REPORT

Nexant

Reimagine tomorrow.

T&D Third-Party Bring Your Own Thermostat Demand Response Pilot Evaluation

Public Version- Final Report

February 8, 2017

Prepared for

Pacific Gas and Electric

Prepared by

Eric Bell, Ph.D.
Managing Consultant

Trevor Cummings
Senior Analyst

Joshua Schellenberg
Vice President

Amanda Stansell
Project Analyst

Michael Sullivan, Ph.D.
Senior Vice President

Nexant, Inc.

1 Executive Summary

Starting in the summer of 2016, PG&E partnered with three technology firms—Vendor 1, Vendor 2, and Vendor 3— to develop the third-party bring your own thermostat (BYOT) pilot designed to leverage already installed smart thermostat technologies as a demand response (DR) resource that could be ramped up in a relatively short amount of time and provide load relief to specific capacity-constrained areas on the grid. The BYOT pilot was part of PG&E's Demand Response Transmission and Distribution pilot and co funded by the Emerging Technologies (DRET) program. Pilot participants were recruited by the three vendors within eight PG&E substation footprints deemed to have local capacity constraints. Customers were offered varying financial incentives by the thermostat manufacturers in exchange for allowing their thermostat to be temporarily setback on event days. This evaluation documents the program design and implementation by each of the three vendors, and provides load impact estimates across the four events that took place in 2016.

1.1 Vendor Recruitment

Among the three vendors, Vendor 1 is the only third-party that also manufactures devices. Vendor 2 and Vendor 3 provide a software platform to various other third-parties who manufacture smart thermostats such as Thermostat 2, Thermostat 3, and Thermostat 4 whose technology could be adopted through direct purchase or through other third-parties such as Third-Party 1 or Third-Party 2. The three vendors, and/or their partners, were responsible for managing customer relationships during the pilot. Vendor responsibilities included:

- recruiting participants from their existing customers (located within the targeted substation footprints);
- providing load curtailment notification prior to and during events;
- dispatching directly to the smart thermostat;
- responding to customer questions and concerns about the assessment; and
- providing compensation to participating customers when appropriate.

Vendor recruitment efforts resulted in very different numbers of recruited and enrolled customers. Table 1-1 presents the total number of recruited and enrolled customers by vendor. Vendor 1 recruited and enrolled the largest amount of customers (639 recruited and 503 enrolled). Vendor 3 and Vendor 2 recruited 64 and 77 respectively and enrolled less than 50 customers each.

The time required to complete the contracting process between PG&E and the three vendors resulted in a compressed timeline for customer recruitment. Interestingly Vendor 3 and Vendor 2 had approximately six weeks in the field to recruit customers while Vendor 1 used approximately two and a half weeks in the field for recruitment. So, Vendor 1 recruited approximately 10 times the number of customers as the other vendors in approximately half the time.

There were significant differences in the respective recruiting processes used by the different vendors (i.e., customer incentives and recruiting processes) and these almost certainly played a significant role in the effectiveness of recruiting efforts. Section 2.1 describes the different approaches in detail.

Table 1-1: Recruitment and Enrollment Counts by Thermostat Vendor

Thermostat Vendor	Recruited Customers	Enrolled Customers	Enrollment %
Vendor 1	639	502	78.6%
Vendor 2	64	39	60.9%
Vendor 3	77	48	62.3%

1.2 Load Impacts

Recruiting ended for the pilot on August 31, 2016; and the four pilot events from the 2016 season all occurred immediately thereafter in September. The average load reductions estimated across the four events for each thermostat vendor are reported in Table 1-2. In interpreting the results in Table 1-2, it is important to keep in mind that the load impact measurements for Vendor 2 and Vendor 3 thermostats are statistically unreliable because of the small number of these devices included in the pilot. So it is really not appropriate to compare the performance of the devices reported in this table. The results for Vendor 2 and Vendor 3 are provided for completeness only.

In the table, the column labeled "Load w/o DR" displays the hourly average per customer load in the absence of a BYOT event. The column labeled "Load w/DR" indicates hourly average per customer load during the BYOT events. The load impact is the difference between the Load w/o DR and Load w/ DR.

The results for the Vendor 1 thermostat are statistically robust and indicate that the average load reduction over all hours on pilot event days was .43[1] kW plus or minus .1 kW with 90% confidence. This is equivalent to ~19% load reduction on average. The load reduction for the Vendor 2 is very similar (.44 kW) – an ~18% reduction. The load impact for the Vendor 3 thermostat was estimated to be ~.16 kW. Both the Vendor 2 and Vendor 3 impacts are statistically significant – meaning that they are very likely greater than zero. However, the magnitude of these impacts is highly uncertain given the small number of observed thermostats in both cases. With less than 50 customers enrolled in both technologies, it is possible a few customers could have a large influence on the results and a larger sample could yield very different outcomes.

Table 1-2: Average Load Reduction per Customer Across All Event Hours

Thermostat Vendor	Load w/o DR (kW)	Load w/ DR (kW)	Impact (kW)	Impact (%)
Vendor 1	2.30	1.87	0.43	18.6%
Vendor 2	2.37	1.94	0.44	18.1%
Vendor 3	1.87	1.71	0.16	8.3%

Load impacts were also estimated for each of the five geographic regions as shown in Table 1-3. The Chico and Santa Rosa regions contained less than 60 customers each, so their estimates are not as

1 Results may be lower than typically seen due to end of season event timing, potentially causing a higher number of customers to not be in cooling mode. Vendor 1 reported average estimated load drop for participating devices of 0.92 kW, which is more in line with average load drops seen from other programs.

robust as the other regions. The Chico region produced the highest average hourly load impact. However, the results were positively biased[2] by approximately .5 kW. This bias is evident during the hours prior to the event where the control group load is approximately .5 kW higher than the treatment group load. Subtracting out this bias still leaves the Chico area with one of the largest impacts across the geographic regions at approximately .5 kW. The San Jose region provides the highest unbiased average percent impact and also had the second best recruitment performance. The San Jose region produced an average hourly impact of .53 kW and a percent impact of 20.9%. Stockton provided the next largest impact at .39 kW (16.3%), followed by Palo Alto at .21 kW (9.9).

Table 1-3: Average Load Reduction per Customer across All Event Hours by Geographic Region

Geographic Region	Load w/o DR (kW)	Load w/ DR (kW)	Impact (kW)	Impact (%)	Number of Participants
Chico	2.48	1.47	1.01[2]	40.7%	56
Palo Alto	2.12	1.91	0.21	9.9%	196
Santa Rosa	1.58	1.30	0.28	17.7%	54
San Jose	2.54	2.00	0.53	20.9%	155
Stockton	2.40	2.02	0.39	16.3%	111

1.3 Conclusions and Recommendations

Based on the successful recruitment and statistically significant load impacts, the objective of the BYOT pilot was met, and it has now been shown that smart thermostats can be leveraged to provide load relief to specific capacity-constrained areas. However, there was not equal performance among the vendors, and for AC load control such as this, geography matters.

As discussed in section 2.1, the vendors in the study used similar approaches to offering the DR program to customers, but got dramatically different results. Vendor 1's marketing approach produced about 10 times more participants than either of the other two vendors in about half of the time. It is impossible to say with confidence why this occurred. Plausible explanations include:

- Vendor 1 offered a significantly higher enrollment incentive than the other vendors (i.e., Vendor 1 = $60, Vendor 2 = $0 and Vendor 3 = $25); and
- The marketing efforts of the three vendors were similar in many respects (i.e., email solicitations), but the messages sent to customers differed in significant ways which may have influenced the response rates. For example, the advertising was done solely under the Vendor 1 brand and specifically said "get Paid $60" for participating. While, the advertising for the other vendors was co-branded with the thermostat manufacturers (adding visual and cognitive complexity); and either didn't offer incentives or offered them as a gift card (which literature shows are discounted by consumers on offer).

[2] The results are statistically significant; however, they are also biased. See footnote 13 for additional information.

As a practical matter, because PG&E pays a fixed price per delivered customer, it should be indifferent to the marketing effectiveness of alternative vendors. However, to the extent that PG&E is interested in encouraging vendors to participate, it should clearly communicate a comparison of the success of the recruiting efforts to all of the vendors to ensure that those that did not produce significant enrollment might be able to do so using marketing and support strategies more like those used by Vendor 1.

It is really only possible to precisely describe the load impacts of the Vendor 1 thermostat. Two findings about the load impacts of the Vendor 1 thermostat stand out as extremely important. First, over the four hour periods under study, the load impacts from controlling the Vendor 1 thermostats (i.e., ~.43 kW) were only slightly smaller than the load impacts obtained from the SmartAC program overall during the previous summer (i.e., ~.51 kW). This suggests that as the market penetration of Vendor 1 thermostats increases or the geographical scope of the program is expanded beyond the pilot substation areas, the program could be a significant demand response resource.

However, like most thermostat load control programs that rely on setback and temperature ramping strategies, the load impacts from controlling the Vendor 1 thermostat diminish as the duration of the load control event increases. In part, this decline in load impact results from the fact that about 40% of the thermostats that started the event did not complete it. That is, occupants opted out at some time during the event. In addition, while Vendor 1 did not provide detailed information about its thermostat control strategy, they state that their control algorithm takes account of the thermal performance of the building envelope to establish the level of pre-cooling and the AC cycling strategy on a unit by unit basis. The declining load reduction as the events wear on suggests that the heat gain may be overcoming the impact of the cycling so that the air conditioner begins to cycle at its normal rate as the event proceeds.

It should be obvious that a diminishing impact with respect to operating time is not a desirable performance characteristic. Future program efforts should focus on improving the performance duration of load control from this technology or development of scheduling strategies that involve the operation of these devices for shorter periods of time when load relief is more critical. PG&E's SmartAC program leverages load control receiver (switch) technology that, utilizing an adaptive algorithm, delivers consistent hour over hour load reduction values. This level of reliability is important when considering locational dispatch or market integration.

The five different geographic regions included in the pilot each yielded different levels of performance in customer recruitment success and magnitude of load impacts. Chico and Santa Rosa yielded significantly fewer customers relative to the other regions and may benefit from increasing the incentive levels to customers in those areas. Chico (~.5 kW[3]) , San Jose (.5 kW), and Stockton (.39 kW) clearly provided the best value at nearly twice the average kW impact per customer compared to Palo Alto (.21 kW). Based on this difference in performance, it could be argued to only continue the pilot in the Chico, San Jose, and Stockton areas. However, conducting a cost effectiveness analysis would be helpful to determine the break even impact per customer required to make the program cost effective at a given level of per customer enrollment incentive. Therefore, should the BYOT pilot be continued for the 2017 season, it likely makes sense to continue the pilot in the Chico, San Jose, and Stockton areas, and potentially in the Palo Alto region if deemed cost effective.

[3] An impact of approximately .5 kW results after removing the observed bias of approximately .5 kW.

2 Overview of Regulatory Backdrop, Program Design, and Implementation

The third-party bring your own thermostat (BYOT) pilot is part of PG&E's Demand Response Transmission and Distribution Pilot and Emerging Technologies (DRET) program, and is an exploration of load management automation that integrates smart thermostats and their manufacturers into a residential automated demand response (ADR) program. The primary objective of the pilot was to understand the ramp up and potential of providing load relief of already installed smart thermostats to specific capacity-constrained areas on the grid. Pilot participants were recruited by the three vendors and their partners within eight PG&E substation footprints deemed to have local capacity constraints through a distribution operations planning process.

BYOT is an element of Phase 2 of PG&E's Transmission and Distribution Pilot.[4] BYOT also fulfills new requirements related to Assembly Bill (AB) 793 Section 717 and was included in PG&E's Advice Letter 3744-G-A/4886-E-A. However, it should be noted that the BYOT pilot is very new with an initial objective of implementing limited testing prior to the filing of the 2018-2022 DR application. Implications for future application in demand response focus on the value of such resources in third party offers under a competitive bidding environment, i.e. the Demand Response Auction Mechanism (DRAM), the Distribution Resource Plan (DRP), Integrated Distribution Energy Resources (IDER), and utility implemented DR resources such as a potentially expanded Capacity Bidding Program into the residential segment.

The BYOT pilot was event-based, meaning that it targeted relatively few hours on days of peak demand. Load reductions were attained on event days from temporary degree setbacks on thermostats, which lead to a reduction in demand for air conditioning, and overall system demand/load. Pilot participants were recruited from the thermostat vendors and their partners' existing customers with the offering of a financial incentive in some cases. Further details regarding the implementation of the pilot are contained in Section 2.1.

[4] PG&E's Transmission and Distribution Pilot was originally approved in the California Public Utilities Commission (CPUC) Decision (D.)12-04-045, and extended in (D.) 14-05-025.

2.1 Pilot Design and Implementation

Three, third-party thermostat suppliers -- Vendor 1, Vendor 2, and Vendor 3 – participated in the pilot. The vendors leveraged customers who had already adopted smart thermostat technologies and were collectively responsible for managing their respective customer relationships. This included recruiting participants from among customers, providing load curtailment notification prior to and during events, responding to customer questions/concerns about the assessment, and, in some cases, providing compensation. As shown in Table 2-1, Vendor 1 is the only third-party among these three vendors that also manufactures devices. Vendor 2 and Vendor 3 provide a software platform to various other third-parties who manufacture smart thermostats such as Thermostat 2, Thermostat 3, Thermostat 4, and Thermostat 5 whose technology could be adopted through direct purchase or through other third-parties such as Third-Party 1 or Third-Party 2.

Table 2-1: Vendors and Smart Thermostat Manufacturers from the Pilot

Contracted Vendor	Smart Thermostat Manufacturer
Vendor 1	Thermostat 1
Vendor 2	Thermostat 2
Vendor 3	Thermostat 3, Thermostat 4, Thermostat 5

Customer recruitment took place during the late summer once contract negotiations between PG&E and each of the three vendors were completed. Ultimately, due to the time required to complete the contracting process, the customer recruitment window was compressed in order to optimize the balance between having as many customers as possible, and still having enough hot summer weather remaining to hold several events. Vendor 2 and Vendor 3 had approximately six weeks in the field to recruit customers, launching their recruitment campaigns on July 13 and July 18, respectively. Vendor 1 started later (on August 15th) -- allowing only about three weeks to recruit its customers.

By design, the pilot called for the recruitment of 2,000 customers. At the end of the recruitment window on August 31, a total of 780 customers had accepted an offer by one of the vendors -- 639 from Vendor 1, 77 from Vendor 3, and 64 from Vendor 2. Not of the all customers who accepted offers were ultimately determined eligible to participate in the pilot. Table 2-2 provides a summary of eligibility screens that were applied to customers who had agreed to participate. Customers were screened by PG&E from participation if they were not located in the eight pilot substations, or had an address in the vendor's records that could not be matched to PG&E records, or were medical baseline customers; or were participants in other demand response programs.

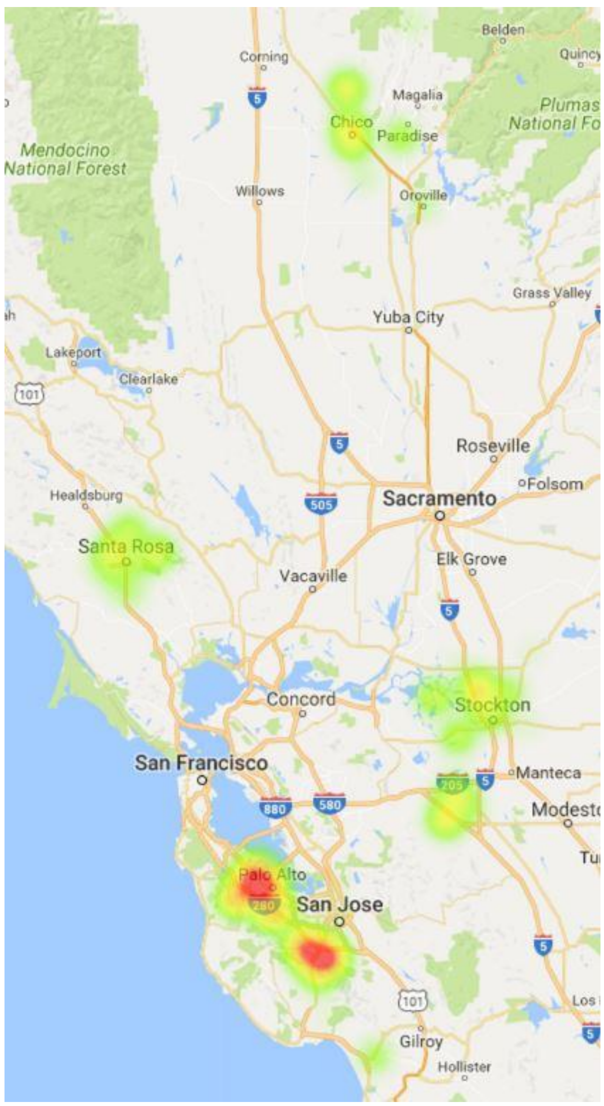
Table 2-2: PG&E Eligibility Screening Criteria

Criteria
Medical Baseline
Ineligible zip code
Invalid Account Number
No Address Match
Direct Participants with CAISO
SmartAC
SmartAC w/Thermostat
SmartAC w/Thermostat, SmartRate
SmartAC, EEsmartThermostat
SmartAC, SmartRate
SmartRate

The vendors were given a list of allowable zip codes in which to recruit customers. These zip codes were identified by PG&E as containing customers that were likely to have been served by the eight targeted substations experiencing local capacity constraints. The substations were chosen for the pilot in order to assess the ability to leverage smart thermostat technologies as a demand response (DR) resource that could be ramped up in a relatively short amount of time.

Figure 2-1 shows a heat map of locations of pilot participants throughout the PG&E service territory. The largest concentration of customers are in areas surrounding Palo Alto and the area south of San Jose as noted by the red areas on the map. The next largest concentrations of customers are found in the Santa Rosa and Stockton areas noted by yellow and green. Finally, there are also small clusters of participants around Chico and south of Stockton, east of Modesto, where Interstates 5 and 580 intersect.

Figure 2-1: Heat Map of Pilot Participant Location



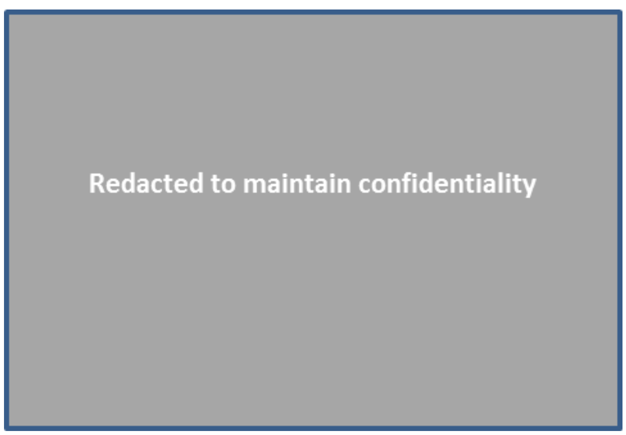
Each of the vendors had a unique program design with a different approach to marketing and customer recruitment, levels of customer incentive, acceptance rates, customer interaction, and thermostat setback strategy. It should be noted that each vendor received the same level of financial compensation per enrolled customers. However, as shown in the different levels of customer incentive, all had very different strategies regarding what to do with the money relative to their customers. The following three sections document the program design for each of the vendors and provide a record of how each of the different programs was implemented. In early November, well after the DR season, questionnaires were sent to each of the vendors requesting details on their marketing approach, program design, customer interface, support services, and overall experience. As of the writing of this report, two of the three vendors provided responses. The details from their responses are integrated into the implementation summary sections below and are the primary source of the content.

2.1.1 Vendor 1 Implementation Summary[5]

Marketing Approach

For recruitment, Vendor 1 sent an email solicitation to all existing Vendor 1 customers who resided in the zip codes associated with the eight substations chosen for the pilot. No other initial screening or specific targeting was applied. As shown in Figure 2-2, the solicitation emphasizes the financial incentive that participants could receive in the form of a $60 Amazon gift card for enrolling in the pilot. Recipients were provided with a link to enroll online. This messaging was similar to other messaging that had proved successful for other programs recruited for by Vendor 1 around the country and no message testing was employed during the short-term recruitment campaign.

Figure 2-2: Vendor 1 Recruitment Email



Recruitment was limited to a two week period in August due to the time required to complete the contracting process. Vendor 1 sent recruitment materials to approximately 7,000 customers in the targeted zip codes on August 15, followed by a second push to the same set of customers (minus those who had signed up) approximately two weeks later on August 29. The content of the two email waves

[5] Note: Large portions of the vendor's response to the questionnaire were integrated into this section.

were very similar. However, the second email was close to the sign up deadline, so key messaging was changed to convey that the deadline for signing up was approaching in order to not miss out.

Of the approximately 7,000 customers who were sent the emails, around 2,400 customers (34%) opened the email, about 1,000 (15%) of the customers clicked on the link, and 639 (9%) accepted the offer. Following the PG&E eligibility screens as noted above in Table 2-2, 502 customers (7%) were ultimately enrolled. More than 200 customers were found to be ineligible to participate in the pilot. Table 2-3 contains the full disposition of Vendor 1's recruitment effort, including a detailed breakout of the ineligible customers.

Table 2-3: Vendor 1 Recruitment Disposition

Vendor 1	Count
Enrollment Target	1,000
Recruitment Offers Made	7,000
# of Customers Accepting Offer	639
# of Customers Eligible/Enrolled	502
# of Customers Found Ineligible	213
Medical Baseline	*32*
Ineligible Zip	*11*
Invalid Account Number	*6*
No Address Match	*2*
Other	*16*
Direct participants with CAISO	*4*
SmartAC	*56*
SmartAC w/Thermostat	*4*
SmartAC w/Thermostat, SmartRate	*1*
SmartAC, EEsmartThermostat	*1*
SmartAC, SmartRate	*19*
SmartRate	*61*

When Vendor 1 was asked how they would improve their marketing approach based on their experience with PG&E in 2016, they responded by stating: "With this pilot we had a very short recruitment time frame (2 weeks). With a longer timeframe, a targeted retail outreach strategy has proven to be very successful[6]." They also stated they would not plan to change their incentive level, "However, if the program was paired and marketed with an up-front EE rebate (for purchasing the thermostat) it would drive significant ongoing program enrollments."

Customer Interface

During each of the four events Vendor 1 sent a notification to the participating thermostats via the internet. Vendor 1 is capable of accepting event requests up to two hours before the start of an event.

[6] Vendor 1 customer satisfaction survey results indicate they have implemented similar programs with at least nine other clients across the US.

Notification of the event was displayed on the thermostat and via the customer's smart phone app[7]. Each Vendor 1 learning thermostat has the potential to respond differently to an event rather than receiving a prescriptive adjustment or predetermined standard number of degrees temperature setback. To accomplish this, each thermostat builds a unique thermal model of the home within which it is operating to ensure that it is controlling the customer's HVAC system well and keeping them comfortable. For each event the device simply receives a notification that an event has been scheduled, and then the device will determine the best strategy for that customer. Depending on how well the home is insulated, that could mean that the AC would do an hour of pre-cooling ahead of the event and a certain level of AC cycling to balance total load reduction with customer satisfaction.

Customers have the option to override an event at any time by simply adjusting their thermostat. Vendor 1 tracks the number of devices that receive the initial event signal, and the number of customers who ultimately adjusted their thermostat. As part of their standard reporting, Vendor 1 provided PG&E the percentage of customers that started and completed the event (providing override information at an aggregated level). Vendor 1 also provided an Achieved Participation percentage which is a metric developed by taking the Actual Number of Minutes Participated over the Total Number of Minutes of the event from all customers. Details on these performance metrics are included in Section 4.1, Load Impacts for Vendor 1.

Program Support Services

Vendor 1 provided customers with a call center back-up in addition to email support in the event they had questions. No calls were received and only approximately three questions were received via email; all were related to when customers should expect to receive their enrollment incentive. No customer complaints were received via phone or email, and no customers dropped out of the program during the 1 month period when events were taking place.

Overall Experience and Outlook

As part of the questionnaire, the vendors were asked about their experience in working with PG&E, and any thoughts and recommendations they had regarding BYOT in future years. When asked about interactions with PG&E they said all communications were succinct and timely. Looking towards the future, Vendor 1 believes that due to the complexity and onerous nature of many market based and third party programs, such as the DRAM, there is an opportunity for PG&E to offer customers a simple BYOT program that would be immediately successful. Vendor 1 states there are more than one hundred thousand devices not currently participating in DR in PG&E's service area which could provide valuable local capacity, and leverage the increasing consumer demand for Vendor 1's products. They feel PG&E's successful history of customer-friendly DR programs would produce a very straightforward customer journey that aligns well with the DSM programs in its portfolio[8].

Vendor 1 also provided additional recommendations to PG&E if it were to pursue BYOT in 2017. They stated with PG&E's potential plan to include a DR incentive alongside an EE incentive, PG&E has an

---

[7] Approximately 90% of Vendor 1 customers pair their device with a Vendor 1 account so they would be able to control it remotely either through the web app or phone apps.

[8] An expansion of the program, particularly if it included an EE incentive to purchase a thermostat, would clearly be beneficial to Vendor 1's core business of selling thermostats.

opportunity to start expanding its BYOT program and hit the ground running in 2018. If an EE rebate were to be made available, Vendor 1 recommends:

- Test and study the impact of IDSM program outreach and messaging to capture the data needed to asses conversion rates for future program estimating/modeling;
- Run additional BYOT events to study load reduction / event performance data; and
- Continue to iterate on the customer journey and engagement touch points.

Vendor 1 suggests all of these would help PG&E refine their BYOT program plans and would be best learned through an expanded test program.

2.1.2 Vendor 2 Implementation Summary[9]

Marketing Approach

For recruitment, Vendor 2 sent Thermostat 2-branded solicitations to customers who resided in the target zip codes of the eight substations that were chosen for the pilot. No other targeting or segmentation strategy was used in determining the list of customers who would receive the offers. Vendor 2 sent multiple versions of its solicitation email in the first round of recruitment to test their individual efficacy, which can all be found in Appendix A. They A/B tested different email messages and landing pages; one set focused on a "savings" message – the other focused on an "eco-friendly/save-the-Earth" message. In the second and subsequent rounds of recruitment, Vendor 2 sent one recruitment email which is below in Figure 2-3. As seen in the figure, the solicitation emphasizes the potential savings of up to $50 a year by updating smart thermostat software to allow for automated demand response initiated by Vendor 2. However, it should be noted that no upfront financial incentive was offered. Recipients were provided with a link to enroll online.

[9] Note: Large portions of the vendor's response to the questionnaire were integrated into this section.

Figure 2-3: Vendor 2 Recruitment Email

Redacted to maintain confidentiality

Recruitment was conducted in three waves via email, with approximately three weeks between each email push. Approximately 2-3,000 customers were contacted in each wave, taking place on July 13, August 9, and August 26. 37% of the customers opened the email; about 10% of the customers clicked on the link, and 64 customers ultimately accepted the offer. Following the PG&E eligibility screens, 39 customers were enrolled. Table 2-4 contains the full disposition of Vendor 2's recruitment effort, including a detailed breakout of the ineligible customers.

Table 2-4: Vendor 2 Recruitment Disposition

Vendor 2	Count
Enrollment Target	500
Recruitment Offers Made	Unknown
# of Customers Accepting Offer	64
# of Customers Eligible/Enrolled	39
# of Customers Found Ineligible	25
Medical Baseline	*7*
Invalid Account Number	*3*
No Address Match	*2*
Other	*1*
Direct market participants with CAISO	*1*
SmartAC	*6*
SmartAC, SmartRate	*1*
SmartRate	*4*

When Vendor 2 was asked how they would improve their marketing approach based on their experience with PG&E in 2016, they responded by stating they would like to utilize multiple customer "touches" – feeling a need to go beyond a 3-part email campaign. They would also prefer to start soliciting much earlier. Additionally, they stated they only collaborated with Thermostat 2 on this pilot and are interested in expanding efforts for a subsequent program and include other OEMs, including OEM 1 and OEM 2 (each of which have their own marketing and enrollment strengths). They appreciated how flexible PG&E was in the enrollment process and found that ease for the customers was critical. For instance, they felt enrollments would have been lower with a Customer Information Service Request (CISR) form or other type of requirement that requires customers to provide additional information. Finally, Vendor 2 stated the promotional strategy for this pilot was for energy savings and a home energy scorecard. However, they would like to understand the optimal incentive rates that other vendors used, and could add that into a future program if PG&E recommends it.

Customer Interface

During events Vendor 2 sent a notification out to the participating thermostats through the internet. Notification of the event was displayed on the thermostat; however there were no other channels for customer notification of the event. The thermostats are programmed to set the temperature back by 4 degrees during events, which is different compared to the Vendor 1 thermostat that takes a more dynamic approach as noted in Section 2.1.1. Timing for pre-cooling before the event and changing set points during the event is determined by the specific parameters that are set and thermodynamics that are specific to the premise, such as house size, and relative to the weather conditions for the house during each event.

Like with the Vendor 1 device, customers have the option to override an event at any time by simply adjusting their thermostat. Vendor 2 tracked the number of devices that received the initial event signal, and the number of customers who ultimately adjusted their thermostat (opted out) during each event. As part of their standard DR event report, Vendor 2 provided PG&E the number of devices that received the event signal and the number of devices that were overridden.

Program Support Services

Vendor 2 provided customers with an FAQ and a "Contact Us" form, but did not provide call center support. Only two emails were received; one to drop out and a second from the same customer to confirm that he had been removed from the service— an event occurred after the request and before the customer could be removed. Vendor 2 did respond back to the customer with a confirmation. The reason provided by the customer for dropping out was that he was no longer interested in participating. No customer complaints were received.

Overall Experience and Outlook

As part of the questionnaire, the vendors were asked about their experience in working with PG&E, and any thoughts and recommendations they had regarding BYOT in future years. When asked about interactions with PG&E, Vendor 2 said they "were very pleased with the communication and contracting process with PG&E and wish we could replicate it with other utilities." Furthermore, "the implementation project and deployment communications were sufficient and timely."

Looking towards the future, Vendor 2 would be very interested in participating in a more broadly applied DR program. They advise it would be very helpful to start the implementation cycle prior to the DR season and to have more flexibility with territory, i.e., zip codes. They also suggest PG&E consider offering a Wi-Fi Thermostat Rebate program in conjunction with this program, if possible. Vendor 2 suggests that a rebate has a strong influence on increasing enrollment and participation rates, and is highly recommended as a consideration for future programs. Thermostat 2 also recommended that PG&E allow for a rebate mechanism to the customer. Thermostat 2 would be interested in participating customers getting a portion of the fee (i.e., financial incentive), but Thermostat 2 says it is very expensive for them to administer; and they suggest it would be far more cost effective if the utility could do it, due to their existing billing relationship with the customer.

2.1.3 Vendor 3 Implementation Summary

Marketing Approach

For recruitment, Vendor 3 sent one of three email solicitations to customers who resided in the area of the eight substations that were chosen for the pilot. Email solicitations were developed and sent to customers depending on the type of smart thermostat they owned. The Thermostat 3, Thermostat 4, and Thermostat 5 email solicitations can be found below in Figure 2-4**Error! Reference source not found.**, Figure 2-5, and Figure 2-6, respectively. As seen in the figures, the solicitations all emphasize the financial incentive that participants could receive in the form of a $25 incentive for enrolling in the program. Recipients were provided with a link to enroll online. Vendor 3 noted they tested multiple incentive structures. However, they stated the recruitment period was not long enough for deploying effective tests.

Figure 2-4: Vendor 3 Recruitment Emails- Thermostat 3

Redacted to maintain confidentiality

Figure 2-5: Vendor 3 Recruitment Emails- Thermostat 4

Redacted to maintain confidentiality

Figure 2-6: Vendor 3 Recruitment Emails- Thermostat 5

Redacted to maintain confidentiality

Vendor 3 sent recruitment materials to customers within the targeted zip codes beginning July 18, 2016 with a goal of recruiting 500 customers. Vendor 3 declined to provide any details regarding the number of customers who received offers, the number of email pushes, the email open rate, or the click through rate; stating the information was proprietary. At the close of the recruitment window on August 31st, 2016, 77 customers had accepted the offer. After PG&E applied the eligibility screens, 49 out of the 77 recruits (64%) were ultimately enrolled. Table 2-5 contains the outcome of the recruitment effort; including the break-down of the customers who accepted the offer but were found to be ineligible.

Table 2-5: Vendor 3 Recruitment Disposition

Vendor 3	Count
Enrollment Target	500
Recruitment Offers Made	Unknown
# of Customers Accepting Offer	77
# of Customers Eligible/Enrolled	49
# of Customers Found Ineligible	28
Medical Baseline	*9*
No Address Match	*3*
Other	*4*
SmartAC	*6*
SmartAC, SmartRate	*1*
SmartRate	*5*

When Vendor 3 was asked how they would improve their marketing approach based on their experience with PG&E in 2016, they responded by providing the following list:

- Broader geography;
- Easier process for co-branding;
- Longer recruiting period;
- Marketing performed by PG&E;
- Employ additional digital marketing tactics like search and social (if more time is available); and
- Message testing

They also responded that they would change their enrollment process to integrate PG&E branding. However, it should be noted that one of the items Vendor 3 has proposed is not necessarily in alignment with the principles of a competitive market and the future state of demand response under Rule 24 as envisioned by the CPUC. As a third- party program, the marketing is up to the vendor, and specifically not performed by PG&E.

Customer Interface

During each of the four events Vendor 3 sent a notification to the participating thermostats via the internet. Customers who have Third-Party 1 devices received an automatic email notification when an event was scheduled. All other thermostat partners did not automatically notify their customers in advance of an event. Customers without Third-Party 1 devices were only made aware an event was underway if they looked at the thermostat itself, or checked the mobile/web applications; depending on the device.

The thermostats are programmed to set the temperature back by four degrees during events, which is similar to the Vendor 2 thermostats, but different compared to the Vendor 1 thermostat that takes a more dynamic approach as noted in Section 2.1.1. Pre-cooling took place for 60 minutes before the event with the thermostats setting the temperature two degrees below normal. Like with the Vendor 1 and Vendor 2 devices, customers have the option to override an event at any time by simply adjusting their thermostat. Vendor 3 tracked the number of devices that received the initial event signal, and the number of customers who ultimately adjusted their thermostat (opted out) during each event. The number of devices that received the event signal and the number of devices that were overridden were made available to PG&E via Vendor 3's DRMS.

Program Support Services

Vendor 3 provided customers with an email address for support inquiries, but did not provide call center support. Only five emails were received; two inquiries regarding unenrollment, one from a customer outside of the target zip codes, and two regarding general program questions. No customer complaints were received, and one customer dropped out of the program following two of the four events. That customer stated they used super-efficient settings, so further curtailment didn't make sense.

Overall Experience and Outlook

As part of the questionnaire, the vendors were asked about their experience in working with PG&E, and any thoughts and recommendations they had regarding BYOT in future years. When asked about interactions with PG&E, Vendor 3 said "The contracting process was typical and did not encounter any issues." Furthermore, "PG&E was accommodating on aligning on the scope of work."

Looking towards the future, Vendor 3 suggests raising the incentive level provided by PG&E to the vendors; stating that the current incentive level is not ideal for recruiting customers. They suggest a higher customer enrollment incentive level could significantly improve program participation. Furthermore, they suggest a three to four year program time horizon would allow vendors to invest the appropriate amount in an upfront customer incentive in order to drive up recruitment.

Vendor 3 also provided additional recommendations to PG&E if it were to pursue BYOT in 2017. They welcome the opportunity to participate in a more broadly applied DR program in the future with PG&E; and provided the following suggestions as a way to improve the pilot to help create a successful more permanent program:

- Adopt one platform to manage all devices (thermostats and potentially other connected devices) in PG&E's territory. This makes recruiting simpler as the entire territory can be marketed to with one message and simplifies PG&E program management efforts as all thermostat manufacturers/service providers are managed under one contract and events are created and managed on one platform.
- Provide a higher incentive and/or longer contract period (e.g. 3 years) to enable vendors to provide customers with a larger upfront incentive.
- Allow for a longer time period for recruiting to allow time to test multiple marketing tactics and messages.
- Expand the program to the entire PG&E service territory.
- Deploy additional marketing directly from the utility and provide the ability for vendors to easily co-brand their program with the PG&E logo.

3 Load Impact Estimation Methodology

One of the primary objectives of the BYOT evaluation was to estimate the load reduction achieved during the four event days for BYOT customers segmented by thermostat vendor. This section summarizes the methodology used to estimate load impacts and the resulting load impacts for each thermostat vendor.

This evaluation used data from the recruited customers in the BYOT program. All of the BYOT customers were called for each of the four events. To estimate load impacts, it is necessary to estimate what energy consumption would have been if customers had not participated in an event: the counterfactual or reference load. This was not a randomized control trial, so it was necessary to construct a suitable control group using statistical matching. Once a suitable control group was created from a group of non-participants, the next step was to use a "difference-in-differences" approach to estimate load impacts.

Nexant selected an applicable control group from customers that did not participate in the BYOT program. Control customers were selected from a pool of non-participant customers that passed several filters that were also applied the BYOT treatment population.

Difference-in-differences helps to yield more precise estimates and can correct for unobserved differences between treatment and control (should they exist). This calculation was done using a fixed-effects regression methodology, which reduces the standard error of the estimates. The underlying approach for difference-in-differences is comprised of the following:

- Measure energy demand for both treatment and control customers on proxy (similar non-BYOT) days;
- Measure energy demand for both treatment and control customers on event days;
- Net out any pre-existing differences between the two groups on proxy days;
- Treatment effects are calculated by taking the difference between the treatment and matched control group in the BYOT event hours and subtracting any difference between the two groups in the event period hours on proxy days.

Additional details on the load impact estimation methodology including the selection of the matched control group and difference-in-differences regression model can be found in Appendix B.

4 Load Impacts

During September, four events were called. All four events ran for four hours and were called from 5 PM to 9 PM. Table 4-1 presents the event dates as well as the average temperatures during event hours and the maximum daily temperatures. The events were originally expected to be dispatched in tandem with SmartRate events, however due to the events not starting until September, the SmartRate program had already been called sufficient times so an alternative strategy for event triggers was developed. Events were based on the localized temperatures of weather stations[10] associated with the eight targeted substations. Table 4-2 contains the average event hour temperatures of the weather stations most closely associated with the BYOT participant population.

Table 4-1: Summary of Event Hours and Temperatures

Event Date	Event Start	Event End	Avg. Event Temperature (F)	Max Daily Temperature (F)
9/7	5:00 PM	9:00 PM	86	92
9/19	5:00 PM	9:00 PM	86	93
9/26	5:00 PM	9:00 PM	89	96
9/27	5:00 PM	9:00 PM	82	92

Table 4-2: Average Event Hour Temperature by Weather Station

Event Date	Average Event Temperature				
	Belmont	Chico	Cupertino	Santa Rosa	Stockton
9/7	84.3	90.1	84.3	86.9	92.5
9/19	84.5	92.6	82.5	84.8	95.0
9/26	86.8	88.0	88.8	88.6	93.5
9/27	79.4	87.6	79.6	76.4	92.3

4.1 Load Impacts for Vendor 1

Table 4-3 summarizes the average impacts for each event hour as well the hour preceding and following the event for Vendor 1 customers. The event hours are highlighted in gray. The impact in the first hour of the event was 0.84 kW and the percent impact was 37.2%. The kW impact decreased by at least 40% each hour as the event progressed. The impact in the final hour of the event was 0.14 kW and the percent impact was 6.4%. In the hour preceding the event, there is a sharp increase in load caused by pre-cooling before the event. Pre-cooling often takes place in events controlled by thermostats to decrease the temperature before the event starts. Additionally, there is “snap back” that occurs in the hour after the event ends. Snap back is defined as when customer energy usage is higher after an event

[10] The weather stations in Table 4-2 are based on the enrolled population and differ somewhat from the weather stations used for event triggers.

than what would be expected if an event had not taken place. An example of what can cause snap back is if the AC compressor has been curtailed from running during the event, the temperature inside the home is likely to have increased a few degrees. In order to return to the desired temperature set point after the event, the AC compressor has to run more consistently for up to the first hour following the event (or longer) in order to bring down the temperature. This can result in increased load in the hours following an event compared to what would typically be expected on a similar non-event day.

Table 4-3: Vendor 1 Average Hourly Load Reduction Per Customer

Hour Ending	Load w/o DR (kW)	Load w/ DR (kW)	Impact (kW)	Impact (%)
17	1.88	2.31	-0.43	-22.9%
18	2.26	1.42	0.84	37.2%
19	2.41	1.93	0.48	19.9%
20	2.33	2.07	0.25	10.9%
21	2.21	2.07	0.14	6.4%
22	1.97	2.36	-0.39	-19.8%

Figure 4-1 provides the average per customer load with DR, load without DR (reference load), and load impact for the average event day for Vendor 1 customers. The average load without DR during event hours was 2.30 kW. The average load with DR during event hours was 1.87 kW. This resulted in an average load reduction of .43[11] kW per customer, representing an 18.6% reduction relative to the reference load. In the first hour following the onset of load control, the load impact of the Vendor 1 thermostat was .84 kW. By the second hour the load impact dropped to .48 kW; by the third hour it dropped again to .25 kW and finally in the fourth hour it had dropped to .14 kW. In part, this decline in load impact results from the fact that about 40% of the thermostats that started the event did not complete it. That is, occupants opted out at some time during the event. In addition, while Vendor 1 did not provide detailed information about its thermostat control strategy, they state that their control algorithm takes account of the thermal performance of the building envelope to establish the level of pre-cooling and the AC cycling strategy on a unit by unit basis. The declining load reduction as the events wear on suggests that the heat gain may be overcoming the impact of the cycling so that the air conditioner begins to cycle at its normal rate as the event proceeds.

In between hours ending 10 and 13, which are outside of the event window, there is a difference between the load with and without DR which is likely due to the higher than usual percentage of customers that are net metered who are enrolled in this program. This anomaly outside of the event window is attributable to the small sample size in combination with the matching process used to select the control group. In a larger sample one would be far less likely to encounter such an anomaly. In the hour preceding the event there is pre-cooling, as the usage spikes prior to the event. The reference load

[11] Results may be lower than typically seen due to end of season event timing, potentially causing a higher number of customers to not be in cooling mode. Vendor 1 reported average estimated load drop for participating devices of 0.92 kW, which is more in line with average load drops seen from other programs.

at the hour before the event (hour ending 17) is 1.88 kW while the load with DR in the hour before the event was 2.31 kW. Additionally, there is snap back that occurs in the hour after the event. The average customer load with DR rises to 2.36 kW in the hour after the event while the reference load in the hour after the event is 1.97 kW.

Figure 4-1: Vendor 1 Average Hourly Load Impact per Customer on Average Event Day

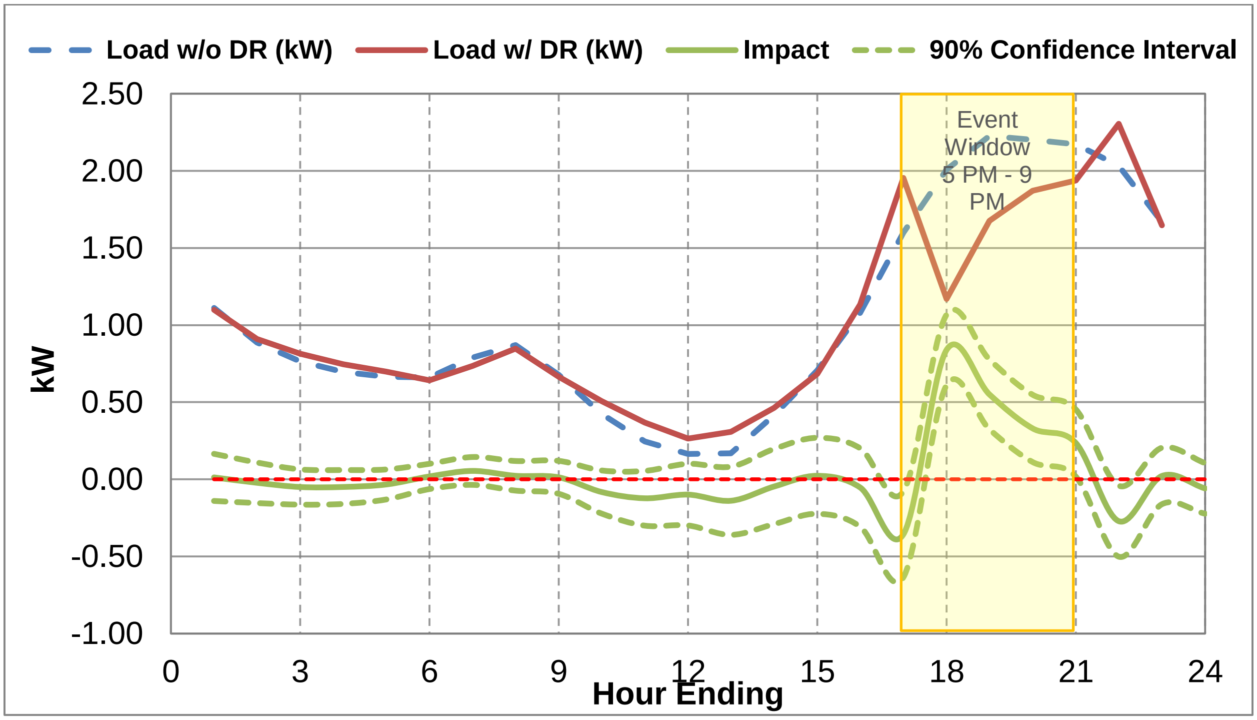


Table 4-4 provides load impacts for each event by hour for Vendor 1 customers. The event hours are highlighted in gray. Uncertainty adjusted impacts are provided in addition to the point estimates for the load impacts. The uncertainty adjusted impacts help provide context to the effect of relatively small sample sizes on the load impact calculations. For each event, the impact decreased as the event progressed; the final hour having the smallest impact. The most reasonable interpretation of this finding is that the thermostat temperature setback is overcome by the temperature in the home, so as the temperature rises the internal temperature gets closer to the temperature that has been set back and the appliance has to cycle just as much as it would have if the temperature had not been set back. This has been seen this with other temperatures setback strategies in the industry, and it requires the controller to keep advancing the temperature setting in order to maintain the load relief. The first event on September 7, 2016 had the largest percent impacts during each hour of the event compared to the same hours on the other event days. The average event impact percentage was 23.1% (0.49 kW) on this day.

Table 4-4: Vendor 1 Load Impacts for Each Event by Hour

Event Date	Hour Ending	Load w/o DR (kW)	Load w/ DR (kW)	Impact (kW)	Impact (%)	Weighted Temp. (°F)	Uncertainty-adjusted Impact - Percentiles				
							10th	30th	50th	70th	90th
9/7	17	1.60	1.96	-0.36	-22.3%	92.4	-0.57	-0.45	-0.36	-0.27	-0.14
	18	2.01	1.17	0.84	41.7%	91.9	0.66	0.77	0.84	0.91	1.02
	19	2.23	1.68	0.55	24.7%	88.9	0.37	0.48	0.55	0.62	0.73
	20	2.20	1.87	0.33	15.0%	84.2	0.16	0.26	0.33	0.40	0.50
	21	2.17	1.94	0.24	10.8%	79.6	0.07	0.17	0.24	0.31	0.41
	22	2.03	2.31	-0.27	-13.4%	76.3	-0.45	-0.35	-0.27	-0.20	-0.09
	Avg. Hourly	2.15	1.67	0.49	23.1%	86.2	0.32	0.42	0.49	0.56	0.67
9/19	17	2.03	2.39	-0.37	-18.0%	92.8	-0.59	-0.46	-0.37	-0.27	-0.14
	18	2.45	1.50	0.94	38.5%	91.7	0.74	0.86	0.94	1.02	1.14
	19	2.56	2.08	0.48	18.7%	88.3	0.28	0.40	0.48	0.56	0.68
	20	2.52	2.23	0.28	11.3%	83.7	0.10	0.21	0.28	0.36	0.47
	21	2.38	2.29	0.09	3.9%	80.5	-0.09	0.02	0.09	0.16	0.27
	22	2.11	2.49	-0.38	-17.9%	77.7	-0.56	-0.45	-0.38	-0.31	-0.20
	Avg. Hourly	2.48	2.03	0.45	18.1%	86.1	0.26	0.37	0.45	0.53	0.64
9/26	17	2.09	2.63	-0.55	-26.2%	95.4	-0.78	-0.64	-0.55	-0.45	-0.32
	18	2.45	1.65	0.80	32.7%	94.6	0.60	0.72	0.80	0.88	1.00
	19	2.57	2.13	0.45	17.3%	92.2	0.25	0.36	0.45	0.53	0.64
	20	2.44	2.25	0.18	7.5%	86.9	0.00	0.11	0.18	0.26	0.36
	21	2.30	2.20	0.10	4.3%	81.9	-0.07	0.03	0.10	0.17	0.27
	22	1.95	2.42	-0.47	-24.0%	78.6	-0.63	-0.53	-0.47	-0.40	-0.31
	Avg. Hourly	2.44	2.06	0.38	15.5%	88.9	0.20	0.31	0.38	0.46	0.57
9/27	17	1.80	2.25	-0.45	-25.3%	91.7	-0.66	-0.54	-0.45	-0.37	-0.24
	18	2.13	1.35	0.78	36.5%	89.3	0.61	0.71	0.78	0.85	0.95
	19	2.28	1.84	0.44	19.3%	84.6	0.27	0.37	0.44	0.51	0.61
	20	2.15	1.93	0.22	10.2%	78.6	0.06	0.15	0.22	0.29	0.38
	21	2.00	1.86	0.14	7.0%	75.0	-0.01	0.08	0.14	0.20	0.29
	22	1.77	2.22	-0.44	-25.0%	72.2	-0.60	-0.51	-0.44	-0.38	-0.29
	Avg. Hourly	2.14	1.75	0.40	18.3%	81.9	0.23	0.33	0.40	0.46	0.56

Table 4-5 contains a summary of how the Vendor 1 thermostats responded to the event notifications throughout each stage of the events. Across the 502 households, there were 644 thermostats. For the average event, the notification was sent out to 644 thermostats, and approximately 638 (99%) received the notification. Of the devices that received the notification, an average of 467 (73%) devices started the event. Only devices set to the "cooling mode" can start the event. Therefore, the difference in number of devices receiving the notification versus starting the event is attributable to thermostats either set to "away" or not in the "cooling mode." Of the devices that started the event, approximately 269 (58%) thermostats completed the event on average. The 58% average completion rate was certainly a contributing factor to the falling load impacts as the hours progressed. However, the overrides by themselves likely do not account for the entire difference observed between the first and last hours of the events.

As a point of comparison, the Vendor 2 average event completion rate was 64%. However, this difference is likely attributable to customer self-selection effects driven by the enrollment incentive decisions made by each vendor. The Vendor 2 customers received no financial incentive to participate, whereas the Vendor 1 customers received $60. Therefore, it is reasonable to conclude the two populations had different motivations to participate in the pilot and as a consequence would likely behave differently. While Vendor 1 did have lower event completion rates relative to Vendor 2, Vendor 1 also had ten times as many customers start the event, and even after factoring in the overrides, had nine times as many customers complete the event.

Table 4-5: Vendor 1 Device Level Status by Event Stage

Event	Sent	Received	Started	Completed
9/7/2016	645	633	440	271
9/19/2016	644	639	466	252
9/26/2016	644	639	482	264
9/27/2016	643	639	480	290
Average	**644**	**637.5**	**467**	**269.25**

4.2 Load Impacts for Vendor 2

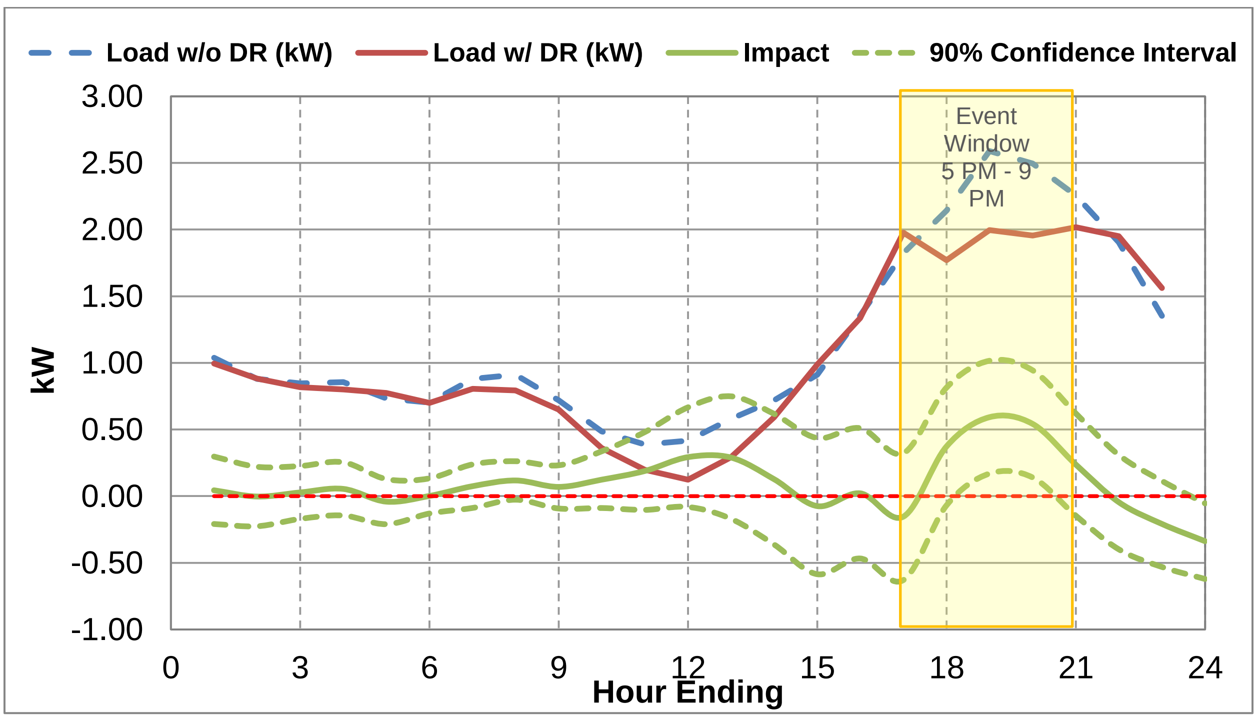
Table 4-6 summarizes the average impacts for each event hour as well the hour preceding and after the event for Vendor 2 customers. The event hours are highlighted in gray. It is important to distinguish the small sample size when looking at these results as there were only 39 customers enrolled by Vendor 2. Similar to Vendor 1 customers, the impact in the last hour of the event was the smallest. The impact in the first hour of the event was .37 kW and the percent impact was 17.4%. The impacts for the first three hours were statistically significant, but the final hour of the event was not statistically significant at the 95% level. In the hour preceding the event, there is a slight increase in load presumably caused by pre-cooling before the event.

Table 4-6: Vendor 2 Average Hourly Load Reduction Per Customer

Hour Ending	Load w/o DR (kW)	Load w/ DR (kW)	Impact (kW)	Impact (%)
17	1.82	1.98	-0.15	-8.4%
18	2.14	1.77	0.37	17.4%
19	2.59	2.00	0.59	22.9%
20	2.50	1.95	0.54	21.7%
21	2.25	2.02	0.24	10.5%
22	1.91	1.95	-0.05	-2.4%

Figure 4-2 provides the average per customer load with DR, load without DR (reference load), and load impact for the average event day for Vendor 2 customers. The average load without DR during event hours was 2.37 kW. The average load with DR during event hours was 1.93 kW. This resulted in an average load reduction of 0.44 kW per customer, representing an 18.1% reduction relative to the reference load. In between hours ending 10 and 13, there is a difference between the load with and without DR which is likely due to the higher than usual percentage of customers that are net metered who are enrolled in this program. The pre-cooling and snap back are relatively minimal in the hours before and after the event.

Figure 4-2: Vendor 2 Average Hourly Load Impact per Customer on Average Event Day



t of the four events at 13.1%. The smallest impact appears to be at least partially attributable to higher customer overrides on the hottest day. The topic of overrides will be discussed in further detail below.

Table 4-7 provides load impacts for each event by hour for Vendor 2 customers. The event hours are highlighted in gray. Uncertainty adjusted impacts are provided in addition to the point estimates for the load impacts. The impacts did not consistently decrease or increase as the event progressed across the four individual events. This inconsistency in the pattern of the load impacts is due to the small sample size being used to estimate load impacts. Similar to Vendor 1 customers, looking at the reference loads throughout the event period indicates that as the event progressed the baseline load was increasing for the first two hours before leveling out over the third or fourth of the event suggesting that there is more potential for load reductions in the middle two hours. This could also simply be a factor of the small sample size. The first event on September 7, 2016 had the largest average percent impact of 24.7%. Although the third event had the highest weighted temperature during event hours, the average percent impact was the smallest of the four events at 13.1%. The smallest impact appears to be at least partially attributable to higher customer overrides on the hottest day. The topic of overrides will be discussed in further detail below.

Table 4-7: Vendor 2 Load Impacts for Each Event by Hour

Event Date	Hour Ending	Load w/o DR (kW)	Load w/ DR (kW)	Impact (kW)	Impact (%)	Weighted Temp. (°F)	Uncertainty-adjusted Impact - Percentiles				
							10th	30th	50th	70th	90th
9/7	17	1.71	1.65	0.06	3.7%	92.9	-0.56	-0.19	0.06	0.32	0.69
	18	2.05	1.51	0.54	26.5%	92.2	-0.04	0.31	0.54	0.78	1.12
	19	2.43	1.73	0.70	28.7%	89.3	0.12	0.46	0.70	0.93	1.27
	20	2.48	1.67	0.81	32.6%	84.4	0.26	0.59	0.81	1.03	1.36
	21	2.21	1.97	0.24	11.0%	79.7	-0.31	0.02	0.24	0.47	0.80
	22	1.95	2.14	-0.19	-9.7%	76.2	-0.73	-0.41	-0.19	0.03	0.35
	Avg. Hourly	2.29	1.72	0.57	24.7%	86.4	0.01	0.35	0.57	0.80	1.14
9/19	17	2.32	2.07	0.25	10.7%	93.6	-0.45	-0.04	0.25	0.53	0.95
	18	2.51	1.81	0.71	28.1%	92.5	0.12	0.46	0.71	0.95	1.29
	19	2.93	2.25	0.68	23.1%	88.9	0.07	0.43	0.68	0.92	1.28
	20	2.68	2.29	0.39	14.6%	84.2	-0.22	0.14	0.39	0.64	1.00
	21	2.45	2.26	0.19	7.9%	80.7	-0.37	-0.04	0.19	0.43	0.76
	22	2.10	1.98	0.13	6.0%	77.9	-0.38	-0.08	0.13	0.33	0.63
	Avg. Hourly	2.64	2.15	0.49	18.4%	86.6	-0.10	0.25	0.49	0.74	1.08
9/26	17	1.76	2.01	-0.25	-14.2%	96.3	-0.90	-0.52	-0.25	0.02	0.40
	18	2.06	1.93	0.13	6.3%	95.0	-0.49	-0.12	0.13	0.38	0.75
	19	2.72	2.32	0.40	14.8%	92.3	-0.23	0.14	0.40	0.66	1.04
	20	2.56	2.19	0.36	14.2%	86.7	-0.21	0.13	0.36	0.60	0.94
	21	2.43	2.01	0.42	17.2%	81.5	-0.12	0.20	0.42	0.64	0.96
	22	1.85	1.92	-0.07	-3.6%	78.0	-0.53	-0.26	-0.07	0.12	0.40
	Avg. Hourly	2.44	2.11	0.33	13.1%	88.9	-0.26	0.09	0.33	0.57	0.92
9/27	17	1.51	2.19	-0.68	-44.8%	91.9	-1.33	-0.95	-0.68	-0.41	-0.02
	18	1.94	1.83	0.11	5.7%	89.5	-0.52	-0.15	0.11	0.37	0.74
	19	2.27	1.68	0.59	26.1%	85.1	0.09	0.39	0.59	0.80	1.09
	20	2.26	1.66	0.60	26.6%	78.6	0.13	0.41	0.60	0.79	1.07
	21	1.93	1.83	0.10	5.0%	74.5	-0.39	-0.10	0.10	0.30	0.59
	22	1.72	1.77	-0.05	-2.9%	71.8	-0.52	-0.24	-0.05	0.14	0.42
	Avg. Hourly	2.10	1.75	0.35	15.9%	81.9	-0.17	0.14	0.35	0.57	0.87

Table 4-8 contains a summary of how the Vendor 2 controlled thermostats responded to the event notifications throughout each stage of the events. Across the 39 households, there were 48 thermostats. For the average event, the notification was sent out to all 48 thermostats. Of the devices that were sent the notification, an average of 45 (93%) devices started the event. Of the devices that started the event, approximately 29 (64%) thermostats completed the event on average. While the 64% event completion rate is higher than the 58% from Vendor 1, on the hottest day the event completion rate fell to 55%. Given the Vendor 2 customers received no financial incentive to participate in the pilot, they may also be more likely to override on very hot days[12].

Table 4-8: Vendor 2 Device Level Status by Event Stage

Event	Sent	Started	Completed
9/7/2016	49	46	33
9/19/2016	48	45	29
9/26/2016	48	44	24
9/27/2016	48	44	31
Average	**48**	**45**	**29**

4.3 Load Impacts for Vendor 3

Table 4-9 summarizes the average impacts for each event hour as well the hour preceding and following the event for Vendor 3 customers. The event hours are highlighted in gray. Similar to Vendor 2, the sample size is a limiting factor in calculating the load impacts for Vendor 3. Only 48 customers were enrolled by Vendor 3. The impacts in the last two hours of the event were close to zero and not statistically significant. Additionally, only the first hour of the event was statistically significant. Given the extremely small sample size, the small and statistically insignificant load impacts don't necessarily imply that the Vendor 3 program wouldn't ultimately result in load impacts more similar to the other vendors if a larger population was available. The impact in the first hour of the event was 0.54 kW and the percent impact was 28.7%. In the hour preceding the event, there is a slight increase in load caused by pre-cooling. There is also snap back in the hour after the event ends.

Table 4-9: Vendor 3 Average Hourly Load Reduction Per Customer

Hour Ending	Load w/o DR (kW)	Load w/ DR (kW)	Impact (kW)	Impact (%)
17	1.62	1.92	-0.30	-18.5%
18	1.88	1.34	0.54	28.7%
19	1.86	1.71	0.15	8.1%
20	1.83	1.91	-0.08	-4.6%
21	1.90	1.89	0.02	0.9%
22	1.61	1.84	-0.23	-14.4%

Figure 4-3 provides the average per customer load with DR, load without DR (reference load), and load impact for the average event day for Vendor 3 customers. The average load without DR during event hours was 1.87 kW. The average load with DR during event hours was 1.71 kW. This resulted in an average load reduction of 0.16 kW per customer, representing an 8.3% reduction relative to the

[12] It should be noted these findings are only directional in nature, and a larger sample size with more events would be needed to further test this theory.

reference load. In the hour before the event, there is some pre-cooling and in the hour after the event there is some snap back.

Figure 4-3: Vendor 3 Average Hourly Load Impact per Customer on Average Event Day

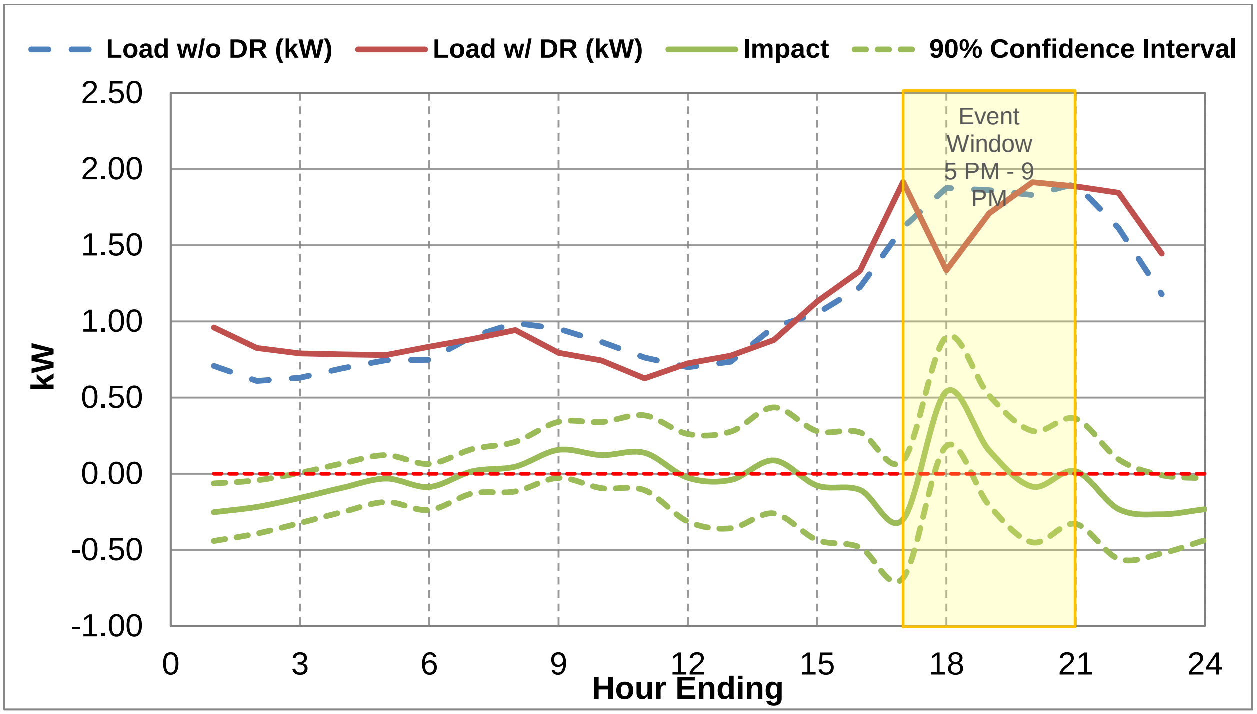


Table 4-10 provides load impacts for each event by hour for Vendor 3 customers. The event hours are highlighted in gray. Uncertainty adjusted impacts are provided in addition to the point estimates for the load impacts. Similar to Vendor 2, the impacts for Vendor 3 did not consistently decrease or increase as the event progressed across the four individual events. Again, this inconsistency in the pattern of the load impacts is due to the small sample size being used to estimate load impacts. The average hourly percent impact ranges from -8.7% to 19.7% between the four events. The second event had the largest average percent impact. Most of the impacts are not statistically significant at this granular of a level. Overall, the load impacts are the smallest of the three vendors.

Table 4-10: Vendor 3 Load Impacts for Each Event by Hour

Event Date	Hour Ending	Load w/o DR (kW)	Load w/ DR (kW)	Impact (kW)	Impact (%)	Weighted Temp. (°F)	Uncertainty-adjusted Impact - Percentiles				
							10th	30th	50th	70th	90th
9/7	17	1.20	1.53	-0.33	-27.6%	92.1	-0.81	-0.53	-0.33	-0.13	0.15
	18	1.49	1.14	0.35	23.5%	91.9	-0.07	0.18	0.35	0.52	0.77
	19	1.53	1.60	-0.07	-4.5%	90.1	-0.56	-0.27	-0.07	0.13	0.43
	20	1.62	1.86	-0.24	-14.6%	86.8	-0.73	-0.44	-0.24	-0.03	0.26
	21	1.84	1.83	0.01	0.5%	82.9	-0.46	-0.18	0.01	0.20	0.48
	22	1.70	1.92	-0.22	-12.8%	79.4	-0.72	-0.42	-0.22	-0.01	0.28
	Avg. Hourly	1.62	1.61	0.01	1.2%	87.9	-0.46	-0.18	0.01	0.21	0.49
9/19	17	1.81	2.00	-0.20	-11.0%	95.3	-0.73	-0.41	-0.20	0.02	0.33
	18	2.11	1.36	0.76	35.9%	94.7	0.28	0.56	0.76	0.95	1.23
	19	2.01	1.60	0.41	20.6%	92.2	-0.06	0.22	0.41	0.61	0.89
	20	2.10	1.72	0.38	17.9%	88.0	-0.11	0.18	0.38	0.57	0.86
	21	2.00	1.91	0.08	4.2%	83.7	-0.39	-0.11	0.08	0.28	0.56
	22	1.76	1.93	-0.17	-9.6%	80.1	-0.65	-0.37	-0.17	0.03	0.32
	Avg. Hourly	2.06	1.65	0.41	19.7%	89.7	-0.07	0.21	0.41	0.60	0.89
9/26	17	1.83	2.15	-0.32	-17.8%	97.2	-0.87	-0.55	-0.32	-0.10	0.22
	18	1.96	1.43	0.53	26.9%	96.3	0.05	0.33	0.53	0.72	1.00
	19	2.13	1.79	0.34	15.8%	94.3	-0.16	0.13	0.34	0.54	0.83
	20	2.05	1.98	0.07	3.4%	89.6	-0.47	-0.15	0.07	0.29	0.60
	21	2.23	1.92	0.31	13.8%	84.0	-0.20	0.10	0.31	0.52	0.82
	22	1.55	1.82	-0.27	-17.1%	80.3	-0.69	-0.44	-0.27	-0.09	0.16
	Avg. Hourly	2.09	1.78	0.31	15.0%	91.1	-0.20	0.10	0.31	0.52	0.81
9/27	17	1.64	1.97	-0.34	-20.7%	93.7	-0.88	-0.56	-0.34	-0.12	0.20
	18	1.94	1.42	0.52	26.8%	92.1	-0.01	0.30	0.52	0.74	1.05
	19	1.77	1.85	-0.08	-4.4%	89.0	-0.56	-0.28	-0.08	0.12	0.41
	20	1.55	2.09	-0.55	-35.4%	83.2	-1.04	-0.75	-0.55	-0.35	-0.06
	21	1.55	1.89	-0.34	-21.7%	79.3	-0.76	-0.51	-0.34	-0.16	0.09
	22	1.44	1.72	-0.28	-19.3%	76.3	-0.67	-0.44	-0.28	-0.12	0.11
	Avg. Hourly	1.70	1.81	-0.11	-8.7%	85.9	-0.59	-0.31	-0.11	0.09	0.37

4.4 Load Impacts by Geographic Region

The vendors and their partners targeted and enrolled customers in the allowable areas identified by PG&E as being served by the eight targeted substations experiencing local capacity constraints. The pilot participants reside in five distinct geographic regions. Table 4-11 presents the distribution of the pilot participants across the different geographic regions and vendors. Each of the five geographic regions in the table encompasses several zip codes in close proximity to one another. The Palo Alto, San Jose, and Stockton regions have the highest concentrations of pilot participants.

Table 4-11: Distribution of Customer Counts across Geographic Region and Vendor

Thermostat Vendor	Customer Count				
	Chico	Palo Alto	Santa Rosa	San Jose	Stockton
Vendor 1	47	180	46	145	80
Vendor 3	5	7	5	5	26
Vendor 2	8	12	5	8	5
All	60	199	56	158	111

Table 4-12 summarizes the average impacts for the average event hour for each of the five distinct geographic areas. It also details how many pilot participants are included in each region. These estimates include customers from all three thermostat vendors. The Chico and Santa Rosa regions contain less than 60 customers each, so their estimates are not as robust as the other regions. The Chico region produced the highest average hourly load impact. However, the results were positively biased[13] by approximately .5 kW. This bias is evident during the hours prior to the event where the control group load is approximately .5 kW higher than the treatment group load. Subtracting out this bias still leaves the Chico area with one of the largest impacts across the geographic regions at approximately .5 kW. The San Jose region provides the highest unbiased average percent impact and also had the second best recruitment performance. The San Jose region produced an average hourly impact of .53 kW and a percent impact of 20.9%. Stockton provided the next largest impact at .39 kW (16.3%), followed by Palo Alto at .21 kW (9.9%). The impact size in kW and percentage terms aligned with the relative size of the reference load in each region; with San Jose having the largest reference load of 2.54 kW, while Stockton had a reference load of 2.4 kW and Palo Alto had the smallest reference load of 2.12 kW.

13 The results are statistically significant; however, they are also biased. This can happen when there are unobserved differences between the treatment and control group. One example of the potential differences could be customers with different sized solar installations. Bias is more likely to occur with small sample sizes when it's difficult to find matches for customers and single customers may have a significant influence on an entire group. When removing the bias of approximately .5 kW, the remaining impact equals approximately .5 kW.

Table 4-12: Average Load Reduction per Customer across All Event Hours by Geographic Region

Geographic Region	Load w/o DR (kW)	Load w/ DR (kW)	Impact (kW)	Impact (%)	Number of Participants
Chico	2.48	1.47	1.01[13]	40.7%	56
Palo Alto	2.12	1.91	0.21	9.9%	196
Santa Rosa	1.58	1.30	0.28	17.7%	54
San Jose	2.54	2.00	0.53	20.9%	155
Stockton	2.40	2.02	0.39	16.3%	111

As noted previously, the primary objective of the pilot was to determine whether this type of resource could be leveraged to address locational constraints. In order to be successful, vendors and their partners needed to be able to recruit a sufficient customer base, and those recruited customers then needed to provide significant load impacts. Based off the recruitment results in Table 4-11, the Chico and Santa Rosa areas lacked sufficient customer response given the marketing strategies that were applied during the pilot. There may be some nuances regarding the size of the existing customer bases in each region, but Chico and Santa Rosa yielded significantly fewer customers relative to the other regions and should likely be omitted from future implementations.

The remaining three regions each have their merits with respect to customer recruitment success and/or load impact size. As noted above, San Jose had the largest impacts, yet Palo Alto had the highest customer recruitment success. When combining the average customer impact with the respective customer counts, the aggregate kW load impact in each region can be calculated. This metric provides a cumulative value that each of the regions is providing; factoring both the customer recruitment success and the average impact magnitude. Table 4-13 shows the aggregate load impacts, where San Jose had the largest impacts at 84 kW, and Stockton and Palo Alto had very similar aggregate impacts to one another at 43.3 kW and 41.8 kW, respectively. Chico had an aggregate impacts of 56.6 kW, but that should be divided in half to account for the bias in the estimation; resulting in an aggregate impact of approximately 28 kW. The similar outcomes from Stockton and Palo Alto illustrate how load impact size can help to offset lower customer recruitment, and vice versa.

Table 4-13: Aggregate Average Hourly Load Reduction by Geographic Region

Geographic Region	Impact per Customer (kW)	Number of Participants	Aggregate Impact (kW)
Chico	1.01[13]	56	56.6
Palo Alto	0.21	196	41.8
Santa Rosa	0.28	54	15.1
San Jose	0.53	155	83.7
Stockton	0.39	111	43.3

When factoring in the recruitment cost per customer, which was uniform across the geographic regions from PG&E's perspective, San Jose, Chico, and Stockton clearly provided the best value at nearly twice the average kW impact per customer. Based on this difference in performance, it could be argued to only continue the pilot in the San Jose, Chico, and Stockton areas. However, two factors should be taken into consideration before drawing such conclusions. First, it's possible that the cost per kW may be below PG&E's comparable capacity costs even with the .21 kW per customer impacts in Palo Alto. Conducting a cost effectiveness analysis would be helpful to determine the break even impact per customer required to make the program cost effective at a given level of customer incentive. Second, with small sample sizes as noted above, there is a level of uncertainty around the size of the load impacts. It is possible with a larger recruitment effort in the future that the impacts in Palo Alto could increase, or the impacts in the other areas may be lower; all based on which customers choose to accept the offer. Therefore, it likely makes sense to continue the pilot in the San Jose, Chico, and Stockton areas, and it may make sense to continue the pilot in the Palo Alto region if deemed cost effective.

5 Conclusions and Recommendations

The primary objective of the BYOT pilot was to determine the viability of leveraging already installed smart thermostat technologies as a DR resource that could be ramped up in a relatively short amount of time, and provide load relief to specific capacity-constrained areas on the grid. In order to be successful, this meant the vendors and their partners needed to be able to recruit a sufficient customer base, and those recruited customers then needed to provide significant load impacts. The results from the pilot show it is clearly possible to successfully recruit customers in targeted areas. However, there was wide variation in recruitment success rates among the vendors, and significant differences in recruitment rates by geographic region as well. Ultimately, all three vendors successfully provided load reductions. But, it isn't possible to compare performance between vendors due to the small sample sizes. Similarly to the recruitment, impacts also varied significantly by geographic region. Based on the successful recruitment and statistically significant load impacts, the objective of the BYOT pilot was met, and it was shown that smart thermostats can be leveraged to provide load relief to specific capacity-constrained areas. However, there was not equal performance among the vendors, and geography matters.

Vendor 1 customers provided consistent and statistically significant load impacts across the four events with an average percent load reduction of 18.6% and an absolute load reduction of 0.43 kW. However, like other thermostat load control programs, the load impacts from controlling the Vendor 1 thermostat diminish as the duration of the load control event increases. The declining load reduction as the events wear on suggests that the heat gain may be overcoming the impact of the cycling so that the air conditioner begins to cycle at its normal rate as the event proceeds. It may also be due to the setback strategy of the vendor; which wasn't explicitly provided. Vendor 1 stated that "each Vendor 1 learning thermostat has the potential to respond differently to an event rather than receiving a prescriptive adjustment or predetermined standard number of degrees temperature setback." Based on this statement, it's not clear exactly what the setback strategy is for the entire event, nor is it clear how the vendor balances customer comfort and the cycling strategy, or how that strategy might change throughout the event to balance customer comfort and load impacts. However, the evidence observed from the load impacts points to some sort of customized descending temperature set back strategy which could account for at least some of the diminishing load impacts.

The Vendor 2 and Vendor 3 customer load impacts are not statistically robust as they were constrained by the significantly smaller sample size of the customers they enrolled. While the Vendor 2 impacts appear to be similar in magnitude to the Vendor 1 results, both the Vendor 2 and Vendor 3 results should be viewed with caution as their 95% confidence intervals are extremely wide and in some cases actually contain zero or negative impacts. Consequently, it is only appropriate to state that both Vendor 2 and Vendor 3 customers appear to have responded to the events; however, a larger customer population is required in order to compare event performance across the three vendors.

From an implementation aspect, each of the vendors took different approaches in customer recruitment and support offered for any customer questions. All three vendors received the same financial compensation per enrolled customer. However, each had very different strategies with what to do with the money relative to their customers. Vendor 1 passed the most money through to its customers via a $60 incentive (Vendor 3 paid $ 25 and Vendor 2 paid $ 0) and was able to recruit ten times the customers in less than half the time of the other two vendors. However, the installed base of customers for each vendor and the number of customers being recruited by the other two vendors wasn't clear, [14] so it isn't possible to compare the actual customer offer acceptance rates between the vendors. It was clear that Vendor 1 is quite experienced in their customer recruitment efforts given they have implemented similar programs with at least nine other clients. Both Vendor 1 and Vendor 2 suggested that pairing a device rebate for a smart thermostat along with the program enrollment offer should significantly boost enrollment. Vendor 3 suggested increasing the incentive made available to customers, and to expand it to a multi-year program. Additionally, all three vendors recommend starting the recruitment prior to the DR event season in order to maximize enrollment and to increase the number of events. None of the vendors received any complaints from customers, indicating that all three vendors had successful implementations.

Should the BYOT pilot be continued for the 2017 season, there is a significant opportunity to learn more about customer acceptance and participation at varying levels of incentives. With a $60 incentive, Vendor 1 had the highest incentive and the highest levels of enrollment. Vendor 3 offered a $25 incentive and Vendor 2 had no enrollment incentive. They both recruited a similar number of customers relative to Vendor 1, but ultimately neither recruited enough to be able to estimate robust load impacts. Perhaps if more time is available for recruitment both could increase their enrollment numbers and yield robust load impacts. It would also be informative to observe the override rates from customers who received a larger incentive compared to customers who received a smaller incentive through a controlled experiment. This information would be very useful for developing cost effective program offerings in the future.

The five different geographic regions included in the pilot each yielded different levels of performance in customer recruitment success and magnitude of load impacts.

[14] Vendor 3 did not provide the number of customers contacted, and Vendor 2 stated approximately two to three thousand customers were contacted in each of three marketing waves, but it isn't clear if the same set of customers were contacted each time, or if unique groups were contacted for each wave.

San Jose and Chico[15] had the largest impacts, yet Palo Alto had the highest customer recruitment success. The aggregate kW load impact from each region provides a cumulative measure of the value achieved in each location. San Jose had the largest impact at 84 kW, and Stockton and Palo Alto had aggregate impacts that were very similar to one another at 43.3 kW and 41.8 kW, respectively. Chico had an aggregate impacts of 56.6 kW, but that should be divided in half to account for the bias in the estimation; resulting in an impact of approximately 28 kW. The similar outcomes between Stockton and Palo Alto illustrate how higher load impact size can help to offset lower customer recruitment, and vice versa.

When factoring in the recruitment cost per customer, which was uniform across the geographic regions, San Jose (.5 kW), Chico (~.5 kw[15]), and Stockton (.39 kW) clearly provided the best value at nearly twice the average kW impact per customer compared to Palo Alto (.21 kW). Based on this difference in performance, it could be argued to only continue the pilot in the San Jose, Chico, and Stockton areas. However, it would be beneficial to conduct a cost effectiveness analysis to determine the break even impact per customer required to make the program cost effective at a given level of per customer enrollment incentive. Therefore, it likely makes sense to continue the pilot in the San Jose, Chico, and Stockton areas, and it may make sense to continue the pilot in the Palo Alto region if deemed cost effective.

[15] Impacts in Chico were statistically significant; however, they were also biased by approximately .5 kW. This results in an average impact of approximately .5 kW when removing the bias, which is still among the largest impacts.

Appendix A Vendor 2 Marketing Emails

A.1 First round - Test Version 1

Redacted to maintain confidentiality

A.2 First round - Test Version 2

Redacted to maintain confidentiality

A.3 First round - Test Version 3

Redacted to maintain confidentiality

Appendix B Load Impact Methodology Details

Continued from Section 3, Load Impact Estimation Methodology

5.1 Selection of Matched Control Group

Customers who signed up to participate in the BYOT program are inherently different from customers who did not sign up to participate in the BYOT program or customers who were not targeted by the thermostat vendors. For this reason, a control group must be constructed using statistical matching. It is possible that the customers who enrolled in the BYOT program had particular characteristics that made them more likely to enroll than customers who did not enroll or customers who were not targeted to enroll. This is particularly important when studying early adopters of a new technology such as smart thermostats who may have very different energy consumption patterns from those of the rest of the population. This type of behavior introduces selection bias because the difference in usage between the two groups caused by characteristics differences could be mistaken as the impact of treatment. A matched control group is the primary source for reference loads which are used to estimate impacts. The method used to assemble the matched control group is designed to ensure that the control group load on events days is an accurate estimate of what load would have been among BYOT customers on event days if an event hadn't taken place. The control group was selected using a propensity score matching model to find customers in the control group pool who had load shapes most similar to BYOT customers.

The BYOT customers had to pass several filters in order to be eligible for the program. When selecting a suitable control group it was necessary to use the same filters that were applied to the treatment group on the control group. The BYOT customers reside in a very select area of PG&E's service territory. The zip codes of the customers were used to create five clusters of zip codes where the BYOT customers reside. Additionally, PG&E provided an AC load category score for each treatment customer. This score was based on an algorithm PG&E developed to predict the magnitude of each customer's AC usage. In order to select a control group pool, Nexant requested a 20:1 ratio of potential control to treatment customers within each zip code cluster and AC likelihood combination of the treatment customers. There was a large amount of treatment customers who were net-metered so an additional 20:1 ratio pool was requested to obtain enough potential control customers that were net-metered.

A probit model was used to estimate a propensity score for each treatment customer and potential control candidate. Observed characteristics such as zip code, AC load category, and load profiles are explanatory variables that are used to predict whether or not a particular customer enrolled in the treatment or not. The probit model outputs propensity scores for each customer indicating how likely they are to be in the treatment group given the observable characteristics used in the model. Treatment customers are matched to a customer in the control group with the most similar propensity score. This process helps eliminate the difference between the treatment and match-controlled group on the matching variables.

In order to select the probit model used to find the best match for each treatment customer, "out of sample" testing was performed to evaluate several different probit model specifications. Out of sample testing involves running each of the different model specifications using all but one of the proxy days, leaving the unused proxy day to test how well the model performed. By leaving a different proxy day out each time the matching selection is run, one is able to see how well the matches look on a day that was not used to select the match. During this process, eleven different model specifications were tested using different observable variables including usage during event hours, average total daily usage, and usage from 12pm to 9pm. For each of the eleven models six different "calipers" were tested. Calipers set a maximum threshold of how large the difference in propensity scores can be for a matched pair. During the matching process, the treatment customers are matched to the control customer who has the most similar propensity score to them. Additionally, treatment customers can only be matched to a control customer in the same zip code cluster and AC load category. If the difference between a treatment customer and control customer's propensity score is higher than the set caliper, the treatment customer will not be matched. Therefore, a caliper sets the standard for how close the matched pairs need to be. The model that was initially selected did not perform well across the Vendor 1, Vendor 3, and Vendor 2 customers at the same time. It was necessary to split out the BYOT customers into smaller segments by vendor to find the optimal probit model to find the closest control customer matches. This provided much closer matches for each of the three thermostat vendor customers.

Figure 5-1 through Figure 5-3 show the results of the matched control group for each of the three thermostat vendors. The Vendor 1 customers match very well to their matched control group on proxy days. The Vendor 3 and Vendor 2 customers do not align as well as the Vendor 1 customers do. This is in large part due to the difference in sample size between the three thermostat vendors. Vendor 1 has over 500 customers, while Vendor 3 and Vendor 2 have less than 50 customers each. With such a small sample, it's very hard to get as close of a match compared to when there are 500 customers. Additionally, the matched control groups selected for Vendor 3 and Vendor 2 provided the closest match on event days leading up to the event hours.

Figure 5-1: Hourly Average Demand for Vendor 1 Customers on Proxy Days

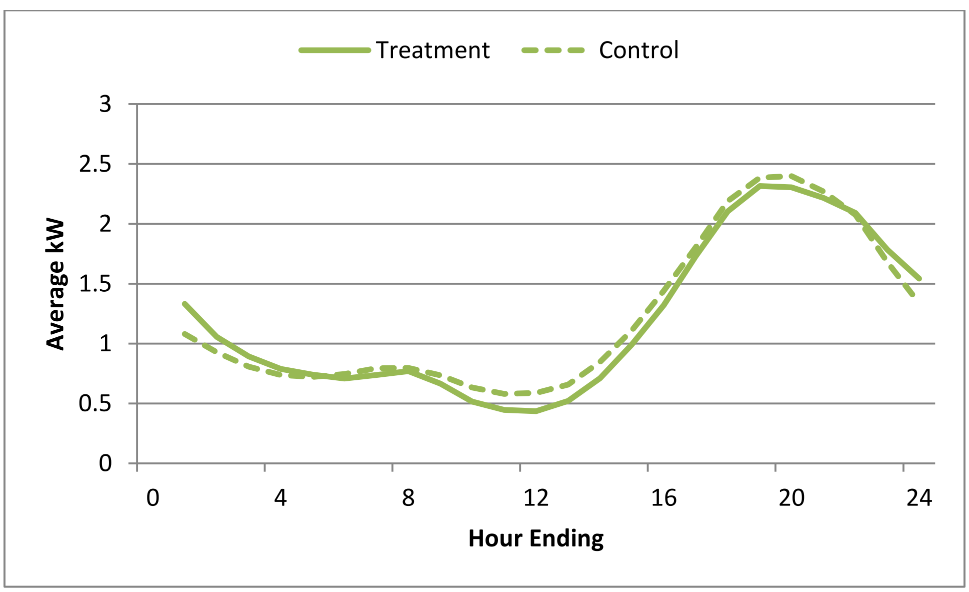


Figure 5-2: Hourly Average Demand for Vendor 2 Customers on Proxy Days

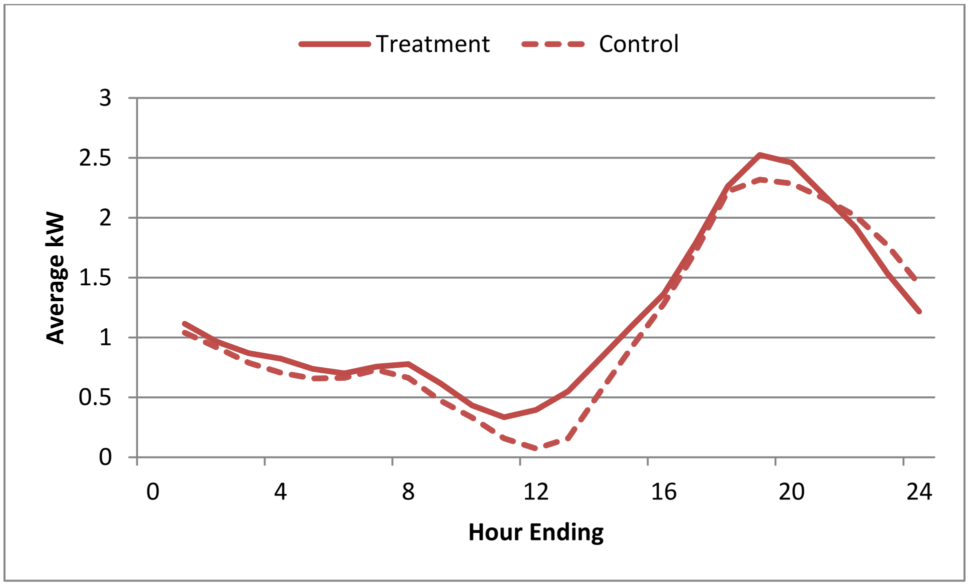
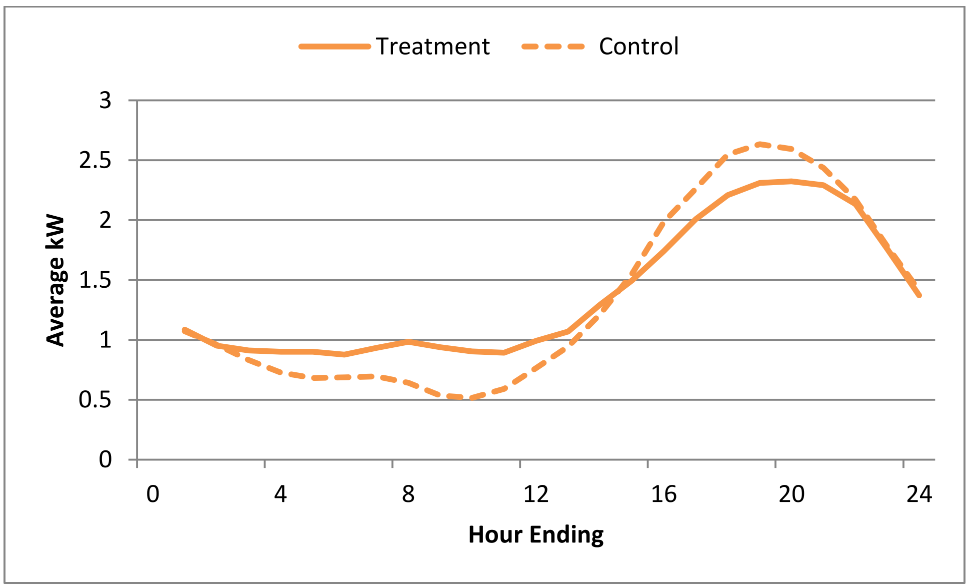


Figure 5-3: Hourly Average Demand for Vendor 3 Customers on Proxy Days



5.2 Difference-in-Differences Regression Models

After a matched control group was created, program impacts were estimated using a difference-in-differences regression model. This methodology is based on the assumption that the program impact is equal to the difference in usage between the treatment and the control groups during the event period, minus any pre-existing difference between the two groups. When using difference-in-differences, the matched control group does not need to perfectly match the treatment group on the proxy days- as was the case with both the Vendor 2 and Vendor 3 customers due to the small sample size. Any differences that may be due to unobservable variables that could not be included in the matching model will be netted out by the differencing. It is a reasonable assumption that any unobservable differences between the treatment and the control groups during the event period hours on proxy days stay the same during the BYOT event hours. Therefore any further difference between the groups in the BYOT event hours is assumed to be the impact of treatment. To reduce any bias that might result from small differences in the loads for customers selected for the matched control groups, a difference-in-differences model was used to estimate load impacts. This regression model is shown in **Error! Reference source not found.** below:

Equation 5-1: Difference-in-Differences Models

$$kW_{i,t} = a + b \cdot Treatment_i + c \cdot Event_t + d \cdot (Treatment_i \cdot Event_t) + u_t + \varepsilon_{i,t} \text{ for } i \in \{1, \dots, n_i\} \text{ and } t \in \{1, \dots, n_t\}$$

Variable	Definition
i, t, n	Indicate observations for each individual *i*, date *t* and event number *n*
a	The model constant
b	Pre-existing difference between treatment and control customers
c	The difference between event and proxy days common to both treatment and control group members[16]
d	The net difference between treatment and control group customers during event days–this parameter represents the difference-in-differences
u	Time effects for each date that control for unobserved factors that are common to all treatment and control customers but unique to the time period
v	Customer fixed effects that control for unobserved factors that are time-invariant and unique to each customer; fixed effects do not control for fixed characteristics such as air conditioning that interact with time varying factors like weather
E	The error for each individual customer and time period
Treatment	A binary indicator or whether or not the customer is part of the treatment or control group
Event	A binary indicator of whether an event occurred that day–impacts are only observed if the customer is enrolled in BYOT (*Treatment* = 1) and it was an event day

The model was estimated using both event days and proxy days, which are nonevent days with similar weather conditions and system load usage as days when events are called. The difference in loads between treatment and control customers for the event period hours on proxy days is subtracted from the differences on BYOT event hours to adjust for any differences between the treatment and control groups due to random chance.

As an extra validation, the simple difference in loads between treatment and control customers during event hours on event and proxy days was calculated to ensure that the regression model produces a similar output. The regression model also reduces the standard errors of the impact estimates compared to those that can be calculated from a simple difference in loads.

[16] In practice, this term is absorbed by the time effects, but it is useful for representing the model logic.