels could well be due to random variation in responses rather than due to an actual difference in discomfort.

Figure F-2: Difference in Means Between Treatment and Control Groups for Residential Respondents with Confidence Interval Bounds



In addition to the questions about discomfort level, the survey contained several questions about the respondents' homes and the demographics of the respondents. The responses to these questions are summarized in Tables F-1 through F-6. The primary conclusion to draw from these tables is that the treatment group and control group have similar distributions for all these variables. This provides confidence that conclusions about discomfort and customer satisfaction that we draw from survey responses are not likely to be due to underlying differences between the control group and treatment group. Also note that survey respondents tend to be older compared to the population – the average age is above 60 for both the treatment and control groups. This response bias is typical of telephone surveys.

Table F-1: “Please tell me which of the following types of buildings best describes your home?”

Type of Home	Control Group	Treatment Group	Total
Single family detached house	86%	83%	85%
Townhouse	4%	6%	5%
Duplex	2%	2%	2%
Apartment	2%	3%	2%
Mobile home	3%	5%	4%
Don’t know/Not sure	3%	1%	2%

Table F-2: “Do you own or rent your home?”

Rent/Own	Control Group	Treatment Group	Total
Own	94%	92%	93%
Rent or Lease	3%	7%	5%
Other	1%	0%	0%
Don’t know/Not sure	2%	1%	2%

Table F-3: “What is your household’s total annual income before taxes?”

Income Level	Control Group	Treatment Group	Total
<$15,000	4%	6%	5%
$15k - 20k	2%	3%	3%
$20k - $30k	11%	14%	12%
$30k - $40k	8%	8%	8%
$40k - $50k	6%	5%	6%
$50k - $75k	10%	12%	11%
$75k - $100k	11%	7%	9%
$100k - $125k	7%	5%	6%
$125k - $175k	7%	5%	6%
> $175k	6%	7%	6%
Don’t know/Not sure	27%	27%	27%

Table F-4: "Which of the following is the highest level of education you have completed?"

Education Level	Control Group	Treatment Group	Total
8th grade or lower	2%	2%	2%
High school	17%	21%	19%
Associates degree, vocational degree	27%	28%	28%
Four year college degree	25%	24%	24%
Graduate or professional	24%	23%	23%
Don't know/Not sure	5%	2%	4%

Table F-5: "Including yourself and children, how many people live in your home at least six months of the year?"

Number of People	Control Group	Treatment Group	Total
1	23%	24%	23%
2	46%	44%	45%
3	10%	16%	13%
4	12%	8%	10%
5	5%	5%	5%
6 or more	4%	3%	4%
Average number of people*	2.4	2.3	2.4

*Assumes six people in households of six or more

Table F-6: "What is your age?"

Age Group	Control Group	Treatment Group	Total
< 25 years old	1%	1%	1%
25 - 34 years old	3%	4%	4%
35 - 44 years old	9%	9%	9%
45 - 54 years old	11%	12%	12%
55 - 64 years old	16%	18%	17%
65 - 74 years old	28%	25%	27%
> 75 years old	26%	29%	28%
Prefer not to answer	5%	3%	4%

Finally, the survey asked about satisfaction with SmartAC. After they were asked about discomfort, customers were asked, "Based on all of your experiences with the SmartAC program so far, how satisfied have you been with the program overall?" and given a range from 1 being "Very Dissatisfied" to 10 being "Very Satisfied." These answers are shown in Table F-7. Customers who were treated on September 6 had an average satisfaction rating of 8.0 and customers who were not treated that day had an average rating of 8.3. The difference is not statistically significant.

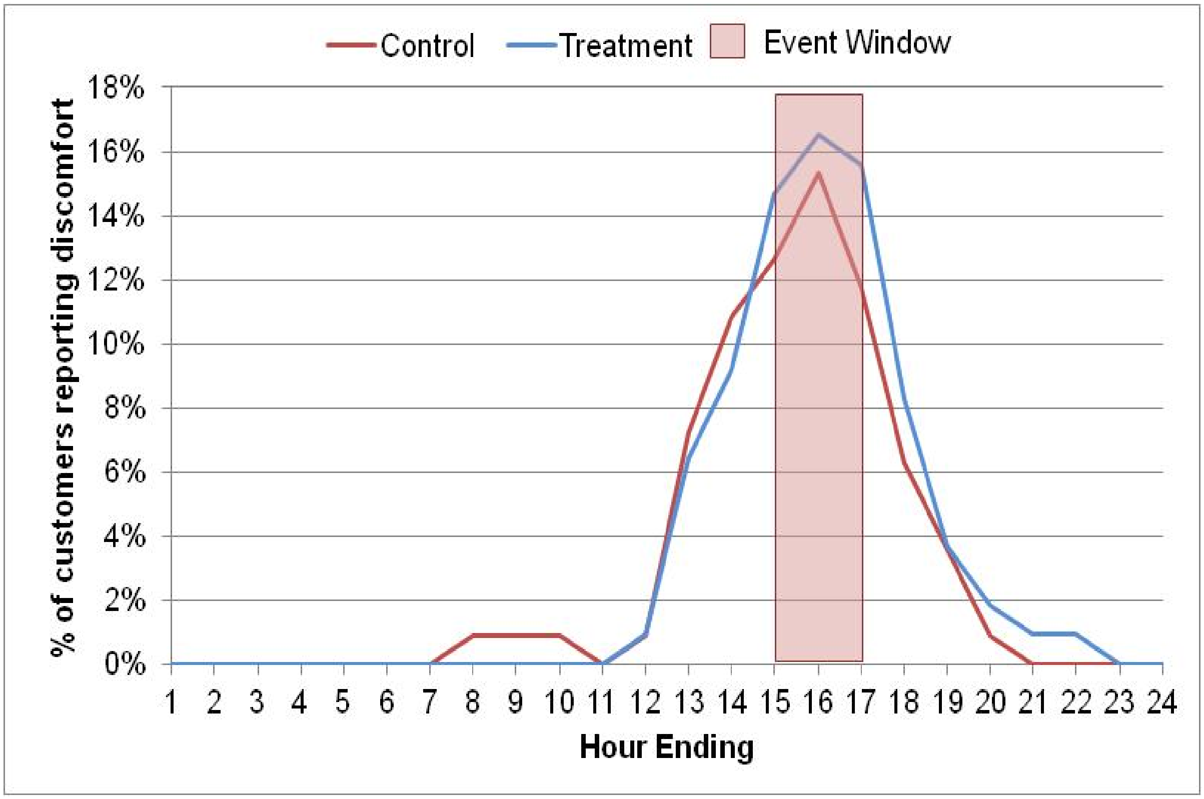
Table F-7: Satisfaction Levels with SmartAC

Group	Mean	Median	Std. Dev
Treatment	8.0	9.0	2.2
Control	8.3	9.0	2.2
Total	8.1	9.0	2.2

F.2. SMB Customers

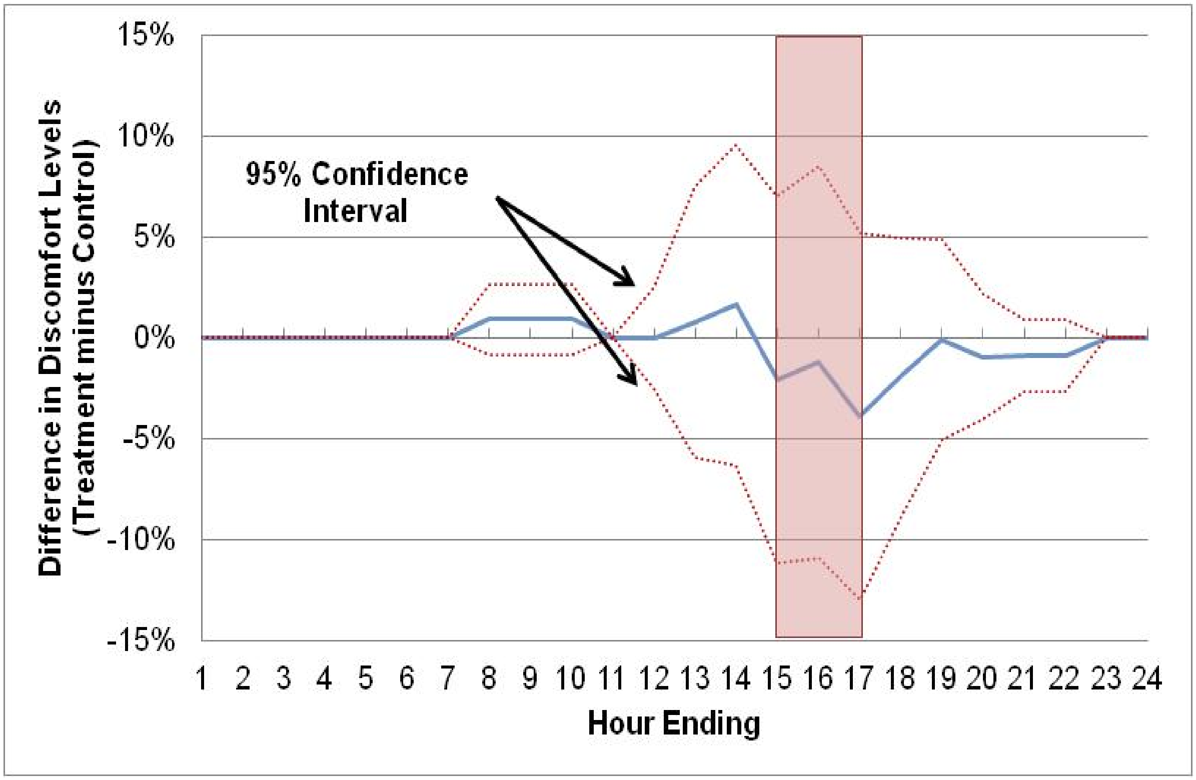
For SMB customers in the treatment group, an average of 16.5% of respondents were uncomfortable over the two-hour event; while an average of 14% of respondents in the control group were uncomfortable during the event. Figure F-3 shows the levels of reported discomfort in this sample for each hour of the day. Both groups reported discomfort peaked near the middle of the event period.

Figure F-3: Reported Discomfort Levels in the SMB Treatment and Control Groups



Just as in the residential case, the difference in reported discomfort levels could be due to random variation, a difference-in-means test was used on this group as well. Figure F-4 shows the difference between the discomfort levels in each group for each hour of the day, along with the 95% confidence interval boundaries. As seen with the residential survey, the confidence interval always includes zero, indicating that the difference in discomfort between the two groups is never statistically significant. This provides confidence that the treatment and control groups are similar in their general tendency to report discomfort.

Figure F-4: Difference in Means Between Treatment and Control Groups for SMB Respondents with Confidence Interval Bounds



F.3. Further Survey Results

Tables F-8 through F-12 show the distributions of various business characteristics across the treatment and control groups. Notice that in each case the distributions are similar suggesting that the difference in reported discomfort levels is not simply due to differences in customer characteristics between the two groups.

Table F-8: "What business sector does your firm belong to?"

Business Category	Control	Treatment	Total
Office/ Professional	36%	44%	40%
Retail/ Sales	24%	14%	19%
Restaurant	13%	17%	15%
Grocery Store	7%	0%	1%
Manufacturing/ Construction	4%	3%	4%
Warehouse	2%	1%	1%
Other	18%	20%	19%
Don't Know / Refused	1%	1%	1%
Total	100%	100%	100%

Table F-9: "Do you lease or own your facility?"

Ownership	Control	Treatment	Total
Lease	44%	48%	46%
Own	54%	50%	52%
Other	1%	0%	0%
Don't Know / Refused	1%	2%	1%
Total	100%	100%	100%

Table F-10: "What is the approximate annual revenue of your business?"

Locations	Control	Treatment	Total
< $250k	19%	22%	20%
$250k - $500k	13%	10%	12%
$500k - $1M	13%	12%	13%
$1M - $2M	6%	5%	6%
$2M - $5M	4%	4%	4%
$5M - $10M	2%	2%	2%
> $10M	1%	2%	1%
Don't Know / Refused	42%	44%	43%
Total	100%	100%	100%

Table F-11: "About how many employees do you have at this location?"

Employees	Control	Treatment	Total
1 to 5	49%	52%	51%
6 to 10	23%	17%	20%
11 to 20	19%	18%	18%
Over 20	6%	12%	9%
Don't know / Refused	3%	1%	2%
Total	100%	100%	100%

Table F-12: "What is your position or title?"

Square Footage	Control	Treatment	Total
Owner / Partner	39%	43%	41%
CEO / President	5%	5%	5%
General Manager/COO	9%	13%	11%
Office or Store Manager	27%	21%	24%
Accountant or Bookkeeper	2%	4%	3%
Administrative	7%	9%	8%
Other	9%	5%	7%
Don't know / Not sure	2%	1%	1%
Total	100%	100%	100%

After they were asked about discomfort, customers were asked, "Based on all of your experiences with the SmartAC program so far, how satisfied have you been with the program overall?" and given a range from 1 being "Very Dissatisfied" and 10 being "Very Satisfied." These answers are shown in Table F-13. Both customers who were treated and customers who were not treated on September 6 had an average satisfaction ration of 7.7. The difference is not statistically significant.

Table F-13: Satisfaction Levels with SmartAC Program

Group	Mean	Median	Std. Dev
Treatment	7.7	8.0	2.5
Control	7.7	8.0	2.3
Total	7.7	8.0	2.4

Appendix G. Temperature Changes During Events

The internal data logs used to assess control device communication success also contain indoor temperature data; this is only true for PCT internal device logs.

Tables G-1 and G-2 below show the average temperature increase during events for PCTs in the samples used to assess control device communication success. Only devices that were in cooling mode and received an event signal were included in the analysis. The number of devices with data varied for each event, but there were roughly 30 devices underlying each average in the residential table and roughly 130 underlying each average in the commercial table.

In both groups, average temperature increases are about 1 degree in the first hour and rise to about 3 degrees in the fifth hour. There is a fairly large amount of variation on a day-to-day basis though. Residential temperature increases are similar to commercial increases, despite commercial customers being controlled with a 30% cycling strategy and residential customers being controlled with a 50% cycling strategy.

Table G-1: Average Temperature Increases During Residential SmartAC/SmartRate Events

Event Date	2-3 PM	3-4 PM	4-5 PM	5-6 PM	6-7 PM
21-Jun-11	1.0	2.0	2.2	2.7	3.8
22-Jun-11	1.1	2.4	3.4	1.8	2.2
6-Jul-11	1.5	2.9	3.8	4.5	3.4
28-Jul-11	0.8	1.5	1.9	2.5	2.9
29-Jul-11	0.7	2.0	2.8	3.4	3.8
17-Aug-11	2.0	3.5	4.3	5.0	3.5
23-Aug-11	1.0	2.2	3.0	3.5	3.6
29-Aug-11	1.9	3.3	3.7	4.1	3.1
2-Sep-11	0.7	1.5	1.7	2.1	2.1
6-Sep-11	0.9	1.7	2.4	2.7	2.6
7-Sep-11	1.1	2.0	2.6	3.2	3.4
8-Sep-11	0.7	1.5	2.0	2.5	2.5
20-Sep-11	0.8	1.8	2.2	2.8	2.8
Total	1.0	2.0	2.6	2.9	3.0

Table G-2: Average Temperature Increases During Commercial SmartAC Events

Event Date	1-2 PM	2-3 PM	3-4 PM	4-5 PM	5-6 PM
21-Jun-11	1.3	2.6	3.3	3.9	4.1
22-Jun-11	0.9	1.8	2.3	2.4	3.2
23-Jun-11	0.7	1.3	1.7	2.0	2.3
24-Aug-11			0.9	2.0	
6-Sep-11			0.7	2.0	
7-Sep-11			1.2	2.1	
8-Sep-11			0.8	1.6	
20-Sep-11			1.4	2.3	
21-Sep-11			1.3	2.2	
Total	1.0	1.9	1.5	2.3	3.2