

Turning Up the Heat (Before Peak): Lessons from a Heat Pump Water Heating Load-Shifting Pilot

Jenny Hughes, Stephan Barsun, Sara Rosenbrock, Amelie Besson and Helen Liu

ABSTRACT

Decarbonizing and decentralizing the energy landscape requires innovative approaches to meet climate goals. The WatterSaver program, launched in 2022, aims to shift energy usage away from peak demand and emissions periods, improving grid reliability and reducing costs for Time-of-Use customers. The program employs two load control strategies and staggered shed periods:

- Basic Load Up: maintains the customer's set point.
- Advanced Load Up: preheats water above the set point to store energy.

Shed periods run from 4-9 PM and 3 PM-12 AM to align with peak Time-of-Use pricing. These strategies aim to increase energy storage during periods of peak renewable production and reduce consumption during peak times, thereby contributing to load flexibility—a key component in integrating Distributed Energy Resources.

Program performance varied due to inconsistent command execution that limited effectiveness. We used hourly panel data regression models with participant fixed effects to evaluate the program's impact. This method combined time series and cross-sectional data to model household water heater usage. Additionally, we evaluated the Randomized Control Trial that defined the baseline.

Our analysis found that Advanced Load Up participants achieved greater load reduction during peak hours than Basic Load Up participants, albeit at the expense of higher overall energy consumption and customer bills. This suggests a need to refine Advanced Load Up algorithms to balance preheating and energy efficiency. Extended shed periods showed higher opt-out rates, likely due to hot water depletion, suggesting a need for improved customer engagement and control strategies to reduce opt-outs.

Background and Introduction

Launched in 2022, with the first water heater enrolled on March 20th, 2022, the WatterSaver Pilot Program manages smart electric water heaters and connectivity devices with the goals of shifting energy usage outside peak demand periods, enhancing grid reliability, and reducing costs for PG&E's Time-of-Use (TOU) customers. Other programs across the United States operate similarly.¹² PG&E has contracted with the Association for Energy Affordability (AEA) to conduct the implementation of the WatterSaver Pilot. Virtual Peaker (VP) is the third party Distributed Energy Resource Management System (DERMS) provider that actively controls enrolled participants. These controls consist of two types of load shifting commands sent daily to participating water heaters:

- Load Up – the water heater is sent a command to heat the stored water to either:

¹ See for a similar study in the SE United States: [Grid-connected heat pump water heater benefits for low-income households in the Southeastern United States](#)

² This is a similar California program that has not yet published an evaluation: <https://energized.edison.com/stories/save-money-with-smart-water-heater-rewards>

- The customer’s desired temperature set point by reducing the deadband before the shed period told to store more energy either by bringing more of the tank to the set point temperature (Basic Load Up or BLU) or to
 - A temperature higher than the customer’s set point but only for customers with a thermostatic mixing valve (TMV) to minimize scalding potential (Advanced Load Up or ALU).
- Shed—The water heater is sent a command to reduce energy consumption; each manufacturer uses its own logic to determine how this is done, but it usually involves lowering the setpoint, lowering the bottom bound of the deadband that the tank temperature can drift to, and/or using more efficient operating modes. The shed commands are the same for both BLU and ALU. Shed periods usually align with the participant’s peak tariff hours.

This Program is a pilot using nascent and evolving technology, so the Program has continually changed and adapted over time as AEA and PG&E learn more about the technology and its nuances. This report analyzes the WatterSaver Program’s enrollment and its impact on energy consumption, customer bills, and emissions for the year from May 1, 2023, to April 30, 2024. It also evaluates the reliability of water heaters in responding to commands and the effectiveness of the Randomized Control Trial in producing statistically significant results.

Enrollment Analysis

As of April 2024, 243 water heaters were actively participating in the Program, with approximately 95% heat pump water heaters (HPWHs) and 5% electric resistance water heaters (ERWHs). Since program inception, 289 unique water heaters have joined, with 46 participants unenrolling. Enrollment is expected to grow to between 1,200 and 4,100 devices by the end of 2025.

The proportion of solar customers, including customers with Net Energy Metering (NEM) represented in the WatterSaver Program, is higher in WatterSaver (65%) than in the general PG&E customer base (approximately 18%).³ Moreover, the representation of Self Generation Incentive Program (SGIP) HPWH participants is also higher in the WatterSaver compared to the overall PG&E population.

Participants are grouped into platoons based on their preferred load-shedding period (typically aligned with their TOU peak hours), water heater type (HPWH or ERWH), load control strategy (Basic Load Up or Advanced Load Up), and sector (residential vs. small business). Once a platoon reaches 35 participants, Virtual Peaker implements a Randomized Control Trial (RCT). In this RCT, each day 10% of participants are randomly selected to act as a control group and do not receive load shifting commands, establishing a baseline.⁴ This enables accurate comparisons between baseline and shifted usage. For this analysis, three platoons were baselined, as shown in Table 1.

Table 1: Program Participation by Platoon as of April 30, 2024

Platoon Characteristics	N	Percent	Baselined
HPWH ALU 4-9 Shed	89	36.6%	Yes, started in March 2023
HPWH BLU 4-9 Shed	86	35.4%	Yes, started in March 2023
HPWH ALU 3-12 Shed	41	16.9%	Yes, started in April 2024

³ Chhabra, Mohit, Powering Change: Understanding California’s residential Electric Rate Challenge and Affordability Solutions, National Resources Defense Council, March 2025

⁴ 35 was picked to ensure that at least 3 water heaters (35 x 10% with some buffer to account for communication failures) were selected for the baseline each day. Platoons with fewer than 35 people are not sent commands to not load shift.

Platoon Characteristics	N	Percent	Baselined
HPWH BLU 3-12 Shed	15	6.2%	Not Yet
ERWH BLU 4-9 Shed	11	4.5%	Not Yet
HPWH Small Bus	1	0.4%	Not Yet
Total	243		

Command Performance

For the Program to be successful, load shift commands must be reliably sent by Virtual Peaker and successfully received by the water heater. These commands are delivered either through third party devices such as CTA-2045 modules or through an app-based connection via the participant's home Wi-Fi or a built-in cellular module.

Analysis of command success rates reveals a few observable trends:

- **CTA-2045 devices**, which typically connect through **cellular networks**, showed the highest reliability, with 98% of shift commands successfully delivered. Because these devices are independent of the participant's home network, they are not affected by Wi-Fi outages or router issues, contributing to their consistency.
- **App-based devices using Wi-Fi** make up the majority of program participants. These devices demonstrated moderate but consistent success across manufacturers, with success rates across manufacturers varying from 13% to about 82% of successful commands. For manufacturers with lower success rates, we were able to work with their teams to troubleshoot what was causing failures and have since observed higher success rates.
- **App-based devices using cellular** are not represented in our customer sample but are offered through some manufacturers and may be present in the second round of analysis.

Performance can vary depending on the quality and stability of the participant's internet connection. Command control failures are not the only factor in how the water heater performs as customers ultimately are able to opt out of the commands if they do not want their water heater to perform load shifting at that time. Figure 1 presents the reasons for command failure by platoon. The 3PM-12AM shed platoons have a noticeable higher rate of opt-outs when compared to other platoons, especially for devices in the 3PM-12AM basic load up platoon. This may be indicative of customers running out of hot water for these longer shed periods. Additionally, the basic load up 4-9PM shed platoon has a slightly higher opt-out rate than advanced load up 4-9PM shed platoon, which may also be driven by running out of hot water during the shed period.

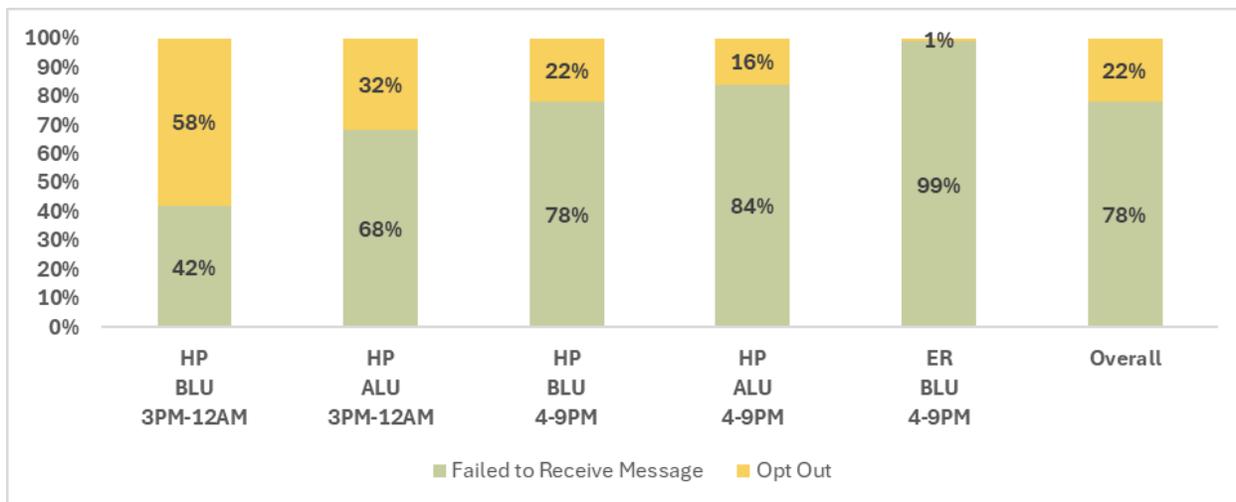


Figure 1: Reason for Failure by Platoon for Residential Devices May 2023 to April 2024

Methodology

This section presents our methodologies to evaluate energy impacts, customer bill impacts, and emissions impacts.

Energy Impacts Methodology

To estimate WatterSaver program ex-post impacts, Verdant used participant information, 15-minute interval telemetry data, and daily event logs provided by PG&E, AEA, and Virtual Peaker. These data were reviewed and quality controlled for analysis.

Impacts were modeled using hourly panel data regression with participant fixed effects. This approach uses data with both time series (datetime) and cross sectional (household) dimensions to model water heater consumption. Devices were grouped by platoon and hour of the day. The models included independent variables for command status (success/fail), weekend indicator, and their interactions. Failed control days were estimated in addition to successful control days because these days are inherently different than baseline days where no controls were sent out. Failed days are defined by the water heater either not responding to a load up or shed command or by the customer opting out of the controls for the day, creating an impact different than both successful command and baseline days. Separate regressions were run for each platoon-hour combination to capture variation in impacts over the course of the day.

Fixed effects were included to control for:

- Participants-level characteristics (e.g. household size, water heating habits, device efficiency)
- Day-level effects (e.g. daily weather variation)

Due to data availability when this analysis was performed, we could not account for seasonality or model by any additional segmentation within the platoon for variables such as climate zone or water heater manufacturer. Additionally, platoons with smaller samples sizes and weekend results in noisier data with broader error bands, especially on weekends when there is less data. In addition to regression modeling, average hourly load shapes were developed by platoon to visualize baseline and load-shifted usage patterns.

The regression model controls for both individual device fixed effects and daily fixed effects. The individual fixed effects account for any unobserved characteristics specific to each device, capturing time invariant constant differences between devices. Examples of characteristics these effects control for include household size, water heating habits, and the inherent efficiency of the water heater. The daily fixed effects account for any variation that are common across water heaters but vary by day, such as daily weather conditions. We include device fixed effects α_i to absorb time-invariant heterogeneity across water heaters and an idiosyncratic error term $\varepsilon_{i,t}$ for random shocks and measurement error; these serve distinct purposes and are both standard in fixed-effects panel models.

Equation 1: Model Specification

$$kWh_{i,t} = \alpha_i + \Gamma_T + \beta_C Success_{i,t} + \beta_F Fail_{i,t} + \beta_W Weekend_t + \beta_{CW}(Success * Weekend)_{i,t} + \beta_{FW}(Fail * Weekend)_{i,t} + \varepsilon_{i,t}$$

Where:

$kWh_{i,t}$ = Actual water heater kWh usage for customer i during date interval t

$Success$ = Indicator variable (0, 1) for command success

$Fail$ = Indicator variable (0, 1) for command fail

$Weekend$ = Indicator variable (0, 1), where 1 is a weekend date

$\beta_C, \beta_F, \beta_W, \beta_{CW}, \beta_{FW}$ = Coefficients estimated by the model, representing the effect of successful events, failed command events, weekend status, successful command and weekend status interacted, and failed command and weekend status interacted, respectively

α_i = Water Heater-specific fixed-effect, capturing all time-invariant characteristics unique to the customer

Γ_T = Daily fixed effect, capturing any variation that is common across water heaters but vary over time

$\varepsilon_{i,t}$ = Idiosyncratic error term, assumed to be normally distributed with mean 0

Model outputs estimate kWh usage relative to the weekday baseline. Weekend impacts are calculated using coefficient combinations (e.g. $\beta_C + \beta_{CW} - \beta_W$ equals the impact of a successful command day on weekends) to isolate the effect of control events on weekends.

Bill Impacts Methodology

To estimate bill cost impacts we needed two pieces of information, 1) whole home usage before and after participation and 2) a way to calculate bill impacts based on rate code. We used the following to estimate or calculate that information:

- To estimate whole home usage before and after participation in the WatterSaver program, Verdant leveraged load shapes developed as part of the NEM 2.0 Lookback Study.⁵ We modeled two illustrative home sizes, a small home whose consumption does not exceed the first usage tier, and a larger home whose usage enters the second (more expensive) tier. These load shapes were used to estimate the baseline whole home usage (or per-participation usage). The energy impacts from the regression model were applied to the estimated baseline whole home usage to create a WatterSaver participant whole home usage (or post-participation usage).⁶

⁵ While 67% of WatterSaver participants have NEM solar, the whole home load shapes for the bill impacts analysis do not reflect the influence of onsite renewable generation.

⁶ Energy impacts used for the bills assume 90% successful commands and 10% RCT baseline to reflect expected command reliability later in the program.

- To calculate bill costs based on rate code we used Verdant’s bill calculator with rates effective in 2024.

Emissions Methodology

This section presents the estimated GHG emission impacts for the WatterSaver Program averaged over one year (May 1, 2023 to April 30, 2024). The marginal grid GHG emissions values used to calculate emission impacts were prepared by WattTime. The data sources and analytic methodology used by WattTime are consistent with the Avoided Cost Calculator and are approved by the CPUC. Assumptions in the Avoided Cost Calculator are updated periodically. Updated assumptions in the 2020 ACC and the 2021 ACC provided motivation for an update to the SGIP GHG Signal calculations. That update resulted in WattTime releasing a new version of the SGIP GHG Signal starting February 1, 2022: Version 2.4.⁷

An emissions impact per hour was calculated using the energy impact presented earlier in this report multiplied by the marginal emissions rate for that hour (kilograms CO₂ / kWh). The WatterSaver program’s goal is to shift energy usage from the peak hours to off peak hours. These peak energy use hours generally line up with GHG emission peaks. Figure 2 displays the marginal emissions in red and the program impact in green for the ALI 4-9PM Shed platoon. It is important to highlight that marginal emissions are still relatively high after the shed period when a snapback typically occurs.

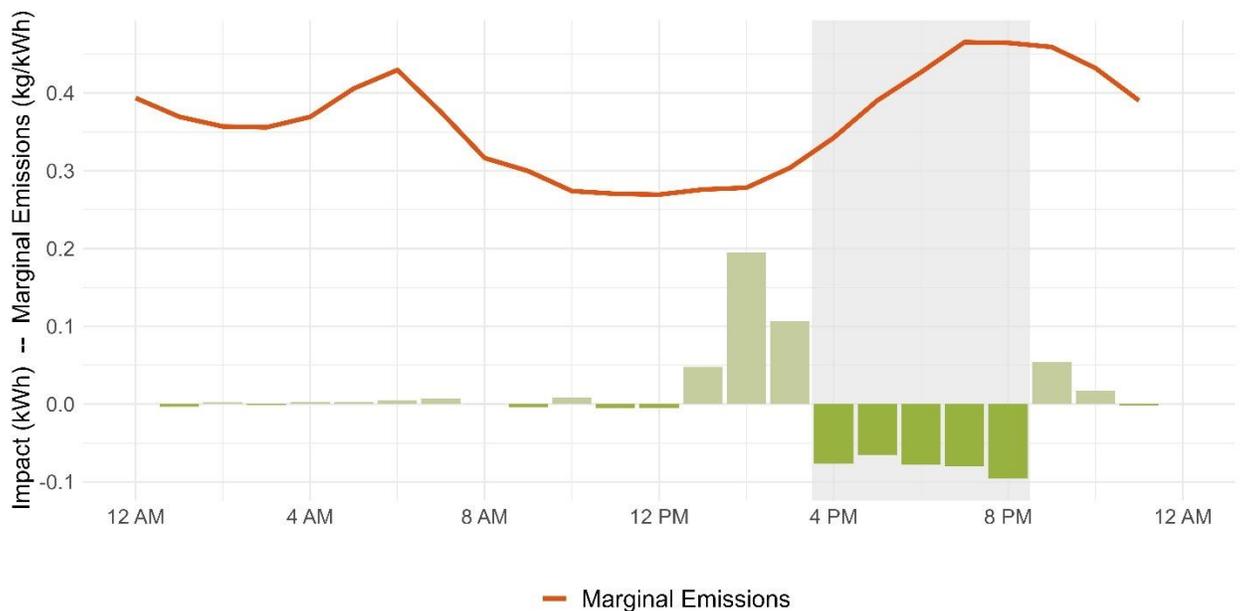


Figure 2: GHG Signal and Energy Impact for Residential HPWH with ALU and 4-9PM Shed

Impacts

Figure 3 displays the average daily kWh usage profile for the ALU 4pm-9pm Shed platoon by weekday/weekend by baseline days, failed command days, and successful command days. These days are defined as:

⁷ <https://content.sgipsignal.com/download-data>. This website provides access to real-time and forecasted marginal Greenhouse Gas (GHG) emissions data for participants in the Self Generation Incentive Program (SGIP). We used ba = “SGIP_CAISO_PGE”, version = 2.0, between May 1, 2023 to April 30, 2024 for this analysis.

- **Baseline** days are defined as the customer water heater purposefully received no commands and operating under non-load shifting circumstances.
- **Successful Command** days vary by platoon and are classified this way when the water heater follows commands to load up and shed that day.
- **Failed Command** days are classified as either missing the load up or shed commands, either due to device communication failures or customer opt outs. These days, some customers still received a successful load up or shed command, but not both. Those partial command days may explain why there is still some evidence of load up or shed on such days.

The three different load shapes (baseline, failed, and successful) vary greatly in the load up period, with peak usage during successful days at 2pm for this platoon. In contrast, the BLU platoon (not pictured) successful days exhibit no significant load up peak and instead is defined by a shed period followed by a very large snap-back peak, in comparison to the ALU platoon.

While load profiles are only illustrative of the changes in use brought on through the WatterSaver program, they are a tool to visualize the impacts prior to running the models defined in the previous section.

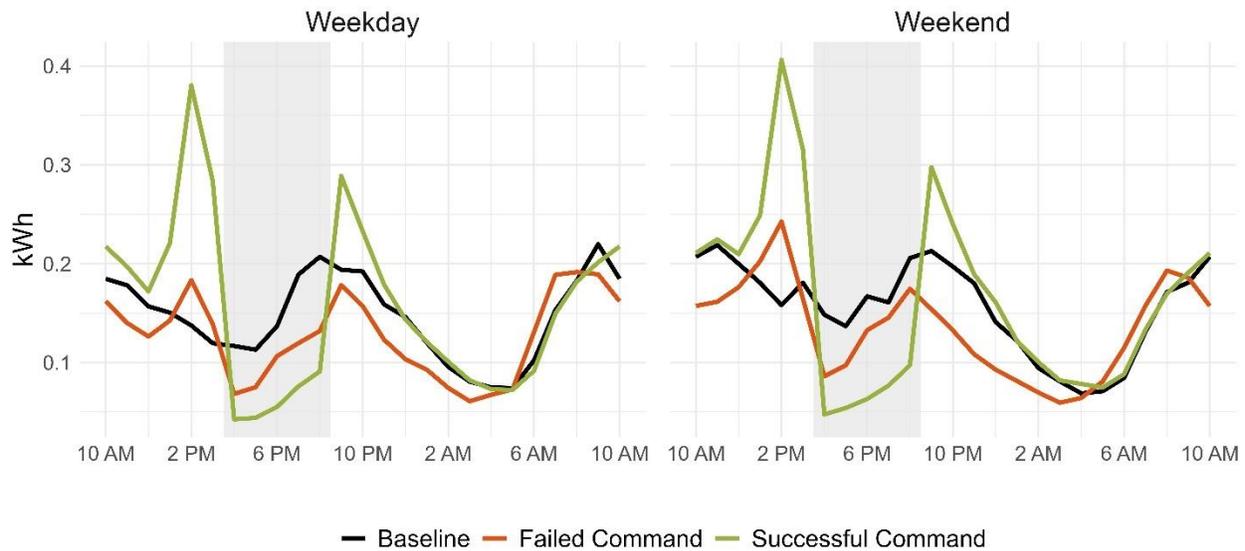


Figure 3: Average Daily kWh Usage for Residential HPWH With ALU and 4-9PM Shed

Figure 4 presents the hourly modeled kW impact of a ALU 4-9PM Shed water heater enrolled in the Wattersaver program. It is important to note that compared to Figure 3 that is an average load shape, Figure 4 displays estimated impacts of load shifting.

- The green line and its confidence bounds (in a lighter shade of green) represent water heaters that responded to all load up and shed commands successfully on the given day.
- The red line and its confidence bounds represent water heaters that did not respond successfully to at least one load up or shed command on the give day (including customers that opted out of load shifting).
- The black line and its confidence bounds represent a baseline day where the load up and shed commands were not sent to the water heater.

The water heater’s baseline usage is presented as the horizontal axis at zero. Deviations from the x-axis (or the black line) represent changes from baseline usage. Impacts below zero indicate reductions in usage relative to the baseline while those above zero are relative increases in energy consumption.

For successful commands, the ALU 4–9 PM platoon exhibits similar reductions in usage during the shed period relative to baseline on both weekdays and weekends. This platoon also shows significant increases in usage during load-up hours compared to baseline, again consistent across weekdays and weekends. During the snapback period, weekday usage rises significantly above baseline, while weekend usage shows some evidence of an increase but is not statistically distinguishable from the baseline. Because weekend usage patterns are more irregular and fewer weekend observations are available, the confidence bounds for weekend estimates are wider than those for weekdays. On failed command days, ALU 4–9 PM customers still display some evidence of load-up, shed, and snapback, suggesting that portions of the commands may have been partially effective.



Figure 4: Modeled Daily kWh Impact for Residential HPWH With ALU and 4-9PM Shed

Table 2 shows the daily weekday and weekend kWh modeled impacts by platoon. The results are the daily kWh impacts accounting for the percentage of devices in baseline, controlled, and failed status over the last year of the analysis period.

Advanced Load Up (ALU) for the 4-9PM platoon has more load reduction during the weekday shed period and moves load to earlier in the day (Load Up Hours) as the ALU HPWHs are storing more energy to prepare ahead of the peak period than BLU HPWHs. Also, for the 4-9PM platoons, BLU has more snapback than ALU as these HPWHs need to make up for the energy lost during the shed period since they did not pre-heat like ALU HPWHs.

Comparing ALU 4-9PM vs. 3PM-12AM platoons also shows some differences. The longer shed platoon has slightly more load up but less reduction in energy during the shed hours. This lower reduction is likely due to three reasons:

- The water heater can’t ‘coast’ as effectively through 9 hours vs. 5 hours without turning on;
- The 3PM-12AM platoon consumes less energy during shed hours than the 4-9PM platoon, resulting in less potential to shed load (possibly due to no winter data when water heating loads tend to be higher);

- The 3PM-12AM ALU platoon has a higher command failure rate with a higher fraction due to opt outs (32% vs 16%) than the 4-9PM platoon, reducing shed impacts. Participants in this platoon are on either E-Elec or EV2a with partial peak periods extending the 4-9PM peak to 3PM-12AM.

Table 2 Modeled Daily kWh Impacts by Control Strategy per Device

Segment	Day Type	kWh Impact During Specific Periods of the Day				Daily Impact 10am-10am	
		Load Hours	Up	Shed Hours	Snapback Hours		Other Hours
HP BLU 4-9PM*	weekday	0.04		-0.20	0.15	-0.03	-0.04
HP BLU 4-9PM*	weekend	-0.02		-0.45	0.19	-0.10	-0.38
HP ALU 4-9PM**	weekday	0.35		-0.38	0.08	0.00	0.05
HP ALU 4-9PM**	weekend	0.34		-0.43	0.06	0.03	0.00
HP ALU 3PM-12AM***	weekday	0.45		-0.28	0.17	-0.01	0.34
HP ALU 3PM-12AM***	weekend	0.36		-0.11	0.07	-0.11	0.21
Weighted Average	all	0.19		-0.32	0.12	-0.03	-0.03

*BLU 4-9PM periods are: 2pm-4pm Load Up, 4pm-9pm Shed, 9pm-10pm Snapback, 10pm-2pm Other

**ALU 4-9PM periods are: 1pm-4pm Load Up, 4pm-9pm Shed, 9pm-10pm Snapback, 10pm-1pm Other

***ALU 3PM-12AM periods are: noon-3pm Load Up, 3pm-12am Shed, 12am-1am Snapback, 1am-noon Other

Table 3 presents the estimated average yearly energy, GHG, and bill cost impacts per device by platoon, as well as a weighted average across the 3 platoons.⁸

Table 3: Impact Results

Segment	Average Yearly Impact per Device					
	kWh Energy	% of Baseline Impact	Annual Energy	GHG Emission kg CO ₂ _equiv	Percent Impact	Bill Cost Impact (\$)
HP BLU 4-9PM	-50.1	-4%		-17.3	-4%	-\$24.81
HP ALU 4-9PM	13.0	+1%		-11.3	-2%	\$13.84
HP ALU 3PM-12AM	110.5	+11%		18.6	+5%	\$49.24
Weighted Average	-11.0	-1%		-11.4	-3%	-\$2.22

Positive (+) values in the table represent increases in the variable. Negative (-) values indicate reductions in the variable.

Randomized Control Trial Baseline Assessment

An important but often overlooked feature of the WatterSaver program is its use of a Randomized Control Trial (RCT) to generate baseline data. The goal is to ensure sufficient baseline days to evaluate program impacts while maintaining statistical balance so that savings remain detectable. The RCT design

⁸ All results presented in Table 3 were weighted by the average participant counts over the year leading up to this report.

activates when a platoon reaches 35 participants, at which point 10% of sites are randomly selected to serve as controls and do not receive shifting commands.

To assess the effectiveness and flexibility of this design, we examined whether relaxing the 35-participant threshold would still yield meaningful results. Our analysis suggests that the current RCT approach generates statistically significant load reductions during shed hours. An exception occurs with ALU participants on weekends in the 3PM-12AM platoon, where only one month of baseline data was available and enrollment was lower. As more baseline data accumulate and participation increases, statistical significance is expected to improve for this subgroup.

Importantly, no systemic bias was detected in the selection of baseline sites, supporting the integrity of the randomization process. Based on power analysis using Cohen's *d*, the 35-participant threshold for baselining appears to be somewhat conservative. Reducing the threshold to 25 participants could still produce statistically valid estimates while enabling more frequent activation of the RCT process.

Conclusions

Our evaluation of the WatterSaver pilot program demonstrates the benefits and some of challenges of using smart load control for residential HPWHs to shift energy away from peak hours. The program's use of both Basic Load Up (BLU) and Advanced Load Up (ALU) strategies allowed for a nuanced comparison of impacts across different configurations. Our overall conclusions include:

Advanced Load Up increased energy shifting but also energy usage: ALU Platoons achieved greater peak load reduction by preheating water ahead of shed periods. However, this came at the cost of increased overall energy use and higher customer bills, especially in the extended 3PM-12AM shed group. BLU devices produced net reductions in energy use, GHG emissions, and customer bills, but experienced post-shed snapback peaks due to lack of preheating.

Different rate structures with a higher differential between peak and off-peak, for example, could reverse the bill increase in the ALU platoon. Also, it is important to note that the annual participation incentive of \$60 (\$5/month) more than offsets the average calculated bill increases for ALU groups. For the 3PM-12AM ALU platoon the increase in kWh energy usage (~11%) also results in higher GHG emissions. However, none of the GHG emissions impacts exceed five percent of the overall GHG emissions driven by the baseline usage of HPWHs.⁹

Opt-outs were higher for customers in the BLU and longer shed platoons, suggesting customer dissatisfaction: Elevated opt-out rates, especially for customers in the 3PM-12AM platoons and among BLU participants, suggest that customer dissatisfaction may be linked to hot water availability, especially during long shed periods. These opt-outs reduce program effectiveness and warrant further investigation into customer comfort and expectations.

Command reliability was higher for customers with cellular-connections compared to WiFi connections: Communication success was highest for CTA-2045 cellular-connected devices, with lower and more variable performance observed for WiFi connected devices. WiFi failures and customer opt-outs together accounted for a significant portion of command failures.

Snapbacks were observed to occur during high-emission hours, leading to increased emissions: ALU platoons with extended shed periods contributed to increased GHG emissions due to elevated consumption and snapback behavior during high-emission evening hours. While BLU devices generally aligned better with emissions reduction goals, they exhibited a strong snapback effect between 9-11PM that still overlaps with high-emission periods.

Delaying the snapback period or tailoring shed durations by user profile may improve emissions outcomes. For example, participants with lower hot water usage during peak hours may tolerate longer

⁹ As calculated based on extrapolating the baseline usage over one year.

sheds. Alternatively, focusing control on the highest-emissions hours later in the shed window may achieve emissions benefits without compromising performance during peak hours.

The randomized control trial appeared to be effective during this pilot: The randomized control trial (RCT) design successfully created a robust and statistically valid baseline. It is important to note how the HPWH performs during the baseline (i.e. is it following another shifting schedule outside of WatterSaver or is it not shifting at all) in determining sample size.

Overall, the WatterSaver pilot shows encouraging results in terms of grid alignment and control strategy experimentation. Further refinements could enhance both customer satisfaction in the program and emissions outcomes.

Recommendations

Based on our evaluation of the program, here we propose several recommendations to improve program outcomes that may be applicable to similar programs:

Strengthen Device Connectivity Monitoring: To ensure reliable performance, the program should continue improving how it monitors and troubleshoots communications issues between water heaters and control systems. Many devices rely on home WiFi, which can be disrupted by changes to passwords or routers. A proactive system to track communication failures would help catch issues early and keep devices responding as intended.

Understand and Reduce Opt-Outs: A significant number of participants are opting out of load shifting, especially those in longer shed periods. This could mean customers are running out of hot water or are unsure about what the program is doing. Gathering feedback through surveys or outreach could help improve customer experience and reduce opt-outs.

Improve Control for Preheating Strategies: Some devices using the ALU strategy used more energy and led to higher bills. The program should fine-tune how much these devices preheat based on typical household use, avoiding unnecessary energy use while still meeting hot water needs. Planned updates to reduce use of the backup heating element and temperature settings are a good step in this direction.

Review the Testing Design as Enrollment Grows: The program's method of testing (using a daily control strategy) is working well overall. When enrollment expands, adjustments may be needed to account for participants in other programs (like SGIP in California that uses time of use load shifting) who may already shift energy use based on pre-programmed controls. These users may need separate evaluation groups or comparisons to get accurate results.

Build a Central Participants Database: As the program grows, it would benefit from a centralized tracking system. This database should include key details about participants and their water heaters (such as enrollment dates, rate plans, device type, and location). This would make it easier to manage the program, respond to issues, and evaluate performance over time.

Limitations

It is important to note that the findings in this paper have limitations due to several factors:

- **Relatively Low Enrollment:** The amount of available data for baseline measurements is affected by enrollment numbers. This had an impact on our ability to observe the effects of seasonal differences and to evaluate every platoon comprehensively, especially for the HPWH 3PM-12AM and the ERHW platoons.

- **Shifting Participants Between Platoons and Changing Platoon Definitions:** The movement of participants between different platoons made it challenging to maintain a stable population over the full analysis period. Additionally, the nascent nature of this technology, the evolving definitions of what constitutes a platoon, and modifications to control strategies have further complicated the evaluation process.
- **Device Communication Issues:** Some manufacturers are experiencing a high rate of communication failure. These failures limit our ability to segment results by manufacturer.
- **Missing Water Heater Usage and Event Log Data:** The absence of crucial data on water heater usage and event logs may indicate underlying device communication issues that are not captured in this analysis. Note that these missing values represent a very small percentage of units (~1%.)
- **Incomplete Tracking Data:** These data were stored in multiple files that were sometimes incomplete or contradictory.

A second round of analysis will be completed at the end of the program that will benefit from likely increases in enrollment and the learning that has occurred during the initial phases of the pilot program, hopefully enabling more robust results and increased segmentation.