Entering Uncharted Water (Heater Demand Response)

Christina Carlson, Apex Analytics, Boulder, CO
Evelyn Martin, Tacoma Power, Tacoma, WA
Lauren Gage, Apex Analytics, Coeur d'Alene, ID

ABSTRACT

In 2022, Tacoma Power, a public utility in the Pacific Northwest, implemented a water heater demand response (WHDR) pilot. The pilot has been in operation for over two years and was designed as a randomized controlled trial (RCT). Using a fast feedback pilot evaluation strategy, the utility has been working in parallel with their implementor and evaluator to explore new demand response (DR) strategies, optimize savings, and flatten customer load. Throughout the duration of this pilot, the utility has successfully implemented thousands of events with very few customer complaints. The first year of the pilot focused on a standardized approach to determine how much peak load could be shifted from enrolled customers within a defined set of events. The second year focused on testing new strategies to see if savings could be further optimized.

Tacoma Power is unique because of its substantial hydroelectric generation capacity. The utility recognizes the likely future capacity constraints of electrification and is proactively testing new technologies and platforms to manage customer load. Through this pilot, Tacoma Power has employed innovative testing strategies to identify impacts in WHDR to support future large-scale DR opportunities. This paper will share the pilot's results and Tacoma Power's experiences with its WHDR pilot.

Introduction

The impetus for this pilot was Tacoma Power's Integrated Resource Plan[1], which identified the need for 10 megawatts (MW) of morning and evening load management. Electric resistance water heating is common in Tacoma Power's service territory, with an estimated 100,000 residential units. In 2021, WHDR was estimated to be roughly half (~53 MW) of Tacoma Power's DR technical potential. Given the high potential, this pilot was designed to explore the ability to reduce peak demand using water heaters. Early on, Tacoma Power recognized that water heater usage consistently peaks in the morning and afternoon, when their customers are home, taking showers, cooking, and doing laundry. These peaks occur at approximately the same time that Tacoma's generation capacity costs[2] are highest, as shown in Figure 1. Demand reductions during hours with higher costs, shown in $/MW, would lead to cost savings for both the utility and ratepayers. Therefore, technologies that can reduce or shift load from morning or afternoon peaks are especially valuable to Tacoma Power.

[1] Available at https://www.mytpu.org/wp-content/uploads/5569A_TP_2024_IRP_Report_Resolution_Cover_0924_OUT.pdf
[2] The cost of having 1 MW of generating capacity available.

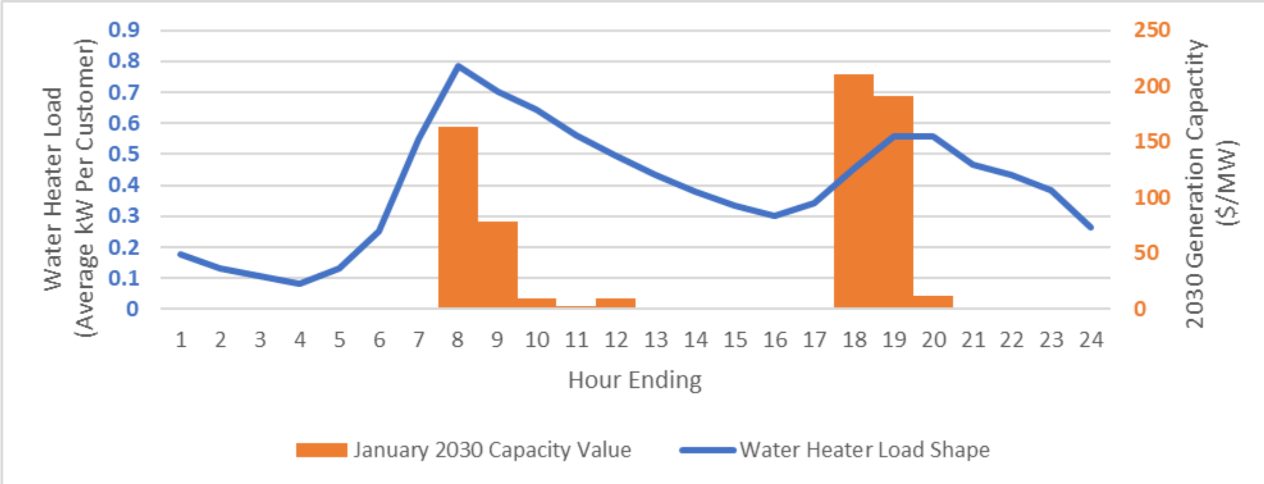

Figure 1. Water Heater Load vs. 2030 Capacity Prices

At the start of the pilot, controllers were installed on roughly 240 electric water heaters in single-family homes across the Tacoma Power territory. These controllers had cellular connectivity and could turn off the water heaters when provided a signal from Tacoma Power. Each household was randomly assigned to either the treatment or control group for the duration of the pilot. Events for the treatment group were pre-scheduled without customer knowledge to determine whether customers would notice when events were called and if events affected customers' water usage.

During DR events, participating water heaters were turned off for predefined durations, resulting in measurable reductions in electric demand (measured in kW). The initial events were scheduled to be three hours long, aligning with system peaks in the morning and afternoon. Once the events concluded and the water heaters resumed operation, a temporary increase in load, referred to as "snapback", was observed as the water in the tanks began to warm up again.

The second year of the pilot tested three main strategies. First, Tacoma Power attempted to mitigate the effects of snapback by staggering the times at which treatment water heaters turned back on. The second strategy was an attempt to increase the impact of events by shortening event durations from three hours to two hours. The final phase of testing was referred to as "Resource Following", which used price forecasts to schedule events across the day when prices were highest, with the goal of optimizing cost savings for the customers.

Evaluation Strategy

The evaluation team (Evaluation) was provided with water heater usage data in 15-minute intervals from both treatment and control groups. For the purpose of this evaluation, the data was averaged to hourly kW. To assess both the load reductions during events and the magnitude of snapback after the events, usage data was leveraged alongside details provided by the implementer. The pilot employed an RCT framework, allowing for robust comparisons between treatment and control participants. Two primary analytical methods were applied:

- **Pooled Regression Models (Sept 2022–May 2024):** Used to evaluate periods with fixed, consistent event timing, enabling precise estimation of average hourly impacts across customer groups. Models were pooled based on their event timing each month.
- **Difference-in-Differences (June 2024–December 2024):** Employed during the Resource Following phase, when events occurred at highly variable times and with differing levels of customer participation. This method helped isolate the treatment effect while controlling for confounding variations in baseline usage.

First Year of Pilot Testing – Standardized Events

Tacoma Power has been testing innovative strategies since 2022 to manage residential water heater load as part of its multi-year DR pilot. The initial phase of the pilot focused on targeting peak usage periods, with two 3-hour events scheduled per event day, typically in the morning and evening, when water heater demand is highest. As shown in Figure 2, per-customer load reductions varied monthly from September 2022 to September 2023. Overall, the average kW reductions of the 3-hour events was -0.26 kW.

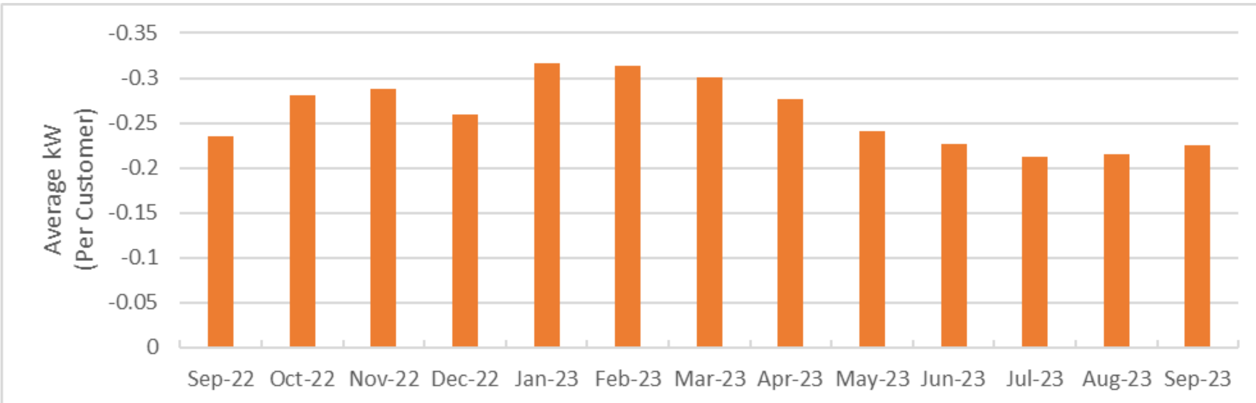

Figure 2. Monthly Event Impacts Measured in Average kW[3]

A seasonality to the results was evident in the first year of results. The greatest impacts occurred during the winter months (January through March 2023), with average reductions around -0.30 kW per customer. These higher savings are likely due to colder inlet water temperatures, which increase water heater energy consumption. In contrast, reductions were smaller in the warmer months (June through August 2023), likely due to lower baseline water heating demand. Despite this seasonality, the DR events consistently produced measurable load reductions throughout the year. Figure 3 depicts these impacts on an hourly basis.

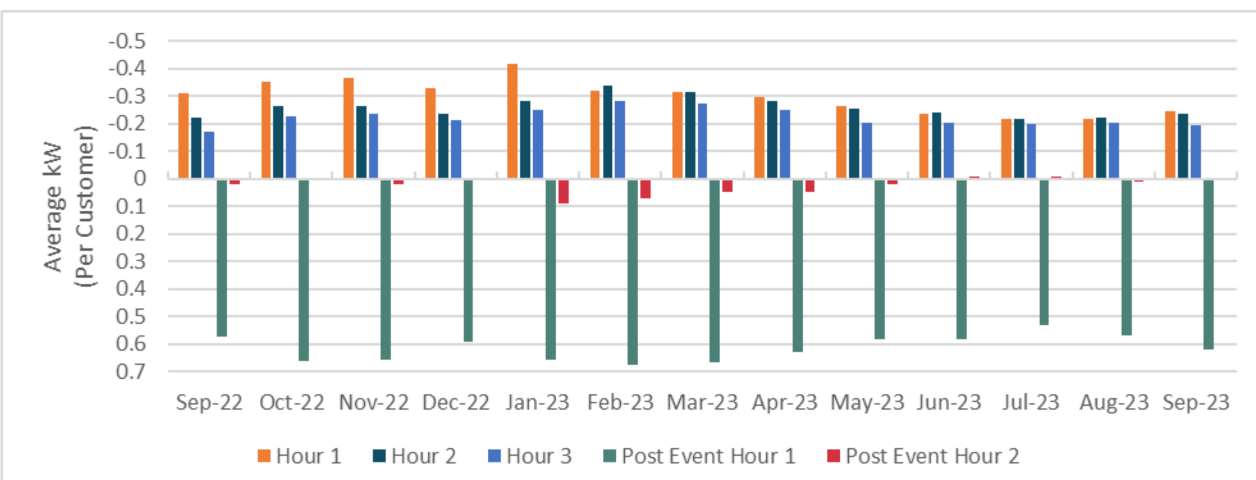

Figure 3. Hourly Impacts During First Year of Pilot

Figure 3 shows the load reductions across the 3-hour event windows (Hours 1, 2, and 3), with the greatest reduction typically occurring during the first and second hours of the event (orange and navy bars). Hour 3 consistently had the smallest kW reductions.

---

[3] Tacoma Power canceled multiple events in December 2022 due to the holiday season and not wanting customers to feel effects of water heater events.

Large snapbacks were also observed, with the first post-event hour showing an average increase of approximately 0.61 kW per customer. This is clearly illustrated in the graph above, where the green bars representing "Post Event Hour 1" are positive across nearly all months, indicating a rebound in demand immediately after water heaters resume operation. The snapback in this testing strategy was concentrated in the first hour following an event. The red bars highlighting "Post Event Hour 2" show minimal impacts to usage two hours after an event took place.

Second Year of Pilot Testing – Varied Strategies

After observing consistent impacts in the first year of the pilot, Tacoma Power began varying the dispatch strategy to optimize kW reductions during events and reduce snapback following events. Several strategies were implemented, as shown in Figure 4 below.

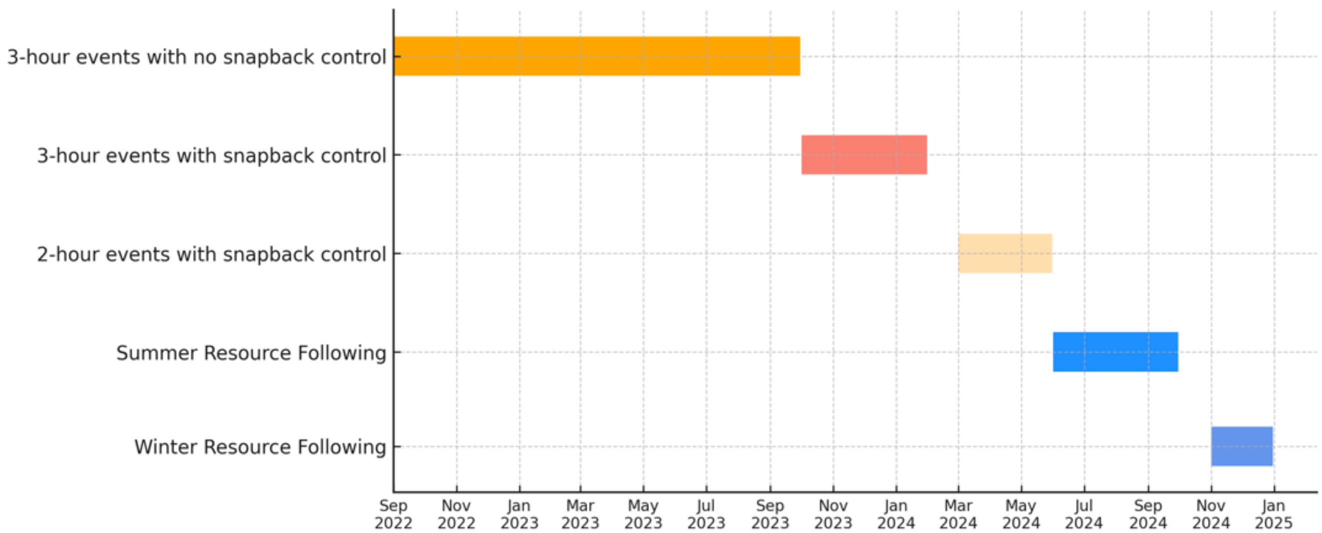

Figure 4. Timeline of Testing Strategies[4]

From October 2023 to January 2024, snapback control was tested by staggering the restoration of water heater operation. This approach was designed to mitigate the sharp load spike typically seen when all units turn on simultaneously. The goal was instead to distribute the snapback over time. After noticing that the third event hour often had the lowest kW reductions, from March to May 2024, event durations were shortened to two hours. These shorter events continued to incorporate snapback control. Throughout the seven-month period of testing snapback controls, a range of snapback delays, from 15 to 60 minutes, were implemented.

Finally, the pilot tested a new strategy, known as Resource Following, which used the implementer's algorithm to determine the timing of events based on wholesale market and capacity prices provided by Tacoma Power. Resource Following events could occur at any time of the day and multiple times a day. The duration of events varied by month:

- June–July: Only 1-hour events scheduled
- August–September: 1- and 2-hour events scheduled
- November–December: 1- and 2-hour events scheduled

[4] Testing strategy from October 2024 omitted from this paper.

One week each month had no events to allow for a baseline against which the evaluation could be compared and a baseline for the implementer's algorithm to understand treatment group usage when events were not running. During an event hour, rather than the entire treatment group receiving an event, individual treatment customers were assigned to be on or off based on the algorithm. This led to some hours having a portion of treatment customers receiving events, and other treatment customers experiencing snapback from an event in the hour prior.

Due to the change in how the events were implemented in these Resource Following months, the evaluation approach was shifted from a pooled regression to a difference-in-differences methodology. Hourly impacts were assessed by looking at the treatment group as a whole, regardless of the fact that not all customers were receiving events at the same time. Due to the varying number of treatment customers receiving events, reductions were calculated in hours where at least half of the customers (40) experienced an active event. Given that fewer customers were receiving events during event hours, a lower average impact was expected, but this evaluation methodology was based on the assumption that the benefit to this staggering of events was important to measure.

During the evaluation of this pilot, differences in the treatment and control groups were studied. Over the duration of pilot, the treatment and control groups slowly decreased in size, from about 120 customers each to around 80 customers each. In January 2024, customers were given the option to opt out of the pilot, which saw a decrease in participation of about 33%. Due to issues with controller failures earlier in the pilot, Tacoma Power and Evaluation decided it would not be beneficial to reassign treatment and control groups after this opt-out period.

It was identified that there were gradual deviations in these groups as customers naturally changed their general usage patterns over time (shown from September 2022 to January 2024 in Figure 5). The gray line indicates treatment usage during non-event hours, and the orange line represents control usage during non-event hours. The difference in the two groups was ultimately mitigated in the modeling and the usage of the difference-in-differences methodology.

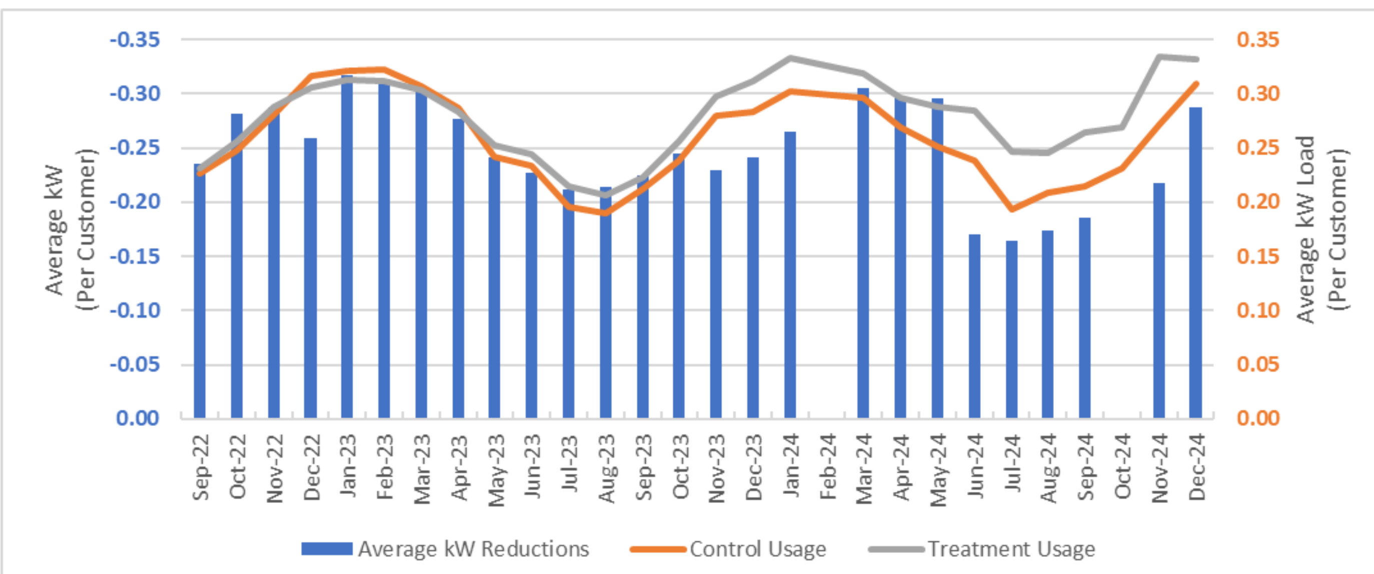

Figure 5. Monthly Impacts in Comparison to Usage During Non-Event Hours

As expected from these controllers, the impacts of the pilot often aligned closely with average usage. The blue bars in Figure 5 demonstrate the kW reductions of each month of the pilot. During the first year of the pilot, September 2022 to September 2023, it can be seen that the impacts almost perfectly line up with the treatment usage (gray line). This demonstrates that the effect of these events is limited by the magnitude of the usage of the water heaters. As testing strategies changed in October 2023, there was some deviation from the baseline treatment usage. These deviations were especially apparent when Resource Following was being implemented from June 2024 to December 2024.

Results

Table 1 shows an overview of the variation in impacts based on the testing strategies. The highest per-customer impacts occurred during the 2-hour events with snapback control (-0.30 kW). These were higher in comparison to the baseline 3-hour events that were evaluated in the first year of the pilot (-0.26 kW). The lowest per-customer impacts occurred during summer Resource Following (-0.17 kW). Impacts in Resource Following months were likely lower due to events occurring during hours with lower usage.

Table 1. Summary Results of Testing Strategies[5]

Standard Event Schedule Strategy	Hourly kW Reduction (Per Customer)	First Hour kW Snapback (Per Customer)	# of Months	# of Events
3-hour events with no snapback control (2023)	-0.26	0.61	13	236
3-hour events with snapback control (Oct 2023-Jan 2024)	-0.24	0.40	4	77
2-hour events with snapback control (Mar 2024-May 2024)	-0.30	0.30	3	66
Summer Resource Following (June 2024-Sept 2024)	-0.17	--*	4	1,084
Winter Resource Following* (Nov 2024-Dec 2024)	-0.25	--*	2	120

The months in which Resource Following was tested had a significantly higher number of events than previous months. In each month of summer of 2024, an average of 271 events were run. This contrasts with the 236 events run over the entire first year of the pilot. Figure 6 shows a more detailed, monthly look at the impacts of these strategies.

[5] Snapback was not calculated during Resource Following testing due to variation in timing of snapback with smaller sample sizes.

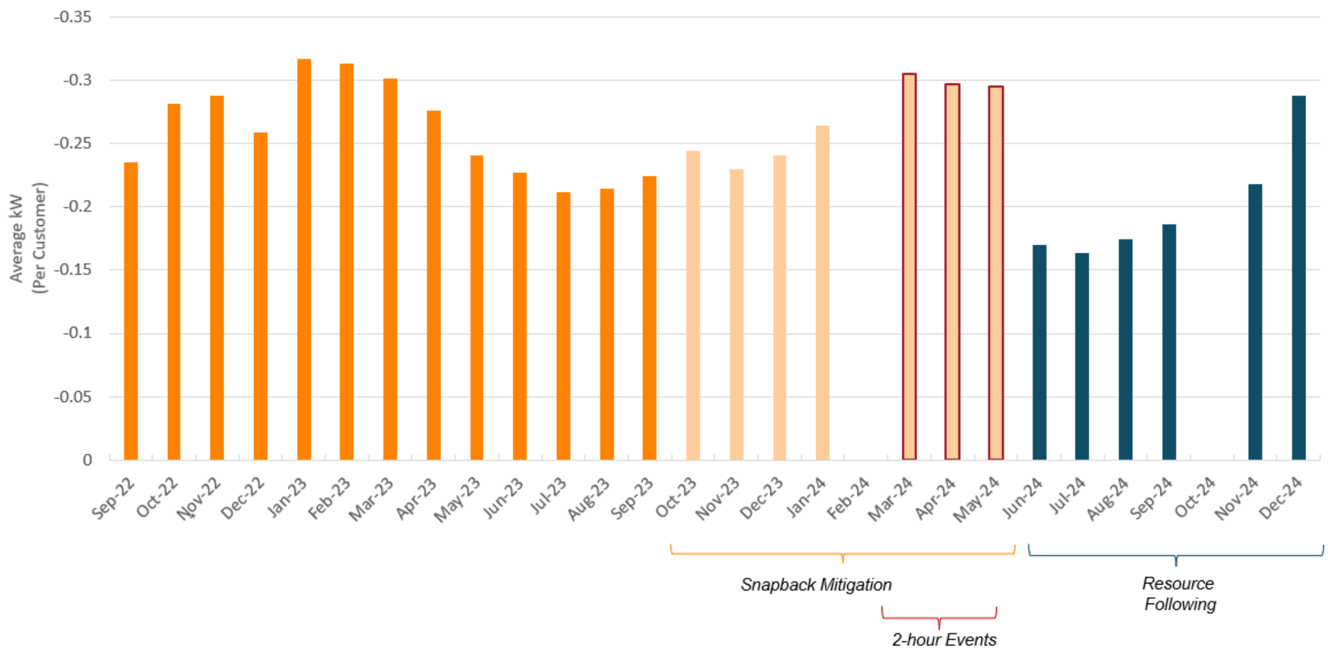

Figure 6. Monthly Average kW Impacts for Full Pilot

Figure 6 depicts the monthly impacts for the entirety of the WHDR pilot. The dark orange bars (September 2022 to September 2023) show the first year's results previously seen in Figure 2. The lighter orange bars with no border (October 2023 to January 2024) show the impacts of 3-hour events with snapback control strategies. These impacts were slightly lower than the impacts seen in 2022. There is still an evident trend of increased impacts as weather gets colder. The lighter orange bars with a red outline (March 2024 to May 2024) show 2-hour events with snapback control. These impacts were higher than March 2023 to May 2023, which can be attributed to the shorter duration events concentrating on the first two hours with the highest reductions. Finally, the dark blue bars (June 2024 to December 2024) depict the impacts from Resource Following testing. The summer months of Resource Following show impacts slightly lower than those seen in 2023. The winter months of Resource Following show impacts beginning to increase. The details of these Resource Following events will be explored further later in this paper.

As stated previously, the highest impacts came during 2-hour events in March 2024 through May 2024. The first two hours of the events often had the largest impacts as seen in Figure 2, so these shorter events appeared to concentrate the savings successfully. This trend can also be seen in the loadshapes, in that there was slightly higher load to be reduced in these months than in the previous year.

Figure 7 below shows the March 2023 through May 2023 events with a 3-hour duration and no snapback control versus March 2024 through May 2024, which had 2-hour events and snapback control strategies deployed. The 2024 months show a clear shift of snapback into Post Event Hour 2 (red bars). This allowed for a reduction in the impact of snapback in Post Event Hour 1 (green bars). In 2023, the Post Event Hour 1 Snapback was 0.61 kW. This reduced to 0.30 kW in 2024.

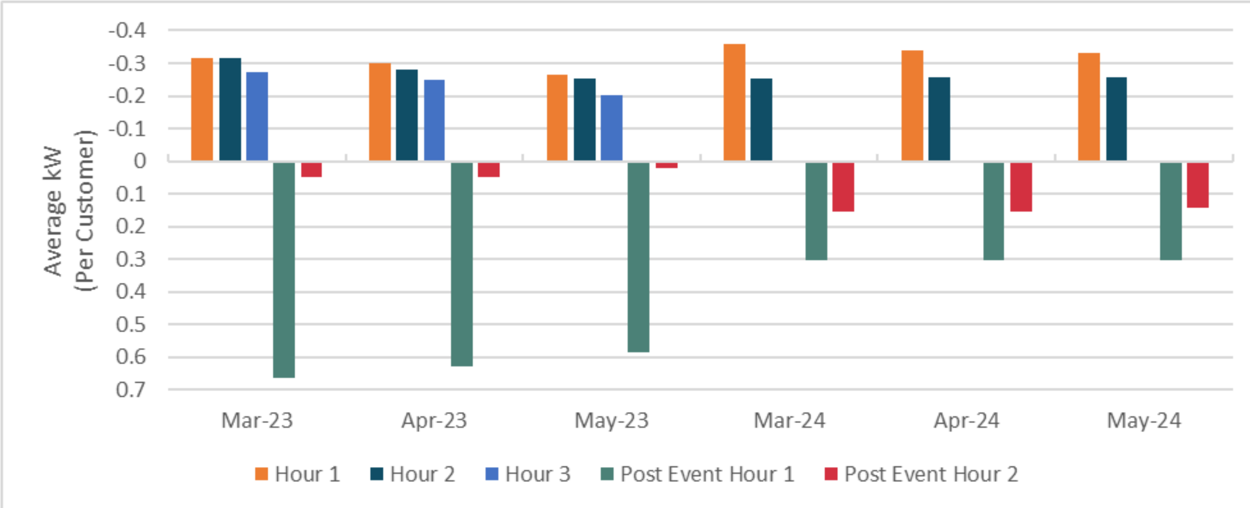

Figure 7. 3-hour Event Impacts in 2023 Compared to 2-hour Event Impacts in 2023

Of the six months that Resource Following was tested, September had the best impacts. In Figure 8 the average hourly impacts are shown layered on top of the hourly price forecast. As seen below, the price forecast used in September (blue shaded area) had a bimodal peak in the evening hours. The algorithm attempted to align events with high price forecasts, and snapback with low price forecasts. There are clear events from 5-6 PM (Hour Ending 18) and from 7-8 PM that line up with the high price forecasts at those hours. The most successful event hour was 7-8 PM with a kW impact of around -0.37 kW. It should be noted that there were events run during hours where usage was not high, for example, at midnight. Events during these hours have lower impacts as there is less load to curtail. Resource Following could be an effective strategy with the right methodology of determining event timing.

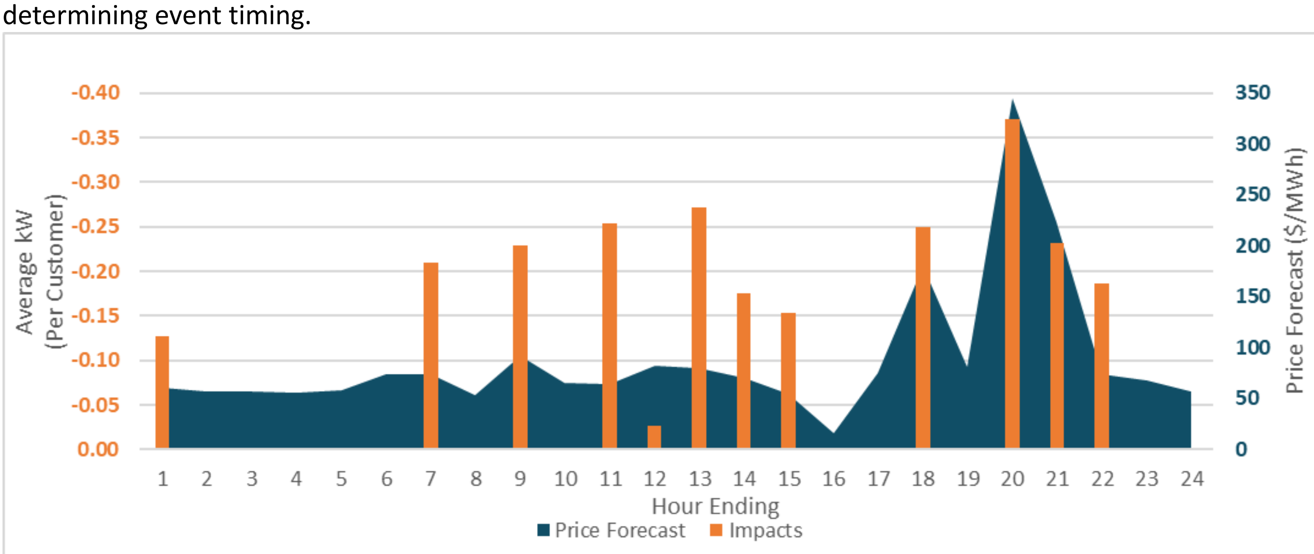

Figure 8. Summer Resource Following Impacts vs. Price Forecast in September

Testing Cost Savings with Resource Following

Tacoma Power was also interested in how much cost savings they could achieve with a WHDR pilot in their territory. During the Resource Following testing, Evaluation looked at per-customer utility cost savings based on the price forecasts provided. An example day is shown in Figure 9 below. Hourly avoided costs were calculated in hours that had at least half of the treatment customers experiencing events or experiencing snapback, as shown below.

*Hourly Cost Savings = Hourly kW Impact [difference in differences of hourly treatment and control usage] * Hourly Price Forecast*

Hours with events had decreases in customer load, which led to positive cost savings. Hours with snapback had increases in load, which led to negative cost savings. These hourly costs were summed up to equal a daily cost savings. Daily cost savings were found to be very low. Figure 9 shows a more successful day in September, and in total the daily cost savings were only $0.32 per customer.

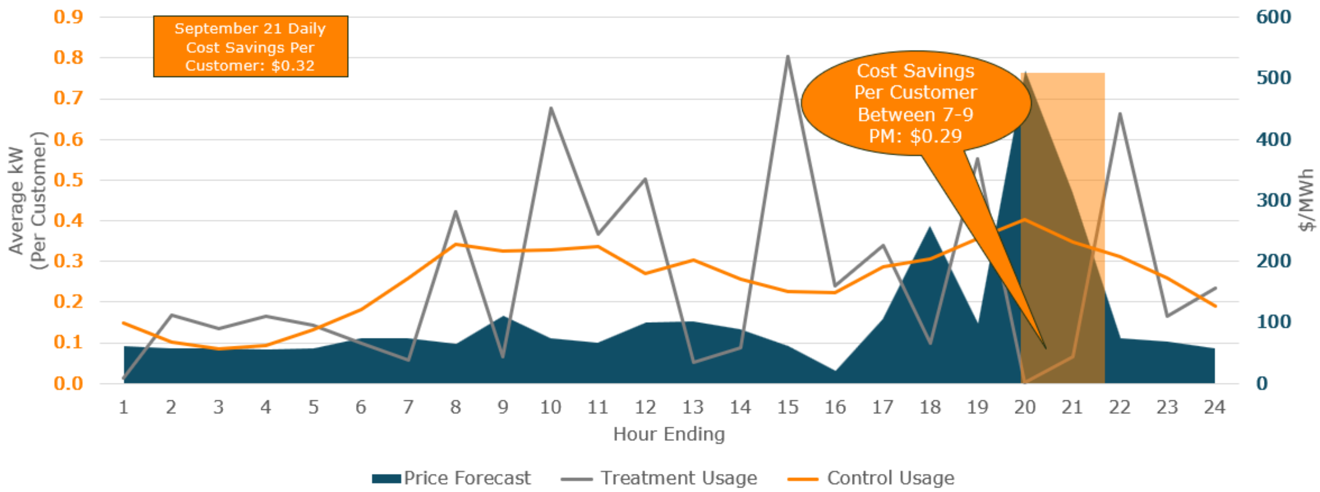

Figure 9: Cost Savings on September 21, 2024

Through looking into these potential cost savings, it was determined that most of the savings were driven by evening events with high price spikes as highlighted in the figure above. September had high evening price forecasts, and the 7-9 PM event led to $0.29 in avoided costs per customer. Based on the algorithm's scheduling of events, this testing was not able to create impactful cost reductions for Tacoma Power as the total cost savings were around $5 per customer per month. In the case of this pilot, the cost savings were highly dependent on the price forecasts and the logic behind the algorithm used to set these event times. Regardless, the events were successful in reducing load, and this was an innovative strategy that has room to be refined in the future.

Customer Satisfaction

Tacoma Power was interested in how customers might feel the effects of this pilot. They sent out a series of surveys throughout the pilot to test customers' satisfaction. These surveys were sent to both treatment and control customers in order to determine if there was a significant difference between the groups.

Participant satisfaction remained high throughout the pilot and did not erode significantly over time. Figure 10 demonstrates the change in customer satisfaction in 2022 at the beginning of the pilot (blue bars) and at the end of the pilot in 2024 (orange bars). There was very little deviation in answers across the years, likely reflecting the unobtrusive nature of the pilot. Only three participants, all of whom were in the control group, contacted Tacoma Power about cold water issues. From the survey's results, 94% of participants understood the purpose and goals of the pilot, and 87% said they would recommend an expanded version to friends or family. Figure 10 shows the results of the customer satisfaction surveys in 2022 and 2024.

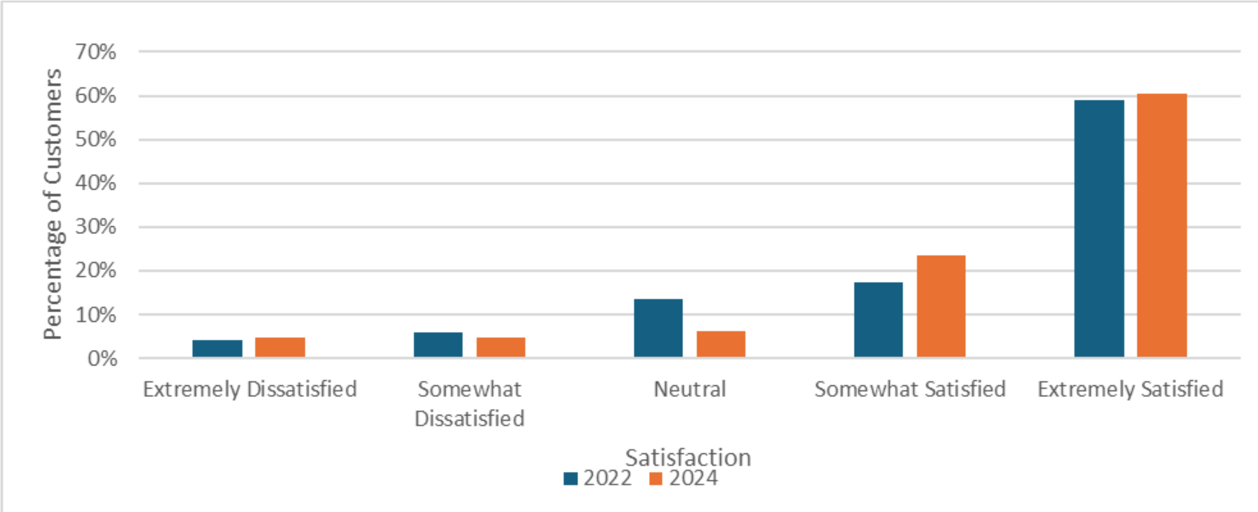

Figure 10. Customer Satisfaction at Beginning and End of Pilot

Successes

From 2022 to 2024, Tacoma Power's WHDR pilot explored a range of innovative strategies for managing water heater loads. Over the 2.5-year period, hundreds of DR events were successfully executed with minimal customer complaints, suggesting that the events were generally non-disruptive to daily routines. The pilot was supported by a robust evaluation framework using an RCT, which led to finding relatively consistent savings across the pilot. A key strength of the pilot was the real-time communication between Tacoma Power and Evaluation, which allowed for results to be quickly understood by Tacoma Power and issues in data to be fixed in real time.

Challenges

At the start of the pilot there were several initial challenges that included early event scheduling issues, frequent hardware failures, and complications arising from changing technologies and data platforms. These challenges were to be expected from a new pilot testing innovative strategies. In 2024, some of these challenges continued, with scheduling conflicts and errors in flagging events as frequency of events increased. Additionally, in order to test more strategies, each variation of testing only occurred for a few months. This challenged Evaluation by having fewer months against which to compare results. Finally, changes in reporting outputs further introduced data quality issues, and at times indirect communication with the implementer resulted in a lack of clarity regarding the details of the testing strategies.

One of the hardware issues from this pilot were relay failures of the controllers, requiring in-field replacement that impacted over 20% of the treatment group (the control group was unaffected). These failures made it impossible for the water heater elements to turn on, effectively disabling participants' water heaters. A solution that was found was that failures were often, if not always, proceeded by a detectable voltage spike. An algorithm was created to check for these voltage spikes daily and participants were contacted immediately, allowing someone to be dispatched to remove the controller. This quick response helped avoid participant discomfort issues. Removing the controller allowed the water heater to work as expected.

Conclusion

Overall, this pilot successfully tested a wide range of WHDR strategies and resulted in consistent impacts across a variety of events. Customers appeared unaffected by the events scheduled, implying that a WHDR program should not affect their daily routines. The most successful event strategies included shorter duration events that occurred during peak customer usage with the addition of snapback mitigation strategies. It was ultimately found that the impacts of a pilot like this are limited by the customers' usage.

One of the biggest challenges surrounded the failures of the controllers themselves. These controllers were expensive to install and had frequent failures. Due to these hardware issues, a large scale WHDR program would be difficult to continue at this time. Despite the hardware challenges, it was determined that water heaters can be used effectively for DR with minimal disturbance from the customers.