Can Residential Water Heater Provide Reliable and Predictable Demand Response Grid Services?

Josh Bode, Demand Side Analytics, Los Angeles, CA
Adriana Ciccone, Demand Side Analytics, San Francisco, CA
Zachary Horvath, Cadmus, Portland, OR
Yoh Kawanami, Hawaiian Electric Company, Honolulu, HI
Phillip Kearns, Cadmus, Portland, OR
Jim Stewart, PhD, Cadmus, Portland, OR
Alex Xie, Demand Side Analytics, Los Angeles, CA

ABSTRACT

In 2021, Hawaiian Electric Company (HECO) evaluated the demand response grid services from residential water heaters obtained through the Grid Services Purchase Agreement (GSPA) #1. The contract between HECO and Open Access Technology International (OATI), a third-party aggregator, provides for the delivery of fast frequency response (FFR), capacity-build, and capacity-reduction demand response grid services on Maui and Oʻahu. Shifted Energy, an OATI subcontractor, retrofitted residential electric resistance water heaters with controllers to implement grid services. The GSPA evaluators, Cadmus and Demand Side Analytics, designed a randomized controlled trial (RCT), the gold standard in evaluation, to evaluate the performance of GSPA #1 on Oʻahu by measuring demand impacts and verifying the accuracy of OATI's settlement calculations, settlement methods, and grid services forecasts. The RCT study took place between January 21, 2021, and June 1, 2021, during which time HECO called 27 capacity-build events and 37 capacity-reduction events, and one FFR event was triggered. Overall, the evaluation shows that the GSPA water heaters delivered the contracted grid services, though issues related to minimizing snapback and ensuring consistent delivery of load building grid services should be addressed before HECO scales the resource.

Introduction

Hawaii has a clean energy goal of 100% electricity sales from renewable sources by 2045 and is increasing its reliance on utility-scale wind and solar power. In 2020, Hawaiian Electric Company (HECO), a vertically integrated utility that provides electricity service on five islands and to 95% of the state's residential customers, generated 35% of its electricity from renewables. To meet the 2045 clean energy goal, HECO will need to find new ways of balancing its grid and addressing issues related to intermittency, the over-and under-supply of power, and ramping from integrating renewable resources. Specifically, HECO is looking for alternative sources of grid services, currently provided by diesel-generating facilities, and is investing in utility-scale battery storage and distributed energy resources (DERs), including grid-interactive water heaters and behind-the-meter batteries with solar photovoltaic (PV).

This paper presents results from the evaluation of HECO's Grid Services Purchase Agreement (GSPA) #1, through which the utility contracted with a third-party aggregator to receive fast frequency response (FFR), capacity-building, and capacity-reduction demand response grid services. This paper reports on the impacts of demand response grid services provided by residential water heaters retrofitted with a grid-interactive controller and the accuracy of the aggregator's settlement calculations.

Background

In March 2019, HECO reached agreement with OATI, Inc., a demand response aggregator, to provide 11 MW of FFR, 1 MW of load building, and 10 MW of load-reducing demand response capacity between 2019 and 2024.[1] The demand response capacity would be provided by a mix of residential water heaters, residential solar PV and battery storage systems, and commercial battery storage systems on the islands of Maui and Oʻahu.

Implementation of the grid services contract was delayed by about one year because the COVID-19 pandemic slowed down participant recruitment and because integration of OATI's and HECO's demand response management systems took longer than expected. As a result, only grid services from residential water heaters on Oʻahu were operational in 2020 and 2021. Shifted Energy, an OATI subcontractor, implemented this part of the GSPA.

Beginning in January 2020, Shifted Energy began enrolling residential customers in water heater demand response. Most GSPA participants were in low- or middle-income multifamily residential buildings. Shifted Energy targeted multifamily buildings because, in most cases, it was possible to coordinate the retrofits of water heaters with a single building manager rather than individual customers and to economize on installation costs by installing multiple controllers at one site. Shifted Energy also enrolled a small number of water heaters in single-family homes.

Shifted Energy retrofitted residential electric resistance water heaters at participating facilities with Tempo smart controllers, which respond instantly to commands to turn the units on or off either from OATI's Grid Services Delivery System or automatically when the water heaters detect a deviation from the desired frequency on HECO's electric distribution system (i.e., over- or under-frequency event). Because water heaters are relatively well insulated, they can store energy, analogous to batteries, and help utilities by absorbing energy during periods of excess power supply. The over-supply of power on HECO's system is a frequent occurrence because of growing integration of solar energy resources. The controllers communicate with the implementer control system through a cellular network connection, avoiding the need to rely on the residential customer's Wi-Fi network. The controller's current transformer collects high-frequency voltage, current, and frequency measurements.

Evaluation

Working with HECO, evaluation consultants Cadmus and Demand Side Analytics (the Cadmus team) set up a randomized controlled trial (RCT) field experiment to evaluate the demand response grid services from water heaters in the GSPA #1. The experiment ran from January 21, 2021, to June 1, 2021, and included 1,463 water heaters on Oʻahu.

At the beginning of the experiment, HECO and the Cadmus team randomly assigned about half of the eligible water heaters to a treatment group (N=733) and the other half to a control group (N=730). Water heaters in the control group did not experience any demand response events and provided a baseline for estimating the impacts of the demand response. Because water heaters were randomly assigned to one or the other group, experiencing a demand response event should not correlate with the energy consumption and other characteristics of water heaters, and comparison of the demand of the treatment and control groups should provide an unbiased estimate of savings. As shown below, the energy consumption of treatment and control group water heaters was not statistically different on non-event days, suggesting that the randomization appears to have resulted in well-balanced groups.

---

[1] Public Utilities Commission of the State of Hawaii (August 9, 2019). Order No. 36467 Approving the HECO Companies' Grid Services Purchase Agreement with Open Access Technology International. https://dms.puc.hawaii.gov/dms/DocumentViewer?pid=A1001001A19H12A85058F00420.

During the RCT, HECO initiated 27 capacity-build and 37 capacity-reduction demand response events. Also, one FFR demand response event on March 29 at 4:53 a.m. HST, lasting about five minutes, was automatically triggered in response to a drop in frequency detected on HECO's power system. All capacity-build events started at 10:00 a.m. and lasted four hours. Capacity-reduction events occurred between 5:00 p.m. and 9:00 p.m. and averaged one hour and 17 minutes in duration. Table 1 shows the number, typical length, and window of the different event types during the RCT.

Table 1. HECO Demand Response Grid Services Events

Event Type	Number of Events	Average Length (hr: min: sec)	Event Window
Capacity Build	27	04:00:00	10:00 a.m. - 2:00 p.m.
Capacity Reduction	37	01:17:50	5:00 p.m. - 9:00 p.m.
FFR	1	00:04:53	4:48 a.m. - 4:53 a.m.

Data

Shifted Energy provided the evaluation consultants with five- or 15-minute interval electricity demand (kilowatt) data for the water heaters in the experiment. The Cadmus team analyzed water heater telemetry data because many participating water heaters were located in master-metered buildings and many were in buildings with individually metered apartments that did not have AMI interval data.

The telemetry interval consumption data were generally complete and clean, with relatively few water heaters having missing or erroneous consumption readings. A small percentage of water heaters did not return reads for more than 30% of the 15-minute intervals, which was likely attributable to poor cellular reception at these locations. After removing these water heaters and four other water heaters in the extreme tails of the distribution, 1,336 water heaters remained in the analysis sample.

The Cadmus team validated the accuracy of the water heater telemetry demand data by installing data loggers that collected one-minute interval kWh readings on nine GSPA participant water heaters and compared the telemetry and logger electricity demand readings. Over part of the RCT from May 12 to May 31, 2021, when the loggers were installed and collecting data, the correlation coefficient between the logger data and telemetry data across all water heaters was 0.999 with a mean absolute percentage error of 3.1% across all five-minute intervals. The mean difference between water heater telemetry and logger data in each hour of the day did not exceed 6 watts. Overall, the team concluded that the GSPA water heater telemetry data accurately measured water heater electricity demand at five-minute intervals, validating the use of these data for impact evaluation.

Validation of the RCT

The Cadmus team also validated the RCT design by comparing the 15-minute interval electricity demand of water heaters in the treatment and control groups on non-event, non-holiday weekdays during the RCT period. Figure 1 shows that electricity demand for the groups tracked each other closely and there were no statistically significant differences between the groups for most intervals, suggesting that the randomization appears to have resulted in well-balanced groups.

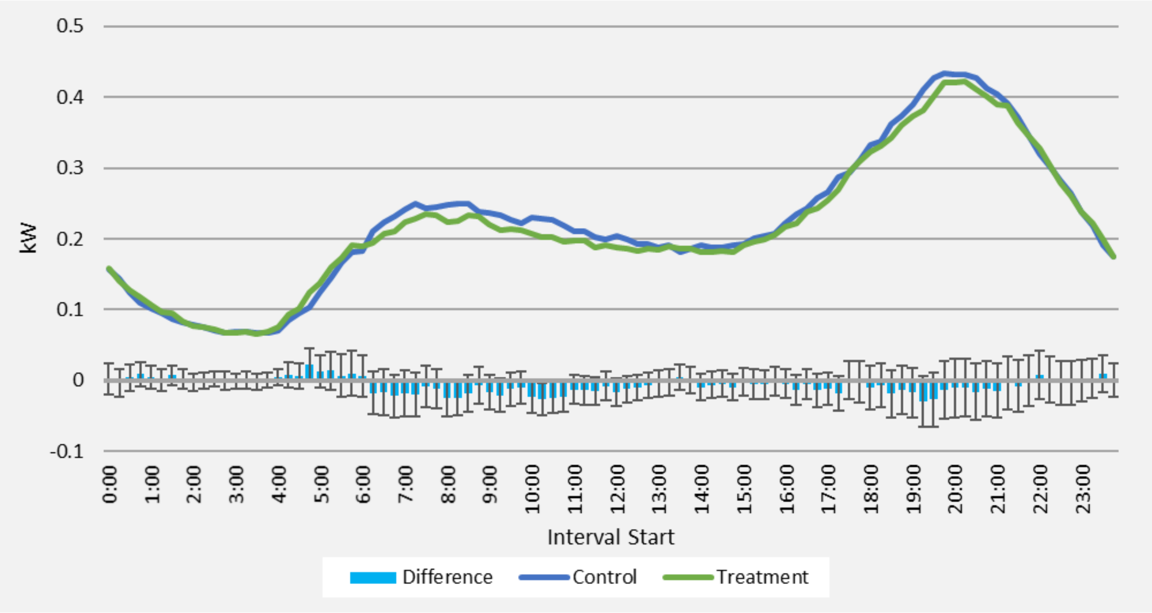

Figure 1. Water Heating Electricity Demand on Non-event, Non-holiday Weekdays during the Experiment. The figure shows water heating electricity demand on non-event, non-holiday weekdays during the experiment for the control and treatment groups. The difference in average consumption per water heater between the groups was estimated in a regression of individual water heater consumption on 15-minute interval of the day fixed effects and interval fixed effects interacted with an indicator for assignment to treatment. The error bars show 95% confidence intervals. Standard errors were clustered on water heaters.

Figure 1 also shows that across all weekday hours, water heaters exhibited relatively low average electricity demand, with demand peaking in the morning at about 0.25 kW and again in the evening at about 0.45 kW. As the typical water heater in the RCT demanded between 3.5 kW and 4.5 kW while heating, the demand data suggest that only a small percentage of water heaters were heating at each point in time.

Demand Response Grid Services Impact Estimates

Figure *2* depicts the impacts of capacity-build (10:00 a.m. to 2:00 p.m.) and capacity-reduction (6:00 p.m. to 8:00 p.m.) demand response events on March 2, 2021. At 8:00 a.m., the treatment group was shut off for the two hours preceding the start of the capacity-build event at 10:00 a.m. to shift water heater load into the midday period. (All capacity-build events during the RCT were preceded by a controlled reduction in load.) During the capacity-build event, the treatment group's average demand was higher than the control group's average demand, reflecting the impacts of the treatment. During the capacity-reduction event, devices in the treatment group were shut off from 6:00 p.m. to 8:00 p.m. After the event ended, the demand of treatment group water heaters snapped back sharply (to over 1 kW in this example) as the devices were allowed to switch back on and resume heating water after being shut off during the first half of the typical evening hot water consumption period.

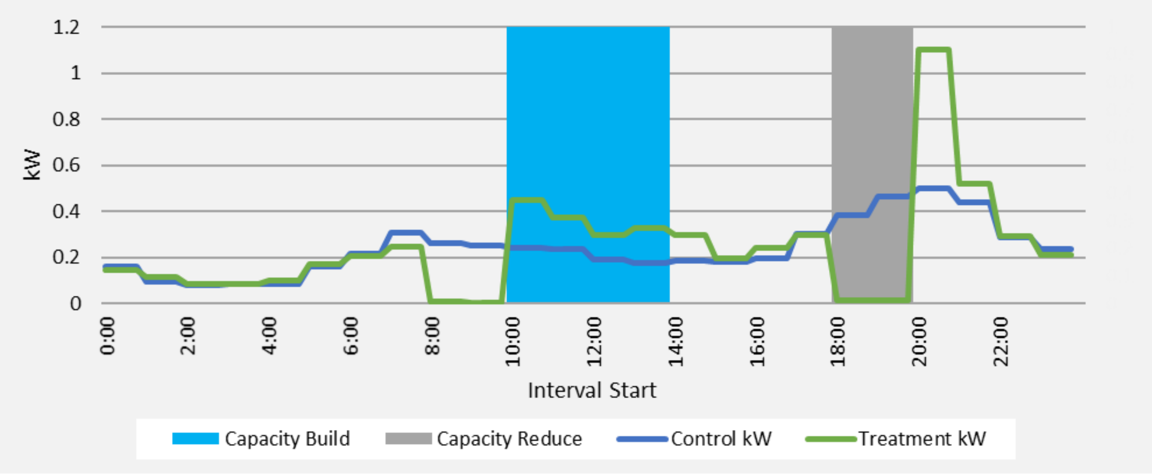

Figure 1. Average Demand During Capacity Build and Capacity Reduction Events, March 2, 2021. Figure shows unconditional mean electricity demand for water heaters in the RCT treatment and control groups on March 2, 2021. Hourly electricity demand calculated using five-minute interval water heater electricity demand telemetry data.

The Cadmus team used panel regression models of water heater electricity demand to estimate the demand impacts from capacity-build, capacity-reduction, and FFR events. A separate regression of 15-minute interval electricity demand was estimated for each event using only data for the event day. Independent variables included in the model were time-of-day (15-minute interval) fixed effects, each device's average non-event day usage (calculated at the monthly level) interacted with time-of-day fixed effects, and assignment to the treatment group status interacted with the time-of-day effects. The coefficients from the time-of-day and treatment interactions represent the impacts from the grid services. The demand response grid services impact estimates presented in this paper are robust; that is, they do not change when alternative estimation methods or model specifications are used.

Capacity Build

Figure 3 and Figure 4 show the average treatment effect estimates (in kilowatts per water heater and as a percentage of the reference load, respectively) for each of the 27 capacity-build events during the RCT study period.

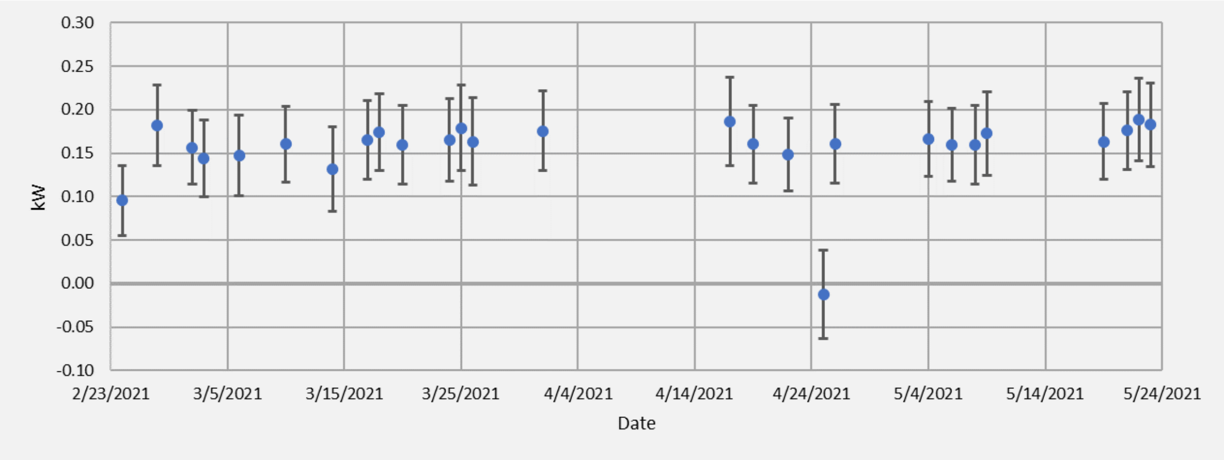

Figure 2. Average Treatment Effect Per Water Heater (kW) for Capacity-Build Events. Estimates obtained from ordinary least squares (OLS) regressions of water heater demand. See text for details. Error bars represent the 95% confidence interval for impact estimates. Standard errors were clustered on water heaters.

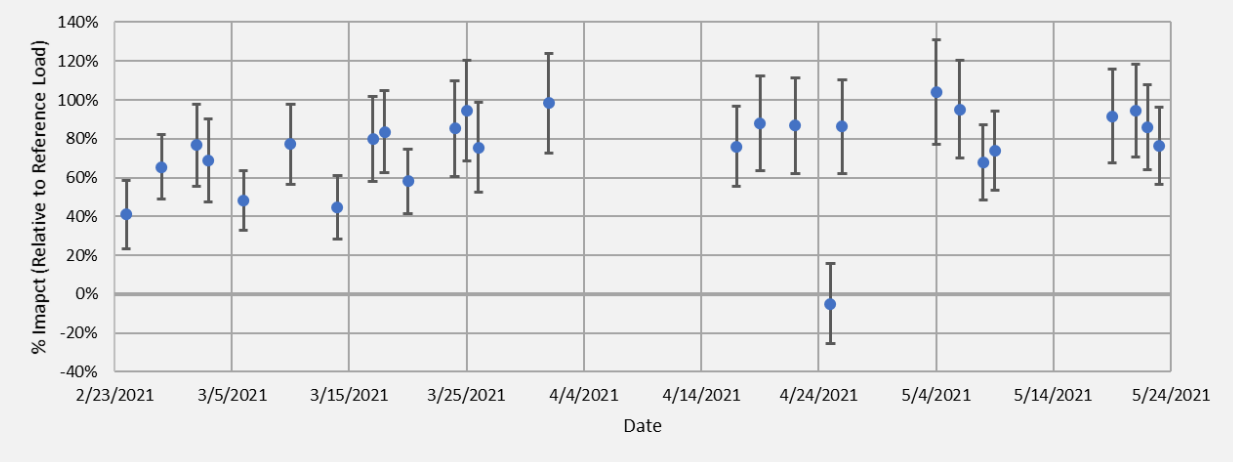

Figure 3. Percentage Treatment Effects for Capacity-Build Events. Percentage impacts estimated as kilowatt impact divided by the average reference load for treatment group water heaters. Error bars represent the 95% confidence interval for impact estimates. Standard errors were clustered on water heaters.

The capacity-build events resulted in impacts per water heater ranging from 0.15 to 0.20 kW and averaging 0.159 kW across hours of all capacity-build events. These impacts represent an average increase of 76% of baseline water-heating demand during the 10:00 a.m. to 2:00 p.m. period. Some events doubled baseline demand, as shown by the percentage impacts exceeding 100%. Capacity-build demand response provided consistent increases in demand across events, with the exception of the April 25, 2021, event (which appears to have failed, producing no statistically significant increase in demand.)

Capacity Reduction

Figure 5 and Figure 6 show the average treatment effect estimates (in kilowatt per water heater and as a percentage of the reference load, respectively) for each of the 37 capacity-reduction events during the RCT study period.

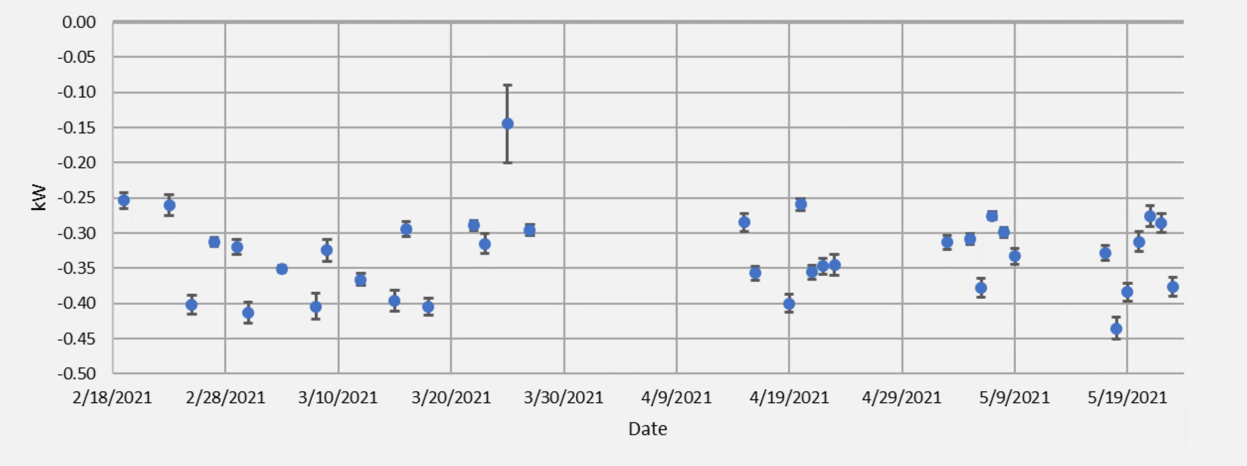

Figure 4. Average Treatment Effect Per Water Heater (kW) for Capacity-Reduction Events. Estimates obtained from OLS regressions of water heater demand. See text for details. Error bars represent the 95% confidence interval for impact estimates. Standard errors were clustered on water heaters.

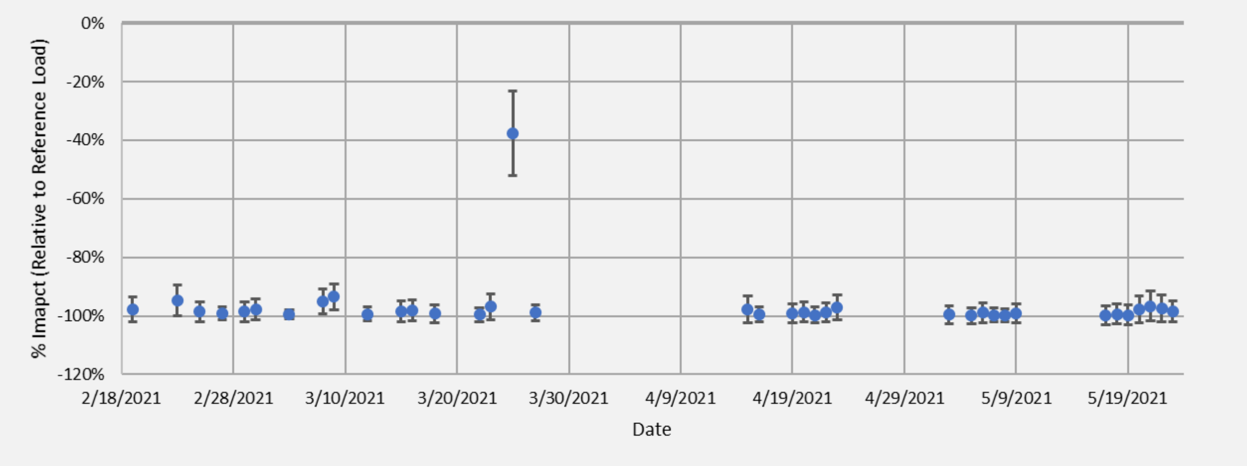

Figure 5. Percentage Treatment Effects for Capacity-Reduction Events. Percentage impacts estimated as kW impact divided by the average reference load for treatment group water heaters. Error bars represent the 95% confidence interval for impact estimates. Standard errors were clustered on water heaters.

During most capacity-reduction events, water heating electricity demand was reduced to nearly zero, showing that most water heaters that would have been operating remained off. On average, electricity demand was reduced by an average of 0.321 kW per water heater, or 95% of the reference load across all events. All but one event (March 25, 2021) reduced demand by 93% or more. Nine of the events reduced demand by 100%.

Fast Frequency Response

During the one FFR event, treatment and control group water heaters were shut off when the under-frequency condition was detected. This FFR event occurred on March 29, 2021, at 4:48:24 a.m., and the frequency returned to normal at 4:53:02 a.m.

FFR was not implemented as an RCT, so the Cadmus team modified its analysis approach to estimate impacts for both the treatment and control groups relative to their predicted baseline demand in each five-minute interval.[2] Figure 7 shows average metered demand and average reference (model predicted baseline) demand per water heater around the FFR event. The figure shows that, in the intervals including the under-frequency detection at 4:48:24 and the following interval, average water heater demand drops nearly to zero. In the following intervals, demand then rises as the controllers allow the water heaters to turn back on.

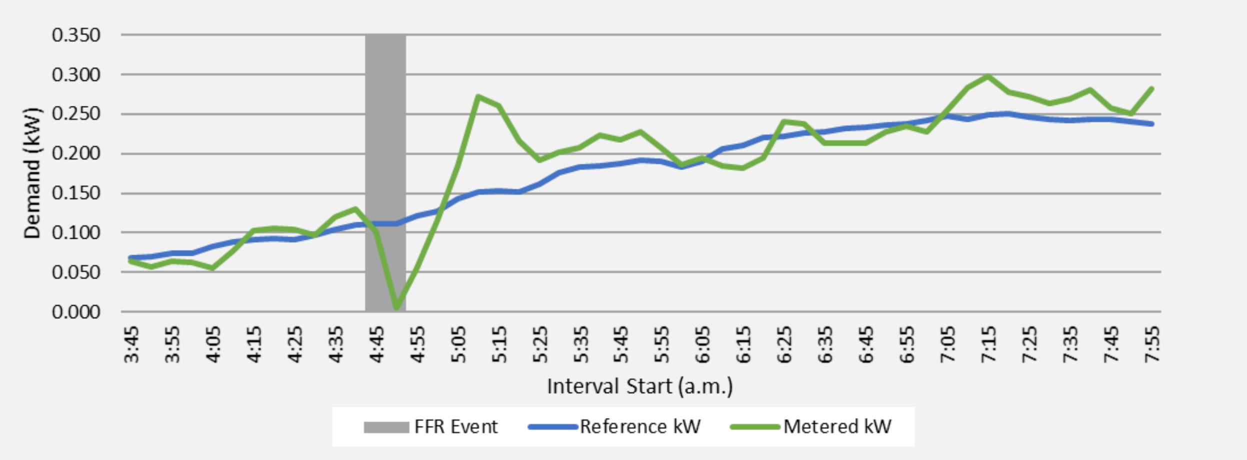

Figure 6. Electricity Demand during FFR Event on March 29, 2021. Reference load obtained from regression analysis of water heater electricity demand. See text for details. The FFR event began at 4:48 a.m. and frequency returned to the normal range at 4:53 a.m.

Because the FFR event started midway through the five-minute interval (4:45 a.m. to 4:50 a.m.), its impact was estimated in two steps. First, the Cadmus team estimated the demand reduction for the whole interval using regression analysis of the interval electricity demand data. The estimated impacts for the first interval of the FFR event and other intervals between 3:45 a.m. and 7:55 a.m. are shown in Figure 9. Second, the team calculated the proportion of the first interval affected by the FFR event and divided the estimated impact for this interval by the proportion to estimate the average FFR impact. Dividing the impact estimate for the first interval (=0.01) by 0.311 (31.1%) yields an FFR demand impact estimate of 0.032 kW. During the second interval of the FFR event, the impact exceeded 0.1 kW.

---

[2] Baseline demand was estimated in an OLS regression of water heater 5-minute interval demand on a set of 5-minute interval of the day fixed effects and the fixed effects interacted with an indicator for the FFR event day. The model was estimated with data for the FFR event day and all non-holiday, non-event days in March 2021 and April 2021. The coefficients on the interval fixed effects are estimates of baseline demand. The coefficients on the interaction variables provide estimates of the FFR kilowatt impacts per water heater.

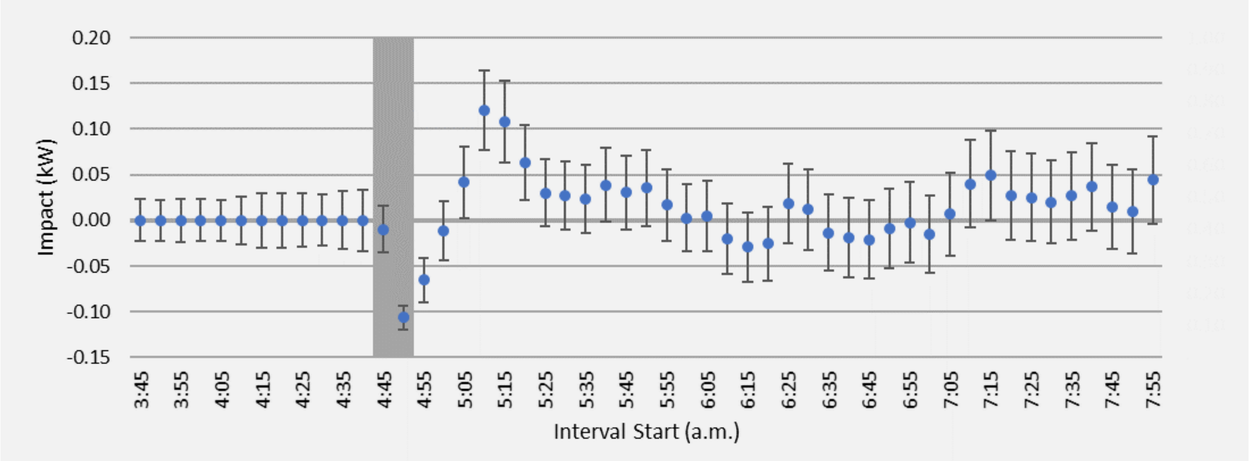

Figure 7. Average Impact Per Water Heater (kW) for FFR Event on March 29, 2021. The kW estimates are for five-minute intervals. Impact estimates are obtained from a regression analysis of water heater electricity demand. See text for estimation details. Error bars show 95% confidence intervals. Standard errors were clustered on water heaters.

The estimates show that FFR grid service provided the expected demand reductions, though the team could not verify whether devices responded as quickly as specified in the GSPA.[3] In addition, the impacts reported here reflect the FFR's performance during just one event, so conclusions concerning its performance are less robust than for capacity build or capacity reduction events, which are based on a much larger sample of events across the study period. Future studies using higher-frequency data could provide additional information about FFR's performance as a grid service under GSPA #1.

Potential Impacts of Demand Response on Grid Stability

Based on the evaluation results, the Cadmus team determined that OATI delivered the grid services within the requirements of the contract. Though the grid services contracted under GSPA1 are relatively small, two aspects of their performance could become issues if the aggregated resource under GSPA1 were expanded:

- Water heater electricity demand after capacity-reduction events was high compared to baseline demand. The GSPA1 capacity ramp rate requirements limit the total amount of post-event snapback; however, HECO could impose more stringent ramping requirements in future GSPAs if the amount of snapback were to risk grid stability.
- Capacity-build events did not show a consistent, sustained increase in load but showed a relatively volatile (i.e., spiky) grid service delivery during the event. For GSPA #1, capacity-build performance is assessed using a one-hour interval demand, which obscures within-hour volatility. For subsequent GSPAs, performance assessment will be measured for 15-minute intervals to encourage a more consistent response during capacity-build events.

---

[3] The frequency of the telemetry data available (five-minute intervals) relative to the duration of the FFR event (less than five minutes), as well as the low average water heating demand during the early morning when this event occurred, make precise estimation of the impacts of FFR during the first interval challenging.

Settlement Calculations Accuracy Assessment

The GSPA #1 specifies baseline methods for calculating demand impacts for performing settlement calculations. Capacity-build and capacity-reduction baselines are to be calculated using a 10-in-10 similar day baseline. FFR baselines are to be calculated using the load in the interval preceding the start of the event. As part of the evaluation, the Cadmus team determined whether the GSPA baseline methods were capable of accurately quantifying the load increases or reductions associated with the grid service events.

The method to conduct this assessment was relatively simple. The Cadmus team picked pseudo-event days: non-event days similar to events but where no event was actually called. For each of these pseudo-event days, the team ran the baseline algorithm used for settlement calculations as if the pseudo-event days were actual event days. Because no event took place, the baseline result could be compared to the true observed loads. Any differences between the two during event periods represent errors attributable to the baseline method. These differences were summarized using statistics representing two key variables of accuracy. The mean percentage error (% error in the table below) summarizes the degree to which the baseline tends to over- or understate the true value on average. The normalized root mean squared error summarizes the variability in error from event to event (or hour to hour within an event). Together, these variables represent the accuracy (% error) and precision (root mean squared error) of a baseline. The best baselines are both accurate and precise.

Table 2 summarizes these key statistics across all pseudo-event days assessed in this analysis. These included six pseudo capacity-build events, eight pseudo capacity-reduction events, and five pseudo FFR events. The table shows that observed usage for the average device during the pseudo-event period, the average baseline usage for the same period, and the normalized RMSE (a measure of precision) and percent bias (a measure of accuracy). The magnitude of bias is less than ±5% in all three grid services. The load impacts calculated by the Cadmus team were approximately 76% for capacity-build events, 95% for capacity-reduction events, and at least 28% for FFR.[4] In all cases, the magnitude of the bias associated with the baseline methods was substantially lower than the magnitude of the reductions delivered in the events themselves. As a result, the Cadmus team finds no need to recommend changes to the GSPA-prescribed baseline method of the baseline algorithms.

Table 2. Baseline Accuracy and Precision Summary by Grid Service

Grid Service	Usage (kW)	Baseline (kW)	Normalized RMSE (%)	Bias (%)
Fast Frequency Response	0.23	0.22	5.88	-2.63
Capacity Build	0.23	0.22	15.22	-4.69
Capacity Reduction	0.32	0.33	7.98	3.68

Conclusions

Using a randomized controlled trial, the Cadmus team evaluated the impacts of demand response grid services from GSPA #1 retrofitted water heaters. Overall, the evaluation showed that the water heaters delivered the contracted grid services. During most capacity-reduction events, water heating electricity demand was reduced to nearly zero, showing that most water heaters that would have been operating remained off. Capacity-build events greatly increased water heater electricity demand relative to baseline demand, often by 90% or more. Only one FFR event was triggered, but it decreased water heating electricity demand in response to detection of an under-frequency event.

---

[4] This result is based on the first five-minute interval of the event.

The evaluation results are very encouraging for utilities considering grid-interactive water heaters as a means of managing frequency events and periods of over- and under-supply of power. However, the evaluation also suggests areas for future research before utilities scale this type of resource. First, utilities will want to prepare for and be ready to manage significant snapback in electricity demand after water heater load reduction events end. Second, utilities will want to work closely with demand response aggregators or service providers to ensure consistent delivery of aggregate demand response grid services during load building events. Third, as only one FFR event was triggered during this evaluation, more research is needed before the performance of this grid service can be evaluated conclusively.