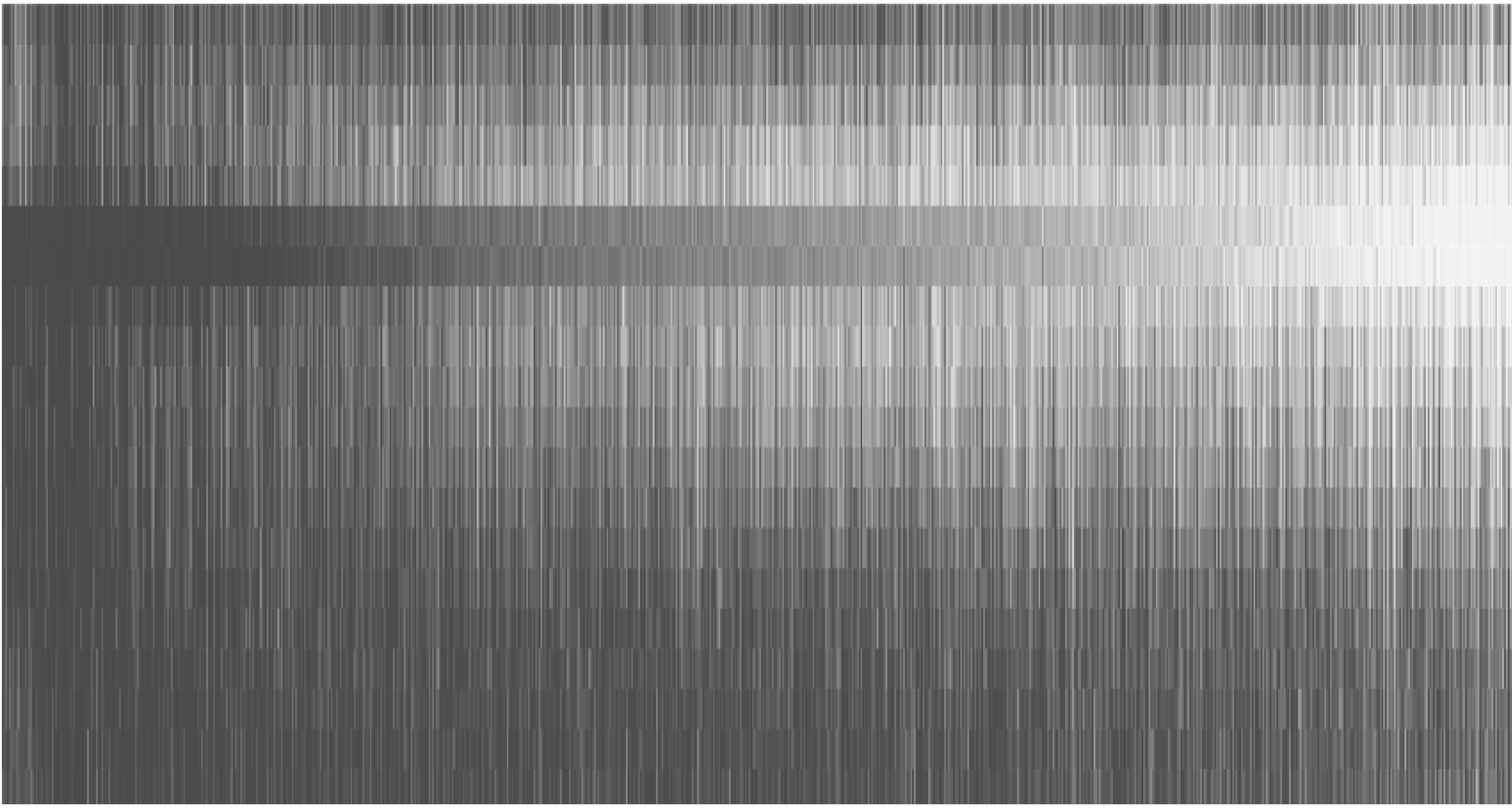
2017 Load Impact Evaluation for Pacific Gas & Electric Company's SmartAC™ Program



April 2, 2018

Prepared for Pacific Gas and Electric Company

Prepared by
Sam Borgeson
Phil Price
David Noonan
Convergence Data Analytics, LLC

Jennifer Mitchell-Jackson
Grounded Research and Consulting, LLC

CALMAC ID: PGE0415

Contents

Cover: Graphical depiction of the impact of the SmartAC™ program on event participant electric loads.

Executive summary

This report documents ex post and ex ante load impacts for Pacific Gas and Electric's (PG&E) SmartAC™ Program for the year 2017. The SmartAC™ program utilizes direct load control switches (switches) and programmable communicating thermostats (PCTs) to reduce electricity demand from central air conditioning units (AC) owned by residential customers during times of peak system usage. This report presents the findings from the impact evaluation of residential customers during the 2017 season, which ran from May through October.

Residential ex post load impacts

For the purposes of this evaluation and report, we define "events" as SmartAC™ hardware activation for a group of participants that share a **date, start and end time, and geography**. There were 16 event days with 36 discrete events called during the 2017 season.

Of these:

- 30 events were serial group events (called using the last digits of the serial numbers, i.e. 0-9, assigned to the control devices)
- 5 events were sub-LAP events (referring to geographic indicators set on each piece of hardware upon its installation)
- 1 event, on May 3rd, was called due to grid conditions with every enrolled device called at the same time.

Serial group events

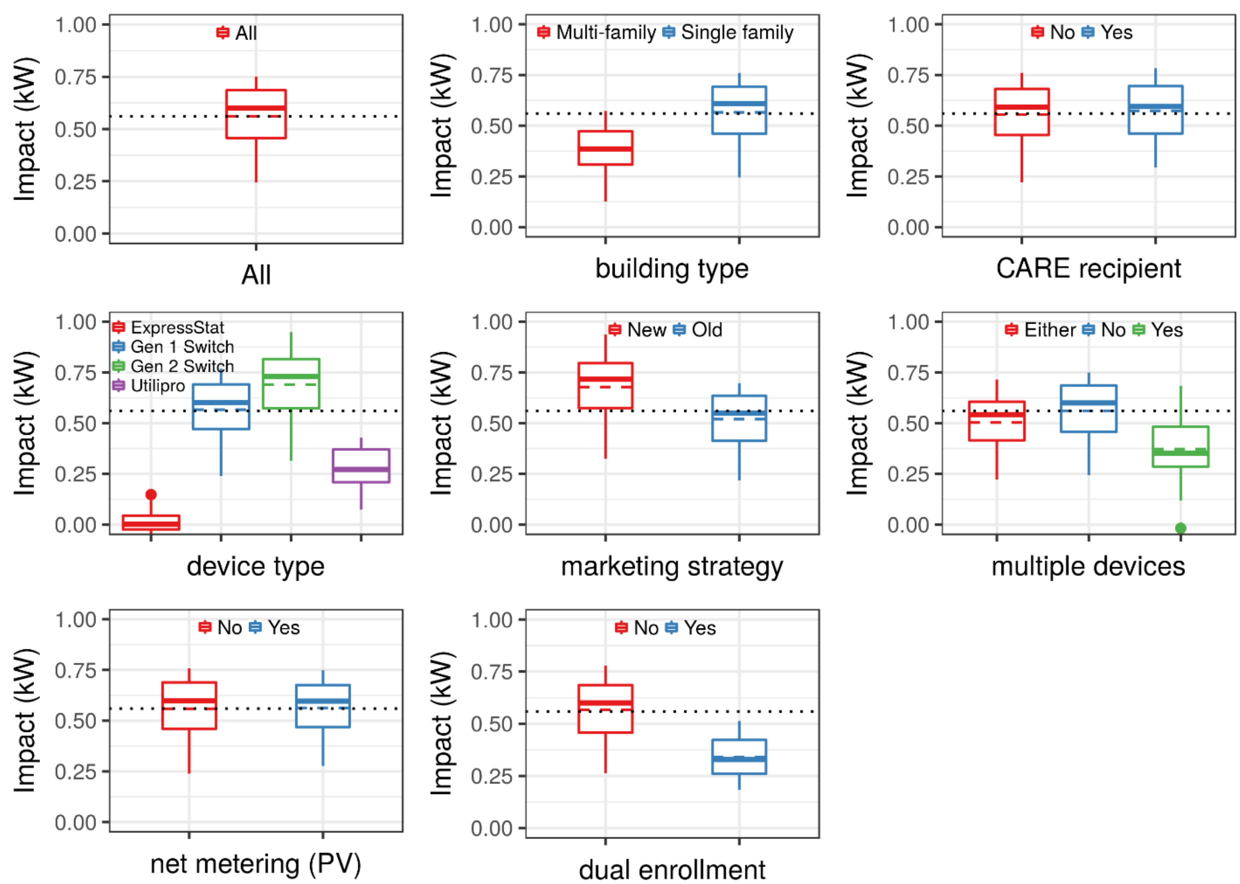
Serial group events are called on just a subset of devices, based on the last digit of their serial numbers (0-9) and events can include any number of these groups. This sets up the evaluation of such events through randomized control trial methods. This year, to improve modeled load reduction estimates for low sample count sub-groups (for example, the program has under 6,000 multi-family customers enrolled), we computed load reductions as the difference in differences between participants and controls on comparison days and event days.

The average serial group event load reduction per household at 5-6 pm (the most common peak hour) was 0.65 kW per household, or about 21% of the whole house load. However, there is wide variation across local capacity areas (LCAs) and days based on the time that the event was called and the average temperature during the event. On average, per household, 2017 ex post results are significantly higher than 2016 ex post results (0.65 kW compared to 0.52 kW). This is in part due to the program's planned removal of low performing participants and in part due to differences in the times and temperatures of the events (i.e., the 2017 season had a greater range of temperatures than 2016, and prior years).

In aggregate, ex post load reductions under the serial group events ranged from 1.7 MW to 12.3 MW depending on the number of groups called and the event conditions.

Figure 1, reproduced from the ex post results section, summarizes the performance of various customer sub-groups across all serial events. A variety of devices has been installed since program inception; older devices are 'Gen 1 Switches' and the two types of legacy (pre-2013) thermostats are ExpressStat and Utility Pro models. Utility Pro, ExpressStat, dually enrolled, multi-family, and multi-device customers perform poorly relative to the full population of participants, with no apparent load reduction at all coming from ExpressStat devices. On the other hand, newer 'Gen 2 Switches' and customers recruited under the 'New' marketing strategy (which is based on AC usage modeling), systematically outperform the full population of participants. Multi-family customers had load impacts 30% lower than the full population of single device owners, on average. Multi-device customers had impacts 35% lower than the full population (multi-device households are discussed further in the recommendations section). The load impacts associated with Gen 2 switches outperformed the full population by 24%. Based on the available evidence, dually enrolled customers underperformed by 27%. Finally, customers recruited with the "New" marketing approach outperformed all customers by 22% and "Old" customers by over 30%.

Figure 1: Event impacts by customer sub-group



Note: Box and whiskers summary of event impact by sub-category, across all hours of all serial group events. Boxes extend from the 25th to the 75th percentile, with the median, or 50th percentile marked with a horizontal line and average marked by a dashed line. The Black dotted line is the average serial group event load impact across all customers.

Other customer characteristics, like whether they have a solar photovoltaic (PV) system or receive low-income California Alternate Rates for Energy (CARE) rates appear not to influence outcomes very much. The CARE result is a departure from previous findings that CARE customers significantly outperform non-CARE. It is generally understood that CARE customers live in hotter climates and have larger cooling loads. The removal of low-AC-consumption customers from the pool of participants removed more non-CARE (21%) than CARE (15%) customers, most likely improving the relative performance of the non-CARE sub-group to the point where the performance gap has been narrowed.

These outcomes collectively suggest that recruitment, device selection, and installations are the key drivers of program outcomes.

Sub-LAP events

PG&E can also trigger events of the SmartAC™ program by California Independent System Operator (CAISO) defined sub-Load Aggregation Point (sub-LAP) grid geography. Beginning in 2018, more events will be triggered by sub-LAP because the SmartAC™ program will be bid into the CAISO energy markets as a Proxy Demand Resource. For program year 2017, sub-LAP events were triggered on five separate days and spanning eight sub-LAPs, most of which were called twice. Because sub-LAP events control the loads of all available customers in that sub-LAP, there is no natural randomized control design available for their evaluation. Instead, they are evaluated with matched control groups assembled from larger pools of potential controls based on similarity to the event participants. Due to the flexibility of evaluating without randomized controls, these matching methods are the subject of increasing interest and discussed in the methods section of this report.

In general, sub-LAP event dispatch is known to suffer from lower response rates than serial group dispatch, due to reliability issues associated with inconsistencies in control device setup/configuration and the paging system that serves as the communication platform for the legacy 1-way communicating technology. In 2017 PG&E began the installation of more reliable 2-way communicating switches which leverage PG&E's AMI network but those were not subject to load control events and have not been included in this evaluation. Load impacts for the existing device mix averaged 0.36 kW (13.4% of reference loads) but range from 0.15-0.18 kW (6.4 to 8.0% of reference loads) in the South Bay and North Coast sub-LAPs to 0.70 to 0.74 (21.3 to 22.0% of reference loads) in the Kern sub-LAP. This heterogeneity underscores the degree to which prevailing local temperature conditions drive load impacts and the potential benefit to the grid of being able to focus responses on specific geographies.

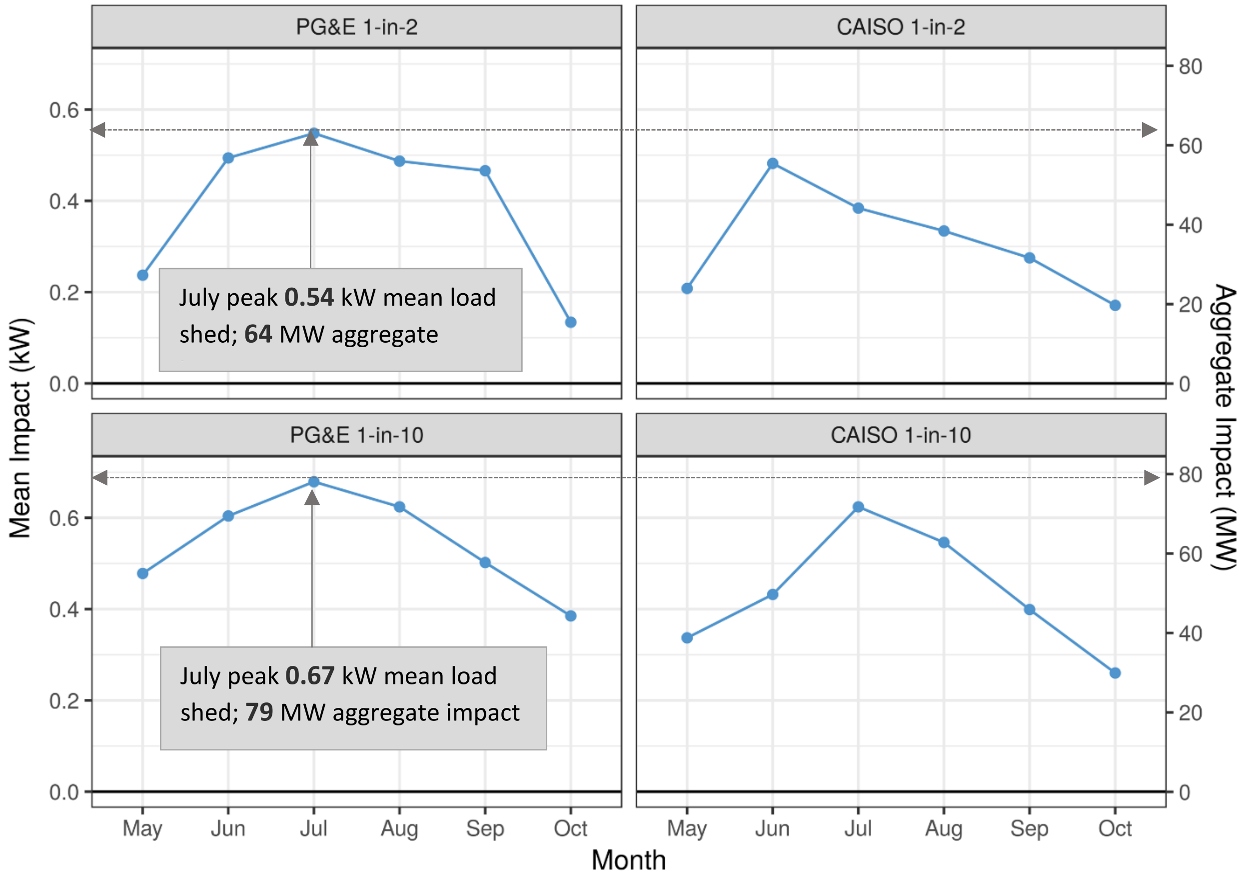
Emergency event 5/3

An emergency event on 5/3 called all participants via serial groups 0-9. However, it was not a particularly hot day and may have preceded the changeover to running AC in the summer for many households. Load impacts, estimated by treating the event as simultaneous sub-LAP events in all sub-LAPs, were estimated at 0.12 kW (7% of reference loads). This result underscores the importance of outside temperature in determining the aggregate resource available via SmartAC™. On the load impact estimation front, this event presented technical challenges due to the lack of resemblance between the event day and the comparison days used for the difference in differences estimate(s), which were selected as the 20 hottest days in each weather station territory that were not SmartAC™ days to begin with.

Residential ex ante load impacts

Ex ante load impact estimates represent the expected average and aggregate load impacts that would occur during a SmartAC™ event under normal (1-in-2) and extreme (1-in-10) weather conditions if all customers were called simultaneously.

Figure 2. Mean and aggregate 2018 ex ante impacts



While average load reductions per household were higher in 2017, the range of aggregate load reductions were lower primarily due to lower participation. Participation in the program fell from 153,000 at the end of 2016, to approximately 117,600 during the 2017 season. This was partly due to the removal of low performers, and partly due to natural attrition that was not back filled for various reasons.

The aggregate ex ante impact of the program has been reduced by about 20% in the hottest months (June through September) in the PG&E 1-in-2 weather year, and by about 12-15% in the PG&E 1-in-10 weather year. As shown in Figure 30, this year's ex ante model is more weather sensitive than prior years. The mean impact per customer is higher in the 2017 analysis than in the 2015 analysis in hot conditions – the PG&E 1-in-10 weather year and the hot months of the other weather years – and comparable or lower in cooler conditions. This increased sensitivity is primarily due to changes in the structure of the ex ante model that improve overall fit and minimize biases associated with specific LCAs and we believe it is a more accurate assessment of the resource. Specifically, the 2017 ex ante model includes independent temperature response terms for the Greater Fresno Area and the Greater Bay Area because models without these terms tend to substantially over-predict the load shed in Fresno and under-predict shed in the Bay Area.

Finally, to increase the statistical support for their models, prior evaluations have trained their models using historical data, including from the program years 2011 through 2013. However, those years feature ex post load impacts systematically higher than measured in subsequent years. This year's decision not to include prior years' data corrected the over-prediction of load impacts caused by those older results.

Other findings: opt-outs

During the season, the fraction of customers who opted out of at least one event is very small: fewer than 2,800 customers opted out at least once (~2.4%) in 2017. Of these, a few hundred people opt out routinely—with 180 of these opting out 10 times or more times during the season. These serial opt-outers would most likely contribute nothing to the program even if called to participate.

Interestingly, most opt-outs are from people who were not participating in an event when they chose to opt out. As such, the perception of being uncomfortable appeared to have a higher relationship to opting out than the level of comfort from the air conditioner being cycled.

Overall, opt-outs do not currently affect the aggregate impact of the program to a significant degree since the number of opt-outs is a very small fraction of the total enrollment.

Conclusions and recommendations

Based on the evaluation results, there appear to be significant benefits to continual program modernization through more targeted recruitment and deployment of the newer technology. The disparity across device types is particularly striking. These serve as a reminder that recruitment, device selection, and installations are critical to program outcomes. Especially because they also rank among the most significant program costs, these are resource-defining aspects of the program.

PG&E will need to weigh investments versus costs but in general this program does provide value as represented in its ability to shed load over the next 10 years. Especially because a lot of the program's costs are front loaded into recruitment, devices, and installations, care should be taken to recruit and retain customers who are good fits for the program and equip them with devices that are expected to be reliable and flexible over time. For customers already in the program, however, retention decisions should consider only the costs of their ongoing participation. Even if they are not the best performers, the money invested in them is already spent. They may well deliver more value than the ongoing costs of keeping them enrolled and dispatched.

The value of this program was also demonstrated during an emergency event on May 3rd when CAISO called a "Stage 1" emergency due to a confluence of reduced generation and electricity import capacity, and a large demand forecasting error leading to shortfalls of power on the grid.

Recommendations to Improve the program

Based on the current research, we offer five recommendations for the program.

Program recommendation	Continue	Explore	Description
Target high users	✓		The program should continue to target homes with higher reference loads and more potential to save on hot days (as they have been doing since 2014). The program will also need to recruit additional participants (at least to account for attrition) to keep aggregate ex ante load reductions up.
Upgrade Hardware	✓		Upgrading hardware, or at least using the newer 2-way AMI switches for new participants, is expected to continue to increase the average impacts per participant. The Gen 2 switches seem to be performing much better. PG&E should continue to replace ExpressStats PCTs, and use improved hardware going forward, where possible.
Characterize sub-LAP events over a range of conditions	✓		The plan for market integrated events for program year 2018 and beyond will dispatch devices at the sub-LAP level in the service of grid balancing. Because it is tied to grid stability, this dispatch will happen over a wider range of operating conditions, including in cooler weather and with off-peak timing. To the extent that 2018 events are not guaranteed to be called under these conditions (it will depend on grid/market status), it is recommended that test sub-LAP events are called to ensure that evaluation data spanning anticipated conditions is available.
Adopt a value-driven framework for customer retention		✓	PG&E should assess the value of load impacts versus the cost of continuing to include legacy customers in the program. If there are no or low additional costs (i.e., all costs to date are sunk) then there may be reason to encourage legacy customers to stick with the program. They will likely lower average load reduction per device, but from a purely value centric perspective they could continue to serve the program.
Explore satisfaction of opt-outers to help with targeting		✓	We recommend additional research on opt-out customers. If customers become dissatisfied if they participate in several multi-hour events in a year and therefore drop out of the program or stop participating in events, it will limit the effectiveness of the program and complicate the ability to predict the load shed in late-season events. Additional research in this area will help with targeting and predicting load shed in the future.

Recommendations to improve load reduction estimates

In addition to the program recommendations, there are two areas where we recommend additional research to inform the load reduction estimates and evaluation approach.

Evaluation Areas	Continue	Explore	Recommendations to improve load reduction estimates
Explore multi-device load reduction		✓	Although they are no longer being recruited, we recommend further exploration of the impacts of legacy multi-device customers to help increase the accuracy of future estimates (both serial and sub-LAP). The established evaluation methods for this program, including those employed in this report, do not to accurately correct the over-representation of multi-device customers as event participants and under-representation as controls. Further, under sub-LAP events, multi-device customers should see all their devices dispatched at once. This should lead to oversized load reduction per-multi-device-customer (as compared to under-performance in serial group events), although most likely more modest load reduction per-device.
Explore other control methods		✓	We recommend additional research to explore synthetic control group matching. We believe this work could have broad benefits to PG&E. Better control-matching would improve sub-LAP event impact estimates, but also improve the performance of any PG&E program evaluation (or other evaluation) that relies on case-control analysis that is not based on randomized controlled trials.

Overview of SmartAC™ Program in 2017

PG&E's SmartAC™ program utilizes direct load control switches on central air (AC) and programmable communicating thermostats (PCTs) at residential premises to reduce electricity demand during times of peak system usage. This report presents the findings from the impact evaluation of residential customers during the 2017 season, which ran from May through October.

Events and event trends over time

When a SmartAC™ event is called, the control devices limit the duty cycles of AC units, thereby reducing demand. For the purposes of this evaluation and report, **we define "events" as SmartAC™ control device activation for a group of participants that share a date, start and end time, and geography**. For example, calling one set of customers for 1 to 3pm on the same day that another set is called 2 to 3pm would qualify as two separate events.

During 2017, there were three types of events called across eight local capacity areas (LCAs):

Serial group events are called using the last digits of the serial numbers, i.e. 0-9, assigned to the load control devices. This scheme partitions enrolled devices into 10 sub-groups randomly distributed within the full population of enrolled devices. There is no reason to believe that any "serial group" is systematically different from any other, so this allows non-participants to serve as randomized controls for participants for a given event. In practice, there can be several events called on the same day, so controls are actually identified as the serial groups not called during the entire day. This randomized control design allows event impacts to be estimated using the difference between the mean load shapes of the control and participant groups. These events can occur at a variety of times and days between May and October.

Sub-LAP events are called using geographic indicators which, in the legacy 1-way devices, was sent only over the air and then beginning in 2014 was programmed by technicians into each piece of hardware upon its installation. Sub-LAPs are portions of the grid defined by the CAISO and are the geographies used for location sensitive markets and pricing.[1] Sub-LAPs are smaller than, and fit within, LCAs, the other grid-relevant geography considered as a part of this evaluation. The SmartAC™ program serves customers in 16 sub-LAPs that fall within 7 named LCAs and one "Other" LCA. The sub-LAP events called in 2017 were all from 4pm to 7pm between July and October.

Grid-induced or emergency events are called when grid conditions threaten its stability and load curtailment can help reduce those stresses. In 2017, there was a single emergency event called on May 3, 2017 at the very beginning of the season. On this day, CAISO called a "Stage 1" grid emergency due to a confluence of reduced generation and electricity import capacity and a large demand forecasting error leading to shortfalls of power on the grid. Due to the urgent nature of the situation, all group serial numbers were called at once - thus there was no control group. The relatively moderate weather conditions and lack of control customers are the defining features of this event from an evaluation perspective.

In total, there were 16 event days with 36 discrete events called during the 2017 season.

---

[1] The sub-LAPs were re-aligned by CAISO to fit within LCAs going into the 2017 program year. http://www.caiso.com/Documents/Sub-LoadAggregationPointRealignmentDiscussionWebConference9-1-16.html

- 30 events were serial group events
- 5 events were sub-LAP events
- 1 event, on May 3rd, was an emergency event called due to grid conditions with every enrolled device called at the same time.

Table 1 below provides the key details for the emergency event and all serial group events. SmartAC™ customers are also allowed to participate in PG&E's critical peak pricing program, SmartRate™. Just over 20% of SmartAC™ customers are also enrolled in PG&E's SmartRate™ program. For these dually enrolled customers, PG&E cycles participants' air conditioners during the SmartRate™ peak period from 2 to 7pm on all days, known as SmartDays™, when critical peak pricing is in effect. For this reason, these customers are not eligible for serial group or sub-LAP event dispatch on SmartDays™ and must be excluded from event analysis. This is why each event listing includes whether it occurred on a SmartDay™. It has been reported in the past and we confirm again this year that SmartRate™ customers tend to shed less load when they are dispatched for SmartAC™ than customers that are only enrolled in SmartAC™ (See Figure 17 in the ex post results for example). This means that per-customer program impacts should be expected to be slightly higher on SmartDays™ than non-SmartDays™.

To ensure that the messages are received and that AC units are being curtailed at the start of each event, control device dispatch signals are sent out 30 minutes prior to the official beginning of each event. An official start of 5pm would see the control signals going out at 4:30pm. **As such, events display apparent "pre-shedding" of loads.**

Table 1: Summary of all 2017 serial group events, plus the emergency event

Event type	Date	Smart day?	Group(s) called	Start	End
emergency	5/3/2017	no	0, 1, 2, 3, 4, 5, 6, 7, 8, 9	7 PM	10 PM
serial group	6/19/2017	yes	4, 9	5 PM	7 PM
			7	5 PM	8 PM
			8	6 PM	7 PM
			0	8 PM	9 PM
	6/22/2017	yes	8	6 PM	7 PM
			4	7 PM	8 PM
	7/7/2017	yes	5, 7	4 PM	7 PM
			6, 8	5 PM	6 PM
			4, 9	5 PM	7 PM
			0	7 PM	8 PM
	7/15/2017	no	1	12 PM	3 PM
			0	3 PM	6 PM
			3	6 PM	9 PM
	7/27/2017	yes	5, 7	3 PM	6 PM

Event type	Date	Smart day?	Group(s) called	Start	End
			6, 8	5 PM	6 PM
			4, 9	5 PM	7 PM
	8/1/2017	yes	8	6 PM	7 PM
			1, 4	7 PM	8 PM
			3	8 PM	10 PM
	8/2/2017	yes	4	4 PM	5 PM
			8	5 PM	6 PM
	8/27/2017	no	0	12 PM	3 PM
			3	3 PM	6 PM
			1	6 PM	9 PM
	8/28/2017	yes	4, 9	5 PM	7 PM
			5, 7	5 PM	8 PM
			6, 8	6 PM	7 PM
			1	8 PM	9 PM
	8/31/2017	yes	8	6 PM	7 PM
			4	7 PM	8 PM

Table 2 below provides the details for all sub-LAP events called during the summer of 2017. Note that sub-LAP events all share the same timing.

Table 2: Summary of all 2017 sub-LAP events

Event type	Date	Smart day?	Group(s) called	Start	End
sub-LAP	7/6/2017	yes	PGF1, PGNP, PGZP	4 PM	7 PM
	7/28/2017	no	PGF1, PGKN, PGNP, PGZP	4 PM	7 PM
	7/31/2017	yes	PGKN, PGNC, PGSI	4 PM	7 PM
	9/11/2017	no	PGP2, PGSB	4 PM	7 PM
	10/24/2017	no	PGP2, PGSB	4 PM	7 PM

Participants and device trends over time

The makeup of both participants and devices over time can affect the magnitude of the impacts. In 2017, the program experienced large changes in the number of participants--and consequently, large changes in the number of devices.

Removal of low-AC-consumption customers and net de-enrollment

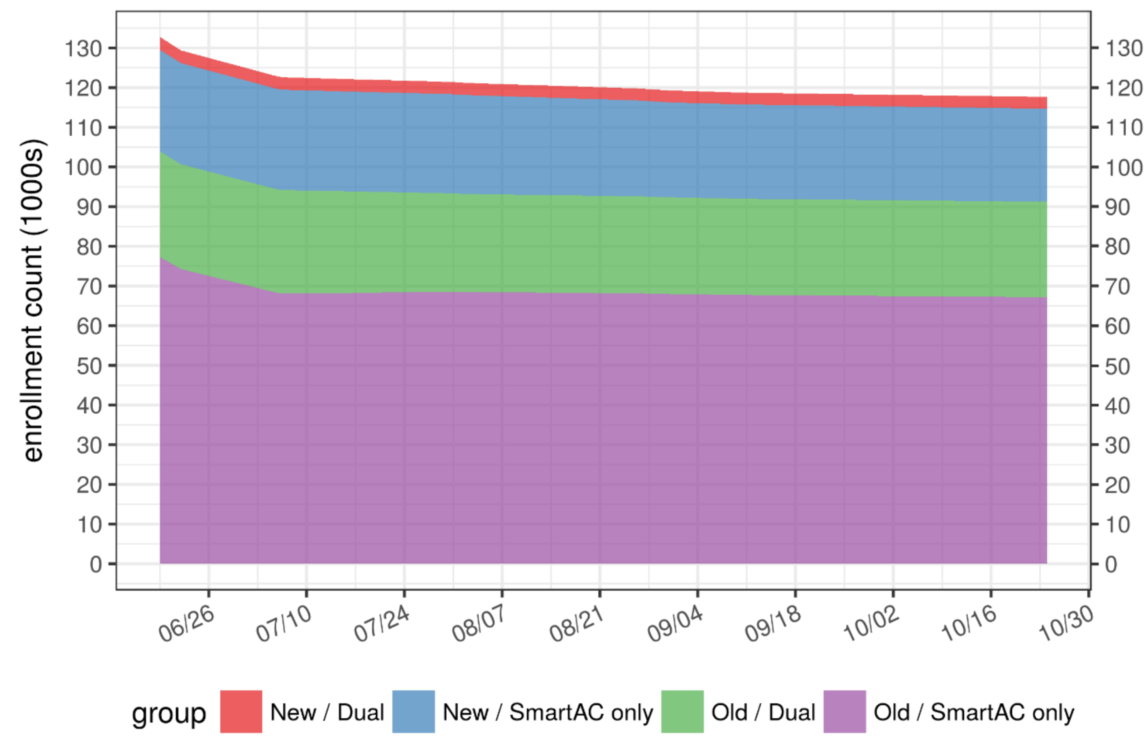
Historically, the largest driver of program attrition has been the automatic de-enrollment of customers when they move. At the end of the 2016 season, there were approximately 153,000 participants, some of which de-enrolled over the off-season. The 2017 season began with 146,500 residential participants.

Approximately 23,800 program participants were de-enrolled during May (13,800) and June (10,000). Roughly 22,000 of these were de-enrolled as low-AC-consumers, with the balance leaving due to natural attrition. The low-AC- consumers, averaged load sheds of 20% of the rest of the population and typically comprised under 5% of aggregate load shed.

There was more net natural attrition in 2017 than previous years due to the strategic retreat from planned “Automoves” recruitment, where customers would be defaulted into SmartAC™ if they were formerly SmartAC™ or moved into a SmartAC™ house. Due to customer pushback, the program discontinued this recruitment tactic very early in the year leading to unusually low recruitment of new customers in 2017. The enrollment for the last event of the 2017 season was approximately 117,600 customers.

The population of SmartAC™ participants can be split up by geography and other attributes. Two important attributes that are relevant to understanding program mechanics and differences in load shed among customer sub-groups are whether customers were recruited using “Old” vs. “New” marketing strategies and whether customers are dually enrolled (“Dual”) in both SmartRate™ and SmartAC™ programs or “SmartAC™ only.” These groups are discussed in more detail in our ex post results and ex ante methods sections, but we introduce them here to provide context. The 2017 program year enrollment trends are illustrated in Figure 3 below.

Figure 3: In-season 2017 program year enrollment changes



The program ended the year with 117,600 customers. Table 3 shows the total number of customers in 2017 and summarizes the differences between enrollments and device counts between program years 2016 and 2017. We provide additional details on 2017 enrollments in the table below to support our 2017 analysis.

Table 3. Participant and Device Overview (end of 2016 compared to the end of 2017)

	2016		2017
Number of customers	**153,400**		**117,600**
SmartAC only	117,900	*New Smart AC only*	67,172
		Old Smart AC only	23,467
Dual enrolled	35,500	*New dual enrolled*	24,063
		Old dual enrolled	2,960
Number of devices	**170,500**		**127,508**
Switches	153,000	*Gen 2 LCR Switches*	28,286
		Gen 1 LCR Switches	86,786
PCTs	17,500	*Utilipro PCTs*	9,374
		Express Stat PCTs	3,062

The location of these customers and devices also affects impacts primarily because climate and temperatures vary across the geographic regions (i.e., LCAs). Table 4 provides a detailed accounting of device counts by type and LCA at the conclusion of the 2017 program year.

Table 4: SmartAC™ control devices at end of 2017 program year

Local Capacity Area	ExpressStat	Gen 1 Switch	Gen 2 Switch	Utilipro	Total
Greater Bay Area	1,259	30,411	7,057	2,422	**41,149**
Greater Fresno Area	1,024	14,926	5,939	1,585	**23,474**
Sierra	198	15,159	4,140	1,325	**20,822**
Other	342	13,533	5,223	1,720	**20,818**
Stockton	226	5,972	2,359	592	**9,149**
Kern	0	3,131	2,552	1,224	**6,907**
North Coast and North Bay	13	3,652	1,016	506	**5,187**
Humboldt	0	2	0	0	**2**
All / total	**3,062**	**86,786**	**28,286**	**9,374**	**127,508**
Note: We do not present Humboldt data in our analyses due to the small number of participants. There are far too few to analyze with statistical methods.					

Other context for 2017

Other factors also affected impacts in 2017. The description of these factors is based on our analysis of available data and is presented here to provide context for the impact results.

- **Weather trends:** 2017 events were slightly hotter on average than in 2016, but similar to 2015. (See upper left graph in Figure 4.)
    - More events at hotter temperatures would be anticipated to increase average impacts for ex post, but the difference in temperatures is modest and the pattern varies by LCA.
    - With unusually cool event hours as well as unusually hot ones, they also spanned a larger range of temperatures than recent years.
- **Variation in event hours:** There were more events in later hours than in 2016 (See Figure 5.) and other prior years.
    - More events at later times of day would be anticipated to shift average results higher because of higher loads, likely driven by higher occupancy in the later hours.

Note that as shown in the figure below, there are much more significant weather differences across LCAs than across years, but the 2017 events were as hot or slightly hotter than events in recent years.

Figure 4: Average event hour temperature by program year and LCA for the four most recent years

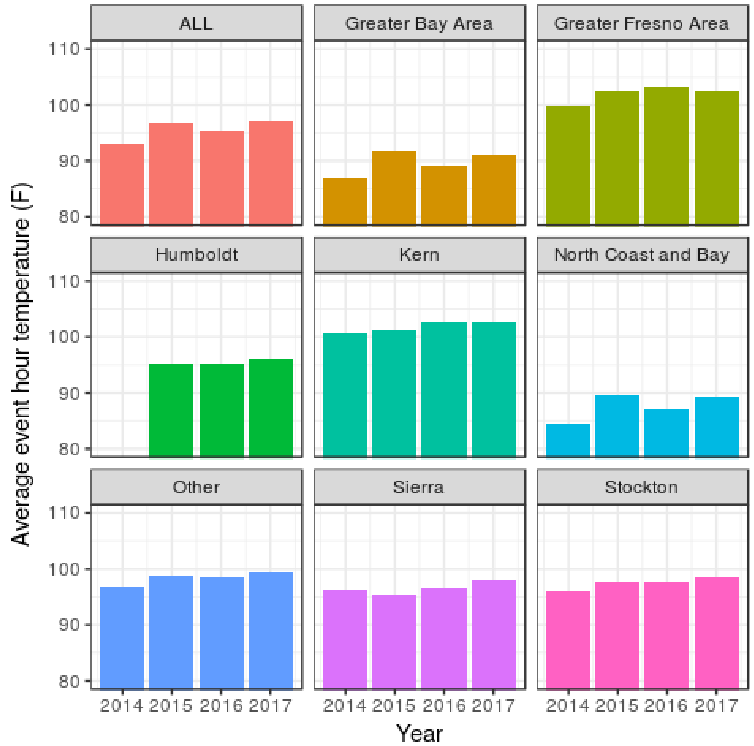
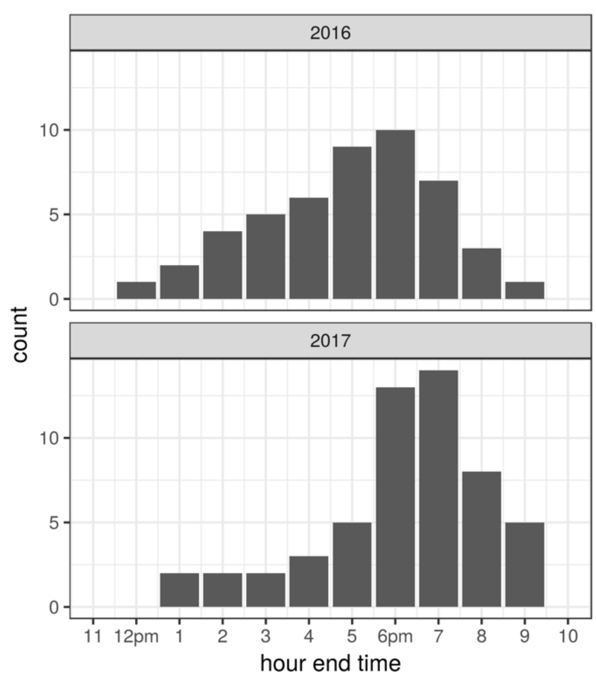


Figure 5: Distributions of serial event hours for 2016 and 2017



Methods

This section summarizes the methods used to perform the ex post and ex ante evaluation analyses that this report presents.

Data requirements

Several categories of data were needed to perform the 2017 SmartAC™ program impact evaluation:

(1) 24 hours of meter data for all SmartAC™ participants for all event days and 20 additional event comparison days, which are defined as the 20 hottest non-event days local to each weather station.
(2) Customer account characteristics for all participants that include enrollment, opt-out, and de-enrollment dates, and all of the customer attributes that are used to separate out sub-groups of customers for focused analysis. See Table 5 for details on customer characteristics of interest.
(3) Local weather data for all customers for all event and event comparison days.
(4) Device type and (de)enrollment timing for each program device.
(5) A large random sample of additional customers (for constructing synthetic control groups for sub-LAP analyses, as discussed below)
(6) Data to characterize each SmartAC™ event and SmartRate™ period, including which customers were enrolled and participated; the start and end time of the event; which customers opted out.
(7) Four years of prior ex post results data to compare our ex post outcomes and ex ante models to previous evaluation results.

Experimental Design for Ex Post Estimation

The schedule of serial group and sub-LAP events was designed to support both evaluation and the use of event outcomes as inputs into predictive ex ante models. Although events mostly tend to be called on the hottest days of the year, the ex ante model benefits from events that span the months of the summer season, represent a range of outside temperatures, ensure that different groups were called for varying numbers and durations of events, ensure that all sub-LAPs were called multiple times, ensure that events were triggered for different times of day.

Analysis of customer sub-groups

For every serial group event and for every LCA, several permutations of customer attributes were varied to isolate and study customer sub-groups with specific characteristics. Table 5 provides a comprehensive set of the customer characteristics used to define sub-groups. A separate estimation of outcomes was made for every distinct value of every attribute, resulting in 9 LCAs x 35 distinct attribute values = 315 sub-groups examined per-event.

Table 5: Customer attributes used to define customer sub-groups in ex post analysis

Customer variable	Customer attribute	Potential values
building_type	building type	Single family, Multi-family
multi	device count	Single device (default), Multiple devices, Single or Multiple devices
care_ind	CARE recipient	Yes, No
device_type	device type	Gen 2 Switch, Gen 1 Switch, Unknown,

		Utilipro, ExpressStat
marketing_strategy	recruitment strategy	New (as of 9/2014), Old
smart_rate	dually enrolled in SmartRate™	Yes, No
net_mtr_ind	net-metered for PV	Yes, No
sub-LAP_id	sub-LAP location	PGF1, PGCC, PGSI, PGNP, PGST, PGEB, PGSB, PGP2, PGKN, PGNB, PGZP, PGFG, PGNC, PGSF, PGHB, PGSA
local_capacity_area	local capacity area location	Greater Fresno Area, Greater Bay Area, Sierra, Other, Stockton, Kern, North Coast and North Bay, Humboldt

Customers with each distinct value of each attribute were isolated for separate analysis for each event. The exception was that **when it was not specified, the device count was defaulted to "Single".**

In the absence of data on building types, the 2017 SmartAC™ evaluation uses customer addresses to determine whether a building is a single-family home or an apartment.[2] This is done with a strategy of finding "apartment words" in the addresses. Certain words are indicative of belonging to a large building, such as a room number or a unit. If an address contains one of these words, the building type is flagged as "multifamily," in the form of an indicator variable in the dataset.

The list of "apartment words" was created by exploring commonly occurring words in the addresses dataset. Addresses containing these words were evaluated for the possibility of false positives; addresses without these words were probed for false negatives. The final list of "apartment words" was chosen based on the lowest rate of these errors and used to create the *building_type* single family vs. apartment indicator variable in the dataset. The words "apt", "unit", "ste" (suite), "bldg", and "#?", where the ? could match any number 1-10 and any letter of the alphabet, were used to identify addresses of apartments. Variants on "lot" and "spc" (space), which were used to flag apartments in prior evaluations, were found via spot checking on Google Maps to most consistently relate to mobile or tract homes and were not used to flag apartments this year. In total, we identified 5969 customers in apartments.

The new vs. old marketing strategy variable identifies customers recruited after a change in approach to recruiting customers was put in place in September of 2014. The newer customers are a better fit for the program and tend to have larger load impacts.

Analysis to support ex ante estimation

As established in previous evaluations, the key drivers of event shed levels, and therefore predicted event outcomes heading into the future are outside temperature, time of day, the groups defined by LCA, marketing strategy, SmartRate™ dual enrollment, and customer count. To support ex ante modeling, a separate run of ex post estimates was made for all 4 permutations of marketing strategy and SmartRate™ enrollment status.

[2] This method matches earlier evaluations.

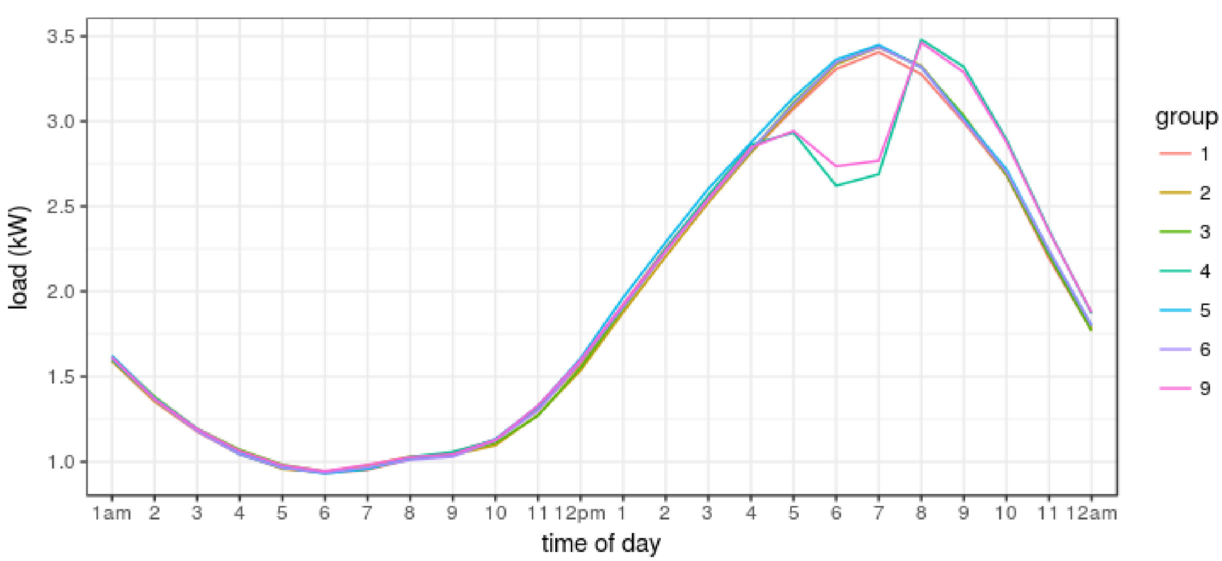
Table 6: Customer attributes used in ex ante modeling

Customer variable	Customer attribute	Potential values
multi	device count	Single device (default), Multiple devices, Single or Multiple devices
marketing_strategy	recruitment strategy	New (as of 9/2014), Old
smart_rate	dually enrolled in SmartRate™	Yes, No
local_capacity_area	local capacity area location	Greater Fresno Area, Greater Bay Area, Sierra, Other, Stockton, Kern, North Coast and North Bay, Humboldt

Ex post difference in differences calculations

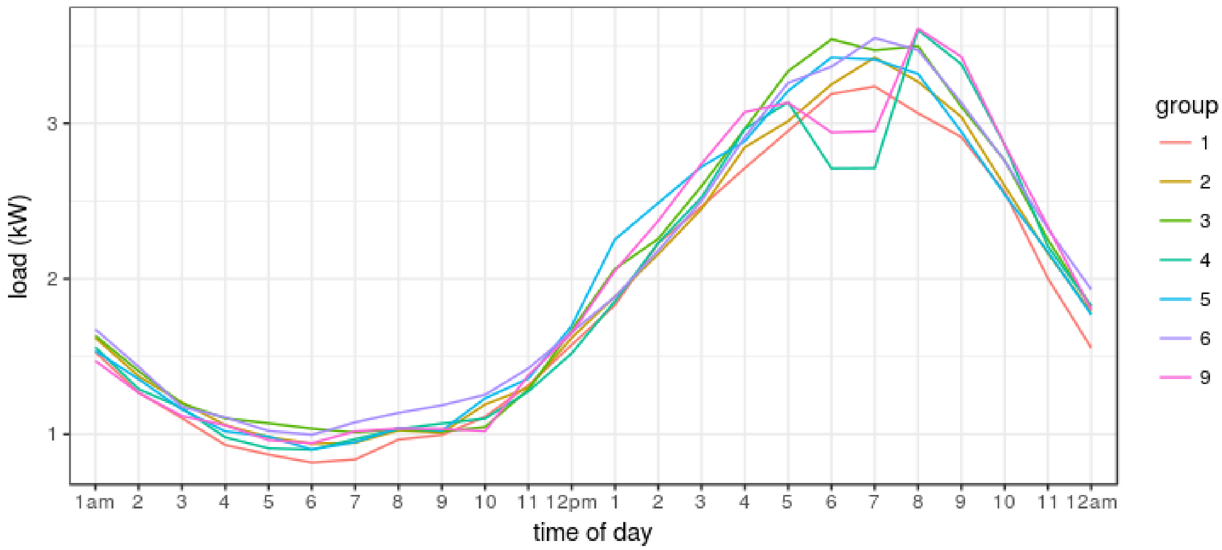
The basic mechanics of an event are illustrated in Figure 6 below. All serial groups of customers have very nearly the same loads prior to the event window. When the event arrives (technically the dispatch signals are sent out 30 minutes prior to the event window to ensure the signal has been received and is impacting cooling loads when the official event begins), the groups being called, numbers 4 and 9 in this case, deviate from their paths, but all the other groups continue without disruption. Here groups 1, 2, 3, 5, and 6 can serve as controls for groups 4 and 9 to quantify the depth of event load reduction and the snapback afterwards.

Figure 6: Average loads on a typical event day for 63,000 customers by group



However, the tight relationship between the event participants and their controls is loosened up as we isolate smaller sub-groups of customers. Figure 7 illustrates this point by recreating Figure 6 using a random sub-sample of 2,000 customers. The effects of variation across individuals don't average out with this sample size.

Figure 7: Average loads on a typical event day for 2,000 customers by group



Assuming some customers are consistently above the average and others below, one way to improve estimated load reduction is to compare how customer loads change on event days relative to non-event days, rather than directly comparing loads. This is called a difference in difference and it is implemented as follows.

A set of 'comparison days' are chosen to be similar to the event day except that no SmartAC™ event was called. For each hour of the event day the following equation is used, where L is the estimated load impact and h is the hour.

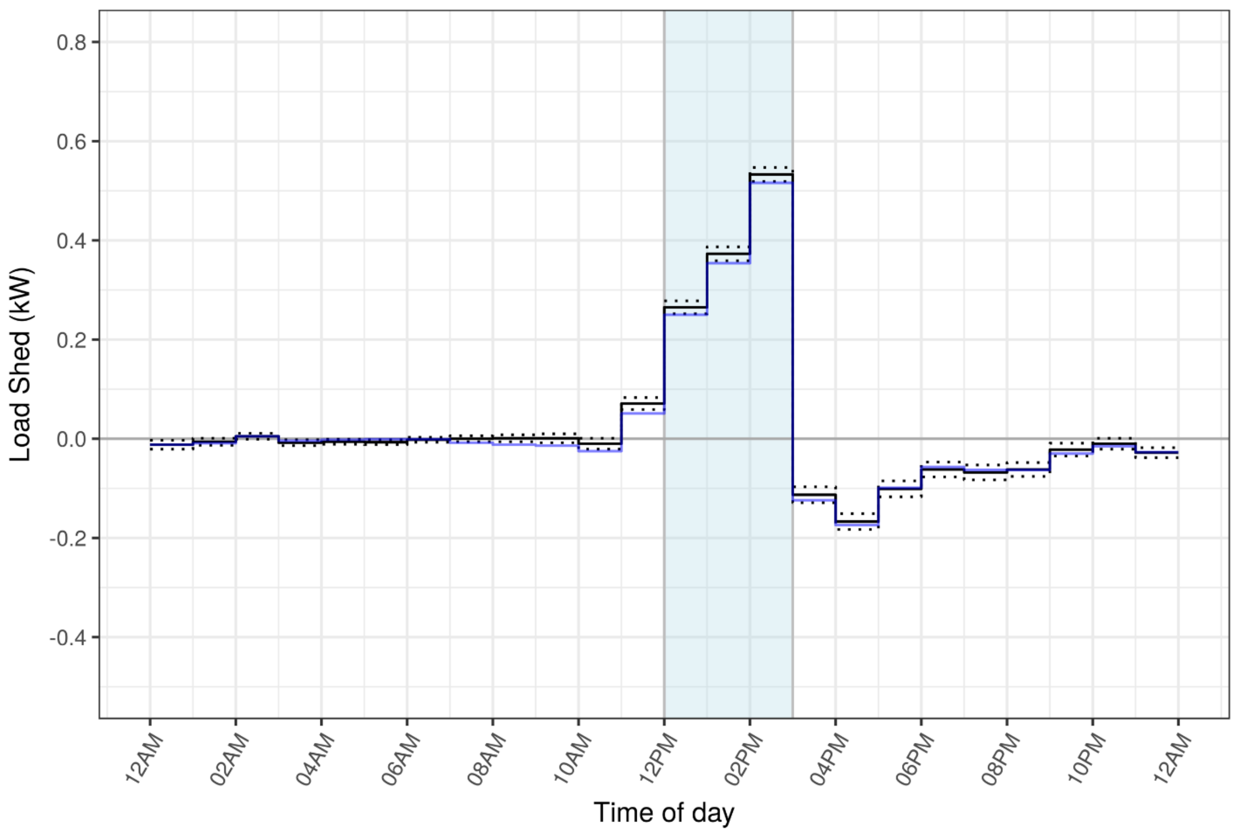
Equation 1: Difference-in-difference estimate

$$L^{event}(h) = \left(kW_{cases}^{event}(h) - kW_{controls}^{event}(h)\right) - \left(kW_{cases}^{matched}(h) - kW_{controls}^{matched}(h)\right)$$

The first term of Equation 1 is a direct comparison of cases and controls during the event hour, a concept that is familiar to most people. The second term adjusts for any difference between cases and controls that is present on the comparison days. The assumption is that if the mean load of the cases is lower or higher than that of the controls on the comparison days, this would lead to an under- or over-estimate of the load shed on the event day. One way to think about this: suppose the analysis were performed for an 'event day' on which no SmartAC™ event actually occurred. The load impact on such a day would be 0 by definition, but this would not be the case if there is a systematic difference between cases and controls. The second term in Equation 1 attempts to adjust for that difference.

The effect of using both terms in Equation 1, rather than just the first term, is small for serial events without filtering down to sub-groups. For those events, cases and controls are assigned using a randomized controlled trial design, so any systematic difference between the load of the cases and of the controls would be due only to sampling statistics: when customers are split into cases and controls the mean load of the two groups will not be *exactly* the same. But the SmartAC™ program is large enough that there are tens of thousands of cases and controls for each event, so the differences in mean load are very small and the second term therefore has a small effect. An example is shown in Figure 8, which shows the estimated load shed for each hour of an event day using just the difference between case and control loads on that day (blue) and the full difference-in-differences estimate (black). Even with over 10,000 cases and 75,000 controls, there is a slight difference between the estimates.

Figure 8: Load shed estimates based on the first term of Equation 1 (blue) and on the full equation (black) for a serial group on an event day.



The figure also shows uncertainty bounds around the difference-in-differences estimate (dashed lines). The uncertainty calculation is most easily understood if the terms of Equation 1 are rearranged into Equation 2:

Equation 2

$$L^{event}(h) = \left((kW^{event}_{cases}(h) - kW^{matched}_{cases}(h)) - (kW^{event}_{controls}(h) - kW^{matched}_{controls}(h)\right)$$

The first term now involves only the cases, and the second term only the controls. The first term in parentheses is the mean, over all of the cases, of the load difference between event day and matched days, and the second term is the same for the controls. The sample-size-related uncertainty in the first term is the variance, over all of the cases, of the load difference between event day and matched days, divided by the number of cases, and similarly for the controls.

The standard error of the impact is:

Equation 3: Standard error in impact estimate for a serial event

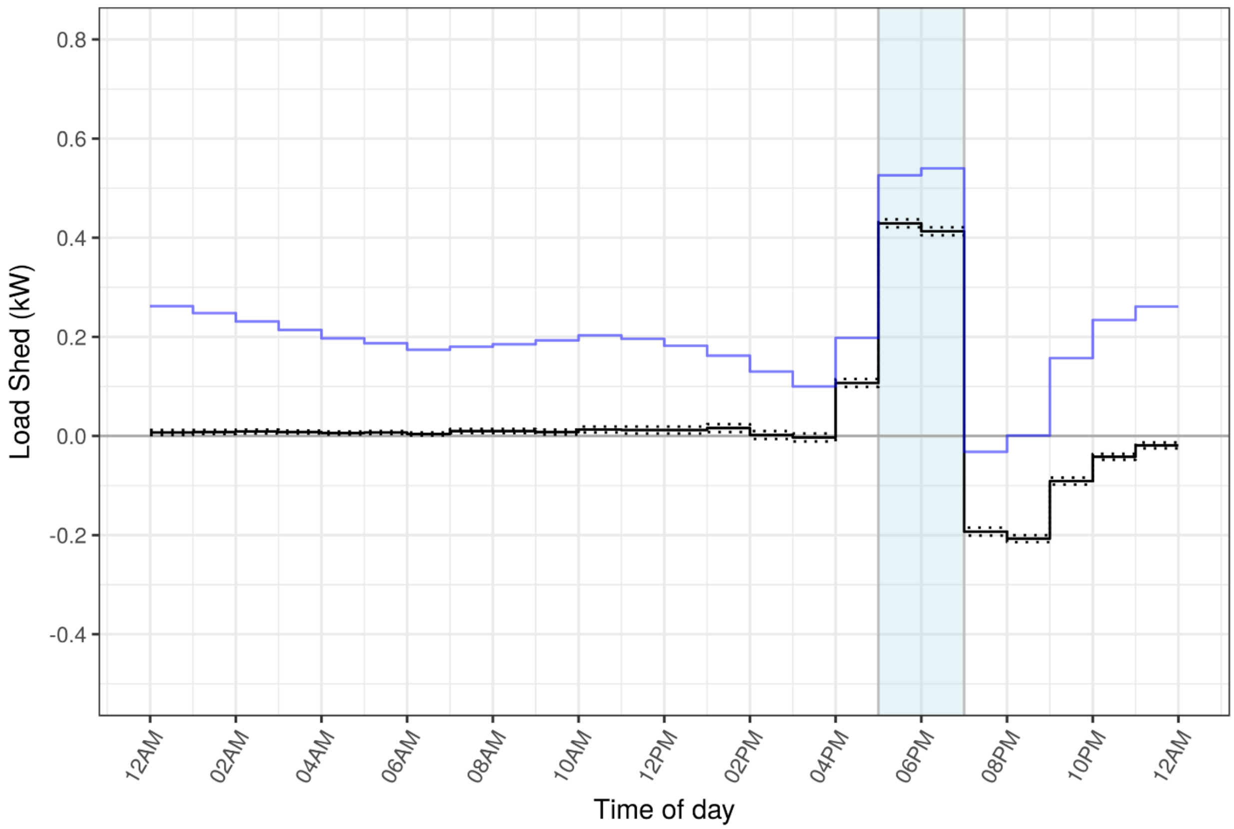
$$SE = \sqrt{\frac{Var(term\ 1)}{n_{cases}} + \frac{Var(term\ 2)/}{n_{controls}}}$$

Although the difference-in-differences approach is only a very slight improvement over the raw difference between cases and controls for a large randomized controlled trial, *the difference is much larger for lower-customer-count sub-groups within serial group events and the sub-LAP events*, for which controls have to be matched to cases.

The goal when choosing controls to match cases is to choose controls such that, if there had *not* been an event, the estimated impact would have been zero. This is true (on average) when cases and controls are randomly selected from the same population, which is why the randomized controlled trial design is so powerful. But when a randomized controlled trial is not possible, a statistical model must be used to choose controls. Figure 9 is the equivalent of Figure 8 for one of the sub-LAP events. The controls were chosen to match the temperature sensitivity of the cases, not the load shape on an average day, and there is a large difference between the load of the cases and the load of the controls even in the hours prior to the event, as shown in blue. The difference-in-differences method adjusts for the systematic difference in the hourly load between cases and controls, allowing us to select controls expected to respond to changing temperatures in the same manner as the event participants, leading to a more accurate estimate.

In other words, electricity usage is comprised of many end-uses and they all vary in magnitude and fraction of total loads across utility customers. The SmartAC™ program only controls AC loads and is therefore only concerned with that portion of total loads. By matching on temperature sensitivities and using a difference in differences approach to measurement, we are matching on the AC loads and isolating them for measurement. Even though the simple difference in loads between participants and controls (blue line) is way off zero (i.e. due to other loads), the way the controls vary from their comparison days is driven by weather changes and the difference in differences provides a nice tight estimate of event impacts on AC loads.

Figure 9: Load shed estimates based on the first term of Equation 1 (blue) and on the full equation (black) for a sub-LAP event.



Impact estimation procedures

Serial group event impact estimation

In 2017, our analysis moved from a simple differences approach for serial group events (used in earlier years) to a difference in differences approach to attempt to make the impact estimates more robust.

Serial group event estimation was performed by implementing a difference in differences procedure following these steps:

(1) Identify all participants in the event and participants in other events on the same day.
(2) Define the control group as all enrolled customers not in either of these categories.
(3) Optionally filter customers by category/sub-category values if the estimate is to focus on a sub-set of customer types. If not, we still apply a default filter that ensures that (a) no SmartRate™ customers are analyzed on SmartDays™ and (b) that default estimates are made for only single device customers.
(4) Optionally filter customers by local capacity area, if the estimate is to focus on a specific geography.
(5) Load meter data from the event day and comparison days for the remaining participants and controls.

(6) For both participants and controls, compute the aggregated mean *comparison* day hourly load shape across all comparison days by taking 24 hour-of-day averages across all relevant customers.
(7) For both participants and controls, compute the aggregated mean *event* day hourly load shape by taking 24 hour-of-day averages across all relevant customers.
(8) Subtract the mean comparison load of the controls from the mean comparison load of the participants and do the same for the event loads.
(9) Subtract the resulting event day difference from the resulting comparison day difference. These 24 hourly values are the difference in differences RCT estimate of the load impact for that event.
(10) Basic statistics, like participant and control counts and other relevant metrics, like the standard errors associated with each hourly impact estimate (see Ex post difference in differences calculations for a description of these calculations), hourly outside temperature, hourly reference load, etc. are also computed and reported alongside the impact estimates.

Key assumptions

1. The foundational assumption required to assume randomized controls is that the last digit of each device serial number will be randomly distributed in the population of enrolled customers such that the groups called and not called for an event should be interchangeable in terms of their aggregate consumption (See Figure 6).
2. A corollary is that the number of participants and controls will be high enough that the aggregate loads do in fact average out the high and low extremes of loads.
3. By adopting a difference in differences evaluation approach, i.e. incorporating comparison days rather than computing a simple difference between participants and controls on the event day only, we make the impact estimates more robust to fixed differences between participants and controls that can emerge as the number of customers in the sub-groups being examined decrease (i.e. when assumption #2 doesn't hold).
4. Difference in difference assumes only that the *difference between comparison days and the event day* of the participants and controls are the same, so constant differences in baseloads, etc. that have nothing to do with the loads shed by the program and will subtract out of the difference between comparison days and the event day data can be overcome.

Sub-LAP event impact estimation

Sub-LAP events, which call all customers in a specific geographic area and therefore lack embedded controls, follow the same estimation procedure as the serial events, except step #2 above (i.e., Defining the control group). For sub-LAP events, sampling the control group for the event becomes a separate pre-processing procedure.

Control matching procedure (substitute these for step 2 above):

The control group in this procedure is selected from a pool of 1M candidate controls for whom we have meter data and weather data from all event days and comparison days as well as sub-LAP assignments and basic customer attributes.

(2.1) Identify the comparison days for event participants. Note that these will differ by weather station, as the comparison days are defined as the 20 hottest non-event days by weather station, and that a given sub-LAP can span the territories of several weather stations. In other words, one weather station is assigned to each customer based on their location but customers that share a sub-LAP can collectively be assigned to multiple stations.

(2.2) Load comparison day data for all participants (i.e. all enrolled customers in the sub-LAP(s) being examined) and all *candidate* controls assigned to the sub-LAP(s) being examined. Because comparison days are defined at the weather station, participants and candidate controls are partitioned by weather station assignments. The control(s) ultimately assigned to each participant is/are drawn from the same weather station territory.

(2.3) Select a “distance metric” that can be used to assess the similarity between two customers’ data. Typical metrics would be the Euclidean distance between average comparison load shapes, or the difference in observed temperature sensitivities between the customers. The later, with pre-computed comparison day temperature sensitivities for all participants and candidate controls is used to generate our official results.

(2.4) Compute the distance metric between all participants within each weather station territory and all candidate controls in that territories.

(2.5) Select the closest (i.e. smallest distance metric) N candidate controls for each participant, where N tends to be 1 or 2 or 3 to produce control groups 1-3x the size of the participant group. Our final results are based on N=2.

(2.6) Pool all the candidate controls selected in this manner into a synthetic control group and continue with the steps of difference in differences impact estimation.

Key assumptions

All of the same assumptions for the serial group apply to the sub-LAP estimation. In addition, in the sub-LAP estimation we also make the following assumptions:

1. The distance metric used for matching controls provides groups that will function well as controls. Specifically, their (non-participating) event day loads would diverge from comparison day loads by the same amount event participants would if they were not participants.
2. Temperature sensitivities (increase in daily total kWh for each degree of daily average temperature) are the key driver of the difference between comparison and event day loads.

This was tested by looking for model estimates of zero impact when running the sub-LAP impact estimation machinery on non-event days (when there is no event impact by definition), and for zero impact during the hours leading up to an event.

Distance metric comparison

We compared matched control groups based on load shapes (as used in prior evaluations) and temperature sensitivities and chose to use a matched control group based on temperature sensitivities.

Matched control groups are constructed by computing a distance metric between each participant and all potential controls and selecting one or more control(s) with the smallest distance value(s). For example, if we wanted to match controls one to one to participants using average load, we would compute the average load for all participants and all potential controls and select a control for each participant with the minimum absolute difference in average load.

Because averaging dampens both load timing and temperature sensitivity, average load is not an ideal metric for matching SmartAC™ controls. While both load shape and temperature sensitivity approaches perform adequately, we believe matching by temperature sensitivity is more robust across a range of outside temperatures due to its focus on AC loads and simplicity. Here we provide details on both approaches and their comparison.

Load shape matching

To match event participants to their closest controls by load shape, we compute 24 hourly load averages across all comparison days for all participants and all potential controls. These average load shapes can then be compared based on their difference in magnitude across all hours. A reasonable distance metric in this setting is the Euclidean Distance, or square root of the sum of the squared differences at every hour, between load shapes. In mathematical notation, the distance D between load shapes L1 and L2 across hours of the day h is:

$$D_{1,2} = \left(\sum_{h=1}^{24} (L1 - L2)^2 \right)^{1/2}$$

We add potential controls with the least distance from participants to the control group.

Temperature sensitivity matching

To match event participants to their closest controls using temperature sensitivity, we run a regression on daily total loads vs mean daily outside temperature with data from all comparison days across all event participants and potential controls. This provides a sensitivity coefficient for each customer with units of kWh/day per °F increase in daily average outside temperature. In other words, it is an estimate of the magnitude of cooling loads on site.

Lining up these calculations requires (1) computing daily total loads for all comparison days (typically 20 per customer) for both participants and potential controls (2) using customer weather station assignments to match customers to a table of daily weather statistics for PG&E's territory. With 1M potential controls, this is computationally intensive in practice, but it can all be pre-computed in advance of any event impact calculations.

We match potential controls with the smallest difference in temperature sensitivity from participants to form the control group.

How does this method compare to load shape matching?

Prior evaluations have matched controls using average load shape on comparison days and found it necessary to implement a same-day adjustment where the pre-event consumption diverged between the participants and controls, explaining *"This small adjustment was applied to all hours of the event day. This adjustment* ***accounts for any potential differences in weather sensitivity between the SmartAC™ customers and control groups****. By controlling for any discrepancies in load on the actual event day, it increases the accuracy of the impact estimates across events."*

In other words, when using load shape matching it is necessary to adjust control group loads because the controls and participants respond differently to changes in outside temperature. The loads of the controls were comparable on comparison days, but the proportion of their loads coming from AC vs. all else were actually different. When weather conditions exceed the comparison days (which event days typically do by definition - events are called on the hottest days of the year when the grid needs them most), the total consumption for customers with larger AC systems will diverge from smaller ones relative to the loads during the comparison days, skewing the impact estimation results and creating artifacts like apparent negative load reduction leading into the event period.

The correction applied for the previous evaluation adds a fixed offset to each hour of the day to line up the hours leading into the event with zero load impact, but this undermines the use of controls by allowing their data to be adjusted using participant data (similar to comparing the participants to their own pre-event data) and it again ignores the fact that the controls will be systematically less responsive to any changes in temperature between the adjustment and event periods.

It is simpler and more appropriate to control for differences in weather sensitivity by employing control group selection methods that minimize those differences the in the first place.

Emergency event impact estimation

The emergency event, called on May 3, was technically invoked using serial groups - all of them were called at once. With every enrolled customer participating in the event, there were no controls against which to measure load reductions, so we treated this event as multiple parallel sub-LAP events. To get estimates of the load impact across all sub-LAPs, we combined the hourly load impacts and uncertainties from each individual sub-LAP estimation using rules for adding normally distributed variables (i.e. their means and standard deviations). This event also presented technical challenges due to the lack of resemblance between the event day and the comparison days used for the difference in differences estimate(s), which were selected as the 20 hottest days in each weather station territory that were not SmartAC™ days to begin with.

Ex ante methods

Ex ante estimation uses a statistical model trained using ex post event impacts to predict program event impacts up to 10 years into the future, given forecasts of weather conditions and future enrollment counts for several types of participants.

Because they control for the natural variability in weather and event timing that influence the ex post results, ex ante estimates also facilitate direct comparison of program outcomes with prior years. The weather conditions used to estimate ex ante event impacts correspond to typical event days and monthly system peak load days in each month under weather conditions that tend to drive peak loads in PG&E's territory and the larger California Independent System Operator (CAISO) system. The conditions are labeled '1-in-2' and '1-in-10' to express how unusual the weather conditions are: 1-in-2 and 1-in-10 mean that the forecasted weather conditions, particularly the high temperatures, are representative of those that would be seen once every 2 or 10 years during the monthly load peaks, respectively, on average. In total, there are four sets of forecasts that drive ex ante estimates: PG&E 1-in-2; PG&E 1-in-10; CAISO 1-in-2; and CAISO 1-in-10.

Like the models fit for prior program evaluations, the 2017 ex ante statistical model predicts the load shed for customer segments, for each hour of the day, for a given set of hourly temperature conditions. The customer segments used for the 2017 evaluation are:

- **Old / SmartAC™ only:** SmartAC-only customers recruited with the marketing strategy used before 2014 (we call these 'old' customers).
- **New / SmartAC™ only**: SmartAC-only customers recruited with the new marketing strategy. These customers shed more load than old customers, on average.

- **Old / Dual:** Dual-enrolled customers (those in both the SmartRate™ and SmartAC™ programs) recruited with the old marketing strategy.
- **New / Dual:** Dual-enrolled customers recruited with the new marketing strategy.

Ex ante projections are generated as follows:

1. Fit a statistical model to the ex post data from the 2017 SmartAC™ events[3] disaggregated by LCA, for the SmartAC-only customers, using data on ex post load shed, time of day, outside temperatures, LCA, and indicators for the Old and New marketing categories.
2. Use the model with enrollment and weather forecast data as inputs to predict the mean load shed for SmartAC-only customers in both marketing categories, for event hours from 1pm through 6pm for PG&E and CAISO and 1-in-2 and 1-in-10 weather scenarios.
3. Multiply the predicted load shed of the SmartAC-only customers by a scaling factor to predict the mean load shed for the dual-enrolled customers; this is done separately for each marketing category. This is the same procedure that has been used historically. The scaling factor is the ratio of the mean impact of the dual customers to the mean impact of the SmartAC-only customers for the events that did not take place on SmartAC™ days. This ratio was 0.77 for the new customers and 0.70 for the old customers.
4. Multiply the predicted mean load shed in each customer segment by the projected number of customers in the segment in order to forecast the aggregate load shed by category; add these to get the aggregate load shed for the entire program.

Dual-enrolled customers have their hardware controlled to serve the SmartRate™ program on 'SmartDays™', i.e. days when there is a SmartRate™ event. This means they do not keep the same event timing as SmartAC™ events on SmartDays™, and, for practical reasons, their ex-post load reductions on SmartRate™ days are quantified by the SmartRate™ evaluation every year. There were only two serial group event days with dual-enrolled customers participating with SmartAC™ event timing and they did not provide enough statistical support for a detailed ex post model estimate. Instead, the load shed for dual-enrolled customers is estimated using a scaling factor relative to the load shed for SmartAC-only customers.

In addition to the ex ante predictions for event hours, we provide predictions for the hours before and after an event that flesh out the 24-hour event day. There is no reason to believe that events trigger impacts prior to their start,[4] so the pre-event impacts should be zero by definition. The period after the event, however, is likely to experience "snapback" where the air conditioning has to work harder than normal in order to cool the residence back to its desired temperature once the event has ended. The model used for predicting snapback is discussed in the Ex Ante modeling appendix B. In the body of this report we focus on the 1 pm – 6 pm "resource adequacy" event period.

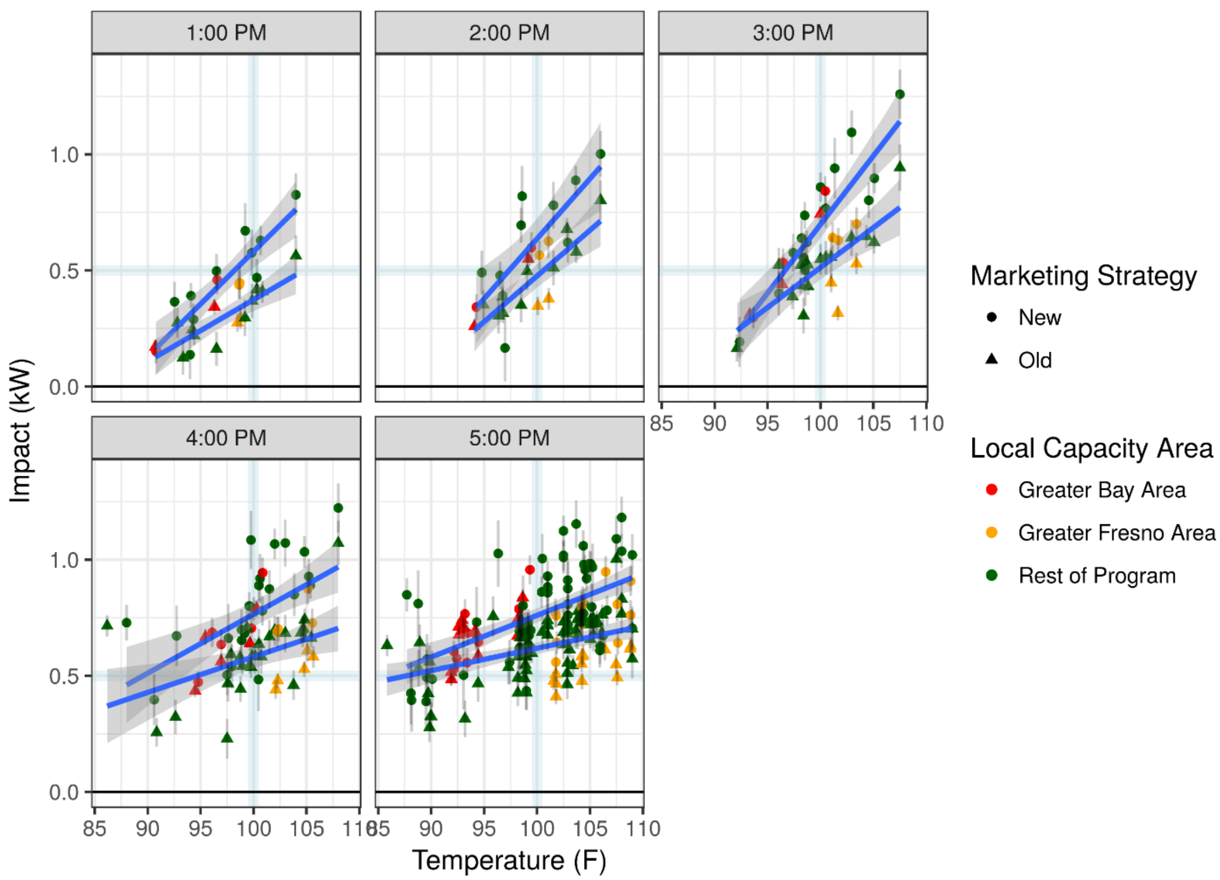
---

[3] Previous evaluations used data from multiple prior program years to fit their ex ante models. For a discussion of why that was not a good strategy for this year's evaluation, see the ex ante modeling Appendix B.

[4] In theory, some people could respond to events that are telegraphed in advance by pre-cooling their homes to prepare for their AC to be scaled back, but the automated nature of SmartAC™ means that there is no expectation of notice prior to events. More importantly, there is little evidence in the ex post results to suggest that people are pre-cooling.

Figure 10 shows the ex post load shed data (as estimated with the randomized control trial methodology) and illustrates how they are used to fit the statistical model that is used for ex ante forecasts. Each point represents the hourly ex post load shed (y-axis) vs. the average outside temperature for the same hour (x-axis) for one serial group event localized to one LCA during one hour of the day. Each of the panels show data from a single hour of the day. For example, the upper left shows the hour 1-2pm and the last panel is for the hour 5-6pm. In each panel, light blue lines cross at a temperature of 100°F and a load shed of 0.5 kW, to facilitate comparing across panels. In each panel, two regression lines are shown: the upper one is the best fit to the data from the customers recruited with the new marketing strategy, and the lower one is the best fit to the data from the older customers.

Figure 10: Mean ex post load shed for each serial group event (y-axis) vs temperature (x-axis), localized by LCA, for customers in SmartAC™ Only, for each hour between 1 and 6pm, labeled with the time at the beginning of the hour.



The figure illustrates several phenomena that the statistical model incorporates in order to produce accurate forecasts:

1. The load shed is strongly related to outdoor air temperature at the time of the event: impacts rise as temperatures rise.
2. The relationship between temperature and load shed is different for old vs new customers: the two slopes on each graph have different slopes.

3. Some hours have higher load shed at a given temperature than others. This is especially true for 5-6pm. The best fit lines have lower slopes (weaker temperature response) offset by significantly higher shed at 90°F (about 0.5 kW) than other hours. We hypothesize that this is the result of people returning home from school and work, causing a systematically different load shed behavior than earlier hours, when shed is more predictable.

The statistical model also takes into account the following:

1. Most data points from the Greater Bay Area fall on or above their corresponding regression lines, suggesting that the model shown in the plot systematically under-predicts the load shed in the Bay Area. Similarly, most Greater Fresno Area points fall below the line, leading to systematic over-prediction. In response, the final model includes Fresno- and Bay Area-specific terms. The temperature slope in the Bay Area is higher than in other regions; that is, each additional degree of outdoor air temperature leads to more additional load shed in the Bay Area than in other local capacity areas.
2. The relationship between load and temperature between midnight and 8 am provides additional information: high overnight temperature is associated with slightly higher load shed, even if the event hour temperature is the same. A term to capture this effect is included in the final model.

Details of the full ex ante model can be found in Appendix B: Detailed description ex ante approach.

Ex post load impacts

As mentioned in the introduction, there were 16 event days with 36 discrete events called during the 2017 season. For the purposes of this evaluation and report, we define “events” as SmartAC™ hardware activation for a group of participants that share a **date, start and end time, and geography**.

Of these:

- 30 events were serial group events (called using the last digits of the serial numbers, i.e. 0-9, assigned to the control devices)
- 5 events were sub-LAP events (referring to geographic indicators set on each piece of hardware upon its installation)
- 1 event, on May 3rd, was called due to grid conditions with every enrolled device called at the same time.

Below we present our overall findings for group events including, serial group events assessed through an RCT design and used to inform our ex ante analysis. We then present findings for sub-LAP events, which were assessed using matched control groups.

Ex Post for serial groups events

The average load reduction per household at 5-6 pm (the most common peak hour) was 0.65 kW per household, or about 21% of the whole house load, which is consistent with, but higher than, 2016 results. However, there is wide variation across LCAs and days based on the time that the event was called and the average temperature during the event. See Table 7 for 5-6pm customer impacts by LCA, compared between 2016 and 2017 and Table 8 for a summary of all 2017 5-6pm event hours.

Table 7. Ex post summary of load reduction: Average per-customer impacts (5-6 pm)

		2016		2017	
All LCAs together	**Average reduction per cust. (5-6 pm)**	**0.55 kW**	**21%**	**0.65 kW**	**21%**
	Range of reduction per cust. (5-6 pm)	0.28-0.71 kW	15% - 23%	0.53 -0.76 kW	19% - 23%
By LCA	Greater Bay Area	0.46	21%	0.65	23%
	Greater Fresno	0.61	19%	0.59	17%
	Humboldt	0.73	27%	Not reported due to small sample	
	Kern	0.73	22%	0.79	23%
	Northern Coast	0.37	20%	0.53	19%
	Other	0.58	20%	0.68	21%
	Sierra	0.59	20%	0.70	21%
	Stockton	0.65	22%	0.66	19%

Note: Several other factors besides LCA can also affect per customers impacts (e.g., SF v MF, CARE, multiple AC units, NEM, etc.). These effects are described in Table 9 and Table 10. Humboldt data is not reported due to the small number of participants in 2017 (de-enrollments and new geographic assignments associated with updated LCAs reduced it from 722 participants in 2016 to 2).

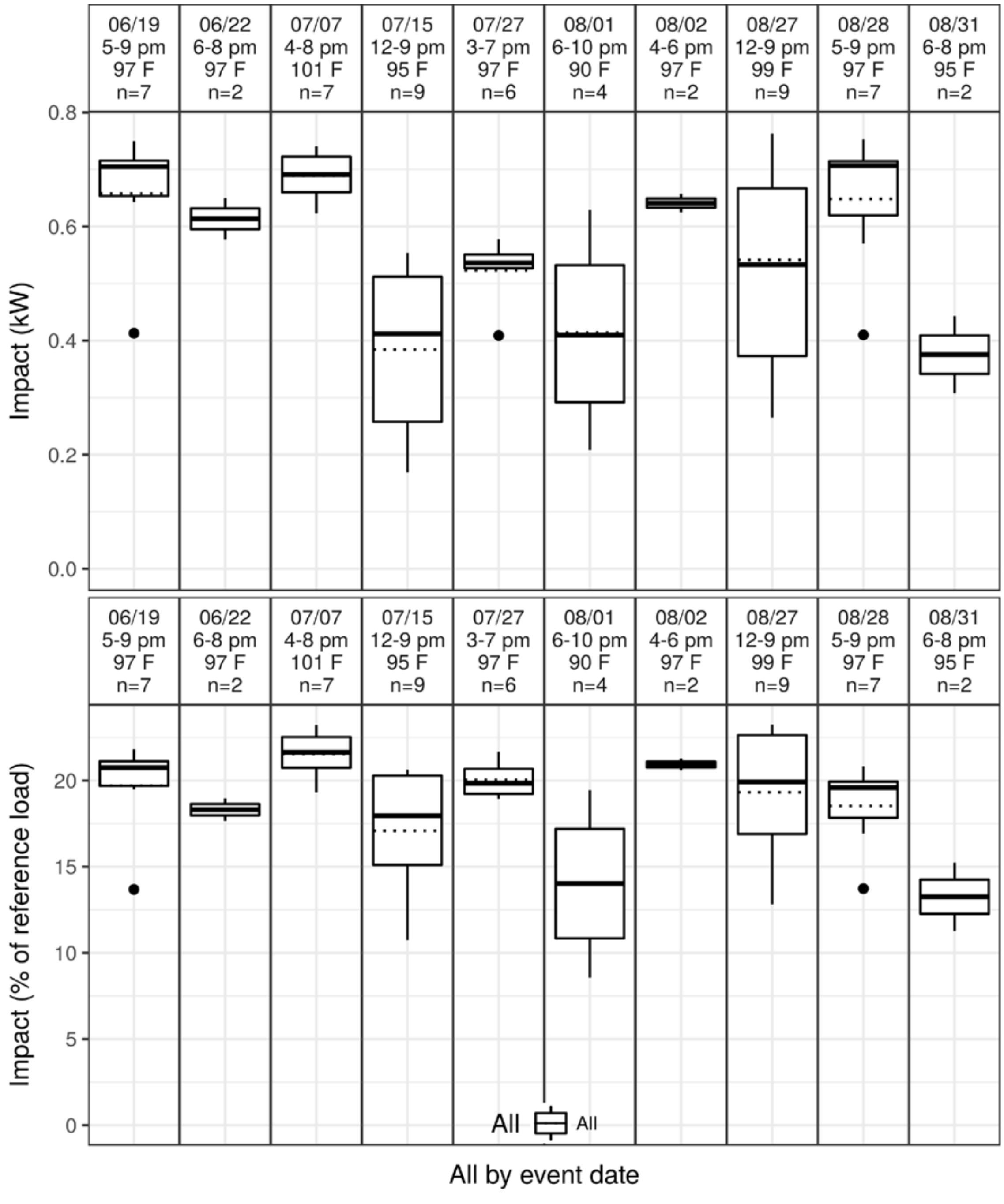
The average hourly impact per customer during 5-6pm (the most common hour) ranged from 0.53 kW on 7/27 to 0.76 kW on 8/27, or 19.1-22.6% of the reference load. As mentioned above, the overall average for the season based on the events called was 0.65 kW, as shown in Table 7.

Table 8. Ex Post loads, impacts, and temperatures for serial group events (5-6pm)

Event Date	Treat Group	Number of Customers Called	Hour End	Ref Load (kW)	Avg. Impact (kW)	Std. Err. of Impact (kW)	% Impact	Aggregate Impact (MW)	Avg. Temp (°F)
06/19	4,9	22,106	18	3.46	0.61	0.01	17.5%	13.4	100
	7	11,042	18	3.46	0.64	0.02	18.4%	7.0	100
07/07	5,7	19,856	18	3.31	0.70	0.01	21.2%	13.9	102
	6,8	**20,028**	**18**	**3.30**	**0.72**	**0.01**	**21.7%**	**14.3**	**102**
	4,9	20,124	18	3.31	0.65	0.01	19.7%	13.1	102
07/15	0	**13,186**	**18**	**2.85**	**0.52**	**0.01**	**18.3%**	**6.9**	**98**
07/27	5,7	19,886	18	2.86	0.58	0.01	20.3%	11.5	97
	6,8	20,075	18	2.85	0.55	0.01	19.1%	11.0	97
	4,9	20,168	18	2.86	0.55	0.01	19.2%	11.1	97
08/02	8	10,145	18	3.33	0.59	0.02	17.9%	6.0	96
08/27	3	13,030	18	3.52	0.65	0.02	18.6%	8.5	103
08/28	4,9	19,986	18	3.67	0.56	0.02	15.2%	11.1	102
	5,7	19,688	18	3.67	0.62	0.02	16.8%	12.1	102
Mean		**17,640**		**3.27**	**0.61**	**0.01**	**18.8%**	**10.8**	**100**

While the tables above report for a common hour (5-6pm), Figure 11 below depicts the range of hourly event impacts across the range of customer sub-categories for all event hours within each of the 10 days when the program called serial events, labeled by date, range of hours, average outside temperature, and count of event hours for each day. The boxes place solid horizontal lines at the 25th, 50th, and 75th percentile values, with dotted lines for the means. As shown in the figure, there is variation in both absolute impacts and impacts as a percentage of reference loads across events. There is also variation across the days throughout the season. This variation is largely driven by event timing and outside temperature, with the highest impacts on hot days with mid-day event hours. For days with a wide range of event hours, especially hours that are either early or late in the day, the spread in hourly impacts in apparent. We explore these drivers of event impacts in the Environmental drivers of ex post results section.

Figure 11: Average event impact (as box and whiskers)



Note: 25th, 50th, and 75th percentiles as solid horizontal lines, with dotted mean for all values. The upper and lower whiskers extend from the 75th and 25th percentile lines, respectively, to the largest/smallest value no further than 1.5 times the IQR (where IQR is the inter-quartile range, or distance between the upper and lower lines of the box). Values beyond that, point are plotted individually as outliers. Labels include the event date, the range of hours the event(s) spanned, the average temperature across all locations during the event(s), and n, the total number of event hours contributing to each day's box and whiskers.

Comparisons to other years

Figure 12 provides a comparison of event impacts between 4 and 5pm vs. the average of all event day temperatures from midnight to 5pm (i.e., mean17, which has been reported on by past evaluations), for serial group events from program years 2014 through 2017, with color coding, regression lines, and gray uncertainties around the regression lines. We present the hour between 4-5pm for comparison to previous reports. The open circles are excluded event data from Fresno, where the signaling network does not perform as well as expected when dispatching events, clouding the patterns that emerge when comparing from one year to the next. The open circles also include a single event from Kern in 2017 (the lowest and hottest open point, where elevated overnight temperatures pushed the mean17 value, but not event temperatures, unusually high).

Overall, 2017 mean load impacts are higher than in earlier years (see figure below). Notice that the regression line for 2017 is roughly parallel to the 2015 and 2016 lines (with 2014 slightly more responsive to outside temperature than other years), but this line is systematically higher. The gray uncertainty range associated with 2017 has little overlap with 2015 or 2016, further suggesting a significant difference. We believe this to be the effect of the program's removal of low performers at the beginning of the program year.

Figure 12: Historical and 2017 4-5pm load impact vs. average of all temperatures from midnight to 5pm (mean17)

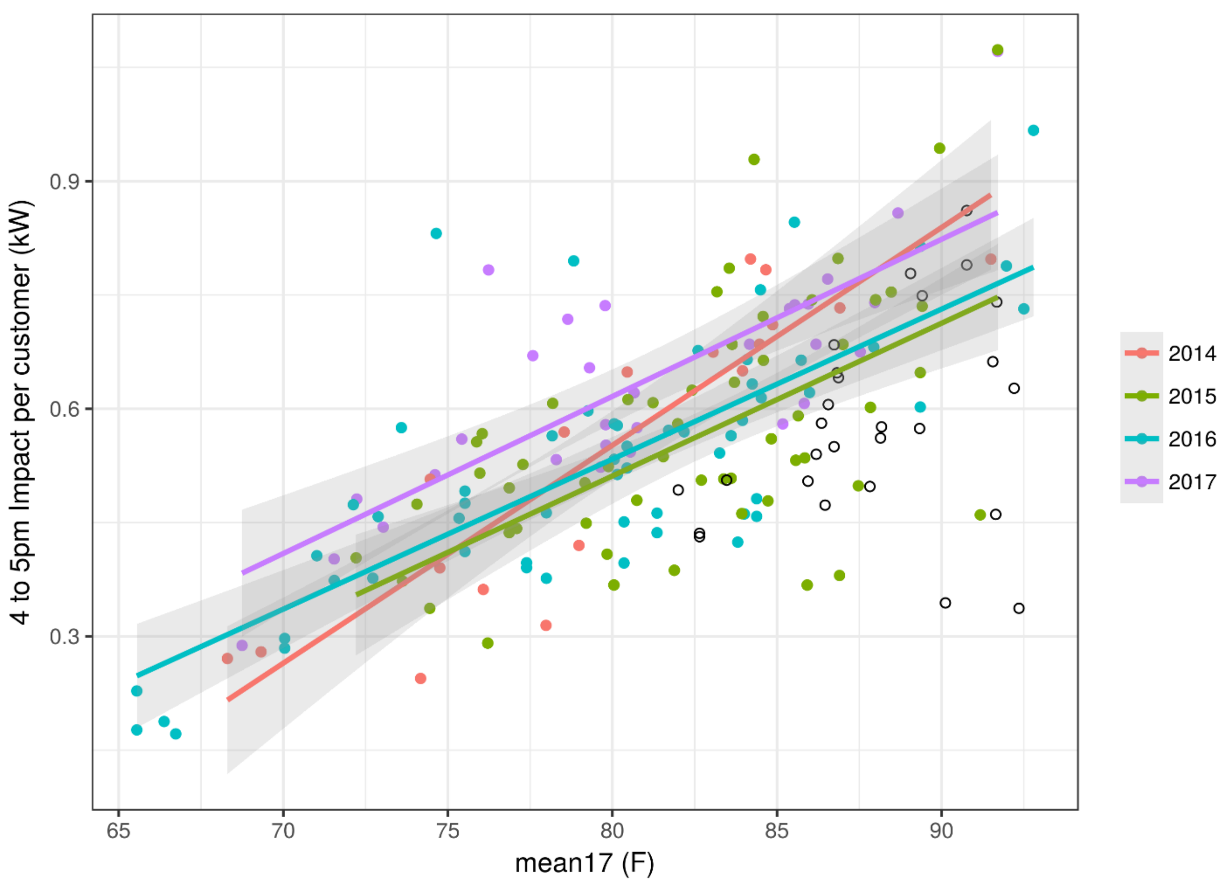
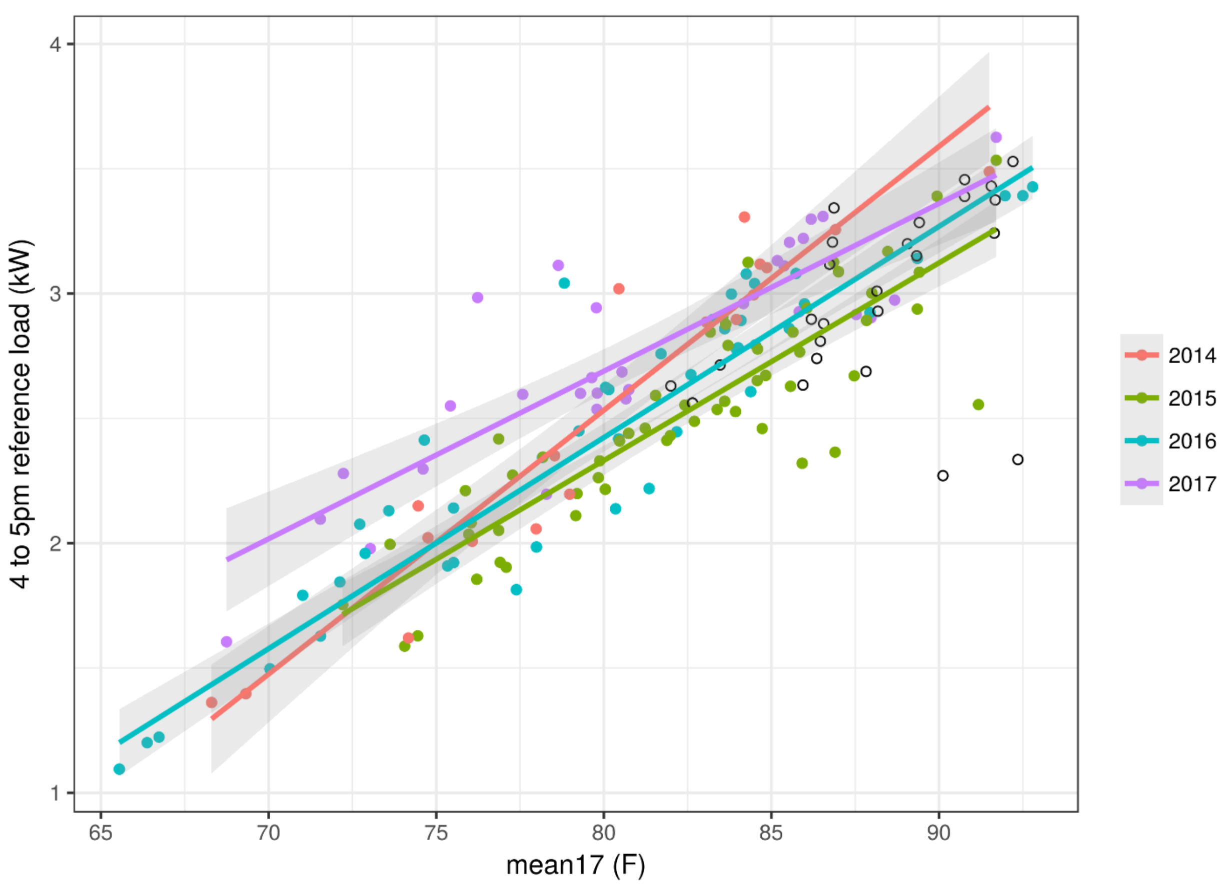


Figure 13 provides a **similar comparison, but of reference loads** between 4 and 5pm vs. mean17 temperature, the average of all event day temperatures from midnight to 5pm, for serial group events from program years 2014 through 2017, with color coding, regression lines, and gray uncertainties around the regression lines. Again, the open circles are excluded event data from Fresno and a single event from Kern in 2017 (the lowest and hottest open point).

As shown in the figure, references loads increase with temperature (in all years), but they are higher among the 2017 participants (and possibly less responsive to temperature) than in 2015 or 2016. Recall that the reference loads come from enrolled, but non-participating, customers for serial group events. Notice that the regression line for 2017 is roughly parallel to the 2015 and 2016 lines (with 2014 slightly more responsive to outside temperature than other years) but the 2017 line is still systematically higher. The gray uncertainty range associated with 2017 has little overlap with 2015 or 2016, further suggesting a significant difference. We believe this to be another effect of the widespread de-enrollment at the beginning of the program year: the average customer is more temperature sensitive than in prior years, so the reference loads during events are higher than in previous years.

Figure 13: Historical and 2017 4-5pm reference load vs. average of all temperatures from midnight to 5pm (mean17)



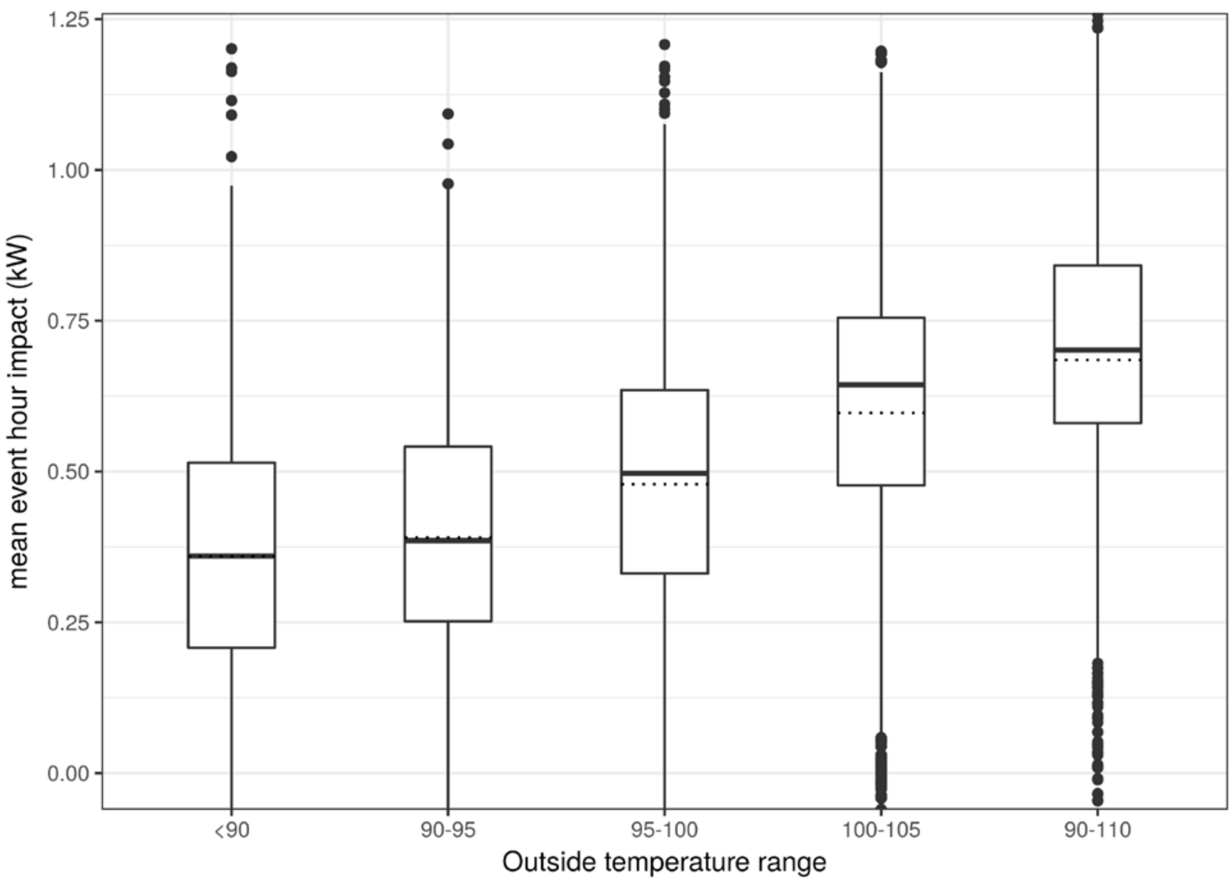
Environmental drivers of ex post results

To understand what drives high and low load impacts, we have created graphics that depict impacts vs. their key drivers. These include impacts as driven by temperature, time of day, and geography.

Temperature

As shown in Figure 14, hours with hotter temperatures tend to have higher load impacts. The figure shows average event load impacts for all participants as a function of outside temperature. The general pattern of higher load impacts in hotter weather can be seen, but the effect is dampened as it starts to level off at lower temperatures.

Figure 14: Outside temperature effect for all event hours across all LCAs in 2017

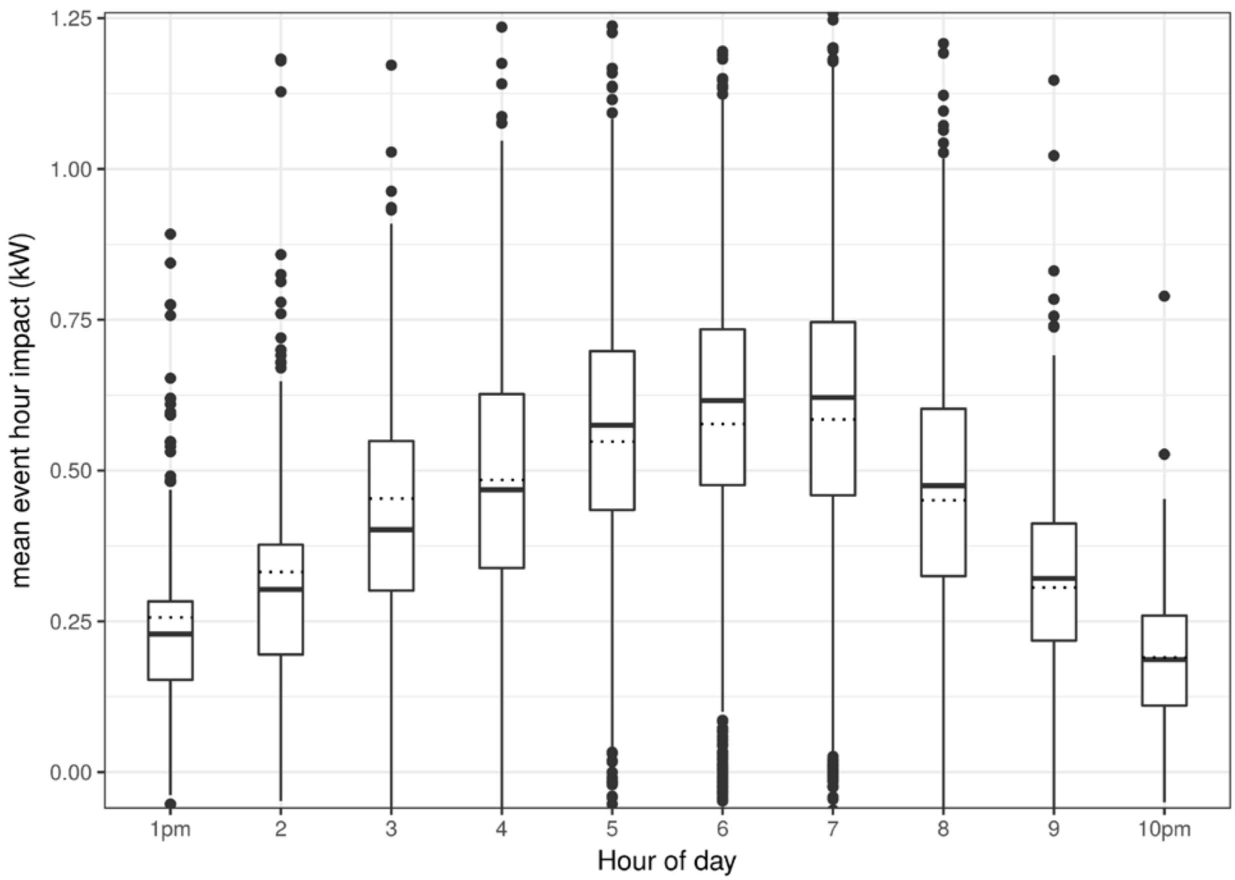


Note: 25th, 50th, and 75th percentiles as solid horizontal lines, with dotted mean) for all customers. (See the note for Figure 11 for a full description)

Time of day

Figure 15 visualizes average event impacts for all participants as a function of hour of day. Based on the figure below, the peak hours for load reduction come between hours ending at 5-7pm, with lower load reduction earlier and later in the day. This is driven to some extent by outside temperatures, but we note that the peak impact hours are later than typical peak daily temperatures (usually around 3pm) and the hours of 5-7pm tend to be when households reach full occupancy after residents return from work, school, and other daytime activities.

Figure 15: Hour of day effect for all event hours across all LCAs in 2017



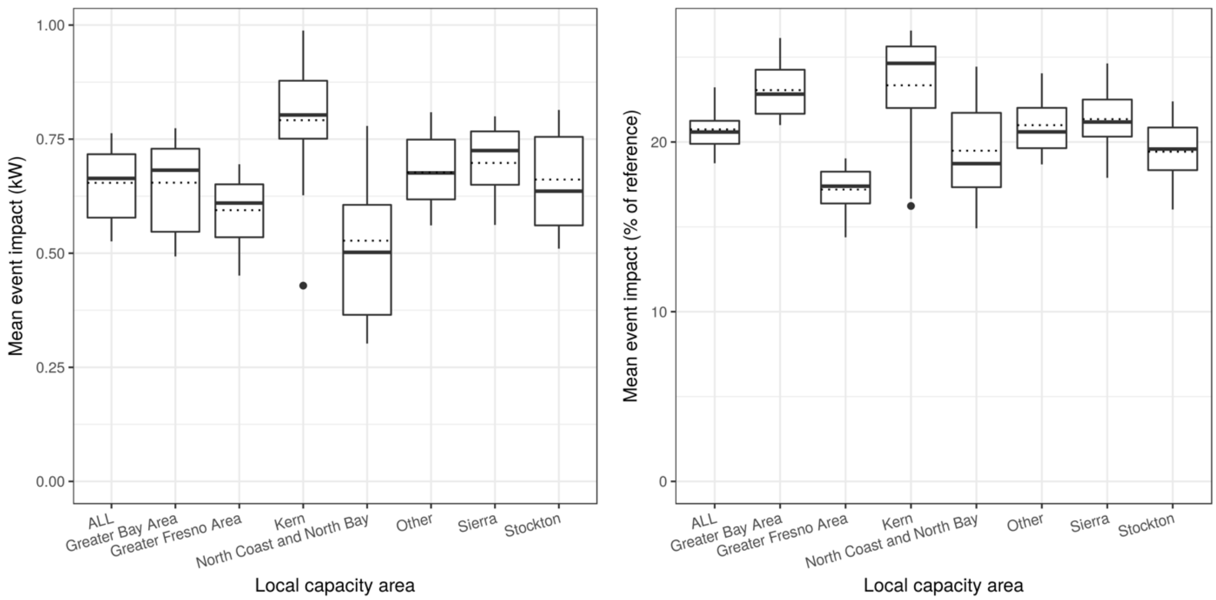
Note: 25th, 50th, and 75th percentiles as solid horizontal lines, with dotted mean for all customers. (See the note for Figure 11 for a full description)

Geography

Figure 16 below shows load impacts for event hours 5-6pm by local capacity area. It displays load impacts in kW and as a percentage of reference loads, but as box and whisker plots.

This figure illustrates that location is a large driver of load impacts. One might assume that climate is the primary source of such variability. Noting that reference loads increase with outside temperature, we do observe that impacts as a percentage of reference loads are fairly tightly clustered around 20%, but the LCAs that deviate from 20% impacts, including the Greater Bay Area and Kern on the high side and the Greater Fresno Area and the North Coast and North Bay on the low side suggest other geographically differentiated factors, like poor paging reception (known to be a problem in Fresno), building style, AC operational norms, enrollment self-selection, etc. play a role in load reductions as well.

Figure 16: Mean event impact (Kw on left and % of reference load on right) from 5-6pm by local capacity area



Note: 25th, 50th, and 75th percentiles as solid horizontal lines, with dotted mean for all customers. (See the note for Figure 11 for a full description)

In addition to the environmental drivers, customer attributes and characteristics also play an important role in determining load impacts. The next section examines the relationship between customer and device characteristics and load impacts.

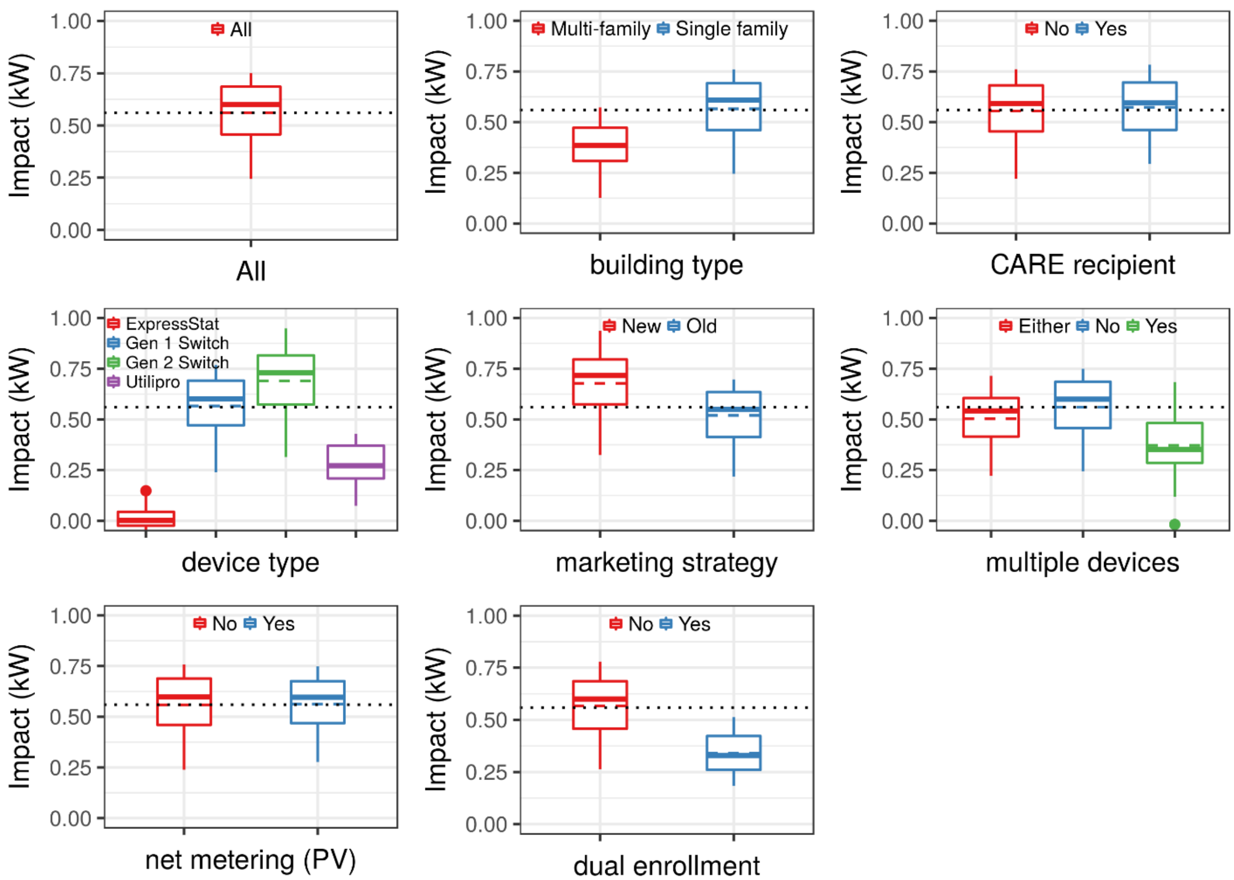
Ex Post findings by sub-group categories

Participants in the SmartAC™ program are not uniform. Their characteristics can vary greatly, with some of these characteristics significantly affecting their potential impacts. Below we provide an overview of our findings by sub-group (or participant characteristics). Note that many of these trends are consistent with past findings; however, we call out where we see differences, or where we are able to use 2017 data to augment past findings. To help guide the program team, we also describe the program implications of these findings.

Figure 17 uses "box and whiskers" plots to show the full range of event-hour impacts for each customer sub-category across all serial group event hours for 2017. The box and whiskers capture variability within sub-group across events largely driven by time of day and outside temperature effects. The figure shows that **Utilipro, ExpressStat, dually enrolled, multi-family, and multi-device customers perform poorly relative to the full population of participants**, with no apparent load reduction at all coming from ExpressStat devices. On the other hand, **newer 'Gen 2 Switches' and customers recruited under the 'New' marketing strategy, systematically outperform the full population of participants**. Other customer characteristics, like whether they have a PV system or receive CARE rates appear not to influence outcomes very much. The former provides a validation of the randomized control design of the program as comparison of loads for customers with vs. without PV will tend to result is dramatic disparities driven by the PV generation.

Based on these results, there appear to be significant benefits to program modernization through more targeted recruitment and better devices and signaling. The disparity across device types is particularly striking. These serve as a reminder that recruitment, device selection, and installations are critical to program outcomes. Especially because they also rank among the most significant program costs, these are resource defining aspects of the program.

Figure 17: Event impacts by customer sub-group



Note: Box and whiskers summary of event impact by sub-category, across all hours of all serial group events. Boxes extend from the 25th to the 75th percentile, with the median, or 50th percentile marked with a horizontal line and average marked by a dashed line. The Black dotted line is the average serial group event load impact across all customers.

Table 9 quantifies the average impact across all event hours in units of kW for all participating sub-groups for every serial group event day and provides a mean across all days for each (see also Table 10 for results presented as relative percentage load reductions). These are the raw ex post evaluation numbers behind the box and whisker plots. Among the quantitative values, we can see that dually enrolled customers only participated in two event days in the 2017 season, providing only a modest amount of data to analyze and draw conclusions from.

Table 9: Event impact by customer sub-category (kW)

Category	Sub-category	Event date 06/19	06/22	07/07	07/15	07/27	08/01	08/02	08/27	08/28	08/31	Mean
All	All	0.64	0.61	0.69	0.38	0.53	0.46	0.64	0.54	0.62	0.38	0.55
Building type	Multi-family	0.44	0.34	0.47	0.30	0.42	0.30	0.37	0.31	0.51	0.29	0.38
	Single family	0.64	0.62	0.69	0.39	0.53	0.47	0.65	0.55	0.63	0.38	0.56
CARE recipient	No	0.63	0.61	0.69	0.37	0.50	0.45	0.66	0.52	0.62	0.38	0.54
	Yes	0.64	0.62	0.67	0.42	0.57	0.49	0.61	0.59	0.63	0.36	0.56
Device type	ExpressStat	0.03	-0.10	0.00	0.02	0.03	-0.02	-0.01	0.06	0.02	0.00	0.00
	Gen 1 Switch	0.65	0.62	0.70	0.38	0.52	0.47	0.67	0.53	0.64	0.38	0.56
	Gen 2 Switch	0.77	0.79	0.82	0.49	0.65	0.58	0.76	0.72	0.74	0.48	0.68
	Utilipro	0.34	0.23	0.36	0.24	0.29	0.20	0.20	0.30	0.34	0.10	0.26
Marketing strategy	New	0.76	0.76	0.80	0.49	0.64	0.58	0.74	0.71	0.72	0.47	0.67
	Old	0.59	0.56	0.64	0.35	0.48	0.42	0.61	0.49	0.59	0.34	0.51
Multiple devices	Either	0.56	0.53	0.64	0.36	0.53	0.42	0.57	0.46	0.52	0.35	0.49
	No (default)	0.64	0.61	0.69	0.38	0.53	0.46	0.64	0.54	0.62	0.38	0.55
	Yes	0.51	0.41	0.52	0.28	0.38	0.29	0.39	0.32	0.32	0.19	0.36
Net metering (PV)	No	0.64	0.61	0.69	0.38	0.53	0.47	0.65	0.54	0.62	0.37	0.55
	Yes	0.63	0.62	0.65	0.43	0.52	0.44	0.61	0.57	0.65	0.40	0.55
SmartRate	No	0.64	0.61	0.69	0.42	0.53	0.46	0.64	0.58	0.62	0.38	0.56
	Yes	NA	NA	NA	0.28	NA	NA	NA	0.40	NA	NA	0.34

Note: All includes only single device customers. This is why “No” under Multiple devices performs at 100%. “No” is equal to All. “Either” represents results when both multi-device customers and single devices customers are included.

Table 10 quantifies average event impacts as the percentage of the full population impacts in units ‘% of All’ for all participating sub-groups for every serial group event day. Note that by default, only single device customers are examined when computing average event impacts. This is why “No” under Multiple devices performs at 100%. This table enables us to see that multi-family customers had load impacts 30% lower than the full population of single device owners, on average. Multi-device customers had impacts 35% lower than the full population (multi-device households are discussed further in the recommendations section). The load impacts associated with Gen 2 switches outperformed the full population by 24%. Based on the available evidence, dually enrolled customers underperformed by 27%. Finally, we are able to report for the first time that the “New” marketing approach outperformed all customers by 22% and “Old” customers by over 30%.

Table 10: Event impact by customer sub-category (% of All customer impacts)

Category	Sub-category	Event date 06/19	06/22	07/07	07/15	07/27	08/01	08/02	08/27	08/28	08/31	Mean
All	All	100	100	100	100	100	100	100	100	100	100	100
Building type	Multi-family	69	56	69	78	79	65	58	57	82	78	69
	Single family	101	102	101	101	101	101	101	101	101	101	101
CARE recipient	No	100	99	101	96	96	97	102	96	100	102	99

	Yes	101	101	97	110	109	106	95	109	101	96	103
Device type	ExpressStat	4	-16	0	5	6	-4	-1	11	3	-1	1
	Gen 1 Switch	102	101	102	99	99	101	105	97	102	103	101
	Gen 2 Switch	121	128	119	127	124	124	118	133	119	128	124
	Utilipro	53	37	53	63	56	42	31	55	55	26	47
Marketing strategy	New	119	124	117	127	122	124	115	131	116	125	122
	Old	94	92	94	92	92	91	95	91	94	91	92
Multiple devices	Either	88	87	94	94	100	90	88	86	84	92	90
	No (default)	100	100	100	100	100	100	100	100	100	100	100
	Yes	81	66	75	72	72	63	61	59	51	50	65
Net metering (PV)	No	100	100	101	98	100	102	101	99	99	98	100
	Yes	99	101	95	112	99	94	95	105	104	107	101
SmartRate	No	100	100	100	108	100	100	100	108	100	100	102
	Yes	NA	NA	NA	73	NA	NA	NA	74	NA	NA	73

Note: All includes only single device customers. This is why “No” under Multiple devices performs at 100%. “No” is equal to All. “Either” represents results when both multi-device customers and single devices customers are included.

The findings by sub-group are consistent with those from past years, with one exception. Prior evaluations have found a larger margin by which CARE recipients outperform non-CARE customers. We observe that CARE customers outperformed for 7 out of 10 event days, with the largest differences tending to come on days with lower overall impacts. This is consistent with the interpretation that CARE customers tend to live in hotter climates and have higher cooling loads on more moderate days, but everyone catches up to them on the hottest days. The most likely explanation for the gains achieved by non-CARE customers this year is the de-enrollment of low performing customers. Over the course of the program year, CARE customers dropped in count by 15% while non-CARE dropped by 21%. This suggests that the non-CARE population performance was improved relative to CARE by dropping a greater proportion of their numbers due to low performance.

Table 11 summarizes the evaluation results by customer sub-group and provides mean values for each segment to allow for comparisons and commentary.

Table 11. Ex post findings: summary by customer sub-group

Subgroup	Total enrollment on last event day 2017	Description of general findings for group	Mean ex post impacts in box & whisker plots
Building type	3,499 MF	Single-family customers produced significantly greater absolute impacts that multi-family premises.	MF 0.38 kW SF 0.56 kW
CARE recipient	33,751 CARE	2016 findings indicate that absolute impacts and reference loads are higher for CARE customers, likely due to hotter climates. There is little difference between the sub-groups in 2017, likely due to the de-enrollment of low performing customers disproportionately improving non-CARE results.	Not CARE 0.54 kW CARE 0.56 kW
Device type	2,645 ExpressStat 8,260 Utilipro 77,789 Gen 1 Switch 26,082 Gen 2 Switch	Gen 2 switches save more than any other devices, and switches seem to perform better than PCTs. Evidence from pre-2016 evaluations suggested that ExpressStat device had stopped working and that these households contributed no impacts. Our results support this finding.	ExpressStat 0.00 kW Utilipro 0.26 kW Gen 1 Switch 0.56 kW Gen 2 Switch 0.68 kW
Marketing strategy	91,235 Old 26,427 New	Impacts for customers recruited using the *new* method appear to be larger on both an absolute and relative basis	Old 0.51 kW New 0.67 kW
Multiple devices	11,652 Customers	Households with multiple devices are estimated to save less energy per device than those with only one device. These households have been omitted from full population ex post results in the past (2015 and earlier) because the can confound the results by being in treatment and control group at same time (2016). Multi-device customers were omitted in 2017.	Multi-device 0.36 kW Single device 0.55 kW (Either 0.49 kW)
Net metered	20,448	No differences. Load reductions are similar, but post event load growth and snapback were found to be higher than non-net metered (in 2016). Net metered customers have a very different load shape, commonly with negative net loads during sunny hours of the day, contributing to "duck curve."	NEM 0.55 kW Not NEM 0.55 kW
Dually enrolled (SmartRate™)	27,023	Dually-enrolled customers typically have lower impacts that SmartAC™ only customers because they use less energy (i.e., lower reference loads). This is likely a combination of the programs attracting people already interested in managing their energy consumption and feedback effects from experience in the programs.	Dual 0.34 kW SmartAC™ 0.56 kW

Ex post for sub-LAP events

Beginning in 2018, more events will be triggered by sub-LAP because the SmartAC™ program will be bid into the CAISO energy markets as a Proxy Demand Resource, with the requirement that load reductions must be bid in at the sub-Load Aggregation Point (sub-LAP) level. As such, it has been important for PG&E to understand the aggregate load reductions that SmartAC™ can provide for each of its sub-LAPs.

The average impacts per customer for each sub-LAP are shown in Table 12 below. Note that the results shown below are based off of a matched control group. As described in our methods sections, our approach matches potential controls with event participants based on their temperature sensitivity during a set of comparison days. Additional event data, spanning a wider range of conditions and further diagnostic analyses will be needed to gain a more thorough understanding of which "flavor" of control matching is able to accurately predict results in the context of SmartAC™.

In general, sub-LAP event dispatch is known to suffer from lower response rates than serial group dispatch, due to poor paging and subsequent inconsistencies in control devices' setup and configuration. Load impacts average 0.36 kW (13.4% of reference loads) but range from 0.15-0.18 kW (6.4 to 8.0% of reference loads) in the South Bay and North Coast sub-LAPs to 0.70 to 0.74 (21.3 to 22.0% of reference loads) in the Kern sub-LAP.

Note that the tables and figures below depict 2017 sub-LAP events, which amounted to only one to two per listed sub-LAP. Some sub-LAPs did not experience any events. Program year 2018 CAISO bidding by sub-LAP should be seen as an opportunity to build up experience with these events.

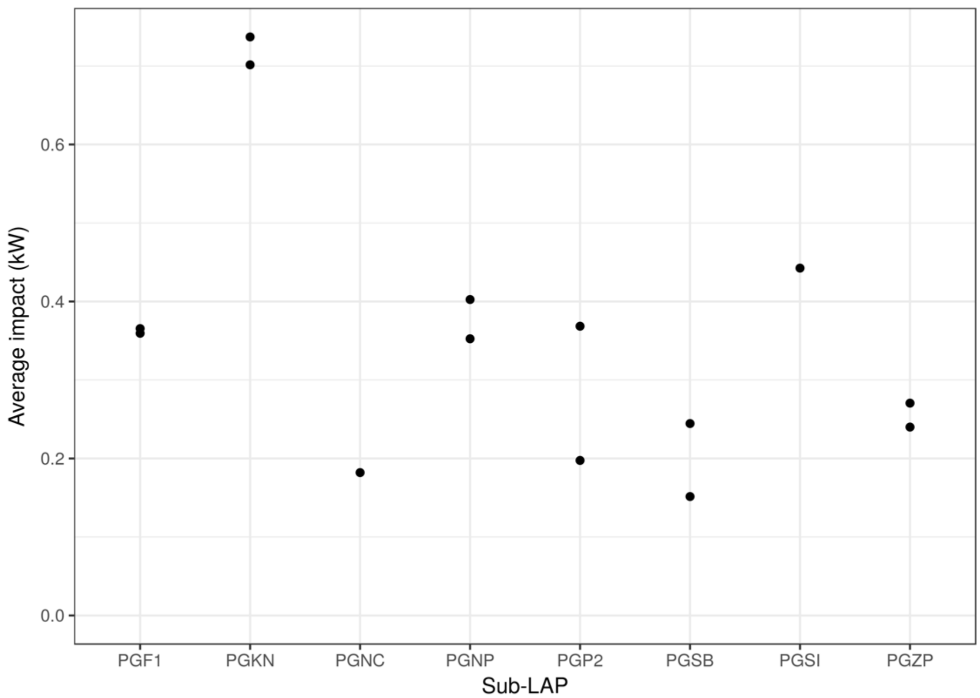
Table 12. Average ex post load impact estimates for 2017 sub-LAP events by sub-LAP

Sub-LAP	Date	Smart Day	# called	# controls	Ref. load (kW)	Average impact (kW)	% impact	Aggregate impact (MW)	Average temperature (F)
PGF1	07/06	TRUE	16,976	28,469	3.23	0.37	11.34	6.20	101
PGF1	07/28	FALSE	20,819	34,029	3.20	0.36	11.24	7.48	103
PGKN	07/28	FALSE	5,984	9,286	3.30	0.70	21.29	4.20	103
PGKN	07/31	TRUE	4,749	7,638	3.34	0.74	22.06	3.50	102
PGNC	07/31	TRUE	752	1,233	2.84	0.18	6.38	0.14	91
PGNP	07/06	TRUE	12,822	20,374	2.89	0.40	13.93	5.16	99
PGNP	07/28	FALSE	16,756	25,434	2.72	0.35	12.96	5.91	97
PGP2	09/11	FALSE	3,355	5,375	2.28	0.37	16.13	1.24	82
PGP2	10/24	FALSE	3,338	5,352	1.41	0.20	14.11	0.66	87
PGSB	09/11	FALSE	7,927	12,940	1.93	0.24	12.65	1.94	83
PGSB	10/24	FALSE	7,880	12,870	1.23	0.15	12.37	1.19	87
PGSI	07/31	TRUE	12,986	22,654	3.08	0.44	14.39	5.75	97
PGZP	07/06	TRUE	1,741	3,041	3.00	0.24	8.01	0.42	101
PGZP	07/28	FALSE	2,294	3,898	2.67	0.27	10.16	0.62	93
Mean			**8,456**	**13,757**	**2.65**	**0.36**	**13.36**	**3.17**	**95**

Note: a map detailing the location and common names for these sub-LAPs can be found in Appendix C: CAISO Sub-LAPs for PG&E Service Territory.

Figure 18 visualizes the sub-LAP event outcomes relative to one another. Here the critical role of local weather and other conditions becomes clear. The load reductions from the Kern sub-LAP are dramatically higher than others, while customers from sub-LAPs in the North Coast (NC), Peninsula (SP), South Bay (SB), and San Francisco (ZP) return lower than average results.

Figure 18: Event load impact by sub-LAP for all events called

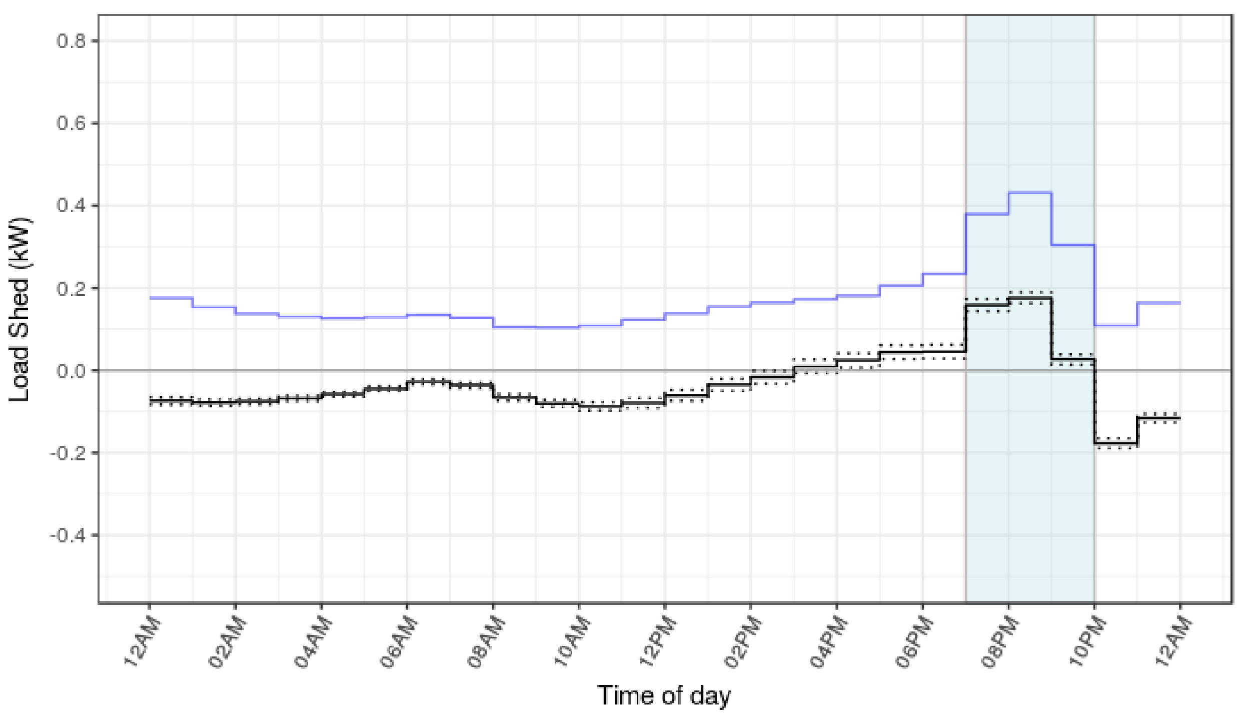


Emergency event on May 3

The emergency event, called on May 3, was technically invoked using serial groups - all of them were called at once. However, it was not a particularly hot day and may have preceded the changeover to running AC in the summer for many households. Load impacts, estimated by treating the event as simultaneous sub-LAP events in all sub-LAPs and combining them into a single ensemble estimate, are visualized in Figure 19. Impacts were estimated at 0.12 kW (7% of reference loads) for the entire event but were closer to 0.17 kW (9.5%) for the first two hours. This result underscores the importance of outside temperature in determining the aggregate resource available via SmartAC™.

This event also presented technical challenges due to the lack of resemblance between the event day and the comparison days used for the difference in differences estimate(s), which were selected as the hottest available non-event days and were not comparable to a fairly normal day in early May. These issues manifest via non-zero pre-event load impacts, but the sharp increase in impact at the from near zero at the start of the event window provides some confidence that the impact estimates are about right.

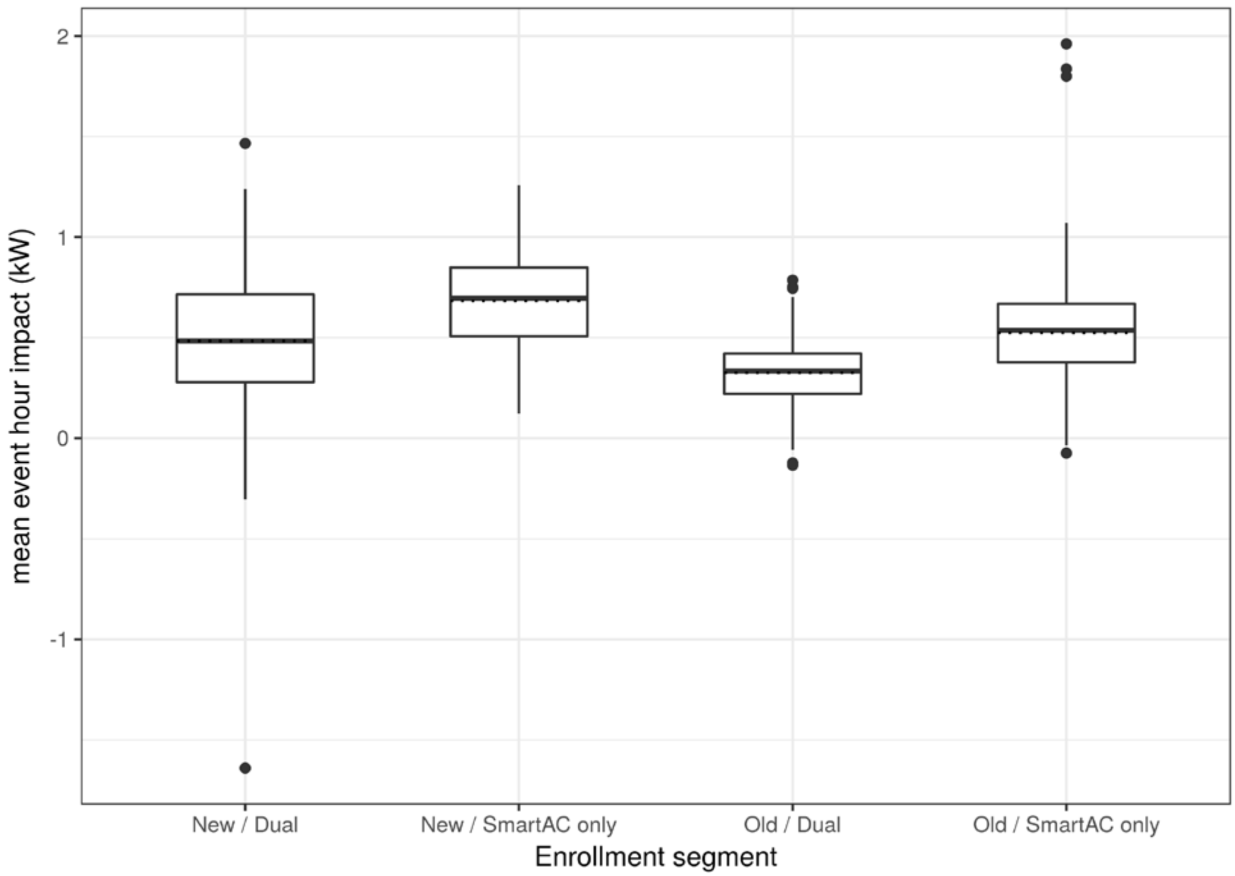
Figure 19: Emergency event estimated per-customer impact



Ex post results for ex ante inputs

Figure 20 depicts ex post results for the segments that are used in our ex ante analysis (next section). As shown in the figure (below), the new marketing segment that participates only in the SmartAC™ program has the highest impacts, with the old marketing segment in dual programs contributing the lowest impacts. These findings are consistent with past ex post results.

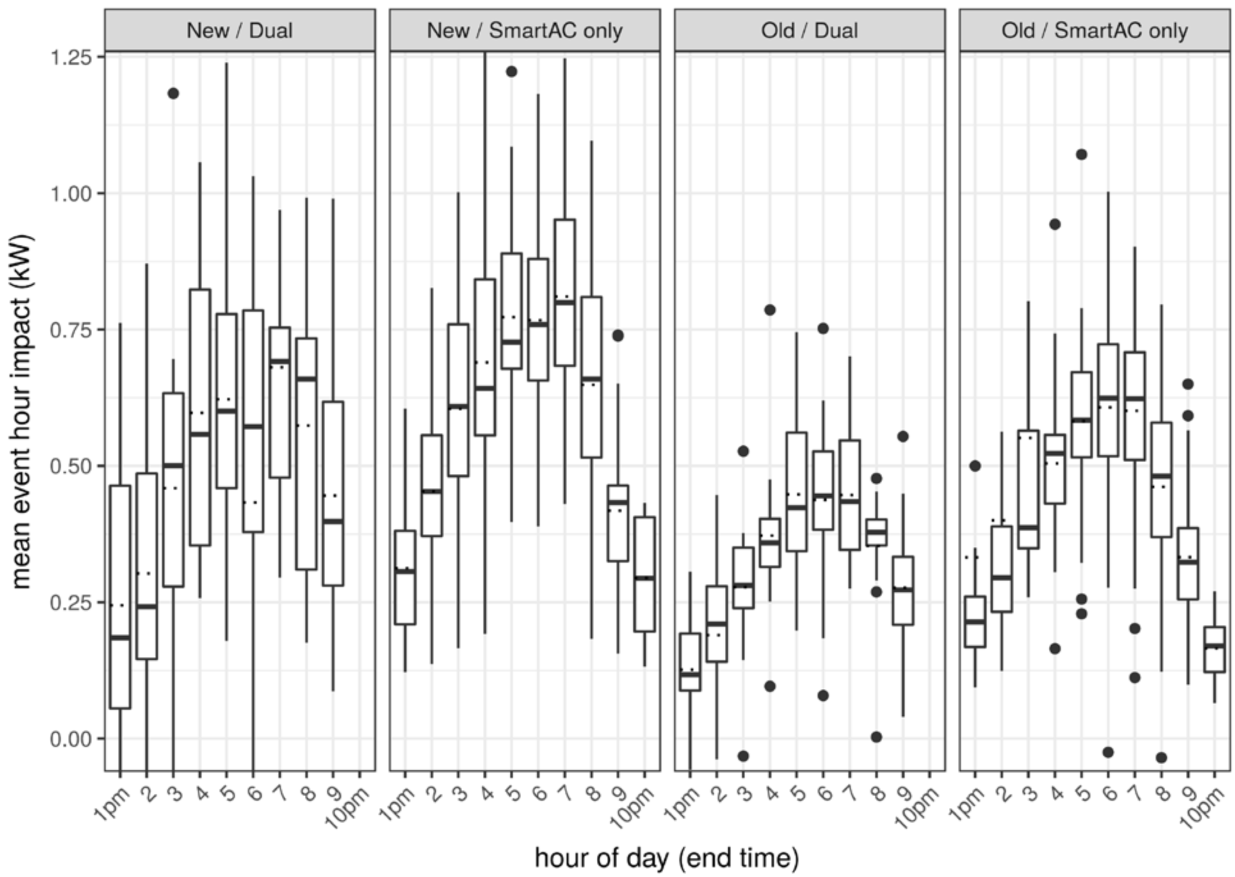
Figure 20: Event impacts for enrollment segments used in ex ante



Note: See the note for Figure 11 for a full description of box and whisker plots

We've already seen that program impacts systematically vary by hour of day, but these hourly patterns across enrollment segments are significantly different from one another (See Figure 21). The new marketing segment that participates only in the SmartAC™ program has the highest impacts, with the old marketing segment in dual programs contributing the lowest impacts, but the relative shapes of the hourly effects are a bit different as well.

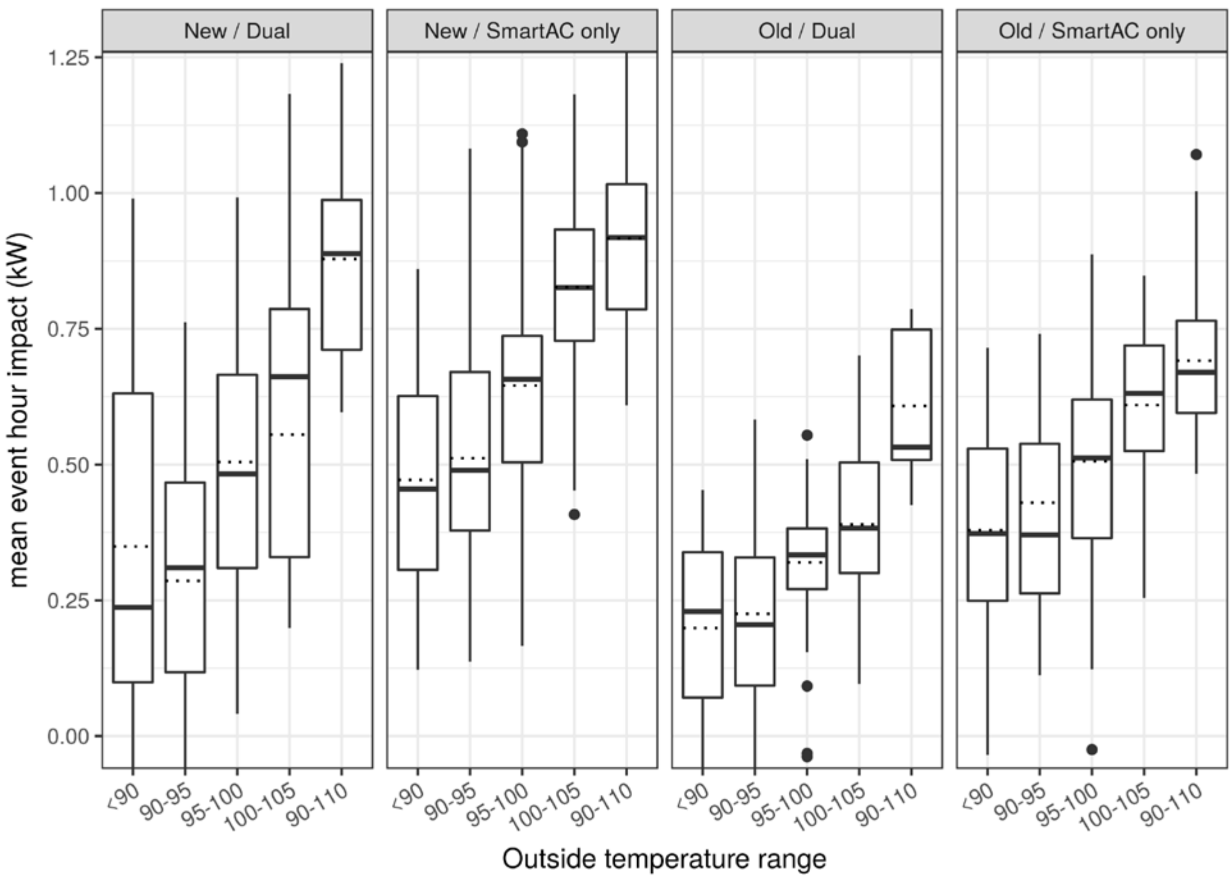
Figure 21: Event impacts for ex ante segments by hour of the day



Note: See the note for Figure 11 for a full description of box and whisker plots

In Figure 22, which shows the patterns of temperature response separately for each enrollment segment, we see that the primary determinant of response is SmartAC™-only vs. Dual enrollment. This provides more evidence for a structural difference in consumption for Duals. Their cooling loads are systematically lower than their SmartAC™-only peers. This is something that would be expected of customers who are smaller or more efficient. Both self-selection and time of use effects would tend to be correlated with efficiency improvements, so that is the current explanation for this effect.

Figure 22: Event impacts for ex ante segments by outside temperature



Note: See the note for Figure 11 for a full description of box and whisker plots

This section was meant to provide an intuitive understanding of the enrollment segments that help improve the predictive power of the ex ante model(s). The performance of these four segments, and the comparison of these segments to the assumptions made for 2017 and beyond, are quantified and discussed further in the ex ante load impact forecasts section that follows.

Ex ante load impact forecasts

This section provides ex ante forecasting results for the SmartAC™ program. Ex ante load impact estimates represent the expected per-customer average and system-wide aggregate load impacts that would occur during a SmartAC™ event under normal (1-in-2 year) and extreme (1-in-10 year) weather conditions if all customers were called simultaneously. Ex ante results serve two purposes: 1) they assist PG&E and the State with long-term resource planning, and 2) they allow PG&E to assess year-to-year changes in the program's effectiveness.

This section first presents mean and aggregate ex ante results, then compares those results to findings from previous years. The results are presented for PG&E and CAISO monthly system peaks over the period from May to October. CAISO, which spans the whole state, has different peak timing from PG&E because of the contribution of southern CA loads. in PG&E's service territory, CAISO peaks tend to happen on slightly cooler days than service territory wide peaks.

Detailed tables of results are in the summary section at the end of the chapter.

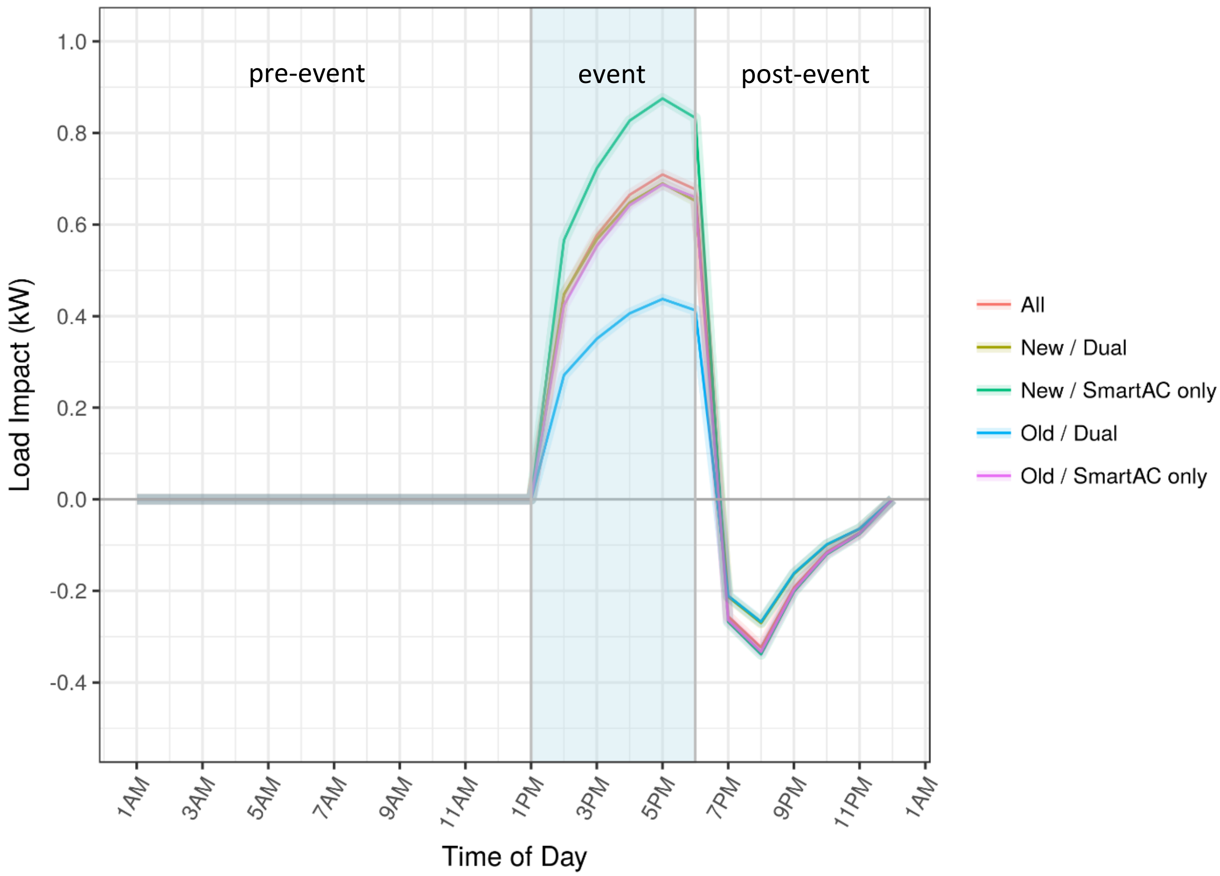
A companion ex ante "table generator" spreadsheet, which is an annex to this report, provides hourly load shed predictions for all of the standardized weather conditions, for May through October, using projected enrollment rates for the next ten years. That granularity of data is not reproduced in this report.

The ex ante model

The ex ante model is described at length in the methods section of this report. One of the key features is that it is fit to the observed outcomes of four customer segments: 1) Old / SmartAC™ only, 2) New / SmartAC™ only, 3) Old / Dual, and 4) New / Dual, where "Old" and "New" refer to the marketing strategy, updated in the fall of 2014, used to recruit new customers and "SmartAC™ only" and "Dual" refer to whether customers are simultaneously enrolled in the SmartRate program, which takes precedence over SmartRate™ control dispatch. The model is then used to predict the mean load shed for customers in each of these four segments in the ex ante weather conditions. The mean load for all customers is the enrollment-count-weighted mean of the mean of each segment. An example of the hourly predictions for each segment, for the weather conditions corresponding to the PG&E 1-in-2 day in August, is shown in Figure 23.

Like prior evaluators, to make the ex ante forecasts we assume that dual-enrolled customers are dispatched in a manner that contributes to SmartAC™ event load reduction. In practice, dual-enrolled customers are dispatched on a different schedule than SmartAC™ customers on SmartRate™ days and most SmartAC™ event days are also SmartRate™ event days. As described earlier in the report, the impacts of dual-enrolled customers on SmartRate™ days are evaluated under the SmartRate™ program and are not quantified in our ex post SmartAC™ analysis.

Figure 23: Predicted load shed by customer segment on the August Peak PG&E 1-in-2 day

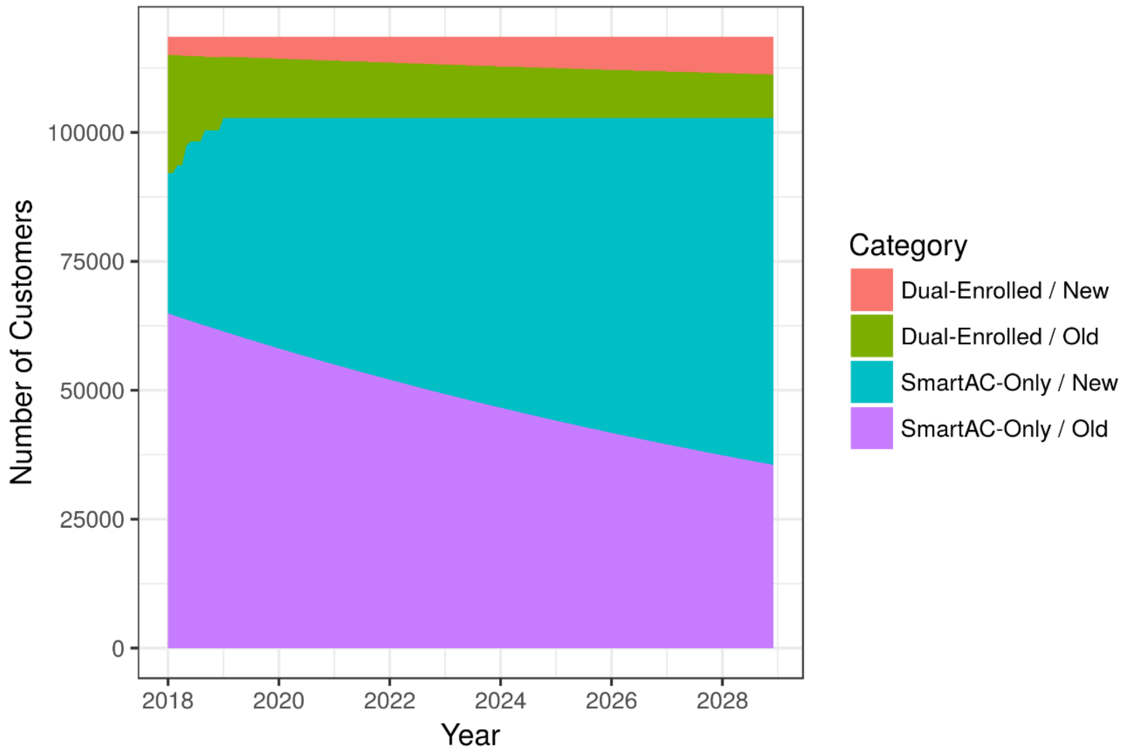


There are three logical periods to a typical SmartAC™ event. During the pre-event hours (12am to 1pm), it is assumed that there are no event impacts[5] and shed is set to zero; we ignore the initial ramp-up that occurs due to the signaling of the event occurring 30 minutes prior to the official event start time. A statistical model is fit to ex post impact data from the event hours (1pm to 6pm) and used to predict the ex ante load during those hours using forecasted weather and enrollment data. A separate model is used to predict the "snapback" after the event, when additional cooling energy is required to return the homes to their usual thermostat set points after the period of AC curtailment. In the ex ante discussions that follow, we focus only on the event hours.

Over the 10-year forecast, PG&E assumes that participation levels are constant at 118,000 participants, but that the makeup of these participants changes over time. The largest projected change is that older SmartAC™ only customers will be replaced by newer SmartAC™ only customers, which have been shown to have more load impact. The dually enrolled customers are also replaced through attrition. (See Figure 24.)

[5] It is possible that customers could pre-cool in anticipation of an event, but we do not see this in practice and there is not explicit outreach notifying customers of upcoming events.

Figure 24. Projected enrollment by customer segment



PG&E provided the yearly projected total enrollment of SmartAC™-only and dual-enrolled customers through 2028. PG&E also provided the number of old and new customers in each category in Fall 2017 and changes between old and new were predicted by assuming that old customers are lost through attrition and replaced by new customers. Due to the de-enrollment of low-AC-consumption customers and reduced recruitment, both of which were unusual occurrences, it was not possible to accurately estimate future attrition rates using 2017 data. The assumed attrition rate is 5.5% per year for SmartAC™-only customers and 3.5% per year for dual-enrolled customers. These are numbers chosen to be consistent with observed attrition documented in past reports.

The enrollment projections for dual-enrolled customers are fairly uncertain due to changes in both the energy markets and PG&E's marketing efforts. Customers who switch to new plans under Community Choice Aggregation (CCA) will no longer be in the SmartRate™ program, and PG&E is not trying to recruit more SmartRate™ customers into the SmartAC™ program. The combination of these factors may lead to declining numbers of dual-enrolled customers, in contrast to the current projection that assumes this number will be steady after 2019.

Ex Ante Load Reduction

Figure 25 shows both the mean impact per customer (axis on left) and the aggregate impact (axis on right) for PG&E and CAISO (1-2 and 1-10 weather years; monthly peak weather), for the population of customers assumed to be enrolled in the summer of 2018. The weather during the CAISO peak is different from the PG&E peak, producing differently shaped (and generally lower) predicted load shed than those for PG&E. The aggregate impact calculation assumes all customers participate in the event: it is equal to the mean impact per customer (values from the axis on the left side of the graphic) times the count of approximately 118,000 customers forecast for program year 2018.

In aggregate, ex ante estimates for the aggregate load shed (averaged over the five-hour event window) reach as high as 79 MW in July of PG&E's 1-in-10 weather year, as shown in the bottom left figure below. The minimum for any of the weather years is 16 MW in October of PG&E's 1-in-2 weather year (top left).

Figure 25: Mean and aggregate ex ante impacts

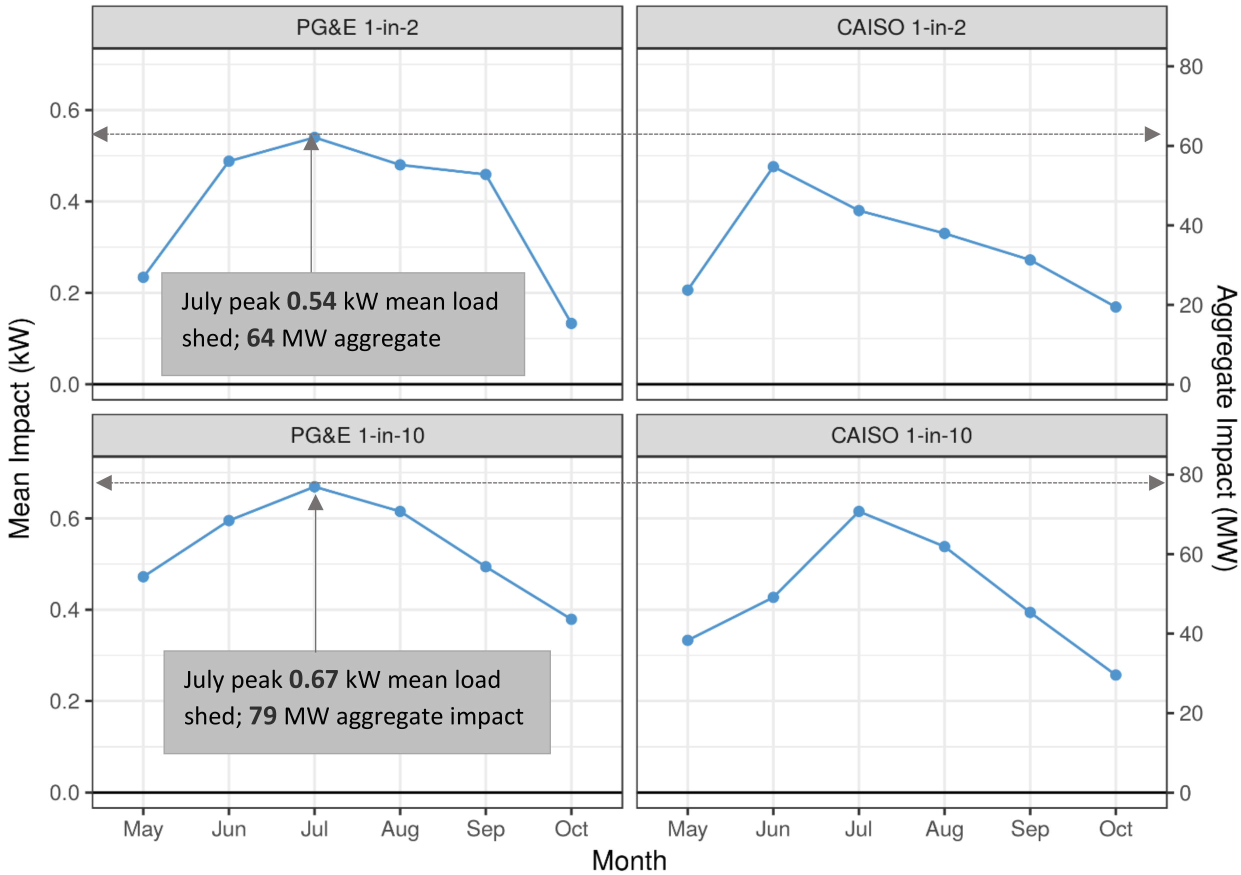


Figure note: The left and right axes apply to all four figures. Mean load shed is indicated on the left, and aggregate on the right, as demonstrated by the July examples in PG&E 1-in-2 and 1-in-10 and the corresponding arrows pointing left and right.

Overall, in each of the scenarios May and October have much cooler weather than the other months and load impacts are strongly correlated with outside temperatures at the time of the event. As a result, the "shoulder months" have significantly lower projected load shed than mid-summer.

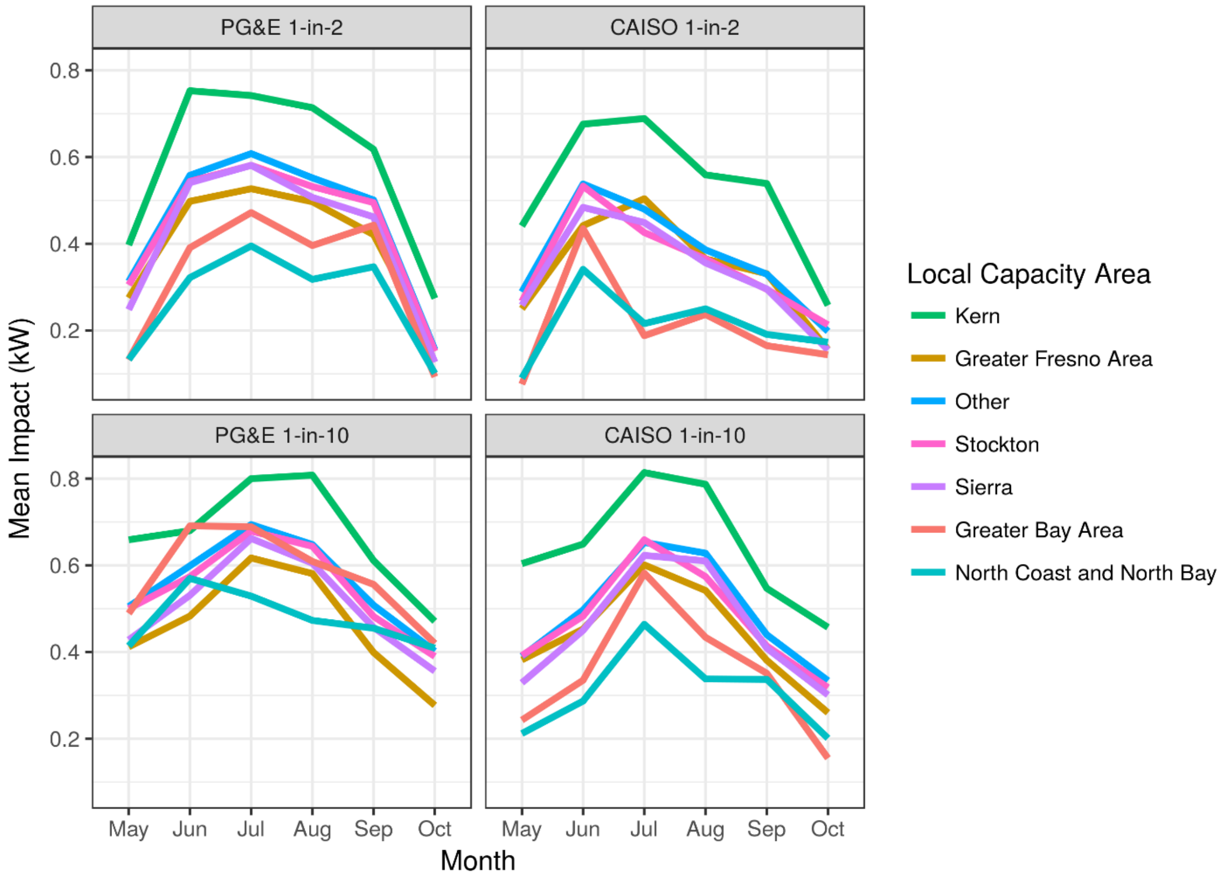
To assist with long-term resource planning, the detailed ex ante data for each month and scenario are shown in tabular form at the end of this ex ante discussion (see Table 15 and Table 16).

Ex ante load reduction by LCA

The per-customer ex ante impacts vary across the local capacity areas. Figure 26 shows mean ex ante load shed by LCA (excluding Humboldt) for each of the weather years. Almost all of the difference between regions is due to differences in weather: a 1-in-10 year is much hotter in Kern than in the North Coast and North Bay. The exceptions are the Greater Bay Area, whose average load shed per degree is higher than the other regions, and the Greater Fresno Area where it is lower. There are also small differences due to differing ratios of new- and old-marketing customers in each LCA. Even though the

Bay Area has higher load shed per degree than other regions, its relatively mild weather makes it a fairly low contributor in terms of mean load shed per customer. Conversely, the Greater Fresno Area experiences relatively hot weather but has only moderate impact per customer due to its low load shed per degree; part of this relatively poor performance is due to problems communicating with the SmartAC™ devices via the pager network in that area.

Figure 26: Mean impact per customer by local capacity area

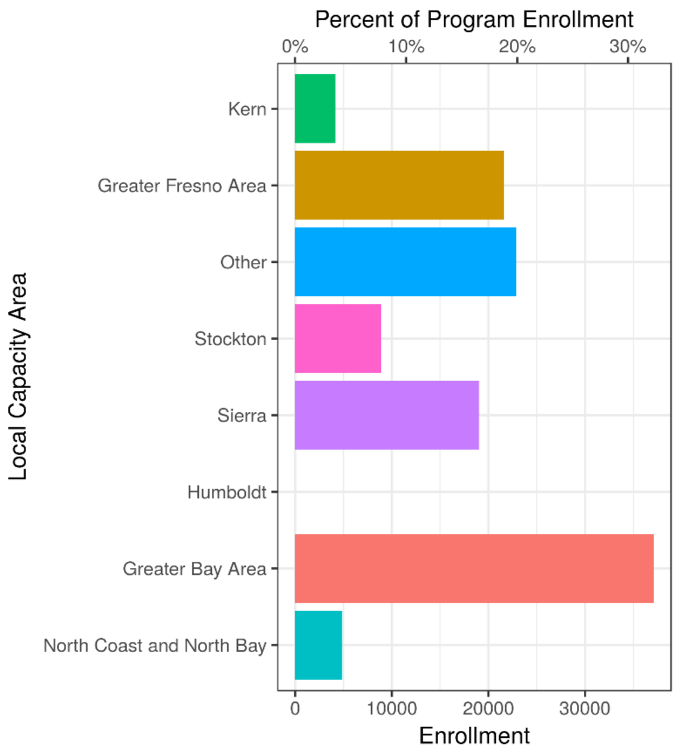


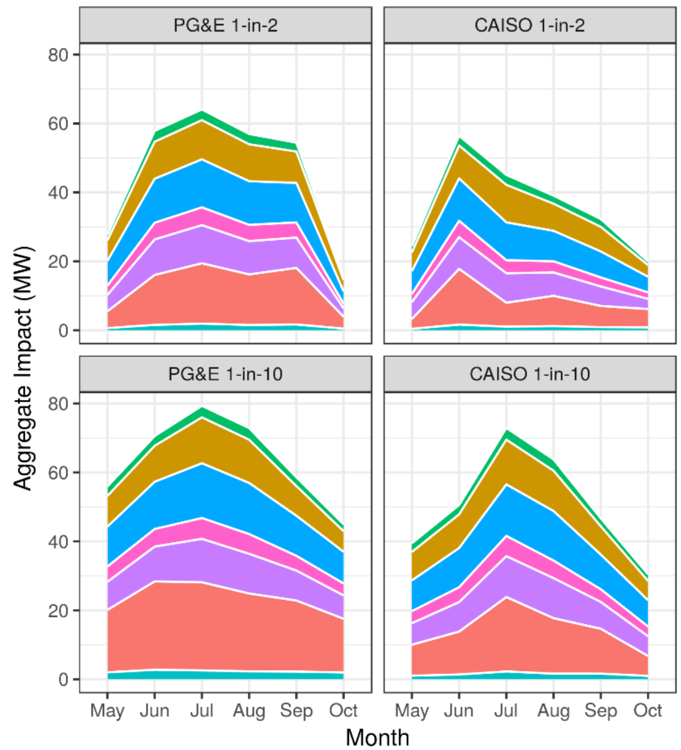
In the figure above, Kern has the highest load shed per customer while 'North Coast and North Bay' generally has the lowest. The order is the same in most months of most weather years except for the Greater Bay Area: the temperature slope is higher in that LCA than the others, so even though it has low mean load shed per customer in the 1-in-2 scenarios, it is above the middle of the pack in the hotter PG&E 1-in-10 weather year.

Like the mean load shed per customer, the aggregate impacts also vary greatly between local capacity areas, due in part to the difference in weather (discussed above) but even more because of the differences in enrollment among LCAs, as shown in Figure 27a, which plots the assumed ex ante enrollment by LCA for 2018; these numbers are close to the actual enrollment at the end of 2017. The bars are in order from highest to lowest mean load shed per customer in July of the PG&E 1-in-2 year (see Figure 26 above).

Figure 27b shows the aggregate load shed by local capacity area; this is the multiple of the load shed per customer (Figure 26) times the number of customers (Figure 27a) . Although Kern and Stockton are high in terms of load shed per customer because of the weather they experience, modest numbers of participants in those LCAs leads to low aggregate load shed, as shown in the right-hand panel. This is especially true of Humboldt, which has only two SmartAC™ customers! North Coast and North Bay contributes little to the program due to both poor load shed per customer (due to mild weather) and a low number of customers. The situation with the Greater Bay Area is more complicated due to its higher weather sensitivity: in the 1-in-2 weather years it contributes less than 33% of the aggregate load shed despite having 33% of the enrollment, but in the PG&E 1-in-10 weather year the Greater Bay Area contributes approximately in proportion to its enrollment. The Greater Fresno Area underperforms at all temperatures.

Figure 27a and b: (a) LCA enrollment sorted from top to bottom by mean PG&E 1-in-2 July load shed and (b) Aggregate load shed by LCA

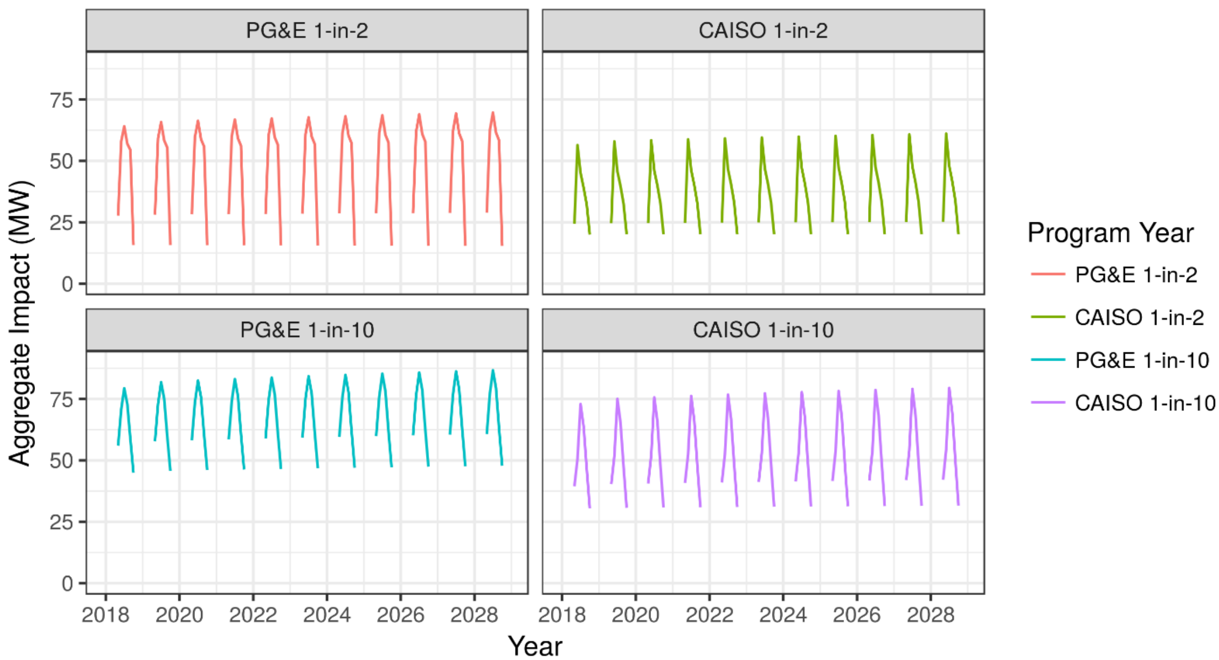




Ex Ante Impacts Over Time

All of the figures above apply to the ex ante predictions for 2018. Predictions for *future years* are slightly different from the predictions for 2018 due to projected changes in the mix of customers, specifically the gradual transition of old customers to new ones through attrition and recruitment. Since new customers shed more load than old customers, on average, this transition is projected to lead to a gradual increase in the impact capacity of the SmartAC™ program as shown in Figure 28. The changes are gradual and of modest impact, with modest increases in impact for each sequential program year.

Figure 28: Ex ante aggregate load shed in future years



Relationship between ex post and ex ante aggregate impacts

Ex post and ex ante aggregate load impacts differ for a variety of reasons. Table 13 summarizes the factors that influence the relationship between ex post observations and ex ante predictions.

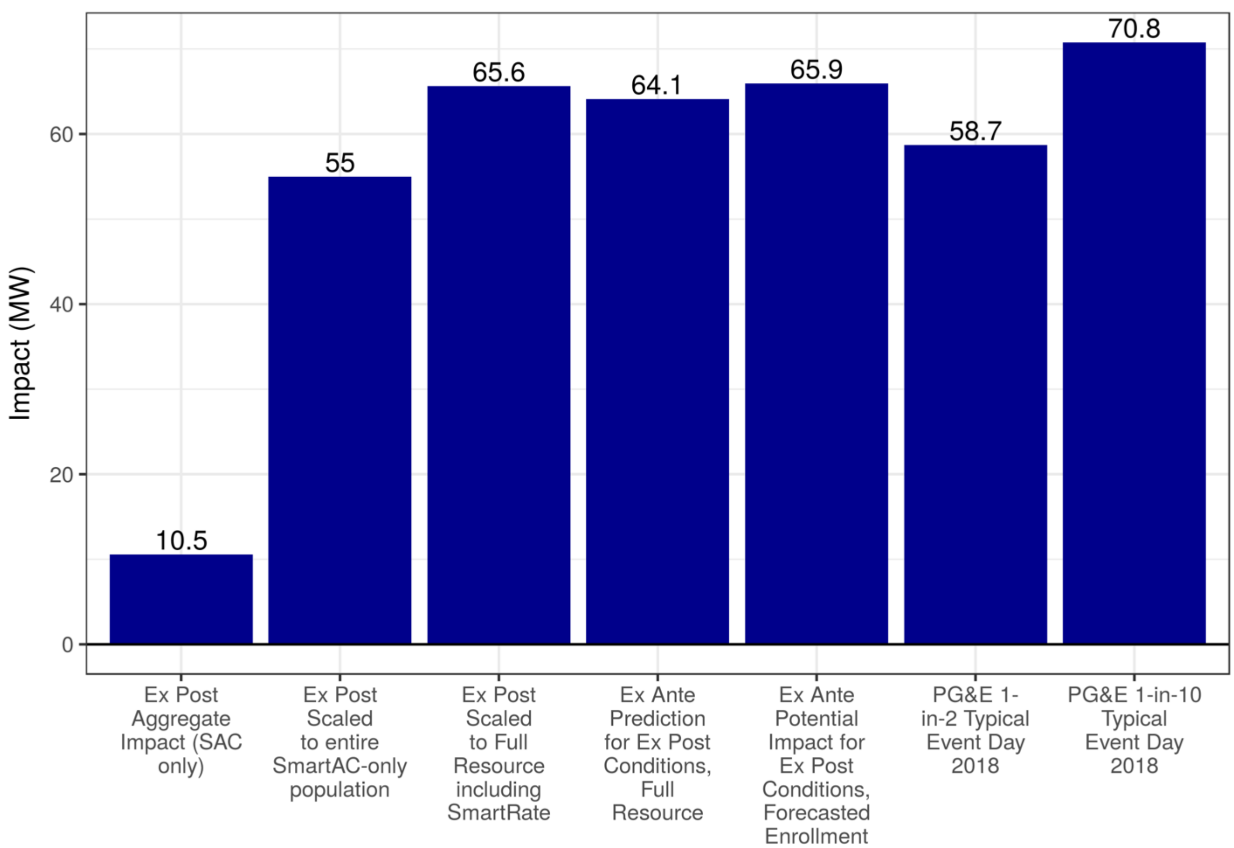
Table 13: Important factors that relate ex post observations to ex ante predictions

Factor	Ex Post	Ex Ante	Magnitude of Impact of Assumptions
% of resource dispatched	10-20% of the program is typically dispatched for each event.	Assumes 100% dispatch.	Biggest impact of all factors
Enrollment	Actual enrollment	Projected enrollment	Potentially large in the later years of the ex ante predictions. The aggregate impact is equal to the mean impact per customer multiplied by the number of customers. The further into the future the enrollment is projected, the more uncertain it is.
Dual-Enrolled in SmartAC™ and SmartRate™	Events are called on both SmartRate™ days and non-SmartRate™ days. Dual customers don't share SmartAC™ dispatch timing on SmartRate™ days.	Assumes all customers are dispatched with the desired event timing.	Medium. Most hot days are SmartRate™ days. Dual-enrolled customers do not have ex post load reduction estimates for SmartAC™ events that occur on SmartRate™ days, so their contribution must be estimated using limited data, but they are in the minority. See Figure 29.
Event window	Typically one or two hours for any single customer.	Uniform ex ante event window is 5 hours, between 1 and 6 pm.	Small. The ex ante statistical model adjusts for the hours of day of the event, although this adjustment is not perfect.
Weather	Variety of weather conditions on different days in different local capacity areas.	Standardized weather conditions. PG&E 1-in-10 days are extremely hot; CAISO 1-in-2 days are much cooler.	Small for most months and scenarios of interest. 2017 event temperatures were not far from PG&E 1-in-2 weather in June, July, and August. Low-temperature months and scenarios are subject to extrapolation error (e.g. CAISO 1-in-2 May and October).
Statistical methods	Impacts are based on the outcome of randomized controlled trials (RCT) with large treatment and control groups.	A statistical model based on the ex post results is used to extrapolate to ex ante weather conditions.	Small for most of the weather scenarios in most months. Some error is introduced when extrapolating to lower or higher temperatures than were experienced in the observed events.

Figure 29 shows the effect of some of the factors discussed above. The blue bar at far left shows the average aggregate impact for the SmartAC™-only customers in the serial events in 2017, for the event-hours from 1 to 6 PM. Only a fraction of the customers participated in any one event, but that aggregate impact can be scaled up to predict what the impact would have been if all of the customers had participated (second bar). The third bar adds on the estimated effect of the dual-enrolled customers. The middle (fourth) bar shows the prediction from the ex ante model, applied to the weather conditions that were experienced during the events, however, the ex ante model was only trained on data from hours spanning 1pm through 6pm (thus only able to predict for those hours) while actual 2017 events included hours outside that range. If the statistical model were perfect and designed to span all event hours, the third and fourth bars would be the same. The next bar scales up the ex ante prediction to adjust for the number of customers expected in the summer 2018. The last two bars show the predictions for the PG&E 1-in-2 and 1-in-10 typical event day, for the number of customers expected in 2018. A much more extensive comparison can be found in Table 17 at the end of this section.

The scaled ex post impact (including SmartRate™) and the ex ante prediction for the conditions that were experienced in the 2017 events are nearly in agreement, differing by less than three percent: the model is accurate for the weather conditions it was trained on. Both the scaled ex post predictions and the ex ante predictions are higher than the PG&E 1-in-2 Typical Event Day because the weather during the 2017 events was more extreme than the 1-in-2 weather year, although not as extreme as 1-in-10 weather. For example, the mean temperature in the Greater Bay Area during the 2017 events was 95.9 F, more than 2.5 degrees higher than the 93.3 F temperature of the 1-in-2 Typical Event Day. The temperatures experienced during the 2017 events were one to two degrees higher than the 1-in-2 Typical Event Day in most of the other regions too.

Figure 29: Factors that relate ex post to ex ante, for the 1 – 6 PM event window

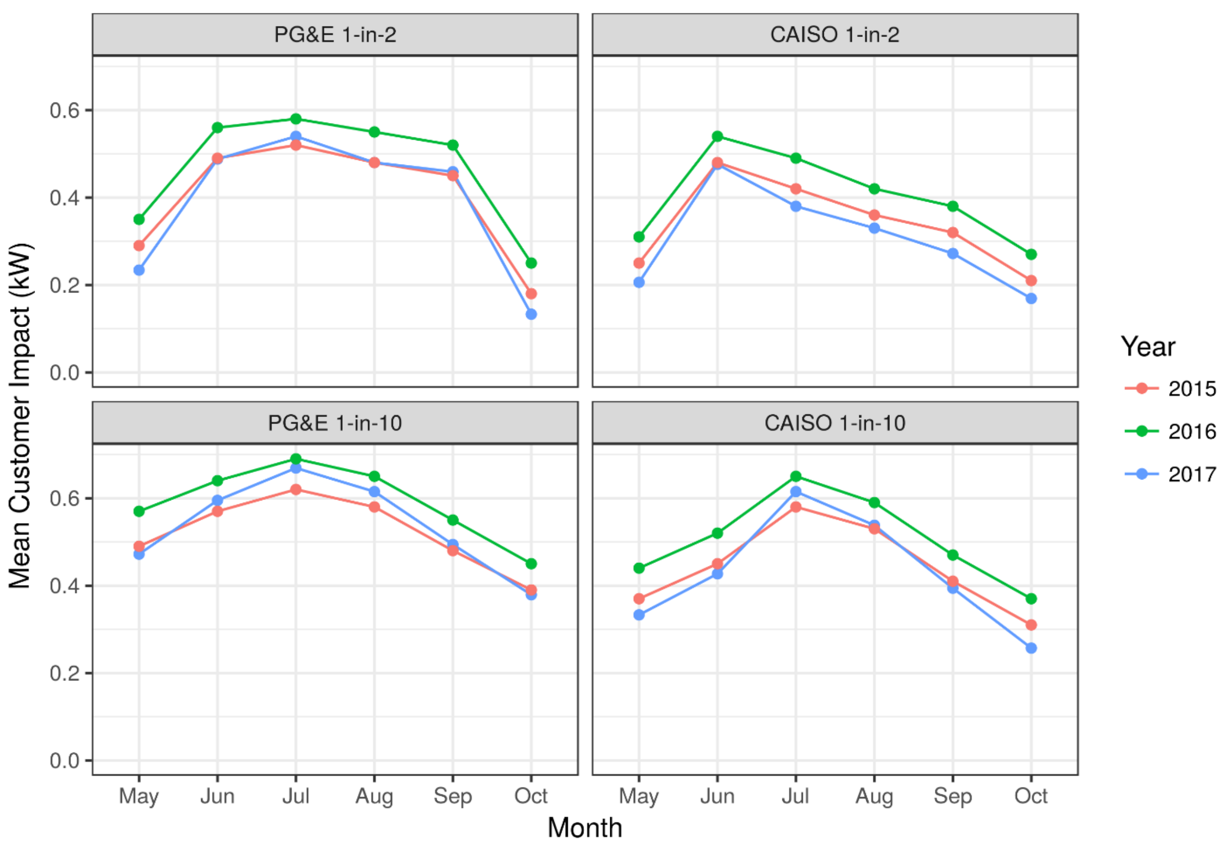


Comparison to previous reports

Figure 30 shows the predicted ex ante mean customer load shed for the four standard weather scenarios, for each month, from three SmartAC™ program evaluations (2015, 2016 and the current year). The blue line shows the predictions based on the current ex ante model; that is, the present report. Previous years are shown in red and green. As discussed in the 'Key differences between years' section, the ex ante mean impact results from the 2016 program evaluation, shown in green, are known to be overestimates, so the changes in the program should be assessed by comparing 2015 (red) to 2017 (blue).

As shown in Figure 30, the mean impact per customer is higher in the 2017 analysis than in the 2015 analysis in hot conditions – the PG&E 1-in-10 weather year and the hot months of the other weather years – and comparable or lower in cooler conditions. The maximum estimated difference in the standard weather scenarios is in July of the PG&E 1-in-10 weather year: the mean impact estimate for the current customers is about 8% higher than the ex ante estimate from the 2015 report. This increase in mean per-customer impact in hot months is primarily due to the active de-enrollment of relatively low-AC-consumption customers. And the increase in the impact is probably larger than 8% because the 2015 estimate was probably too high, for reasons discussed below in the section "Methods of forecasting per-customer loads".

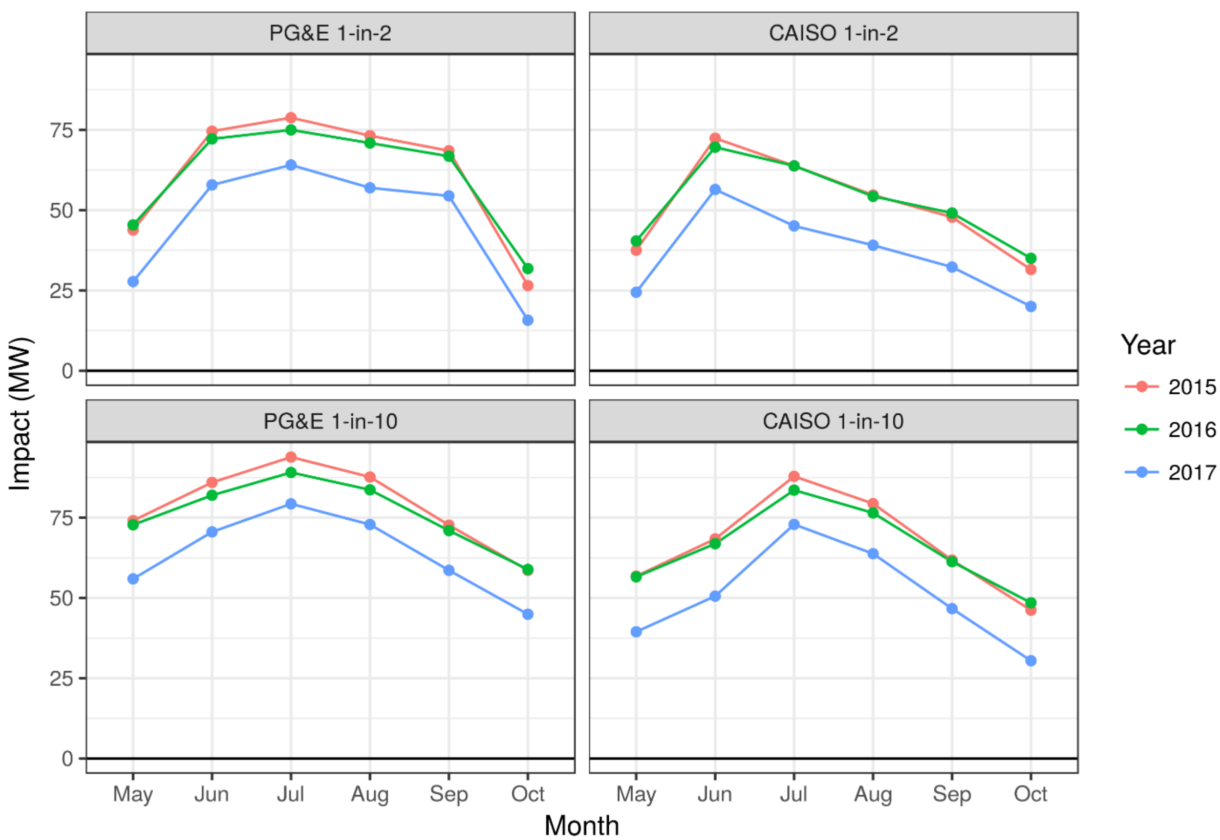
Figure 30. Average ex ante impacts: Comparison across program years 2015-2017



Note: Assess changes in program years by comparing 2015 to 2017: the 2016 results are known overestimates.

Figure 31 shows the predicted aggregate ex ante impact, which is simply the mean impact per customer multiplied by the number of customers. The aggregate ex ante impact is substantially lower now than it was in previous years. The explanation is as follows: Although the per-customer impact has gone up somewhat in the hot months, the number of customers in the SmartAC™ program has decreased by almost 25% compared to the end of 2015, from over 150,000 to about 117,600 customers. On average, the customers who remain in the program shed more load in hot weather than was true in past years, but this effect is not large enough to overcome the reduction in the total number of customers in the program.

Figure 31. Aggregate Ex Ante Impact Estimates: Comparison Across Report Years 2015-2017



.

The current ex ante forecast that is produced from the 2017 program review differs from past forecasts for two major reasons:

1. **Forecasted enrollment:** The projected participation numbers have dropped significantly since earlier reports, which affects the aggregate impacts from the program. At the end of the PY 2017 season, enrollment was down by nearly 36,000 customers relative to the end of PY 2016. A drop of 22,000 was due to the planned de-enrollment of low-AC-consumption customers, which was anticipated in the evaluation of the 2016 SmartAC™ program. The balance of

customers was lost through attrition and were not replaced by new enrollments as planned due to the cancellation of a program to enroll new customers by default; see the discussion of this issue in the section of the program overview titled “Removal of low-AC-consumption customers and net de-enrollment.”

2. **Method of forecasting per-customer loads.**
    - The statistical model builds on the average (or per participant) PY 2017 ex post results only; it does not include results from previous years as was done in the past. The rationale for this choice is that the makeup of participants has changed significantly over time given natural attrition, targeted marketing, and the deliberate de-enrollment of customers who had low load shed. Furthermore, the earliest program years’ ex-post estimates (i.e. 2012-2013) were significantly higher than more recent ones, so including these tends to bias the forecast upwards despite no recent attainment of similar load shed performance[6].
    - The statistical model quantifies the different load shed behavior of customer recruited using “Old” and “New” marketing and applies this difference when making the ex ante forecasts. This was not done in the analysis of the 2015 program because there were not yet enough new customers to provide accurate estimates, and it was not done in 2016 because it was recognized that the planned de-enrollment of old, low-AC-consumption customers would change the average behavior of the old customers who remained.
    - The statistical model quantifies and controls for the different load shed behavior of customers in the Greater Bay Area and in the Greater Fresno Area compared to other customers.
    - There are some additional minor differences between the 2017 model and the previous models; see Appendix B on the ex ante model for details.

Each of the modeling differences is discussed in more detail, below.

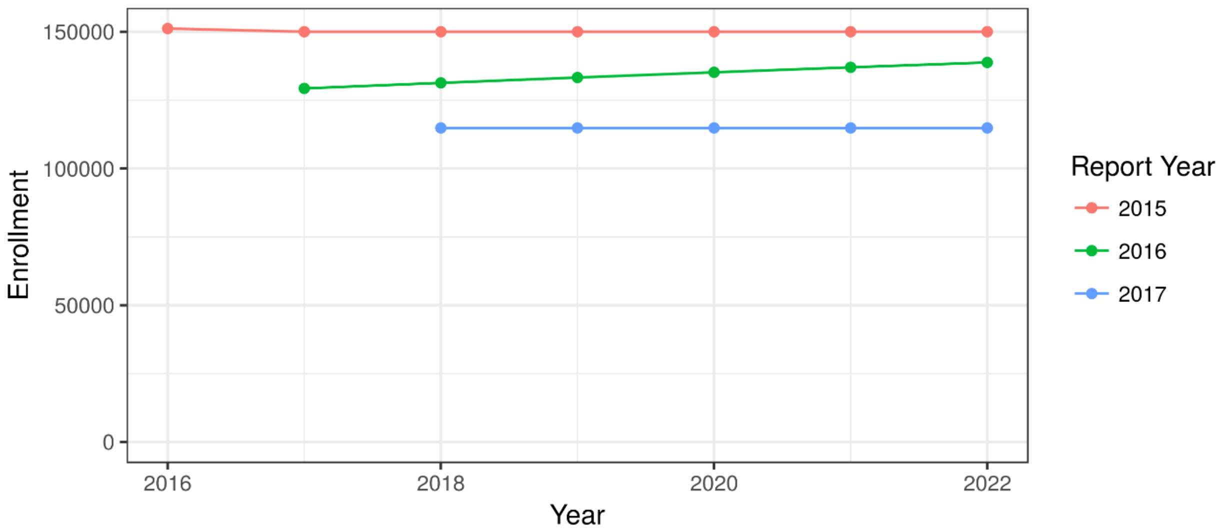
Forecasted enrollment

The number of program participants changes with time, so each year there is a different projection of the future enrollment in the program. These projections are provided by PG&E. Figure 32 shows the projected future enrollment that was used for the ex ante predictions for the PY 2015 and 2016 reports and for the present report. Numbers used in the 2016 report reflect the expected effect of the removal of the assumed low-load-shed customers that took place during early 2017. At the time the 2016 report was written, that change was expected to reduce the number of customers in the program to about 129,000 in 2017, with numbers expected to increase steadily due to recruitment of new customers. The recruitment strategy for PY 2017, known as “Automoves” was to preserve enrollment for customers that move and to automatically enroll customers moving into a home with a SmartAC™ enabled device. That strategy was cut short after customer pushback, and it was not immediately replaced by an alternative recruitment policy. In the absence of new recruitment, the total program enrollment hit 117,600 in Fall of 2017. The present report uses PG&E projections that assume that in future years the enrollment will remain near its current number, with a constant total enrollment of about 118,600.

---

[6] In fact, we suspect the ex post modeling from those early years may have been biased upwards by one assumption or another.

Figure 32: Projected future SmartAC™ enrollment by report year



Methods of forecasting per-customer loads

1. The present analysis uses a statistical model that was fit to data from 2017 only. Previous reports combined data across multiple years in order to reduce the statistical noise associated with having a small number of events. There is year-to-year variability in program performance, for a variety of factors including the changing mix of customers as people enroll or de-enroll, the failure of old switches and communication devices, the introduction of new switches and devices for new enrollees, different estimation methods and assumptions, and so on. **Per-customer impacts in 2012 and 2013 were higher than in subsequent years, so including those years when fitting the model led to models that over-predict load sheds when applied to the current program.**
2. The present analysis treats the Greater Bay Area and Greater Fresno Area differently from the others: all of the other Local Capacity Areas are assumed to have the same temperature slope – that is, the same increase in load shed per degree of outdoor air temperature – but the Greater Bay Area and Greater Fresno Area each have a different slope. **This was done because, empirically, a model that does not include this feature tends to substantially over-predict the load shed in the Greater Fresno Area and under-predict it in the Greater Bay Area.** Previous years fit a single slope for all of the areas. This modeling choice makes little difference when predicting the program-wide load shed for temperatures similar to those that were experienced during the events that were used to fit the model, but it does make a difference when extrapolating to higher or lower temperatures. Appendix B, on the ex ante model, includes a comparison of model predictions with and without special handling of the Greater Bay Area and Greater Fresno Area.
3. The present analysis fit a statistical model that estimates different temperature slopes for customers recruited using “Old” and “New” marketing methods. The 2015 analysis did not distinguish between these customer groups, largely because at the time the number of new customers was rather small. The 2016 analysis recognized that over 20,000 old customers were going to be de-enrolled due to poor performance (i.e. low load shed during events), so they

attempted to adjust for that change. The analysis assumed that the customers who were de-enrolled from the program in the spring of 2017 provided no load shed at all, so that removing these customers would not affect the aggregate load shed: *"This is because those customers being removed from the program are **assumed to have no impacts**, such that removing them **should not have an effect on aggregate impacts**."* 2016 report p.55, emphasis ours. Even with the information available at the time, this assumption was very aggressive – most customers with working hardware save *something* - and subsequent analysis demonstrated that the low-AC-consumption customers had load impacts of just over 20% of their fellow enrolled customers. The previous report further assumed that all of those customers were enrolled as SmartAC-only and had been recruited with the old marketing strategy. The assumption that the aggregate load shed would be unchanged led, mathematically[7], to the implication that each 'Old' SmartAC™-only customer who remains in the program – each remaining customer who was recruited under the old marketing strategy – sheds slightly *more* load, on average, than a 'New' customer. Empirically, this proved not to be the case: in the 2017 events, each old customer shed about 20% less load on average than a typical new customer.

The differences in the approach to estimating impact per customer are summarized in Table 14.

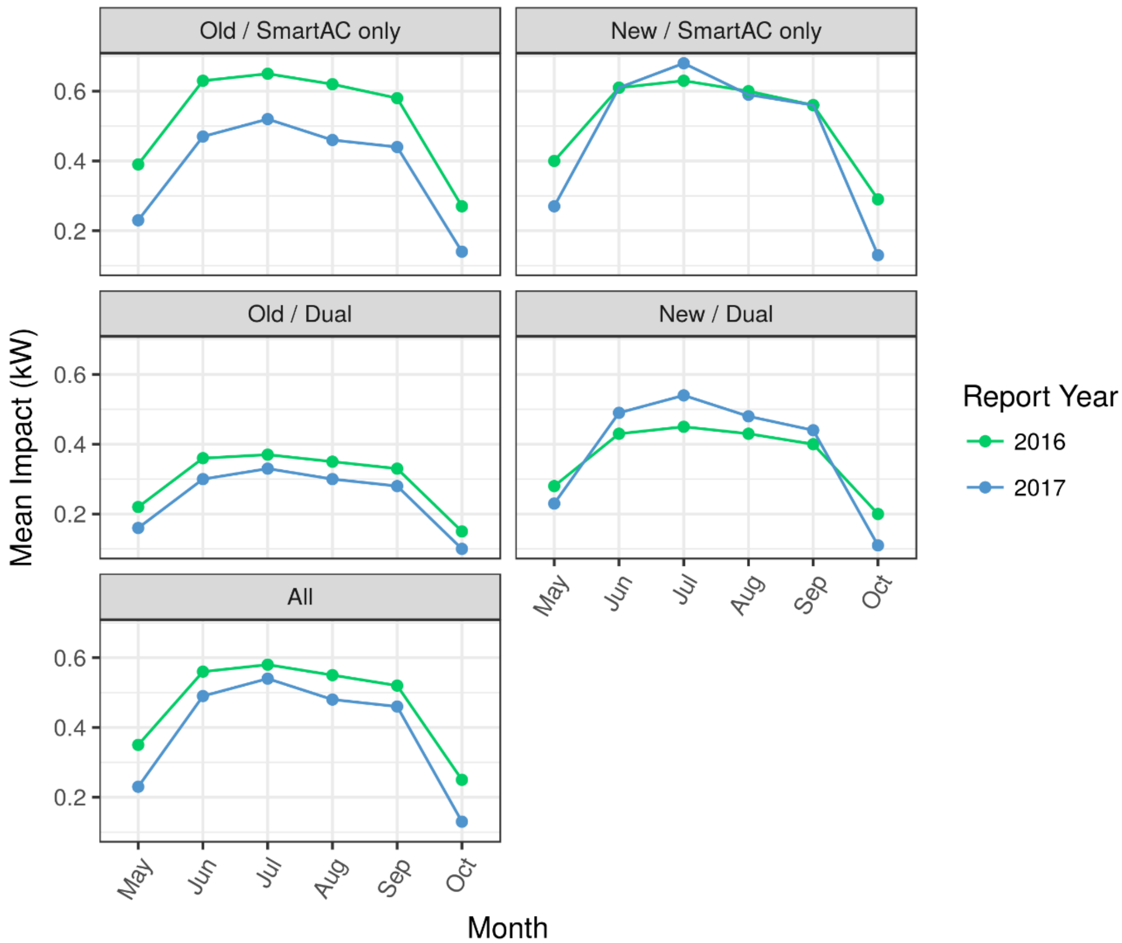
Table 14: Number of participants and method of forecasting: Comparison across report years

Report Year	Assumed number of ex ante SmartAC™ participants for the next program year	Method of forecasting the mean impact per customer
2015	150,000	Fit to data from 2011 through 2015. No distinction between old and new customers. All local capacity areas treated the same.
2016	130,000	Fit to data from 2012 through 2016. Ex post data on "New" customer impact available for the first time. Mean impact of "Old" customers was assumed to increase (due to de-enrollment of low-impact customers) so that the aggregate impact is the same with 130,000 customers as it was the previous year with 150,000 customers. All local capacity areas treated the same.
2017	118,600	Fit to 2017 data only. Separate slopes for "Old" and "New" customers Different temperature response allowed for Greater Bay Area, Greater Fresno Area, compared to all other LCAs.

[7] (Aggregate impact) = (mean impact of old customers) x (number of old customers) + (mean impact of new customers) x (number of new customers). Only old customers were being de-enrolled, so the mean impact of new customers and the number of new customers stayed the same. By assumption, the aggregate impact also stayed the same. For that to happen in spite of the drop in the number of old customers, the mean impact of old customers remaining in the program would have to increase. Given the magnitude of the decrease of the number of customers, the mean impact of the remaining old customers would have had to slightly exceed that of the new customers. See Table 5-2 of the 2016 report.

Ex ante forecasts for the PG&E 1-in-2 weather year are shown in Figure 33 for each customer segment and for all segments combined (weighted by the number of customers in each segment). As can be seen in the figure, the fit to the 2017 data predicts much less load shed from the 'Old SmartAC™-only' customers than was assumed in the 2016 report (upper left panel). Since about 55% of customers are in this customer segment, the result is that the ex ante mean load shed per customer for the entire program is lower in the 2017 analysis than in the 2016 analysis.

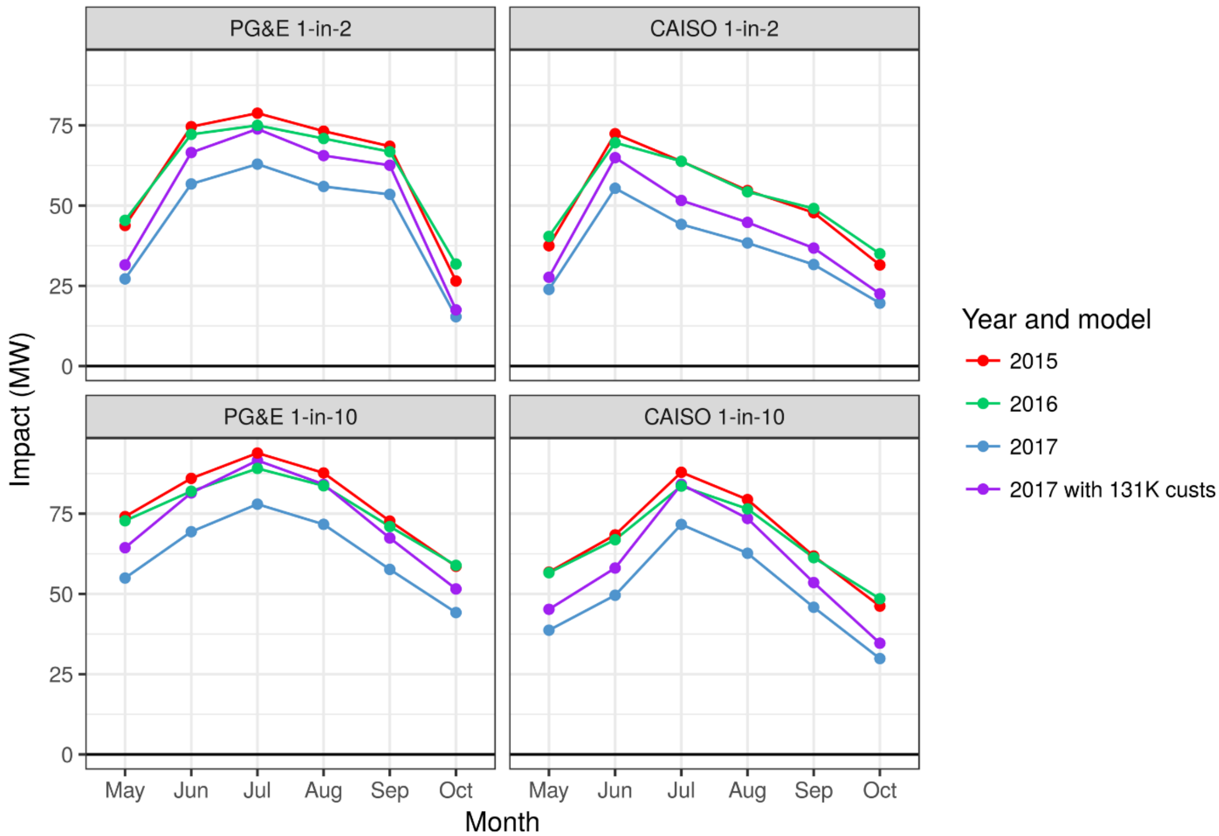
Figure 33: Compare ex ante mean impact by segment for PG&E 1-in-2 weather year



Key differences between years

Most of the change in the system-wide aggregate ex ante impacts is due to the changes in participation. Figure 34 shows what would happen to the 2017 ex ante predictions for 2018 if the number of customers in the program in 2018 was 131,000, as was anticipated in the 2016 report, rather than the 118,600 that is now assumed to be the enrollment in Summer 2018. (See the discussion in the Forecasted enrollment section for an explanation of why the target of 131,000 was not attained). The "missing" 13,000 customers were assumed to be in the "new SmartAC™-only" category, since that was who the program intended to recruit. As the plot shows, if that number of additional customers were in the program the discrepancy between the 2016 forecast and the 2017 forecast would have been much smaller in all of the hotter months. Over all of the scenarios, the median difference between the 2016 and 2017 ex ante predictions in June, July, and August is 30% and the mean is 35%; the discrepancy in the cooler months is even greater. If the number of new enrollments had met expectations, the discrepancy in the hot months would have had both a mean and median of 15**%. About half of the reason the potential load shed in the peak months is less than was predicted in 2016 is due to the lower-than-expected enrollment**. Almost all of the rest of the discrepancy is due to the remaining old customers failing to provide as much load shed as they were assumed to provide and, for cooler months, the increased temperature sensitivity of the 2017 ex ante model.

Figure 34. Ex Ante Predictions comparing 2017 with actual enrollments vs. intended enrollments



Detailed summary of ex ante impacts

The detailed ex ante impact estimates for all customers, and by segment, are shown in table form below (see Table 15 and Table 16). These, as well as additional ex ante details in the appendices, are provided to help with long-term resource planning. As described earlier, the highest estimated aggregate impact occurs on the PG&E 1-in-10 July peak day, with an average impact of 79 MW and a peak hourly impact of 91 MW (see Table 16).

Table 15. Mean hourly SmartAC™ ex ante load impact estimates by segment, weather year and day type (event period 1-6 pm)

		Day type	All	New Marketing, SmartAC™ only	Old Marketing, SmartAC™ only	New Marketing, Dually Enrolled	Old Marketing, Dually Enrolled
PG&E	**1-in-2**	Typical Event Day	0.49	0.61	0.47	0.49	0.31
		May Peak	0.23	0.27	0.23	0.23	0.16
		June Peak	0.49	0.61	0.47	0.49	0.30
		July Peak	**0.54**	**0.68**	**0.52**	**0.54**	**0.33**
		August Peak	0.48	0.59	0.46	0.48	0.30
		September Peak	0.46	0.56	0.44	0.44	0.28
		October Peak	0.13	0.13	0.14	0.11	0.10
PG&E	**1-in-10**	Typical Event Day	0.59	0.73	0.57	0.57	0.36
		May Peak	0.47	0.58	0.46	0.46	0.29
		June Peak	0.60	0.74	0.58	0.57	0.36
		July Peak	**0.67**	**0.83**	**0.65**	**0.65**	**0.41**
		August Peak	0.61	0.76	0.59	0.60	0.38
		September Peak	0.49	0.60	0.48	0.46	0.29
		October Peak	0.38	0.45	0.37	0.35	0.23
CAISO	**1-in-2**	Typical Event Day	0.36	0.44	0.35	0.36	0.23
		May Peak	0.21	0.24	0.20	0.21	0.14
		June Peak	**0.48**	**0.59**	**0.46**	**0.47**	**0.30**
		July Peak	0.38	0.46	0.37	0.39	0.25
		August Peak	0.33	0.40	0.32	0.32	0.21
		September Peak	0.27	0.31	0.27	0.27	0.19
		October Peak	0.17	0.18	0.18	0.15	0.12
CAISO	**1-in-10**	Typical Event Day	0.49	0.61	0.48	0.49	0.31
		May Peak	0.33	0.40	0.33	0.33	0.22
		June Peak	0.43	0.52	0.41	0.43	0.27
		July Peak	**0.62**	**0.77**	**0.59**	**0.61**	**0.38**
		August Peak	0.54	0.67	0.52	0.54	0.33
		September Peak	0.39	0.47	0.38	0.38	0.25
		October Peak	0.26	0.29	0.25	0.25	0.18

Table 16: SmartAC™ ex ante load impact estimates by weather year and day type (event period 1-6 PM)

			Per Customer (kW)		Aggregate (MW)	
		Day type	Mean Hourly Impact	Max Hourly Impact	Mean Hourly Impact	Max Hourly Impact
PG&E	**1-in-2**	Typical Event Day	0.49	0.60	58.2	71.4
		May Peak	0.23	0.44	27.7	51.7
		June Peak	0.49	0.60	57.8	70.7

		July Peak	**0.54**	**0.64**	**64.1**	**76.1**
		August Peak	0.48	0.60	57.0	70.8
		September Peak	0.46	0.58	54.4	68.2
		October Peak	0.13	0.33	15.7	39.4
PG&E	**1-in-10**	Typical Event Day	0.59	0.68	70.2	81.0
		May Peak	0.47	0.57	55.9	68.0
		June Peak	0.60	0.69	70.6	81.9
		July Peak	**0.67**	**0.77**	**79.3**	**91.2**
		August Peak	0.61	0.71	72.9	84.1
		September Peak	0.49	0.61	58.6	72.3
		October Peak	0.38	0.53	44.9	63.2
CAISO	**1-in-2**	Typical Event Day	0.36	0.52	42.7	62.2
		May Peak	0.21	0.39	24.4	46.6
		June Peak	**0.48**	**0.60**	**56.4**	**70.8**
		July Peak	0.38	0.53	45.1	62.5
		August Peak	0.33	0.51	39.1	60.3
		September Peak	0.27	0.47	32.3	55.3
		October Peak	0.17	0.38	20.0	45.1
CAISO	**1-in-10**	Typical Event Day	0.49	0.60	58.3	71.5
		May Peak	0.33	0.50	39.5	59.2
		June Peak	0.43	0.57	50.6	67.1
		July Peak	**0.62**	**0.71**	**72.9**	**84.1**
		August Peak	0.54	0.63	63.8	74.7
		September Peak	0.39	0.54	46.7	63.7
		October Peak	0.26	0.44	30.5	51.6

Table 17 below walks through the details of starting with event-level aggregate estimates and aggregating up through the full population and then extrapolating aggregate estimates using ex ante methods applied to historical and then forecasted weather data. Enrollment declined throughout the summer, so the relationship between the final two columns changed with time: at the time of the first event in the table there were 133,000 customers in the program, and by the end there were 119,600; the last column assumes there will be 118,300 customers in the program in the summer of 2018.

Key features of the relationships shown in the table are summarized in Figure 29.

Table 17: Differences between Ex Post and Ex Ante Impacts, hours 1 – 6 PM

Date	Hour	Mean17	% of resources dispatched[8]	Ex Post Aggregate Impact (SAC only)	Ex Post Scaled to entire SmartAC-only population	Ex Post Scaled to Full Resource including SmartRate	Ex Ante for Ex Post Conditions, Full Resource	Ex Ante for Standard Event Window, Ex Post Conditions	Ex Ante for Standard Event Window, Ex Post Conditions, Projected Enrollment
06-19	5 - 6 PM	87.45	30	18.48	70.44	84.17	83.12	83.12	74.25
07-07	4 - 5 PM	83.89	20	10.75	61.29	74.14	81.06	79.31	76.73
07-07	5 - 6 PM	83.89	60	35.14	66.37	80.28	77.56		
07-15	1 - 2 PM	79.79	10	2.66	23.15	27.53	25.14	50.91	49.42
07-15	2 - 3 PM	79.79	10	3.85	33.51	40.02	42.47		
07-15	3 - 4 PM	79.79	10	5.46	47.42	56.39	55.36		
07-15	4 - 5 PM	79.79	10	6.27	54.42	64.67	63.79		
07-15	5 - 6 PM	79.79	10	6.52	56.66	67.39	67.79		
07-27	3 - 4 PM	79.10	20	6.70	38.27	45.98	50.22	58.19	56.75
07-27	4 - 5 PM	79.10	20	8.68	49.59	59.58	58.96		
07-27	5 - 6 PM	79.12	60	27.14	51.34	61.68	65.38		
08-02	4 - 5 PM	83.18	10	5.20	58.38	70.04	59.25	63.91	62.55
08-02	5 - 6 PM	83.30	10	5.45	61.37	73.63	68.56		
08-27	1 - 2 PM	84.12	10	4.35	37.92	44.81	47.33	69.09	68.43
08-27	2 - 3 PM	84.12	10	6.26	54.53	64.37	63.75		
08-27	3 - 4 PM	84.04	10	7.60	66.06	78.08	75.03		
08-27	4 - 5 PM	84.04	10	8.47	73.63	87.26	81.77		
08-27	5 - 6 PM	84.04	10	8.75	76.03	90.12	77.56		
08-28	5 - 6 PM	84.62	40	22.59	63.87	76.64	74.12	74.12	73.50
Average		**82.26**	**19**	**10.54**	**54.96**	**65.62**	**64.12**	**68.38**	**65.95**

[8] These are based on the number of serial groups involved. In practice the actual count of event participants tends to be a slightly lower percentage of actual enrolled participants. Actual enrollments, and not the % of resource dispatched are used to compute the aggregated impacts.

Other key findings: Opt-outs

In the context of the SmartAC™ program, opt-outs are actions taken by customers to exempt themselves from program participation for the day via web interface or call center. The fraction of customers who opted out of even one event in the entire year is very small: fewer than 2,800 customers (2.2-2.4% of enrolled customers) opted out at least once. However, such actions are only taken by the most pro-active customers, so it is reasonable to expect that the timing of opt-outs and who choose to do so expresses patterns relevant to understanding the motivations and morale of the broader based of less pro-active program participants as well.

Customers opt out of the program on days that are and are not event days: Figure 35 presents the number of opt-outs as a function of time, encoded in red on event days and blue on non-event days. By definition, opting-out on a non-event day has no effect. Event days do see elevated opt-out behavior, but so do days near them and other hot days. Note the relatively large number of opt-outs in early September even on non-event days. Some customers may think they are, or will be, participating in an event when they are not, and so may believe the discomfort that motivated opting out was caused by the program when it has another source entirely.

Figure 35: Number of opt-outs by date

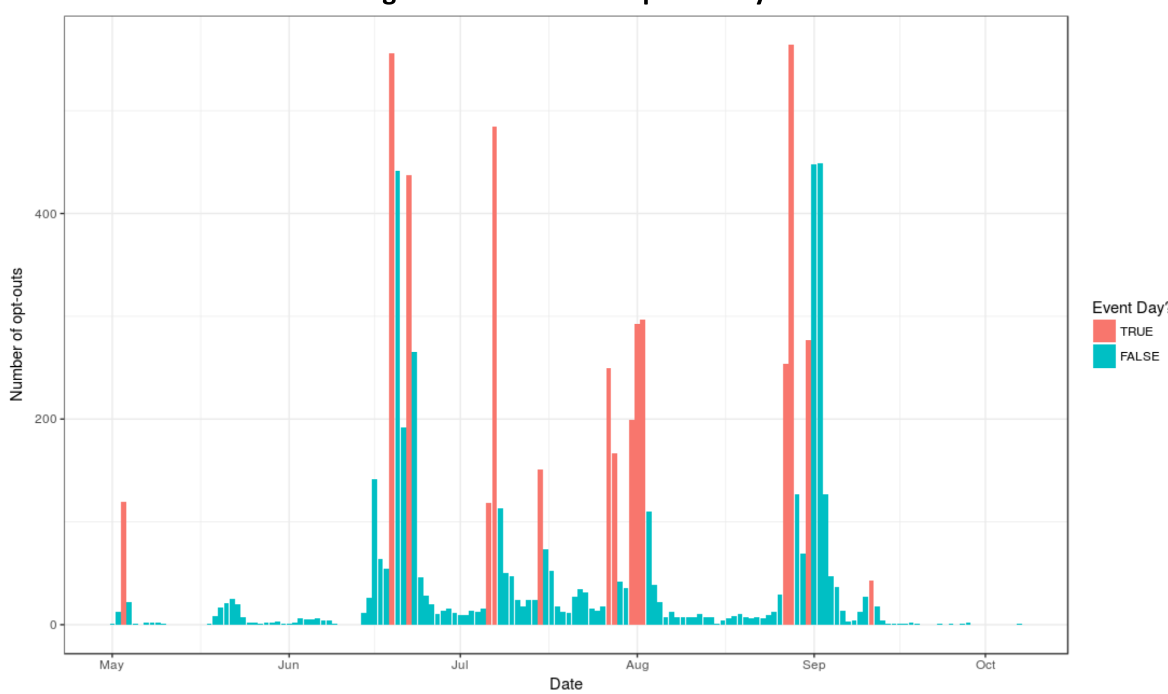


Figure 36 shows the number of opt-outs for each customer, in order by the number of times they opted out (top panel). At the far left are a few customers who each opted out more than 80 times. About 180 customers opted out ten times or more. The bottom panel of the figure shows the cumulative number of opt-outs (the integral of the upper plot). The 500 customers who opted out the most are responsible for more than 5300 opt-outs, which is well over half of the 8234 opt-outs that took place during the year.

Figure 36: Two views of customers opt out

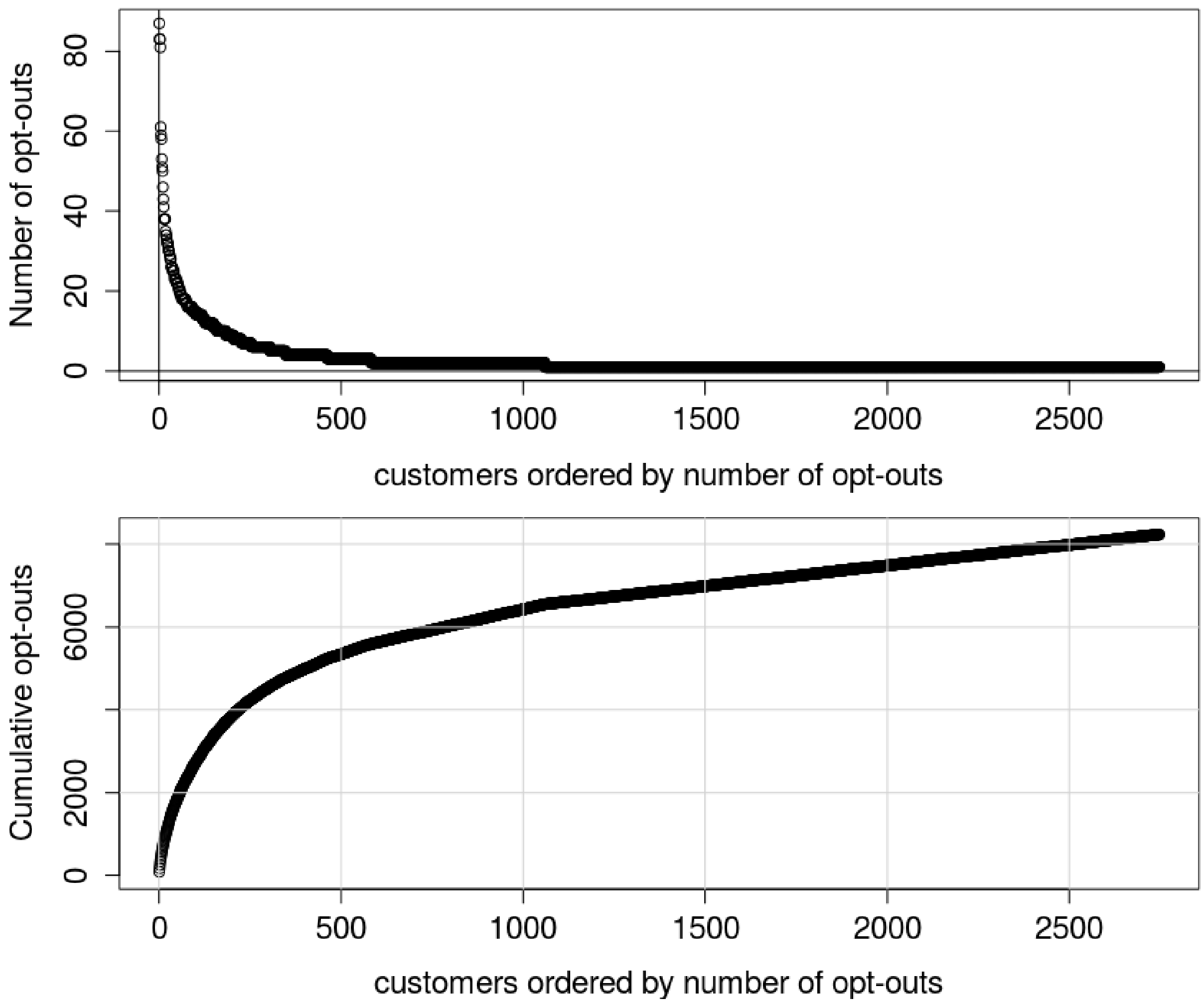
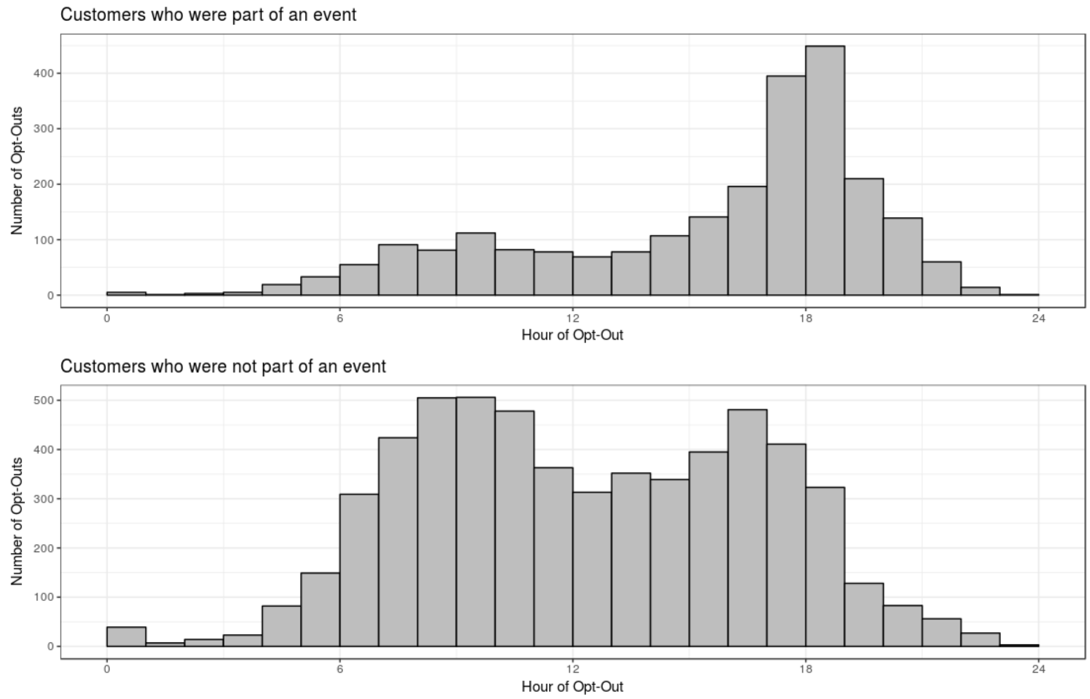


Figure 37 shows the time of day that people opted out, separately for people who were and were not part of an event. **Most opt-outs are from people who were not participating in an event**. These people opted out fairly uniformly throughout the day (bottom panel). Customers who were part of an event and chose to opt out were much more likely to do so in the late afternoon or early evening. This difference in timing suggests that participants were specifically responding to the events and that discomfort likely played a role in the decision. Perhaps some of these customers returned from work and found that their house was uncomfortably hot.

Figure 37: What time of day did people opt out?



Overall, opt-outs do not currently affect the aggregate impact of the program to a significant degree, since the number of opt-outs is a very small fraction of the total enrollment. Most opt-outs do not happen because the program is making people uncomfortable: the majority of opt-outs are from customers who are not participating in an event, although they may think they are.

A few hundred people opt out routinely and therefore would contribute nothing to the program even if called to participate. They are likely unhappy with their arrangement and are certainly expending significant effort to avoid participation.

Conclusions and recommendations

Based on our evaluation results, there have been significant benefits to the program through recent modernization efforts including more targeted recruitment and better devices and signaling. The poor performance of thermostat devices, gains from Gen 2 switches, and the "New" marketing strategy are particularly striking. Another finding, consistent with past results, was the under-performance of customers also enrolled in the SmartRate™ program, likely due to self-selection with unusually motivated savers agreeing to be part of both programs and having durable efficiency responses to the time of use prices associated with SmartRate™. This is quite possibly an example of a non-virtuous interaction between efficiency and demand response. Collectively, these observations serve as a reminder that **recruitment, device selection, and proper installation and configuration are critical to program outcomes**. Especially because they also rank among the most significant program costs, they are resource-defining drivers of the program. They are also all determined at the earliest stages of customer engagement with the program.

On the recruitment front, the most notable development in the 2017 program year was the cancellation of the "Automoves" initiative, leading to much higher than anticipated net de-enrollment from the program for the year. This outcome was a disappointment in terms of aggregate potential but there were important lessons about customer perceptions of the program and its role in their lives in the pushback that led to the cancellation of Automoves. The recruitment “gap” for 2017 leaves room for future recruitment that makes strategic adjustments to tactical methods and the program population.

The other major change to the program was the de-enrollment of 22,000 customers with low air conditioning loads. Per-customer load impacts increased compared to recent years, so this change appears to have had the desired overall effect. Most of the de-enrolled customers had been contributing at least a little load shed, so this change, coupled with limited new enrollment, decreased aggregate load shed. There are some additional shifts that are likely attributable to the changed mix of enrolled customers. The gap between CARE and non-CARE customers was dramatically narrowed, apparently due to a greater proportion of non-CARE customers (21% vs. 15%) eliminated as low-AC-consumers. There is also evidence in the ex post results and ex ante analysis that event outcomes are more temperature sensitive than in the past, which would be expected after eliminating less responsive customers, although this observation is complicated by other structural changes to the ex ante model that improve the fit and lead to greater modeled temperature sensitivity. One consequence of more temperature-sensitive customers is greater reductions in load impacts for cooler events.

Across the board, the lessons are thematically the same: the SmartAC™ resource is shaped by who is enrolled. To continue to shape and adapt the program to the changing needs of the utility and the grid, PG&E will need to weigh investments, primarily in recruitment, retention, and hardware, in comparison to resource value. Direct load control of AC units is one of the most dispatchable and measurable demand side resources, so the program will clearly be able to offer a valuable and quantifiable resource over the coming years. One aspect of the value of this program was demonstrated during an emergency event on May 3rd when CAISO called a “Stage 1” emergency due to a confluence of reduced generation and electricity import capacity, and a large demand forecasting error leading to shortfalls of power on the grid. SmartAC™ Contributed two hours of 14-15MW load shed on a day significantly cooler than usual event days.

Program level recommendations and recommendations for future research are outlined below.

Program recommendations

Targeting. The program should continue to target homes with higher reference cooling loads and more potential to save on hot days (as they have been doing since 2014). The program will also need to recruit additional participants to recapture lost customer count in 2017 and should focus on customers with high temperature responsiveness in hotter climates.

Upgrading hardware*.* Upgrading hardware to ensure reliable communications and durability for new participants is expected to continue to increase per-participant event impacts. PG&E should continue to replace the under-performing PCTs and use 2-way switches going forward. As some PCTs be updated “over the air,” we recommend contracting with third party vendors for PCT participation in SmartAC™, especially if the recruitment can be done for customers whose install costs have already been incurred and whose AC consumption magnitude and patterns are known.

Removing participants. Before removing additional participants, PG&E should assess the full impact on aggregate ex ante load reduction versus the cost of continuing to include these low performers in the program. If there are no additional costs (i.e., all costs to date are sunk) then there may be reason to keep these low performers if they contribute to load reductions at some level. They will have a downward pull on the average reduction per device but could nevertheless deliver more value than they cost to keep enrolled.

Recommendations for future research

Based on our findings, there are three areas where we are recommending additional research:

Learning from opt-outs and de-enrollment

Customer program defection for the day (opt-outs) or forever (de-enrollment) reduce load impacts from the program and strand recruitment, device, and maintenance costs. Efforts should be made to understand these customer actions better. The patterns in each can be considered customers voting on program characteristics through their own actions. Some opt-outs provide needed flexibility and preserve valuable customer enrollment. Others appear to be a form of quasi-de-enrollment by customers who would be better off leaving the program. Our preliminary analysis with the data on hand demonstrates correlations between opt-outs and temperature, the period during events, and the one or two days following events. We anticipate patterns in de-enrollment that also reflect the experience of customers during and following events. In sum, such information should help provide a barometer for what types of customers are not “happy” participants in the program and are more likely to defect, either dampening event performance or stranding initial recruitment and installation costs. Such information should prove valuable in informing future customer recruitment and retention strategies.

Better characterization of control matching method performance

Sub-LAP events, evaluated using matched control groups, are growing in importance for the program. In past SmartAC™ program evaluations, synthetic control-matching has relied on matching each event participant to a “control” customer that has a similar mean daily load shape. This method of choosing controls implicitly assumes that program response is determined by the full load shape, yet the program targets only AC loads. Controls with smaller AC loads than participants (offset by higher non-AC loads) can still make good matches by load shape. When events are called on hot days, such controls with have systematically lower AC response to the hot weather, resulting in artificially low reference loads and under-prediction of event impacts.

The work to date applied a new matching approach for sub-LAP events based on regressed “thermal response” in the form of daily kWh per mean degree of outdoor temperature during comparison days. This model appears to perform better than load shape matching across a wider range of temperature conditions than used to make the matches, but there are many other potential permutations and improvements that could be made. We recommend developing a metrics of control matching performance in the SmartAC™ context (i.e. looking for zero estimated load impact on non-event days and during pre-event periods, and similarity to RCT-based impact estimates when applied to serial group events), followed by the development and systematic comparison of a range of control matching strategies. Our recommended emphasis would be on strategies that can take into account time of day and seasonal and climatic variability at the same time. Better control-matching would improve the performance of any PG&E program evaluation (or other evaluation) that relies on case-control analysis that is not based on randomized controlled trials, so this work could have broad benefits to PG&E.

Further exploration of multi-device customers

Although multi-device customers are no longer being recruited, they are still present in the pool of enrolled customers. From an evaluation perspective, multi-device customers are problematic due to their increased likelihood of being event participants and decrease likelihood of being controls, coupled with their tendency to contribute lower than average load impacts (at least to serial group events). To date, evaluators have either set aside multi-device customers (2013-2015; 2017 to date) or included them in event estimates without correcting control sampling (2016). It's likely that multi-device customers would out-perform their single-device peers when all their devices are called, as happens in sub-LAP events: most of these customers are large consumers electricity for air conditioning, and although their load shed per device trails that of other customers they probably more than make up for this by having more devices.

Looking to the future

All of the above recommendations could be understood as preparation for SmartAC™ to continue providing value under the changing conditions of the grid over the coming years. With variable renewables making deeper incursions into the supply mix, the need for load reductions on the grid is expected to be less reliably driven by peak temperatures. At the same time, resources like grid tied batteries are offering a hardware-based alternative to demand side management. The costs of those battery systems are likely to provide a new backstop price on load curtailment resources like SmartAC™. The program should also keep an eye on these developments to ensure it provides a valuable service to the grid for years to come.

Appendix A: Detailed ex post results

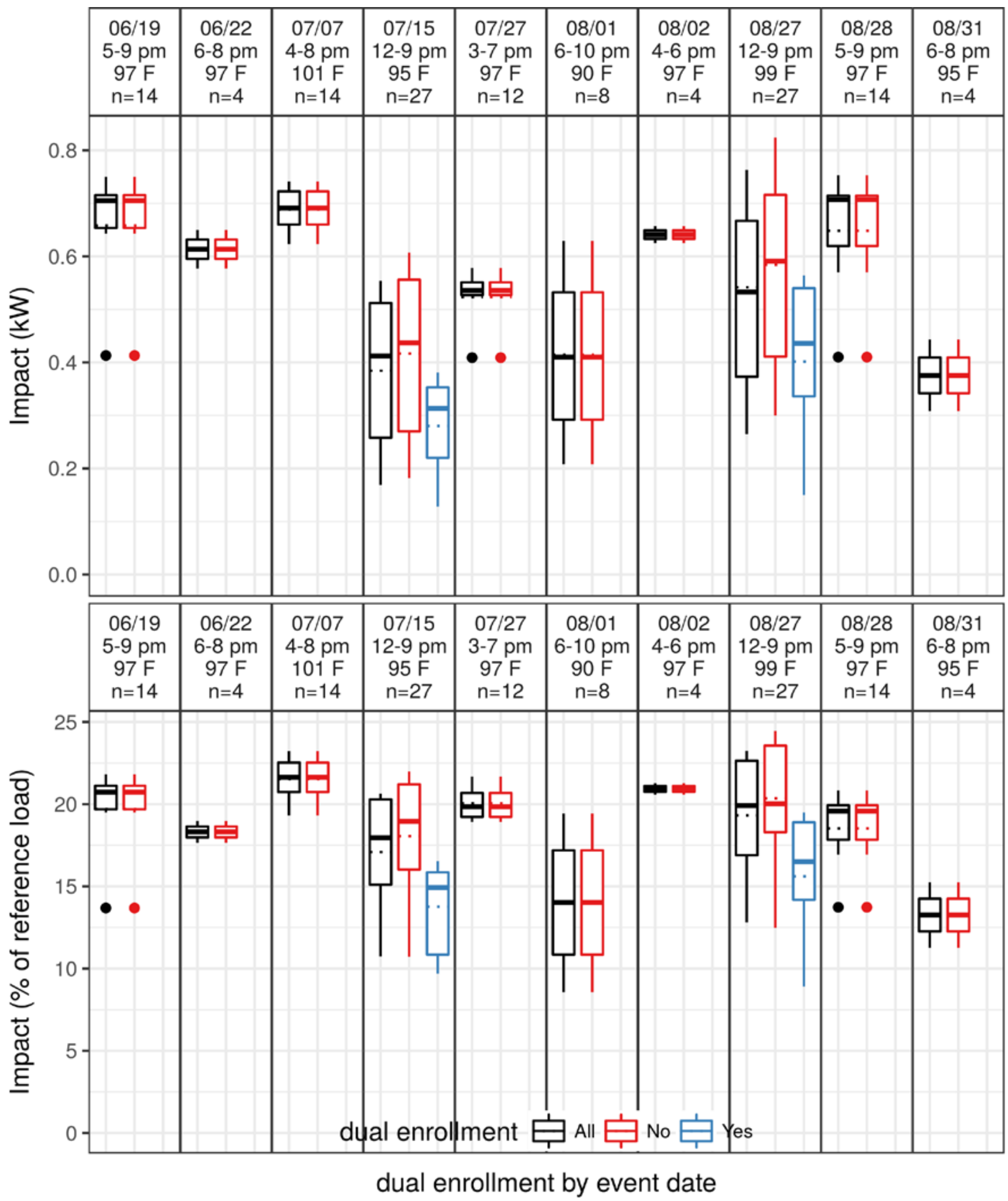
This appendix makes extensive use of box and whisker plots to illustrate variability of load impacts across event hours within each day of serial group event activity. Each day is labeled with the date, range of hours during which events were active, the average outside temperature across all event hours, and the number of event hours, n, across all relevant customer sub-groups used to construct the day's box plots. As a reminder, for boxplots 25th, 50th, and 75th percentiles as solid horizontal lines, with dotted line mean for all values. The upper and lower whiskers extend from the 75th and 25th percentile lines, respectively, to the largest/smallest value no further than 1.5 times the IQR (where IQR is the inter-quartile range, or distance between the upper and lower lines of the box). Values beyond that, point are plotted individually as outliers.

In all cases, the figure presented for each customer sub-category provides results as the average load impact per customer in kW (above) and percentage of reference loads (below). The upper panel shows absolute performance, but, assuming reference loads and load impacts are both proportional to outside temperature, the bottom panel shows a more weather normalized view of outcomes across days.

Dually enrolled customers (SmartAC™ and SmartRate™)

There were only two non-SmartDays™ when SmartAC™ events were called. From this small sample in Figure 38, it is clear that dually enrolled customers (blue) have lower load impacts than their SmartAC™-only peers. This effect is likely attributable to self-selection bias of customers willing to enroll in both programs and the energy usage and control changes customers are likely to put in place when faced with time of use rates.

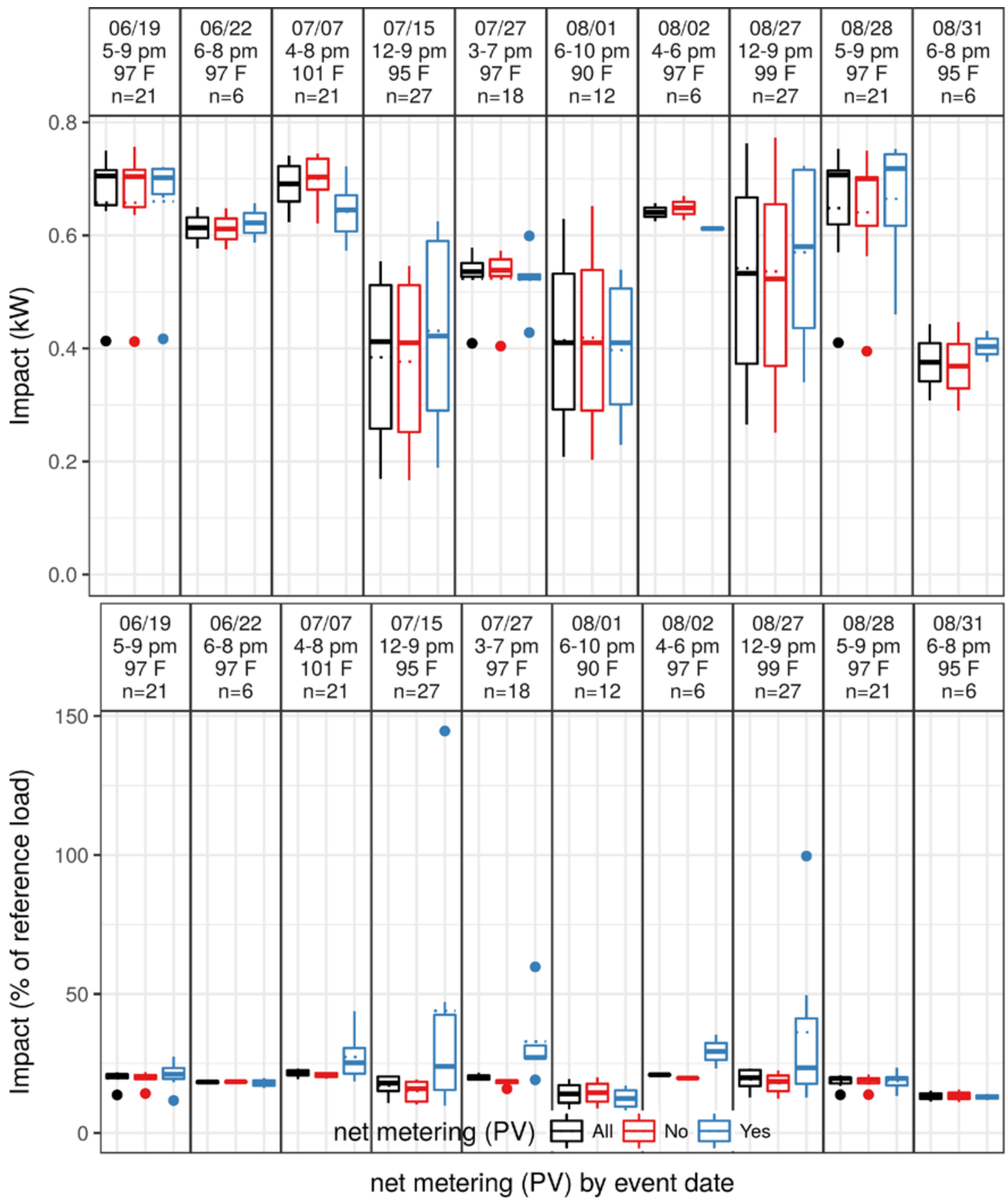
Figure 38: Load impacts for all event days by dual program enrollment



Net-metered customers

Referencing Figure 39, there does not seem to be a systematic difference between net-metered customers in blue (i.e. PV owners) and everyone else in red, but we do see that group under-performing on 7/7 and out performing somewhat on 8/27 and 8/28. Note also that the net reference loads for PV owners would be expected to be smaller than the general population. We see that effect illustrated with the net-metered sub-group tending to have greater load reductions as percentage of reference loads.

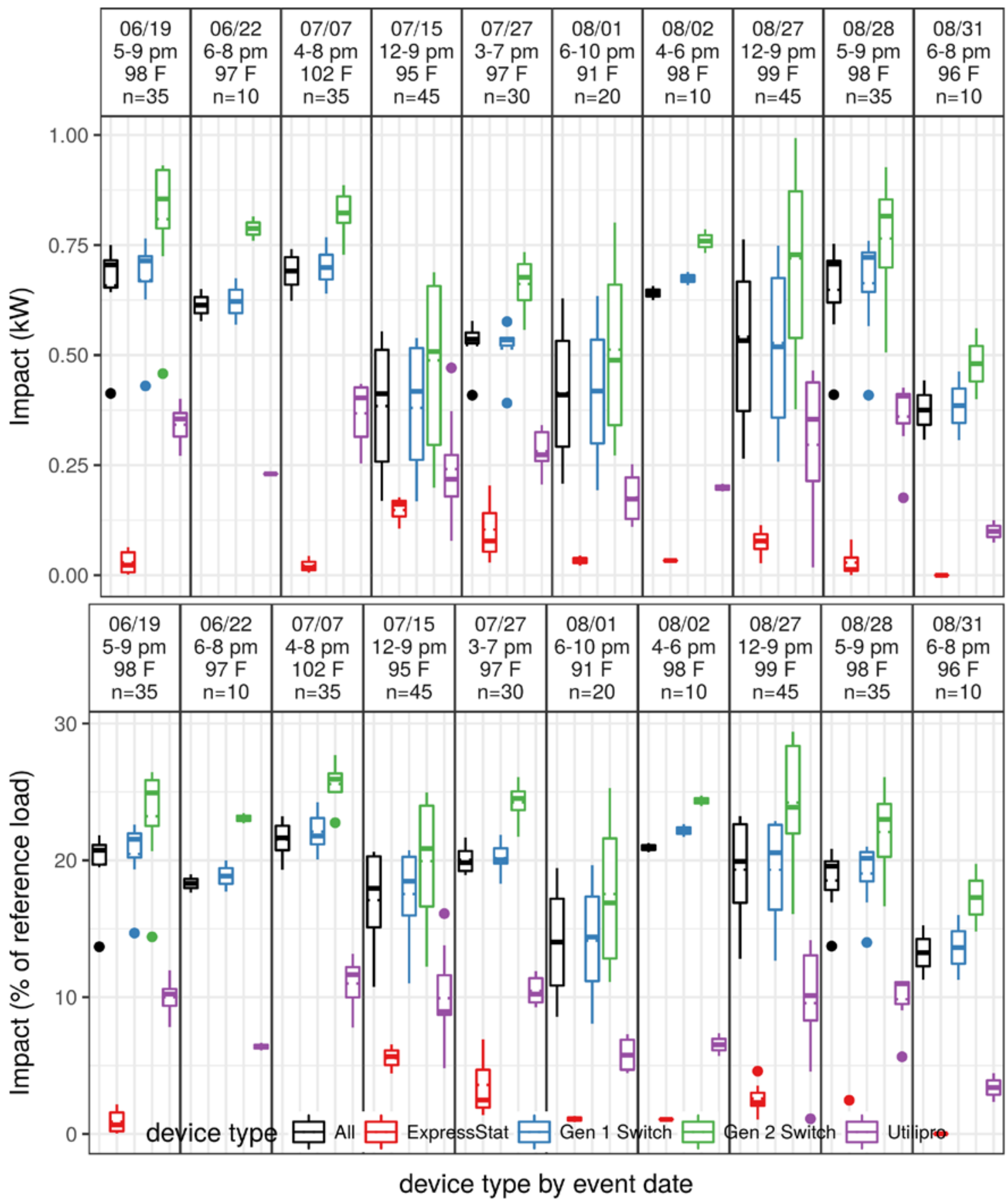
Figure 39: Load impacts for all event days by net-metering status



Device type

Figure 40 illustrates event hour outcomes by day and device type. Among the various devices, Gen 2 devices show the highest impacts during the peak, followed by Gen 1 devices and the Utilipro. As discussed in earlier program reports, ExpressStat devices do not appear to contribute load reduction. The programmable thermostats associated with the program fail to deliver adequate load impacts, whereas the newest hardware is out-performing the general population by a wide margin. We note that the newer hardware is also in the hands of better program recruits due to "New" targeted marketing.

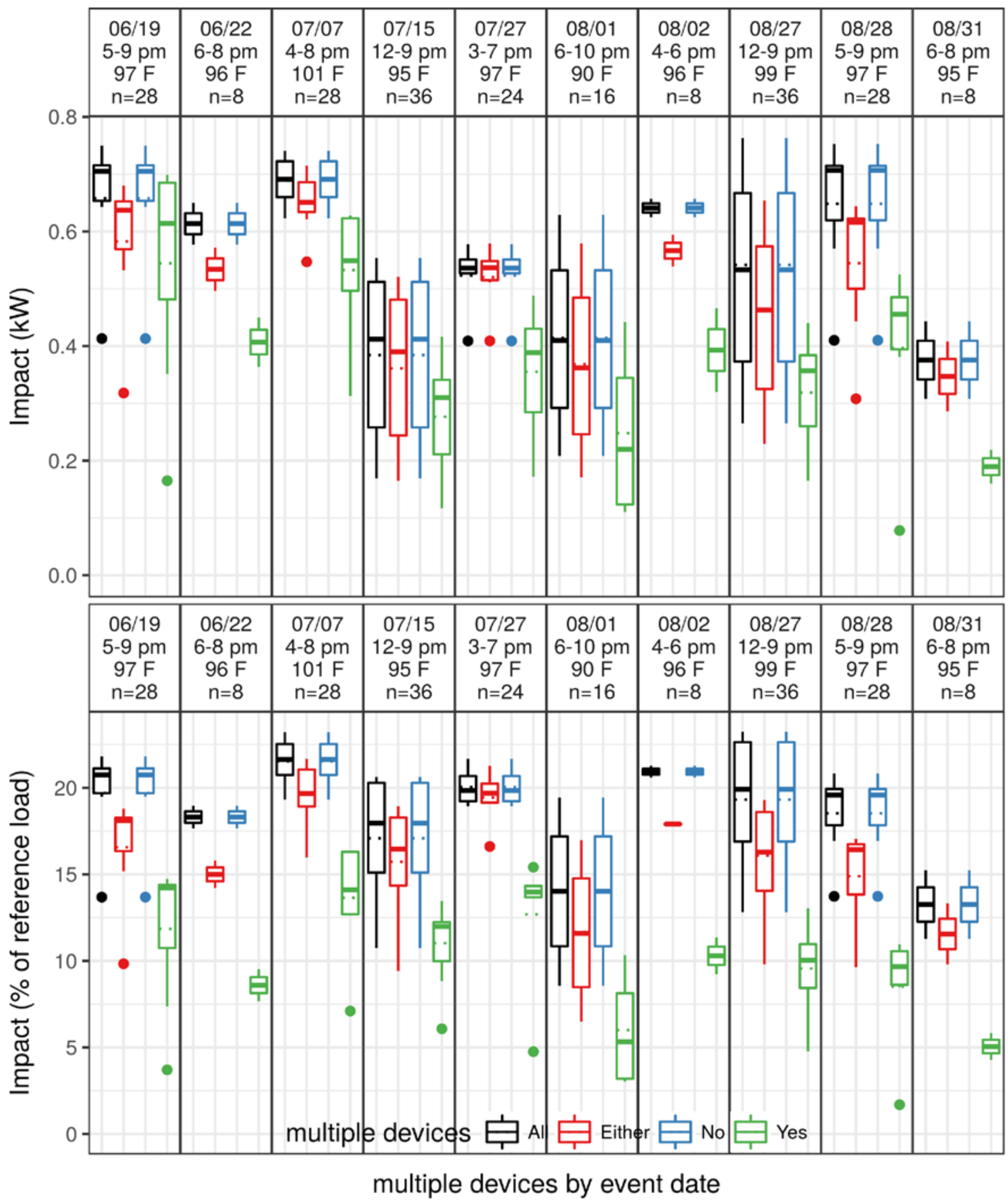
Figure 40: Load impacts for all event days by device type



Households with multiple control devices

Figure 41 compares load impacts for customers with ("Yes") and without ("No") multiple devices in their homes. First it must be noted that ex post results default to measuring single device households, so the "No" category matched the outcome for all customers. The under-performance of multi-device households is consistent with past program results. We believe this is a function of (a) a wider variety of device capacities in multi-device households – some are small and (b) only a subset of devices tending to be called by serial group with the others continuing to operate and pick up some of the slack.

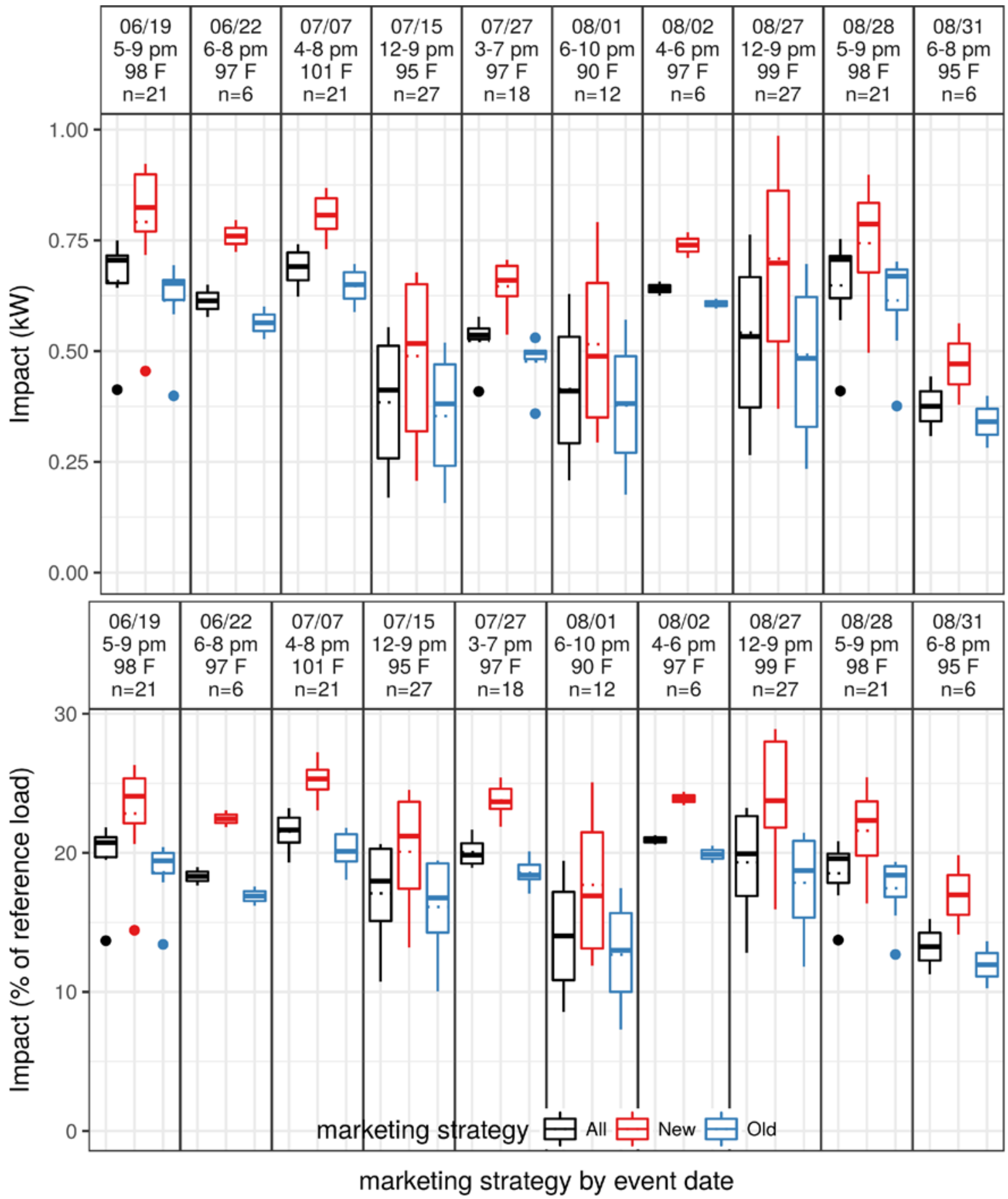
Figure 41: Load impacts for all event days by device count



Targeted marketing strategy

Figure 42 compares outcomes for customers enrolled under "Old" (blue) and "New" (red) marketing strategies. The new marketing is out-performing the older customers and this is one of the highest impact effects we've seen among customer sub-groups. It is certainly due to higher quality recruits, but we also suspect part of the effect is due to newer/improved hardware and vice-versa. Naturally the newer customers were more likely to have newer Gen 2 Switches that out-perform the Gen 1 model, so the recruitment and device hardware are mutually supportive.

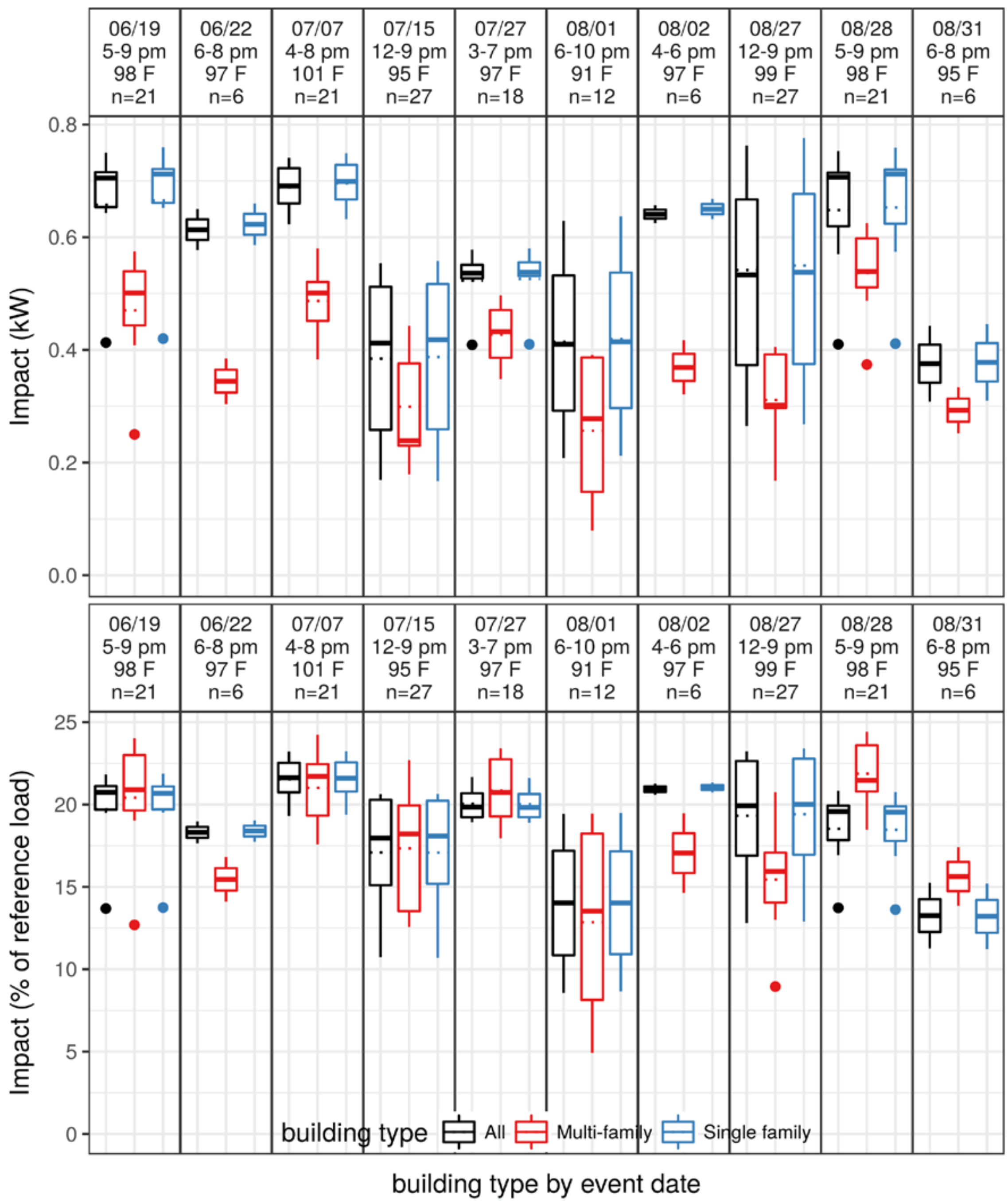
Figure 42: Load impacts for all event days by marketing strategy used for recruitment.



Multi-family

Figure 43 compares outcomes for Single (blue) and Multi-family (red) customers. The single-family customers vastly outnumber the multis and share the outcomes of "All" customers. Apartments tend to be smaller than free standing homes and they have less surface area exposed to the outdoors vs. neighboring apartments above, below, and alongside of them. All of these factors should lead to lower cooling requirements. This is corroborated by the fact that the multi-family customers typically perform as well as or better than the other groups as a percentage of the reference loads of their peers.

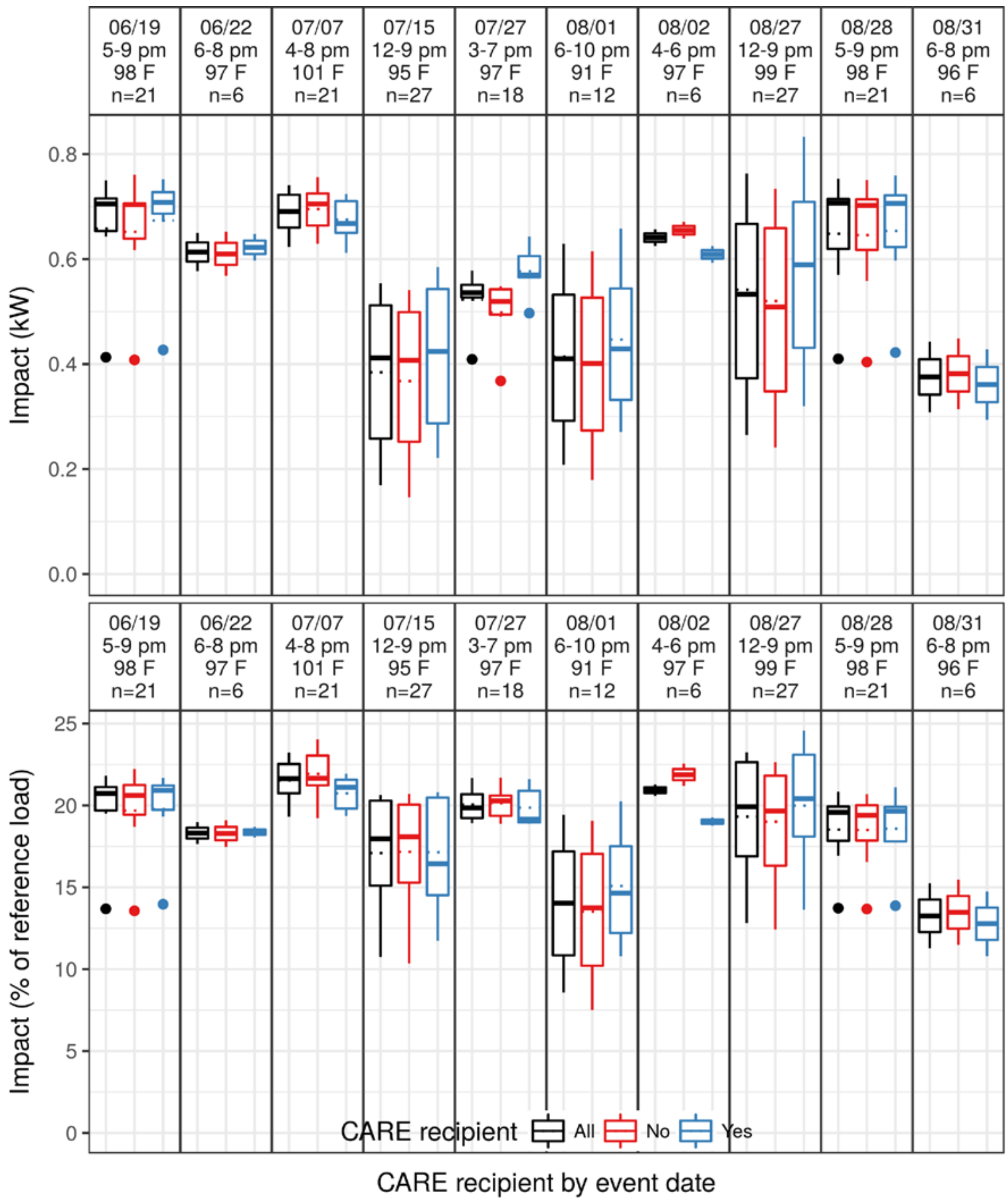
Figure 43: Load impacts for all event days by building type (multi vs single family)



CARE customers

Figure 44 compares outcomes for CARE recipients (blue) and non-CARE (red). The main news here is that the two groups have similar outcomes in both kW and percent of reference loads terms. In the past, CARE recipients have had larger load impacts than non-CARE customers, largely due to more of them living in hotter climates. We believe the change this year is the result of de-enrollment of low-AC-consumption customers (21% of non-CARE but only 15% of CARE customers were dropped).

Figure 44: Load impacts for all event days by CARE status



Appendix B: Detailed description ex ante approach

Data preparation

The first step in fitting the ex ante model is to select what data to use.

Incomparability of previous years' data

The statistical model builds on the average (or per participant) PY 2017 ex post results only; it does not include results from previous years as was done in the past. The rationale for this choice is that the make-up of participants has changed significantly over time given natural attrition, targeted marketing, and the fact that early in the PY 2017 season the program lost more than 30,000 households, most of them due to a decision to de-enroll customers with relatively low load average load shed. Furthermore, the earliest program years provided significantly higher load shed than more recent ones. Their inclusion tends to bias the forecast upwards. Finally, PG&E changed the assignment of customers to

Exclusion of certain events

We exclude data from the Sierra local capacity area from 3-5 PM on 8/2/2017. At that time there was a severe thunderstorm warning across the entire region. The load shed is quite high, in spite of the outdoor temperature being the lowest in the entire dataset. The result is that this event is a statistical 'outlier', whose data would have a disproportionate effect on the statistical model. The conditions that led to the unusual load shed behavior of that event in that local capacity area appear to be anomalous, so those data points were excluded from the ex ante modeling. Those points can be seen in Figure 45 which shows all of the ex post load shed data on which the ex ante model is based. Each point represents one marketing category (old or new customers) in serial group in one LCA. Each panel shows a different hour, labeled by the start of the hour. Data from the Greater Bay Area and the Greater Fresno Area are shown in different color for reasons we discuss below. In each hour, a blue horizontal line at 0.5 kW and a blue vertical line at 100 F are shown to help compare across hours. Also shown in each hour is the best-fitting line is shown for the new customers (circles, upper line) and the old customers (triangles, lower line). Gray bands around the regression lines indicate the uncertainty in the slope and intercept of those lines.

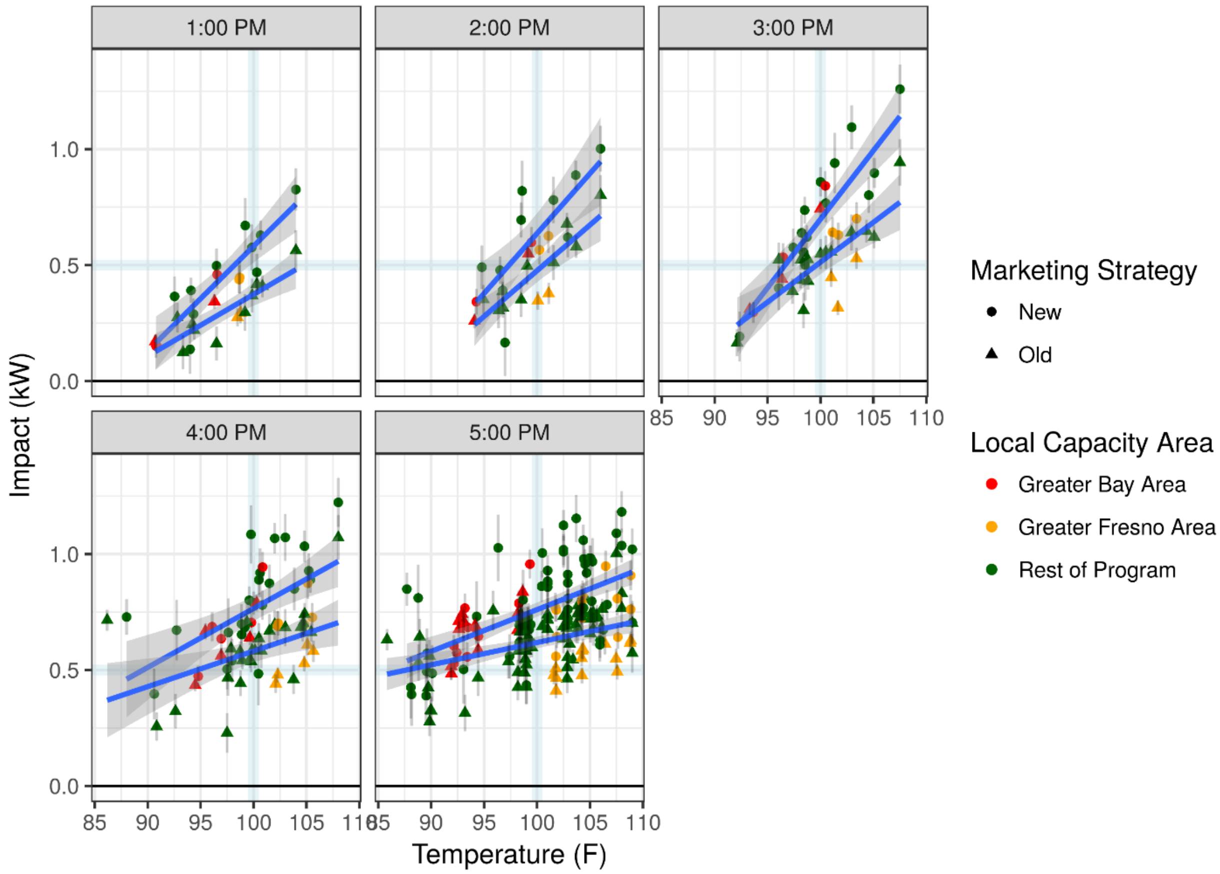
At any temperature, the data points from the Greater Fresno Area have the lowest observed load shed. The anomalous Sierra event data are at the far left in the 4:00 PM plot. Data from this event stand out less in the 5:00 PM plot, but they are still somewhat extreme: they are the triangle at extreme left and the highest of the two circles below 90 F.

Use of SmartAC-only customers when fitting the model

Customers who are dual-enrolled in the SmartAC™ and SmartRate™ programs do not share load reduction timing with SmartAC™ events on days that are also SmartRate™ days and their ex post load reductions are quantified as part of the SmartRate™ evaluation effort. In 2017 there were only two SmartAC™ days that were not also SmartRate™ days, so there is not enough ex post data on SmartRate™ customers to reliably fit a model to them. Instead, we follow the procedure used in previous years: the ex ante model was fit to the SmartAC™ customers only. Ex ante predictions for SmartRate customers assume that they have a per-customer impact that is a fraction of their SmartAC™-only counterparts: within a given local capacity area, new dual-enrolled customers are assumed to generate 77% as much load impact as the SmartAC™-only customers; for old dual-enrolled customers the ratio is 68%. These figures are based on the ratios on the two SmartAC™ event days that were not SmartRate™ days. These ratios are rather uncertain due to being based on only two days of data and small numbers of dual-enrolled customers in some load capacity areas.

Characteristics of the ex post data

Figure 45: Mean load shed for each serial event group vs Temperature, for customers in SmartAC™ Only, for each hour



1. As expected, the ex post data in Figure 45 show that there is a strong relationship between outdoor air temperature in a given hour and the load reduction if a SmartAC™ event takes place in that hour.
2. There is also a strong effect associated with the hour of the day, with different hours having different load shed at a given temperature. The slope with temperature is about the same in all of the first four plots, for both old and new customers, but the 5:00 PM plot (the hour from 5-6 PM) has much shallower slopes.
3. Additionally, it's clear that there is a substantial difference between customers recruited using 'old' and 'new' marketing (fit with the lower and upper line on each plot). The extent to which this is a causal relationship is not clear: customers identified via new marketing also have new control devices, which are more likely to work properly than the older devices.
4. A weak effect is expected from temperatures prior to the event hour: if temperatures are cool overnight and warm up throughout the day to a peak of 95 F, the cooling load will be lower than if temperatures were warm overnight and warmed farther throughout the day. In the latter

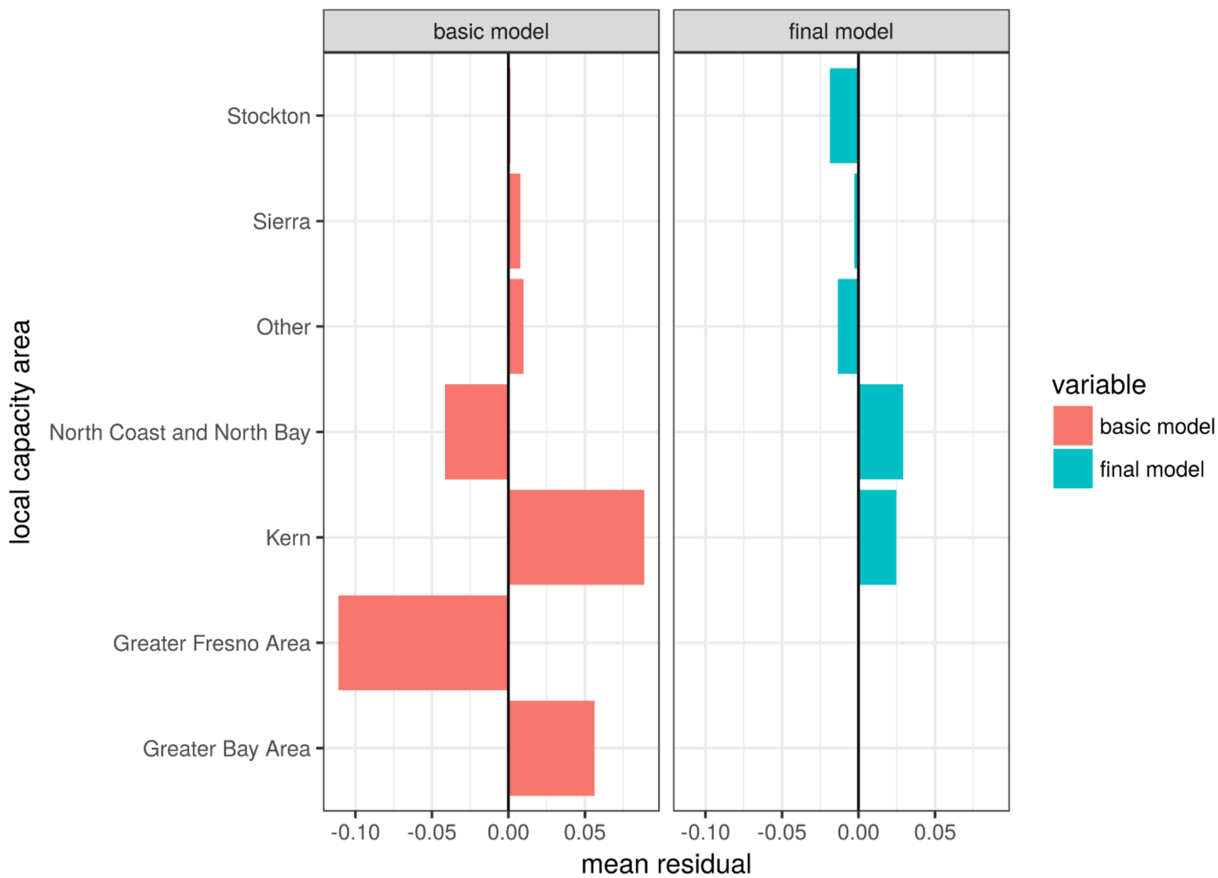
case the cooling load will be higher than in the former case, and thus the load shed will be higher if a customer experiences a SmartAC™ event. This effect was included via an additional temperature variable called 'mean8', which is the mean outdoor temperature over the first eight hours of the day.

The effects listed above constitute the major observed characteristics and expected characteristics.

Statistical modeling

A linear regression model was constructed to capture all of the features of the data that are noted above; we call this the 'basic purely linear' model (we explain the 'purely linear' terminology, below). Figure 46 shows the mean residual (actual impact minus predicted impact, in kW) by LCA for the basic model and for the final model, which includes additional terms as discussed below; Humboldt has only two SmartAC™ customers and is excluded from the plot.

Figure 46: Mean residual by LCA, for models without and with LCA-specific terms



As Figure 46 illustrates, the impact is substantially lower than predicted by the basic model (left panel) in the Greater Fresno Area, and higher than predicted in the Greater Bay Area and in Kern. Communication with the SmartAC™ control devices is known to be problematic in the Greater Fresno Area: some devices do not receive the signal to participate in events, so it is understandable that the load shed per customer there is lower than elsewhere.

To adjust for the difference in load shed behavior between the these two LCAs the other LCAs, we add an additional constant term and temperature slope that apply only to Bay Area customers, and an additional temperature slope that applies only to Fresno Area customers. There is no evidence that the Fresno Area customers need a separate constant term.

The 'final purely linear' regression model is defined by the following equation. (The 'purely linear' terminology is described below).

$$
\begin{aligned}
L_h &= \alpha_h \delta_h + \beta_T T_h + \beta_8 T_8 + \alpha_{old}\delta_{old} + \alpha_{Bay}\delta_{Bay} \\
&+ \beta_{h18}\delta_{h18} T_h + \beta_{old}\delta_{old} T_h + \beta_{bay}\delta_{bay} T_h \\
&+ \beta_{Fresno}\delta_{Fresno} T_h + \epsilon
\end{aligned}
$$

where:

L_h is the predicted impact (the load shed) in hour *h.*

ϵ is the residual error for each estimate.

α_h is the constant term for hour *h*; this is the end hour, e.g. hour 14 is the hour ending 14:00, i.e. the hour from 1-2 PM. (This notation is used for historical reasons).

δ_h is an indicator which is 1 (thus activating its coefficient) for data points in hour *h* and 0 for other hours. We use this δ notation for other indicator variables as well.

β_T is the coefficient of the temperature in all hours *h*.

β_8 is the coefficient of the mean temperature over the first eight hours of the day (midnight to 8 AM).

α_{old} is the offset for customers recruited under the old marketing strategy; data points for those customers are flagged with an indicator variable δ_{old}

α_{Bay} is the offset for customers in the Greater Bay Area, δ_{Bay} indicates data points from that LCA.

β_{h18} is the additional temperature coefficient that applies in 'the hour ending at 18:00', i.e. 5-6 p.m.; data points for those hours are flagged with an indicator variable δ_{h18}.

β_{old} is the additional temperature coefficient that applies to old customers, as indicated by δ_{old}.

β_{bay} is the additional temperature coefficient that applies to customers from the Greater Bay Area, as indicated by δ_{Bay}.

β_{Fresno} is the additional temperature coefficient that applies customers from the Greater Fresno Area, as indicated by δ_{Fresno}.

All of the α terms have units of kW; all of the β terms have units of kW per degree Fahrenheit; $T_h$and T_8 are measured in degrees Fahrenheit.

In fitting the regression, the ex post data points were given statistical weights related to their standard error (see the ex post section of this report for the uncertainty calculation). The assigned weights were:

$$w = \frac{0.04^2}{0.04^2 + \sigma^2}$$

where σ is the standard error (in kW). Weighting all points equally would give equal influence to the two customers in Humboldt as to the tens of thousands of customers in the Bay Area. Weighting by $1/\sigma^2$ would be correct if the model were expected to fit perfectly except for variation caused by sampling error, i.e. if the model would fit perfectly if sample sizes were extremely large. In fact, though, we know that some LCAs behave differently from others, and some days or event-hours will differ from what would be predicted by any combination of explanatory variables. The weighting parameter that we set to 0.04 kW in the equation above is intended to approximate the amount that the *actual* load impacts (as opposed to the *observed* load impacts, which are subject to sampling error) typically differ from their predicted values. A formal statistical estimate of that parameter would require a much more involved analysis and would likely have little influence on the results, as we checked by varying the parameter by a factor of 2 and confirming that this made little difference in the model predictions.

Table 18: Measures of model fit for the final purely linear model

R2	Adj. R2	RMSE
0.979	0.979	0.169

Table 19: 2017 ex ante model regression coefficients and p-values:

Term	Coeff.	Std. Err.	p-value	p-value flags: p < 0.1 : * p < 0.01: ** p < 0.001: ***
$\boldsymbol{\alpha_{14}}$	-3.989	0.279	1.04e-36	***
$\boldsymbol{\alpha_{15}}$	-3.930	0.284	6.27e-35	***
$\boldsymbol{\alpha_{16}}$	-3.880	0.284	4.55e-34	***
$\boldsymbol{\alpha_{17}}$	-3.833	0.287	4.74e-33	***
$\boldsymbol{\alpha_{18}}$	-1.796	0.228	4.82e-14	***
$\boldsymbol{\beta_T}$	0.042	0.003	2.47e-33	***
$\boldsymbol{\beta_8}$	0.005	0.001	8.98e-05	***
$\boldsymbol{\alpha_{old}}$	1.040	0.230	8.21e-06	***
$\boldsymbol{\alpha_{Bay}}$	-0.997	0.400	1.32e-02	*
$\boldsymbol{\beta_{h18}}$	-0.020	0.003	3.18e-14	***
$\boldsymbol{\beta_{old}}$	-0.012	0.002	3.95e-07	***
$\boldsymbol{\beta_{bay}}$	0.011	0.004	6.45e-03	**
$\boldsymbol{\beta_{Fresno}}$	-0.002	0.000	3.84e-26	***

For example, consider predicting the load shed for the following customers and conditions:

Event hour from 2 – 3 PM (event hour 15); old customers from Sierra; Temperature at that hour = T_h= 95 F; mean temperature from midnight to 8 AM = T_8= 55 F. The customers are not in the Greater

Fresno Area (so δ_{Fresno}= 0) or the Greater Bay Area ($\delta_{Bay} = 0$), and the hour is not hour 18 ($\delta_{oh18} = 0$), but they are old customers so δ_{old}= 1.

Filling in the terms, we find:

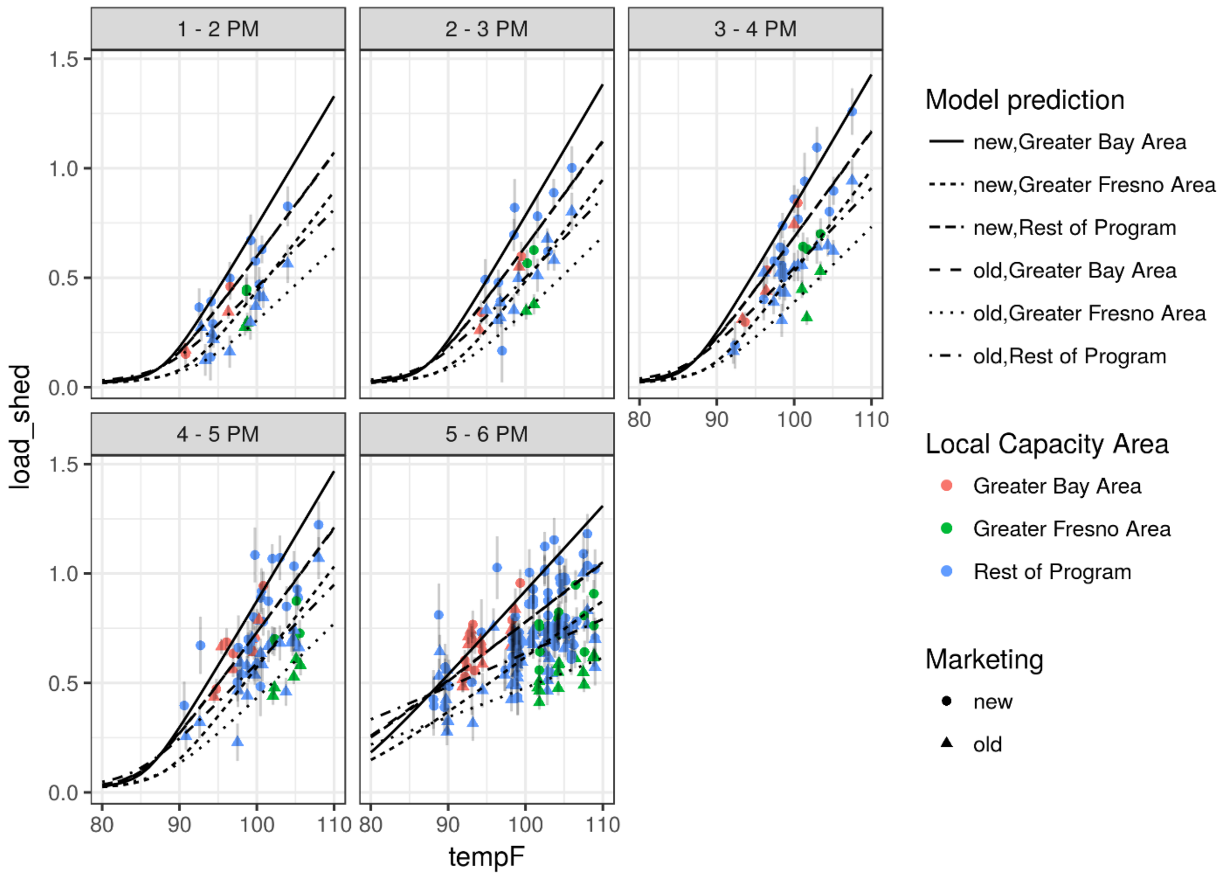
When we fill in the numbers for the terms above, in the same order as the table, the resulting prediction is:

$L_h =$ -3.989 + 0.042 * 95 + 0.005 * 55 + 1.040 – 0.011 * 95 = 0.271 kW.

Both the basic purely linear model and the final purely linear model generate negative predicted load shed for temperatures below about 85 F. Since the SmartAC™ program only restricts air conditioning load if the air conditioner is running more than half the time, one would expect that for any given customer there will be a temperature below which the air conditioner is in use but the SmartAC™ program has no effect because the air conditioner was running less than half the time anyway. As the temperature decreases more and more customers will fall into this category, so the load shed will approach zero. But there are surely many customers whose air conditioners are still running more than half the time at temperatures of 85 F or below. Lacking data that explore those lower temperatures, we applied a low-temperature adjustment that has the right qualitative form and has nearly no effect on the predictions for temperatures that were experienced in the 2017 events. Applying this adjustment to the 'basic purely linear' model and the 'final purely linear model' result in the 'basic' model and the 'final' model. Applying both the LCA-specific terms and the low-temperature term results in the 'final' model.

Figure 47 shows the resulting prediction for the final model. The low-temperature adjustment has a noticeable effect on only a few of the ex ante predictions, specifically the 1-in-2 predictions for May and October, which have relatively low temperatures. Without the adjustment those predictions would be too low. Since we have no information about the actual curve at temperatures below about 90 F, there is additional uncertainty in the ex ante predictions for those months and weather years: plausible curves lead to ex ante impact estimates that are about 10% higher or lower than the impacts we report.

Figure 47: Ex Post data, with model predictions



Returning to the discussion of the LCA-specific terms: The mean residual by LCA for the final model is shown in the right-hand panel of Figure 46. Addition of the LCA-specific terms eliminates the mean residual in the Greater Bay Area and the Greater Fresno Area, and also reduces the mean residual in the other LCAs: without those terms, the model was forced to compromise, over-predicting in some LCAs and under-predicting in others. In principle additional LCA-specific terms could be added in order to reduce or eliminate the mean residual in the other LCAs as well, but adding so many additional parameters would likely lead to 'overfitting', i.e. fitting noise rather than signal, especially for the LCAs with relatively few customers.

Figure 48 shows the ex ante predictions from the basic model and the final model, for the standard weather years. The only difference between the models is the handling of the Greater Bay Area and the Greater Fresno Area: the basic model treats all local capacity areas the same, whereas the final model fits different temperature slopes to these LCAs, as seen in Figure 47. (The Greater Bay Area also gets a different vertical offset). As the figure illustrates, the two models make nearly the same predictions for the hotter PG&E 1-in-2 months, because the temperatures in that weather year are close to what was actually experienced in the 2017 events and are therefore well fit on average by either model. But there is a large difference between the models when extrapolating to either cooler or hotter temperatures than were experienced in 2017: the final model is more temperature-sensitive, yielding lower predicted impact in cooler conditions (the CAISO 1-in-2 weather year) but higher predicted impact in hotter conditions (PG&E 1-in-10). This distinction may be relevant when comparing across report years as well: previous years did not use LCA-specific terms in their models.

Figure 48: Comparison of Basic and Final model

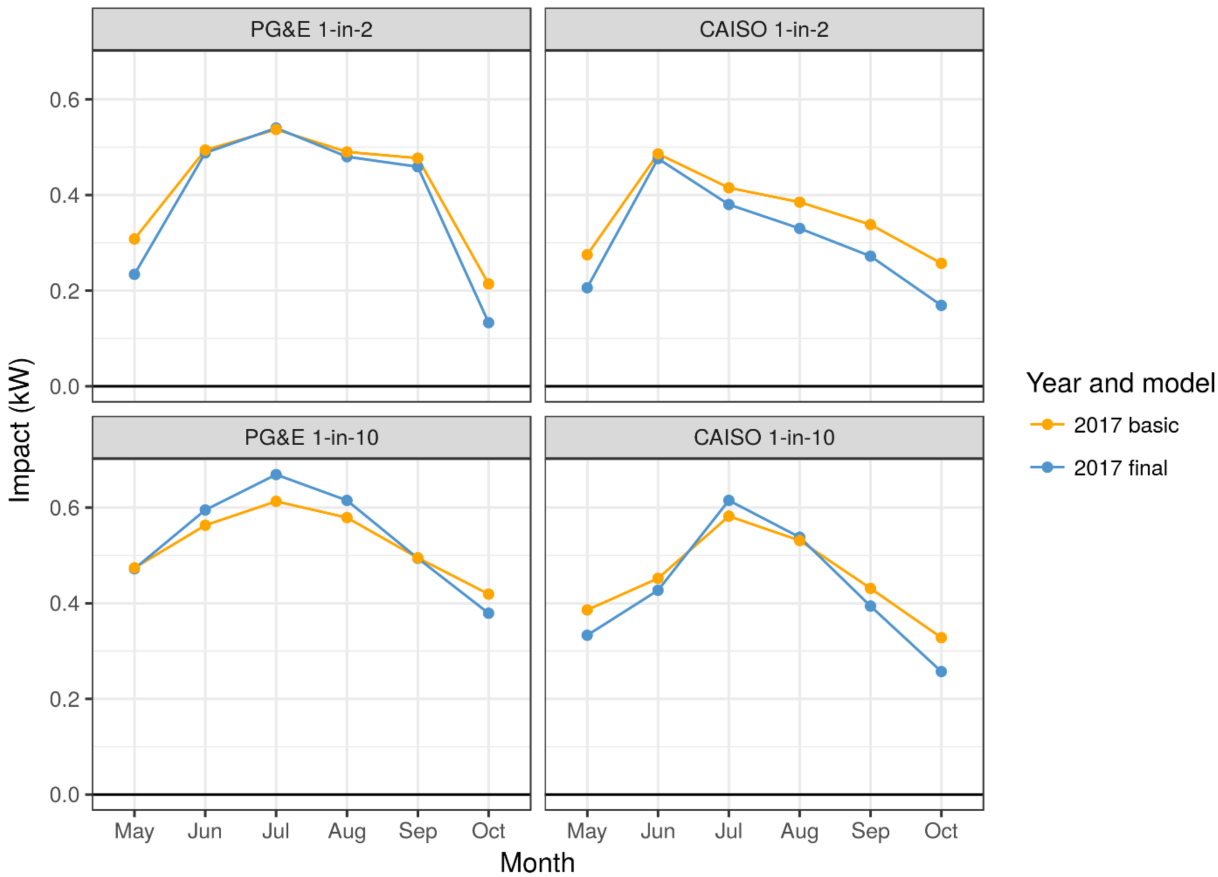
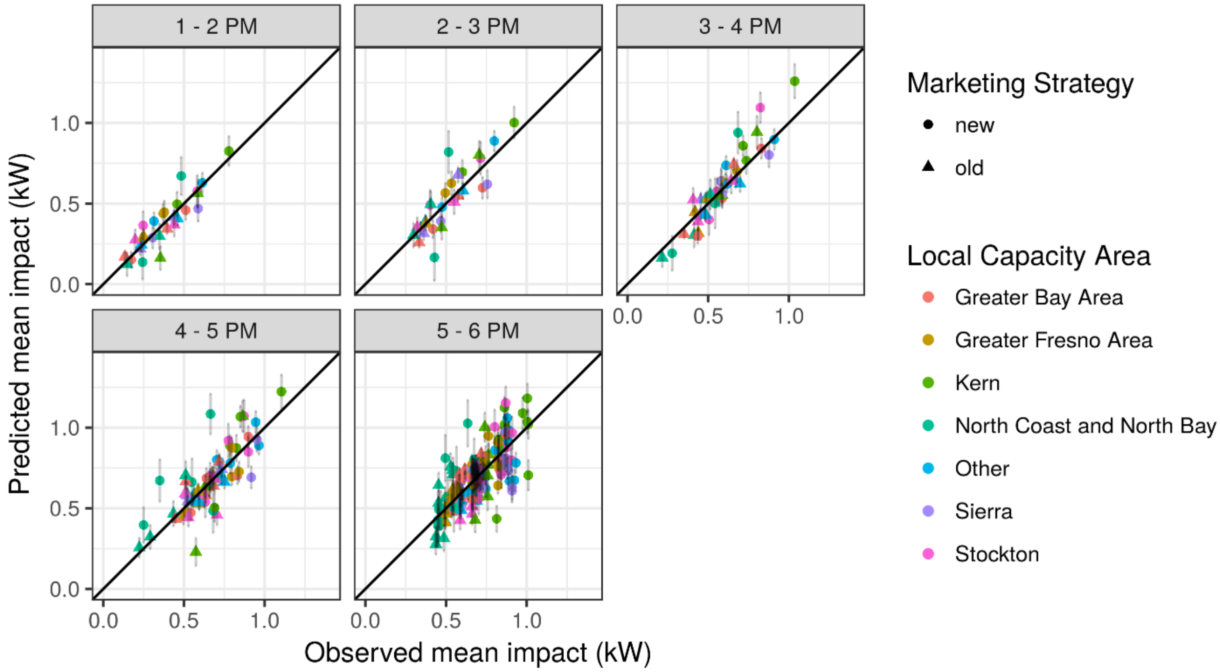


Figure 49 shows the ex post load shed versus the predicted load shed from the final model. This plot gives an idea of how accurately the model can predict the load shed for a given event in a given LCA in a given hour. There is some uncertainty in the ex post numbers, as indicated by the error bars on the data points, so even if the model fit perfectly the observed data points would not fall on the 45-degree line.

Figure 49: Load shed vs predicted load shed from final model



Snapback model

The main quantities of interest in the ex ante analysis are the predicted impacts in the standard event window from 1 to 6 PM, so that is where the bulk of the modeling and analysis effort was expended. Simple reference load and snapback models were implemented.

The snapback model predicts the load impact in the snapback period (after 6 PM) as follows:

$$L_t = \gamma_t L_{18} + \theta T_{75} t + \epsilon$$

where:

t is the number of hours after the end of the event;
L_{18} is the load at the end of the event (6:00, or 18:00 hours);

γ_t is a constant, different for each hour t,

θ is a constant with units of kW per degree F;

T_{75} is a temperature term equal to (temperature – 75 F) if the temperature exceeds 75 F, and 0 otherwise.

Term	Coeff.	Std. Err.	p-value	p-value flags: p < 0.1 : * p < 0.01: ** p < 0.001: ***
γ_t	2.956e-01	8.323e-03	< 2e-16	***
γ_t	-3.703e-01	8.840e-03	< 2e-16	***
γ_t	2.127e-01	8.960e-03	< 2e-16	***
γ_t	1.200e-01	9.052e-03	< 2e-16	***
γ_t	-6.962e-02	9.462e-03	2.62e-13	***
θ	3.561e-04	8.419e-05	2.43e-05	***

Metrics of model Fit:
Residual standard error: 0.05577 kW
Multiple R-squared: 0.7115, Adjusted R-squared: 0.7107

Reference load model

The model that predicts reference loads in the ex ante timeframe is structured as:

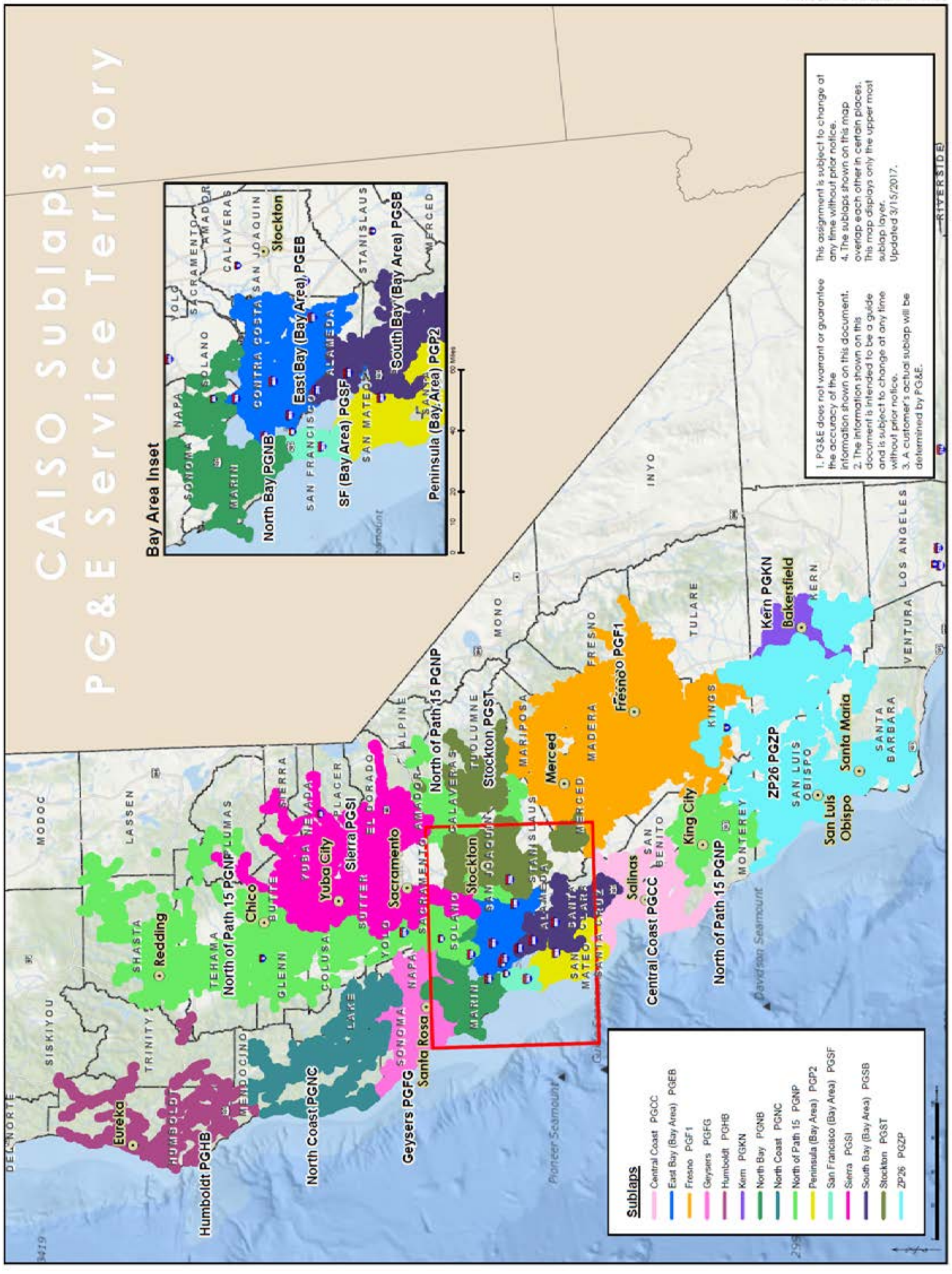
$$L_h = \alpha_h \delta_h + \beta_T T_h + \alpha_{old}\delta_{old} + \alpha_{dual}\delta_{dual} + \alpha_{dual,old}\delta_{dual,old} + \epsilon$$

where all of those parameters have the same meaning as in the final model, and

$\alpha_{dual,old}$ is a parameter (units of kW) representing an additional offset for dual-enrolled customers who were enro lled under the old marketing strategy.

We do not tabulate all 51 model parameters here, but summary statistics are as follows:
Residual standard error: 0.1193
Multiple R-squared: 0.9876, Adjusted R-squared: 0.9876

Appendix C: CAISO Sub-LAPs for PG&E Service Territory



This image was taken from the PDF map of the new sub-LPAs as of 2017 (accessed 3/2018). https://www.pge.com/pge_global/common/pdfs/save-energy-money/energy-management-programs/demand-response-programs/2018-demand-response/2018-demand-response-auction-mechanism/PGE-Sub-Lap-Map-201703.pdf